

Selection on Alleles Affecting Human Longevity and Late-Life Disease: The Example of Apolipoprotein E

Fotios Drenos^{*‡}, Thomas B. L. Kirkwood

Institute for Ageing and Health, Newcastle University, Newcastle Upon Tyne, Tyne, United Kingdom

Abstract

It is often claimed that genes affecting health in old age, such as cardiovascular and Alzheimer diseases, are beyond the reach of natural selection. We show in a simulation study based on known genetic (apolipoprotein E) and non-genetic risk factors (gender, diet, smoking, alcohol, exercise) that, because there is a statistical distribution of ages at which these genes exert their influence on morbidity and mortality, the effects of selection are in fact non-negligible. A gradual increase with each generation of the $\epsilon 2$ and $\epsilon 3$ alleles of the gene at the expense of the $\epsilon 4$ allele was predicted from the model. The $\epsilon 2$ allele frequency was found to increase slightly more rapidly than that for $\epsilon 3$, although there was no statistically significant difference between the two. Our result may explain the recent evolutionary history of the epsilon 2, 3 and 4 alleles of the apolipoprotein E gene and has wider relevance for genes affecting human longevity.

Citation: Drenos F, Kirkwood TBL (2010) Selection on Alleles Affecting Human Longevity and Late-Life Disease: The Example of Apolipoprotein E. PLoS ONE 5(4): e10022. doi:10.1371/journal.pone.0010022

Editor: Matt Kaeberlein, University of Washington, United States of America

Received: January 14, 2010; **Accepted:** March 13, 2010; **Published:** April 2, 2010

Copyright: © 2010 Drenos, Kirkwood. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors thank the Dr Hadwen Trust for financial support (<http://www.drhadwentrust.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: F.Drenos@ucl.ac.uk

‡ Current address: Division of Cardiovascular Genetics, Department of Medicine, University College London, London, United Kingdom

Introduction

Evidence points towards the existence of a strong heritable component of human longevity [1,2]. Around a quarter to a third of the variability of lifespan can be attributed to the action of genes [3,4,5,6,7]. One of the best examples of a gene affecting survival in old age is the apolipoprotein E gene *APOE*. Situated on human chromosome 19 at locus 19q13.32 (<http://genome.ucsc.edu/>, assembly (hg18)), the *APOE* gene is 3.6 kb long containing four exons and coding for a 317 amino-acid polypeptide that after cleavage gives rise to a 299 amino-acid long mature protein [8,9]. Apolipoprotein E (APO E) is a member of a diverse family of carrier proteins specializing in lipoprotein particle formation, secretion, transport, binding and metabolism [10]. APO E is synthesized in many different regions of the body such as the liver, brain (primarily astrocytes), skin, macrophages and steroidogenic organs [11].

APO E has three major and more than thirty minor isoforms, the latter being mostly linked with disease. The three common alleles are epsilon (ϵ) 2, 3 and 4 producing three homozygous ($\epsilon 2/\epsilon 2$, $\epsilon 3/\epsilon 3$ and $\epsilon 4/\epsilon 4$) and three heterozygous ($\epsilon 2/\epsilon 3$, $\epsilon 3/\epsilon 4$ and $\epsilon 2/\epsilon 4$) genotypes [12]. The difference between the alleles lies at two amino acid residues 112 and 158; with the most common allele $\epsilon 3$ having cysteine and arginine at these residues, respectively, while $\epsilon 2$ has cysteine and $\epsilon 4$ arginine at both locations [9]. Carriers of different alleles show differences in the incidence of coronary artery disease (CAD), peripheral atherosclerosis, Alzheimer disease, possibly stroke and even ability to recover from trauma [11,13,14,15,16]. The underlying mechanism for the action of the gene involves an inter-domain interaction between the amino- and carboxyl- terminals of the protein, producing isoform-specific

lipoprotein preferences such that E2 and E3 protein isoforms bind preferentially to HDL (the “good” cholesterol), while the E4 isoform shows a preference for VLDL [17]. The APO E2 isoform is defective in binding the LDL receptor, although it retains its ability to interact with LDL-receptor-related protein and other related receptors [11]. A small number of $\epsilon 2$ homozygotes suffer from a condition known as type III hyperlipoproteinemia, characterized by accumulation of cholesterol-rich remnant lipoproteins due to incomplete catabolism of chylomicrons and VLDL, leading to premature atherosclerosis. The fact that although 90% of the patients are $\epsilon 2$ homozygotes, but only 5% of the $\epsilon 2$ homozygotes suffer from the condition, suggests that further factors are required for the phenotypic expression of the condition [18,19].

Generally in Caucasian populations survival to advanced age is more likely for carriers of the $\epsilon 2$ allele than for $\epsilon 3$ homozygotes and less likely for carriers of the $\epsilon 4$ allele [20,21,22,23,24]. The $\epsilon 4$ allele is strongly associated with increased risk of coronary heart disease (CHD) [14,25,26,27,28], while for the $\epsilon 2$ allele there is some, though weaker, evidence for a protective effect [29,30]. The APOE polymorphism, has also been implicated in Alzheimer’s disease (AD), which is the commonest form of senile dementia [31,32]. With the $\epsilon 4$ allele shown, in a number of studies, to be associated with both familial and sporadic forms of the disease, causing higher incidence and earlier age of onset, and affecting its pathology and rate of progression [33,34]. Normal age-associated cognitive decline, and mild cognitive impairment and the risk of its conversion to AD have also been repeatedly associated with the presence of the $\epsilon 4$ allele [35].

One of the major ideas in the evolutionary theory of ageing [36] is the suggestion that, because the force of natural selection

declines with age, alleles with deleterious effects seen only at older ages can reach higher frequencies than those that have their effects earlier in life [37]. Therefore, if a gene exerts an effect only after the end of the reproductive phase of the lifespan it has been thought unlikely that it could have been subject to significant direct selection pressure, and this would seem to be the case for genes affecting late-life diseases such as AD and CHD [38,39]. Nevertheless, the worldwide abundance of $\epsilon 3$, as compared to $\epsilon 4$ which, from studies in non-human primates appears to be the ancestral allele [40,41], suggests that selection has acted upon these alleles. Here, we use a simulation approach based on the known risk factors for cardiovascular disease (CVD) to examine the hypothesis that the evolution of the *APOE* gene was, and still is, driven by its role in lipid metabolism and its subsequent effects on health.

