Testing two evolutionary theories of ageing by using public genome-wide data

Juan Antonio Rodríguez Pérez

TESI DOCTORAL UPF / 2016

DIRECTORS DE LA TESI

Dra. Elena Bosch Fusté

Dr. Arcadi Navarro Cuartiellas

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

upf. Universitat Pompeu Fabra Barcelona

"Genetics play a huge part in who we are.
But we also have free will."

Aidan Quinn (American actor, b. 1959)

A Carlos, Josefa y Lela.

Acknowledgments

Desde que pisé el departamento de Biología Evolutiva de la UPF y sostuve las tesis de anteriores estudiantes entre las manos, siempre había soñado con el momento de escribir los agradecimientos. Sería un momento mágico que clamaba un "¡se acabó!": sonaría música triunfalmente épica, todo se movería a cámara lenta; respiro aliviado, caminando sin mirar atrás, como en las películas, mientras una explosión lo manda todo al carajo y me quito las gafas de sol al viento. E iría plasmando mis memorias de color arco iris de estos últimos 4 años y pico en un par de páginas en las que mencionaría todo lo que os quiero y toda esa

Ese momento ha llegado. No tengo ni puñetera idea de lo que hay que escribir, ni color rosa ni arco iris, ni Pantone 448 C. Mariano Rajoy habla de sabe-dios-qué en la radio, todo se mueve a mil por hora, y respiro como un asmático gordo saltando vallas, para poder imprimir y depositar a tiempo un petate de doscientas y pico hojas que tenga un mínimo de coherencia.

A pesar de todo esto, este asmático gordo puede hacer un esfuerzo para acordarse de gentes y demás.

En primer lugar, gracias a Elena y a Arcadi, el *seny* i la *rauxa* (ya escogéis vosotros quién es quién), por la oportunidad de trabajar con vosotros y desarrollar este trabajo. También por todo lo que me habéis enseñado y lo que he aprendido, no sólo en el ámbito científico. Gracias, de verdad. Ojalá que dure.

A Urko, por enseñarme que en esta vida más allá de los GWAS hay tortugas apareándose y chimpancés que hablan; así como también agradecer sus enseñanzas neoliberales.

Al 419 y a todos los que pasaron por allí. Desde Ferrán cuando todavía no mandaba, a Josephine por sus Stormtroopers-waffels o esa cosa tan deliciosa. Sin olvidar a Óscar Ramírez y en especial a David Hughes, for being such a good man, in every possible aspect.
Mención especial para el núcleo duro del 419; Marina, Nino y Diego por compartir ricas horas de discusión sobre la dulce y divertida nada.
A Nino; por ser compañero de viaje doctoral de inicio a fin.

A Marina; porque algún día abandonará el Perl.
Y a Diego, por enseñarme pedalear; consciente e inconscientemente.

Al Taper Club, por ser más que un club. No me convence mucho cómo suena eso, pero se entiende, y es cierto.

A Tara, por iluminar una sosa portada con tu horrípido arte. Hubiera preferido un limón. Xa paparemos unha mariscada das boas coma dous pepes. E con rape e caixa de bombós.

Para ti, Anna, que te la has metido toda.

A las compis, Vicky y Vero, por haberme abierto la puerta hace ahora dos años. ¡Y qué dos años!

A Elisa, sobre todo a Elisa. Sin ella, esto no sería así.

A los cadáveres que se han quedado por el camino. Esta tesis también contiene átomos vuestros.

Y a Clàudia por ser compañera de aventuras, y de algo más, durante este último año.

En los artículos científicos, los últimos autores siempre son los mejores.
Aquí va la familia, en especial madre y madrina porque siempre creyeron, incluso más que yo, que esto se haría realidad.
Al tío Juan. Por guiarme siempre, dentro y fuera de lo académico.
Y gracias, Hilda, por acompañarme y cuidarme unos años en esta ciudad.

Y a ti, querido lector, que si no calculo mal, no pasarás la página. Aunque espero equivocarme en, por lo menos, tres ocasiones.

Juan A. Rodríguez

Barcelona, 15 de Septiembre 2016

Imagen de cubierta:
Taracido, Adrián. *Fragmentos de un cuerpo desechado* (Detalle). 2015 ©

Abstract

Old age comes coupled with frailty and disease and, thus, the ageing of the World's population has spurred the interest on the causes and mechanisms of senescence. Senescence has long been a mystery, with no single universally accepted theory accounting for its ultimate evolutionary causes (if indeed these causes exist). Perhaps two of the most popular evolutionary explanations proposed so far are the Mutation Accumulation Theory proposed by Peter Medawar in 1951 and the Antagonistic Pleiotropy Theory, suggested by George C. Williams in 1957. The large amount of data derived from Genome-Wide Association Studies (GWAS) obtained over the last decade allows testing both theories, provided that they can make predictions in terms of the genetic architecture of complex human disease. However, if we want to take advantage from GWAS data, we need to assure that they are sound, replicable and that they contain information that is useful for our purposes. This PhD thesis deals with both goals: we first assess the quality and replicability of information on genome-disease associations and then we use it to explore the Mutation Accumulation and Antagonistic Pleiotropy theories of senescence. Knowledge about the impact of these theories will be important for an increasingly ageing population.

Resumo

A meirande parte das enfermidades xorden a idades avanzadas, e polo tanto, o envellecemento global da poboación motivou o interese nas causas e os mecanismos da senescencia. A senescencia foi de sempre un misterio, sen ningunha explicación para a súa derradeira causa evolutiva, se é que existe. Se cadra, dúas das máis populares explicacións propostas polo de agora sexan a Teoría da Acumulación de Mutacións, suxerida por Peter Medawar no 1951 e a Teoría da Pleiotropía Antagonista, formulada por George C. Williams no 1957. A enorme cantidade de datos derivados dos estudos de asociación de xenoma completo obtidos na última década permítenos testar ámbalas dúas teorías dado que delas despréndense predicións sobre as enfermidades a niveis xenéticos. Inda que, se queremos aproveitarnos destes datos temos que garantir que son sólidos e replicábeis, e que poden achegar verdadeira información de cara aos nosos obxectivos. Esta tese de doutorado comprende ámbalas dúas fins: Primeiro controlamos a calidade e replicabilidade dos datos xenómicos e logo utilizámolos para explorar as dúas citadas teorías da senescencia. O coñecemento achegado polo presente traballo sobre o impacto das dúas teorías serán de grande utilidade para unha poboación en continuo envellecemento.

Prologue

"You there, Ephialtes...
I hope you may live forever."

Leonidas (Spartan king, b. 540 a. C, attributed)

Spartans believed that those dying in battle deserved higher honors than those who die (from age-related conditions) after a long life. That is why the Spartan king Leonidas cursed Ephialtes to live a long life with the burden of his treason after discovering his betrayal, because life is also about suffering.

How have we shifted from an often fatalistic vision of long living in ancient Greece era to the current knowledge about ageing and our attempts to achieve a healthier and longer lifespan? In the present PhD thesis we will move across the dynamic vision of longevity through time, from Sumerian legends (2,000 BC) to the cutting-edge technologies that currently let us read genomes and find the genetic keys to a better and healthier ageing. Leveraging on the recent bloom of public tools and genomic data, which allow researchers to perform new tasks in an unprecedented way, I aim to help unraveling some of the explanations for human ageing and why does it happen the way it does.

To my knowledge, this is the first time that an exhaustive and applied analysis on ageing theories from a genome-wide perspective is performed, with mostly satisfactory results.

Index

1. INTRODUCTION

"*Do not try to live forever.*
You will never succeed."

Bernard Shaw (Irish writer, b. 1856);
The Doctor's Dilemma (1946)

1.1 HISTORICAL FACTS ON LONGEVITY AND DISEASE: FIRST THEORIES ABOUT AGEING

1.1.1 Some legends and myths on ageing

Since the dawn of civilizations we have been craving for eternal living. The oldest reference to the (by now) impossible enterprise of attaining endless youth can be found in the **Sumerian legend of Gilgamesh**(George and Mosley 2010), dating from 2100 BC. The legend narrates how the Sumerian king Gilgamesh struggled to find a formula to reverse ageing. At the end, he realized that such an effort made no sense and that death was unavoidable so he gave up on his quest. This pro-longevity dream has been recurrent across civilizations in time. However, it has not been free of criticism. A debate between opposed visions of longevity was already enunciated in an ancient Egyptian papyrus, the papyrus Smith (Feldman and Goodrich 1999), known to contain a recipe for a rejuvenation elixir. However, it also contains a statement raising awareness of the limited possibility of reaching more than 60 years, a deed

reserved only to those blessed by divinities. Actually, the average life expectancy in ancient Egypt was about 35 years, and one-third of the population died during childhood, so reaching more than 60 years was not very common (Filer 1995). As I will discuss later, living longer than the menopause age was not common until very recently, therefore age-related diseases after the sixth decade were not usually observed. This is an essential concept to bear in mind in order to understand the postulates of the evolutionary theories of ageing.

Humans have been interested in the possibility of extending their lifespan since very ancient times; however, it was not until Classical Greece that the **first theories of disease and ageing** started to develop. **Ancient Greek mythology** eventually represented the derived problems of living forever through mythological stories, such as the myth of the Trojan prince Tithonus. Eos, the goddess of dawn, fell in love with Tithonus, who was a mortal. In her attempt for maintaining Tithonus as young as he was, she asked Zeus to confer him immortality by feeding him with nectar, the divine drink of the Olympian gods, but she forgot to ask for eternal youth. Eos and Tithonus lived together for years, but after several centuries he grew so old that he could not move and withered away so much that had to be locked in a room. Then, Eos finished his pain by transforming Tithonus into a grasshopper, and sent him back to Earth (Finch 2010). The distinction between living forever and being eternally young and vigorous illustrates the difference between healthy *vs.* unhealthy ageing. Indeed, we not only desire to attain a longer lifespan but a longer and healthy one, that is, to avoid the age-related diseases and frailty usually linked to senescence. Other stories like the Endymion myth raised similar concerns about attaining endless youth (Jeune 2002).

According to the Greek historian Herodotus, the legendary Aethiopian tribe termed Macrobians ("big life", in Greek) (Finch 2010) were supposed to attain their large lifespan of 120 years by bathing in a supposed "fountain of eternal youth". Everyone who bathed there came out fresh, glossy and younger. This legend may have given birth to the popular tale that a fountain of eternal youth exists somewhere. According to the legend, during the Spanish colonial times in Central America (16th century) Ponce de León, the Governor of Puerto Rico, was told by Native Americans that such a fountain was located in the Bimini Island (Bahamas) and became obsessed with finding it to reacquire his youthfulness. Obviously, he did not succeed.

Already in the 13th century, the Franciscan friar Roger Bacon, following the Biblical example of Methuselah, who lived until 969 years, hypothesized that humans were programmed to live up to ~1,000 years. He argued that the constant decline in longevity and increasing frailty were due to the corruption and ignorance transmitted from one generation to the next, causing thus a deterioration in lifespan until reaching a maximum of 80 years, an obvious observation linking again senescence and pathology, despite the wrong interpretation.

1.1.2 A more sensible view of longevity

Even if pro-longevitists dominated the Enlightenment period, it has not been always like this. A more prudent opposed view, even pessimistic sometimes, existed already since the Greco-Roman era.

This more realistic vision of an ephemeral life, supported by **Aristotle, Galen, Empedocles** and **Hippocrates** (Table 1), fostered the advancement of science and modern medicine leading them to envisage

the **first theories of ageing** and disease (Jeune 2002). Nonetheless, their scientific ideas about medicine and ageing contrasted with those inherited from previous centuries (from the 8th to 5th century BC). In ancient Greece, the belief that lifespan was away from the control of mortals and determined by gods was widespread. **Homer** left texts exemplifying this vision, in which fate curtailed lifespan (Finch 2010). Homer thought that the immortality of gods was achieved through diet. Their special nourishment on nectar and ambrosia was the key to their eternal life, in contrast with the bread and wine reserved for humans that caused them to senesce. Actually, we currently know that the only intervention that can delay the onset of the symptoms of ageing and extend lifespan is caloric restriction (Heilbronn and Ravussin 2003). It is interesting to realize that ancient Greeks were already linking diet and lifespan.

We may ask ourselves now whether this belief of mortality defined by diet enlightened **Empedocles** to cite, maybe for the first time in history, the use of drugs as a treatment for ageing-related conditions in one of his poems (Empedocles and Leonard 1908). Still, Empedocles believed that supernatural forces controlled other aspects of life such as the weather.

A further evolved and opposed vision to Empedocles was the naturalistic vision of **Hippocrates**, considered the father of modern medicine, who stated that all diseases had natural causes. As Hippocrates stated, exercise, drink and food precisely chosen and combined were the key to guarantee optimum health (Nutton 2007). Regarding ageing, the **Hippocratic Corpus** (a collection of medical works) established a link between advanced ages and susceptibility to disease, considering ageing as an important factor influencing the treatment of conditions. The prevalent medical theories in ancient Greece mainly discoursed upon the so-called **antagonistic humors**. According to Hippocrates, these humors formed the base of all living beings. Diseases and weaknesses resulted from an

imbalance between these four fluids, namely: blood, black and yellow bile and phlegm. Even Aristotle wrote some of his essays following this humor conception. In the same line, Alcmaeon (Longrigg 1993) is thought to have enunciated that disability and health depend on a disequilibrium of heat and moisture, which at the same time goes fading with age, just like an oil lamp. Each creature carries a specific level of these conditions, which vanishes with age (Jeune 2002; Gilleard 2015). As I will show, the idea of disease resulting from an imbalance was rescued ~2,000 years later when George C. Williams suggested his Antagonistic Pleiotropy Theory of Senescence, which postulates that the key of senescence is to be found on mutations that have two opposite effects on the organism: beneficial when young and harmful when old. That is, there is a tradeoff between the effects of mutations during different periods of life, but more about this theory later on. In more quantitative terms, and as the ancient Greece period was coming to an end, consciousness that lifespan had a set limit was raising. In words of the musician and poet Mimnermus:

"Would that my fated death might come at sixty,
unattended by sickness and grievous cares."

The idea suggests disability and disease from the seventh decade on. However, we know that the onset for diseases curtailing lifespan always came much earlier than the seventh decade, ages that were not easy to reach at the time. Details about the onset of senescence and age-related diseases will be discussed in sections 1.3 and 4.

1.1.3 Onset of Galenic thinking

Galen (1st century CE) (Table 1), the Roman physician, surgeon and philosopher is considered the most accomplished medical researcher of antiquity. Inspired by Aristotle, Plato and Hippocrates, he inherited the theory of humors and further developed it to formulate his **Hippocratic doctrine** (Gilleard 2015; Nutton 2005). This combined the four humors, the four elements (air, earth, fire, water) and the four qualities (heat, coldness, moisture and dryness) deriving in, probably, the first theory of ageing ever. In his theory, ageing was correlated with progressive coldness and dryness because old humans run out of humors, and as a result decline in organ function, vitality and muscle strength. Galen did not consider ageing as a disease, but as a natural process. He smartly noted that coldness not only affected the body, but also the mind, after seeing many people becoming demented at extreme old ages. Contrarily, "humid" and "warm" individuals will have greater chances of living a long life. This theory was further extended by the Islamic world philosophers, being Avicenna (10-11th century CE) the most important in this task (Shoja et al. 2011). In particular, his *Canon of Medicine* was one of the most exhaustive descriptions of Galen's natural philosophy. This **Galenic model of ageing** endured for centuries. It was the ground for several essays written during the Renaissance, and kept influencing ideas on age and the healthy ageing process well into the 18th century, and possibly beyond.

1.1.4 Challenging Galenism

Galen, by combining Aristotle's philosophical work with the knowledge he gained with dissections, created a system that explained the structure and function of the human body. Galen's treatises were the basis of medical education in Europe well into the 11th century. Despite an overall accuracy, **certain scholars criticized Galen's works**. Andreas Vesalius was initially educated as a Galenist, but after moving from Paris to Padua, he started dissecting human corpses by himself and comparing his observations to Galen's affirmations. Soon, Vesalius realized that Galen had committed some errors of anatomical nature. Of relevance, Galen mentions that the sternum was composed out of seven smaller bones, when it is formed out of just three segments. Also, he stated that the humerus was the longest bone of the body. At the light of this, it was obvious that Galen had never properly compared a femur with a humerus (Vesalius 1543). Later, in 1628, William Harvey's works on blood circulation were published, which also discredited Galen when affirming that blood was generated in the heart and the liver (Harvey 1628). There were certainly a lot of misunderstandings in Galen's work. Indeed, Galen never opened a human corpse to explore. This was prohibited by Roman laws, and Galen could only use monkeys and cows to perform dissections. However, it was neither Vesalius nor Harvey's intention to discredit Galen. Certainly, the Galenic model of healthcare and Galenic "hygiene" were both the foundations for one of the first English monographs on geriatrics, written by John Floyer of Lichfield (Floyer 1724).

During the 18-19th centuries medical thinking evolved, mainly in the way a patient was seen: from being considered a human, to a group of organs, to an accumulation of cells. This was called the transition from the "sick-

man" to "diseased tissue" (Jewson 2009). Unfortunately, this paradigm shift did not come accompanied by an interest in re-formulating or enunciating new ageing theories. This was only possible after the arrival of the **Cell Theory** in 1838, contributed by Matthias Schleiden and Theodor Schwann, which conferred the fundaments for the correct understanding of human body and disease.

Already into late-modern times, it is widely acknowledged that the foundations for modern geriatrics were developed after the works of Marjory Warren (1897–1960), the mother of British geriatric medicine (Denham 2011). Beyond writing more than 20 articles on geriatric medicine (Matthews 1984), she took care of hundreds of elderly people, who were despised by general physicians for considering them lost causes. But she went further and was able to devise new treatments and strategies to easy their age-related impairments.

Date	Name/Works	Civilization	Event or claims
2650 BCE	Gilgamesh	Sumerian	Struggled to find a cure to ageing, but soon realized he would never find one
1501 BCE	Papyrus Smith	Egyptian	Opposed views of attaining longevity.
--	Tithonus & Endymion (Myths)	Greek	Warnings on eternal youth
4-5th c. BCE	Aristotle, Empedocles & Hippocrates	Greek	A more rational view of longevity
1st c. CE	Galen	Roman	First theories of diseases and ageing
11th c. CE	Avicenna	Persian	Developed and spread Galen's theories
13th c. CE	Roger Bacon	Modern	Humans are initially programmed to live 1,000 years
16th c. CE	Andreas Vesalius & William Harvey	Modern	Improved and corrected Galen's theories
16-17th c. CE	René Descartes, Francis Bacon & others	Modern	Believed that it was possible to be super-centenarians
19th c CE	Matthias Schleiden & Theodor Schwann	Modern	Developed Cellular Theory, setting bases for new ageing and disease theories.
20th c. CE	Marjory Warren	Modern	Basis for modern geriatrics

Table 1: Summary of the most notorious events, works and characters mentioned along section 1.1, displayed in chronological order. Abbreviations: *c.:* century, *BCE*: before current era, *CE*: current era.

1.2 QUANTIFYING AGEING

"A belief in hell and the knowledge that every ambition is doomed to frustration at the hands of a skeleton have never prevented the majority of human beings from behaving as though death were no more than an unfounded rumor."

Aldous L. Huxley (British writer, b. 1894);
Themes and Variations (1950)

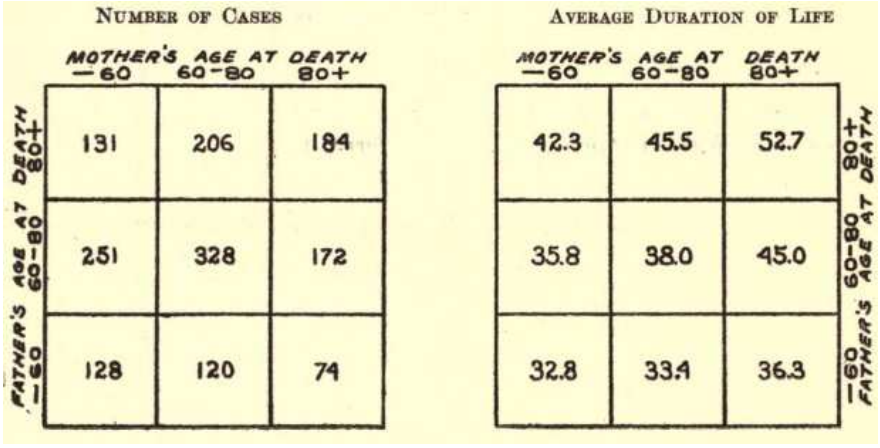
Even before discovering the first genes that influence longevity, or before the first modern evolutionary theories of ageing were formulated, pioneer researchers were already using sophisticated experimental setups to study ageing: From human genealogies to identify a heritable component for longevity, to organismic models (both *in vivo* and *in vitro*) to understand the influence of several parameters in ageing and *how* it progresses in animal cells.

1.2.1 Graham Bell and the inheritance of longevity

The man initially acknowledged to have invented the telephone, **Alexander Graham Bell**, devoted some time, by the end of his life, to study and analyze human genealogies and lifespan. His idea was to explore the inheritance of longevity; that is, whether coming from a long-lived family would increase your odds of dying later. To this aim, he studied the Hyde genealogy, a list of all the ~8,800 descendants of William Hyde, one of the early North American settlers, who died in 1681. This study, published in 1918 as: "The Duration of Life and Conditions Associated with Longevity" (Graham Bell 1918) was a compendium of age records

and analyses of longevity inheritance in the compiled pedigrees. The most striking conclusion from this work was that longevity is inheritable. This is something we know today as a fact, with heritability of longevity estimated to be up to 25% (Skytthe et al. 2003; Hjelmborg et al. 2006), but for the time it may not have looked that trivial. Going into detail, Graham Bell observed that longevity was inherited from both parents; however, a higher heritability was associated with paternal age. In particular, when the father lived to be over 80, 11.3% of the offspring lived to be 80 or older. In contrast, where the mother lived to be >80, only 7.7% of the offspring lived to be >80 (Figure 1). This was confirmed almost a century later, with the observation that paternal age at birth is positively associated with longer telomeres (extremes of the chromosomes; shortening as we age, see details on section 1.4) in offspring and that this effect is cumulative across multiple generations, with subsequently delayed effects of ageing in each generation (Eisenberg, Hayes, and Kuzawa 2012; Njajou et al. 2007). This supports the triple connection between telomere length-aging-lifespan suggesting robust genetic influence, possibly via an imprinting mechanism, on telomere length regulation. In any case, Graham Bell's observations were possibly due to chance, given the little statistical power available to observe them in the dataset he used. However, he went further with his studies on longevity. Motivated by his surprising result that longevity was (at least partially) inheritable, Graham Bell started to compile longevity records from the Washington D.C. area by asking kids in schools how old were their parents and grandparents. He collected these data in what he called "a human stud-book". His goal was that when this project was finished, children from long-lived parents would be able to look up the names, contact and marry those whose ascendants had also a long lifespan to ensure their future descendants would also live longer. Even if this task may sound quite "eugenic", it was never Graham Bell's

intention, not at least in the widely used pejorative sense of the word. Indeed, at the time, mentally handicapped and antisocial people were sterilized by law, something Graham Bell personally found disgusting (Leroi 2005).



