"A Biomathematics Investigation of Genetic Mutations Causing Alzheimer's Disease"

An Honors Thesis

by

Dannielle L. Skander

California, Pennsylvania

2017

California University of Pennsylvania

California, Pennsylvania

We hereby approve the Honors Thesis of

Dannielle L. Skander

Candidate for the degree of Bachelor of Arts

Date	Faculty
4/25/17	Louise Nicholson, PhD Honors Thesis Advisor
4/25/17	Maggie Habeeb, PhD Second Reader
4/25/17	Craig Fox, PhD Associate Director, Honors Program Honors Advisory Board
25 April 2017	M. G. Aune, PhD Director, Honors Program

A Bioinformatics Investigation of Genetic Mutations Causing Alzheimer's Disease

Dannielle Skander

Introduction

Alzheimer's disease is a devastating disease that is estimated to affect more than 5 million Americans currently[1]. It is estimated to effect approximately 47 million people worldwide, only to have this number continue to increase to 76 million by 2030[2]. Alzheimer's disease is most common in Western Europe, with North America falling close behind, and is least common in Africa[3]. In this thesis study, I will be examining various genetic risk factors associated with Alzheimer's disease and their prevalence in different populations.

Cost of Alzheimer's disease to Society

Alzheimer's disease is the most expensive disease in America, costing more than cancer and heart disease[4]. The cost of caring for patients in 2016 was approximately $236 billion[3]. The global cost is about $604 billion[2] but is estimated to increase to $1.1 trillion by 2050[1]. In addition to the costs of society, caregivers also have a huge cost. Caregivers spend approximately $9.7 billion of their own in extra health care costs for their patient. They give around 17.9 billion hours of unpaid care[3] and these hours value over $230 billion[1]. Patients with Alzheimer 's disease, in the last five years of life, spend much more than patients with cancer or heart disease. On average, they spend approximately $111,000 more[4]. The cost to society, the patients and the caregivers will continue to rise as more people develop Alzheimer's disease.

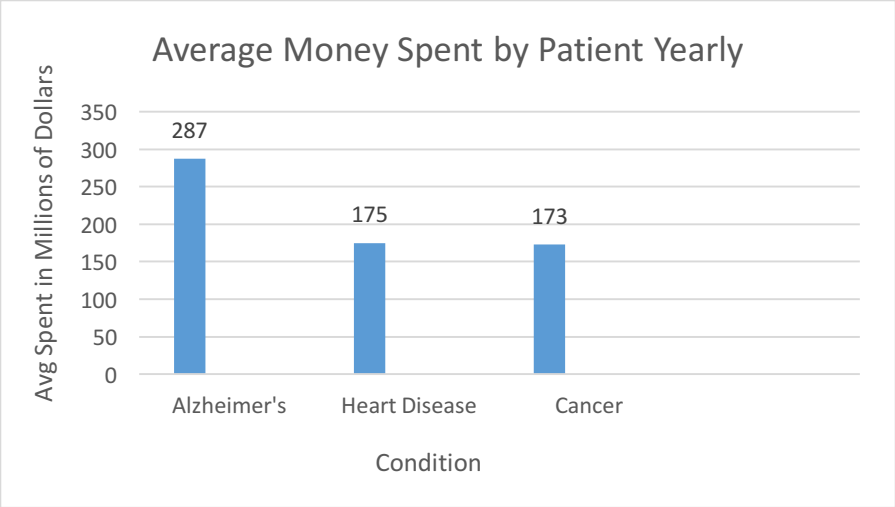

Figure 1: Average money spent by patients having Alzheimer's, heart disease and cancer[5].

The history of Alzheimer's disease

Alzheimer's disease was discovered by Alois Azheimer in 1906[6.] Alzheimer was a psychiatrist and neuroanatomist during his time working at the Frankfurt Psychiatric Hospital. He discovered Alzheimer's disease after he observed one of his patients, Auguste D[7]. She was well until March 1901 when her husband said she developed paranoia and started having difficulties handling money and remembering things[8]. This patient also developed some personality changes[6]. When Alois Alzheimer observed her, he said she had severely impaired recall memory and could not recall something she had just said moments before[8]. After her death, Alzheimer was able do to an autopsy using some brain material. He discovered changes in cell and tissue structures in the brain, which are now known as plaques and neurofibrillary tangles[9]. Another one of the patients, Josef F, was diagnosed with Alzheimer's upon his death. When Alzheimer looked at the brain, he found only plaques and no neurofibrillary tangles.

Alzheimer diagnosed both patients, among others, with Alzheimer's despite the slight difference in brain material. In later years, scientists and doctors have reexamined the patient's brain material and found that Alzheimer was correct in diagnosing them both patients with the disease. Scientists determined that the difference in brain alterations found was due to the development of the disease[7].

When Dr. Alzheimer did an autopsy on his patient, he noticed shrinkage and abnormal deposits around the neurons[6]. After significant technological and scientific advances, scientists were able to study the brain of Alzheimer's patients in much more detail than before. In 1984, the plaques Dr. Alzheimer noted were found to be beta amyloid protein plaques. A few years later in 1986 scientists found out that the tangles were made of the tau protein[7] . These will be discussed below.

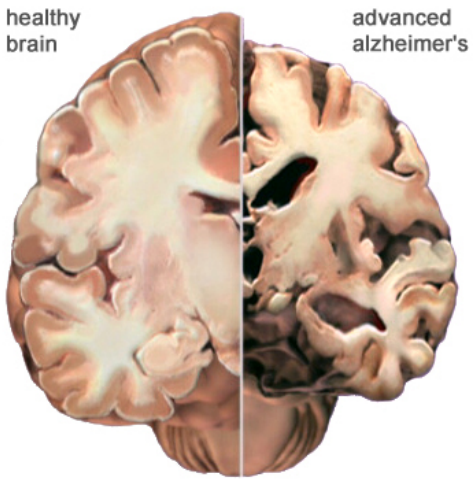

Figure 2: Picture of a healthy brain vs. a brain with Alzheimer's Disease[10] obtained by More Brain Changes, Alzheimer's Association, 2011

Pathophysiology of Alzheimer's disease

Alzheimer's is a devastating irreversible disease affecting neurons in the brain. The damage normally starts in the hippocampus, which is very important for memory formation, and spreads from there. The connections between neurons will weaken and because of this, the neurons will eventually die. As described above, the disease is characterized by plaques, made of the amyloid protein, and neurofibrillary tangles, consisting of the tau protein, in the brain. The amount of the tau protein found in the neurofibrillary tangles is proportionate to the degree of memory loss a patient experiences. Scientists are not certain about the exact function the amyloid plaques play, but they believe that the high concentration of plaques found on the hippocampus and cerebral cortex in Alzheimer's patients plays a role in the neuronal degenerative process[11] Figure 2 illustrates the tangles and plaques that are found in the brain of a patient with Alzheimer's disease.