Methods

Risk factors

The selection of modifiable risk factors to include in the simulation was based on current knowledge of lifestyle parameters affecting cardiovascular disease (CVD) while keeping in mind the need for independence between the risk factors, simplicity and availability of data. There are two categories of risk factors associated with CVD: *non-modifiable* risk factors, such as *APOE* genotype, and gender, and *modifiable* risk factors, such as smoking, unhealthy diet, lack of physical activity and high alcohol consumption [42]. For the non-modifiable risk factors, the estimated genotype relative risk used was from Gerdes et al [29], while the gender difference in CVD was based on the results of Panagiotakos et al [43]. To quantify the effect of diet, we used the five clusters described by Millen et al [44] in the Framingham study (Heart Healthy, Light Eating, Wine and Moderate Eating, High Fat, and Empty Calorie) [44,45]. For alcohol, smoking and exercise, we used the risk estimates from Stampfer et al [46].

Genotype-environment interaction

A number of studies found evidence of interaction between the *APOE* genotype and each of the modifiable risk factors considered in the model (diet, [9,47,48,49]; alcohol [50,51,52]; smoking [53,54]; and exercise [55]), although the results are sometimes conflicting. Importantly, their precise interaction in quantitative terms remains unknown and will ultimately require a series of studies in large populations before the levels of genotype-environment interactions can be estimated with adequate precision.

Here we use the association of *APOE* and diet to illustrate the principles used to account for the unknown interaction terms. Our first assumption was that, except *APOE*, the risk factors considered were independent of each other. Then we assumed that the relative risk weighted average of the risk factor, for all *APOE* alleles, was approximately 1, i.e. equal to the reference $\epsilon 3\epsilon 3$ genotype. According to the Odds Ratios (OR) given in Gerdes et al [29] the weighted average is ($\epsilon 2$ OR $\times \epsilon 2$ freq + $\epsilon 3\epsilon 3$ OR $\times \epsilon 3\epsilon 3$ freq + $\epsilon 4$ OR $\times \epsilon 4$ freq; $0.88 \times 0.11 + 1 \times 0.71 + 1.1 \times 0.18 = 1.005$). Treating both variables as ordinal, we can fit a model between the two factors; in this case a quadratic model gave the best fit to the data. Assuming that all genotypes will have the same OR at some hypothetical level of the modifiable risk factor and considering that the overall OR between the *APOE* genotypes is dependent on their weighted difference for each category of diet intake, making use of the correspondence of the overall weighted mean to the $\epsilon 3\epsilon 3$ genotype, we can work out the beta coefficient for each of the other *APOE* genotypes. Using similar techniques we

computed the interaction components for all of the remaining risk factors.

Calculating risk

Despite OR overestimating risk for common diseases such as CVD [56,57], many researchers do use it to report their results. To approximate the relative risk (RR) from the OR we used the very popular methodology of Zhang and Yu [58] as being both the simplest and the one that requires the least information for the dataset used. Despite the method's slight inaccuracy in calculating confidence intervals and its inability to account for confounding factors, its balance between simplicity and precision make it an especially useful tool [56,57].

The total risk of an individual could be calculated as the product of the gender relative risk and all four modifiable risk factors, corresponding to the individual's specific genotype. We used a proportional hazard model to estimate the mortality schedule of an individual having relative risk R at a specific age x . According to this, if $\mu_1(x)$ is the hazard of death of genotype 1 at age x , and $\mu_2(x)$ is the hazard of death of a genotype 2 at the same age x , then $\mu_1(x) = R \times \mu_2(x)$, where R expresses the relative risk [29].

Simulating evolutionary change

Our simulated populations comprised individuals with randomly assigned genotype, gender and lifestyle parameters from a typical western population. The random draws were done using the MT19937 pseudorandom generator, a variation of the "Mersenne Twister" generator (GSL Reference Manual). Equal number of males and females were generated and subsequently paired as couples randomly. These couples reproduced to age 50 or until either of the individuals died. The reproductive schedule was the same as what is found in contemporary western populations, with a series of binomials distributions describing the chance of a successful birth for each year of life and the probability decreasing with increasing adult age. The age of death was obtained from a Gumbell distribution based on demographic tables for the UK (Office for National Statistics) and a relative risk corresponding to the life style parameters of the simulated individual. Each offspring was given a genotype based on the genotypes of its parents, and all the produced genotypes were pooled into a matrix constituting the initial conditions for the next generation. The process was repeated as many times as the number of generations required. The entire simulation was written in C++ and performed using a 16 CPUs Unix cluster.

Statistical analysis of the results

The Mathematica 4.1 package was used to collect analyse the output of the simulations. To minimize random variation of the results, a set of 60 runs, each comprising 200,000 individuals followed for 50 generations, was used. We summarized the data calculating the mean and standard deviation of all 60 runs for each generation and each allele. A random walk in one dimension was used to obtain the average change per generation, assuming that the change of the allele frequencies was linear and could be described by a simple equation. As long as the alleles do not have a pleiotropic effect balancing the selection for or against them, their frequencies in a stable environment were expected to increase until they were fixed or became extinct in the population. A simple equation describing the change could be written as:

$$\alpha_t = \alpha_{t-1} + d + \sum_t \quad (1)$$

where α_t is the frequency at time or generation t , α_{t-1} is frequency

in the previous generation, d is the change between two consecutive points in the series and Σ_i is random noise normally distributed. In order to calculate d we estimated a y_{jt} such that:

$$y_{jt} = \alpha_{j,t+1} - \alpha_{jt}$$

with simulation run $j = 1, \dots, 60$

and generation number $i = 1, \dots, 49$

returning a matrix of 49×60 elements. From this, the mean d and its 95% confidence intervals (95% CIs) could be calculated. If the 95% CIs are positive then there is evidence for a significant increase of the allele frequency with each generation, while if negative the allele goes towards extinction. In the case that 0 is included within the confidence interval, the change of frequency cannot be considered statistically significant and selection is either extremely weak or due to chance. The data were also transferred into Minitab statistical software to test for statistical significance between and within different variants of the simulation.

Results

Starting with *APOE* genotype frequencies of 1.7% $\epsilon_2\epsilon_2$, 11.6% $\epsilon_2\epsilon_3$, 55.8% $\epsilon_3\epsilon_3$, 1.9% $\epsilon_2\epsilon_4$, 25.1% $\epsilon_3\epsilon_4$, and 3.9% $\epsilon_4\epsilon_4$ and modifiable risk factor distributions as reported in Millen et al [44] and Stampfer et al [46], and assuming the same population distribution between males and females, the mean life expectancy at birth for the simulated individual was 75.7 years with females having 5.4 years longer expected lifespan than males. The differences in mortality between *APOE* genotype classes as fitted in the model can be seen in Figure 1, with ϵ_2 carriers living 1 (OR 0.9) year more, and ϵ_4 carriers living 1.2 (OR 1.13) years less, than the $\epsilon_3\epsilon_3$ reference genotype. In figure 2, the curves obtained when the interaction between genotype and diet was considered can be

seen, while table 1 has the calculated relative risk for each combination of diet and *APOE* genotype. For the force of natural selection on the *APOE* gene, we observed a positive selection for the ϵ_2 (mean frequency change per generation 3.352×10^{-4} , 95% CI 2.661×10^{-4} to 4.043×10^{-4}) and ϵ_3 (mean 2.785×10^{-4} , 95% CI 1.809×10^{-4} to 3.761×10^{-4}) alleles against ϵ_4 (mean -6.137×10^{-4} , 95% CI -6.946×10^{-4} to -5.327×10^{-4}), with the ϵ_2 frequency increasing slightly more rapidly than that for ϵ_3 , although there was no significant difference between them (Figure 3).

