Genetically Informed Longevity

Tom Bakos, Marc Klibanow, Nicholas Schork, Ali Torkamani, and Ashley Van Zeeland

Presented at the Living to 100 Symposium

Orlando, Fla.

January 8–10, 2014

Genetically Informed Longevity

AUTHORS:

Tom Bakos, FSA, MAAA: Consulting actuary with more than 45 years' experience in the life and health insurance industry. An expert in insurance industry intellectual property and nontraditional insurance subject matter, he has focused recently on genetic research into the statistical relationships between genetic variation and disease susceptibility and the use of this new information in risk-selection processes. Mr. Bakos coauthored "Gene Knowledge," an article published in the January/February 2001 issue of *Contingencies Magazine*.

Marc Klibanow, CPA: Former chief executive officer and chief financial officer in the biotech industry with experience on Wall Street. Mr. Klibanow provides a unique perspective of both disciplines. He initiated and co-developed an early direct-to-consumer Consumer Laboratory Improvement Amendments safe-harbored genetic test for CCR5. He is also the inventor of a process (patent application pending) involving the curation of a genetic database for the life insurance industry.

Nicholas Schork, Ph.D.: Professor and Director, Human Biology, J. Craig Venter Institute. Formerly, tenured professor at the Scripps Research Institute and director of bioinformatics and biostatistics at the Scripps Translational Science Institute. An expert in statistical genetics, epidemiology and large-scale integrated data analysis, he is also a founding member of the National Institute on Aging's Longevity Consortium, a 10-year running consortium devoted to understanding the influence of genetic factors on the human lifespan.

Ali Torkamani, Ph.D.: Director of genome informatics and drug discovery at the Scripps Translational Science Institute and an expert in computational biology and genetics. He develops and applies computational techniques for the generation, analysis and interpretation of genomic data, with a focus on the discovery and interpretation of the genetic determinants of human disease and the translation of those findings to therapeutic strategies.

Ashley Van Zeeland, Ph.D., MBA: Former director, strategic partnerships at the Scripps Translational Science Institute, current CEO of Cypher Genomics. She provides the critical link between the life science industry and academia and is an expert in commercialization of life science technologies.

In addition to the above-listed affiliations, the authors are co-founders of Genecast Predictive Systems LLC. Genecast's mission is to apply the information contained in genetic variation to life and health risk selection and management processes in order to achieve superior risk assessment, mortality, medical and financial results. As a group, the authors have long-term experience and practice in the fields of actuarial science, biomedical science, biotech, business development, chemistry, finance, molecular and experimental medicine, biostatistics, statistical and computational genomics, risk analysis and risk prediction, the interpretation of DNA sequence variation, neuroscience, neurobiology and genetics of developmental disorders, and the contribution of rare genetic variants to various human diseases.

Contents

Executive Summary

Whatever else it may be, life is a chemical process the end of which, for the most part, is signaled by death. It is autonomous, self-sustaining and self-repairing until, for reasons debated herein, it fails. It has been built by natural selection but a knowledge and understanding of how those chemical processes work may inform unnatural interventions aimed at lengthening each individual's life and keeping it healthy.

Genetic information and its interpretation is rapidly becoming available to individuals directly or through their doctors because it is useful in diagnosing and treating disease and mitigating health and life risk. This inexpensive, readily available genetic information will inevitably inform and influence decision-making processes affecting life, health and longevity.

Genes code for proteins that participate in metabolic processes operating within our bodies and determine how living organisms respond to environmental exposure or change. Individual humans are genetically similar enough to be part of the same species but, nevertheless, differ in important ways such that these genetic differences can inform and be applied to make decisions regarding how specific individuals can maximize their life expectancy and remain healthy. These decisions may involve environmental change or avoidance mechanisms designed to prevent assaults on vulnerable metabolic pathways that affect aging and longevity. These decisions may also be directed toward the most biologically efficient disease treatment or intervention actions designed to prolong healthy life.

Knowledge of how metabolic pathways operate and how they are disrupted by genetic variation will lead to more effective drugs designed to address specific genetic variation. These drugs may do that by mitigating the effect of adverse symptoms created by genetic variation or by influencing alternative metabolic pathways in order to overcome the disruptions caused by adverse genetic variation.

Longevity and longer life spans are supported by genetically informed decisions that defer death.

Introduction

When a significant number of essential biochemical, metabolic and physiologic processes fail, we die. These failures can be caused by bodily injury or disease. It can be argued that no one actually dies of old age, although we get old and we eventually die. Rather, individuals always die of something. Knowing how aging works and how we can avoid bodily injury and disease can inform a health management process aimed at preventing death, promoting healthy life and extending life spans. Given that the essential biochemical, metabolic and physiologic processes which fail and cause death are all under genetic control, there can be no reasonable doubt that genetically controlled and supplied biological processes operating within the trillions of cells in a human body point to inherent genetic variation as a factor underlying inter-individual differences in the rate of aging and longevity.

Senescence is a theory of aging that argues living cells are “programmed” to fail after some number of cell replications or because in some other way evolved physiological and anatomical design is inconsistent with continued survival. This theory applies on the cellular level as well as to an entire organism. If correct, it explains why all living organisms ultimately have finite life spans. There is, however, debate on how such a postulated, programmed function is expressed within living organisms and what influences it. In any event, the arguments for programmed senescence are consistent with the assertion that our life spans and longevity are under genetic control. Senescence would simply be one of the “somethings” causing humans to die.

Aging and longevity, while often thought of as separate processes, are clearly related. The former moves us biologically closer to death while the latter moves the target death further into our futures. Neither one is inherently good or bad. For example, the first step of the aging process is essential for moving humans from the state of a totally dependent infant at birth to an independent, mature adult. Longevity processes serve us and communities best only when they lengthen a functioning, i.e., healthy and productive, adulthood in which knowledge and experience accumulated throughout life can be used to synthesize new opportunities and improve society in ways that will sustain and enhance life.

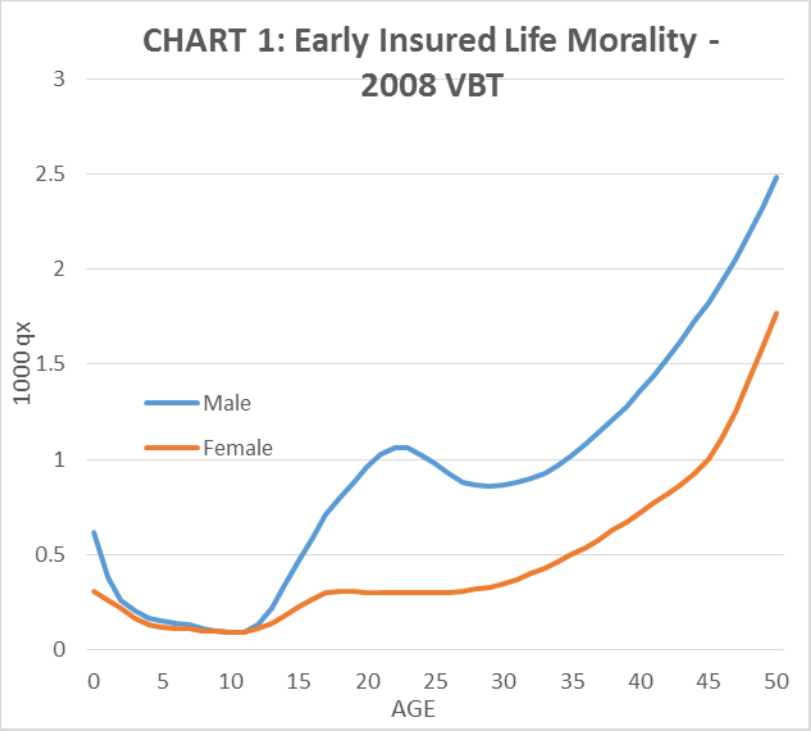
We believe aging and longevity have genetic origins and, importantly, that inter-individual differences in the rate of aging and longevity are attributable, in part, to inherent genetic variations. When coupled with an individual’s environmental exposures and insults, an understanding of how genetic variations influence life-sustaining biochemical, metabolic and physiologic processes can be used to make better-informed decisions intended to improve the odds of anyone living a longer and healthier life.

Aging

Aging, as we define it here, means more than just getting old or putting on years and essentially getting worn out and eventually dying. As actuarial mortality experience studies on humans indicate, aging and longevity can be divided into three stages as reflected in mortality curves. The first stage is essentially associated with “growing up” and early childhood development. The second stage is associated with the more overt aging process as in “getting old.” The third stage includes events associated with advanced old age leading to eventual death.

Growing Up

Charts 1 and 2 show early mortality rates graphed by age. Chart 1 is based on insured life mortality[1] for a newborn at age 0 and shows mortality rate changes through attained age 50.



Males and females both start life with mortality rates that are on average equivalent to the mortality rate of a 20-year-old. As can be seen, mortality rates decrease during infant (age 0–1) and childhood (age 1–5) years. In fact, mortality rates improve during early years of life, reaching their lowest or best level around ages 10–12.