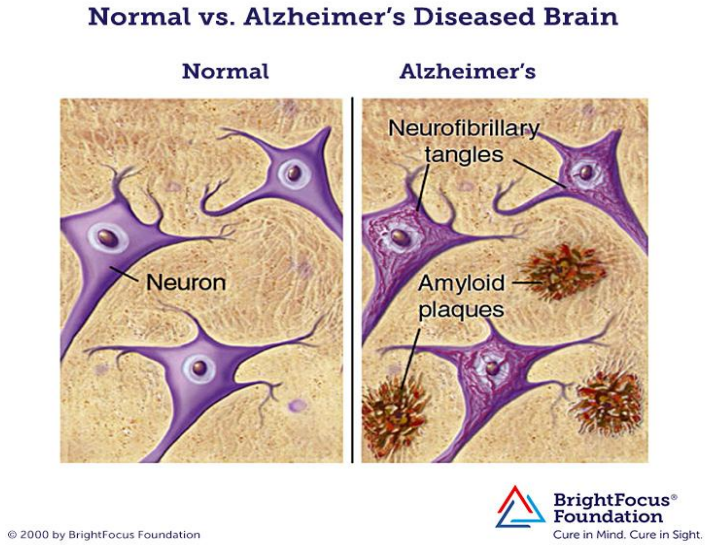

Figure 3: Brain plaques and tangles. Left side is a normal brain and the right side is the amyloid plaques and neurofibirillary tangles found in the brain of an Alzheimer's patient[12]. Picture obtained by BrightFocus Foundation.

Activated glial cells surround the amyloid plaques in diseased patients; these cells are responsible for secreting a large amount of inflammatory molecules, such as pro-inflammatory cytokines, chemokines, and reactive oxygen species (ROS)[13]. The molecules released impair the normal neurophysiologic conditions, causing problems relating to cognition, learning and memory. Other biological processes, such as dysfunction of lysosomal/proteasomal degredation, mitochondrial dysfunction and oxidative stress, have been associated with the disease [14].

This is a progressive disease, which means that as the disease progresses through the brain, the symptoms and damage done keep worsening. The most common symptom is memory loss[11]. Some other symptoms in addition to memory loss are personality and behavioral changes, impaired judgement, wandering, paranoia and language problems[9]. The symptoms may not be experienced by all patients and if experienced, may not be experienced to the same degree.

One hypothesis about the pathogenesis of the disease is the amyloid cascade hypothesis. This hypothesis describes the cleavage of the amyloid precursor protein (APP) which leads to overproduction, oligomerization, and later the deposition of the amyloid beta protein aggregates in the central nervous system[15]. The oligomerization is amyloid beta (Aβ) is thought to initiate the sequence of events that cause the degeneration of neuronal synapses. This degeneration causes inflammation and the death of many neurons.

T Lymphocytes are thought to play an important role in the neuroinflammatory processes of Alzheimer's. There are increased levels of peripheral T cells in postmortem brains of patients when compared to brain tissue from other

neurodegenerative diseases. While scientists believe T cells play a role, they cannot yet tell if it is a damaging or helping affect. T cells specific for Aβ1-40 are found in healthy individuals but T cells specific for Aβ1-42 are found in individuals with the disease[15].

Genes associated with Alzheimer's disease

There are two categories of Alzheimer's disease, early onset and late onset. Early onset happens earlier in life and is believed to be caused by genetic mutations. Patients that are diagnosed with early onset Alzheimer's may also experience myoclonus, which is muscle twitching and spasm [16]. Late onset Alzheimer's typically occurs after the age of 65 and is not definitely known to be caused by genetic mutations, although there are genetic risk factors involved. Three genes have been identified to date that, when mutated, can cause early-onset Alzheimer's disease. These genes are Presenilin 1, presenilin 2, and apolipoprotein E.

The normal functions of presenilin 1 (PSN1) include autophagy, maintenance of calcium homeostasis and the meditation of correct interactions between the endoplasmic reticulum and mitochondria [17]. PSN 1 also has functions concerning the amyloid protein. The amyloid plaques that are characteristic of Alzheimer's disease are formed by the accumulation of amyloid-beta, a neurotoxin that is produced by the breakage of amyloid-beta precursor protein (APP) [18]. Presenilin 1 induces this breakage when the gene is mutated. Inheriting a mutation in the presenilin 1 (PSN1) gene guarantees that a person will develop Alzheimer's disease.

A deficiency of presenilin 2 (PSN 2) is associated with inflammatory effects in microglia [19]. When neuroinflammation of microglia occurs, neurotoxic and

neuroprotective consequences occur to the central nervous system. Scientists believe that the loss of PSN 2 functions contribute to the inflammatory characteristics of Alzheimer's disease. A mutation in the presenilin 2 (PSN2) gene gives a 95% chance of developing the disease[20]. Both PSN1 and PSN 2 genes also code for proteolytic enzymes that cleave APP into the amyloid-beta and other fragments, hence why there is a build up when there is a mutation in the genes [14].

The last gene scientists know to be associated with Alzheimer's disease is the apolipoprotein E (ApoE) gene[21]. There are three forms of the gene that a person can inherit, namely, the e2, e3, and e4 forms. The e4 form gives the person the highest chance of developing Alzheimer's [20]. If a person inherits only one copy, they are three times as highly (compared to e3) to develop the disease but if they inherit two copies, their chances increase eight to twelve fold. The most prevalent genetic risk factor for the disease is the e4 allele of the gene[22]. The functions of the gene include lipid transport throughout the body and damaged tissue repair[23]. The gene also plays a role in neuronal development and plasticity and has an effect on the nutrient intake conditions[24].

There have been a multitude of other studies conducted on genes possibly associated with late-onset Alzheimer's disease. Several of the genes identified, including CDK5, LMO4, PTEN and TGFβ 1, increase the abnormal protein aggregation and other characteristics consistent with that of the disease [14]. These are similar to the effects seen with mutations in ApoE. A further nine genes that are important for the pathology of Alzheimer's disease were identified in 2016. These 9 new loci involve the

genes ABCA7, BIN1, CASS4, CD33, MEF2C, MS4A6A, PICALM, SORL1 and ZCWPW1[14].

Prognosis

Alzheimer's disease starts to become noticeable around age 65. The difference in average life expectancies between different populations will have an effect on the number of individuals that live to express the disease. Africa, where the average life expectancy is only 51, may not see as much of the disease because the people are not living to an age where symptoms become noticeable (Figure 4). The life expectancy will be taken into account when considering the prevalence of Alzheimer's disease in the different populations included in this study.