NUMBER OF CASES

FATHER'S AGE AT DEATH \ MOTHER'S AGE AT DEATH	−60	60−80	80+
80+	131	206	184
60−80	251	328	172
−60	128	120	74

AVERAGE DURATION OF LIFE

FATHER'S AGE AT DEATH \ MOTHER'S AGE AT DEATH	−60	60−80	80+
80+	42.3	45.5	52.7
60−80	35.8	38.0	45.0
−60	32.8	33.4	36.3

Figure 1: Table from Graham Bell's book (1918). It shows the original numbers and demonstrates how life duration depends on parent's age, being more influenced by the paternal than the maternal part (Left: number of individuals tested; right: average duration of their lives).

1.2.2 First experiments on factors influencing lifespan in *Drosophila*

If one types the word "ageing" in PubMed and goes to the oldest record, you will find John Howard Northrop, an American biochemist, winner of the Nobel Prize for his late works on enzymology. In his early years, Northrop was interested in how several parameters influenced lifespan in *Drosophila sp.* In his work published in 1925, Northrop evaluated the influence of light intensity in the **growth and the lifespan of the fruit fly** (Northrop 1925). His experiments showed that different light intensities

had an effect on duration of the different developmental stages (larval and imago). Anyway, this was not the first experiment on fly lifespan performed by Northrop. In a previous experiment, Northrop and Loeb showed that there is an optimum temperature over which life duration increases (9 ºC), until it creates a plateau (at 19 ºC) (Loeb and Northrop 1916). A year later, in 1917, he showed that total duration of life was correlated with starvation (Northrop 1917).

Northrop was not the only one in realizing that nutrition influenced longevity. As mentioned above, ancient Greeks believed that the immortality of gods was due to the nectar and ambrosia diet they followed. In 1935, Clive Maine McCay and colleagues (McCay, Crowell, and Maynard 1935) were the first to show experimentally that **caloric restriction led to a prolonged lifespan** and delay of age-related diseases, a finding still puzzling today's scientists (Heilbronn and Ravussin 2003). Their fundamental experiments were performed in rats, turning them into the first animal model of ageing, even if we know today that this discovery applies to a wide variety of other organisms, including our own species (Heilbronn and Ravussin 2003). Amongst the first possible interpretations proposed for this prolonged lifespan after caloric restriction was the **free radical hypothesis of ageing**, formulated around 20 years later by Denham Harman (Harman 1956). He postulated that ageing and age-associated decay resulted from the deleterious effect of free radicals (ions or molecules with an unpaired electron in the external layer) generated from cell metabolism. Still, the initial 1956 paper contained only just a suggestion of the free radical hypothesis as a possible mechanism behind senescence. The definitive experimental proof of how free radicals cause ageing did not arrive until 9 years later, in a paper by Harman himself (Harman 1965). In detail, free radicals are generated as a by-product of the

respiratory enzymes involved in the direct processing of molecular oxygen and from the action of catalase on hydrogen peroxide. Among others (Figure 2), these processes yield harmful oxygenated compounds and peroxides formed from water and hydrogen peroxide by removal of one hydrogen atom, commonly known as Reactive Oxygen Species (ROS) (Stein and Weiss 1948). An imbalance in levels of these ions will later impair the correct functioning and replication capabilities of the cell (Finkel and Holbrook 2000). I cannot help recalling, once more, the analogy between this imbalance and the theory of the antagonistic humors devised by Hippocrates and others in ancient Greece times (Figure 2).

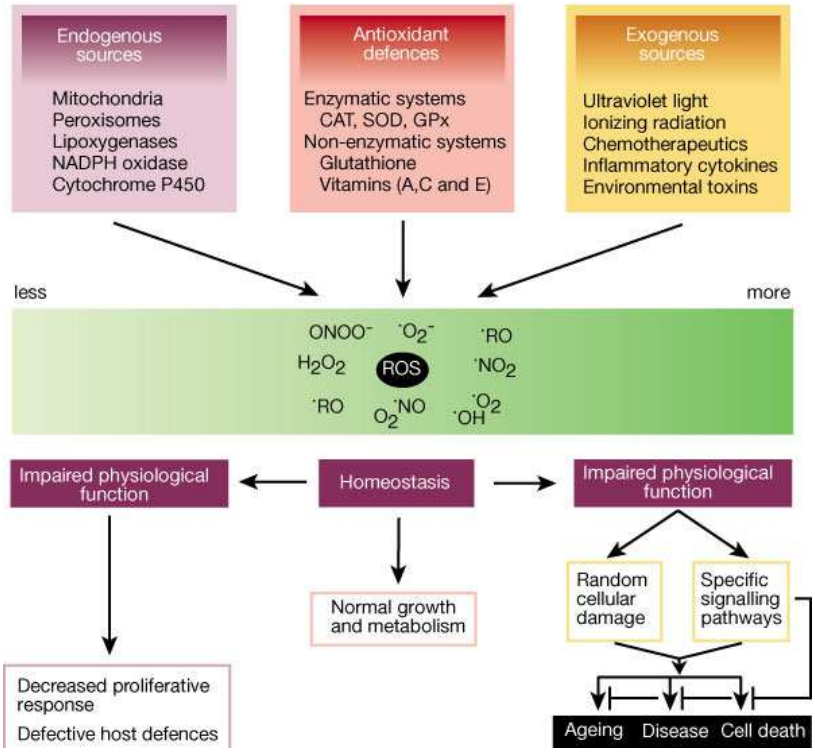


Figure 2: Effect of Reactive Oxygen Species (ROS) in living cells. Optimum levels of ROS are required for homeostasis, but an imbalance will cause impairment of physiological function, leading to ageing and disease when an excess is observed (Figure adapted from Finkel and Holbrook 2000).

As mentioned before, caloric restriction implies a reduction in metabolic rate, which at the same time entails a drop in circulating levels of free radicals. But this is not the only perturbation the body suffers after caloric restriction. Other main physiological effects of food restriction in the organism include improvement of insulin responsiveness, triggering of hormone secretion and modification of the neuroendocrine and sympathetic nervous system function. Caloric restriction also modulates the gene expression profiles in muscle, heart and brain (C. K. Lee, Weindruch, and Prolla 2000; C. K. Lee et al. 2002). It is hypothesized that any (or a combination) of these biological alterations are the triggers of a delay in ageing related symptoms (Heilbronn and Ravussin 2003).

1.2.3 Cells age: How? More hypotheses

By the time the free radical hypothesis of ageing was being developed, another explanation on *how* ageing occurs was parallelly and unintentionally being unveiled. By then, vertebrate cells in culture were believed to be immortal, only dying because of inaccurate culture protocols. This was enunciated by the Nobel laureate Alexis Carrel (Carrel 1921) and this view was completely undisputed by then. This misleading dogma would be challenged by **Leonard Hayflick**, a young North American anatomist, who was recruited in 1958 to run the Wistar Institute's cell-culture laboratory. Hayflick repeatedly failed in growing cell cultures, so he decided to undertake a series of careful experiments to proof Carrel wrong. He devoted more than three years in generating a series of clever experiments clearly showing that his protocols were not erroneous and suggested that **cells under culture could only reach a certain number of divisions**. The observation that some cell division counting machinery was operating behind the ageing process was a

completely new concept (Hayflick 1965). This restricted number of cell divisions in culture is what today we know as the "*Hayflick limit*" coined in 1974 by Sir Macfarlane Burnett, Australian Nobel laureate. This counting mechanism was concluded out of two of Hayflick's findings. One, cultured fetal cells only go through a characteristic number of population doublings, and two; cryogenically preserved cells can recall the number of times they have already divided, once they are unfrozen. Certainly, a still unknown molecular mechanism was responsible for this peculiar clock; however, the golden era of molecular biology had not yet dawned.

Still, the process of DNA replication was already known and it was not a mystery that cells were not able to copy the ends of linear DNA, termed telomeres. This flaw of not being able to copy the full strand at the 3' end was known as the end-replication problem (Watson 1972). Even if telomeres were known since 1938 (Muller 1962; McClintock 1941) their function and structure were unknown. It was not until 1978 that the experiments led by Elizabeth Blackburn and Joseph Gall revealed that the **telomeres** of the protozoan *Tetrahymena termophila* were composed of a series of repeats of the six nucleotides TTGGGG (Blackburn and Gall 1978). This was also confirmed for mammals, but the repeat sequence was slightly different: TTAGGG (Moyzis et al. 1988). This series of tandem repeats were in charge of avoiding the chromosomes to fray, just like the small plastic pieces at the tip of shoelaces do. But since these sequences degrade with each replication, once they are exhausted, cells stop replicating. However, the link between senescence and telomere shortening was not established until relatively few years ago. Once again, Blackburn and colleagues discovered the reason behind why non-senescent organisms did not have issues with telomeric attrition: **telomerase** (Greider and Blackburn 1985), an enzyme that elongates and re-synthesizes telomeres. Active telomerase was found in immortal cell

lines and in a high number of human tumors. However, in humans telomerase is only expressed during the embryonic period to account for the many required cells, and in cells that need a high replacement rate, such as sperm cells or lymphocytes, but not in somatic cells. The expression of transfected telomerase in somatic cells induces telomere length maintenance and this makes them exceed their maximum lifespan by more than five times (Bodnar et al. 1998). There is no doubt today that the senescent process described by Hayflick is a product of telomere shrinkage, but nevertheless it is not its only cause (López-Otín et al. 2013) (Figure 3).

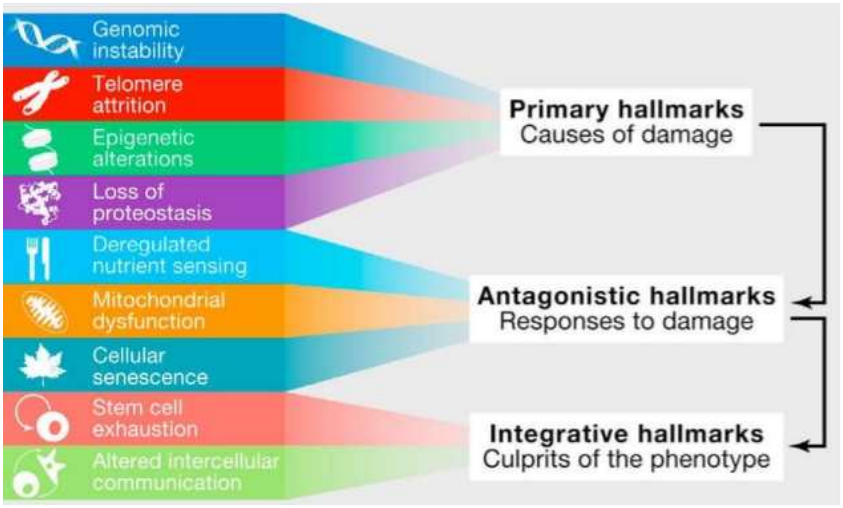


Figure 3: Hallmarks of ageing. Besides genomic instability as a consequence of physical DNA damage and telomere attrition, other causes triggering the senescence process have been proposed. Epigenetic alterations in DNA methylation patterns, post-translational modification of histones, and chromatin remodeling contribute to the process. Loss of quality control in protein folding structures also spurs ageing. Adapted from López-Otín et al. (2013).

Longevity intra- and interspecies is partially genetically controlled, but we do not know exactly what accounts for it and why does it happen. The next section will be devoted more to the "why" do we age than to the

"how", by digging deeper into the formulation and evidence for the ageing theories that are the bases of the present PhD thesis.

1.3 EVOLUTIONARY THEORIES OF AGEING

"*Live fast, die young,*
and have a good-looking corpse."

Willard Motley (African-American writer, b. 1909);
Knock on Any Door (1947)

1.3.1 An unsolved problem of biology

"An unsolved problem of biology" is Sir **Peter Medawar**'s title for a conference he gave in London in 1951 (Medawar 1951). Of course, he was alluding to the ageing or senescence process. After **August Weismann**'s evolutionary theories, Medawar was the next to advance the knowledge on the problem. In brief, Weismann's arguments on ageing mention that the decline in fitness of a population is due to the old individuals, which are harmful for the population, and that thus the healthy and young individuals must displace the old ones (Weismann 1891). Medawar considered that there was some truth on such a reasoning and thus refined it to build his theory of senescence. Medawar started by reckoning that the main causes of diseases at the beginning of the 20th century were infectious diseases such as pneumonia and tuberculosis, which posed a quite common impairment in life duration. Today, we have largely expanded our lifespan, which has almost doubled since 1900 (Roser 2016). As a consequence, the main causes of death nowadays are cancers and cardiovascular diseases. A hundred years ago, it was quite rare to see these diseases prevail in any population, but mainly because these diseases have a typical onset in middle-late ages, which were not easily

reached by the early 20th century (R. Caspari and Lee 2004; Roser 2016). Ancient Greeks and Egyptians were already aware that reaching post-menopausal ages was not common, and that people at such ages were often afflicted with disease. This post-menopausal morbidity is fundamental for the evolutionary ageing theories that will be detailed next.

In Medawar's own words:

"We have already entered a new era in the biological history of the human race."

Maybe this was the view that motivated Medawar to formulate his hypothesis on ageing, relying on the previous works of Ronald A. Fisher on death rates in relation with reproductive values (Fisher 1930). With this, Medawar developed his ageing theory, which would be the **first evolutionary model of ageing in mammals** (Medawar 1946; Medawar 1951). In particular, assuming that genes act at different times in life, then a gene whose action increases survival during reproductive periods will affect fitness more positively than a gene showing effects after reproductive periods. This observation led Medawar to formulate his main claim: that the strength of natural selection declines with age. That is, the action of natural selection will be more effective in improving performance in fertile stages than later in life. As a result, late acting deleterious alleles will accumulate at higher frequencies, a postulate which recalls the name given to the theory, the "**Mutation-Accumulation Theory of Senescence**" (MA from now on) (Rose 1991).

Medawar went further with this idea by suggesting that the scenario could also be slightly different: the very same allele could exert its effects at different periods in life. This implies that a genetic variant increasing

fitness early in life during reproductive stages could be favored despite a negative trade-off late in life, when the organism is no longer able to reproduce. This argument was the base for the development of another ageing theory, named the "**Antagonistic Pleiotropy Theory of Senescence**", (AP from now on) extended, refined and further developed by **George C. Williams**, in 1957 (G. C. Williams 1957) although the full name of the theory was coined by Michael R. Rose in his 1991 book (Rose 1991). These two ageing theories will be central in the course of the present work.

Both theories predict an early period without senescence with mortality independent of age, to decline towards a state in which mortality rates increase, the most important hallmark of senescence (G. C. Williams 1957). However, according to the MA theory, deleterious effects of late acting genes merely accumulate unopposed by selection after the reproductive stages, while according to the AP theory these deleterious late acting genes have been positively selected because of a beneficial effect early in life being maintained in the population as the result of adaptation (Gavrilov and Gavrilova 2002).

1.3.1.1 Pleiotropy

The name of one of the two main theories I will be dealing with along this work implies the term "pleiotropy". **Pleiotropy** is a quite complicated phenomenon with contrasting meanings and interpretations. In general terms, it makes reference to a gene affecting two or more traits. However, multiple variations on the definition of pleiotropy have been proposed including the application of the "pleiotropic" adjective to a molecule, an allele, a genetic marker, a gene or any other entity related to diseases or

complex traits. For this dissertation I will focus on the classical definition of the action of a single genetic variant upon more than one biological process (Plate 1910). In context, the AP theory is called like that because a same mutation affects several traits, particularly exerting protective effects from disease early in life, but acting detrimentally later on.

1.3.2 Conditions and postulates of the AP theory

In his 1957 paper, Williams formulated his theory under the assumption that **senescence**, depicted in the original paper as physiological decay once sexual maturity is reached, is an undesired character in human evolution. It brings disease, decay and in last instance, death. However, it is prevalent to all humans. There should be another force which favors its establishment, and variations in senescence patterns should come as a result of a balance between both forces. Williams proposed that this other force acting in synergy with senescence was natural selection, coming from the action of mutations, or alleles, with differential effects on fitness at different ages. In a few words, genetic variants associated with late onset diseases (at the post-reproductive age) segregate in a population partly because they have a positive effect on fitness while in reproductive periods. I am quoting here an easy example to better understand the application of the theory. In words of Williams (1957):

"*...we might imagine a mutation arising that has a favorable effect on the calcification of bone in the development period, but which expresses itself in a subsequent somatic environment in the calcification of the connective tissue of arteries.*"

For this to apply there should be four conditions or factors that have to be assumed (G. C. Williams 1957):

- The living being should have a soma, key for the reproductive process, but which is not transmitted either sexually or asexually to the offspring. Briefly, somatic and reproductive cells should be different entities.

- Natural selection should be capable of acting on alternative alleles.

- If somatic effects in an individual are seen at different life stages, it necessarily involves pleiotropic genes with contrary effects on fitness at different ages. However, Williams does not mention a single example of such genes, although he was already aware of simple pleiotropic *Drosophila* genes with influences in lifespan (E. Caspari 1952; Gonzales 1921). The big bang on genetics of ageing would be still to come. See section 1.4.
- And finally, the probability of reproduction should drop as we age.

Complementarily, in his seminal paper Williams enumerates a series of nine **testable deductions from the AP theory**, or expectations that will fulfill the theory:

1. *Senescence should be seen where conditions specified in the theory are met, and never where the assumptions are absent.*

The theory asserts that senescence is a derived characteristic of the soma. We should just find it whenever a soma is present. But: How to deal with asexually reproducing organisms? They are all simply copies of the same organism, and even if there is a certain "corporal", deterioration suffered is never alike the senescence suffered by metazoans. Besides sexually reproducing organisms, other species are capable to either show or not

senescence independently of fulfilling the assumptions of the theory (Jones et al. 2014).

2. *Low adult mortality rates should be correlated with low senescence rates and vice versa.*

Mortality rates in a population should be the best proxy for senescence. It has been observed in several species that mortality imposed by environmental conditions (extrinsic senescence) fosters intrinsic senescence (Stearns et al. 2000; Bryant and Reznick 2004; P. D. Williams et al. 2006). As an example, Williams mentions flying birds as species with low intrinsic senescence due to low mortality rates, probably because of the fact that they fly, and can escape predators easier (Healy et al. 2014).

3. *Senescence onsets earlier in organisms that reproduce during the whole life, than in those that start reproducing after a certain age is reached.*

4. *In sexually dimorphic species, senescence onsets faster in the sex with the higher mortality rates and less pronounced increase in fecundity.*

It is widely known that, in general, males present higher rates of senescence.

5. *Senescence comes always as a general deterioration.*

It is not just an issue of a specific organic system. Otherwise, the selection against deterioration in that system would free from senescence the other systems, which is not real.

6. *Post-reproductive period should be absent from the normal life-cycle of any species.*

Studies have shown that only ~50% of the individuals in Late Paleolithic went over 8 years of life, and during the Neolithic only a half reached 26 years (Ascádi and Nemeskéri 1970). Also, less than 1,000 years ago, life expectancy at birth was no higher than 29 years, and Native Americans rarely survived past 50 years, age that is currently reached by ~95% of population from developed countries.

7. *Time of reproductive maturation should mark the onset of senescence.*

8. *A fast development should correlate with a faster senescence*

The sooner an individual reaches maturity, the sooner senescence will start.

9. *An increased longevity should be the result or consequence of reduced fitness in youth* (Figure 4).

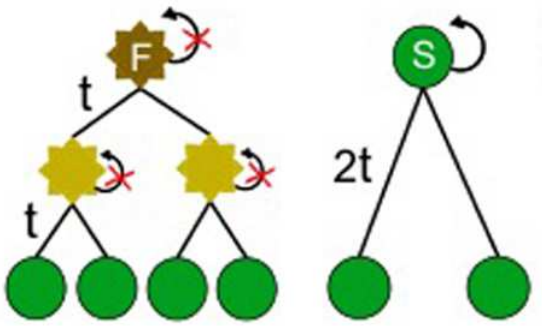


Figure 4: Fitness effects of the phenomenon of antagonistic pleiotropy. Individual *F* carries mutations allowing it to reproduce double than *S* in the same time interval, despite the cost of dying earlier, an inconvenience posed by those same mutations favoring reproduction early in life.

1.3.3 Early tests of the Theory

Even if the theory was well received in the scientific community, it was not until 1980 that the theory was experimentally proved, using *Drosophila* flies (Rose and Charlesworth 1980). **Brian Charlesworth** and Michael R. Rose, two British evolutionary biologists, developed a series of experiments (Rose and Charlesworth 1980) to test the MA and AP theories. They did not find much support for MA theory, but they did it for the AP theory, by comparing egg counts in two populations of *Drosophila*; one reproducing early and the other later. They showed that increases in later reproductive output are associated with reduced early reproductive performance, lower reproductive rate and increased longevity. Also, they observed negative genetic correlations between fecundity and longevity. These observations are both consistent with the AP theory of senescence. But it would not be until fifteen years later that Charlesworth proved that the AP theory is mathematically consistent (Charlesworth 1994). One year after the first experimental proof, the same authors obtained again negative genetic correlations between early fecundity and longevity in *Drosophila* (Rose and Charlesworth 1981), in tune with the AP theory again, although no confidence intervals were provided for these correlations. Several pieces of evidence came later on that were based on different methodologies: between-line correlations in laboratory populations, sib analysis of quantitative genetic variation and chromosomal extraction experiments (Rose 1991). Also, evidences came from other species, such as the turf grass *Poa annua* and the seaweed fly *Coelopa frigida*, for which analogous observations were made (Law 1979; Butlin and Day 1985). Wrapping all this up, it seems that the AP theory measured as negative correlations between early fecundity and longevity, at least in some *Drosophila* populations, does actually apply. In his 1991

book (Rose 1991), Rose highlights two of the clearest demonstrations of the AP theory. The first of them is the *abnormal abdomen* allele in *Drosophila mercatorum*; a mutant that causes a great increase in fecundity, but at the cost of a reduced lifespan (Templeton and Rankin 1978).