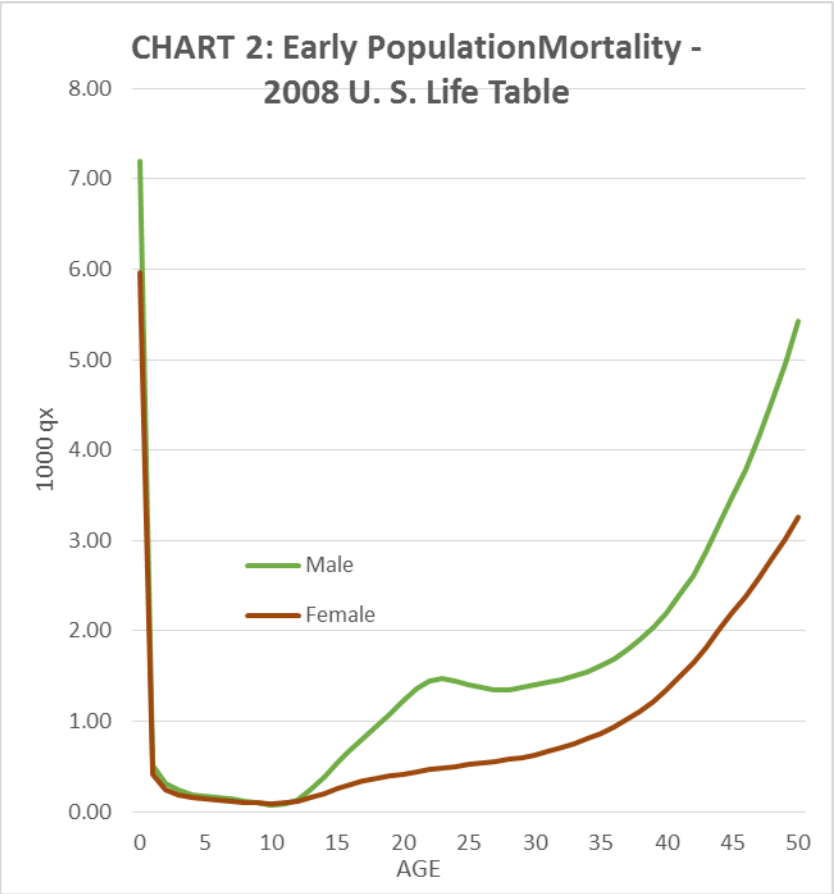
This pattern is evident whether looking at insured life data (as in chart 1) or at population data (as shown in chart 2) that has not been scrubbed by a selection process to remove all but "standard" risks. Although in population mortality data the variation in mortality is more extreme because of the absence of a selection process, the basic, underlying pattern remains.

The population mortality data in chart 2 was taken from the male and female 2008 U.S. Life Tables.[2]

[1] 2008 Valuation Basic Table (VBT) male and female, age nearest birthday (ANB), nonsmoker ultimate mortality for issue age 0. In constructing the 2008 VBT, insured life data for juvenile ages (ages 0–18) was sparse. It was noted in Preferred Valuation Basic Table Team, "2008 Valuation Basic Table," that male insured life mortality was about 78 percent of population mortality. For females, this percentage was 83 percent. Actual mortality experience was used to calculate the mortality rates for attained age 0. For attained ages 1–10, 78 percent and 83 percent (for males and females, respectively) of population mortality was used. From that point, mortality was graded into insured life experience mortality at attained age 25. This mortality table was chosen because it clearly shows the pattern or mortality rates at younger ages as adjusted to reflect the insurance selection effect. Similar patterns can be found in other mortality data.

[2] The mortality tables in Arias, "United States Life Tables, 2008," (table 2 for males and table 3 for females) are on an age last birthday (ALB) basis and are aggregate mortality tables, which show mortality at attained age only. Mortality rates are unaffected by a selection process as in insured life mortality tables.

The U.S. population mortality rates graphed in chart 2 are generally representative of early population mortality levels in socially and economically developed nations where maternal and early child health care is available to eliminate overtly environmental/exposure causes of death related to poor nutrition, poor sanitary conditions or infectious disease.



Among the leading causes of high early (i.e., ages 0–5) mortality rates in the U.S. population[3] are congenital abnormalities, low birth weight, accident (unintentional injury) and, for infants, birth trauma and sudden infant death syndrome (SIDS).

These early causes of death are closely related to genetic variations and are not entirely driven by social or environmental factors. When an insurance selection process is applied to produce an insured life mortality table, as shown in chart 1, medical underwriting operates to select only "standard" mortality risks. This underwriting process eliminates the extremes created by inclusion of obviously unhealthy lives and serves to emphasize an underlying pattern of mortality progression in the early stages of the aging process, i.e., early development or growing up, for which many of the leading causes of death have genetic underpinnings.

The pattern of mortality rates during early development shows a relatively high rate just after birth decreasing to a lowest level at about ages 10–12. One cause of infant deaths, SIDS, is little understood, although researchers are exploring possible genetic links.[4] Looking for a genetic contribution for a condition so little understood is difficult but the clustering of SIDS deaths within families tends to indicate a genetic association. Observation has lead researchers to

[3] From USA Life Expectancy, "USA Causes of Death by Age and Gender," wherein deaths are ranked by age range and gender.

[4] SIDS is the sudden unexpected death of an infant prior to one year of age, typically during sleep. It is thought to be 5 percent explained by genetic inborn errors in metabolism. See Opdal and Rognum, "Sudden Infant Death Syndrome," and Courts and Madea, "Genetics."

conclude that mutations in genes giving rise to metabolic disorders known to threaten the efficiency of biological processes necessary for early development may either cause or be a predisposing factor for SIDS.[5] As researchers learn more through continued research, information on genetic variation related to SIDS and causative effects may inform decisions on how specific genetically identified infants may be more susceptible to SIDS and/or how SIDS deaths may be prevented and, thereby, improve longevity.

Thus, the mortality patterns in early development or the growing-up stage of aging involves, first, surviving to ages 10–12 where the lowest mortality rates are experienced. Generally, male and female mortality rates are nearly identical to age 10–12 but rates for males at birth are slightly higher. After age 10–12, male mortality rates rise to a peak in the mid-20s where they begin a second decline to a rate higher than female mortality at about age 30. After ages 10–12, male mortality will always, on average, be higher than female mortality, a difference reasonably attributable not to the phenotypic differences between males and females but to the underlying genotypic differences and the effects those have on cellular metabolism.

After age 30, male average mortality rates begin a monotonic increase to 1.0 and death. Long-lived individuals can live into their 100s but average mortality rates increase by age because a larger percentage of survivors die with each year of age attained. The significant peak in mortality for ages 20–24 for males is primarily due to accidents, homicide and suicide.[6] Deaths from these causes typically arise from risk-taking behavior, which is more prevalent in males than in females. Studies have shown that about 20 percent of the individual variation in risk-taking can be explained by genetic variation.[7] Clearly, high death rates for males at those ages may also be driven by socio-economic factors. These leading causes of death also apply to females in the age range 15–24, although the absolute level of mortality from those causes is lower than for males and creates only a slight blip in mortality at around age 17 for females as shown in shown in chart 1 and very slightly evident in chart 2.

[5] Variation in a number of genes have been identified as possible contributors to SIDS, including MCAD, KVLQT1, SCN5A, VEGF and IL-6.

[6] See Heron, "Deaths: Leading Causes," table 1.

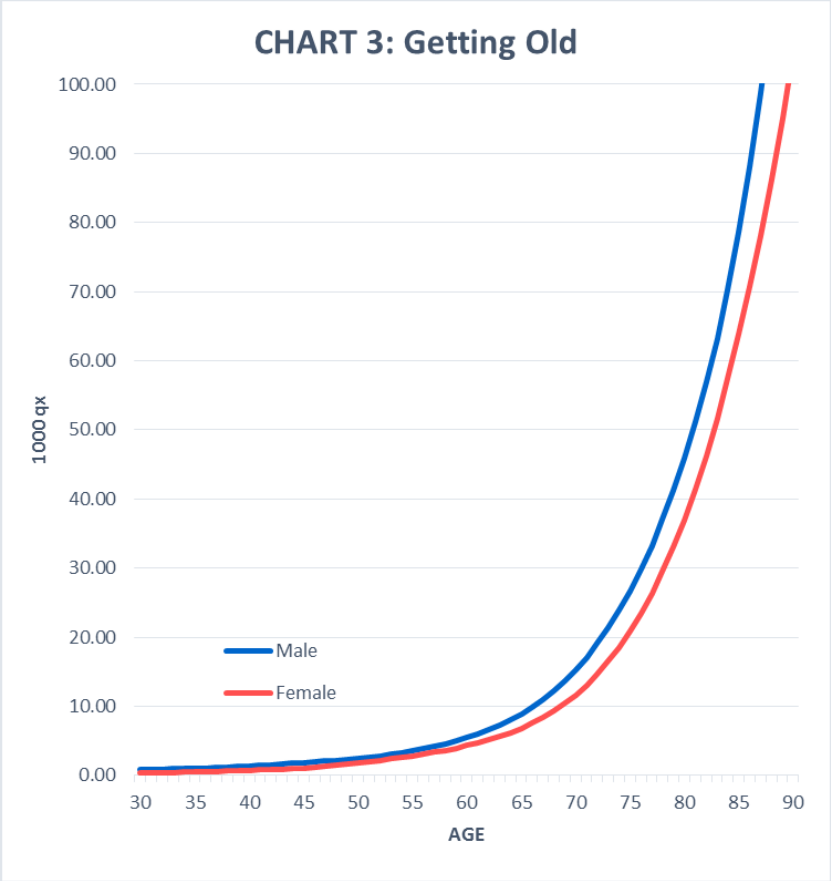
[7] See OMIM, "601969." Risk-taking involves making decisions under conditions of uncertainty and involves often only an implied evaluation weighing potential beneficial vs. adverse outcomes. Indeed, many such decisions made by adults or individuals unaffected by genetic variation appear to be made with no concern for harm avoidance.

Several factors that may increase risk-taking behaviors have been suggested, including, in adolescence, the relative immaturity of neural connections related to behavior regulation, inhibitions and decision-making, and genetic variability in dopamine transmission (related to DRD4 gene).

Risk-taking behavior may be observed in nonlethal situations as well. The SLC6A4 gene (also known as SERT or 5-HTT) regulates the transport of serotonin, a neurotransmitter found in the central nervous system. Because of genetic variation, serotonin levels can be affected, leading to financial risk-taking or other aggressive behavior. A repeat length polymorphism in the promoter of this gene has been shown to affect the rate of serotonin uptake and may play a role in SIDS.

Getting Old

The second stage in the aging process is associated with "getting old," which can be defined as the period from age 30–90. Chart 3 is an extension of chart 1 showing mortality rates between ages 30 and 90 for males and females.[8] During this period, male mortality always exceeds female mortality. There is a gradual increase in mortality rates by age to about age 60. Between ages 60 and 70, the slope of the mortality curve first exceeds 1 and, after that, mortality rates for both males and females increase by year more rapidly. Clearly, beyond age 60 the aging process begins to seriously affect the body's ability to resist disease.



As with the early development/growing-up aging stage, causes of death during the getting-old phase are very informative. The tables[9] below show the top five causes of death for males and females for decennial age ranges within this age 30–90 period.