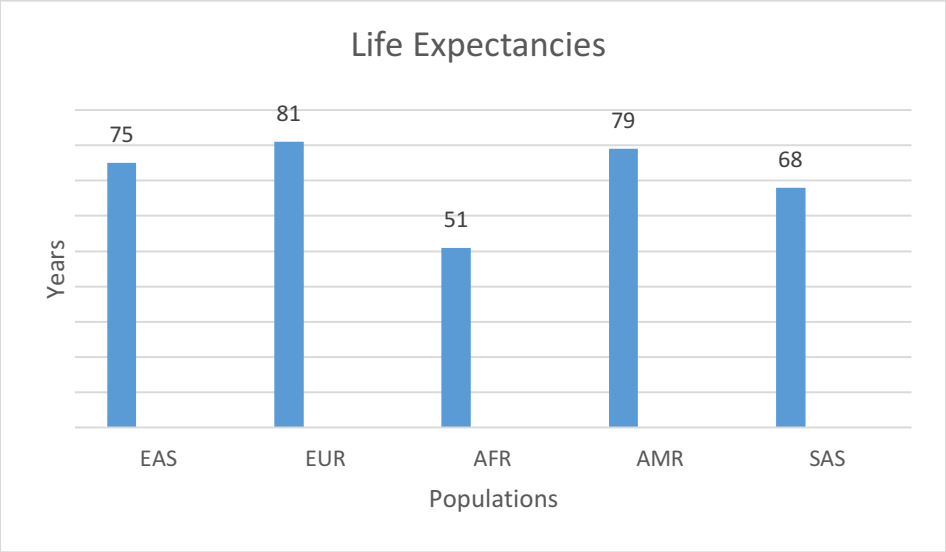

Figure 4: Average life expectancies of different populations[25]. EAS-Eastern Asia; EUR-Europe, AFR-Africa, AMR-America, SAS-Southern Asia

There is currently no treatment for Alzheimer's disease. There is a lack of knowledge about the disease and a lack of early Alzheimer's disease biomarkers, which together hinders the treatment of the disease[15]. When treatments are introduced, the disease has already spread enough throughout the brain to interfere with daily tasks. Doctors can prescribe drugs that slow down the progression of the disease, but the efficacy of the drug may vary between patients. Other drugs that doctors prescribe deal with the symptoms that patients have, such as for their sleeping and anxiety problems.

Methods

Dataset: Genetic information was obtained from the dbSNP Short Genetic Variation database at the National Center for Biotechnology Information. Data was collected on single nucleotide polymorphisms (SNPs) within the presenilin 1 (PSN1), presenilin 2 (PSN2), and apolipoprotein E (ApoE) genes.

Analysis: Each SNP allele was categorized according to the effects that they had on protein function and disease phenotype, and those most likely to result in disease were chosen for further study. Frequency data was obtained for each SNP allele, and the Hardy Weinberg equation ($p^2+2pq+q^2$) was used to calculate the percentage of people estimated to have the mutation. This was calculated for different sub-populations (EAS-Eastern Asia; EUR-Europe, AFR-Africa, AMR-America, SAS-Southern Asia), where data was available. Data on the current number of individuals living in each population was obtained (Worldometers.info), as well as the percentage of the population over 65 (data.worldbank.org), and these were used to calculate the predicted number of affected individuals in each population.

Results

Presenilin 1, presenilin 2 and apolipoprotein E were the three genes chosen for this study. A database search was performed to identify genetic variation (SNPs) within the genes, and SNP frequencies in different populations. When the NCBI database was searched for presenilin 2 mutations for Homo sapiens was searched, no results were found. For this reason, Presenilin 2 was left out of the research.

When looking at the NCBI database, the results were limited to Homo sapiens. The results were further limited to missense, nonsense and frameshift mutations, excluding synonymous mutations. In synonymous mutations there is a single nucleotide change but the change does not affect the amino acid produced. A missense mutation is a mutation where one nucleotide is changed, changing the amino acid produced but not affecting the rest of the protein, while a nonsense mutation is a mutation in which a single nucleotide change created a stop codon, causing the rest of the protein sequence to be lost. In frameshift mutations, one nucleotide is added or deleted so the rest of the nucleotides shift over, thus changing the entire protein from that point on. The hits focused on were the ones that had any clinical significance and the ones that produced data for different populations, Eastern Asia (EAS), Europe (EUR), Africa (AFR), America (AMR) and Southern Asia (SAS). Data was also collected on the number of individuals in each population, as well as the number of individual over 65, the age at which Alzheimer's typically appears.

The Hardy Weinberg equation, $p^2+2pq+q^2$, was used to calculate the expected frequency of the genetic mutation. The equation describes and predicts genotype frequencies given a specific allele frequency. P is the variable used for the dominant

allele and q is for the recessive allele. The p^2 term describes the homozygous dominant individuals, the 2pq is the heterozygous individuals and the q^2 is the homozygous recessive individuals. Only the first section, p^2+2pq was used because this disease is dominant. This frequency was multiplied by the current number of individuals in the population to get the projected number of individuals that have the mutation (Figure 6 and 7, blue bars). To get a better estimate of the number of people in each population with Alzheimer's disease, I multiplied the projected number of people by the percentage of individuals in each population living over 65 (Figure 6 and 7, orange bars). The life expectancy factor decreased the amount of people having the disease because most people will not be living to an age where the symptoms become prevalent.

The total number of individuals with each mutation is an upper bound estimate. The formula $P(A \cap B \cap C)$ less then or equal to $P(A)+P(B)+P(C)-P(A \cap C)-P(A \cap B)-P(B \cap C)+P(A \cap B \cap C)$ is used to calculate the exact number of individual with the mutations. Because we do not have enough on the likelihood that a patient has another genetic mutation on top of the one they already have, we cannot go any further then the upper bound estimate. This estimate is the most amount of people that will have the mutation. We assume that because the genetic mutations are not exclusive, this number will be smaller.

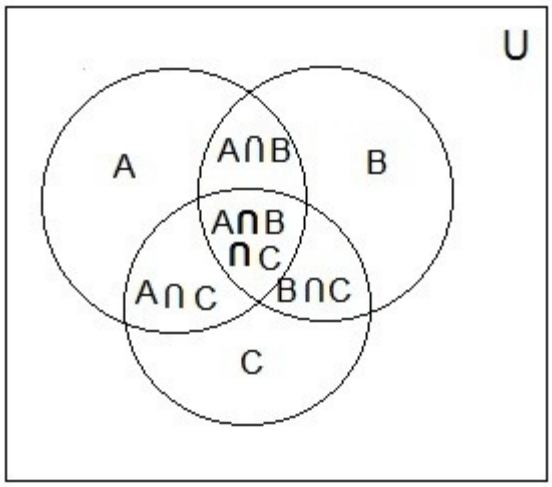

Figure 5: Venn Diagram showing the intersection of (A ∩B ∩C)

ApoE:

I identified 240 SNPs in the ApoE gene. Of these many were excluded, because they were synonymous mutations and would not affect the protein produced, and 35 non-synonymous (missense, non-sense, frameshift) mutations were chosen for further study.