Considering that only a fraction of the population will suffer from *APOE* related diseases, a variant of the model with 30% of the total mortality attributed to CVD was also tested. Again, the results pointed to a positive selection of the ϵ_2 and ϵ_3 alleles (ϵ_2 mean 0.846×10^{-4} , 95% CI 0.497×10^{-4} to 1.195×10^{-4} ; ϵ_3 mean 0.846×10^{-4} , 95% CI 0.316×10^{-4} to 1.377×10^{-4}) and a negative selection for the ϵ_4 allele (mean -1.692×10^{-4} , 95% CI -2.140×10^{-4} to -1.244×10^{-4}) which, as expected, was approximately one third of the change observed previously.

We performed a number of simulation runs with a variety of different starting conditions both to test the sensitivity of the model and analyse the likely changes under different lifestyle choices. The description of all the simulation variants used with their mean life expectancy at birth can be seen in table 2. The changes in the distribution of the modifiable risk factors explored, except the eradication of smoking, were relatively modest and considered possible to occur, if not already present in certain subgroups of western populations. It is important to note the unexpected decrease in mean lifespan when heavy alcohol consumption is lowered and the slight increase in longevity when non- and very light drinkers decrease. This effect is due to the U-shaped relationship between alcohol consumption and health and the re-arrangement of the population distribution across the categories of risk.

The estimates for the mean frequency change for each *APOE* allele per generation, under all the different models considered assuming at the same time that only 30% of the population is affected by any *APOE* related diseases, are presented in table 3. As

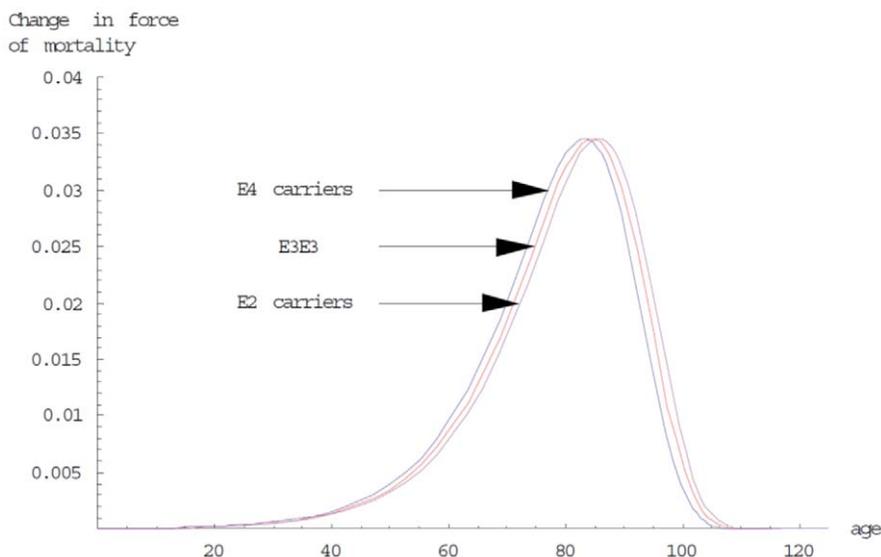


Figure 1. Differences in the distribution of mortality for the three *APOE* genotype classes. These are estimated using a proportional hazard model as described in the methods. The ϵ_2 carriers have an OR of 0.9 and survive a year more than the $\epsilon_3\epsilon_3$ carriers. In contrast, ϵ_4 carriers have a decreased lifespan of 1.2 years due to an OR of 1.13 compared to the reference genotype. doi:10.1371/journal.pone.0010022.g001

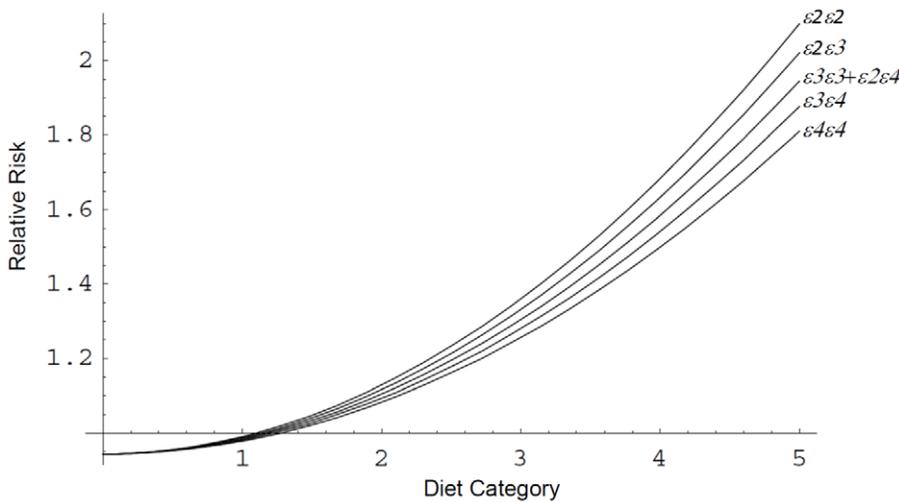


Figure 2. Inferred interaction between APOE genotype and diet. Diet categories are considered on an integer scale from 1 to 5 with Heart Healthy, Light Eating, Wine and Moderate Eating, High Fat, and Empty Calorie clusters [44,45]. We assumed that for a hypothetical diet category 0 there is no distinction in risk between the genotypes.
doi:10.1371/journal.pone.0010022.g002

can be seen in the table, the selection against the $\epsilon 4$ allele and the increase in the frequency of the $\epsilon 3$ and $\epsilon 2$ alleles are robust and remained significant under all the different versions of the simulation. A notable exception was the loss of significance for the increase of the $\epsilon 3$ frequency when the effect of smoking is removed from the population (mean 0.259×10^{-4} , 95% CI -0.057×10^{-4} to 0.574×10^{-4}). The extreme change of removing all smokers in the population, while in the same time leaving the percentage of CVD associated mortality at 30%, reaches the statistical power limits of our current simulation to observe significance for the selection of the $\epsilon 3$ allele in a population of 20,000. For such extreme cases, a corresponding change should also be made to the mortality schedule and the fraction of the population affected by that change.