Note for the getting-old age range of 30–90 the leading causes of death are diseases[10] with notable genetic determinants for both males and females and there is a decreasing influence of

[8] The mortality rates in chart 1 come from the 2008 Valuation Basic Table (VBT) male and female, ANB, nonsmoker ultimate mortality for issue age 0.

[9] See Heron, "Deaths: Leading Causes," table 1.

[10] **Coronary heart disease** (CHD) is the build-up of plaque related to high cholesterol levels in the coronary arteries, resulting in atherosclerosis. This can lead to heart attacks or heart failure as blood flow to the heart is reduced, resulting in death.

Cerebrovascular disease (CVD) is a disease of the blood vessels supplying blood to the brain. This can be caused by embolisms (blockages), aneurisms (stretched or swelled blood vessels that may rupture) or hypertension (high blood pressure), resulting in stroke and death. Atherosclerosis may also cause CVD.

Respiratory, or **chronic lower respiratory disease**, includes emphysema, chronic bronchitis, asthma and other chronic inflammatory diseases of the airways. Tobacco smoke is a principal irritant causing respiratory disease.

accident and suicide among the leading causes of death as individuals age. This is not because the rate of death from accident has gone down per se but because the rate of death attributable to diseases has significantly increased, supplanting accident among the top five causes of death.

	Table 1A. Males						
	Age	35–44	45–54	55–64	65–74	75–84	85+
Rank	1	Accident	CHD	Cancer	Cancer	Cancer	CHD
	2	CHD	Cancer	CHD	CHD	CHD	Cancer
	3	Cancer	Accident	Accident	Respiratory	Respiratory	CVD
	4	Suicide	Suicide	Respiratory	CVD	CVD	Respiratory
	5	Homicide	Liver disease	Diabetes	Diabetes	Diabetes	Alzheimer's

	Table 1B. Females						
	Age	35–44	45–54	55–64	65–74	75–84	85+
Rank	1	Cancer	Cancer	Cancer	Cancer	CHD	CHD
	2	Accident	CHD	CHD	CHD	Cancer	Cancer
	3	CHD	Accident	Respiratory	Respiratory	Respiratory	CVD
	4	Suicide	CVD	CVD	CVD	CVD	Alzheimer's
	5	CVD	Liver disease	Diabetes	Diabetes	Alzheimer's	Respiratory

Dying and Longevity

Human beings, of course, are not immortal and the ultimate result of aging is death. Humans are, however, long-lived relative to most other mammals. In fact, humans have life spans often extending well past age 100. Recent mortality tables reflect longer life spans resulting from improvements in the environment and medical advances by extending the terminal age to 115 or greater.

Clearly, an important element in the survival of a species is having members of that species survive at least to adulthood and sexual maturity for reproductive purposes. The genes we have inherited over the evolutionary process through natural selection that control and fuel life

---

Liver disease refers, in general, to liver dysfunction. The liver is the largest organ in the body and is necessary to filter harmful substances from the blood and serves an important function in food digestion and energy production.

Diabetes is a group of metabolic diseases defined by high blood sugar as a result of either a lack of insulin produced by the pancreas (Type 1) or an inability to use the insulin produced within cellular chemical pathways (Type 2). Diabetes can result in death if not treated.

Cancer, or **malignant neoplasm**, is a broad group of diseases that involve uncontrolled cell growth which adversely affects the functioning of organs and, as a result, leads to death.

Alzheimer's disease is a form of dementia or inability of the brain to function normally, resulting in memory loss and loss of thinking abilities. It worsens with age.

through sexual maturity are complemented by genes selected for contributing to parenting. Human infants, in particular, and children, in general, could not survive early life without parental oversight. Therefore, natural selection has provided humans with natural lifetimes that on average extend well beyond sexual maturity. The grandmother theory of longevity maintains that post-menses longevity in grandmothers allows them to devote resources to the survival of their grandchildren.[11] Some data indicates two additional offspring per additional decade. This provides some evolutionary pressure for the promotion of longevity in populations.

Evidence that there is an inherited or genetic component in aging is found by looking at interspecies lifespan and recognizing that genes involved with growth, development and cellular repair are generally highly conserved up the evolutionary chain. Nature invented them once and used them again and again across various species. An old mouse, dog or human may all have the same telltale age-related signs but, a mouse is old at 2, a dog at 15 and a human at 80. Humans take a relatively long time to reach adulthood and also have a relatively long life. In this light, the evolved design that assures survival at least to reproductive ages and through parenting also has a carryover effect.

Antagonistic pleiotropy is a theory that proposes a single gene may control for more than one trait—one that is beneficial and at least one that is detrimental. To the extent that this effect exists within the genetic structure of organisms, natural selection alone can never lead to "perfection." For example, there is research[12] showing that genes associated with sexual reproduction, an obviously beneficial function for species survival, may also be detrimental to long life. This tradeoff cannot be avoided, hence the belief that senescence or old age is built or "programmed" into our genetic structure. Since the post-reproductive disadvantage does not bear upon reproductive fitness, it is not eliminated by natural selection.

There are other growth and development metabolic pathways that could be examples of this. For example, mTOR, an abbreviation for mechanistic (or mammalian) target of rapamycin, is the name given to a gene, a protein and a metabolic pathway. Its name is derived from rapamycin, an antifungal drug discovered on and named after the island of Rapa Nui[13] in the 1970s. Rapamycin was also discovered to be an immune system suppressant useful in organ transplantation. This usefulness spurred research into discovering how rapamycin worked. Typically, such drugs work by binding with receptors either on the cell membrane or in cell cytoplasm. By binding with a receptor location, the drug prevents another chemical molecule from binding there and can, therefore, affect the chemical processes occurring within a cell. The mTOR protein was determined to be the target of rapamycin. The mTOR protein is coded

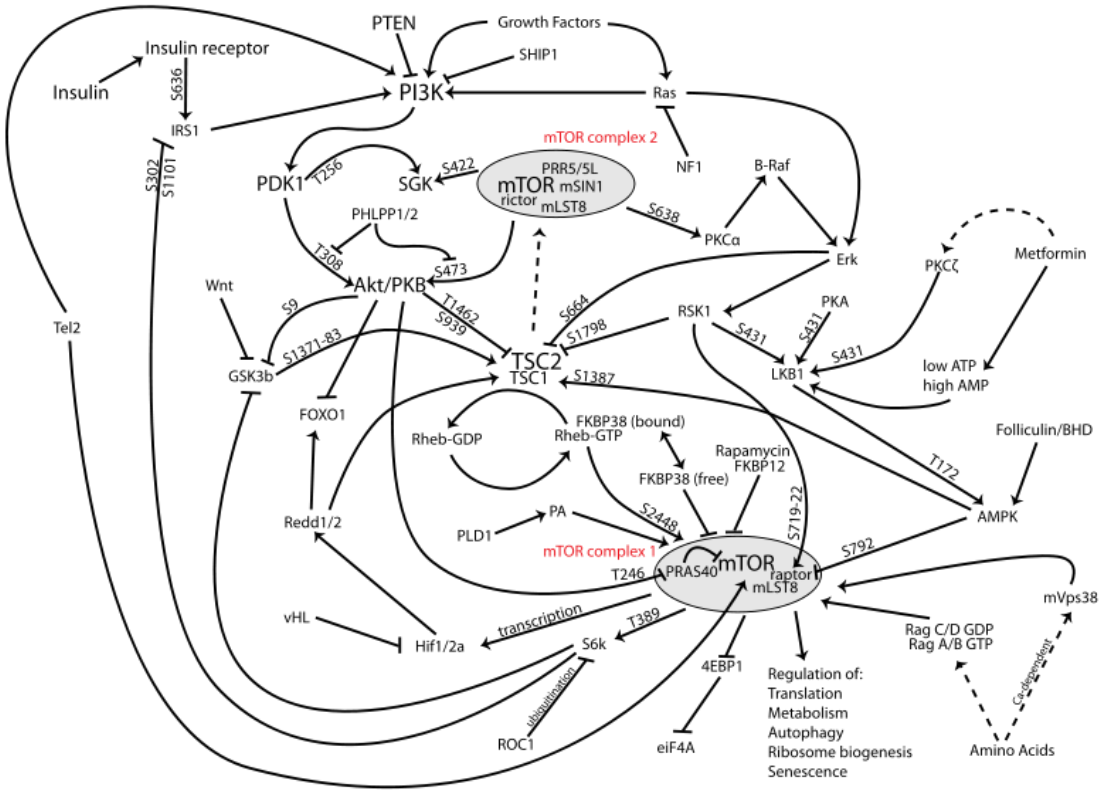
[11] See Lahdenpera et al., "Fitness Benefits."

[12] See McCormick et al., "New Genes." This paper shows the removal of germline stem cells in *Caenorhabditis elegans* (roundworms) and *Drosophila* (fruit flies) increases life span.

[13] Rapu Nui is also known as Easter Island.

for by the mTOR[14] gene. The mTOR protein is related to a family of proteins that function as a master regulator of cellular growth in response to nutrient or hormonal stimulants.

The following is a diagram[15] of the mTOR signaling or metabolic pathway.



The drug rapamycin, by interfering with the mTOR protein receptors that regulate cell growth, can improve longevity by inhibiting chemical processes which might otherwise have occurred mimicking dietary restrictions that, in the absence of malnutrition, are known to extend lifespan in many species. In addition, it is known that reduced insulin/IGF-1-like signaling (IIS) can increase longevity. The mTOR protein receptor is linked to IIS in multiple ways.[16] This signaling while active and beneficial in childhood is thought to be detrimental in adulthood, making it another example of antagonistic pleiotropy supporting the senescence aging theory. The drug rapamycin has been shown to inhibit the mTOR protein signaling pathways and, therefore, have a positive effect on longevity but not without adverse side effects.

An argument for programmed senescence being under genetic control and effectively driven, at least in part, by genetic expression includes the idea it is the developmental processes

[14] The mTOR gene has also been called FRAP, FRAP1 and FRAP2. Clearly, "mTOR" became the name assigned to the protein, the gene that codes for the protein and the metabolic pathway because of the association and involvement of rapamycin.