The other main group of evidences for the AP theory comes from the nematode *Caenorhabditis elegans*. Experiments performed using this small nematode worm have revealed the **first gene ever isolated directly affecting lifespan**, *age-1* (Friedman and Johnson 1988). Mutant worms developed an increase in lifespan, but also a reduction in early fertility. This first case of genetically-derived longevity boost will be subject to longer discussion in section 1.4. Walker et al. (2000) carried out an elegant experiment in a set of wild type and *age-1* mutants. Initially, the two groups were grown in normal conditions and no evidence for differential trade-offs involving longevity and fertility was seen. On the contrary, if both strains were raised in a harsh environment, wild type worms showed increased fertility at early ages at the cost of shorter lifespan. The conclusion we can take from this last paper is that the AP theory only "works" under natural wild conditions. Indeed, several authors raised this concern, leading to the conclusion that Williams' statements were slightly incomplete (Abrams 1993; P. D. Williams and Day 2003; Reznick et al. 2004).

Despite all these positive proofs, already predicted by Williams, negative correlations between reproduction and longevity were not systematically observed. On his initial paper, Williams argued that senescence started right after reproductive maturation is reached, establishing thus a connection between "faster development and early senescence". Contrarily, Economos and Lints (1986) demonstrated that *Drosophila*

specimens with different growth rates did not present variation in total lifespan. The evolutionary gerontologist Caleb E. Finch also raised in 1990 several concerns about the generality and applications of the AP theory in his book (Finch 1990).

1.3.4 Recent empirical evidence

Since the 2000's we have witnessed a great number of accurate comprehensive experiments proving or refuting instances of the AP theory. According to Williams' second avowal, a high rate of extrinsic mortality implies an earlier onset of senescence and vice versa. Using *Drosophila,* this was demonstrated in one of the best empirical supports for this ageing model (Stearns et al. 2000). However, translation of the findings to natural populations might be misleading (P. D. Williams et al. 2006). Problems such as population density issues might arise, in the sense that a high extrinsic mortality rate reduces the population, and surviving individuals might benefit from more resources. A couple of experiments illustrating the response of wild populations to extrinsic mortality rates were performed in early 2000's. The first experiment (Bryant and Reznick 2004) compared rates of senescence between two natural populations of guppies (*Poecilia reticulata*) living in a water stream where predation was barely present. However, one of the two populations had been previously moved there from a high predation brook. Confirming theoretical predictions, the latter population presented higher rates of senescence as a consequence of adaptation to extrinsic pressure. Contrarily, when these populations were reared in laboratory conditions, the derived population from high predation conditions showed slower senescence rates compared to the other wild population (Reznick et al. 2004).

All this variety of data leads to a fundamental question: how can senescence be measured in a consistent way that can be compared across model organisms? On his 1957 paper, Williams proposed that the best proxy for senescence was mortality rate. Nonetheless, several other indices have been suggested (P. D. Williams et al. 2006). In the previously mentioned paper by Reznick et al. (2004) paper, the authors used the "fast start response" as a proxy for senescence. This is an indicator of the ability of fishes to escape predators, an ability that should decrease with age. When they evaluated this response in both types of fish, the native high predation population showed a steeper deterioration for this measure. Together with the experiment of *C. elegans* set under starvation that I mentioned above (Walker et al. 2000), this remits us to the previous observations that Williams' theory, even if true, cannot explain the enormously varied and complex array of life histories that can be observed in nature.

Further evidence from long-term observational studies of vertebrates comes from Hendry et al. (2004), showing that senescence in a natural red salmon population (*Oncorhynchus nerka*) is predicted by the AP theory. Higher vertebrates were also used as models to test the AP theory. In one study performed in 2006 by Charmantier et al., the authors used life-history data coming from a swan cohort closely monitored during 36 years. They found that increased early reproduction brings faster senescence because of genetic tradeoffs, fulfilling the expectations of the AP theory.

However, predictions of the AP theory in humans have not been widely and convincingly performed. Only certain tests and instances of

antagonistic pleiotropic behaviors were observed in humans, but only concerning particular diseases and genes.

1.3.5 Mutation accumulation and antagonistic pleiotropy in humans

A certain number of human diseases were shown to evolve either under the MA or the AP theories. Here, I will briefly review two of the most well known: Huntington's disease and haemochromatosis.

1.3.5.1 Huntington's disease

Huntington's disease (HD) is a rare autosomal dominant neurodegenerative disease, which generally onsets at post-reproductive stages. Antagonistic pleiotropy has been suggested (Carter and Nguyen 2011; Albin 1993; Sørensen, Fenger, and Olsen 1999) as a mechanism to explain its prevalence and manifestation in late ages (Carter and Nguyen 2011). Increased fitness (as measured by number of siblings) in HD affected individuals was already reported during the sixties and seventies (Shokeir 1975; Albin 1993) even before the causative molecular mechanism of the intergenerationally increasing number of CAG trinucleotide repeats in HD was known (MacDonald et al. 1993). At the same time, HD patients have been shown to present reduced rates of some cancers when compared to controls (Nesse 2001; Sørensen, Fenger, and Olsen 1999). Such reduction has been associated to an increase in *p53* tumor suppressor activity (Eskenazi, Wilson-Rich, and Starks 2007), which induces a higher apoptosis rate. While this higher apoptosis rate may be the cause for the neurodegeneration episodes in HD, it has been

suggested that it may also be protective for progression of malignant cancer cells. Thus, a yet unknown mechanism increasing fertility and a higher cell apoptosis rate in HD patients represent two potential cases of antagonistically pleiotropic behavior in carriers of this condition. However, an explanation for the increasing fitness benefit has not been fully provided for this trait, and MA could provide a more suitable and conservative explanation for it (Charlesworth 1994).

1.3.5.2 Haemochromatosis

Hereditary haemochromatosis (HH) is an autosomal recessive disorder, which increases the amount of blood iron with age (Adams and Barton 2007). It is more frequent in populations of European ancestry (~10%) (Distante et al. 2004), where 90% of affected patients present the C282Y missense mutation (rs1800568) in the *HFE* gene (Weinberg 2008). Affected individuals benefit from the associated enhanced dietary iron absorption during their early reproductive ages. This additional iron load results in a more efficient immune system, particularly against *Salmonella typhi* and *Mycobacterium tuberculosis*. Notably, the diseases caused by these pathogens (typhoid fever and tuberculosis, respectively), although prevalent throughout human evolution, expanded with the crowded conditions of large cities (Weinberg 2008). The deleterious effects of the C282Y mutation appear after the fertile period, when increased iron accumulation can represent up to a 33% morbidity increase for C282Y homozygotes (Whitlock et al. 2006). In particular, conditions like certain infections, neoplasias, cardiomyopathy and other neurodegenerative, endocrine and orthopedic diseases have been suggested to be fostered and worsened by raised iron loads (Weinberg 2007).

In spite of such more or less clear examples of human diseases fulfilling the expectations of the MA or AP theories, **evaluation of the AP theory using human demographic data are quite scarce** and barely shed some light on whether the theory applies to us as a species, at least in terms of particular Williams' postulates. Initially matching theoretical expectations, one study (Westendorp and Kirkwood 1998) found out that individuals live longer, especially women, at the cost of reproductive success. This was seen as the fulfillment of one of the predictions of the AP theory, but several methodological flaws were identified in that study (Gavrilov and Gavrilova 1999; Gavrilov and Gavrilova 2002) and, additionally, the authors failed to reconcile their observations with previous work returning opposite results and in fact did not even cite it (Le Bourg et al. 1993). All these contradictory results call for further, deeper and more accurate research to definitely verify George C. Williams' predictions.

1.3.6 A third main theory of ageing

The senior author of the aforementioned polemical paper, **Thomas L. Kirkwood**, a South African evolutionary biologist, proposed a third main theory to explain ageing: The Disposable Soma Theory. **The Disposable Soma Theory** (DS from now on) was formulated in 1977 (Kirkwood 1977) and was later developed by himself and Holliday (Kirkwood and Holliday 1979). It frames senescence as a resource allocation problem by suggesting that organisms must choose between dedicating energy to favor reproduction, thus accelerating senescence, or investing it in soma preservation and growth with the trade-off of an increased lifespan and delayed senescence. Depending on the outcome of both processes certain

options will produce higher and more optimal fitness than others. Kirkwood recognizes that DS it is just a special instance of the AP theory (Kirkwood and Holliday 1979):

"The disposable soma theory is, in a sense, a special case of Williams' (1957) pleiotropic gene hypothesis, the gene in question controlling the switch to reduced accuracy in somatic cells. The good effect of the gene is the reduced investment of resources in the soma, while the bad effect is ultimate somatic disintegration or ageing"

Williams (1957) stated that no organism could manifest senescence unless his soma was separated from the germ line. Given the nature of the DS theory, it is important to make the separation here between the somatic and reproductive part of an organism. The germline must be kept at a level that preserves life across the generations, whereas the soma just acts to preserve the actual generation. In brief, soma is just a larger container for reproductive material.

Particularly, *Hydra sp.* (Cnidaria) are **non-senescent organisms** if maintained under benign conditions and asexual reproduction (Jones et al. 2014; Schaible et al. 2015). A peculiarity of the *Hydra sp.* is that its germline is also its soma, so both are effectively one, making Williams right once more, according to his first postulate. Schaible et al. (2015); Schaible, Sussman, and Kramer (2014) suggested that a certain instance of DS might still be applied to *Hydra.* In the case of the species *H. magnipapillata* the natural continuous cell turnover replaces its cells completely after 3-4 weeks, resulting into a very effective way of preventing the accumulation of damage. At the same time this happens, resources are also invested in renewing reproductive cells, thus preventing senescence (Schaible et al. 2015). In short, hydras have no soma to be disposable (Kirkwood 2001). However, a clearer example may apply to a certain *Hydra* species, *H. oligactis*, who enters a fast-senescence period right

after sexual reproduction is triggered by detrimental environmental conditions, dying within ~150 days. This last species has to deal with a true resource allocation problem, following the DS theory, by stopping self-repairing and then investing all resources in sexual reproduction (Yoshida et al. 2006; Brien 1953). But the non-senescent *Hydra sp.* is more the exception than the standard rule to ageing patterns in the wild.

It is crucial to realize that all three theories explained so far (MA, AP and DS) can act at the same time and are not mutually exclusive, as recognized by many authors (e.g. Gavrilov and Gavrilova 2002).

1.3.7 Is senescence is a universal phenomenon?

According to the British evolutionary biologist **William D. Hamilton**, senescence is a demonstrated universal phenomenon (Hamilton 1966). In this paper Hamilton supports the vision that "***senescence is an inevitable outcome of evolution***", with increasing mortality and fertility loss once reproductive maturity is reached. That statement was and is harshly criticized, and fostered a large debate (Vaupel et al. 1998; Vaupel et al. 2004; Rose et al. 2007), in which I will dig into later. In his key paper on senescence Hamilton attempts to explain the aging process by assessing how the time of action of a gene induces different effects on the fitness of the organism. He shows that natural selection can shape survival and fertility differently, forcing them to evolve slightly independently. To formulate and prove his theory, Hamilton relied upon the works of Ronald A. Fisher (Fisher 1930), centered on age-specific fertility and survival probabilities and analyzed the relationship between fertility and

survival rate with age. Later works of John B. S. Haldane (Haldane 1941), Medawar (Medawar 1951; Medawar 1946), and G. C. Williams (G. C. Williams 1957) used this conception, but none of them formally demonstrated Fisher's predictions. Hamilton was the first to mathematically prove this right.

Criticism of the "Hamiltonian" vision of senescence was started by two seminal papers back in 1992. Curtsinger et al. (1992) and Carey et al. (1992) demonstrated that *Drosophila* and medflies (*Ceratitis capitata*) do not show the expected patterns of senescence. Both studies found that there is no specific increase in mortality after the attainment of sexual maturity, contrasting with the life-span paradigm conceptualized by Hamilton. A further exhaustive literature exploration of death rates in six animal species by Vaupel et al. (1998) showed that mortality rates, namely senescence, do not universally increase after sexual maturity. This generated a sub-debate about the extant several types of senescence: **Positive senescence**, as depicted by Hamilton, **negligible senescence**; already defined by Finch (1990), and **negative senescence**. This latter term was lastly introduced by Vaupel, Baudisch and colleagues in three seminal works (Vaupel et al. 2004; Baudisch 2005; Baudisch 2008) and stands for death rates falling with age, instead of increasing. In a theoretical framework, Vaupel et al. (2004) showed that it was conceptually possible to develop negative senescence for certain species under certain conditions, proving Hamilton's predictions were not always correct. Further theoretical demonstrations by Baudisch (2005) proved that the MA theory may have been relatively unimportant during human evolution, casting more doubts on the Hamiltonian vision of senescence.

But supporters for the Hamiltonian vision of senescence countered this vision. Finch (2009) noted that one cannot not claim that centenarians do not show negligible senescence, and Rose (Rose et al. 2007) demonstrated that even if the models by Vaupel and Baudisch are interesting for life-history models, they emphasize the magnitude of their findings at the light of clonally reproducing organisms for which the Hamiltonian models cannot apply. According to Rose (Rose et al. 2007), Hamilton's equations had been misinterpreted. The explanation was latent in Hamilton's original paper, but formally developed later (Mueller and Rose 1996; Rauser, Mueller, and Rose 2006).

To reconcile both visions of senescence, and to better understand the evolution of ageing in our own and other species, published in 2014 a comprehensive study of **ageing patterns across the tree of life** (Jones et al. 2014). This study revealed the need of a broader understanding of ageing in the wild and showed that classical ageing theories might not suffice to explain the enormous amount of variation of life-history trajectories analyzed. By comparing mortality and reproduction on 46 species of animals and plants they revealed the extant high diversity of senescence patterns in living organisms, ranging from those that show "classic" senescence, to those that show negative senescence like the desert tortoise (*Gopherus agassizii*) or plants like the white mangrove (*Avicernia marina*); and those that ignore senescence, such as the aforementioned *Hydra sp.* (Figure 5).

These mechanisms oriented to revert, to accelerate or to ignore ageing must be, at least partially, genetically determined. The following section will look in more close detail at the genetics of ageing and at the inner mysteries of a long-standing and fascinating conundrum.

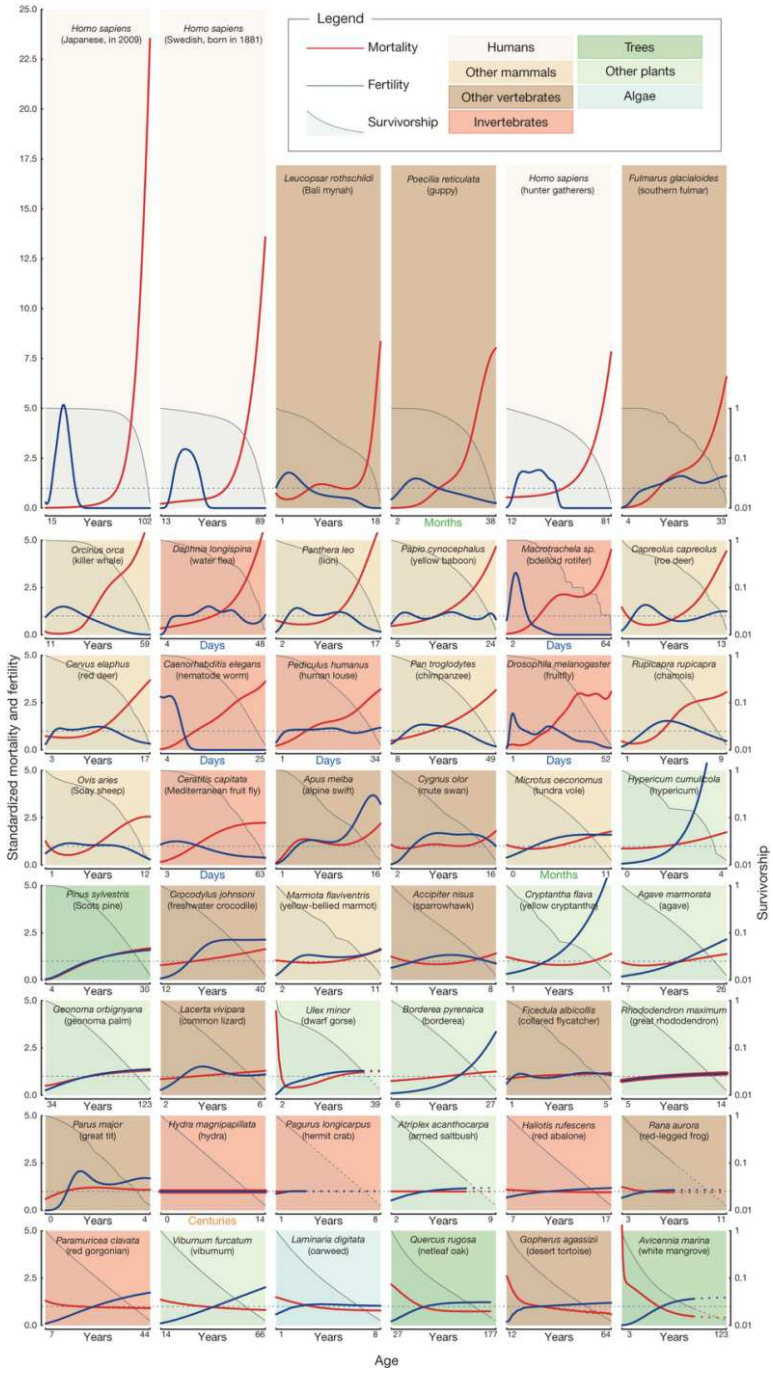


Figure 5: Patterns of senescence in the wild for 46 species across the tree of life. Moving down the graph, the rate of senescence is decreasing, until the last row where we see instances of organisms with negative senescence. Adapted from Jones et al., (2014).

1.4 THE GENETICS OF AGEING

"Some things are more precious
because they don't last long."

Oscar Wilde (Irish writer, b. 1854);
The Picture of Dorian Gray (1890)

The genetic architecture of a trait is used to define the underlying genetic basis of a phenotypic trait, composed by the distribution of causal variants, allele frequencies, effect sizes and the patterns of pleiotropy, dominance and epistasis they maintain (Visscher, Mcevoy, and Yang 2010).

In contrast with Graham Bell's times, the knowledge on genetics accrued during the last five decades has allowed researchers to study diseases and ageing from an internal perspective. Several genotype-phenotype association strategies are now available: from classical linkage mapping methods to high-throughput genome-wide association studies, that allow finding genetic associations. The initial studies on the genetics of ageing already pinpointed genes that control lifespan, but these single genes cannot fully explain the complex nature of the processes of ageing, longevity and their related diseases.

Currently, genome-wide association studies (GWAS) are the choice method to find the polygenic bases not only of common diseases, but also of such complex traits. Moreover, public access to the new wealth of genome-wide disease data places us in an optimal position to test

hypotheses such as the two evolutionary theories I am presenting in this thesis.

1.4.1 Longevity is (also) in your genes

On present times no one challenges the view that longevity has a relatively high inheritable component. These observations were already made, even if not with ideal precision, by Graham Bell a hundred years ago. As we have seen, evolutionary theories consider senescence as physiological decay, and undoubtedly it comes with increased risk of certain age-related diseases. Senescence could be conceptualized like any another complex trait, such as height, body mass or cancer. Apart from scarce large effect variants, it is shaped by lots of genetic variants with medium to small effects contributing to the phenotype in several ways. This is the kind of variants that GWAS are empowered to detect. A phenotype varying between individuals in a population can be affected by environmental conditions and genetics. The variation in a trait attributed to genetics is termed **heritability** (H^2). It is formally defined as the proportion of phenotypic variation (Vp) that is explained by variation in genetic values (Vg), or $H^2 = Vg/Vp$ (Wray and Visscher 2008).

Estimated from studies in twins, heritability of survival over eighty years old is estimated to be over 25% (Skytthe et al. 2003; Hjelmborg et al. 2006), with genetics playing a larger role for people living after the top 5% of maximum longevity (Tan et al. 2008; Sebastiani, Solovieff, et al. 2012; Schoenmaker et al. 2006). **Heritability of longevity** differs between sexes, estimated to be 0.23 for males and 0.20 for females (Herskind et al. 1996), although some studies point up to 0.50 (Yashin et al. 1999). From

this, we can easily deduce that senescence or longevity also have their own characteristic genetic architectures.

1.4.2 Unraveling the genetic architecture of complex traits: The Common Disease – Common Variant hypothesis.

Disease phenotypes have classically been the main focus of studies investigating the genetic architecture of complex traits. For instance, the first experimental studies in the early 50s and 60s found strong associations between common variations in *ABO* and *HLA* genes and several types of cancer and autoimmune diseases (Bodmer and Bonilla 2008). Interestingly, these alleles present large effect sizes along with large allele frequencies. For instance, 8% of individuals of European ancestry carry the HLA-B27 allele associated to ankylosing spondylitis with huge original effect sizes reported (Schlosstein et al. 1973). In the 80s, the study of complex disease genetics with positional cloning methods ramped up thanks to the availability of genomic maps through restriction fragment length polymorphism (Botstein et al. 1980). However, the most successful findings were spotted as false positives by later studies (Botstein and Risch 2003) and the main successes were shown to be restricted to **familial forms of disease** (e.g. *BRCA1* and risk for young-onset breast cancer). The failure of **linkage methods** when applied to complex phenotypes prompted a fierce debate on the methodology necessary to unravel disease causal variants. In the late 90s, several landmark studies explored the methodological challenge of unraveling disease-associated variants of low risk (Lander 1996; Risch and Merikangas 1996; Chakravarti 1999; Reich and Lander 2001). For instance, Risch and Merikangas compared linkage to association methods and showed the latter would be more empowered

to find low-risk alleles in the absence of allelic heterogeneity. The authors also acknowledged that strict significance thresholds would be necessary to avoid an inflation of false positives (since large numbers of polymorphisms across the whole genome ought to be tested).

A variety of papers also explored the issue of disease architecture from the perspective of disease evolution and from the use of simulations (Pritchard 2001; Reich and Lander 2001; Rosenberg et al. 2010). These studies explored wide ranges of the effects of risk variants on the fitness of carriers (*i.e.* the effect a risk variant will have upon the probabilities of survival and reproduction of an individual) to answer questions on the frequency and number of alleles expected in loci playing a major role in the architecture of complex traits. These studies showed that a regime of weak purifying selection would prevent the fixation of disease variants, leaving them at intermediate frequencies. Given its correlation with heterozygosity (Visscher et al. 2012) these alleles would contribute most of the genetic variance of disease. In other words, simulation studies suggested that a relatively restricted set of common variants (frequencies >0.05) could account for a large proportion of the heritability. All these ideas helped in the establishment of the **Common Disease - Common Variant (CD/CV)** hypothesis. This paradigm states that common variants in susceptibility genes account for most of the genetic risk of complex disease and other polygenic traits, including ageing. The slightly deleterious effects of these variants in the fitness of affected individuals explain their relatively high allele frequencies. Their frequency would counterweight the low relative risks of these variants, thus explaining the large prevalence of complex disease in modern populations. Along with strong criticisms from some authors (Weiss and Terwilliger 2000) the CD/CV hypothesis was established as the paradigm in human genetics

and paved the way for the wave of linkage disequilibrium-based association studies.