Discussion

We have shown that under certain environmental conditions, such as those often found in Western populations, the APOE gene is likely to be under the action of natural selection. According to the results obtained, the $\epsilon 2$ and $\epsilon 3$ alleles are increasing with each successive generation at the expense of the $\epsilon 4$ allele, which is slowly being removed from the population. The selection against the $\epsilon 4$ allele was found to be robust, despite changes to the initial

conditions of the simulation. We failed to find any significant difference between the increase in frequency of the $\epsilon 2$ and $\epsilon 3$ alleles, denoting their very similar effects on survival. This accords with meta-analysis studies which have found it difficult to identify any statistically significant risk differences between the carriers of $\epsilon 2$ allele and the reference $\epsilon 3\epsilon 3$ genotype, unless a large number of cases is available, mainly due to the low frequency of $\epsilon 2$ in the populations considered [15].

Despite the difficulty in observing and measuring the force of natural selection as revealed by change of allele frequencies in specific polymorphisms, especially in humans, alternative methods are available to obtain evidence for its action indirectly. APOE seems to be one of the least variable human genes studied, despite the average neutral mutation rate [59]. Comparing the chimpanzee and human APOE genes, revealed that $\epsilon 4$ is probably the ancestral allele [59]. Indeed, most great apes carry only the $\epsilon 4$ allele, although a change of arginine to threonine at position 61 gives this isoform an affinity similar to the human $\epsilon 3$ [40,60,61]. It is unclear when the human $\epsilon 4$ allele arose, but a comparison between coding changes in orthologous genes in five species (Mouse, Rat, Dog, Chimpanzee and Human) showed positive selection for APOE along the hominid lineage [62]. Using the haplotypes defining the three genotypes in a coalescent model, Fullerton et al [59] proposed that $\epsilon 3$ diverged from the $\epsilon 4$ haplotype around 200,000 years ago and that $\epsilon 3$ is showing evidence of increasing in frequency, relative to $\epsilon 4$. Assuming a constant selective pressure and a generation time of ~ 20 years, given an $\epsilon 3$ frequency of around 0.75, a very simple calculation shows that the required increase of $\epsilon 3$ per generation is 7.5×10^{-5} , close to the order of magnitude predicted from the simulation for the current change of the allele frequency. Using a similar argument, we can further propose that since $\epsilon 2$ and $\epsilon 3$ alleles seem to be under a similar pressure from natural selection in our simulation and $\epsilon 2$ has a lower frequency in human populations, $\epsilon 2$ arose later than $\epsilon 3$. Again, analysis of APOE haplotypes using a reduced median network revealed that $\epsilon 2$ was derived from the $\epsilon 3$ allele within the last 80,000 years [59].

Assuming a simple and uniform selection pressure on APOE is rather over simplistic. Current theories suggest a two step evolution of the locus, one to explain the early evolution of the

Table 1. Calculated relative risk for each combination of diet category as described by Miller et al (2001) and APOE genotype.

APOE genotype	22	23	33	34	44
Hypothetical diet 0	0.944	0.944	0.944	0.944	0.944
Heart Healthy	0.979	0.981	0.984	0.987	0.990
Light Eating	1.083	1.093	1.104	1.116	1.129
Wine and Moderate Eating	1.256	1.280	1.304	1.332	1.360
High Fat	1.499	1.541	1.584	1.633	1.683
Empty Calorie	1.811	1.877	1.944	2.020	2.099

doi:10.1371/journal.pone.0010022.t001

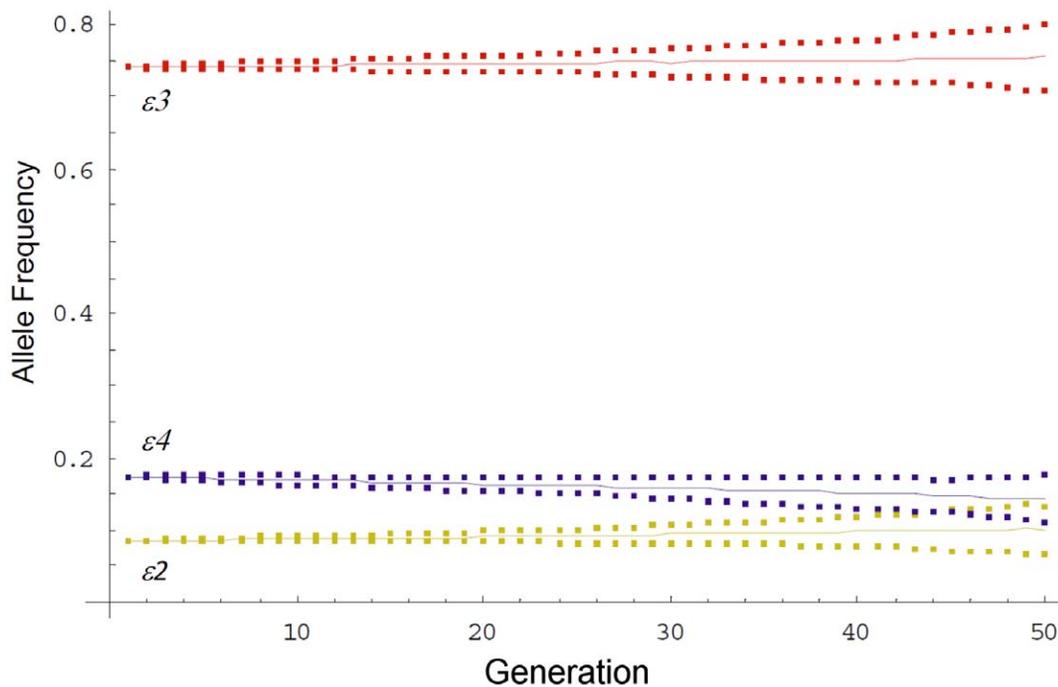


Figure 3. Change in frequency of the three *APOE* alleles with each generation. The long dash line represents the $\epsilon 3$ allele, the solid line the $\epsilon 4$ and the short dash line the $\epsilon 2$. The squares, stars and triangles points represent the 95% Confidence Intervals of the $\epsilon 3$, $\epsilon 4$ and $\epsilon 2$ frequencies respectively.

doi:10.1371/journal.pone.0010022.g003

human alleles and a second step to account for their current global frequencies. Despite the very sparse evidence for both, and their rather conflicting views for the selection pressures applied, the proposed mechanisms are interesting and reveal the importance of *APOE* in human evolution. Finch and Stanford [63] and Finch and Sapolsky [60] suggested that *APOE* is a “meat adaptive” gene that permitted increased consumption of animal tissue during hominid evolution while conferring resistance to the associated risks such as hypercholesterolemia and infections, allowing for the extension of human lifespan. In their view, this increase in meat eating

provided humans with a solution to seasonal deficits of micronutrients. In addition, relatively safe consumption of nutrient dense meat is believed to have led to an increase of body size, without restricting activity, and conferred the extra energy required for the cerebral expansion of early humans [64,65]. On the other hand, the frequencies of the three common alleles vary widely between populations of different ethnic backgrounds. In general $\epsilon 3$ is globally the most common allele, with Africans having the lowest allele frequency ($\leq 70\%$) compared to Caucasians (70–80%) and East Asians ($\geq 80\%$) [66]. In Europe, and between populations of Caucasian descent, there is a north-to-south gradient of decreasing $\epsilon 4$ frequency opposite to the spread of agriculture [67,68]. This has led Corbo and Scacchi [69] to suggest that *APOE* is a thrifty gene, with the $\epsilon 4$ allele being advantageous under seasonal periods of starvation, due to its elevating effect of cholesterol which would otherwise be too low, but detrimental in areas where carbohydrates were readily available, such as those with a long history of agriculture, thus the north-to-south pattern observed in Europe. Doubts have been raised about the lack of seasonal starvation in food producing societies [70], while the hypothesised selection for $\epsilon 4$ in pre-agricultural times is opposite to the one suggested by Finch and Sapolsky [60,63].