[15] Wikimedia Commons, "File:MTOR-pathway-v1.7.svg."

[16] See Zarse et al., "Impaired Insulin," and Johnson, Rabinovitch, and Kaeberlein, "mTOR is a Key Modulator."

continuing or decelerating that cause aging. Change in gene expression over a lifetime is part of those developmental processes and reveals certain genes switched on and off at certain times, clearly a regulated nonrandom process. The fact that gene expression can be both advantageous at times and disadvantageous at other times leading to aging[17] and ultimately death is another example of antagonistic pleiotropy. It is well understood that metabolic pathways are under genetic control and can be affected by genetic variation.

More evidence of a genetic influence on aging is the synchronicity of aging. An organism through an aging process (i.e., ignoring disease or genetic variation causing specific metabolic failures) is affected by a general deterioration of all of its metabolic processes over time. Aging is not just the result of random failure of metabolic processes. Deterioration of the metabolic pathways act as an internal clock to synchronize and generalize the effects of aging. Some recent studies have also implicated neural signaling and related neural genes in the internal clock process and aging.[18]

A central theme in a number of longevity studies involves the insulin and insulin growth factor pathways, IGF1/IGF2.[19] Comparative studies in the roundworm *C. elegans* have shown that inhibiting the function of the gene DAF2 can multiply the lifespan of *C. elegans* several times. DAF2 has common ancestors with human genes involved in the IGF1 and IGF2 signaling pathways. Inhibiting the function of this pathway has a powerful effect on longevity in both roundworms and, to a lesser extent, humans. This effect can also be triggered by dietary intervention such as caloric restriction or other interventions that affect the level of certain chemicals in the body. These interventions affect the way genes express themselves.[20]

[17] From the abstract of Somel et al., "MicroRNA, mRNA, and Protein Expression": "Changes in gene expression levels determine differentiation of tissues involved in development and are associated with functional decline in aging."

[18] From the abstract of Zhang at al., "Hypothalamic Programming": "Ageing is a result of gradual and overall functional deteriorations across the body; however, it is unknown whether an individual tissue primarily works to mediate the ageing progress and control lifespan."

And from the abstract of Walter et al., "A Genome-Wide Association Study": "Human longevity and healthy aging show moderate heritability (20%–50%)."

[19] Kaplan et al., "A Genome-Wide Association Study."

[20] Epigenetics is the study of gene expression that can be changed through functional modification of the signaling chemicals present in a cell. Even with no change in a gene's nucleotide (DNA) sequence, the gene can function differently, or turn on or off, because of epigenetic mechanisms. Core gene regulation epigenetic mechanisms include histone acetylation and methylation, microRNAs and methylation. For example, histone acetylation and deacetylation affects the reading of the DNA molecule. Histones are proteins around which the long DNA molecule spools, making its code more accessible.

Eventually, our human biologic systems do wear out or turn off[21] and cease to function properly or efficiently, resulting in disease and eventually death. This "wearing out" is most pronounced at advanced ages and exacerbates disease processes that may have started earlier in life.

Prolonging Life

Although modern health care has provided diagnostic methods, tests and tools that discover or identify faults in our biologic systems and provide treatments to correct or nullify the effects of disease or infection once discovered, we, as a species, have hardly overcome death. Modern medicine has contributed to a deferral of deaths that otherwise might have occurred without its intervention and, as a result, has dramatically improved expected human life span and longevity but has not eliminated death.

Despite this dramatic improvement in expected human life span, modern medical techniques are limited as they are based on generalities derived from observation, testing or trials on large groups of patients only after these patients become symptomatic for a specific disease. The modern definition of "disease" is itself problematic, as it reflects a generalized set of conditions wherein some things, but not everything, are common across individuals thought to suffer from the same disease. For example, cancer, characterized generally as unregulated cell growth, comes in over 100 types and even those types have subtypes associated with unique genetic characteristics. This heterogeneity creates enormous problems for the treatment and prevention of diseases like cancer and is one reason it is not possible, currently, to stave off disease and prolong life uniformly and for all individuals. Consider that cancers are thought to arise when normal cells acquire genetic alterations through damage or mutation which affect the normal operation of chemical processes within the cell that control or suppress cell division, DNA repair or apoptosis.[22] Without specific information about the genetic variation in specific

[21] This turning off may be attributed to simple wear and tear but, more specifically, may result from the evolutionary design of the genotype we have inherited that is "programmed" or built to turn off slowly over time. See a discussion of the theory of senescence on page 4.

[22] Apoptosis is a natural process of programmed cell death that generally functions to eliminate cells no longer necessary for normal function. A typical adult loses about 50 billion–70 billion cells per day due to apoptosis.

cancer cells even of the same type, it becomes a trial-and-error process to find drugs that target, in order to suppress, the function of proteins produced by cancer-mutated genes.

Family history can indicate a high probability of inherited diseases or conditions in an individual and provide guidance on what to look for and possible preventive or management processes. However, family history is inexact and unnecessary when advances and reduced costs associated with genetic sequencing provide an opportunity for genetically informed or personalized medicine. Genetics can impact life expectancy in two distinct ways: by dictating one's inherent susceptibility to any one of a number of fatal diseases or by affecting pathways related to systemic, semi-synchronized decay and increased susceptibility to disease associated with the "wearing down" of the human body associated with aging.

Longevity is improved by the avoidance of death—for an individual, as well as for the population as a whole because

Metabolic, Signaling & Regulatory Pathways

Metabolism involves the occurrence of thousands of chemical reactions within each cell of the body. These reactions may result in the breakdown of large molecules (catabolism), typically to produce energy or to create smaller molecules useful in other chemical processes, or they may result in the synthesis of larger molecules (anabolism) to, for example, create bone or muscle tissue. Each of the thousands of such series of chemical reactions is identified as a metabolic (or cellular) pathway and is dependent on the product of genes to function effectively. These chemical processes can be simple or quite complex.

Examples:

The ***glycolysis metabolic pathway*** oxidizes glucose in order to obtain adenosine triphosphate (ATP), a molecule used within cells to transport energy for use in other metabolic pathways. Hexokinase is an enzyme or protein essential in this chemical process and is coded for by a gene.

The ***citric acid, tricarboxylic acid (TCA) or Krebs' cycle*** uses the catalytic effect of pyruvate dehydrogenase, a complex of three enzymes, in an oxidative process to produce acetyl-CoA, a precursor to guanosine triphosphate (GTP) and other intermediaries. GTP is an essential energy-transfer molecule readily converted into ATP.

Metabolic processes such as these are essential for life to continue and they are dependent on appropriate genetic input.

of accumulated individual improvements. Life expectancy[23] at birth, based on the 2008 VBT mortality data, is age 81.2 for males and 85.1 for females. Preventing an infant or childhood death will not erase death entirely since humans are not immortal and other factors contribute to mortality after early development. An individual saved from death in infancy or early childhood will die of something else later. However, postponing death adds years of life to the life-expectancy calculation, increasing the overall statistic that is a measure of longevity in a population and benefiting the individual life as well.

Despite substantive changes in overall mortality from birth until middle age, death can occur at any age for many reasons. Understanding why and how death occurs and what aspect of those causes are controlled by genetic variation or, simply, the genetically controlled chemical processes of our evolved physiological and anatomical design, will allow us to make genetically informed decisions to defer death and prolong life. Genetic information does not simply tell us *about* ourselves—explaining why our eyes are a certain color or that our genome makes us more or less susceptible, statistically, to a particular condition or disease. Research is telling us more about how genetic variation found in individuals through a personalized medical approach affects the processes of individual lives and how that information can be added to traditional health care information and practices, to enhance, detail and make more effective the medical care specific individuals receive.

What is Death and How Do We Avoid It?

Adult human beings are multicelled organisms—estimated to have well over 50 trillion cells. These 50 trillion cells are highly specialized. Although each has an identical set of genetic information, depending on the specialized activity, stress and need, different cellular chemical processes are turned on or off. Within each cell occur multiple, simultaneous series of chemical reactions that allow each cell to contribute in its own way to the overall organism's continued growth and survival. Cells do this through cell division and through maintaining healthy cell function. Together, these metabolic processes have been organized into a number of chemical or metabolic pathways controlled or supplied by the products of genes.

A prevalent cause of age-related death is the breakdown and impairment of metabolic pathways over time, like those highlighted in the sidebar above, such that they become less efficient. These metabolic processes become disabled or impaired and eventually fail for a number of reasons. For example, the accumulation of genetic damage or mutation in genes, chromosomes and mitochondria can affect or overwhelm the proper or efficient functioning of the metabolic pathways that control the maintenance and repair of basic cellular function. The accumulation of high molecular weight insoluble aggregates (e.g., cholesterol) over time may create physical changes that cause organs to stop working efficiently—skin wrinkles, vision or hearing may fail, heart or brain function becomes inefficient, plaque accumulates in blood vessels, and disease runs its course. Aging through the getting-old stage is, no doubt, a complex

[23] Life expectancy is calculated in the traditional actuarial way here using complete years of life. It is the age at which half of a population will have died.

process involving many little things that add up and affect different people differently and in many different ways, makes survival more of a challenge.

We experience aging as a normal part of life. It may be argued that the evolved physiology of the human body is inconsistent with immortality. One theory uses telomere length as a predictor of the number of times cells can divide before they lose the ability to do so, which would impose a maximum life span on all animals, including humans. This would be an example of programmed senescence and, because of genetic variation, may affect individuals differently. Not all cells in the human body need to divide to maintain effective function but enough do that the inability to be able to would cause a catastrophic failure of the organism.

It is also clear humans are subject to environmental assault not ideal for healthy function. The effect of these assaults on cellular function may differ from individual to individual because of genetic variation and cannot, in all cases, be recognized as assaults as they are occurring until many years of accumulation have taken hold. An understanding of how genes operate to control and fuel metabolic processes and how genetic variation can affect different individual's ability to survive under variable conditions can be used to inform the decisions each person makes or the interventions necessary to maximize the health and longevity potential of each.