1.4.3 Hunting the first ageing genes

Senescence as a complex and common trait can be framed within the CD/CV hypothesis, in the sense that many intermediate frequency variants contribute to it. However, the first steps in the identification of ageing-related genes were made by isolating some large effect genes.
The first keystone on the puzzle was set in 1983. Klass, Nguyen, and Dechavigny (1983) used populations of the nematode *C. elegans* to identify mutations in a gene named ***age-1***, which controlled lifespan, but authors thought that this extension was due to caloric restriction, the only procedure confirmed to boost lifespan by then (Masoro 1985). Some years later (Friedman and Johnson 1988), *age-1* would be truly confirmed as a gene directly controlling mean and maximum lifespan. The recessive mutant allele *age-1*(hx546) in *C. elegans* posed an increase of 40-65% in mean lifespan and 60-110% increase in maximum lifespan. This lifespan lengthening does not come alone and implies pleiotropic effects by reducing hermaphrodite fertility, fitting Williams' predictions (1957), although this may be a cause of slower ageing, rather than its consequence. This gene codes for a downstream component of the PI3-kinase/PDK/Akt pathway, which possibly affects spermatogenesis or other metabolic processes resulting in decreased fertility.

The second clear-cut example of a gene discovered to modulate ageing is the *daf-2* gene, also from *C. elegans* (Kenyon et al. 1993). Curiously, it also belongs to the PI3-kinase/PDK/Akt pathway. When any of the three

studied mutations within this gene was knocked out, nematode's maximum lifespan extended more than double. However, the activity of a second downstream gene, *daf-16* is also needed. The authors have hypothesized that blocking the signaling on these genes will produce an analogous effect to that produced when caloric restriction is applied (Kenyon et al. 1993). Caloric restriction and related metabolic pathways seem to be a major determinant of longevity in most of the species, including humans, nematodes and flies (Heilbronn and Ravussin 2003).

Drosophila, another "gold standard" laboratory animal, also seems to manifest extended lifespan when manipulating some of these metabolic related genes (Helfand and Rogina 2003; Tatar, Bartke, and Antebi 2003). The highly similar response observed after modifying metabolic genes in worms, mice and flies demonstrates the high evolutionary conservation of pathways affecting ageing across species.

1.4.4 Linkage disequilibrium mapping and the candidate gene era

At the turn of the new millennium and parallel to the ongoing Human Genome Project, linkage disequilibrium-based association mapping was established as the choice method to unravel the genetic bases of complex traits. **Linkage disequilibrium** (LD) (Figure 6) is the nonrandom segregation of two or more alleles at different loci during meiosis (Hill 1974; Lewontin and Kojima 1960).

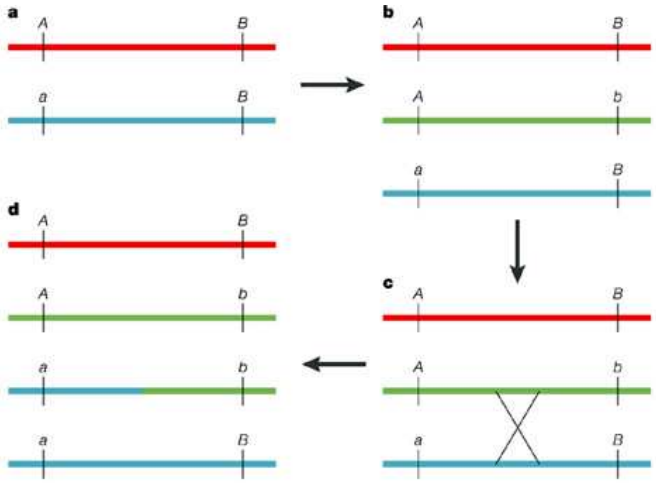


Figure 6: Linkage disequilibrium representation. |**a**| Polymorphic locus with alleles *A* and *a*. |**b**| A mutation occurs changing allele *B* to *b*. This happens on a single chromosome with either allele *A* or *a* at the first locus (*A* in this case). So, only three out of the four possible haplotypes will be observed in the population. The *b* allele will always be found with the *A* allele at the adjacent locus on the same chromosome. |**c**| The association between alleles at the two loci will gradually be disrupted by recombination. |**d**| This will create the fourth possible haplotype and an eventual decline in LD among the markers in the population as the recombinant chromosome (*a*, *b*) increases in frequency. Adapted from Ardlie, Kruglyak, and Seielstad (2002).

These loci are said to be in linkage disequilibrium when the frequencies of gametes with a pair of alleles *x* and *y* is different from the product of their corresponding allele frequencies (Slatkin 2008). The **LD-based association mapping** method aims at establishing statistically supported associations between genetic markers and the phenotype of interest. In contrast to classical genetic linkage studies (Bodmer and Bonilla 2008), association mapping ascertains the transmission of the phenotype with particular alleles instead of with the entire loci. In the context of disease studies, genetic associations arise when specific alleles are more frequent in cases (individuals with a certain phenotype) than in controls (without the phenotype) (Ziegler, Konig, and Pahlke 2010; Hästbacka et al. 1994). Associations between alleles and phenotypes arise either when the tested marker is causal (direct association), or in the more likely scenario that the

variant under study presents strong LD with the marker that in turn causes disease (indirect association).

In consequence, candidate association studies present two requirements regarding genetic variation: First, prior biological knowledge is necessary to select loci that "make biological sense" (Thomas 2004) or contain clues of participation in the etiology of the phenotype of interest. Second, dense spacing of markers is necessary to cover extensively the ascertained candidate loci. Two broad classes of genetic variations can be distinguished in our genomes according to their size: Single nucleotide polymorphisms (SNPs) and structural variation. SNPs are just point mutations of a single base. SNPs are the marker of choice for association studies due to their abundance in our genome (>62 million, according to dbSNP), but an understanding of the strength of LD in human genomes is necessary for a proper design of population-based association studies. A very initial study based on simulations estimated that LD would not extend, on average, beyond 3 kilobases (Kruglyak 1999). Subsequent studies based on actual data established the existence of discrete haplotype blocks that extend for several tens of kilobases before being punctuated by recombination hotspots (Daly et al. 2001). The possibility of characterizing haplotype blocks by the genotyping of a low number of SNPs (named **tagSNPs**) that may be shared across populations (Gabriel et al. 2002) prompted the establishment of the HapMap project in 2002. The widespread availability of SNP markers expedited the publication of candidate gene association studies using case-control designs (Hirschhorn et al. 2002; Hirschhorn and Altshuler 2002). In this design, frequencies of variants at the ascertained SNPs are compared in populations of cases and controls. Their penetrance and relative risks can be approximated in

association studies by calculating the strength of association using, for instance **Odds Ratios** (OR) (Figure 7) in the case of binary traits.

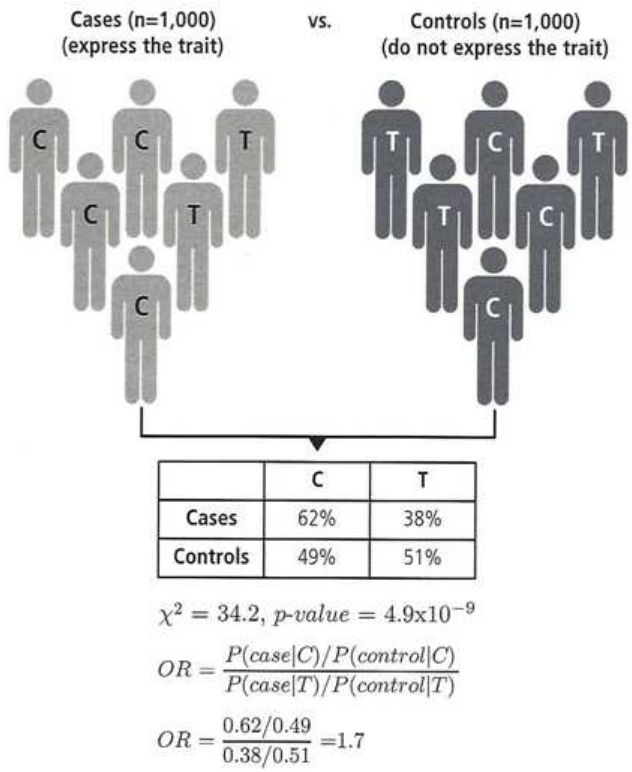


Figure 7: Case-control association study design. Note how ORs are calculated.

Conditions that are better approximated as quantitative traits (*e.g.* height or blood pressure) can be assessed through linear regression methods. The identification and replication of several associations through candidate gene approaches, such as *CTLA4* and type 1 diabetes or *NOS2* and Crohn's disease, created a huge wave of hype in the community.

1.4.5 Candidate genes for ageing

More in line with the present work, ageing or senescence can also be a phenotype of interest to tackle by association mapping. After all, "senescent people" are a group of individuals that share phenotypic and genotypic characteristics. Pioneering studies tested loci previously known or suspected to have a role in human ageing or longevity. Mainly, mutations close to *APOE* and *FOXO3A* genes were consistently replicated (Christensen, Johnson, and Vaupel 2006; Willcox et al. 2008; Schächter et al. 1994), but also some others (Budovsky et al. 2013; Stessman et al. 2005). The *APOE* gene was postulated as a candidate because a certain allele notoriously increases the risk for Alzheimer's and heart disease (Christensen, Johnson, and Vaupel 2006; Schächter et al. 1994; Davignon, Gregg, and Sing 1988). Ortholog of *FOXO3A* is known to increase lifespan in *C. elegans*, *Drosophila* and mice (Murphy 2006).

Despite the handful of genes identified using these more classical methods, the genetic variance explained for senescence phenotypes most likely does not come only from a handful of loci, but from the combined action of many common polymorphisms, usually of a small effect size. Definitely, the true complexity of the genetic architecture of ageing, senescent or longevity phenotypes cannot be captured through traditional linkage, familiar association, or candidate gene methods.

1.4.6 A call for a gold standard technique: Genome-Wide Association Studies (GWAS)

The plethora of **questionable associations that infested the field of candidate-gene studies** could have ruined the prospects of ever discovering the genetic basis of disease and other complex traits. Criticism was independently raised by John P. A. Ioannidis et al and Joel N. Hirschhorn et al. (Ioannidis et al. 2001; Hirschhorn et al. 2002) Ioannidis *et al.,* reviewed 370 studies and 36 genetic associations, showing that ORs in first and subsequent replication attempts were only modestly correlated, with a considerable number of cases in which this discrepancy went beyond what would be expected by chance. They also evaluated up to which extent the so-called *"winners' curse"* was real. This is: results obtained in the first studies tend to overestimate real ORs, when compared to subsequent studies, highlighting thus spurious inflated findings. Outcomes of this experiment revealed up to eight associations initially published by prestigious journals that were affected by this phenomenon. Nonetheless, other associations remained solid after the analysis.

Similarly, Hirschhorn et al. (2002) reviewed >600 associations reported as positive findings. They showed that only 6 out of 166 associations studied more than 3 times were replicable. A wealth of statistical, biological and sociological reasons has been put forward to explain replication failures. Among these; population stratification, linkage disequilibrium, gene-gene and gene-environment interactions or weak genetic effect and lack of power were proposed as putative failures of candidate-gene association studies.

The poor outcome of candidate-gene association studies called for a more trustable system of evaluating gene-disease connections. **Genome-wide association studies' (GWAS)** era had arrived.

1.4.6.1 What is in a GWAS?

Despite contrary predictions (Hirschhorn et al. 2002), technological development of commercial **chips for high-throughput genotyping** improved the availability of polymorphisms to a density of up to several SNPs per kilobase making it feasible to look for common variants associated to complex traits by means of GWAS. Additionally, unraveling of genome-wide LD patterns informed about the possibility to capture ~80% of the predicted **>10 million of common SNPs** segregating in human populations with a scattered selection of just 0.5 to 1 million SNPs (Visscher et al. 2012). As in candidate gene association studies, GWAS rely on comparing allelic or genotypic frequencies between cases and controls, but on the whole genome. Significant departures of expected frequencies between both groups will result in an associated signal (Figure 8). Most traits, such as human height or even ageing, tend to present continuous variation across individuals. This is because they are controlled by large numbers of genes and each causal variant explains a tiny fraction of the overall phenotypic variation. In this regard, GWAS have arisen as one of the most powerful tools to unravel the alleles that underlie individual phenotypic variation.

The first GWAS were published in 2005 and 2006 (Klein et al. 2005; Dewan et al. 2006). Even if using few markers and samples (<100,000 SNPs genotyped upon <200 individuals), both studies managed to find common variants associated to age-related macular degeneration thanks to the large effect size of causal variants (OR>2). In 2007, the Wellcome

Trust Case Control Consortium published a GWAS for seven different diseases using a set of 3,000 shared controls (WTCCC 2007). The WTCCC paper became a landmark due to the large number of samples used (2,000 per disease), the confirmation of the small effect of population stratification in Europeans, the setting of genome-wide significance thresholds and the replication of previous signals. Dropping costs of commercial arrays and the development of computational tools helped in the exponential increase of published GWAS (Welter et al. 2014; Clarke et al. 2011).

GWAS present two key differences with respect to candidate gene studies. First, there is an inherent issue related to numbers. Previously, a few tens of markers and, at most, a few hundreds of individuals were analyzed. In contrast, **GWAS studies test millions of markers** (many of which are just imputed, that is, *inferred* using statistical techniques) in thousands of individuals gathered from diverse cohorts. Second, **GWAS are said to be "hypothesis-free"**: they do look for common risk effects, but they do so by looking at SNPs scattered across the genome and without any *a priori* list of candidate loci (Ziegler, Konig, and Pahlke 2010). Thus, the adjustment for multiple testing constitutes an important decision in GWAS. The two most used cut-offs are 5×10^{-7} (WTCCC 2007) and 5.0×10^{-8} (strict Bonferroni), but there are several available methods available to select proper significance thresholds (McCarthy et al. 2008) (Figure 8).

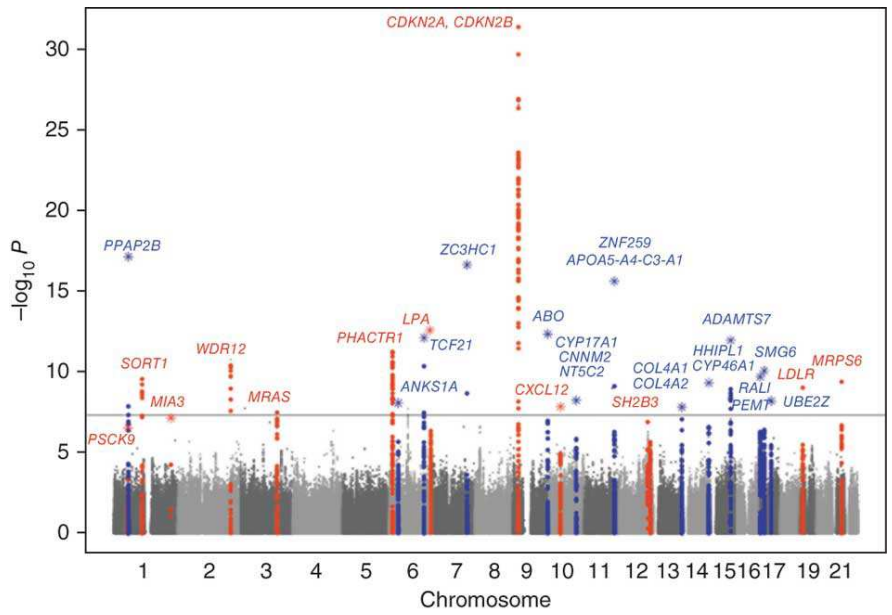


Figure 8: Manhattan plot showing *p*-values from a GWAS. *X*-axis represents SNPs ordered by physical distance and *Y*-axis represents the *p*-value attained in GWAS. SNPs over the line are those considered genome-wide significantly associated (5×10^{-7}) with the trait, according to the WTCCC (2007).

At the time of writing this thesis, more than 2,500 published studies were recorded in the public catalog of GWAS (NHGRI-EBI GWAS Catalog, www.ebi.ac.uk/gwas/) maintained by the European Bioinformatics Institute (Welter et al. 2014). However, GWAS present a heavy bias towards using exclusively individuals of European genetic ancestry. In 2010, an early survey of GWAS publication patterns found that >80% of studies did not use any cohort of non-European ancestry (Rosenberg et al. 2010), although this bias has decreased in recent years (Fu, Festen, and Wijmenga 2011). Interestingly, non-European GWAS present challenges related to imputation ability, genomic coverage, and statistical power (due to the ascertainment of SNPs in commercial arrays). However, several reasons fuel the case for GWAS generalization across populations to better achieve the objectives of complex disease mapping (Rosenberg et al. 2010). The ability to detect disease variants can vary if they have

different effect size or present disparate allele frequencies across populations (Adeyemo and Rotimi 2009). Additionally, the use of diverse populations allows taking advantage from the variation in LD across ethnical groups and thus help in the fine mapping to narrow the location of causal variants (Visscher et al. 2012). In any case, comparisons of GWAS replicability have reported high rates of concordance across populations (Marigorta and Navarro 2013).

1.4.7 Knowledge gained from GWAS and potential applications.

By mid-2016 there were more than 4,000 loci that have been robustly associated to disease with a *p*-value $\leq 5.0 \times 10^{-8}$ (Welter et al. 2014). The number of loci identified for each disease has increased exponentially when compared to associations discovered and replicated through candidate gene approaches. This observation emphasizes the limitations of an approach based on biological candidates compared to the "hypothesis-free" design of GWAS. Pathway analysis of the discovered loci has provided unsuspected insights into the biological mechanisms of human complex phenotypes. The range of new understanding applies to specific traits (*e.g. IL23R* role in ankylosing spondylitis) as well as shared etiology across phenotypes (*e.g.* loci associated to disparate autoimmune disorders).

After the confirmed low replicability of candidate gene studies GWAS were eventually seen in a similar pessimistic way. In the next section I will dig a bit into this debate. Evaluating the reliability of GWAS by studying its replication patterns through time, and their ability to predict disease genetic risk, are two of the results derived from this PhD thesis. These are

important preconditions to be met if one wants to use these data to perform the task of examining two evolutionary theories of ageing relying in such data.

1.4.8 Translating GWAS findings to clinic

The prospects for personalized genetic medicine depend on the ability to **translate true GWAS findings into the clinical setting**. One immediate possibility lies in the development of genetic profiles based on GWAS risk markers available through consumer testing (Jakobsdottir et al. 2009; Kraft et al. 2009; S. H. Lee et al. 2008). Potentially, genetic profiles based on risk markers may distinguish between high-risk and low-risk groups of individuals for particular disease phenotypes. So far, additive models constructed from a few GWAS variants have been shown to lack the power to pinpoint individuals that will develop a given complex disease (Jakobsdottir et al. 2009).

The current performance of predictions for individual genomes remains unclear (Burga and Lehner 2012; Jelier et al. 2011) and only the combination of larger sample sizes and improved genomic coverage will raise to the hopes for personalized genetic testing.

Results derived from this PhD thesis highlight the importance of integrating evolutionary biology into medicine, and how disease prediction accuracy of current genetic association results could, in the future, help informing decisions in healthcare. These and other complementary results will be further discussed in section 3.2.

1.4.9 Novel strategies uncovering senescence factors

Progress in the genomics area accelerated once the human genome was completed in 2001 (Lander et al. 2001), and continues to advance as high-throughput-omics technologies become more accessible (Goodwin, McPherson, and McCombie 2016). **Accessible "Big Data" public databases** (Lappalainen et al. 2015; Mailman et al. 2007) allowed researchers to develop further work reusing genome-wide data for new purposes and lines of research, with optimal results (Pickrell et al. 2016; Fortney et al. 2015).

All along the writing of this thesis I have been linking the concepts of **disease and senescence**. Both of them tend to come together, although one does not necessarily imply the other (Erikson et al. 2016). An inherent problem to study ageing through GWAS is that centenarians are uncommon (less than 0.02% in the US Census Bureau for 2010) and gathering a large and reliable cohort of individuals can be costly. However, it is true that genetic variants associated with age-related disease are more likely to be associated with longevity (Fernandes et al. 2016). Taking advantage of this principle, Fortney et al. (2015) gathered genome-wide data from a meta-analysis of 24 disease and traits related with ageing. Using genome-wide *p*-values from each of these traits they selected the top scoring SNPs and created a QQ plot with the *p*-values obtained for these SNPs in the largest GWAS for longevity by that time (n=801 centenarians) (Sebastiani, Solovieff, et al. 2012) against what could be expected by chance. This allowed them to identify loci for 9 diseases genetically overlapping with longevity, emphasizing the value of public data to discover new loci related to this condition. This represents a remarkable advance, considering that the only hit consistently replicated in

GWAS of ageing were only a handful of variants near *APOE/TOMM40*, also involved in Alzheimer (Sebastiani, Solovieff, et al. 2012; Deelen et al. 2011; Deelen et al. 2014; Broer et al. 2015). Additionally, another locus at chromosome 5q33.3 was associated with longevity at genome-wide significance in one study, but has not been replicated yet (Deelen et al. 2014).

A way to bypass the mentioned limitation of gathering a significant amount of centenarian samples is to use indirect methods. Pilling et al. (2016) used public genotype data (UK Biobank; n=75,000) where age at death for parents of middle-aged genotyped individuals was collected. They conducted a GWAS on longevity using an individual's parents age of death as a proxy and were able to replicate previously ageing-related loci, many of them related also to disease

1.4.10 High-throughput sequencing and the genetics of senescence

Next-generation sequencing has also turned into a very useful and promising tool to unravel the heritability of diseases (Eichler et al. 2010). The interest of uncovering genetic factors that contribute to age-related diseases spurred the motivation of discovering **particular variants, either common or rare, leading to super-centenarian ages**. The sequencing of two super-centenarians (>114 years old) (Sebastiani et al. 2012) revealed that rare, but also common variants, play a role in longevity. Both of the individuals sequenced suffered from ageing-related conditions. This prompts us to make a distinction between common ageing (involving disease) and healthy ageing, a distinction that Greek myths already made.

Recently, Erikson et al. (2016) sequenced a cohort (n=600) of healthy aged people (median age: 84 years), that is, free from a certain spectrum of age-related conditions. The main finding was that **healthy ageing is a different phenotype from longevity**. This was partially a consequence to enrichment with disease protective genetic variants. Protection against cognitive decline was among the culprits of achieving a vigorous ageing. In other words, fit ageing is not simply the result of pure luck, a finding that has been recently reconfirmed (Schwarz et al. 2016).