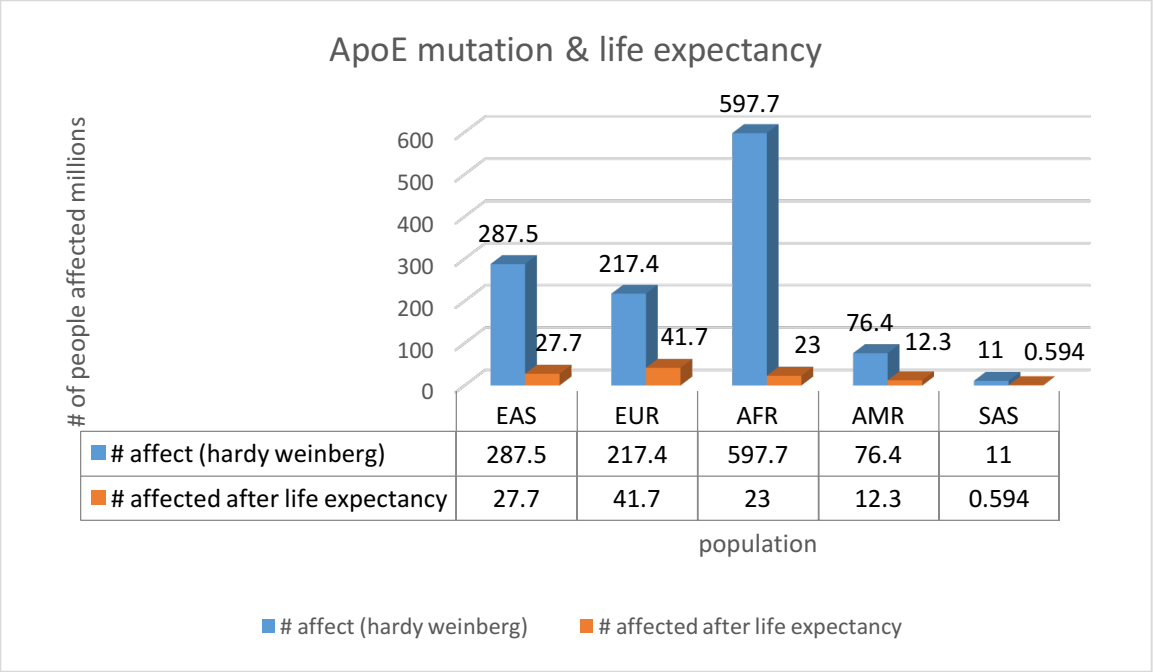

Figure 6: The number of individuals having the Apo E mutation that will be affected after life expectancy is taken into consideration

The ApoE mutation is expected to affect 104,306,378.6 individuals within the Eastern Asia, Europe, Africa, America and Southern Asia populations. The number of individuals affected by the mutation significantly drops once life expectancy is introduced as a variable.

Table 1: The exact number of individuals expected to have Alzheimer's in each population after life expectancy is introduced.

ApoE Mutation (population)	
Population	Number of people with Alzheimer's
EAS	27,654,193.53
EUR	41,716,996.54
AFR	23,041,663.99
AMR	11,299,062.9
SAS	594,461.6691
TOTAL	104,306,378.7

All of the known pathogenic mutations for Apo E were gathered from the database. Allele frequencies were given and used to calculate the expected frequency of affected individuals (homozygous dominant and heterozygous, from the Hardy-Weinberg equation) for the given population. Table 2 is the frequency of each mutation for the world population for the pathogenic cases. Only pathogenic cases were analyzed for table 2.

Table 2: The frequencies for all Apo E mutation data found for aggregated populations

Frequencies for Apo E mutations with aggregated populations	
Mutation number	Frequency
1	0.001648%
2	0.003294%
3	0.003560%
4	0.008859%
5	0.086159%
6	0.039966%
7	0.020089%

PSN 1:

I identified 209 SNPs in the PSN1 gene. Of these many were excluded, because they were synonymous mutations and would not affect the protein produced, and 15 non-synonymous (missense, non-sense, frameshift) mutations were chosen for further study.

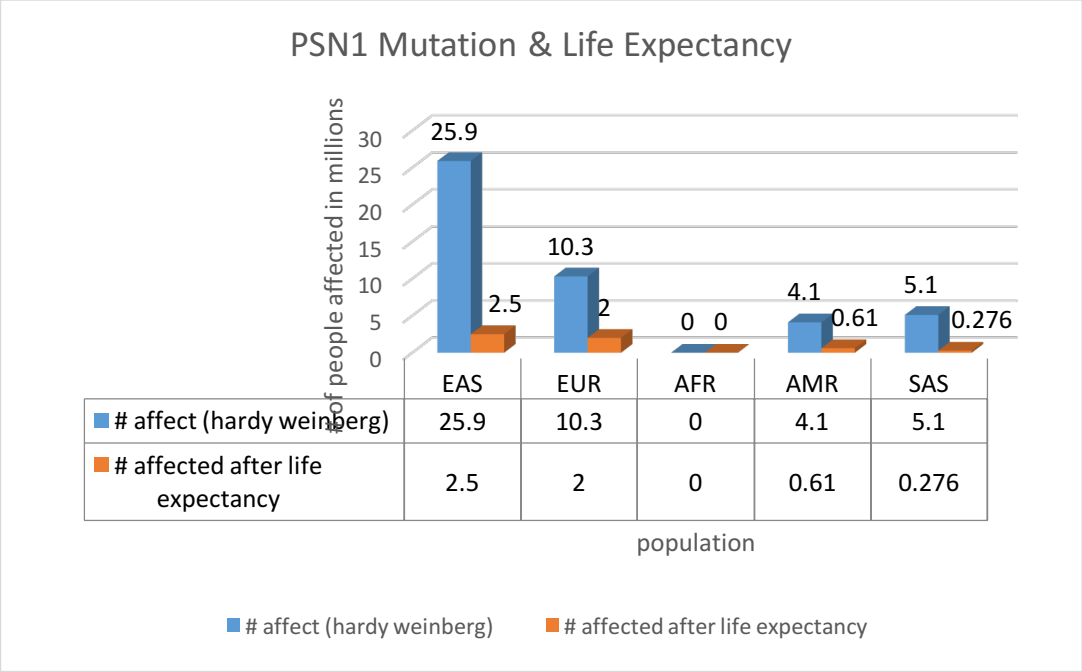

Figure 7: the number of individuals having the PSN1 mutation that will be affected after life expectancy is taken into consideration

The PSN1 mutation is expected to affect 5,362,974.684 individuals within the Eastern Asia, Europe, America and Southern Asia populations. There are no individuals in the African population affected by mutations in the PSN 1 gene. The number of individuals

affected by the mutation significantly drops once life expectancy is introduced as a variable.