Using Genetic Information

There can be no doubt that variation among individuals in a human population regarding health, longevity and aging is related in some way to genetic variation among those individuals. Research continues to add knowledge about the links of genetic variation to disease. Cystic fibrosis, for example, is known to be an autosomal[24] recessive genetic disorder. Two copies of a mutation of the CFTR gene are required (one inherited from each parent) for the disease to manifest itself. Normally, one copy will make an individual a carrier with no symptoms of the disease.[25] The CFTR gene encodes for the protein cystic fibrosis transmembrane conductance regulator (the CFTR protein), a 1,480 amino acid-long protein required to regulate the transference of chloride and sodium ions across epithelial membranes.

The most common mutation of CFTR is a deletion of three nucleotides at position 508 of the protein, resulting in the loss of the amino acid phenylalanine (abbreviated F) at position 508. Hence the mutation is called ΔF508 (or F508del). This mutation accounts for approximately two-thirds of the worldwide occurrence of cystic fibrosis—90 percent in the United States.

[24] "Autosomal" means the gene is not on either the X or Y sex chromosomes. The CFTR gene is located on chromosome 7 between base pairs 117,120,016 and 117,308,718 (GRCh37.p10) at position 7q31.2. It is 188,702 nucleotides long.

[25] Individuals with one copy of a CFTR mutation are called heterozygous. CFTR mutations have been in the human population for upward of 50,000 years and many theories exist as to how such a lethal variant could have existed for so long. These theories center around the fact that one copy, though it does not result in disease, does provide a heterozygote advantage by providing resistance to other disease like cholera, typhoid or diarrhea associated with lactose intolerance—problematic after the domestication of cattle prior to other mutations that resulted in lactose tolerance. Carriers of the CFTR mutation also have some resistance to tuberculosis.

However, there are well over 1,000 mutations to CFTR, each causing different versions of cystic fibrosis as well as other diseases. Different defects in the CFTR gene result in milder or more severe forms of the disease because of the differing effects these mutations have on the production of the protein.[26] Although there is no known cure for cystic fibrosis, there are treatments designed to manage the symptoms. With so many versions of the disease, knowing which version one is dealing with genetically is important.

A new drug, Ivacaftor,[27] was approved by the U.S. Food and Drug Administration (FDA) in January 2010 for use by individuals with the G551D mutation of cystic fibrosis—about 4.5 percent of cases. The G551D mutation affects the protein by substituting the amino acid glycine (abbreviated G) in position 551 with aspartic acid (abbreviated D). Ivacaftor differs from other medication designed to treat symptoms of cystic fibrosis in that it acts as a potentiator, i.e., it's presence allows the defective protein produced by the G551D mutation to function as if normal.[28]

The bottom line is that appropriate treatment to extend life or provide a cure for diseases like cystic fibrosis[29] which have many variations would not be possible without genetic information.

However, there are some limitations to the current use of genetic information, especially for more common diseases. For example, massive genetic studies have uncovered about 25 common genetic markers associated with Type 2 diabetes risk,[30] yet these markers, in aggregate, explain only about 10 percent of the genetic susceptibility to Type 2 diabetes. While these markers can, on average, be used to identify individuals at heightened risk for Type 2 diabetes, the utility of these risk predictions and their implications in terms of therapeutic intervention for a specific individual are somewhat limited. It should be noted that these genetic markers were discovered using a specific technique—genome-wide association studies (GWAS)—described below, and, by definition, are not comprehensive studies into the genetic architecture of disease.

Decreasing Cost of Genetic Testing Makes Use Practical

The working draft of the human genome was announced in 2000 and completed in 2003 at a total cost of about $3 billion. The follow-up International HapMap Project, aimed at

[26] For example with ΔF508, the protein may not fold properly, affecting its function in the cell. Other mutations may cause a protein to be too short or too long, to not use energy properly for normal function, to degrade too quickly or to not be produced in sufficient quantity.

[27] A trade name for Ivacaftor is Kalydeco, which is developed by Vertex Pharmaceuticals (http://www.kalydeco.com).

[28] See Eckford et al., "Cystic Fibrosis."

[29] See, for example, Wikipedia, "Imatinib." Imatinib is marketed by Novartis under the name Gleevec as a treatment for a number of leukemia type cancers. The drug targets only specific cancer cells and inhibits a cascade of chemical reactions that would otherwise result in uncontrolled growth of the cells and therefore cancer, allowing them to die by apoptosis.

[30] Voight et al., "Twelve Type 2 Diabetes susceptibility loci identified through large-scale association analysis."

characterizing individual differences in the human genome, was initiated in October 2002. The goal was to catalog genetic variants that tag, or act as surrogate markers for, the genetic variation contained within co-inherited regions of the genome named haplotype blocks. These haplotype blocks result from the fact that the human genome is transmitted through generations in large nonrandom chunks, leading to co-transmission of genetic variants residing in those haplotype blocks. Thus, for common genetic variants, ascertaining the identity of a subset of genetic variants—known as tag–single-nucleotide polymorphisms (SNPs)—within a haplotype block is sufficient for a reasonable inference of the state of other known genetic variants within that block. The ability to cheaply and efficiently interrogate tag-SNPs spawned a tremendous investment in GWAS aimed at associating tag-SNPs with disease risk.

Genome-wide association studies were enabled by genotyping chips, a simple high-throughput technology that enables the determination of the state, or genotype, of an individual across a large number of known and targeted sites of genetic variation. Genotyping chips, interrogating hundreds of thousands of tag-SNPs, were utilized in a tremendous number of GWAS—leading to a catalog of disease-risk variants, each of which modestly contributes to the overall heritability of a disease. The findings from these GWAS are what populate consumer genetics tests offered by companies such as 23andMe.[31]

During the GWAS era of human genetics, a battery of companies such as Solexa (purchased by Illumina), 454 Life Sciences and Complete Genomics began developing next-generation sequencing platforms. These platforms have rapidly gained accuracy and efficiency, dropping the cost of a human genome sequence from $3 billion at the time the human genome project was initiated, to $100 million in 2001, to $50,000 in 2009, to about $5,000 at present day. This capability spawned a more detailed exploration of the landscape of genetic variants populating the human genome, enabling the identification of rare but physiologically important genetic variants through large-scale genome-sequencing projects, such as the 1,000 Genomes Project. The variants from these studies populate modern genotyping chip platforms, which are capable of interrogating upward of about 2 million genetic variants at a fraction of the cost of genome sequencing. Although not currently used in personal or consumer genetic tests, these chips can reveal the state of known, rare and high-risk disease-associated genetic variants. In the near future, the cost of genome sequencing will make feasible the routine practice of genome sequencing for the simultaneous revelation of common, low-risk genetic variants, rare high-risk genetic variants and other genetic variants relevant to health status, longevity or drug-response phenotypes.

Data Capture Platforms and Databases

As mentioned above, genetic data is currently captured via two biological assays:

1. Genotyping chips, which are designed to interrogate specific, known genetic markers; and

[31] http://www.23andme.com.

2. Sequencing platforms, which produce the entire DNA sequence of an individual either in a specific genomic region or across the whole genome sequence of an individual.

Standard data analysis tools convert the raw data produced by these platforms to one of a few relatively standard reporting formats whose basic elements include the chromosome, physical location and nucleotide identities of the observed genetic variants. In basic personalized medicine and consumer genetics applications, the physical location information is used as a key to cross-reference against databases that contain known disease-risk variants, also indexed by the physical location of variants. These databases generally contain information that is published and publicly accessible; thus the barrier to entry in terms of key intellectual property is low. The curation, analysis, synthesis and interpretation of those publically available data sources are key to developing genetics applications.

Adoption of Genetics in Personalized Medicine and Consumer Genetics

With the exception of a few genetic tests for late-onset diseases, such as BRCA gene testing for breast cancer, the current application of genetic tests in personalized medicine and consumer genetics is through the use of genotyping chips interrogating common low-risk disease variants identified through GWAS. This set of variants also includes pharmacogenetic variants, which are genetic variants that can dramatically influence the efficacy and safety of drugs.

Direct-to-consumer genetic testing by companies like 23andMe provide consumer information on individualized drug response, disease susceptibility, inherited conditions[32] or traits. Having this information may cause an individual to purposefully analyze diseases for which an elevated risk is indicated and seek medical help and advice regarding symptoms that might have otherwise gone unnoticed or to change habits, environment or exposures in response to indicators provided with respect to drug response or traits. This has obvious potential for improving health and longevity, putting off death that might otherwise come from an unattended aging process.

In addition, medical practices are adopting genetically informed approaches. MD Revolution "integrates genetics, metabolics and traditional medicine to develop a personalized health plan with proven outcomes."[33] Personalis "is a contract research organization and genome-scale diagnostics services company pioneering genome guided medicine."[34] Medical providers such as these apply information contained in individual genetic variation to better understand and, therefore, treat or cure diseases or conditions related to differences in metabolism, disease susceptibility and effectiveness of drug therapies.

[32] For example, on cystic fibrosis, 23andMe tests for the presence of 31 of the most common mutations of the CFTR gene. There are more than 1,000.

[33] From the MD Revolution website, http://mdrevolution.com/.

[34] From the Personalis website, http://www.personalis.com/company/overview.

The Level of Genetic Knowledge

Genetic research is beginning to produce interesting results due to the lessening costs associated with identifying genetic variation. It is clear, however, that as of today the data contains much more information than we currently know. With about 3.1 billion base nucleotide pairs[35] in the human chromosome, the billions more in precursor organisms (which help to shed light on the function of genes in the human body), and the relatively short time researchers have had tools sufficient to give them practical access to genetic variation, discovery is only at the beginning.

Today's research efforts, for the most part, focus on finding genotypes associated with phenotypes. That is, research is aimed at finding a statistical relationship (e.g., in GWAS) between genetic variation and how that variation is biologically expressed in an organism. From that statistical association with a biological process, the impact of genetic variation can be imputed to one or more metabolic processes known or suspected to be associated with the phenotype. These associations then allow researchers to dig deeper into the metabolic processes as, for example, with cystic fibrosis discussed above and determine metabolically how genetic variation affects them.