Sequencing strategies were also applied to reveal secrets of longevity in other organisms known to present extraordinary patterns of senescence. Among others, two clear examples of long-lived organisms had their **genomes sequenced**, namely ***Hydra magnipapillata*** (already discussed in section 1.2) (Chapman et al. 2010) and the **bowhead whale** (Keane et al. 2015). In the *Hydra* genome paper authors did not describe any direct genetic controller responsible for immunity to senescence; rather, they found that *Hydra*'s genome is characterized by bursts of transposable element expansions, horizontal gene transfer, trans-splicing, and simplification of gene structure and gene content. A combination of these elements can be responsible for the peculiar life cycle of the organism.

On the contrary, the genome of the bowhead whale (longest lived mammal), shed light on its longevity by identifying positively selected genes and whale specific coding mutations associated with cancer and aging, besides other minor and major genetic changes related to several physiological processes, which may explain its long life.

These approaches call for a diversified perspective if we wish to fully understand ageing and senescence, abandoning the anthropocentric vision

of current research in the field given the high diversity of senescence patterns across the tree of life (Figure 5) (Jones et al. 2014). Resources on a wide variety of genetics of ageing through the tree of life fostered initiatives like AnAge (de Magalhães et al. 2009; de Magalhães, Curado, and Church 2009), where all the genes that had been at least once related to ageing or senescence phenotypes in any organism are gathered.

2. OBJECTIVES

"*And meanwhile time goes about its immemorial work of making everyone look and feel like shit.*"

Martin Amis (British writer, b. 1949);
London Fields (1989)

The increasing ageing of the World's population has spurred the interest on the causes and mechanisms of senescence, the physical decay of organisms with age which implies infirmity, physiological decay and increased disease risk. Senescence has long been a mystery and, as of today, there is no single universally accepted theory that accounts for the ultimate evolutionary purpose of senescence (if indeed there is one). Two of the most popular evolutionary explanations proposed so far are the Mutation Accumulation Theory, proposed by Medawar in 1951, and the Antagonistic Pleiotropy Theory, suggested by Williams in 1957. However, an exhaustive assessment at genome-wide level of the impact of their effects in the senescence of our species has not yet been carried out. This is **the main objective** of the present work.

Within this context, this PhD thesis presents **four specific goals**:

1. To test the soundness of GWAS findings by assessing both, their replicability patterns across time and their quality when used to predict disease risk.

2. To study comorbidity from the perspective of the genetic risk of complex disease considering, for the first time, the age of onset of diseases as a fundamental factor. Derived results may unveil a series of potentially medically relevant comorbidities, especially those involving early and late onset diseases.

3. To test whether the Mutation Accumulation Theory of Senescence can explain the patterns of mutation-selection balance shown by disease-associated alleles, discovered by GWAS in current human populations.

4. To assess, for the first time, up to which point the predictions of the Antagonistic Pleiotropy Theory of Senescence are met at the genome-wide level in humans.

As explained in the introduction, the fulfillment of the first specific goal is critical for us to ensure the proper formal testing of the two theories of ageing discussed in this work. Only if GWAS data are sound does it make sense tackle the other goals and use these data to test evolutionary ageing theories.

3. RESULTS

3.1.

Marigorta UM, Rodriguez JA, Navarro A. GWAS replicability across time and space. In: Appasani K, editor. Genome-Wide Association Studies. Cambridge: Cambridge University Press; p. 53–66. DOI: 10.1017/CBO9781107337459.006

The standard procedure to validate scientific findings is by assessing their replicability in subsequent studies. In our case, if we wish to use GWAS data for further goals, we need to confirm that they are reliable and robust. This exercise is the main aim of this first results section, where we evaluated the replicability of the GWAS discoveries across time.

3.2.

Rodríguez JA, Marigorta UM, Navarro A. Integrating genomics into evolutionary medicine. Curr Opin Genet Dev. 2014 Dec;29:97–102. DOI: 10.1016/j.gde.2014.08.009

Evolutionary medicine is an emerging field whose origins date back to George C. Williams and which considers illness in the context of evolutionary adaptations. The two cornerstone evolutionary theories of senescence that we are trying to demonstrate in the present thesis constitute two evolutionary explanations for disease in old ages. Once we confirmed that GWAS produced solid and replicable results, we aimed to evaluate up to which point this information can be translated to the medical field by, for instance, helping to predict genetic disease risk and informing healthcare decisions from an evolutionary point of view. The suitability of these data to inform complex disease risk is another key condition to satisfy if we wish to anticipate disease onset. Still, two main problems that thwart the widespread application of genomics into evolutionary medicine will be discussed.

3.3.

Rodríguez JA, Marigorta UM, Hughes DA, Spataro N, Bosch E, Navarro A. Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. Nat Ecol Evol. 2017 Jan 30;1(3):55. DOI: 10.1038/s41559-016-0055

(Under Review in: *Nature Ecology and Evolution*)

Two of the most widely known evolutionary theories to explain senescence are tested here using, for the first time, genome-wide human complex disease data. We found clear evidences that actual data fit the predictions of both theories.

4. DISCUSSION

"That you get ugly when you get old. It's all perfectly simple.
In fact I can tell you how it's going to go.
Everything seems fine until you're about 40.

Then something is definitely beginning to go wrong."

Martin Amis (British writer, b. 1949);
when asked about ageing and beauty.

The last chapter of the introduction highlighted both, the opportunity that public data pose to conduct research in genomics of ageing and the interesting and unexpected results that can be produced. Public availability of these data leaves us in an optimum position to test two evolutionary ageing theories in an unprecedented way.

This discussion will be articulated in three parts. The first two parts (sections 4.1 and 4.2) will be dedicated to elaborate on the main findings reported in results section 3.1 and 3.2 and will discuss how GWAS have produced consistent and mostly solid results, whether these can be directly translated to the medical field and the possible drawbacks of such a step. The last section of the discussion (4.3) will be finally devoted to demonstrate that by using this robust and trustworthy public data I can shed new light over a problem that has remained unsolved problem for centuries.

4.1 RELIABILITY OF GWAS DATA

A final purpose of the quest for disease-related genetic markers is to take profit from them to help anticipating the risk, onset and progression of diseases. This is what we know as medical genomics. GWAS have fostered this field since they arose about ten years ago. However, the soundness and quality of GWAS data are crucial if we wish to push forward this emerging field. Replicability of any GWAS finding in an independent set of individuals constitutes a standard key step in validating new discoveries. GWAS entered the genomic scene called to be the solution to the lack of replicability inherited from the candidate-gene era (Lohmueller et al. 2003). However, a systematic evaluation of the replication in GWAS since they appeared had not been yet fully performed until the completion of this PhD.

A first exploratory analysis on GWAS' replication patterns was carried out earlier in our research group, focusing in geographical replication. In particular, earlier work by Marigorta et al. (2011) showed that candidate-genes' lack of replicability was due to populational differences. In a later and deeper analysis, Marigorta and Navarro (2013) using exclusively GWAS variants demonstrated that common risk variants are highly shared within Eurasian populations.

The first GWAS were performed with relatively low sample sizes (Klein et al. 2005) discovering variants of high effect size. However, high effect size variants are not very common and will be efficiently captured even by early low-powered studies. GWAS were aiming to disentangle the genetic architecture of complex diseases, which usually involves many variants

with medium to small effects, requiring sample sizes several orders of magnitude higher than those seen in the first GWAS.

Most of the success of unmasking small effect variants was due to meta-analyses, where individuals from several independent studies were combined to increase discovery power, as a consequence of increasing sample size (Appendix 1). The first wave of these studies was performed during a time (around 2010) in which skepticism about GWAS was encouraged by the fact that the heritability of the diseases under study explained by GWAS hits was below expectations. In that sense, these meta-analyses were crucial to reassure the validity of GWAS. Following this, we currently know that hits returned by the leading WTCCC study in the first semester of 2007 (WTCCC 2007) have been subsequently replicated over and over again, as have many other hits coming from other GWAS, as we can see in Figure 12. That figure shows that results from GWAS are not only comprising new discoveries, but also abundant replicas of previous findings. This picture contrasts sharply with the landscape generated during the candidate gene era, during which barely 3.6% of results were robustly replicable (Hirschhorn et al. 2002).

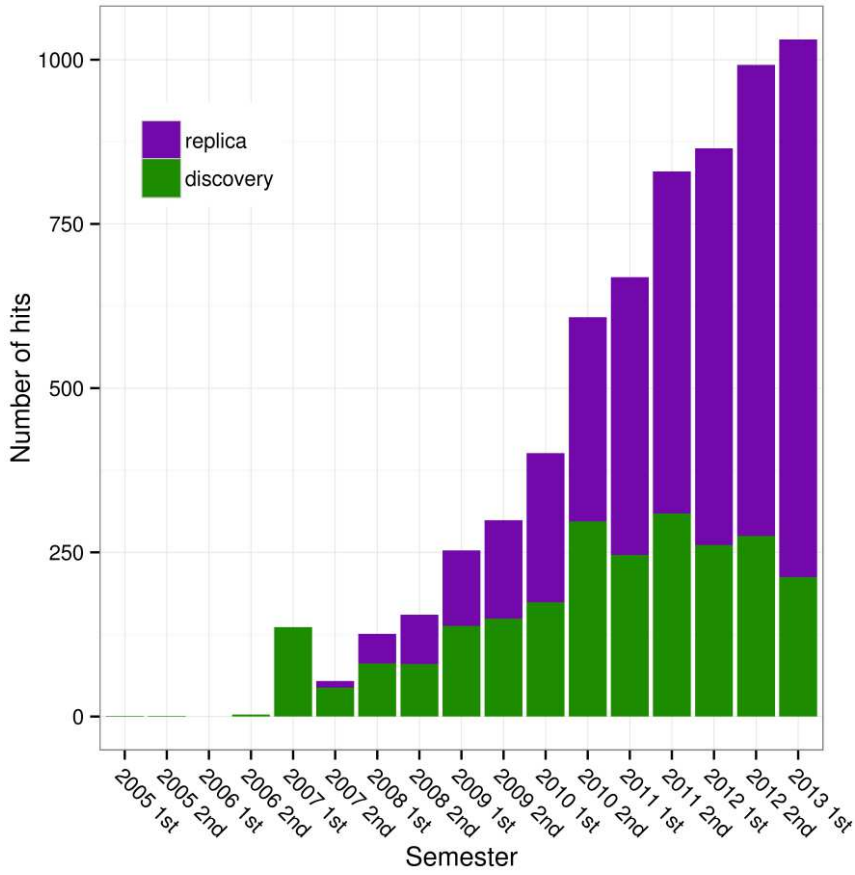


Figure 12: Results derived from GWAS during the period 2005-2013. I represent newly discovered loci and subsequent replicas (considered independent if lying under 200 kilobases from an existing hit), for Eurasian populations for the 46 diseases with the highest number of loci discovered. Replicas are cumulative across semesters.

An interesting view can be gained from replicability trends when comparing different groups of diseases (Figure 13). Findings on mental-related phenotypes did not present any replica until 2011, when large sample sizes became available, reflecting their highly polygenic component.

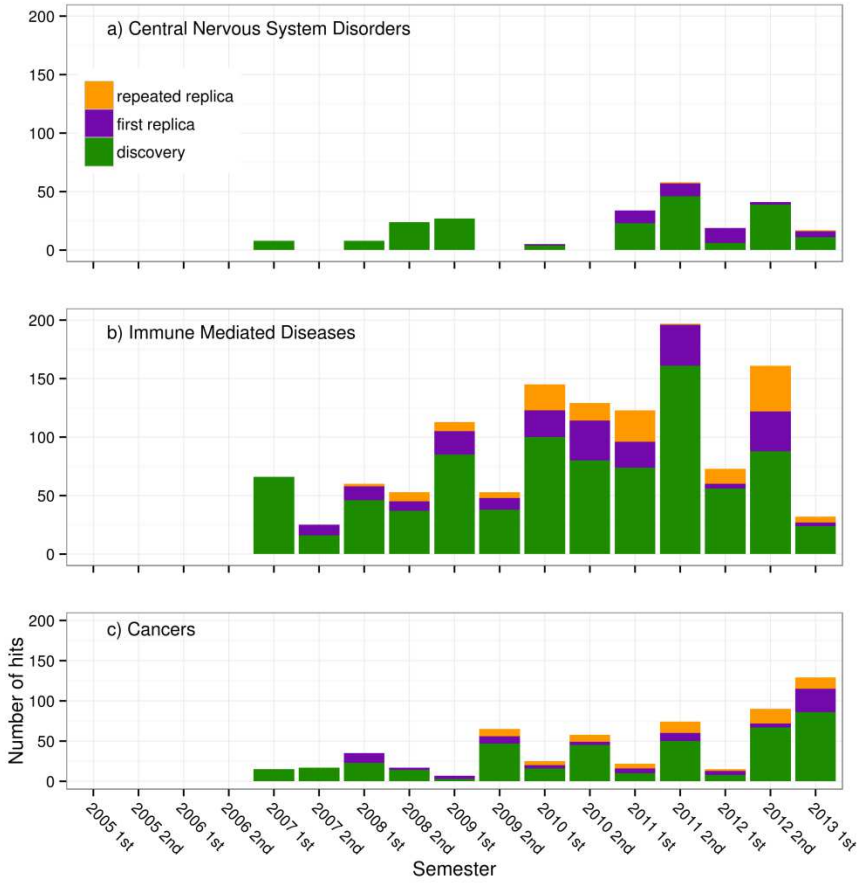


Figure 13: Numbers of discovered loci, first replicas, and repeated replicas for three different categories of disease. (a) Two disorders of the CNS, namely schizophrenia and bipolar disorder; (b) immune-mediated disorders (n = 17); (c) cancer (n = 11).

On the contrary, immune mediated phenotypes quickly got their replicas, while cancers were mostly benefited by meta-analyses that could be conducted at quite regular time intervals. The distinct replication patterns observable for different types of disease reflect their underlying genetic architectures, shaped by natural selection and other phenomena in ancient times. This highlights the need of accounting for evolutionary approaches if we want to translate these findings to a clinical setting, as emphasized next.

4.2 INTEGRATING GENOMICS INTO EVOLUTIONARY MEDICINE

The field of medical genomics offers an interesting perspective about the origin of some diseases. Essentially, evolution provides the context that helps explaining why we get sick the way we do (Nesse and Williams 1998; Nesse and Williams 1994). After all, what we see today as a disease that impairs reproductive fitness may have been placed in a totally different scenario in the past, reflecting adaptations to ancient environments or conditions. The idea that natural selection, and broadly, evolutionary biology, can help us understand human disease was firstly enunciated by **Paul W. Ewald** (Ewald 1980), **Randolph M. Nesse** and **George C. Williams** (G. C. Williams and Nesse 1991), who started the field of "**evolutionary medicine**". They and others (G. C. Williams and Nesse 1991; Nesse and Williams 1994; Nesse and Stearns 2008) emphasize that disease should be actually seen as a combination of at least three elements: the evolutionary history of the human lineage; changing environment of the past and the current evolutionary forces to which we are exposed. A deeper understanding of these three elements and their interactions will be fundamental for the advance of medicine. We have just seen that a high number of signals discovered by GWAS had a replica in subsequent studies, showing that indeed (some) truth lies in them. Gaining the insight offered by the study of the genomics of disease will be of key relevance for the field. When both approaches are combined the result is "**evolutionary genomics medicine**" (EGM).

The MA and AP theories propose an explanation for senescence and age-related diseases, offering an evolutionary explanation to a medical problem. These two theories were proposed more than 50 years ago.

However, despite all previous work on both theories, public genome-wide data on genetic association to complex disease had never been put in the context of understanding senescence. Public availability, robustness and demonstrated replicability allow carrying out detailed tests of both, MA and PA.

In what follows, I will discuss how two obstacles currently hinder the direct integration of evolutionary knowledge in medical decisions and how by overcoming them we could foster EGM.

4.2.1 Dissociation genotype-phenotype

The first obstacle stems from the dissociation between genotype-phenotype. This problem derives from the enormous difficulties in the translation of the observable phenotypes to the molecular machinery behind them. The molecular pathways responsible for even the most obvious phenotypic adaptations, such as the textbook examples of our opposable thumb and capacity for language, remain largely unknown (Lewontin 1974; Culotta and Pennisi 2005; Nielsen et al. 2005). Being able to bridge this gap between observed genetic variation and phenotypic variance is not a straightforward task, but it is vital if we want to take profit from the discoveries made.

Anyway, how translatable are genetic markers to real phenotypes? Up to which point? By using public data (WTCCC 2007) we show that genetic discoveries made during the highpoint of the GWAS era (2008-2014) had a significant positive impact when used to predict if an individual will develop a disease. Besides being replicable, GWAS data have shown a real value when translated to clinical terms. As expected, we obtained different

results for the five conditions studied, which pointed to differential genetic architectures and varied environmental or genetic contributions to the traits. This pattern is coherent with what we obtained previously when evaluating replicability across groups of diseases.

The general increase in prediction ability observed over time was mainly due to larger sample sizes (See Figure 3 in the article, section 3.2), reinforcing the general patterns also seen previously (Appendix 1). Prediction ability for Crohn's disease and bipolar disorder was particularly increased, even though other conditions may need larger sample sizes or a refinement of the approaches to account for environmental variance to detect the infinitesimal architecture of these traits.

4.2.2 Dissociation fitness-phenotype

To apply the principles of EGM (G. C. Williams and Nesse 1991; Nesse 2008), we also need to consider how disease, health and fitness were associated in the past. This leads to the second obstacle: Dissociation fitness-phenotype. An important proportion of the diseases or responses of our body that we consider detrimental exist because they eventually represented a protective response. Ideally, what EGM needs is a proxy of the impact of a disease into reproductive fitness. This is an important issue if we wish to certainly test evolutionary medicine theories, particularly in the case of the AP theory. As explained in the introduction, the AP theory suggests that deleterious mutations associated with early onset diseases will impair fitness during or after the fertile periods. In this sense, we are taking for granted that most of the early onset conditions of today, which we are using for our study, were detrimental to fitness in the

past. I cannot help but thinking that, for example, vitiligo probably did not result in too much impairment at the time of reproduction during Paleolithic times. Some other pathologies, such as childhood cancers, were clearly detrimental. However, the impact of some conditions in reproductive fitness remains intriguing. In our work we have compared the current prevalence of seven disorders across five large world areas. Results are showing that even very close regions have strikingly different prevalence of the same disease (Figure 14), reflecting the assorted population histories of regions and the difficulties of inferring net fitness impact of a disease.

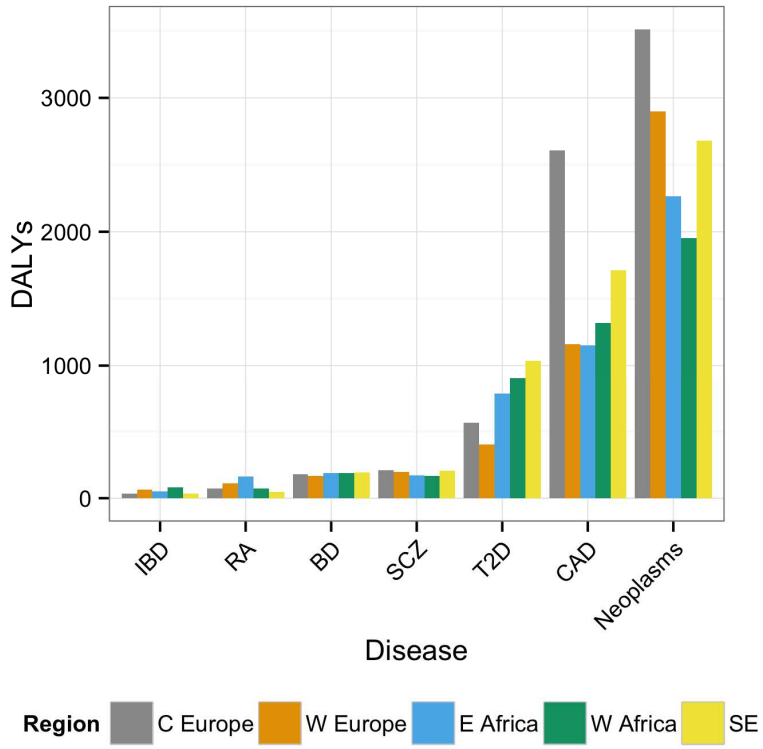


Figure 14: Disability-Adjusted Life Years (DALYs) lost to certain diseases across the World. DALYs are units used by the World Health Organisation as a health statistic. One DALY can be thought of as one lost year of "healthy" life. DALYs are provided in rate/100,000, that is, out of 100,000 DALYs lost in the population how many of correspond to each disease in each region. Data was gathered from the Institute for Health Metrics and Evaluation (www.healthdata.org/).

In the end, the differential evolutionary histories of *Homo sapiens* populations shape differential susceptibilities to diseases.

Recent examples have shown that particular environments in ancient, but also in current times, induce particular molecular adaptations in populations. For example, strong evidence supports a high geographical heterogeneity in effect sizes for the same variant between cohorts analyzed for obesity in the United States and in Europe (Kilpeläinen et al. 2011). Further insights can be obtained by performing GWAS in other populations than the standard ones. Moltke et al. (2014) identified a non-synonymous variant in Greenland Inuit people conferring risk for type 2 diabetes with an effect size several times larger than regular signals. The possibility that this variant was selected as protective because of the "feast and famine" episodes experienced in the past favors the *thrifty genotype* hypothesis (Neel 1962). This hypothesis recently gained further support with the discovery of a missense variant associated with body mass index having a huge effect size (1.36-1.45 kg/m² *per* effect allele) in Samoan islanders (Minster et al. 2016). However, the high preponderance of these alleles may sometimes come with a fitness cost. A study on an Arctic population (Clemente et al. 2014) revealed selective sweeps around a gene carrying a non-synonymous variant, whose derived functional allele is related with diseases such as hypoketotic hypoglycemia and high infant mortality, but which still segregates at high frequencies in current populations. The authors suggest that this variant increased in frequency over the last 6–23 ky, probably after favoring adaptation to a high-fat diet or to cold life conditions.

This vision of disease shaped by evolutionary histories leads us to recall now the main objective of the present PhD: to test evolutionary theories of senescence. Despite unavoidable caveats already discussed and the still

long journey to real application of EGM in clinical settings, the availability of quality data sets us now in an optimal position to investigate the influences of these cornerstone theories in framing human disease and senescence. I have stated in several places that an approach to the problem such as the one we developed here had never been tried before. The main reason is, of course, that it is only recently that the required data are at hand.

4.3 EVIDENCE THAT MUTATION ACCUMULATION AND ANTAGONISTIC PLEIOTROPY INFLUENCE HUMAN SENESCENCE AND DISEASE

4.3.1 What do our results show?

The rationale behind this last but central piece of work from the present PhD thesis was to provide the first systematic, genome-wide evidence for evolutionary theories of senescence in our species. This will shed new perspectives on the long-term debate about the causes of senescence and, may help revealing unexpected connections between apparently unrelated pathologies.