Although it is clear that variation in the *APOE* locus has been, and probably still is, under the control of selection, the fact that its main effects are beyond the reproductive lifespan of humans, has posed an interesting puzzle for the mechanisms through which selection is applied. Finch and Sapolsky [60,71] suggested that the spread of $\epsilon 3$ in human populations is due to the effect of a mother’s late survival on the fitness of her offspring. The grand-mother hypothesis, as it is known, is centred on the idea that the mother-child food sharing seen among hunter-gatherers may allow post-reproductive grandmothers to enhance their daughters’ fertility, thus elevating their own fitness and increasing the selection for

Table 2. The simulated average lifespan in years of each variant considered in the absence of non APO E related mortality.

Run	Parameters	Mean Lifespan
Default	Default	75.69
Default 2	Default values with 30% CVD associated mortality	75.69
Diet 1	Heart Healthy and Light Eating groups decreased by 20%	74.93
Diet 2	High Fat and Empty Calorie groups decreased by 20%	75.98
Alcohol 1	Light and non-drinkers decreased by 25%	75.97
Alcohol 2	Heavy drinkers decrease by 25%	75.54
Exercise 1	Entire population moderately active (2.5h/week)	76.19
Exercise 2	Most active groups decreased by 25%	75.41
Smoking 1	Non-smokers decreased by 25%	75.57
Smoking 2	No smokers in the population	82.95

doi:10.1371/journal.pone.0010022.t002

Table 3. Mean change of the frequency of each allele for the variants of the model used together with the 95% CI of the standard error of the mean in each case.

Run	APO E allele	Mean change of allele frequency per generation	95% confidence intervals for the mean	
Default	$\epsilon 2$	0.0003352	0.0002661	0.0004043
	$\epsilon 3$	0.0002785	0.0001809	0.0003761
	$\epsilon 4$	-0.0006137	-0.0006946	-0.0005327
Default 2	$\epsilon 2$	0.0000846	0.0000497	0.0001195
	$\epsilon 3$	0.0000846	0.0000316	0.0001377
	$\epsilon 4$	-0.0001692	-0.0002140	-0.0001244
Diet 1	$\epsilon 2$	0.0000900	0.0000547	0.0001252
	$\epsilon 3$	0.0000899	0.0000360	0.0001438
	$\epsilon 4$	-0.0001798	-0.0002257	-0.0001340
Diet 2	$\epsilon 2$	0.0000809	0.0000480	0.0001139
	$\epsilon 3$	0.0000643	0.0000130	0.0001157
	$\epsilon 4$	-0.0001453	-0.0001905	-0.0001000
Alcohol 1	$\epsilon 2$	0.0000836	0.0000524	0.0001148
	$\epsilon 3$	0.0000818	0.0000326	0.0001309
	$\epsilon 4$	-0.0001654	-0.0002081	-0.0001226
Alcohol 2	$\epsilon 2$	0.0000847	0.0000500	0.0001195
	$\epsilon 3$	0.0000588	0.0000074	0.0001103
	$\epsilon 4$	-0.0001436	-0.0001880	-0.0000991
Exercise 1	$\epsilon 2$	0.0000895	0.0000555	0.0001235
	$\epsilon 3$	0.0000629	0.0000108	0.0001149
	$\epsilon 4$	-0.0001524	-0.0001959	-0.0001088
Exercise 2	$\epsilon 2$	0.0001088	0.0000749	0.0001427
	$\epsilon 3$	0.0000637	0.0000115	0.0001158
	$\epsilon 4$	-0.0001725	-0.0002171	-0.0001279
Smoking 1	$\epsilon 2$	0.0001283	0.0000897	0.0001668
	$\epsilon 3$	0.0000974	0.0000383	0.0001564
	$\epsilon 4$	-0.0002256	-0.0002766	-0.0001746
Smoking 2	$\epsilon 2$	0.0000406	0.0000205	0.0000607
	$\epsilon 3$	0.0000259	-0.0000057	0.0000574
	$\epsilon 4$	-0.0000665	-0.0000936	-0.0000393

doi:10.1371/journal.pone.0010022.t003

long postmenopausal lifespan [72,73,74]. Finch and Sapolsky [60,71] argued that this evolutionary advantage will have caused selection for an *APOE* allele that will have delayed neuropathology and mortality even if its effects are evident later in life (for counterarguments see [75,76]). Charlesworth [77], on the other hand, proposed that *APOE* is an example of balanced polymorphism with the variation in the locus maintained due to the antagonistic effects of the alleles, where increased late life risk is offset by advantages in younger ages, in accordance with the antagonistic pleiotropy theory of ageing. Martin [76] extended the idea of antagonistic pleiotropy for *APOE* suggesting that the $\epsilon 4$ allele could be advantageous in cases of infections with pathogens requiring host lipids for survival. Parasites like *Trypanosome brucei* not able to carry out de novo liposynthesis may depend upon host LDL for acquisition of lipids, with the high affinity of $\epsilon 4$ for some lipids actually hindering their uptake by the pathogen. Our model indicates that, although these hypotheses might work at a secondary level, they are not strictly necessary to explain the selection of *APOE* alleles. The *direct* effect of *APOE* on the mortality

or morbidity of the population is sufficient, at least in contemporary Western populations, to produce a selection differential between the three alleles. The simulation predicts that the $\epsilon 3$ and $\epsilon 2$ alleles are driven towards fixation, a conclusion supported by the Fullerton et al [59] observations for the mutation at site 3937, characterising the $\epsilon 3$ allele.