Phenotypes (diseases, for example) expressed by variations in a single gene are called monogenetic, are relatively easy to find and, when found, are almost completely predictive of the monogenetic disease or trait they are associated with. Therefore, genetic variation can accurately reveal much about these simple disease phenotypes such as sickle cell anemia or Huntington's disease. Complex phenotypes not linked to a single genetic variation, however, are much more difficult to resolve and are by far much more prevalent.

Genetic research, as noted above, seeks associations between phenotype expression and genetic variation in the groups of people being studied, which is difficult when there is not a one-to-one relationship between genetic variation and a trait being studied. Traits or diseases like cardiovascular disease, Alzheimer's, or aging and longevity, for example, are made complex when many genetic variations at locations throughout the human genome each have a small contributory effect on the phenotype. In addition, gene-gene or gene-environment interactions are likely to have an impact on the expression of complex traits.

When dealing with life insurance risk and collecting data reflective of an association between genetic variation and mortality or morbidity inaccuracies in cause of death or illness, coding may confound the resulting database, making associations more difficult. These inaccuracies result from broad groupings of phenotypical expression that share a common name and are assumed to be the same disease but that actually have multiple etiologies (i.e., causes or

[35] The length of the human genome can only be stated approximately since the state of technology today makes it difficult to correctly sequence all base nucleotide pairs in the correct order with 100% accuracy. In addition, because of insertions, deletions and copy number variation mutations, the human genome is not the same length for everyone.

origins) even though they look the same. For example, cancer is a broad phenotypic category for a class of diseases with multiple, distinct genetic associations.

The effect of prenatal or perinatal environmental conditions may also be a confounding factor with respect to genetic expression. Perhaps the largest confounders in many genetic association studies are epigenetic effects, i.e., the inherited changes in phenotype caused by gene expression mechanisms rather than differences in the underlying DNA.

It is important to recognize that disease and mortality risk built off genetic information is merely probabilistic, in part because our current knowledge of the effects of genetic variation are built off statistical models and because genetic variation is only partially contributory to phenotype expression. Genetic variation is indicative of only a propensity for disease within a range of probability indicated by an interpretation of research results—quantified by a statistical model and/or algorithm. Environment and mitigating actions all affect the phenotypical expression of genetic variation. That is why information on genetic variation merely informs; it does not predict.

Variant Classes and How Genetic Information Informs

Genes related to aging and longevity have been identified and listed in the GenAge[36] database. These lists have been compiled through extensive manual curation of research on aging and longevity in which a gene's association with aging or longevity must be unambiguous. Typically, genes are included as associated with either aging or longevity as a result of manipulation of the gene as opposed to mere correlation with age. Appendix A shows a listing of human genes associated to aging or longevity; appendix B lists genes analyzed for their possible association with human longevity and the conclusions reached with respect to each analysis.

The fact that genetic variation along with other factors is related to aging and longevity is discussed in a 2012 *Nature* supplement.[37] There are many research articles linking genetic variation to aging and longevity,[38] and appendices A and B provide links to articles describing research that connected the listed gene to aging and longevity. The attached paper[39] also provides an analysis of how genetic variation in one gene, FOXO3, can affect longevity.

GWAS Disease-Associated Tag Markers

Genome-wide association study disease-associated tag markers are genetic markers (i.e., naturally occurring variations in the DNA sequence possessed by individuals) known to be

[36] The GenAge database is a product of the Human Ageing Genomic Resources (HAGR). More information can be found at http://genomics.senescence.info/about.html.

[37] *Nature*, "Ageing," supplement.

[38] See, for example, Yashin at al., "Joint Influence."

[39] Bakos et al., "The Effect of FOXO3," describes the metabolic processes affected by genetic variation in the FOXO3 gene and how this affect may be used to modify mortality tables based on the presence of the genetic variation.

associated with increased (or decreased) risk of specific diseases.[40] Many markers are known to be associated with general longevity.[41] Individually these variants occur commonly in the population but each individually contributes weakly to the risk of disease or life span. Ultimately, the aggregate impact of GWAS tag markers within an individual can explain a small but significant adjustment in an individual's disease risk. These markers can be, and are, routinely interrogated by genotyping chips and DNA sequencing technologies, such as those employed by consumer genetics test providers.

Known High-Risk (Causative or High-Risk) Variants

Known high-risk variants are genetic markers shown to be unequivocally associated, or even causally associated, with increased risk of specific diseases.[42] Individually these variants occur rarely in the population but each one contributes dramatically to disease risk when identified in an individual. While each variant is individually rare, numerous rare variants can impact a specific gene—for example, there are over 1,000 known mutations in the CFTR gene associated with cystic fibrosis. Uncommon (approximately a 1 percent allele frequency) high-risk causative variants can be interrogated on high density genotyping chips using special methodologies; however, these variants are most often identified by targeted gene-sequencing tests.

Predicted High-Risk (Causative or High-Risk) Variants

Predicted high-risk variants are genetic markers similar to known high-risk variants but are only predicted or suspected, rather than known, to be associated with disease. Predictions are based on algorithms that combine biochemical and structural information about a variant within a known disease-associated gene, the population frequency of that variant and the predicted impact of the variant upon gene function in order to stratify variants into risk classes based on prediction confidence. These predictions are derived from computationally based genetic variant classification engines such as those developed by Cypher Genomics Inc. The aggregate impact of known high-risk disease variants and predicted high-risk disease variants upon disease incidence is generally thought to be larger than that of GWAS tag markers—however, confirmation is not currently feasible given the heterogeneity and low frequency of rare variants.

Genetic Risk Assessment and Types of Evidence

In calculating genetically informed disease and mortality risks, a number of very specific methodological issues and strategies should be considered. Each is outlined briefly below.

---

[40] Hindorff et al., "Potential Etiologic."

[41] Deelen et al., "Genome-Wide Association Study," and Yashin at al., "Joint Influence."

[42] See Online Mendelian Inheritance in Man, http://omim.org/, which allows one to search various diseases or traits for associations with genetic variations.

- Direct associations with longevity: Some genetic markers have been shown to be associated with human life span directly using data on longitudinal studies.[43] Such studies are rare given the complications and costs surrounding their implementation.

- Direct associations with survival to old age: Genetic markers have been identified that are associated with an individual's surviving or living to an old age (e.g., living to 100 or older).[44] Such markers are clearly associated with longevity but are not directly associated with mortality or the incidence rate of death, requiring appropriate modeling to determine their association with age-specific mortality probabilities.

- Direct associations with disease: As noted, many genetic markers have been associated with diseases and incidence of disease.[45] Many of these diseases are in turn associated with mortality and life span. Thus, in drawing conclusions about the probability of death within a certain age-interval based on a genetic marker associated with disease, one must factor in that disease's association with the incidence of death.

- Associations with disease risk factors: Many genetic markers have been shown to be associated with disease- and/or health-related factors, such as cholesterol level, cigarette smoking and even diet.[46] The consideration of such markers in an assessment of mortality incidence must accommodate assumptions and modeling about the relationship of those factors to disease as well as association of disease to mortality.

- Polygenic modeling: It has been shown that some conditions are influenced by the collective effects of many genes, each with near infinitesimal effects. These effects can be captured through the use of high-density genotyping or sequencing technologies and can be applied to the analysis of the genetic determinants of life span (if the relevant cohorts exist), healthy aging, disease risk and disease risk factors.

Integrated Risk Assessment: Predictive Biomarkers vs. Surrogate Endpoints

Of crucial importance in the assignment of disease and mortality risk to an individual or group of individuals is sensitivity to, and accommodation of, the difference between predictive biomarkers and surrogate endpoints and the need to integrate the two. Predictive biomarkers are markers of disease or life span that are fixed and do not change over time, such as a person's genetic constitution. Surrogate endpoints are measures that provide insight into whether a disease process is present and changes over time as a result of, for example, interventions with respect to cholesterol level or lifestyle changes. Each of these should be considered when evaluating a person's disease- or mortality-risk status, to the degree possible

[43] Albani et al., "Modulation of Human Longevity."

[44] Atzmon et al., "Lipoprotein Genotype," and Willcox et al., "FOXO3A Genotype."

[45] Yashin at al., "Joint Influence."

[46] Ibid.

given available information on an individual. An individual may possess genetic variants that predispose him or her to heart disease but not have any signs heart disease is present. This person would be at risk for heart disease based on his or her genetic profile. However, if that person has an elevated cholesterol level (i.e., surrogate endpoint measure consistent with the presence of subclinical disease) at the time of his or her genetic analysis and it is revealed he or she has an elevated genetic risk (i.e., predictive biomarker risk), then that person's overall risk should be upwardly modified. Risk modeling that combines predictive biomarker and surrogate endpoint measures is not trivial in the absence of longitudinal data[47] that considers both, but is possible.

Genetically Informed Disease-Risk Assessment

Complex disease results from the combined small effects of many genetic variations—perhaps hundreds or thousands of differences in base nucleotide pairs that make up the 3.1 billion base pair human genome. In addition, gene-gene or gene-environment interactions may also play a role. As noted, there have been a number of studies that have identified associations between particular genetic variants that some individuals possess and increased disease risk.[48] By combining the genetic variants associated with any disease, a single risk assessment or classification model for that disease can be fashioned. Such a classification model could be used to predict whether, on the basis of his or her genetic profile, an individual is likely to have a particular disease. A few recent publications have considered the overall potential for such combined classification models using a standard statistical measure known as the "Area Under the Curve" (AUC) within the context of "Receiver Operator Characteristic" (ROC) curve analyses.[49] Essentially, the AUC represents the probability that a randomly chosen individual with a disease is correctly rated or ranked as more susceptible to that disease than a randomly chosen individual without the disease. The AUC is a function of many things, including the prevalence of the condition assessed, the frequency of the factors used to assess the condition and the reliability of the measurement of the susceptibility factors. An AUC of 1.0 suggests that a classification model is perfect and an AUC of 0.5 suggests the classification model does not work at all (i.e., it is no better than a coin flip as to whether someone is likely to have the disease or not). A risk-assessment model with an AUC value of 0.85 or greater is typically seen as being clinical meaningful and useful.