Our main results show at least five things:

1. That risk alleles with a notable impact on fitness (OR $\geq$ 2) for late onset diseases have higher frequencies than alleles increasing the risk of early onset diseases. This favors the MA hypothesis because it is consistent with natural selection keeping early onset deleterious alleles under harsher purifying selection (see Figure 1A in the manuscript, section 3.3). In

particular, when using age 40 or younger as a cut-off for the early-late onset transition, frequencies of early onset diseases are always significantly lower, reflecting the higher efficiency of natural selection during the early ages of a cohort.

2. That SNPs associated to late onset diseases explain a higher proportion of the variance of their associated diseases. Again, this favors the MA hypothesis because stronger effects can be tolerated by weaker purifying selection. Differences between both early and late onset groups stop being significant at age 50, emphasizing again a biological barrier around 40-50 years (see Figure 1B in the manuscript, section 3.3).

3. That there is an excess of early-late antagonistic pleiotropies. In other words, some mutations that harm the organism at old ages also seem to have beneficial effects on health at younger ages. This favors the AP theory because it predicted the existence of such mutations.

4. That we found suggestive evidence of the action of natural selection in 3 loci out of 19 involved in antagonistic early-late pleiotropies. The early onset diseases of these pleiotropies imply really strong consequences to reproductive fitness, that is, theory would predict that a protective effect would be advantageous and favored despite detrimental consequences later on (Table 2).

EARLY ONSET			LATE ONSET				
SNP	**Disease**	**Ancestral Allele**	**SNP**	**Disease**	**Ancestral Allele**	**Derived is early protective**	**Gene**
rs1295686	Asthma	T	rs20541	Psoriasis	G	YES	*IL13*
rs1295686	Atopic dermatitis	T	rs20541	Psoriasis	G	YES	*IL13*
rs20541	Hodgkin's lymphoma	G	rs20541	Psoriasis	G	NO	*IL13*
rs2157719	Glioma	C	rs523096	Glaucoma	A	YES	*CDKN2A*
rs2157719	Glioma	C	rs564398	Type 2 diabetes	T	YES	*CDKN2A*
rs2157719	Glioma	C	rs7865618	Coronary heart disease	A	YES	*CDKN2A*
rs2157719	Glioma	C	rs1412829	Nasopharynx carcinoma	A	YES	*CDKN2A*
rs11755724	Multiple sclerosis	A	rs11755724	Age-related macular degeneration	A	YES	*RREB1*

Table 2: Early-late antagonistic pleiotropies detected in the genes showing evidences of natural selection: *IL13*, *CDKN2A* and *RREB1*. Note that almost all of the protective alleles are derived, compared to chimpanzee.

One of these three genes, *CDKN2A,* is particularly outstanding because of the clear role it plays in ageing and disease. Several molecular signals for recent positive selection have been found around this gene. Among others, the most interesting were two selective sweeps, one complete and another incomplete, surrounding this locus. Further information, analysis and images regarding its molecular biology and natural selection patterns can be found in the Supplementary Information section 7 of the manuscript (section 3.3).

Recently, three papers (Fortney et al., 2015, Pilling et al., 2016, Erikson et al., 2016) highlighted the role of *CDKN2A* in ageing (either healthy or poor) and exceptional longevity. Interestingly, Fortney et al. identified four loci (*APOE/TOMM40*, *CDKN2B/ANRIL*, *ABO* and *SH2B3*) associated with exceptional longevity. Except for *ABO*, the other three were involved in 30 of our pleiotropies (four of them being antagonistic

early-late), which makes sense because we found evidences that ageing-related genes carry a significant excess of pleiotropies (Table 1 in the paper, section 3.3).

In the previous section I discussed the different prevalence of diseases across populations and the need for an EGM approach to understand it. The main early onset disease related with *CDKN2A* is glioma, a deadly child brain tumor accounting for up to 24% of child cancers in UK (Cancer Research UK 2016). Quite clearly, this poses a large handicap in terms of reproduction; now and 10,000 years ago. Prevalence in terms of DALYs (Disability-Adjusted Life Years) lost to this cancer in three large world regions (Figure 15) are strikingly dissimilar; being practically absent in Sub-Saharan Africa and Southeast Asia, where the protective allele is virtually fixed, but mildly prevalent in Europe, where the allele segregates at intermediate frequencies.

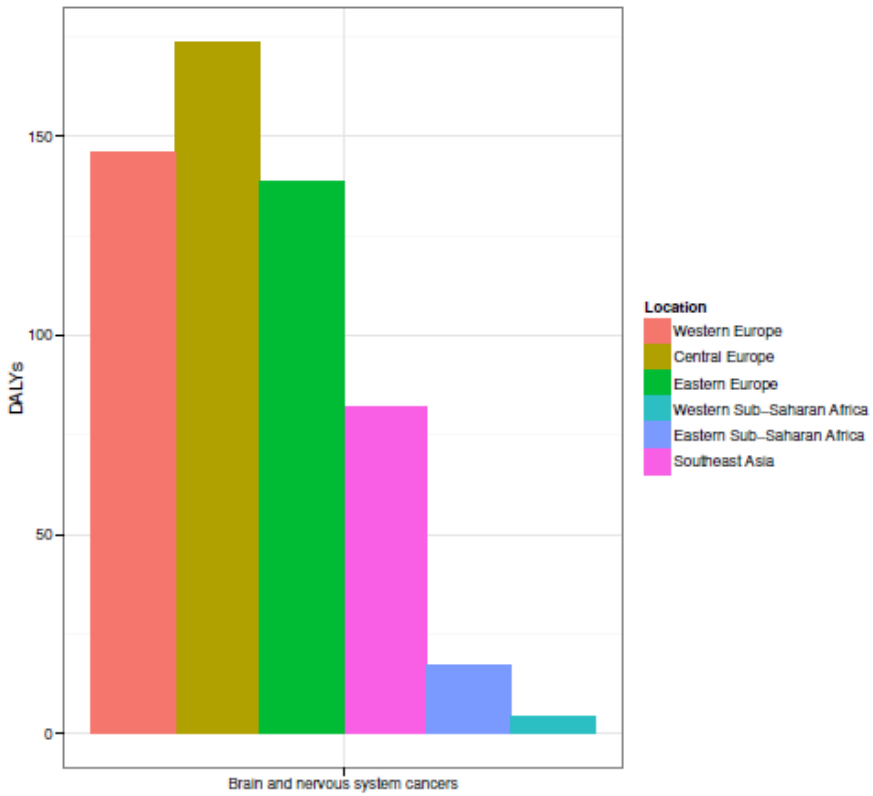


Figure 15: Disability Adjusted Life Years (DALYs) for brain and nervous system cancers in three world regions. DALYs are provided in rate/100,000. Data was gathered from the Institute for Health Metrics and Evaluation (www.healthdata.org/).

If the higher frequencies of the protective allele for glioma are reflected in fewer DALYs lost in Africa and Asia, the opposite pattern should also be expected. That is, DALYs lost to the four late onset diseases linked to *CDKN2A* (type 2 diabetes mellitus, ischemic heart disease, glaucoma and nasopharynx cancer) should be significantly higher both in Sub-Saharan Africa and Southeast Asia than in Europe, a prediction that is clearly fulfilled (Figure 16).

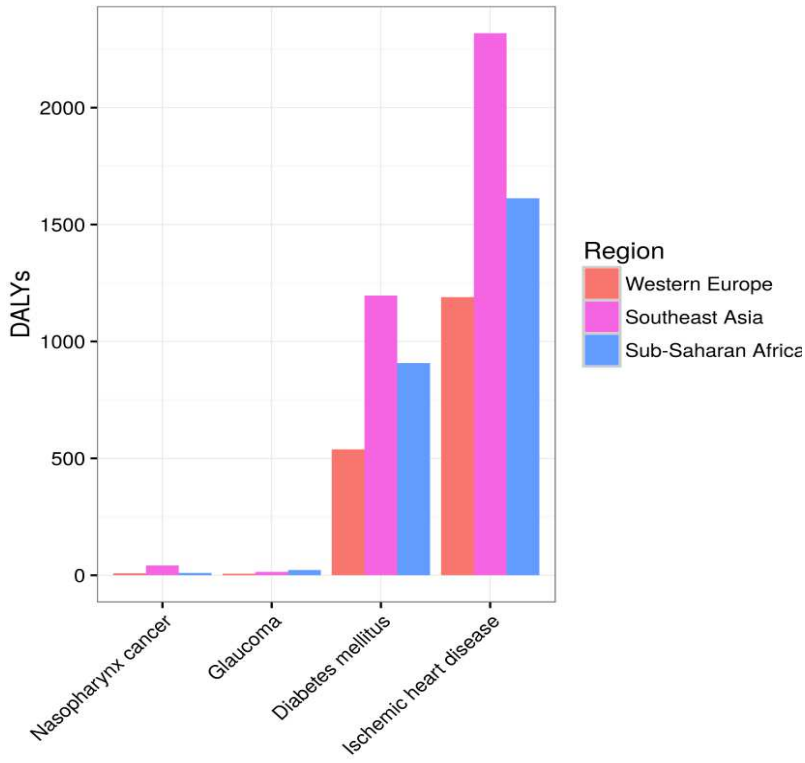


Figure 16: Disability Adjusted Life Years (DALYs) lost to the four late onset conditions involved in pleiotropies in the *CDKN2A* gene. Note the lower DALYs lost in Europe for all four conditions.

5. That genes harboring such early-late antagonistic pleiotropies present a very significant overlap with several sets of genes that have been related to senescence and ageing. In addition, pleiotropic regions and regions that change methylation state with age (differentially methylated regions; DMRs) present a statistically significant overlap. Again, this favors the AP theory because it strongly links age-related antagonistic pleiotropy with senescence (see Table 1 in the manuscript, section 3.3)

In conclusion, as predicted theoretically by **Peter Medawar** and **George C. Williams**, we found clear evidences that senescence and disease are influenced by the phenomena of mutation accumulation and antagonistic pleiotropy.

4.3.2 Placing results into a biological context

Around the period where the excess of antagonistic pleiotropies early-late is more significant (40-50 years old) a physiological decay arises with the onset of menopause, even though, in current populations, this is not linked with an increase in mortality. Still, evaluation of medical records in industrialized societies with long life spans such as the United States revealed that these ages go together with an intense increase of cardiovascular diseases, a leading cause of mortality (Parker et al. 1997; Heron et al. 2009). In terms of some related parallel aspects of post-reproductive senescence, Reznick et al. (2004) realized through laboratory experiments that guppies (*Poecilia reticulata*), if originally coming from a high-predation water stream, showed a faster decay in predator escape responses after the reproduction period ends, making them slower, as it should be if the AP theory is fulfilled. (Williams' postulate number 2). In lab mice, it has also been demonstrated that at the time of follicular depletion (resembling women's menopause) there is a higher aging-related physiological decline and an increasing incidence of disease (Finch 1990; Finch 2007).

Theoretical and fossil evidences show that reaching the age of menopause may not have been common amongst ancient hominids. *Australopithecus sp.* rarely lived beyond 46.6 years (Carey 2003), and fossil register for *Homo erectus* shows that they seldom got over 50 (McHenry 1994), although theoretical analysis by Carey (2003) based on regressions of longevity over

body mass index in several primates, suggests that they might have reached 62 years. Further work developed by (R. Caspari and Lee 2004) demonstrates that reaching old ages took place rather late in human evolution. Therefore, unless menopause was coming much earlier in ancient than in present times, the probabilities of reaching the menopause age were low, making it only possible if individuals lived long enough (Sievert 2006) justifying the existence of a biological barrier.

Our results fit within the general trend in humans, but that does not rule out further plausible senescence patterns, as discussed all along this thesis. Unfortunately, in our case we cannot show how human senescence patterns in terms of the AP or the MA theories will progress further than 60 years old because of the scarcity of data and the consequent drop of statistical power.

4.3.3 Some (possible) future steps

It is important to state clearly that we are not showing that the AP or the MA theories are the only evolutionary explanations for senescence in humans or in the living world at large. Other theories could also be important, but our data precludes testing them.

In order to further test MA and AP theories, a straightforward and logical next step would be to repeat the same exercise, but this time using all *p*-values (not only top 10^{-5} hits from the GWAS Catalog) for a set of diseases available under request in full-fledged databases such as the dbGAP (database of Genotypes and Phenotypes, NCBI based) or the EGA (European Genome and phenome Archive, EBI-CRG based). Even

if *p*-values above 10^{-5} do present high levels of statistical noise, the fact that we are exploring second order relationships linking several diseases immediately suggests ways to mine these data.

In our manuscript (section 3.3) we used disease trajectories derived from medical records (Jensen et al. 2014) to see up to which extent pleiotropies overlapped with true disease comorbidities. A certain overlap was seen, reflecting the potential of the approach. However, we could not assess the full spectrum of pleiotropies because the disease trajectories were only gathered within an interval of 14 years, and many of the diseases in pleiotropies have an onset farther apart than these years. To circumvent this, we could potentially use other health/genealogical records, such as those from deCODE genetics Inc., an Icelandic company whose objective was to create the world's first population-wide genomic biobank by collecting data from the whole population of Iceland (~300,000). A particularity of this enterprise is that biological samples, together with the highly accurate medical history and genealogy since the last 1,000 years are put together. Gaining access to this data could allow further exploration, not only of the AP and MA theories but also of other hypothesis around senescence.

Even if we did not find an excess of signatures of natural selection in our set of pleiotropies, some striking cases of antagonistic pleiotropies under selection have been identified. Experimental analysis of these cases could help deciphering the mechanisms behind pleiotropy and consequently, behind disease. Particularly, given our own evidences and those from the literature, *CDKN2A* offers an excellent opportunity to create a mouse model containing the human variants we identified (Table 2) and study their phenotypic effects. Using recombinase-mediated genomic

replacement (Wallace et al. 2007), we could insert human haplotypes, both protective and risk, in mice embryos and see how mortality, disease, ageing patterns and fertility vary between those animals carrying the early and the late onset disease risk haplotypes. The success in this project would represent not only definitive proof of the AP theory, but a major leap forward in evolutionary systems biology.

4.4 CONCLUDING REMARKS

In summary, the work derived from this PhD has demonstrated, at least, five things:

1. That the public GWAS data required to achieve our main objective and accumulated during the last decade are sound and replicable.

2. Genome-wide data applied to prediction of disease risk works increasingly well, even if some important concerns when translating these hits into clinic are raised.

3. Our approach to identify pleiotropies has the potential to reveal comorbidities between apparently unrelated diseases, with special interest in those occurring between different life periods.

4. The mutation accumulation and the antagonistic pleiotropy phenomena influence human ageing and disease.

5. We identified some remarkable candidate loci namely *CDKN2A, RREB1* and *IL13* that participate in antagonistic pleiotropies between early and late onset diseases, which also carry strong molecular signals of positive selection.

5. BIBLIOGRAPHY

Abrams, P A. 1993. "Does Increased Mortality Favor the Evolution of More Rapid Senescence?" *Source: Evolution* 47 (3): 877–87. doi:10.2307/2410191.

Adams, P C, and Barton J C. 2007. "Haemochromatosis." *Lancet* 370: 1855–60. doi:10.1016/S0140-6736(07)61782-6.

Adeyemo, A, and Rotimi C. 2009. "Genetic Variants Associated with Complex Human Diseases Show Wide Variation across Multiple Populations." *Public Health Genomics* 13 (2): 72–79. doi:10.1159/000218711.

Albin, R L. 1993. "Antagonistic Pleiotropy, Mutation Accumulation, and Human Genetic Disease." *Genetica* 91 (1–3): 279–86.

Ardlie, K G, Kruglyak L, and Seielstad M. 2002. "Patterns of Linkage Disequilibrium in the Human Genome." *Nature Reviews Genetics* 3 (4): 299–309. doi:10.1038/nrg777.

Arnold, M, Hartsperger, M L, Baurecht H, Rodríguez, E, Wachinger, B, Franke, A, Kabesch, M, et al. 2012. "Network-Based SNP Meta-Analysis Identifies Joint and Disjoint Genetic Features across Common Human Diseases." *BMC Genomics* 13 (1): 490. doi:10.1186/1471-2164-13-490.

Ascádi, G, and Nemeskéri J. 1970. *History of Human Lifespan and Mortality*. Akadémiai Kiadó.

Baudisch, A. 2005. "Hamilton's Indicators of the Force of Selection." *Proceedings of the National Academy of Sciences* 102 (23).

———. 2008. *Inevitable Aging? Contributions to Evolutionary-Demographic Theory*. 1sted. Springer-Verlag Berlin Heidelberg.

Blackburn, E H., and Gall J G. 1978. "A Tandemly Repeated Sequence at

the Termini of the Extrachromosomal Ribosomal RNA Genes in Tetrahymena." *Journal of Molecular Biology* 120 (1): 33–53. doi:10.1016/0022-2836(78)90294-2.

Bodmer, W, and Bonilla C. 2008. "Common and Rare Variants in Multifactorial Susceptibility to Common Diseases." *Nature Genetics* 40 (6): 695–701. doi:10.1038/ng.f.136.

Bodnar, A G, Ouellette, M, Frolkis, M, Holt, S E, Chiu, C P, Morin, G B, Harley, C B, Shay, J W, Lichtsteiner, S, and Wright W E. 1998. "Extension of Life-Span by Introduction of Telomerase into Normal Human Cells." *Science (New York, N.Y.)* 279 (5349): 349–52. doi:10.1126/science.279.5349.349.

Botstein, D, and Risch N. 2003. "Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease." *Nature Genetics* 33 Suppl (march): 228–37. doi:10.1038/ng1090.

Botstein, D, White, R L, Skolnick, M, and Davis R W. 1980. "Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms." *American Journal of Human Genetics* 32 (3): 314–31. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1686077&tool=pmcentrez&rendertype=abstract.

Brien, P. 1953. "La Pérennité Somatique." *Biological Reviews* 28 (3): 308–49. doi:10.1111/j.1469-185X.1953.tb01381.x.

Broer, L, Buchman, A S, Deelen, J, Evans, D S, Faul, J D, Lunetta, KL Sebastiani, P. et al. 2015. "GWAS of Longevity in CHARGE Consortium Confirms APOE and FOXO3 Candidacy." *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 70 (1): 110–18. doi:10.1093/gerona/glu166.

Bryant, M J, and Reznick D. 2004. "Comparative Studies of Senescence in

Natural Populations of Guppies." *The American Naturalist* 163 (1): 55–68. doi:10.1086/380650.

Budovsky, A, Craig T, Wang J, Tacutu R, Csordas, A, Lourenço, J, Fraifeld, VE, and de Magalhães, J P. 2013. "LongevityMap: A Database of Human Genetic Variants Associated with Longevity." *Trends in Genetics*, August, 6–7. doi:10.1016/j.tig.2013.08.003.

Burga, A, and Lehner B. 2012. "Beyond Genotype to Phenotype: Why the Phenotype of an Individual Cannot Always Be Predicted from Their Genome Sequence and the Environment That They Experience." *The FEBS Journal* 279 (20): 3765–75. doi:10.1111/j.1742-4658.2012.08810.x.

Butlin, R K, and Day T H. 1985. "Adult Size, Longevity and Fecundity in the Seaweed Fly, Coelopa Frigida." *Heredity* 54 (July 1984): 107–10. doi:10.1038/hdy.1985.14.

"Cancer Research UK." 2016. Accessed November 1. http://www.cancerresearchuk.org/.

Carey, J R. 2003. *Longevity: The Biology and Demography of Life Span*. Princeton (New Jersey): Princeton University Press.

Carey, J R, Liedo P, Orozco D, and Vaupel, J W. 1992. "Slowing of Mortality Rates at Older Ages in Large Medfly Cohorts." *Science (New York, N.Y.)* 258 (5081): 457–61. doi:10.1126/science.1411540.

Carrel, A. 1921. "Age and Multiplication of Fibroblasts." *Journal of Experimental Medicine* 34 (6): 599–623. doi:10.1084/jem.34.6.599.

Carter, A J R and Nguyen A Q. 2011. "Antagonistic Pleiotropy as a Widespread Mechanism for the Maintenance of Polymorphic Disease Alleles." *BMC Medical Genetics* 12 (1): 160. doi:10.1186/1471-2350-12-160.

Caspari, E. 1952. "Pleiotropic Gene Action." *Evolution* 6 (1): 1–18.

Caspari, R., and Lee S H. 2004. "Older Age Becomes Common Late in

Human Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 101 (30): 10895–900. doi:10.1073/pnas.0402857101.

Chakravarti, A. 1999. "Population Genetics — Making Sense out of Sequence." *Nature Genetics Supplement* 21 (january): 56–60. doi:10.1038/4482.

Chapman, J A, Kirkness, E F, Simakov, O, Hampson, S E, Mitros, T, Weinmaier, T, Rattei, T et al. 2010. "The Dynamic Genome of Hydra." *Nature* 464 (7288): 592–96. doi:10.1038/nature08830.

Charlesworth, B. 1994. *Evolution in Age-Structured Populations*. 2nded. Cambridge University Press.

Charmantier, A, Perrins, C, McCleery, R H, and Sheldon, B C. 2006. "Quantitative Genetics of Age at Reproduction in Wild Swans: Support for Antagonistic Pleiotropy Models of Senescence." *Proceedings of the National Academy of Sciences of the United States of America* 103 (17): 6587–92. doi:10.1073/pnas.0511123103.

Christensen, K, Johnson T E, and Vaupel J W. 2006. "The Quest for Genetic Determinants of Human Longevity: Challenges and Insights." *Nature Reviews. Genetics* 7 (6): 436–48. doi:10.1038/nrg1871.

Clarke, G M, Anderson, C A, Pettersson, F H, Cardon, L R, Morris, A P, and Zondervan K T. 2011. "Basic Statistical Analysis in Genetic Case-Control Studies." *Nature Protocols* 6 (2): 121–33. doi:10.1038/nprot.2010.182.Basic.

Clemente, F J J, Cardona, A, Inchley, C E E, Benjamin M M, Jacobs, G, Pagani, L, Lawson, D J J. et al. 2014. "A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations." *American Journal of Human Genetics* 95 (5): 1–6. doi:10.1016/j.ajhg.2014.09.016.

Culotta, E, and Pennisi, E. 2005. "Evolution in Action." *Science* 310.

Curtsinger, J W, H H Fukui, D R Townsend, and J W Vaupel. 1992.

"Demography of Genotypes: Failure of the Limited Life-Span Paradigm in Drosophila Melanogaster." *Science* 258 (5081): 461–63. doi:10.1126/science.1411541.

Daly, M J, Rioux, J D, Schaffner, S F, Hudson, ,T J, and Lander E S. 2001. "High-Resolution Haplotype Structure in the Human Genome." *Nature Genetics* 29 (2): 229–32. doi:10.1038/ng1001-229.