Table 3: The exact number of individuals expected to have Alzheimer's in each population after life expectancy is introduced

PSN1 Mutation (after life expectancy)	
Population	Number of people with Alzheimer's
EAS	2,493,154.959
EUR	1,982,707.613
AFR	0
AMR	610,297.185
SAS	276,814.9267
Total	5,362,974.684

All of the known pathogenic mutations for PSN1 were gathered from the database. Allele frequencies were given and used to calculate the Hardy Weinberg frequency for the given population. Table 4 is the frequency of each mutation for the world population for the pathogenic cases. Only pathogenic cases were analyzed for table 4.

Table 4: The frequencies for all PSN1 mutation data found for aggregated

Frequencies for PSN1 mutations with aggregated populations	
Mutation number	Frequency
1	0.003294%
2	0.001648%

Discussion

I project that the ApoE mutation will affect approximately 104,306,378.7 individuals and the PSN1 mutation will affect approximately 5,362,974.684 individuals across the nation. The total number of individuals I project that will be affected by either of these mutations is 109,669,353.4. This number is based on the current world population and will fluctuate with the change in world population. This approximation of 109 is an upper bound estimate of the number of individuals having Alzheimer's. The 109 million is based on the assumption that each person in the world can only have one mutation causing the disease, which we know is not true. For example, person A could have three mutations that cause Alzheimer's but in my data, I am considering her as 3 different people with the disease because of her three mutations. This projection is also assuming that every missense, nonsense and frameshift mutation is pathogenic, which might not be the case. Depending on where the mutation occurs on the protein, it may or may not have a pathogenic effect. Current literature estimates that there will be 76 million people worldly having the disease[2]. This is significantly smaller than my projection, most likely due to my estimate being an upper bound and assuming every mutation to be pathogenic.

The individuals that comprise each population in the database could also skew the data. It is unlikely that the sampling is completely random so it may not be sampling every population equally. The African population is seeing a zero frequency for one mutation which could be caused by the lack of resources in this area. There may be only a small number of individuals capable of getting their genome sequenced, so the

sample may not accurately represent the genetic variation present in the entire population.

Current literature shows Europe having the highest prevalency of Alzhiemer's with America following behind[3]. My data supports Europe having the highest but does not support America being in second. According to the data I produced, Eastern Asia would be the second. I would expect that Southern Asia and Africa would not spend as much time, money or resources on Alzheimer's disease research than Europe, America or Eastern Asia would. After life expectancy was factored in, the number of individuals dropped significantly for Eastern Asia and Africa. If the health care improves in these parts of the world, they should see an increase in the life expectancies. If these life expectancies increase to about 65, they will see much more people with the disease. At that time, I would expect Eastern Asia and Africa to spend more resources and time on the disease.

Natural selection is the differential reproduction of genotypes. It unlikely to have been taking place for any mutations in PSN1 or ApoE for any population, given that Alzheimer's is a relatively recent finding, first identified only in 1906. The symptoms of the disease occur later in life, after a reproductive age. While natural selection is probably not happening, selection might be; selection here being the choosing whether to reproduce or not. People are now getting to see their parents and grandparent develop the disease because the life expectancies are much better than they used to be. Seeing this, people might choose to get genetic testing done to see if they have a mutation in PSN1, PSN2, or ApoE.

One problematic issue with Alzheimer's disease is that there is no cure, but the treatment would need to start years before symptoms become noticeable. Classical gene therapy is when scientists are able to delete an entire gene out of the DNA. This would not be beneficial to patients with Alzheimer's because the genes that cause the disease have very important functions. If the gene was lost, the patient could have a multitude of other problems. One possible future treatment could be the CRISPR technology. CRISPR uses a protein, CAS 9, and guide RNA to edit genes[26]. The protein and guide RNA go into the DNA, find the mutated section of DNA, remove the damaged part and replace with the correct set. The benefit to this is that the genes are not completely lost, and therefore, do not lose their entire function.

Alzheimer's disease currently costs the nation a total of $604 billion[2], which isn't including the $9.7 billion extra that families pay for care takers and other health costs[3]. Because the number of patients with the disease are expected to increase, so is the cost to the nation. The cost is estimated to be about $1.1 trillion by 2050[1] if the prevalence does not decrease. The nation, especially regions such as eastern Asia and Europe, needs to increase their efforts on finding a cure for the disease. Specifically, I believe the focus should be more on the ApoE gene because it affects a significantly higher amount of people across the nation. A cure would significantly help decrease the amount of people having the disease and consequently, the cost to society.

Appendix. Data Tables

Results from PSN1, pathogenic group

Function	Allele	"correct" allele	Amino	"correct" amino	codon position	amino position	A	T	C	G	-	population	p^2+2p
frame shift	GG	-	Glu	Val	2	288				(GG)0.00001647	0.99998355	ExAc aggregated	0.003
missense	C	G	Arg	Gly	1	280	NA	NA	NA	NA	NA	NA	NA
frame shift	-	G	Ile	Asp	3	232	NA	NA	NA	NA	NA	NA	NA
missense	G	A	Ser	Asn	2	229	NA	NA	NA	NA	NA	NA	NA
nonsense	G	T	-	Tyr	3	220	NA	NA	NA	NA	NA	NA	NA
nonsense	T	C	-	Gln	1	216		0.00000824	0.99999177			ExAc aggregated	0.001
frame shift	A	-	Ser	Ala	3	168	NA	NA	NA	NA	NA	NA	NA

Results for PSN 1 population, part 1

Chromosome position	mRNA position	Cluster ID	Clinical Sig	Function	allele	"correct" allele	amino	"correct " amino	codon position	amino position	A	T	C	G	-	Population	p^2+2pq	# of pop	# affected	% of population over 65
33014755	1063	rs534306255		missense	G	A	Ser	Asn	2	329		1	0			EAS	0	1623153468	0	9.62%
												1	0			EUR	0	739107476	0	19.19%
												1	0			AFR	0	1237666164	0	3.86%
												0.99859995	0.0014			AMR	0.279804%	362456416	1014167.55	14.79%
												1	0			SAS	0	63688069	0	5.42%
												0.99997526	0.00002471			ExAc aggregated	0.004942%			
33015729	927	rs576075856		missense	G	A	Val	Met	1	284		1	0			EAS	0	1623153468	0	
												1	0			EUR	0	739107476	0	
												1	0			AFR	0	1237666164	0	
												0.99859995	0.0014			AMR	0.279804%	362456416	1014167.55	
												1	0			SAS	0	63688069	0	
												0.99998355	0.00001647			ExAc aggregated	0.003294%			
33015753	903	rs543405986		missense	A	G	Arg	Gly	1	276		0	1			EAS	0	1623153468	0	
												0	1			EUR	0	739107476	0	
												0	1			AFR	0	1237666164	0	
												0	1			AMR	0	362456416	0	
												0.001	0.99899995			SAS	0.19990%	63688069	127312.45	
												0.00000824	0.99999177			ExAc aggregated	0.001648%			
33015788	868	rs142248153		missense	G	A	Ser	Asn	2	264		1	0			EAS	0	1623153468	0	
												0.99900001	0.001			EUR	0.1999%	739107476	1477475.84	
												1	0			AFR	0	1237666164	0	
												1	0			AMR	0	362456416	0	
												1	0			SAS	0	63688069	0	
												0.99995881	0.00004118			ExAc aggregated	0.008236%			