[47] A longitudinal study is a type of research study in which the same group of lives is observed over a long period of time. A longitudinal study examining the effect of genetic variation over time and eliminates extraneous differences that might be present when different groups of individuals are studied as, for example, in a cross-sectional study. Therefore, the effect of surrogate endpoint measures can more definitively be attributed to a disease process and not some other difference, such as generational, cultural, environmental or geographic, in the populations being studied. However, cross-sectional studies in which different individuals are studied might still provide useful results if the differences between the groups being studied are known and accounted for.

[48] Manolio, "Genomewide Association," and Manolio, Brooks, and Collins, "A HapMap Harvest."

[49] Pepe, *The Statistical Evaluation.*

To better understand table 2 below, it is important to understand "heritability" as used therein means the proportion of a phenotypic variation that results from the additive effect of hundreds to thousands of genetic variations. Genetic variation may contribute to a phenotype but genetic variation may not be the only or sole contributor to phenotypic expression of a trait or disease. For example, environment, diet or random chance are examples of nongenetic factors that, along with genetic variation, can play a role in the phenotypic expression of traits.

Heritability is a population parameter. Looking at table 2, one can say that for individuals in the population exhibiting a Type 2 diabetes phenotype, 60 percent of that expression is due to the additive effect of genetic variation shared by those individuals. The rest is nongenetic in origin, or due to environment. For example, it is estimated that height is approximately 80 percent heritable. It is highly likely that short parents will not give rise to tall children. However, although it is highly likely that tall children will result from tall parents, it is not at all certain since only 80 percent of height is a result of the cumulative effect of genetic variation. Early childhood nutrition will play a role. On the other hand, the ability to speak a particular language is not a heritable trait in the genetic sense. Humans are genetically endowed with the ability to speak a language but the fact that humans converse in many languages is nongenetic.

Information about how much of the risk for a particular disease can be explained by genetic variations has been put forth in the scientific literature based on classical heritability and family studies. Given this information, it is theoretically possible to determine the AUC for a disease classification model in which some fraction of the genetic component of a disease is explained through the identification of individual genetic variants associated with that disease. Table 2, which is adapted and extended from the calculations provided by two 2010 publications,[50] provides AUC values for diseases contributing to the leading causes of death listed in tables 1A and 1B of this paper. In table 2, the Heritability column shows the classic, overall, heritability factor for each disease trait listed. The disease traits shown in table 2 are all complex traits, meaning their phenotypic expression could be contributed to by thousands of genetic variants in addition to environment. Table 2 assumes that some fraction (25 percent, 50 percent or 100 percent) of the total number of the genetic variants contributing to the genetic component of a complex disease are identified and present. This fraction is represented as the percentage of the heritable component of disease susceptibility explained by those variants.

Table 2

AUC values for risk prediction for settings in which sets of genetic variants associated with a leading cause of mortality have been identified

Disease	Heritability	25%	50%	100%
Type 2 diabetes	0.60	0.75	0.84	0.94
Prostate cancer	0.44	0.72	0.80	0.90
Breast cancer	0.44	0.71	0.79	0.89

[50] Wray et al., "The Genetic Interpretation," and So and Sham, "A Unifying Framework."

Coronary artery disease	0.72	0.75	0.84	0.95
Asthma	0.37	0.71	0.79	0.88
Alzheimer's disease	0.70	0.78	0.87	0.96
Suicide	0.40	0.82	0.91	0.98
Alcoholism	0.55	0.72	0.80	0.91

Key:

AUC: Area under the curve associated with receiver-operator characteristic (ROC) curve analysis

25%, 50% and 100%: percentage of the heritable component of disease explained by identified associated genetic variants

In table 2, the columns labeled 25 percent, 50 percent and 100 percent show AUC values with respect to the diseases or traits listed in each row for individuals with, respectively, 25 percent, 50 percent or 100 percent of the genetic variants found to be contributors to the genetic component of the trait. As noted above, AUC values are relative measures of the likelihood that the percentage of genetic variation associated with a trait found in an individual is predictive of that individual having the trait or disease. An AUC of 1.0 would suggest certainty. An AUC of .50 would be no better than flipping a coin. An AUC greater than .50, however, as shown in table 2, indicates that having a larger percentage of the variants associated with a disease trait increases the control genetic variation has on the expression of the trait and, therefore, the heritability of the trait.

Note that the categories for the leading causes of death listed in tables 1A and 1B are broad (e.g., cancer) and subsume many different specific types of disease (e.g., breast cancer, prostate cancer) that have not all been equally well studied, some of which are listed in table 2. Also note that although there are genetic variants that contribute to phenomena such as suicide, they are not as well characterized. In addition, although the fraction of the genetic component of a disease explained by currently identified, unequivocally associated variants for many diseases is small—on the order of what amounts to 5–20 percent of the heritable component for the diseases listed in table 2[51]—research efforts are growing such that more genetic variants associated with disease will be found. Table 2 therefore really does reflect the potential that genetically informed disease-risk assessment models can have and suggests a reliability over currently used models.

[51] See, for example, Do et al., "Comparison of Family History."

From the introduction: "Recent estimates of the proportion of heritability explained by known susceptibility variants across a survey of ten complex diseases (Alzheimer disease, bipolar disorder, breast cancer, coronary artery disease, Crohn disease, prostate cancer, schizophrenia, systemic lupus erythematosus, Type 1 diabetes, and Type 2 diabetes) have ranged from 0.4% to 31.2% [5]. These proportions highlight the sobering reality that only a fraction of the genetic contributions to disease have yet been discovered."

The footnote [5] above is a reference to So et al., "Evaluating the Heritability."

Concluding Remarks

Finally, it is important to emphasize three things.

1. Just having a disease does not dictate dying from it; in fact, as emphasized throughout this paper, intervention and therapeutic strategies exist that can mitigate the life-threatening nature of a disease.
2. Many diseases are correlated, such that if an individual is genetically susceptible to one disease, that individual is likely at greater risk for other diseases as well.
3. True disease and mortality risk are functions of genetic and environmental factors as well as their interactions. This is reflected in the fact that even if all the genetic variants were identified for a particular disease, i.e., 100 percent of the heritable component for that disease, the risk assessment classifier would still not be perfect and as such have an AUC < 1.0.

Thus, comprehensive and reliable risk-assessment models need to consider both genetic and nongenetic factors.

Social, Political and Privacy Issues

Until recently, the collection and use of genetic information for any purpose was a theoretical issue. The expense of obtaining genetic information had been a significant prohibition of its use, particularly in insurance risk assessment but also even in medical practices. For example, in an American Academy of Actuaries' 1997 *Issues Brief*,[52] the following conclusion was reached: "At present [i.e., 1997], the complexity and expense of genetic tests make premature the widespread use of genetic testing in underwriting."

In 2003, the 13-year Human Genome Project completed the first sequencing of a complete human genome for a reported cost of $3 billion—about $1 per nucleotide base pair. Although a complete sequencing of a human genome is not necessary to get useful genetic information, clearly these expenses only 10 years ago were a prohibitive factor to any routine use of genetic information in a business process or even in most medical treatment.

The cost today for a complete sequencing of the human genome is pushing down on $1,000 and genetic testing used to find genetic markers (1,000 or more) that can be used to find genetic variation associated with disease are $50 or less and are now available directly to the general public by companies such as 23andMe and GenePlanet[53] for about $99.

[52] American Academy of Actuaries, "Risk Classification in Voluntary Life Insurance," and a number of AAA publications have addressed the use of genetic information in insurance risk-selection processes. See also American Academy of Actuaries, "Genetic Information and Medical Expense Insurance," "Genetic Information and Voluntary Life Insurance," "Risk Classification in Individually Purchased Voluntary Medical Expense Insurance," "The Use of Genetic Information in Disability Income and Long-Term Care Insurance" and "Risk Classification in Voluntary Individual Disability Income and Long-Term Care Insurance."

[53] http://www.geneplanet.com/

The ease with which genetic information can be acquired and its value to medicine is likely to make it a prevalent part of medical databases and, although genetic information is probabilistic and not determinative, it is useful in assessing life and health risk. Insurers and others who price risk may be at a disadvantage relative to their customers if only one party in a contract situation has access to the information.

The Health Insurance Portability and Accountability Act of 1996 (HIPPA)[54] and Genetic Information Nondiscrimination Act (GINA)[55] in general protect the privacy of health information, but they do so principally in the health insurance market. Genetic information covered in GINA (there is no special category for genetic information in HIPPA) is defined to include not only information derived from genetic tests on an individual or family members but also "the manifestation of a disease or disorder in family members of such individual," i.e., family history.

While this paper is not directed at the social, political or privacy aspects of the use of genetic information in risk assessment or selection, those are obviously concerns with respect to any planned use of genetic information. The author[56] of one paper has addressed concerns about use of genetic information in insurance and found no justification for regulation. Clearly, genetics is a very rapidly progressing field of research with obvious value in medical care and the use of genetic information will become widespread and probably better accepted as more is learned.

[54] This act was amended in January 2013.

[55] This act was signed into law in May 2008 and last amended in March 2013.

[56] Manson and Conko, "Genetic Testing."

References

Albani, Diego, Eleonora Ateri, Stefano Mazzuco, Alice Ghilardi, Serena Rodilossi, Gloria Biella, Fausta Ongaro, et al. "Modulation of Human Longevity by SIRT3 Single Nucleotide Polymorphisms in the Prospective Study 'Treviso Longeva (TRELONG).' " *Age* 36 (2014): 469–78. doi:10.1007/s11357-013-9559-2.

American Academy of Actuaries (AAA). "Genetic Information and Medical Expense Insurance." Public Policy Monograph, June 2000. http://www.actuary.org/files/geneticmono.4.pdf/geneticmono.4.pdf.

———. "Genetic Information and Voluntary Life Insurance." *Issue Brief*, Spring 1998. http://www.actuary.org/files/genet.4.pdf/genet.4.pdf.

———. "Risk Classification in Individually Purchased Voluntary Medical Expense Insurance." *Issue Brief*, February 1999. http://www.actuary.org/files/publications/risk%20classification%20021999.pdf.

———. "Risk Classification in Voluntary Individual Disability Income and Long-Term Care Insurance." *Issue Brief*, Winter 2001. http://www.actuary.org/files/issue_genetic_021601.4.pdf/issue_genetic_021601.4.pdf.