Davignon, J, Gregg R E, and Sing C F. 1988. "Apolipoprotein E Polymorphism and Atherosclerosis." *Arteriosclerosis, Thrombosis, and Vascular Biology* 8 (1): 1–21. doi:10.1161/01.ATV.8.1.1.

de Magalhães, J P, Budovsky, A, Lehmann, G, Costa, J, Li, Y, Fraifeld, V, and Church G M. 2009. "The Human Ageing Genomic Resources: Online Databases and Tools for Biogerontologists." *Aging Cell*. doi:10.1111/j.1474-9726.2008.00442.x.

de Magalhães, J P, Curado, J, and Church G M. 2009. "Meta-Analysis of Age-Related Gene Expression Profiles Identifies Common Signatures of Aging." *Bioinformatics* 25 (7): 875–81. doi:10.1093/bioinformatics/btp073.

Deelen, J, Beekman, M, Won Uh, H, Broer, L, Ayers, K L, Tan, Q, Kamatani, Y. et al. 2014. "Genome-Wide Association Meta-Analysis of Human Longevity Identifies a Novel Locus Conferring Survival beyond 90 Years of Age." *Human Molecular Genetics* 23 (16): 4420–32. doi:10.1093/hmg/ddu139.

Deelen, J, Beekman, M, Won Uh, H, Helmer, Q, Kuningas, M, Christiansen, L, Kremer, D. et al. 2011. "Genome-Wide Association Study Identifies a Single Major Locus Contributing to Survival into Old Age; the APOE Locus Revisited." *Aging Cell* 10 (4): 686–98. doi:10.1111/j.1474-9726.2011.00705.x.

Denham, M J. 2011. "Dr Marjory Warren CBE MRCS LRCP (1897–1960): The Mother of British Geriatric Medicine." JOUR. *Journal of*

Medical Biography 19 (3): 105–10. doi:10.1258/jmb.2010.010030.

Dewan, A, Liu, M, Hartman, S, Shao-Min Zhang, S, Liu, D T L, Zhao, C, Pancy, O, Tam, S, et al. 2006. "HTRA1 Promoter Polymorphism in Wet Age-Related Macular Degeneration." *Science (New York, N.Y.)* 314 (5801): 989–92. doi:10.1126/science.1133807.

Distante, S, Robson, K J H, Graham-Campbell, J, Arnaiz-Villena, A, Brissot, P, and Worwood, M. 2004. "The Origin and Spread of the HFE-C282Y Haemochromatosis Mutation." *Human Genetics* 115 (4): 269–79. doi:10.1007/s00439-004-1152-4.

Economos, A C, and Lints F A. 1986. "Developmental Temperature and Life Span in Drosophila Melanogaster. I. Constant Developmental Temperature: Evidence for Physiological Adaptation in a Wide Temperature Range." *Gerontology* 32 (1): 18–27. http://www.ncbi.nlm.nih.gov/pubmed/3081405.

Eichler, EE, Flint, J, Gibson, G, Kong, A, Leal, S M, Moore, J H, and Nadeau, J H. 2010. "Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease." *Nature Reviews. Genetics* 11 (6). Nature Publishing Group: 446–50. doi:10.1038/nrg2809.

Eisenberg, D T , Hayes, M G, and Kuzawa, C W. 2012. "Delayed Paternal Age of Reproduction in Humans Is Associated with Longer Telomeres across Two Generations of Descendants." *Proc.Natl.Acad.Sci.U.S.A* 109 (1091–6490 (Electronic)): 10251–56. doi:10.1073/pnas.1202092109.

Empedocles, and Leonard, W E. 1908. *The Fragments of Empedocles*. Chicago: The Open Court Publishing Company.

Erikson, G A, Bodian, D L, Rueda, M, Molparia, B, Scott, E R, Scott-Van Zeeland, A A, Topol, S E, et al. 2016. "Whole-Genome Sequencing of a Healthy Aging Cohort." *Cell* 165 (4). Elsevier Inc.: 1002–11. doi:10.1016/j.cell.2016.03.022.

Eskenazi, B R., Wilson-Rich, N S, and Starks, P T. 2007. "A Darwinian Approach to Huntington's Disease: Subtle Health Benefits of a Neurological Disorder." *Medical Hypotheses* 69 (6): 1183–89. doi:10.1016/j.mehy.2007.02.046.

Ewald, P W. 1980. "Evolutionary Biology and the Treatment of Signs and Symptoms of Infectious Disease." *Journal of Theoretical Biology* 86 (1): 169–76. http://www.ncbi.nlm.nih.gov/pubmed/7464170.

Feldman, R P, and Goodrich J T. 1999. "The Edwin Smith Surgical Papyrus." *Child's Nervous System* 15 (6–7): 281–84. doi:10.1007/s003810050395.

Fernandes, M, Wan, C, Tacutu, R, Barardo, D, Rajput, A, Wang, J, Thoppil, H. et al. 2016. "Systematic Analysis of the Gerontome Reveals Links between Aging and Age-Related Diseases." JOUR. *Human Molecular Genetics* , September. doi:10.1093/hmg/ddw307.

Filer, J. 1995. *Disease*. Texas A&M University Press.

Finch, C E. 1990. *Longevity, Senescence, and the Genome*. Edited by University of Chicago Press. 1st ed. Chicago: Academic Press.

———. 2007. *The Biology of Human Longevity*. *The Biology of Human Longevity*. http://www.scopus.com/inward/record.url?eid=2-s2.0-84882159379&partnerID=40&md5=e0ff31f6cc25e6cbcbe2ded42bba1476.

———. 2009. "Update on Slow Aging and Negligible Senescence - A Mini-Review." *Gerontology* 55 (3): 307–13. doi:10.1159/000215589.

———. 2010. "Evolving Views of Ageing and Longevity from Homer to Hippocrates: Emergence of Natural Factors, Persistence of the Supernatural." *Greece and Rome* 57 (2): 355–57.

Finkel, T, and Holbrook, N J. 2000. "Oxidants, Oxidative Stress and the Biology of Ageing." *Nature* 408 (6809): 239–47. doi:10.1038/35041687.

Fisher, R A. 1930. "The Genetical Theory of Natural Selection." *Genetics* 154: 272. doi:10.1038/158453a0.

Floyer, J. 1724. *Medicina Gerocomica; or the Galenic Art of Preserving Old Men's Health*. London.

Fortney, K, Dobriban, E, Garagnani, P, Pirazzini, C, Monti, D, Mari, D Atzmon, G. et al. 2015. "Genome-Wide Scan Informed by Age-Related Disease Identifies Loci for Exceptional Human Longevity." *PLoS Genetics* 11 (12). doi:10.1371/journal.pgen.1005728.

Friedman, D B, and Johnson, T E. 1988. "A Mutation in the Age-1 Gene in Caenorhabditis Elegans Lengthens Life and Reduces Hermaphrodite Fertility." *Genetics* 118 (1): 75–86. doi:10.1016/S0960-9822(00)00522-4.

Fu, J, Festen, E A M, and Wijmenga, C. 2011. "Multi-Ethnic Studies in Complex Traits." *Human Molecular Genetics* 20 (R2): 206–13. doi:10.1093/hmg/ddr386.

Gabriel, S B, Schaffner, S F, Nguyen, H, Moore J M, Roy, J, Blumenstiel, B, Higgins, J. et al. 2002. "The Structure of Haplotype Blocks in the Human Genome." *Science (New York, N.Y.)* 296 (5576): 2225–29. doi:10.1126/science.1069424.

Gavrilov, LA, and Gavrilova, NS. 1999. "Is There a Reproductive Cost for Human Longevity?" *Journal of Anti-Aging Medicine*. doi:10.1089/rej.1.1999.2.121.

Gavrilov, LA, and Gavrilova, NS. 2002. "Evolutionary Theories of Aging and Longevity." *TheScientificWorldJournal* 2: 339–56. doi:10.1100/tsw.2002.96.

George, A, and Mosley, F. 2010. "The Epic of Gilgamesh." *Online* 1 (0): 1–27. doi:10.1017/CBO9781107415324.004.

Gilleard, Chris. 2015. "Ageing and the Galenic Tradition: A Brief Overview." *Ageing and Society* 35 (35): 489–511.

doi:10.1017/S0144686X13000834.

Gonzales, B M. 1921. "Experimental Studies on the Duration of Life. VIII. The Influence Upon Duration of Life of Certain Mutant Genes of Drosophila Melanogaster." *American Naturalist* 57 (651): 289–325.

Goodwin, S, McPherson, J D and McCombie, W R. 2016. "Coming of Age: Ten Years of next-Generation Sequencing Technologies." JOUR. *Nat Rev Genet* 17 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 333–51. http://dx.doi.org/10.1038/nrg.2016.49.

Graham Bell, A. 1918. *The Duration of Life and Conditions Associated with Longevity*. 1st ed. Washington, D. C.: Genealogical record office.

Greider, C W, and Blackburn, E H. 1985. "Identification of a Specific Telomere Terminal Transferase Activity in Tetrahymena Extracts." *Cell* 43 (2 PART 1): 405–13. doi:10.1016/0092-8674(85)90170-9.

Haldane, J B S. 1941. *New Paths in Genetics*. George Allen & Unwin.

Hamilton, W D. 1966. "The Moulding of Senescence by Natural Selection." *Journal of Theoretical Biology* 12 (1): 12–45. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2982254&tool=pmcentrez&rendertype=abstract.

Harman, D. 1956. "Aging: A Theory Based on Free Radical and Radiation Chemistry." *Journal of Gerontology* 11 (3): 298–300. doi:10.1093/geronj/11.3.298.

———. 1965. "The Free Radical Theory of Aging: Effect of Age on Serum Copper Levels." JOUR. *Journal of Gerontology* 20 (2): 151–53. doi:10.1093/geronj/20.2.151.

Harvey, W. 1628. "Exercitatio Anatomica de Motu Cordis et Sanguinis in Animalibus."

Hästbacka, J, de la Chapelle, A, Mahtani, M M, Clines, G, Reeve-Daly, M P, Daly, M, Hamilton, B A et al. 1994. "The Diastrophic Dysplasia

Gene Encodes a Novel Sulfate Transporter: Positional Cloning by Fine-Structure Linkage Disequilibrium Mapping." *Cell* 78 (6): 1073–87. doi:10.1016/0092-8674(94)90281-X.

Hayflick, L. 1965. "The Limited in Vitro Lifetime of Human Diploid Cell Strains." *Experimental Cell Research* 37 (3): 614–36. doi:10.1016/0014-4827(65)90211-9.

Healy, K, Sive Finlay, G, Kane, A, Kelly, S B A, McClean, Kelly, D J, Donohue, I, Jackson, A L, and Cooper, N. 2014. "Ecology and Mode-of-Life Explain Lifespan Variation in Birds and Mammals." JOUR. *Proceedings of the Royal Society B: Biological Sciences* 281 (1784). http://rspb.royalsocietypublishing.org/content/281/1784/20140298.abstract.

Heilbronn, L K, and Ravussin, E. 2003. "Calorie Restriction and Aging: Review of the Literature and Implications for Studies in Humans." *The American Journal of Clinical Nutrition* 78 (3): 361–69. http://ajcn.nutrition.org/content/78/3/361.abstract.

Helfand, S L, and Rogina, B. 2003. "Genetics of Aging in the Fruit Fly, Drosophila Melanogaster." *Annual Review of Genetics* 37: 329–48. doi:10.1146/annurev.genet.37.040103.095211.

Hendry, A P, Morbey, Y E, Berg, O K, and Wenburg, J K. 2004. "Adaptive Variation in Senescence: Reproductive Lifespan in a Wild Salmon Population." *Proceedings. Biological Sciences / The Royal Society* 271 (1536): 259–66. doi:10.1098/rspb.2003.2600.

Heron, M, Hoyert, D L, Murphy, S L, Xu, J, Kochanek, K D, and Tejada-Vera, B. 2009. "Deaths: Final Data for 2006." *National Vital Statistics Reports : From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*. Vol. 57. doi:May 8, 2013.

Herskind, A M, McGue, M, Holm, N V, Sorensen, T I A, Harvald, B, and

Vaupel, J W. 1996. "The Heritability of Human Longevity: A Population-Based Study of 2872 Danish Twin Pairs Born 1870-1900." *Human Genetics* 97 (3): 319–23. doi:10.1007/s004390050042.

Hill, W G. 1974. "Estimation of Linkage Disequilibrium in Randomly Mating Populations." *Heredity* 33 (2): 229–39. doi:10.1038/hdy.1979.10.

Hirschhorn, J N, and Altshuler, D. 2002. "Once and Again—Issues Surrounding Replication in Genetic Association Studies." JOUR. *The Journal of Clinical Endocrinology & Metabolism* 87 (10). The Endocrine Society: 4438–41. doi:10.1210/jc.2002-021329.

Hirschhorn, J N, Lohmueller, K, Byrne, E, and Hirschhorn, K. 2002. "A Comprehensive Review of Genetic Association Studies." *Genetics in Medicine : Official Journal of the American College of Medical Genetics* 4 (2): 45–61. doi:10.109700125817-200203000-00002.

Hjelmborg, J B, Iachine, I, Skytthe, A, Vaupel, J W, McGue, M, Koskenvuo, M, Kaprio, J, Pedersen, N L, and Christensen K. 2006. "Genetic Influence on Human Lifespan and Longevity." *Human Genetics* 119 (3): 312–21. doi:10.1007/s00439-006-0144-y.

Ioannidis, J P A, Ntzani, E E, Trikalinos, T A, and Contopoulos-ioannidis, D G. 2001. "Replication Validity of Genetic Association Studies John." *Nature Genetics* 29 (October): 306–9. doi:10.1038/ng749.

Jakobsdottir, J, Gorin, M B, Conley, Y P, Ferrell, R E, and Weeks D E. 2009. "Interpretation of Genetic Association Studies: Markers with Replicated Highly Significant Odds Ratios May Be Poor Classifiers." Article. *PLoS Genet* 5 (2): e1000337. doi:10.1371/journal.pgen.1000337.

Jelier, R, Semple, J I, Garcia-Verdugo, R, and Lehner, B. 2011. "Predicting Phenotypic Variation in Yeast from Individual Genome Sequences."

Nature Genetics 43 (12): 1270–74. doi:10.1038/ng.1007.

Jensen, A B, Moseley, P L, Oprea, T I, Ellesøe, S E, Eriksson, R, Schmock, H, Jensen, P B, Jensen, L J, and Brunak, S. 2014. "Temporal Disease Trajectories Condensed from Population-Wide Registry Data Covering 6.2 Million Patients." *Nature Communications* 5 (May): 4022. doi:10.1038/ncomms5022.

Jeune, B. 2002. "Living Longer - But Better?" *Aging Clinical and Experimental Research* 14 (2): 72–93. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0036019370&partnerID=40&md5=4e4573ebb62906c9253c310ec1aee379.

Jewson, N. D. 2009. "The Disappearance of the Sick-Man from Medical Cosmology, 1770-1870." *International Journal of Epidemiology* 38 (3): 622–33. doi:10.1093/ije/dyp180.

Jones, O R, Scheuerlein, A, Salguero-Gómez, R, Camarda, C G, Schaible, R, Casper, B B, Dahlgren, J P, et al. 2014. "Diversity of Ageing across the Tree of Life." *Nature*, December, 5–10. doi:10.1038/nature12789.

Keane, M, Semeiks, J, Webb, A E, Li, Y I, Quesada, V, Craig, T, Madsen, L B et al. 2015. "Insights into the Evolution of Longevity from the Bowhead Whale Genome." *Cell Reports* 10 (1): 112–22. doi:10.1016/j.celrep.2014.12.008.

Kenyon, C, Chang, J, Gensch, E, Rudner, A, and Tabtiang, R. 1993. "A C. Elegans Mutant That Lives Twice as Long as Wild Type." *Nature* 366 (6454): 461–64. doi:10.1038/366461a0.

Kilpeläinen, T O, Qi, L, Brage, S, Sharp, S J, Sonestedt, E, Demerath, E, Ahmad, T, et al. 2011. "Physical Activity Attenuates the Influence of FTO Variants on Obesity Risk: A Meta-Analysis of 218,166 Adults and 19,268 Children." *PLoS Medicine* 8 (11).

doi:10.1371/journal.pmed.1001116.

Kirkwood, T B. 1977. "Evolution of Ageing." *Nature* 170: 201–4.

———. 2001. *Time of Our Lives: The Science of Human Aging*. 1sted. Oxford University Press.

Kirkwood, T B, and Holliday, R. 1979. "The Evolution of Ageing and Longevity." *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)* 205 (1161): 531–46. doi:10.1098/rspb.1979.0083.

Klass, M, Nguyen, P N, and Dechavigny, A. 1983. "Age-Correlated Changes in the DNA Template in the Nematode Caenorhabditis Elegans." *Mechanisms of Ageing and Development* 22 (3–4): 253–63. doi:10.1016/0047-6374(83)90080-5.

Klein, R J, Zeiss, C, Chew, E Y, Tsai, J Y, Sackler, R S, Haynes, C, Henning A K, et al. 2005. "Complement Factor H Polymorphism in Age-Related Macular Degeneration." *Science (New York, N.Y.)* 308 (5720): 385–89. doi:10.1126/science.1109557.

Kraft, P, Wacholder, S, Cornelis, M C, Hu F B, Hayes, R B, Thomas, G, Hoover, R, Hunter, D J, and Chanock, S. 2009. "Beyond Odds Ratios — Communicating Disease Risk Based on Genetic Profiles." *Nature Reviews Genetics* 10 (4): 264–69. doi:10.1038/nrg2516.

Kruglyak, L. 1999. "Prospects for Whole-Genome Linkage Disequilibrium Mapping of Common Disease Genes." *Nature Genetics* 22 (2): 139–44. doi:10.1038/9642.

Lander E S. 1996. "The New Genomics: Global Views of Biology." *Science (New York, N.Y.)* 274 (5287): 536–39. doi:10.1126/science.274.5287.536.

Lander E S, Linton, L M, Birren, B, Nusbaum, C, Zody, M C, Baldwin, J, Devon, K, et al. 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* 409 (6822): 860–921.

doi:10.1038/35057062.

Lappalainen, I, Almeida-King, J, Kumanduri, V, Senf, A, Dylan Spalding, A, ur-Rehman, S, Saunders, G, et al. 2015. "The European Genome-Phenome Archive of Human Data Consented for Biomedical Research." *Nature Genetics* 47 (7). Nature Publishing Group: 692–95. doi:10.1038/ng.3312.

Law, R. 1979. "The Cost of Reproduction in Annual Meadow Grass." *The American Naturalist* 113 (1): 3–16. doi:10.2307/2459939.

Le Bourg, E, Thon, B, Légaré, J, Desjardins, B, and Charbonneau, H. 1993. "Reproductive Life of French-Canadians in the 17-18th Centuries: A Search for a Trade-off between Early Fecundity and Longevity." *Experimental Gerontology* 28 (3): 217–32. doi:10.1016/0531-5565(93)90030-H.

Lee, C K, Allison, D B, Brand, J, Weindruch, R, and Prolla, T A. 2002. "Transcriptional Profiles Associated with Aging and Middle Age-Onset Caloric Restriction in Mouse Hearts." *Proc Natl Acad Sci U S A* 99 (23): 14988–93. doi:10.1073/pnas.232308999.

Lee, C K, Weindruch, R, and Prolla, T A. 2000. "Gene-Expression Profile of the Ageing Brain in Mice." *Nature Genetics* 25 (3): 294–97. doi:10.1038/77046.

Lee, S H, Van Der Werf, J H J, Hayes, B, Goddard, M E, and Visscher P M. 2008. "Predicting Unobserved Phenotypes for Complex Traits from Whole-Genome SNP Data." *PLoS Genetics* 4 (10). doi:10.1371/journal.pgen.1000231.

Leroi, A M. 2005. *Mutants: On Genetic Variety and the Human Body*. Penguin Books.

Lewontin, R C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press.

Lewontin, R C, and Kojima K. 1960. "The Evolutionary Dynamics of

Complex Polymorphisms." *Evolution*. doi:10.2307/2405995.

Loeb, J, and Northrop J H. 1916. "Is There a Temperature Coefficient for the Duration of Life?" *Proc Natl Acad Sci U S A* 2 (8): 456–57.

Lohmueller, K E, Pearce, C L, Pike, M, Lander E S, and Hirschhorn J N. 2003. "Meta-Analysis of Genetic Association Studies Supports a Contribution of Common Variants to Susceptibility to Common Disease." *Nature Genetics* 33 (2): 177–82. doi:10.1038/ng1071.

Longrigg, J. 1993. *Greek Rational Medicine: Philosophy and Medicine from Alcmaeon to the Alexandrians*. Routledge.

López-Otín, C, Blasco, M A, Partridge, L, Serrano, M, and Kroemer, G. 2013. "The Hallmarks of Aging." *Cell* 153 (6): 1194–1217. doi:10.1016/j.cell.2013.05.039.

MacDonald, M E, Ambrose, C M, Duyao, M P, Myers, R H, Lin, C, Srinidhi, L, Barnes, G, et al. 1993. "A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes." *Cell* 72 (6): 971–83. doi:10.1016/0092-8674(93)90585-E.

Mailman, M D, Feolo, M, Jin, Y, Kimura, M, Tryka, K, Bagoutdinov, R, Hao, L, et al. 2007. "The NCBI dbGaP Database of Genotypes and Phenotypes." *Nature Genetics* 39 (10): 1181–86. doi:10.1038/ng1007-1181.

Marigorta, U M, Lao, O, Casals, F Calafell, F, Morcillo-Suárez, C, Faria, R, Bosch, E, et al. 2011. "Recent Human Evolution Has Shaped Geographical Differences in Susceptibility to Disease." *BMC Genomics* 12 (1): 55. doi:10.1186/1471-2164-12-55.

Marigorta, U M, and Navarro, A. 2013. "High Trans-Ethnic Replicability of GWAS Results Implies Common Causal Variants." *PLoS Genetics* 9 (6): e1003566. doi:10.1371/journal.pgen.1003566.

Masoro, E J. 1985. "Nutrition and Aging—A Current Assessment."

JOUR. *The Journal of Nutrition* 115 (7): 842–48. http://jn.nutrition.org/content/115/7/842.short.

Matthews, D A. 1984. "Dr. Marjory Warren and the Origin of British Geriatrics." *Journal of the American Geriatrics Society* 32 (4): 253–58. http://www.ncbi.nlm.nih.gov/pubmed/6368651.

McCarthy, M I, Abecasis, G A, Cardon, L R, Goldstein, D B, Little, J, Ioannidis, J P A and Hirschhorn J N. 2008. "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges." *Nature Reviews. Genetics* 9 (5): 356–69. doi:10.1038/nrg2344.