This simulation study is considering the pressure of natural selection on a population with demographics equal, or similar, to what is found in contemporary populations. It will be wrong to extrapolate our results in order to interpret the early evolution of *APOE*, since modern populations experience a very different environment compared to our early ancestors. Nevertheless, our basic conclusion that certain genetic polymorphisms, such as *APOE*, exhibiting their main effect later in life can still be under the action of natural selection, still holds. Interestingly, analysis of a Danish cohort of individual born between 1895 and 1899, showed that carriers of the $\epsilon 4$ allele had increased early mortality compared to those carrying the $\epsilon 2$ allele as suggested by our simulation [78]. It is currently unclear how far back our results can

be considered as valid, but given the early age (>35 years of age) that the *APOE* effects on mortality can be seen in the Danish cohort, we believe that this can extend in our recent history.

In our simulation we cannot account for any pleiotropic effects of *APOE*. We described the effects of the gene on lifespan only through its relation to elevated risk for CVD. *APOE* has a much wider role than just lipid metabolism, including immunoregulation and susceptibility to infections [60,79,80,81]. We chose not to include these effects within the simulation since reliable quantitative data concerning their impact on lifespan do not exist. *APOE*, though its role in steroidogenesis, has also been shown to influence fertility in pre-industrial populations [82], old Italians (>80 years of age) [83] and North European males [84], with the $\epsilon 2$ carriers having less children. The small number of $\epsilon 2$ subjects together with the many cultural and socioeconomic factors affecting current human reproduction makes it difficult to say if this effect is indeed real and relevant in the model used. A limitation of our model was the lack of data for the interaction between the *APOE* genotypes and each of the modifiable risk factors considered, which forced us to use a very approximate method to calculate the unknown interactions. Further epidemiological work is required to elucidate the precise quantitative relationship of the *APOE* gene with the parameters of the model so that the predictions can be made more accurate. While, for simplicity, we assumed that the rest of the risk factors are independent of each other, although it has been recognised that these tend to cluster, to some degree, in individuals of high risk [85].

Despite some limitations, our results indicate that although the main effect of *APOE* is seen after the end of the reproductive lifespan of humans, the relatively rare events of early mortality or

morbidity are enough to produce selection against the $\epsilon 4$ allele. Considering these findings in the light of the evolutionary theory of ageing and longevity, particularly as regards the ‘mutation accumulation’ of late acting deleterious alleles, we can identify an important perspective on this concept. Since most longevity-associated genes, such as *APOE*, will cause a distribution around a mean for the corresponding mortality, evolution will rarely, if ever, manage to push deleterious gene action entirely beyond the reach of selection, although as envisaged by Medawar there will be a continual selection pressure to postpone such action to later and later ages. The gradual postponement of the detrimental gene effect will take place simultaneously in a wide array of mortality associated genes. Any single mortality distribution exposed significantly more often to the action of selection will be moved towards older ages faster, until it reaches a balance with all the other detrimental genotypes. This process will continue to drive the longevity of the organism until the age-related mortality becomes non-significant in relation to the age-independent mortality. In this respect the evolution of longevity, although based on selection of individual genes, is more of an evolutionarily coordinated transfiguration of the whole genome to achieve the maximum lifespan in the given conditions as hypothesized by Hamilton [38].

Author Contributions

Conceived and designed the experiments: FD TBLK. Performed the experiments: FD. Analyzed the data: FD. Contributed reagents/materials/analysis tools: FD. Wrote the paper: FD TBLK.

References

- Perls TT, Bubrick E, Wager CG, Vijg J, Kruglyak L (1998) Siblings of centenarians live longer. *Lancet* 351: 1560–1560.
- Hjelmborg JV, Iachine I, Skytthe A, Vaupel JW, McGue M, et al. (2006) Genetic influence on human lifespan and longevity. *Human Genetics* 119: 312–321.
- Herskind AM, McGue M, Holm NV, Sorensen TIA, Harvald B, et al. (1996) The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics* 97: 319–323.
- Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL (1998) The effect of genetic factors for longevity: A comparison of identical and fraternal twins in the Swedish twin registry. *Journals of Gerontology Series A-Biological Sciences and Medical Sciences* 53: M441–M446.
- Cournil A, Legay JM, Schachter F (2000) Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18–20th centuries. *Proceedings of the Royal Society of London Series B-Biological Sciences* 267: 1021–1025.
- Gudmundsson H, Gudbjartsson DF, Kong A, Gudbjartsson H, Frigge M, et al. (2000) Inheritance of human longevity in Iceland. *European Journal of Human Genetics* 8: 743–749.
- Skytthe A, Pedersen NL, Kaprio J, Stazi MA, Hjelmborg JVB, et al. (2003) Longevity studies in GenomEUtwin. *Twin Research* 6: 448–454.
- Wernette-Hammond CE (1989) Glycosylation of human apolipoprotein E. The carbohydrate attachment site is threonine 194. *The Journal of biological chemistry* 264: 9094–9101.
- Hagberg JM, Wilund KR, Ferrell RE (2000) APO E gene and gene-environment effects on plasma lipoprotein- lipid levels. *Physiological Genomics* 4: 101–108.
- van Bockxmeer FM (1994) Apoe and Ace genes - Impact on human longevity. *Nature Genetics* 6: 4–5.
- Smith JD (2002) Apolipoproteins and aging: emerging mechanisms. *Ageing Research Reviews* 1: 345–365.
- Rall SC, Jr., Weisgraber KH, Mahley RW (1982) Human apolipoprotein E. The complete amino acid sequence. *Journal of Biological Chemistry* 257: 4171–4178.
- Utermann G, Hardewig A, Zimmer F (1984) APOLIPOPROTEIN-E PHENOTYPES IN PATIENTS WITH MYOCARDIAL-INFARCTION. *Human Genetics* 65: 237–241.
- Song YQ, Stampfer MJ, Liu SM (2004) Meta-analysis: Apolipoprotein E genotypes and risk for coronary heart disease. *Annals of Internal Medicine* 141: 137–147.
- Bennet AM, Di Angelantonio E, Yc Z, Wensley F, Dahlin A, et al. (2007) Association of apolipoprotein E genotypes with lipid levels and coronary risk. *Jama-Journal of the American Medical Association* 298: 1300–1311.
- Zhou WD, Xu D, Peng XX, Zhang QH, Jia JP, et al. (2008) Meta-analysis of APOE4 allele and outcome after traumatic brain injury. *Journal of Neurotrauma* 25: 279–290.
- Dong LM, Wilson C, Wardell MR, Simmons T, Mahley RW, et al. (1994) Human Apolipoprotein-E - Role of Arginine-61 in Mediating the Lipoprotein Preferences of the E3-Isoform and E4-Isoform. *Journal of Biological Chemistry* 269: 22358–22365.
- Marz W, Hoffmann MM, Scharnagl H, Fisher E, Chen M, et al. (1998) Apolipoprotein E2 (Arg136→Cys) mutation in the receptor binding domain of apoE is not associated with dominant type III hyperlipoproteinemia. *Journal of Lipid Research* 39: 658–669.
- Henneman P, van der Smant-de Beer F, Moghaddam PH, Huijts P, Stalenhoef AFH, et al. (2009) The expression of type III hyperlipoproteinemia: involvement of lipolysis genes. *European Journal of Human Genetics* 17: 620–628.
- Schachter F, Fauredelanef L, Guenot F, Rouger H, Froguel P, et al. (1994) Genetic associations with human longevity at the Apoe and Ace loci. *Nature Genetics* 6: 29–32.
- Blanche H, Cabane L, Sahbatou M, Thomas G (2001) A study of French centenarians: are ACE and APOE associated with longevity? *Comptes Rendus De L Academie Des Sciences Serie III-Sciences De La Vie-Life Sciences* 324: 129–135.
- Rea IM, Mc Dowell I, McMaster D, Smye M, Stout R, et al. (2001) Apolipoprotein E alleles in nonagenarian subjects in the Belfast Elderly Longitudinal Free-living Ageing Study (BELFAST). *Mechanisms of Ageing and Development* 122: 1367–1372.
- Frisoni GB, Louhija J, Geroldi C, Trabucchi M (2001) Longevity and the epsilon 2 allele of apolipoprotein E: The Finnish Centenarians Study. *Journals of Gerontology Series A - Biological Sciences and Medical Sciences* 56: M75–M78.
- Lewis SJ, Brunner EJ (2004) Methodological problems in genetic association studies of longevity - the apolipoprotein E gene as an example. *International Journal of Epidemiology* 33: 962–970.
- Wilson PWFMD, Myers RHP, Larson MGS, Ordovas JMP, Wolf PAMD, et al. (1994) Apolipoprotein E Alleles, Dyslipidemia, and Coronary Heart Disease: The Framingham Offspring Study. *Jama-Journal of the American Medical Association* 272: 1666–1671.