———. "Risk Classification in Voluntary Life Insurance." *Issue Brief,* Spring 1997. http://www.actuary.org/files/riskclas.4.pdf/riskclas.4.pdf.

———. "The Use of Genetic Information in Disability Income and Long-Term Care Insurance." *Issue Brief*, Spring 2002. http://www.actuary.org/files/publications/genetic_25apr02.pdf.

Arias, Elizabeth. "United States Life Tables, 2008." *National Vital Statistics Reports* 61, no. 3 (Sept. 24, 2012). http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_03.pdf.

Atzmon, Gil, Marielisa Rincon, Clyde B. Schechter, Alan R. Shuldiner, Richard B. Lipton, Aviv Bergman, and Nir Barzilai. "Lipoprotein Genotype and Conserved Pathway for Exceptional Longevity in Humans." *PLoS Biology* 4, no. 4 (2006): e113. doi:10.1371/journal.pbio.0040113.

Bakos, Tom, et al. "The Effect of FOXO3 Gene Variants on Human Longevity." Unpublished.

Courts, Cornelius, and Burkhard Madea. "Genetics of the Sudden Infant Death Syndrome." *Forensic Science International* 203, no. 1–3 (Dec. 10, 2010): 25–33. doi:10.1016/j.forsciint.2010.07.008.

Deelen, Joris, Marian Beekman, Hae-Won Uh, Quinta Helmer, Maris Kuningas, Lene Christiansen, Dennis Kremer, et al. "Genome-Wide Association Study Identifies a Single Major Locus Contributing to Survival Into Old Age: The APOE Locus Revisited." *Aging Cell* 10, no. 4 (2011): 686–98. doi:10.1111/j.1474-9726.2011.00705.x.

Do, Chuong B., David A. Hinds, Uta Francke, and Nicholas Eriksson. "Comparison of Family History and SNPs for Predicting Risk of Complex Disease." *PLoS Genetics* 8, no. 10 (2012): e1002973. doi:10.1371/journal.pgen.1002973. http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002973

Eckford, Paul D. W., Canhui Li, Mohabir Ramjeesingh, and Christine E. Bear. "Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Potentiator VX-770 (Ivacaftor) Opens the Defective Channel Gate of Mutant CFTR in a Phosphorylation-Dependent but ATP-Independent Manner." *Journal of Biological Chemistry* 287, no. 44 (2012): 36639–49. doi:10.1074/jbc.M112.393637. http://www.jbc.org/content/287/44/36639.long

Heron, Melonie. "Deaths: Leading Causes for 2009." *National Vital Statistics Report* 61, no. 7 (Oct. 26, 2012).

Hindorff, Lucia A., Praveen Sethupathy, Heather A. Junkins, Erin M. Ramos, Jayashri P. Mehta, Francis S. Collins, and Teri A. Manolio. "Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits." *Proceedings of the National Academy of Sciences* 106, no. 23 (2009): 9362–67. doi:10.1073/pnas.0903103106.

Kaplan, Robert C., Ann-Kristin Petersen, Ming-Huei Chen, Alexander Teumer, Nicole L. Glazer, Angela Doring, Carolyn S. P. Lam, et al. "A Genome-Wide Association Study Identifies Novel Loci Associated With Circulating IGF-I and IGFBP-3." *Human Molecular Genetics* 20, no. 6 (2011): 1241–51. doi:10.1093/hmg/ddq560.

Lahdenpera, Mikka, Virpi Lummaa, Samuli Helle, Marc Tremblay, and Andrew F. Russell. "Fitness Benefits of Prolonged Post-Reproductive Lifespan in Women." *Nature* 428, no. 6979 (March 11, 2004): 178–81. http://www.ncbi.nlm.nih.gov/pubmed/15014499

Manolio, Teri A. "Genomewide Association Studies and Assessment of the Risk of Disease." *New England Journal of Medicine* 363, no. 2 (2010): 166–76. doi:10.1056/NEJMra0905980.

Manolio, Teri A., Lisa D. Brooks, and Francis S. Collins. "A HapMap Harvest of Insights Into the Genetics of Common Disease." *The Journal of clinical investigation* 118, no. 5 (2008): 1590–605. doi:10.1172/JCI34772. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2336881/

Manson, Neil A., and Gregory Conko. "Genetic Testing and Insurance: Why the Fear of 'Genetic Discrimination' Does not Justify Regulation." Issue Analysis, no. 4, Competitive Enterprise Institute (2007). http://cei.org/pdf/5855.pdf.

McCormick, Mark, Kan Chen, Priya Ramaswamy, and Cynthia Kenyon. "New Genes That Extend *Caenorhabditis elegans*' Lifespan in Response to Reproductive Signals." *Aging Cell* 11, no. 2 (2012): 192–202.

Nature. "Ageing." *Nature Outlook* 492, no. 7427 supplement (2012): S1–S40. http://www.nature.com/nature/outlook/ageing/.

Online Mendelian Inheritance in Man (OMIM). "601969: Novelty Seeking Personality Trait." http://www.omim.org/entry/601696. Accessed May 2013.

Opdal, Siri H., and Torleiv O. Rognum. "The Sudden Infant Death Syndrome Gene: Does It Exist?" *Pediatrics* 114, no. 4 (2004): e506-e512. doi:10.1542/peds.2004-0683. http://pediatrics.aappublications.org/content/114/4/e506.full.html

Pepe, Margaret Sullivan. *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford Statistical Science Series 28. Oxford: Oxford University Press, 2004.

Preferred Valuation Basic Table Team, a subgroup of the American Academy of Actuaries/Society of Actuaries Joint Preferred Mortality Project Oversight Group. "2008 Valuation Basic Table." Report presented to the National Association of Insurance Commissioners' Life and Health Actuarial Task Force, Orlando, Fla., March 2008.

Simon C. Johnson, Peter S. Rabinovitch, and Matt Kaeberlein. "mTOR is a Key Modulator of Ageing and Age-Related Disease." *Nature* 493, no. 7432 (2013): 338–45. doi:10.1038/nature11861.

So, Hon-Cheong, Allen H.S. Gui, Stacey S. Cherny, and Pak C. Sham. "Evaluating the Heritability Explained by Known Susceptibility Variants: A Survey of Ten Complex Diseases." *Genetic Epidemiology* 35, no. 5 (2011): 310–17. doi:10.1002/gepi.20579.

So, Hon-Cheong, and Pak C. Sham. "A Unifying Framework for Evaluating the Predictive Power of Genetic Variants Based on the Level of Heritability Explained." *PLoS Genetics* 6, no. 12 (2010): e1001230. doi:10.1371/journal.pgen.1001230. http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1001230

Somel, Mehmet, Song Guo, Ning Fu, Zheng Yan, Hai Yang Hu, Ying Xu, Yuan Yuan, et al. "MicroRNA, mRNA, and Protein Expression Link Development and Aging in Human and Macaque Brain." *Genome Research* 20, no. 9 (2010): 1207–18. doi:10.1101/gr.106849.110.

USA LifeExpectancy. "USA Causes of Death by Age and Gender." http://www.worldlifeexpectancy.com/usa-cause-of-death-by-age-and-gender. Accessed May 2013.

Voight, Benjamin F., Laura J. Scott, Valgerdur Steinthorsdottir, Andrew P. Morris, Christian Dina, Ryan P. Welch, Eleftheria Zeggini, et al. "Twelve Type 2 Diabetes Susceptibility Loci Identified Through Large-Scale Association Analysis." *Nature Genetics* 42, no. 7 (2010): 579–89. doi:10.1038/ng.609. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3080658/pdf/ukmss-34421.pdf

Walter, Stefan, Gil Atzmon, Ellen W. Demerath, Melissa E. Garcia, Robert C. Kaplan, Meena Kumari, Kathryn L. Lunetta, et al. "A Genome-Wide Association Study of Aging." *Neurobiology of Aging* 32, no. 11 (2011): 2109.e15-2109.e28.

Wikimedia Commons. “File:MTOR-pathway-v1.7.svg.” Attributed to Charles Betz. Last updated August 31, 2012. http://en.wikipedia.org/wiki/File:MTOR-pathway-v1.7.svg.

Wikipedia. “Imatinib.” http://en.wikipedia.org/wiki/Imatinib. Accessed May 2013.

Willcox, Bradley J., Timothy A. Donlon, Qimei He, Randi Chen, John S. Grove, Katsuhiko Yano, Kamal H. Masaki, D. Craig Willcox, Beatriz Rodriguez, and J. David Curb. “FOXO3A Genotype is Strongly Associated With Human Longevity.” *Proceedings of the National Academy of Sciences* 105, no. 37 (2008): 13987–92. doi:10.1073/pnas.0801030105.

Wray, Naomi R., Jian Yang, Michael E. Goddard, and Peter M. Visscher. “The Genetic Interpretation of Area Under the ROC Curve in Genomic Profiling.” *PLoS Genetics* 6, no. 2 (2010): e1000864. doi:10.1371/journal.pgen.1000864. http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000864

Yashin, Anatoliy I., Deqing Wu, Konstantin G. Arbeev, and Svetlana V. Ukraintseva. “Joint Influence of Small-Effect Genetic Variants on Human Longevity.” *Aging* 2, no. 9 (2010): 612–20. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2984609/

Zarse, Kim, Sebastian Schmeisser, Marco Groth, Steffen Priebe, Gregor Beuster, Doreen Kuhlow, Reinhard Guthke, Matthias Platzer, C. Ronald Kahn, and Michael Ristow. “Impaired Insulin/IGF1 Signaling Extends Life Span by Promoting Mitochondrial L-Proline Catabolism to Induce a Transient ROS Signal.” *Cell Metabolism* 15, no. 4 (2012): 451–65. doi:10.1016/j.cmet.2012.02.013.

Zhang, Guo, Juxue Li, Sudarshana Purkayastha, Yizhe Tang, Hai Zhang, Ye Yin, Bo Li, Gang Liu, and Dongsheng Cai. “Hypothalamic Programming of Systemic Ageing Involving IKK-β, NF-κB and GnRH.” *Nature* 497, no. 7448 (2013): 211–16. doi:10.1038/nature12143.