Results for PSN 1 population, part 2

33015703	953	rs72555746		synonymous	A	G	Gly	Gly	3	292	0.002			0.99800003		EAS	0.3996%	1623153468	6486121.26	
											0.004			0.99599999		EUR	0.7984%	739107476	5901034.09	
											0			1		AFR	0	1237666164	0	
											0.0029			0.99710006		AMR	0.579159%	362456416	2099198.95	
											0.0399			0.9601		SAS	7.82080%	63688069	4980916.5	
											0.00762692			0.99237305		ExAc aggregated	1.51957%			
											0.00265487			0.99734515		ESP Cohort	0.530269%			
											0.0025			0.9975		CSAgilent	0.499375%			
33017477	747	rs181671227		missense	A	G	Ile	Val	1	224		0.005	0.995			EAS	0.9975%	1623153468	16190955.8	
												0	1			EUR	0	739107476	0	
												0	1			AFR	0	1237666164	0	
												0	1			AMR	0	362456416	0	
												0	1			SAS	0	63688069	0	
												0.00013178	0.99986821			ExAc aggregated	0.026354%			
33017523	701	rs546177345		synonymous	C	T	Asp	Asp	3	208	0.99900001			0.001		EAS	0.1999%	1623153468	3244683.78	
											1			0		EUR	0	739107476	0	
											1			0		AFR	0	1237666164	0	
											1			0		AMR	0	362456416	0	
											1			0		SAS	0	63688069	0	
											0.99999177			0.00000824		ExAc aggregated				
33017588	636	rs186325627		missense	C	A	His	Asn	1	187		1		0		EAS	0	1623153468	0	
												0.99799997		0.002		EUR	0.39960%	739107476	2953473.47	
												1		0		AFR	0	1237666164	0	
												1		0		AMR	0	362456416	0	
												1		0		SAS	0	63688069	0	

Results from Apo E pathogenic, part 1

osome position	mRNA Position	Cluster ID	Clinical Sig	Function	allele	"correct" allele	amino	"correct" amino	codon position	amino position	A	T	C	G	-	Population	p^2+2pq	# of pop	#
44907777	177	rs121918392	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G	Lys	Gly	1	21	0.001			0.99900001		EAS	0.1999%	1623153468	
											0			1		EUR	0	739107476	
											0			1		AFR	0	1237666164	
											0			1		AMR	0	362456416	
											0			1		SAS	0	63688069	
											0.00006592			0.99993408		ExAc Aggregated	0.013184%		
44907807	207	rs201672011	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G	Lys	Glu	1	31	0			1		EAS	0	1623153468	
											0			1		EUR	0	739107476	
											0.008			0.99919999		AFR	1.60512%	1237666164	
											0.0029			0.997100006		AMR	0.579159%	362456416	
											0			1		SAS	0	63688069	
											0.00014828			0.9995817		ExAc Aggregated	0.029646%		
44907843	243	rs121918399	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	T	C	Cys	Arg	1	143		0.00000824	0.99999177			ExAc Aggregated	0.001648%		
44907853	253	rs769452	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	C	T	Pro	Leu	2	46		1	0			EAS	0	1623153468	
												0.99599999	0.004			EUR	0.79840%	739107476	
												1	0			AFR	0	1237666164	
												1	0			AMR	0	362456416	
												1	0			SAS	0	63688069	
												0.99759495	0.00240507			ExAc Aggregated	0.480436%		
												0.99867254	0.00132743			ESP Cohort	0.265310%		
												0.99349999	0.0065			CSAgilent	1.29577%		
44907894	294	rs28931576	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	G	A	Ala	Thr	1	60	0.99998355			0.00001647		ExAc Aggregated	0.003294%		
44908601	421	rs11083750	other, Alzheimer's disease, hyperlipoproteinemia type 3, lipoprotein glomerulopathy, sea blue histocyte disease	missense	A	C	Gln	Pro	2	102			0.99999177	0.00000824		ExAc Aggregated	??????		
44908645	465	rs28931577	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G	Thr	Ala	1	117	NA	NA	NA	NA	NA	NA	NA		
44908660	480	rs587778876	untested, major depressive disorder, unipolar depression, seasonal affective disorder	missense	A	C	Met	Leu	1	122	NA	NA	NA	NA	NA	NA	NA		
44908684	504	rs429358	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	C	T	Arg	Cys	1	130		0.91369998	0.0863			EAS	16.5152%	1623153468	
												0.8449001	0.15509999			EUR	28.6144%	739107476	
												0.7321997	0.2678			AFR	46.3883%	1237666164	
												0.89630002	0.1037			AMR	19.6646%	362456416	
												0.9131	0.0869			SAS	16.6248%	63688069	
												0.89976752	0.10023248			ExAc Aggregated	19.0418%		
44908747	567	rs587778877	untested, major depressive disorder, unipolar depression, seasonal affective disorder	missense	A	C	Met	Leu	1	151	0.0000178		0.99998218			ExAc Aggregated	0.003560%		
44908979	799	rs121918396	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	nonsense	A	G	-	Trp	2	228	NA	NA	NA	NA	NA	NA	NA		