McCay, C M, Crowell, M, and Maynard, L A. 1935. "The Effect of Retarded Growth upon the Length of Life Span and upon the Ultimate Body Size." *The Journal of Nutrition* 5 (3): 63–79. DOI: doi:2520283.

McClintock, B. 1941. "The Stability of Broken Ends of Chromosomes in Zea Mays." *Genetics* 26 (2): 234–82. doi:10.1038/378739a0.

McHenry, H M. 1994. "Tempo and Mode in Human Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 91 (15): 6780–86. doi:10.1073/pnas.91.15.6780.

Medawar, P B. 1946. "Old Age and Natural Death." *Modern Quart.* 2: 30–49.

———. 1951. "Un Unsolved Problem of Biology." In *An Inaugural Lecture Delivered at University College, London*. University College London.

Minster, R L, Hawley, N L, Su, C-T, Sun, G, Kershaw, E E, Cheng, H, Buhule, O D, et al. 2016. "A Thrifty Variant in CREBRF Strongly Influences Body Mass Index in Samoans." JOUR. *Nat Genet* 48 (9). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 1049–54. http://dx.doi.org/10.1038/ng.3620.

Moltke, I, Grarup, N, Jørgensen, M E, Bjerregaard, P, Treebak, J T, Fumagalli, M, Korneliussen, T S, et al. 2014. "A Common Greenlandic TBC1D4 Variant Confers Muscle Insulin Resistance and Type 2 Diabetes." *Nature* 512 (7513): 190–93. doi:10.1038/nature13425.

Moyzis, R K, Buckingham, J M, Cram, L S, Dani, M, Deaven, L L, Jones, M D, Meyne, J, Ratcliff, R L, and Wu J R. 1988. "A Highly Conserved Repetitive DNA Sequence, (TTAGGG)n, Present at the Telomeres\rof Human Chromosomes." *Proc. Natl. Acad. Sci. U S A.*

Mueller, L D, and Rose, M R. 1996. "Evolutionary Theory Predicts Late-Life Mortality Plateaus." *Proceedings of the National Academy of Sciences of the United States of America* 93 (26): 15249–53. doi:10.1073/pnas.93.26.15249.

Muller, H J. 1962. "Studies of Genetics: The Selected Papers of H. J. Muller." In *Studies of Genetics: The Selected Papers of H. J. Muller*, 384–408. Bloomington: Indiana Univ. Press.

Murphy, C T. 2006. "The Search for DAF-16/FOXO Transcriptional Targets: Approaches and Discoveries." *Experimental Gerontology* 41 (10): 910–21. doi:10.1016/j.exger.2006.06.040.

Neel, J V. 1962. "Diabetes Mellitus: A 'thrifty' genotype Rendered Detrimental By 'progress'?" *American Journal of Human Genetics* 77 (8): 694-703-3. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2557712&tool=pmcentrez&rendertype=abstract.

Nesse, R M. 2001. "On the Difficulty of Defining Disease: A Darwinian Perspective." *Medicine, Health Care, and Philosophy* 4 (1): 37–46. http://www.ncbi.nlm.nih.gov/pubmed/11315418.

———. 2008. "Evolution: Medicine's Most Basic Science." *The Lancet* 372 (December 2008): S21–27. doi:10.1016/S0140-6736(08)61877-2.

Nesse, R M, and Stearns, S C. 2008. "The Great Opportunity: Evolutionary Applications to Medicine and Public Health." *Evolutionary Applications* 1 (1): 28–48. doi:10.1111/j.1752-4571.2007.00006.x.

Nesse, R M, and Williams, G C. 1994. *Why We Get Sick: The New Science of Darwinian Medicine*. Vintage.

———. 1998. "Evolution and the Origins of Disease." *Scientific American* 279 (5): 86–93. http://www.ncbi.nlm.nih.gov/pubmed/9796548.

Nielsen, R, Williamson, S , Kim, Y, Hubisz, M J, Clark, A G, and Bustamante, C. 2005. "Genomic Scans for Selective Sweeps Using SNP Data." *Genome Research* 15 (11): 1566–75. doi:10.1101/gr.4252305.

Njajou, O T, Cawthon, R M, Damcott, C M, Wu, S-H, Ott, S, Garant, M J, Blackburn, E H, Mitchell, E D, Shuldiner, A R, and Hsueh, W-C. 2007. "Telomere Length Is Paternally Inherited and Is Associated with Parental Lifespan." *Proceedings of the National Academy of Sciences of the United States of America* 104 (29): 12135–39. doi:10.1073/pnas.0702703104.

Northrop, J H. 1917. "On the Influence of Food and Temperature upon the Duration of Life." *Journal of Biological Chemistry* 32: 103–21.

———. 1925. "The Influence of the Intensity of Light on the Rate of Growth and Duration of Life of Drosophila." *The Journal of General Physiology* 9 (1): 81–86.

Nutton, V. 2005. "The Fatal Embrace: Galen and the History of Ancient Medicine." *Science in Context* 18 (1): 111–21. doi:10.1017/S0269889705000384.

———. 2007. *Ancient Medicine. Nature*. Vol. 77. doi:10.2307/1169967.

Parker, S L, Tong, T, Bolden, S, and Wingo P A. 1997. "Cancer Statistics, 1997." *CA: A Cancer Journal for Clinicians* 47 (1): 5–27.

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=Retrieve&list_uids=8996076&dopt=abstractplus\npapers://5fddb7df-536e-4c08-949a-4282cc2a64a2/Paper/p18604.

Pickrell, J K, Berisa, T, Liu, J Z, Segurel, L, Tung, J Y, andHinds, D A. 2016. "Detection and Interpretation of Shared Genetic Influences on 42 Human Traits." JOUR. *Nat Genet* 48 (7). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 709–17. http://dx.doi.org/10.1038/ng.3570.

Pilling, L C, Atkins, J L, Bowman, K, Jones, S E, Tyrrell, J, Beaumont, R N, Ruth, K S et al. 2016. "Human Longevity Is Influenced by Many Genetic Variants: Evidence from 75,000 UK Biobank Participants." *Aging* 8 (3): 1–24. doi:10.1101/038430.

Plate, L. 1910. "Genetics and Evolution." In *Festschrift Zum Sechzigsten Geburtstag Richard Hertwigs*. Jena (Germany): Fischer.

Pritchard, J K. 2001. "Are Rare Variants Responsible for Susceptibility to Complex Diseases?" *American Journal of Human Genetics* 69 (1): 124–37. doi:10.1086/321272.

Rauser, C L, Mueller, L D and Rose, M R. 2006. "The Evolution of Late Life." *Ageing Research Reviews* 5 (1): 14–32. doi:10.1016/j.arr.2005.06.003.

Reich, D E, and Lander E S. 2001. "On the Allelic Spectrum of Human Disease." *Trends in Genetics*. doi:10.1016/S0168-9525(01)02410-6.

Reznick, D N, Bryant, M J, Roff, D, Ghalambor, C K, and Ghalambor, D E. 2004. "Effect of Extrinsic Mortality on the Evolution of Senescence in Guppies." JOUR. *Nature* 431 (7012): 1095–99. http://dx.doi.org/10.1038/nature02936.

Risch, N, and Merikangas, K. 1996. "The Future of Genetic Studies of Complex Human Diseases." *Science (New York, N.Y.)* 273 (5281): 1516–17. doi:doi: 10.1126/science.273.5281.1516.

Rose, M R. 1991. *Evolutionary Biology of Aging. American Journal of Human Genetics*. Vol. 1990. New York: Oxford University Press. doi:10.1017/CBO9781107415324.004.

Rose, M R, and B. Charlesworth. 1980. "A Test of Evolutionary Theories of Senescence." *Nature* 287 (September): 141–42. doi:10.1038/287141a0.

———. 1981. "Genetics of Life History in Drosophila Melanogaster. I. Sib Analysis of Adult Females." *Genetics* 97 (1): 173–86.

Rose, M R, Rauser, C L, Benford, G, Matos, M, and Mueller, L D. 2007. "Hamilton's Forces of Natural Selection after Forty Years." *Evolution* 61 (6): 1265–76. doi:10.1111/j.1558-5646.2007.00120.x.

Rosenberg, N A, Huang L, Jewett, E M, Szpiech, Z A, Jankovic, I, and Boehnke, M. 2010. "Genome-Wide Association Studies in Diverse Populations." *Nature Reviews. Genetics* 11 (5). Nature Publishing Group: 356–66. doi:10.1038/nrg2760.

Roser, M. 2016. "Life Expectancy." *OurWorldInData.org*. https://ourworldindata.org/life-expectancy/.

Schächter, F, Faure-Delanef, L, Guénot, F, Rouger, H, Froguel, P, Lesueur-Ginot, L, and Cohen, D. 1994. "Genetic Associations with Human Longevity at the APOE and ACE Loci." *Nature Genetics* 6 (1): 29–32. doi:10.1038/ng0194-29.

Schaible, R, Scheuerlein, A, Danko, M J, Gampe, J, Martinez, D E and Vaupel, J W. 2015. "Constant Mortality and Fertility over Age in Hydra." *Proc Natl Acad Sci U S A* 112 (51): 15701–6. doi:10.1073/pnas.1521002112.

Schaible, R, Sussman, M, and Kramer B H. 2014. "Aging and Potential for Self-Renewal: Hydra Living in the Age of Aging - A Mini-Review." *Gerontology*. doi:10.1159/000360397.

Schlosstein, L, Terasaki, P, Bluestone, R, and Pearson, C M. 1973. "High

Association of an HL-A Antigen, W27, with Ankylosing Spondylitis." *The New England Journal of Medicine* 288 (14): 704–6. doi:10.1056/NEJM197304052881403.

Schoenmaker, M, de Craen, A J M, de Meijer, P H E M, Beekman, M, Blauw, G J, Slagboom, P E, and Westendorp, R G J. 2006. "Evidence of Genetic Enrichment for Exceptional Survival Using a Family Approach: The Leiden Longevity Study." *European Journal of Human Genetics : EJHG* 14 (1): 79–84. doi:10.1038/sj.ejhg.5201508.

Schwarz, F, Springer, S A, Altheide, T K, Varki, N M, Gagneux, P, and Varki, A. 2016. "Human-Specific Derived Alleles of CD33 and Other Genes Protect against Postreproductive Cognitive Decline." *Proceedings of the National Academy of Sciences*, no. April 2016: 201517951. doi:10.1073/pnas.1517951112.

Sebastiani, P, Riva, A, Montano, M, Pham, P, Torkamani, A, Scherba, E, Benson, G, et al. 2012. "Whole Genome Sequences of a Male and Female Supercentenarian, Ages Greater than 114 Years." *Frontiers in Genetics* 2 (January): 1–28. doi:10.3389/fgene.2011.00090.

Sebastiani, P, Solovieff, N, DeWan, A T, Walsh, K M, Puca, A, Hartley, S W, Melista, E, et al. 2012. "Genetic Signatures of Exceptional Longevity in Humans." *PLoS ONE* 7 (1). doi:10.1371/journal.pone.0029848.

Shoja, M M, Rashidi, M R, Tubbs, R S, Etemadi, J, Abbasnejad, F, and Agutter, P S. 2011. "Legacy of Avicenna and Evidence-Based Medicine." *International Journal of Cardiology*. doi:10.1016/j.ijcard.2010.10.019.

Shokeir, M H. 1975. "Investigations on Huntington's Disease in the Canadian Prairies. I. Prevalence." *Clinical Genetics* 7 (4): 345–48. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med1&NEWS=N&AN=123837.

Sievert, L S. 2006. *Menopause: A Biocultural Perspective (Studies in Medical Anthropology)*. Rutgers University Press.

Skytthe, A, Pedersen, N L, Kaprio, J, Stazi, M A, Hjelmborg, J V, Iachine, I, Vaupel, J W and Christensen K. 2003. "Longevity Studies in GenomEUtwin." *Twin Research : The Official Journal of the International Society for Twin Studies* 6 (5): 448–54. doi:10.1375/136905203770326457.

Slatkin, M. 2008. "Linkage Disequilibrium--Understanding the Evolutionary Past and Mapping the Medical Future." *Nature Reviews. Genetics* 9 (6): 477–85. doi:10.1038/nrg2361.

Sørensen, S A, Fenger, K, and Olsen, J H. 1999. "Significantly Lower Incidence of Cancer among Patients with Huntington Disease: An Apoptotic Effect of an Expanded Polyglutamine Tract?" *Cancer* 86 (7): 1342–46. doi:http://dx.doi.org/10.1002/%28SICI%291097-0142%2819991001%2986:7%3C1342::AID-CNCR33%3E3.0.CO;2-3.

Stearns, S C, Ackermann, M, Doebli, M, and Kaiser, M. 2000. "Experimental Evolution of Aging, Growth, and Reproduction in Fruit Flies." *Proceedings of the National Academy of Sciences of the USA* 97: 3309–13. http://www.pnas.org/cgi/reprint/97/7/3309.

Stein, G, and Weiss, J. 1948. "Chemical Effects of Ionizing Radiations." *Nature* 161: 650.

Stessman, J, Maaravi, Y, Hammerman-Rozenberg, R, Cohen, A, Nemanov, L, Gritsenko, I, Gruberman, N, and Ebstein, R P. 2005. "Candidate Genes Associated with Ageing and Life Expectancy in the Jerusalem Longitudinal Study." In *Mechanisms of Ageing and Development*, 126:333–39. doi:10.1016/j.mad.2004.08.025.

Tan, Q, Hua Zhao, J, Zhang, D, Kruse, T A, and Christensen, K. 2008. "Power for Genetic Association Study of Human Longevity Using

the Case-Control Design." *American Journal of Epidemiology* 168 (8): 890–96. doi:10.1093/aje/kwn205.

Tatar, M, Bartke, A, and Antebi, A. 2003. "The Endocrine Regulation of Aging by Insulin-like Signals." *Science* 299 (5611): 1346–51. doi:10.1126/science.1081447.

Templeton, A R, and Rankin, M. 1978. "Genetic Revolutions and the Control of Insect Populations." In *The Screw-Worm Problem*, edited by R.H. Richardson. Austin: University of Texas Press.

Thomas, D C. 2004. *Statistical Methods in Genetic Epidemiology*. Oxford: Oxford University Press.

Vaupel, J W, Baudisch, A, Dölling, M, Roach, D A, and Gampe, J. 2004. "The Case for Negative Senescence." *Theoretical Population Biology* 65 (4): 339–51. doi:10.1016/j.tpb.2003.12.003.

Vaupel, J W, Carey, J R, Christensen, K, Johnson, T E, Yashin, A I, Holm, N V, Iachine, I et al. 1998. "Biodemographic Trajectories of Longevity." *Science* 280 (5365): 855–60. doi:10.1126/science.280.5365.855.

Vesalius, A. 1543. "De humani corporis fabrica." Ex officina Ioannis Oporini. http://www.biodiversitylibrary.org/item/191384.

Visscher, P M, Brown, M A, McCarthy, M, and Yang, J. 2012. "Five Years of GWAS Discovery." *American Journal of Human Genetics* 90 (1). The American Society of Human Genetics: 7–24. doi:10.1016/j.ajhg.2011.11.029.

Visscher, P M, Mcevoy, N and Yang, J. 2010. "From Galton to GWAS: Quantitative Genetics of Human Height." *Genetics Research* 92 (5–6): 371–79. doi:10.1017/S0016672310000571.

Walker, D W, McColl, G, Jenkins, N L, Harris, J, and Lithgow , G J. 2000. "Evolution of Lifespan in C. Elegans." *Nature* 405 (6784): 296–97. doi:10.1038/35012693.

Wallace, H A C, Marques-Kranc, F, Richardson, M, Luna-Crespo, F, Sharpe, J A, Hughes, J, Wood, W G, Higgs, D, and Smith, A J H. 2007. "Manipulating the Mouse Genome to Engineer Precise Functional Syntenic Replacements??with Human Sequence." *Cell* 128 (1): 197–209. doi:10.1016/j.cell.2006.11.044.

Watson, J D. 1972. "Origin of Concatemeric T7 DNA." *Nature: New Biology* 239 (94): 197–201. doi:10.1038/newbio239197a0.

Weinberg, E D. 2007. "Iron Loading in Humans: A Risk Factor for Enhanced Morbidity and Mortality." *Journal of Nutritional & Environmental Medicine* 16 (1): 43–51. doi:10.1080/13590840601167685.

———. 2008. "Survival Advantage of the Hemochromatosis C282Y Mutation." *Perspectives in Biology and Medicine* 51 (1): 98–102. doi:10.1353/pbm.2008.0001.

Weismann, A. 1891. "Essays Upon Heredity and Kindred Biological Problems." *Oxford: Clarendon Press* 1: 1–5. doi:10.1007/s13398-014-0173-7.2.

Weiss, K M, and Terwilliger J D. 2000. "How Many Diseases Does It Take to Map a Gene with SNPs?" *Nature Genetics* 26 (october): 151–57. doi:10.1038/79866.

Welter, D, Macarthur, J, Morales, J, Burdett, T, Hall, P, Junkins, H, Klemm, A et al. 2014. "The NHGRI GWAS Catalog, a Curated Resource of SNP-Trait Associations." *Nucleic Acids Research* 42 (1): D1001-6. doi:10.1093/nar/gkt1229.

Westendorp, R G, and Kirkwood, T B 1998. "Human Longevity at the Cost of Reproductive Success." *Nature* 396 (December): 743–46. doi:10.1038/25519.

Whitlock, E P, Garlitz, B A, Harris, E L, Beil, T L, and Smith, P R. 2006. "Screening for Hereditary Hemochromatosis: A Systematic Review

for the U.S. Preventive Services Task Force." *Annals of Internal Medicine* 145 (3): 209–23. http://www.ncbi.nlm.nih.gov/pubmed/16880463.

Willcox, B J, Donlon, T A, He, Q, Chen, R, Grove, J S, Yano, K, Masaki, K H, Willcox, D C, Rodriguez, B, and Curb, J D. 2008. "FOXO3A Genotype Is Strongly Associated with Human Longevity." *Proceedings of the National Academy of Sciences of the United States of America* 105 (37): 13987–92. doi:10.1073/pnas.0801030105.

Williams, G C. 1957. "Pleiotropy, Natural Selection, and the Evolution of Senescence." *Evolution* 11: 398–411. doi:10.2307/2406060.

Williams, G C, and Nesse, R M. 1991. "The Dawn of Darwinian Medicine." *The Quarterly Review of Biology* 66 (1): 1–22.

Williams, P D, and Day, T. 2003. "Antagonistic Pleiotropy, Mortality Source Interactions, and the Evolutionary Theory of Senescence." *Evolution* 57 (7): 1478–88. doi:10.1111/j.0014-3820.2003.tb00356.x.

Williams, P D, Day, T, Fletcher, Q and Rowe, L. 2006. "The Shaping of Senescence in the Wild." *Trends in Ecology and Evolution* 21 (8): 458–63. doi:10.1016/j.tree.2006.05.008.

Wray, N R, and Visscher, P M. 2008. "Estimating Trait Heritability Estimating Heritability : Caveats." *Nature Education Knowledge* 1 (2008): 1–4.

The WTCC Consortium. 2007. "Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls." Article. *Nature* 447 (7145): 661–78. doi:10.1038/nature05911.

Yashin, AI, De Benedictis, G, Vaupel, J W, Tan, Q, Andreev, K F, Iachine, I A, Bonafe, M, et al. 1999. "Genes, Demography, and Life Span: The Contribution of Demographic Data in Genetic Studies on Aging and Longevity." *American Journal of Human Genetics* 65 (4):

1178–93. doi:10.1086/302572.

Yoshida, K, Fujisawa, T, Shan Hwang, J, Ikeo, K and Gojobori, T. 2006. "Degeneration after Sexual Differentiation in Hydra and Its Relevance to the Evolution of Aging." *Gene* 385: 64–70. doi:10.1016/j.gene.2006.06.031.

Ziegler, A, Konig, I K, and Pahlke, P. 2010. *A Statistical Approach to Genetic Epidemiology*. Weinheim: VCH, Wiley.

6. APPENDIX 2: OTHER WORK PARALLELLY DEVELOPED AND CONTRIBUTED TO DURING THIS PHD (2012-2016)

1. Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, Prado-Martinez J, **Rodríguez JA**, Rasmussen S, Quilez J, Ramírez O, Marigorta UM, Fernández-Callejo M, Prada ME, Encinas JM, Nielsen R, Netea MG, Novembre J, Sturm RA, Sabeti P, Marquès-Bonet T, Navarro A, Willerslev E, Lalueza-Fox C.

Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European.

Nature. 507, 7491 (2014)

2. Olalde I, Sánchez-Quinto F, Datta D, Marigorta UM, Chiang CW, **Rodríguez JA**, Fernández-Callejo M, González I, Montfort M, Matas-Lalueza L, Civit S, Luiselli D, Charlier P, Pettener D, Ramírez O, Navarro A, Himmelbauer H, Marquès-Bonet T and Lalueza-Fox C.

Genomic analysis of the blood attributed to Louis XVI (1754-1793), king of France.

Scientific Reports 4, 4666 (2014)

3. Marigorta UM, **Rodríguez JA** and Navarro A.

GWAS: a milestone in the road from genotypes to phenotypes. Book chapter in: Genome-Wide Association Studies: From Polymorphism to Personalized Medicine, Cambridge University Press, Cambridge (2016).

4. Dopazo J, Amadoz A, Bleda M, Garcia-Alonso L, Alemán A, García-García F, **Rodríguez JA**, Daub JT, Muntané G, Rueda A, Vela-Boza A, López-Domingo FJ, Florido JP, Arce P, Ruiz-Ferrer M, Méndez-Vidal C, Arnold TE, Spleiss O, Alvarez-Tejado M, Navarro A, Bhattacharya SS, Borrego S, Santoyo-López J and Antiñolo G.

267 Spanish exomes reveal population-specific differences in disease-related genetic variation

Molecular Biology and Evolution 33, 5 (2016).

5. Mieth B*, Kloft M*, **Rodríguez JA***, Sonnenburg S, Vobruba R, Morcillo-Suarez C, Farre X, Marigorta UM, Fehr E, Dickhaus T, Blanchard G, Schunk D, Navarro A and Müller KR

(*Shared first author)

COMBI — Combining Multiple Hypothesis Testing with Machine Learning Increases the Statistical Power of a Genome-wide Association Study of Seven Major Diseases

Under Review in *Scientific Reports*

6. Spataro N, **Rodríguez JA**, Navarro A and Bosch E

Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology.

Under Review in *Human Molecular Genetics*

7. Mandage R, Telford M, **Rodríguez JA**, Farré X, Layouni H Marigorta UM, Cundiff C, Heredia-Genestar JM, Navarro A,

Santpere G

Genetic factors affecting EBV load in lymphoblastoid cell lines derived from the 1000 Genome Project samples: A GWAS on EBV load

Under Review in *Human Molecular Genetics*