26. Lahoz C, Schaefer EJ, Cupples LA, Wilson PWF, Levy D, et al. (2001) Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. *Atherosclerosis* 154: 529–537.
27. Stengard JH, Pekkanen J, Ehnholm C, Nissinen A, Sing CF (1996) Genotypes with the apolipoprotein epsilon 4 allele are predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men. *Human Genetics* 97: 677–684.
28. Wilson PWF, Schaefer EJ, Larson MG, Ordovas JM (1996) Apolipoprotein E alleles and risk of coronary disease - A meta-analysis. *Arteriosclerosis Thrombosis and Vascular Biology* 16: 1250–1255.
29. Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW (2000) Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: Apolipoprotein E gene is a “frailty gene”, not a “longevity gene”. *Genetic Epidemiology* 19: 202–210.
30. Kolovou G, Yiannakouris N, Hatzivassiliou M, Malakos J, Daskalova D, et al. (2002) Association of apolipoprotein E polymorphism with myocardial infarction in Greek patients with coronary artery disease. *Current Medical Research and Opinion* 18: 118–124.
31. Strittmatter WJ, Saunders AM, Schmechel D, Pericakvance M, Enghild J, et al. (1993) APOLIPOPROTEIN-E - HIGH-AVIDITY BINDING TO BETA-AMYLOID AND INCREASED FREQUENCY OF TYPE-4 ALLELE IN LATE-ONSET FAMILIAL ALZHEIMER-DISEASE. *Proceedings of the National Academy of Sciences of the United States of America* 90: 1977–1981.
32. Saunders AM, Strittmatter WJ, Schmechel D, Georgehslop PHS, Pericakvance MA, et al. (1993) ASSOCIATION OF APOLIPOPROTEIN-E ALLELE EPSILON-4 WITH LATE-ONSET FAMILIAL AND SPORADIC ALZHEIMERS-DISEASE. *Neurology* 43: 1467–1472.
33. Roses AD (1997) Apolipoprotein E, a gene with complex biological interactions in the aging brain. *Neurobiology of Disease* 4: 170–185.
34. Huang YD (2006) Apolipoprotein E and Alzheimer disease. *Neurology* 66: S79–S85.
35. de Carli CMD, Miller BLMD, Swan GEP, Reed TP, Wolf PAMD, et al. (2001) Cerebrovascular and Brain Morphologic Correlates of Mild Cognitive Impairment in the National Heart, Lung, and Blood Institute Twin Study. *Archives of Neurology* April 58: 643–647.
36. Kirkwood TBL, Austad SN (2000) Why do we age? *Nature* 408: 233–238.
37. Medawar PB (1952) *An Unsolved Problem in Biology*: Lewis London.
38. Hamilton WD (1966) The moulding of senescence by natural selection. *Journal of Theoretical Biology* 12: 12–45.
39. Charlesworth B (1994) *Evolution in Age-structured Populations*. Cambridge: Cambridge University Press.
40. Hanlon CS, Rubinsztein DC (1995) Arginine Residues at Codon-112 and Codon-158 in the Apolipoprotein-E Gene Correspond to the Ancestral State in Humans. *Atherosclerosis* 112: 85–90.
41. Mahley RW, Rall SC (1999) Is epsilon 4 the ancestral human apoE allele? *Neurobiology of Aging* 20: 429–430.
42. Petersen S, Rayner M (2002) *Coronary Heart Statistics Foundation BH*, ed. London, UK: Department of Public Health.
43. Panagiotakos DB, Pitsavos C, Chrysohou C, Stefanadis C, Toutouzas P (2002) Risk stratification of coronary heart disease in Greece: Final results from the CARDIO2000 epidemiological study. *Preventive Medicine* 35: 548–556.
44. Millen BE, Quatromoni PA, Copenhaver DL, Demissie S, O'Horo CE, et al. (2001) Validation of a dietary pattern approach for evaluating nutritional risk: The Framingham Nutrition Studies. *Journal of the American Dietetic Association* 101: 187–194.
45. Millen BE, Quatromoni PA, Nam BH, O'Horo CE, Polak JF, et al. (2002) Dietary patterns and the odds of carotid atherosclerosis in women: The framingham nutrition studies. *Preventive Medicine* 35: 540–547.
46. Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC (2000) Primary prevention of coronary heart disease in women through diet and lifestyle. *New England Journal of Medicine* 343: 16–22.
47. Ordovas JM (1999) The genetics of serum lipid responsiveness to dietary interventions. *Proceedings of the Nutrition Society* 58: 171–187.
48. Ordovas JM, LopezMiranda J, Mata P, PerezJimenez F, Lichtenstein AH, et al. (1995) Gene-diet interaction in determining plasma lipid response to dietary intervention. *Atherosclerosis* 118: S11–S27.
49. Masson LF, McNeill G, Avenell A (2003) Genetic variation and the lipid response to dietary intervention: a systematic review. *American Journal of Clinical Nutrition* 77: 1098–1111.
50. Kauma H, Savolainen MJ, Rantala AO, Lilja M, Kervinen K, et al. (1998) Apolipoprotein E phenotype determines the effect of alcohol on blood pressure in middle-aged men. *American Journal of Hypertension* 11: 1334–1343.
51. Puddey IB, Rakic V, Dimmitt SB, Burke V, Beilin LJ, et al. (1999) Apolipoprotein E genotype and the blood pressure raising effect of alcohol. *American Journal of Hypertension* 12: 946–947.
52. Corella D, Tucker K, Lahoz C, Coltell O, Cupples LA, et al. (2001) Alcohol drinking determines the effect of the APOE locus on LDL-cholesterol concentrations in men: the Framingham Offspring Study. *American Journal of Clinical Nutrition* 73: 736–745.
53. Humphries SE, Talmud PJ, Hawe E, Bolla M, Day INM, et al. (2001) Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study. *Lancet* 358: 115–119.
54. van der Meer IM, Witterman JCM (2002) Apolipoprotein E genotype, smoking and cardiovascular disease. *Journal of Hypertension* 20: 2327–2329.