Results from Apo E pathogenic, part 2

44909057	877	rs199768005	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	T		Glu	Val		2	254	0.00043089	0.99956912				ExAc Aggregated	0.086159%	
44909080	900	rs140808909	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G		Lys	Glu		1	262	0.003			0.99700004		EAS	0.59910%	1623153
													0			1		EUR	0	739107
													0			1		AFR	0	1237666
													0			1		AMR	0	362456
													0			1		SAS	0	63688
													0.00015836			0.99984163		ExAc Aggregated	0.031669%	
44909083	903	rs190853081	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G		Lys	Glu		1	263	0.003			0.99700004		EAS	0.59910%	1623153
													0			1		EUR	0	739107
													0			1		AFR	0	1237666
													0			1		AMR	0	362456
													0			1		SAS	0	63688
													0.00015833			0.99984169		ExAc Aggregated	0.031663%	
44909101	921	rs267606661	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	G	C		Gly	Arg		1	269			0.99980015	0.00019985		ExAc Aggregated	0.039966%	
														0.00000833	0.99999166			ExAc Aggregated	?????	
44909171	991	rs121918398	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	G		His	Arg		2	292	NA	NA	NA	NA	NA	NA	NA	
44909236	1056	rs28931579	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	C	A		Arg	Ser		1	314	0.99989957		0.00010045			ExAc Aggregated	0.020089%	
44908751	571	rs28931578	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	C		Gln	Pro		2	152	NA	NA	NA	NA	NA	NA	NA	
44908756	576	rs121918393	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	T		Ser	Cys		1	154		0.00003587	0.99996412			ExAc Aggregated	?????	
44908783	603	rs769455	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	T	C		Cys	Arg		1	163		0	1			EAS	0	1623153
														0	1			EUR	0	739107
														0.025	0.97499996			AFR	4.93750%	1237666
														0.0058	0.99419999			AMR	1.15664%	362456
														0	1			SAS	0	63688
														0.0012276	0.99877238			ExAc Aggregated	0.245369%	
44908784	604	rs121918397	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	A	C		His	Pro		2	163	NA	NA	NA	NA	NA	NA	NA	
44908786	606	rs121918394	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	C	G		Gln	Glu		1	164	NA	NA	NA	NA	NA	NA	NA	
44908804	624	rs267606662	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	C	G		Pro	Ala		1	170	NA	NA	NA	NA	NA	NA	NA	
44908822	642	rs7412	pathogenic; Alzheimer's disease, hyperlipoproteinemia type 3, liboprotein, glomerulopathy, sea blue histocyte disease	missense	T	C		Cys	Arg		1	176		0.1002	0.8998			EAS	19.0360%	1623153
														0.0626	0.9370004			EUR	12.1231%	739107
														0.1029	0.89709997			AFR	19.5212%	1237666
														0.0476	0.95239997			AMR	9.29342%	362456
														0.044	0.95600003			SAS	8.60640%	63688

Results from Apo E population, part 1

romosome sition	mRNA position	Cluster ID	Clinical Sig	Function	allele	"correct" allele	amino	"correct" amino	codon position	amino position	A	T	C	G	-	population	p^2+2pq	pop #	# affected
44906655	147	rs144354013		missense	G	A	Ala	Thr	1	11	1	0				EAS	?????	1623153468	
											1	0				EUR	?????	739107476	
											1	0				AFR	?????	1237666164	
											0.99859995			0.0014		AMR	0.279804%	362456416	1014167.55
											1	0				SAS	????	63688069	
											0.99995881	0.00004118				ExAcAggregated	?????		
											0.99998355			0.00001647		ExAcAggregated	0.003294%		
											0.99900001	0.001				CSAgilent	?????		
44906664	159	rs559532612		missense	A	G	Thr	Ala	1	14	0	1				EAS	0	1623153468	0
											0	1				EUR	0	739107476	0
											0	1				AFR	0	1237666164	0
											0.0014	0.99859995				AMR	0.279804%	362456416	1014167.55
											0	1				SAS	0	63688069	0
44907768	168	rs533904656		missense	A	G	Thr	Ala	1	18	0.003			0.99700004		EAS	0.599100%	1623153468	9724312.43
											0			1		EUR	0	739107476	0
											0			1		AFR	0	1237666164	0
											0			1		AMR	0	362456416	0
											0			1		SAS	0	63688069	0
											0.00010714			0.99989283		ExAcAggregated	0.021427%		
													0.00001649	0.99998349		ExAcAggregated	??????		
44907777	177	rs121918392	pathoge	missense	A	G	Lys	Gly	1	21	0.001			0.99900001		EAS	0.1999%	1623153468	3244683.78
											0			1		EUR	0	739107476	0
											0			1		AFR	0	1237666164	0
											0			1		AMR	0	362456416	0
											0			1		SAS	0	63688069	0
											0.00006592			0.99993408		ExAc Aggregated	0.013184%		
44907807	207	rs201672011	pathoge	missense	A	G	Lys	Glu	1	31	0			1		EAS	0	1623153468	0
											0			1		EUR	0	739107476	0
											0.008			0.99919999		AFR	1.60512%	1237666164	19866027.1
											0.0029			0.99710001		AMR	0.579159%	362456416	2099198.95
											0			1		SAS	0	63688069	0

Results from Apo E population, part 2

44907853	253	rs769452	pathoge	missense	C	T	Pro	Leu	2	46		1	0			EAS	0	1623153468	0
												0.99599999	0.004			EUR	0.79840%	739107476	5901034.09
												1	0			AFR	0	1237666164	0
												1	0			AMR	0	362456416	0
												1	0			SAS	0	63688069	0
												0.99759495	0.00240507			ExAc Aggregated	0.480436%		
												0.99867254	0.00132743			ESP Cohort	0.265310%		
												0.99349999	0.0065			CSAgilent	1.29578%		
44907908	308	rs370594287		missense	C	?	His	?	3	64			0.001	0.99900001		EAS	0.1999%	1623153468	3244683.78
													0	1		EUR	0	739107476	0
													0	1		AFR	0	1237666164	0
													0	1		AMR	0	362456416	0
													0.001	0.9989995		SAS	0.19990%	63688069	127312.45
													0.0014826	0.99985176		ExAc Aggregated	0.296696%		
											0.00000824			0.99999177		ExAc Aggregated	?????		
44908542	362	rs557845700		missense	T	G	Ile	Met	3	82		0		1		EAS	0	1623153468	0
												0		1		EUR	0	739107476	0
												0		1		AFR	0	1237666164	0
												0.0014		0.99859995		AMR	0.279804%	362456416	1014167.55
												0		1		SAS	0	63688069	0
44908592	412	rs577618688		missense	G	A	Arg	Gln	2	99	1			0		EAS	0	1623153468	0
											1			0		EUR	0	739107476	0
											0.99849999			0.0015		AFR	0.299775%	1237666164	3710213.74
											1			0		AMR	0	362456416	0
											1			0		SAS	0	63688069	0
											0.99997526			0.00002471		ExAc Aggregated	0.004942%		
44908684	504	rs429358	pathoge	missense	C	T	Arg	Cys	1	130		0.91369998	0.0863			EAS	16.5152%	1623153468	268067042
												0.8449001	0.15509999			EUR	28.6140%	739107476	211488213
												0.7321997	0.2678			AFR	46.3883%	1237666164	574132293
												0.89630002	0.1037			AMR	19.6646%	362456416	71275604.4
												0.9131	0.0869			SAS	16.6248%	63688069	10588014.1

Results from Apo E population, part 3

44908690	510	rs11542041		missense	A	T	Ser	Cys	1	132		0	1			EAS	?????	1623153468	
												0	1			EUR	?????	739107476	
												0.0008	0.99919999			AFR	?????	1237666164	
												0	1			AMR	?????	362456416	
												0	1			SAS	?????	63688069	
												0.0000174	0.9999826			ExAc Aggregated	?????		
44908705	525	rs573658040		missense	T	C	Cys	Arg	1	137		0	1			EAS	0	1623153468	0
												0	1			EUR	0	739107476	0
												0	1			AFR	0	1237666164	0
												0	1			AMR	0	362456416	0
												0.001	0.99899995			SAS	0.19990%	63688069	127312.45
												0.0000176	0.99998242			ExAc Aggregated	0.003520%		
44908708	528	rs543363163		missense	A	G	Ser	Gly	1	138	0.001			0.99900001		EAS	0.1999%	1623153468	3244683.78
											0			1		EUR	0	739107476	0
											0			1		AFR	0	1237666164	0
											0			1		AMR	0	362456416	0
											0			1		SAS	0	63688069	0
											0.0000088			0.99999188		ExAc Aggregated	0.001760%		
44908750	570	rs531939919		missense	T	C	Trp	Arg	1	152		0	1			EAS	0	1623153468	0
												0	1			EUR	0	739107476	0
												0	1			AFR	0	1237666164	0
												0	1			AMR	0	362456416	0
												0.001	0.99899995			SAS	0.19990%	63688069	127312.45

References

1. Alzheimer's Association. (2016). Latest Alzheimer's Facts and Figures. Alzheimer's Association. <http://www.alz.org/facts/>.
2. Alzheimer’s Association. (2014). Alzheimer's and Dementia: Global Resources. Alzheimer's Association. <https://alz.org/global/>.
3. Alzheimers.net. (2017). Alzheimer's Statistics. A Place for Mom, Inc. <http://www.alzheimers.net/resources/alzheimers-statistics/>.
4. Alzheimer’s Association. (2017). Fact Sheet. Alzheimer's Association. <http://act.alz.org/site/DocServer/2012_Costs_Fact_Sheet_version_2.pdf?docID=7161>.
5. Alzheimer’s Association. (2017). Fact Sheet: Costs of Alzheimer’s to Medicare and Medicaid. <http://act.alz.org/site/DocServer/2012_Costs_Fact_Sheet_version_2.pdf?docID=7161>.
6. Alzheimer's Association. (2017). Alzheimer's & Brain Research Milestones Research Center. Alzheimer's Association <http://www.alz.org/research/science/major_milestones_in_alzheimers.asp>.
7. Hippius, Hanns, and Gabriele Neundörfer. (2003). The Discovery of Alzheimer’s Disease. Dialogues in Clinical Neuroscience. 5 (1) 101-108.
8. Ryan, Rosser, Fox. (2015). Alzheimer's Disease in the 100 Years since Alzheimer's Death. Brain: A Journal of Neurology. 138 (12) 3816-3821.
9. National Institutes of Health. U.S. Department of Health and Human Services. (2016). Alzheimer's Disease Fact Sheet. National Institutes of Health. U.S.

Department of Health and Human Services. <https://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-fact-sheet>.

10. Alzheimer's Association. (2011). More Brain Changes. Alzheimer's Association. <http://www.alz.org/braintour/healthy_vs_alzheimers.asp>.
11. Querfurth, LaFerla. (2010). Alzheimer’s Disease. N Engl J Med. 362 329-344.
12. BrightFocus Foundation. (2000). Normal vs. Alzheimer's Diseased Brain. <http://www.brightfocus.org/alzheimers/infographic/amyloid-plaques-and-neurofibrillary-tangles>.
13. Zhang, Li, and Toa, Song. (2015). Inflammation in Alzheimer's Disease and Molecular Genetics: Recent Update. Archivum Immunologiae Et Therapiae Experimentalis. 63 (5) 333-344.
14. Chandrasekaran, Sreedevi and Danail Bonchev. (2016). Network Topology Analysis of Post-Mortem Brain Microarrays Identifies More Alzheimer's Related Genes and Micrornas and Points to Novel Routes for Fighting with the Disease. PLOS One 11 (1) e0144052. doi.org/10.1371/journal.pone.0151122
15. Mietelska-Porowska, Anna and Urszula Wojda. (2017). T Lymphocytes and Inflammatory Mediators in the Interplay between Brain and Blood in Alzheimer's Disease: Potential Pools of New Biomarkers. Journal of Immunology Research. 2017 4626540. doi.org/10.1155/2017/4626540
16. Living With Early Onset Alzheimer's Disease. (2015). Cleveland Clinic. <https://my.clevelandclinic.org/health/articles/living-with-early-onset-alzheimers-disease>.

17. Ben-Gedalya, Moll, Bejerano-Sagie, Frere, Cabral, Friedmann-Morvinski, Slutsky, Burstyn-Cohen, and Marini, Cohen. (2015). Alzheimer's Disease-Causing Proline Substitutions Lead to Presenilin 1 Aggregation and Malfunction. The EMBO Journal. 34 (22) 2820-2839.
18. Zhang, Fang-Fang and Jing Li. (2015). Inhibitory Effect of Chloroquine Derivatives on Presenilin 1 and Ubiquilin 1 Expression in Alzheimer's Disease. International Journal of Clinical and Experimental Pathology. 8 (6) 7640-7643.
19. Jayadev, Case, Alajajian, Eastman, and Moller, Garden. (2013). Presenilin 2 Influences Mir146 Level and Activity in Microglia. Journal of Neurochemistry. 127 (5) 592-599.
20. Ture and Alzheimer’s Association. (2017). 2017 Alzheimer’S Disease Facts And Figures. Alzheimer’s Association. 13 325-373. <http://www.alz.org/documents_custom/2017-facts-and-figures.pdf>.
21. Puthiyedth, Riveros, and Beretta, Moscato. (2016). Identification of Differentially Expressed Genes through Integrated Study of Alzheimer's Disease Affected Brain Regions. PLOS One 11 (4) e0152342. doi.org/10.1371/journal.pone.0152342
22. Dolejší, Liraz, Rudajev, Zimcik, and Dolezal, Michaelson. (2016). Apolipoprotein E4 Reduces Evoked Hippocampal Acetylcholine Release in Adult Mice. Journal of Neurochemistry. 136 (3) 503-509.
23. Wozniak, Iparraguirre, Dirks, Deb-Chatterji, Pflugrad, Goldbecker, Tryc, Worthmann, Gess, Crosset, Forton, and Taylor-Robinson, Weissenborn. (2016). Apolipoprotein E-E4 Deficiency and Cognitive Function in Hepatitis C Virus-Infected Patients. Journal of Viral Hepatitis. 23 (1) 39-46.

24. Lee, Ha, Lee, Moon, Chung, and Kim. Mun. (2016). Apolipoprotein E Genotype Modulates Effects of Vitamin B12 and Homocysteine on Grey Matter Volume in Alzheimer's Disease. Psychogeriatrics: The Official Journal of the Japanese Psychogeriatric Society. 16 (1) 3-11
25. The World Bank Group. (2017). Population, total. The World Bank Group. http://data.worldbank.org/region/east-asia-and-pacific.
26. Yourgenome. (2017). What is CRISPR CAS-9. Yourgenome. http://www.yourgenome.org/facts/what-is-crispr-cas9