55. Bernstein MS, Costanza MC, James RW, Morris MA, Cambien F, et al. (2002) Physical activity may modulate effects of ApoE genotype on lipid profile. *Arteriosclerosis Thrombosis and Vascular Biology* 22: 133–140.
56. McNutt LA, Wu CT, Xue XN, Hafner JP (2003) Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* 157: 940–943.
57. Robbins AS, Chao SY, Fonseca VP (2002) What's the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *Annals of Epidemiology* 12: 452–454.
58. Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Jama-Journal of the American Medical Association* 280: 1690–1691.
59. Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, et al. (2000) Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *American Journal of Human Genetics* 67: 881–900.
60. Finch CE, Sapolsky RM (1999) The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. *Neurobiology of Aging* 20: 407–428.
61. Raffai RL, Dong LM, Farese RV, Weisgraber KH (2001) Introduction of human apolipoprotein E4 “domain interaction” into mouse apolipoprotein E. *Proceedings of the National Academy of Sciences of the United States of America* 98: 11587–11591.
62. Vamathevan J, Hasan S, Emes R, Amrine-Madsen H, Rajagopalan D, et al. (2008) The role of positive selection in determining the molecular cause of species differences in disease. *BMC Evolutionary Biology* 8: 273.
63. Finch CE, Stanford CB (2004) Meat-adaptive genes and the evolution of slower aging in humans. *Quarterly Review of Biology* 79: 3–50.
64. Milton K (1999) A hypothesis to explain the role of meat-eating in human evolution. *Evolutionary Anthropology* 8: 11–21.
65. Milton K. The critical role played by animal source foods in human (Homo) evolution; 2002 Jun 24–26. Washington, D.C.: Amer Inst Nutrition, 3886S–3892S.
66. Gerdes LU, Klausen IC, Sihm I, Faergeman O (1992) APOLIPOPROTEIN-E POLYMORPHISM IN A DANISH POPULATION COMPARED TO FINDINGS IN 45 OTHER STUDY POPULATIONS AROUND THE WORLD. *Genetic Epidemiology* 9: 155–167.
67. Gerdes LU, Gerdes C, Hansen PS, Klausen IC, Faergeman O, et al. (1996) The apolipoprotein E polymorphism in Greenland Inuit in its global perspective. *Human Genetics* 98: 546–550.
68. Corbo RM, Scacchi R, Mureddu L, Mulas G, Alfano G (1995) APOLIPOPROTEIN-E POLYMORPHISM IN ITALY INVESTIGATED IN NATIVE PLASMA BY A SIMPLE POLYACRYLAMIDE-GEL ISOELECTRIC-FOCUSING TECHNIQUE - COMPARISON WITH FREQUENCY DATA OF OTHER EUROPEAN POPULATIONS. *Annals of Human Genetics* 59: 197–209.
69. Corbo RM, Scacchi R (1999) Apolipoprotein E (APOE) allele distribution in the world. Is APOE 4 a “thrifty” allele? *Annals of Human Genetics* 63: 301–310.
70. Prentice AM, Rayco-Solon P, Moore SE. Insights from the developing world: thrifty genotypes and thrifty phenotypes; 2004 Jul 05–08. Dublin, IRELAND: Cabi Publishing, 153–161.
71. Sapolsky RM, Finch CE (2000) Alzheimer's disease and some speculations about the evolution of its modifiers. *Alzheimer's Disease: A Compendium of Current Theories*. pp 99–103.
72. Hawkes K (2004) Human longevity - The grandmother effect. *Nature* 428: 128–129.
73. Hawkes K. Grandmothers and the evolution of human longevity; 2002 Apr. BuffaloNy: Wiley-Liss, 380–400.
74. Hawkes K, O'Connell JF, Jones NGB, Alvarez H, Charnov EL (1998) Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences of the United States of America* 95: 1336–1339.
75. Kennedy GE (2003) Palaeolithic grandmothers? Life history theory and early Homo. *Journal of the Royal Anthropological Institute* 9: 549–572.
76. Martin GM (1999) APOE alleles and lipophylic pathogens. *Neurobiology of Aging* 20: 441–443.
77. Charlesworth B (1996) Evolution of senescence: Alzheimer's disease and evolution. *Current Biology* 6: 20–22.
78. Ewbank DC (2004) The APOE gene and differences in life expectancy in Europe. *Journal of Gerontology Series a-Biological Sciences and Medical Sciences* 59: 16–20.
79. Mahley RW, Rall SC (2000) Apolipoprotein E: Far more than a lipid transport protein. *Annual Review of Genomics and Human Genetics* 1: 507–537.
80. Mahley RW, Weisgraber KH, Huang YD (2009) Apolipoprotein E: structure determines function, from atherosclerosis to Alzheimer's disease to AIDS. *Journal of Lipid Research* 50: S183–S188.
81. Drenos F, Talmud PJ, Casas JP, Smeeth L, Palmen J, et al. (2009) Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk. *Human Molecular Genetics* 18: 2305–2316.
82. Corbo RM, Ulizzi L, Scacchi R, Martinez-Labarga C, De Stefano GF (2004) Apolipoprotein E polymorphism and fertility: a study in pre-industrial populations. *Molecular Human Reproduction* 10: 617–620.
83. Corbo RM, Scacchi R, Cresta M (2004) Differential reproductive efficiency associated with common apolipoprotein E alleles in postreproductive-aged subjects. *Fertility and Sterility* 81: 104–107.

84. Gerdes LU, Gerdes C, Hansen PS, Klausen IC, Faergeman O (1996) Are men carrying the apolipoprotein epsilon 4- or epsilon 2 allele less fertile than epsilon 3 epsilon 3 genotypes? *Human Genetics* 98: 239–242.
85. Twisk JWR, Kemper HCG, van Mechelen W, Post GB (2001) Clustering of risk factors for coronary heart disease: The longitudinal relationship with lifestyle. *Annals of Epidemiology* 11: 157–165.