

# A statistical model that calculates the life time risk of Alzheimer's disease using Bayesian Networks

Seunghyun Yoo<sup>1</sup> and Changwon Yoo<sup>2</sup>

<sup>1</sup>Cypress Bay High School, Weston, FL., <sup>2</sup>Department of Biostatistics, Florida International University, Miami, FL.  
Email: cyoo@fiu.edu

**Abstract:** Alzheimer's is a growing problem within today's society. The risk of Alzheimer increases exponentially with each decade an individual ages, and its fatality rate is hundred percent. Furthermore, no concrete methods of identifying the development of Alzheimer during a patient's lifetime exists, only through post-mortem analysis can we definitely conclude if the victim succumbed to Alzheimer's. Additionally, if the symptoms of Alzheimer's can just be offset by merely five years, trillions of dollars can be saved from healthcare costs (O'Connor 2010).

The disease itself is fairly well documented: it has been shown that Amyloid-  $\beta$  proteins as well as phosphorylated Tau proteins are important in the role of developing Alzheimer's disease (Ihara 1986; Soscia 2010). Recent research provides that the buildup of amyloid-  $\beta$  plaques may be a result of an overactive immune system (Kolata 2010), and that the amyloid-  $\beta$  plaques themselves may even be anti-microbial particles that build up after being over-exposed (Soscia 2010). However, despite the research that has been done on Alzheimer's, little funding is going into researching a cure, or a way to lessen the impact of Alzheimer's on our aging population. Given the fact that every penny the National Institutes of Health spends on Alzheimer research, healthcare providers spend \$3.50 caring for Alzheimer's patients, further coupled with the fact that every second, a baby-boomer reaches his 65th birthday, Alzheimer's is a problem that must be addressed immediately (O'Connor 2010).

In order to address the problem, we have used Bayesian networks to identify which genes play an important role in Alzheimer's disease, and to illustrate the interactions among relevant genes. We have identified six human gene expression case control studies using microarray data and found 11 relevant genes that are related to Alzheimer's disease, four of which are strongly associated with Alzheimer's: TGFB111, LTF, TLX2, and LTB4R. We then analyzed this data, to identify pertinent interactions between the 11 genes.

Using the data collected, we further calculated the lifetime risk, as well as the odd ratio of developing Alzheimer's disease, given the expression levels of combinations of genes. Females (aged 65+) with over-expressed TGFB111 and LTF, combined with under-expressed TLX2 and LTB4R showed a heightened 61% remaining lifetime risk of developing Alzheimer's disease, while males (aged 65+) with the same expression patterns of the genes showed a heightened 44% remaining lifetime risk of developing Alzheimer's disease. Furthermore, odd of developing Alzheimer's disease for an individual with normally expressed LTF and TLX2, under-expressed TGFB111, and over-expressed LTB4R is 25 fold higher than that of an individual with the same expression patterns of the genes except for an over-expressed TLX2. Given these findings we can help create effective methods of diagnosing patients with Alzheimer's disease and ultimately help create effective treatments for Alzheimer's patients.

**Keywords:** *Bayesian Networks, Alzheimer Lifetime Risk, key genes of Alzheimer's Disease*

## 1. INTRODUCTION

Alzheimer's is an emerging disease that can have serious consequences if left unchecked. Already, we see that Alzheimer's carries a 100% fatality rate, and can easily be confused with dementia and other neurodegenerative illnesses. More worrisome is the lack of effective diagnosis of Alzheimer's, with many doctors being forced to do "trial-and-error" experiments to try to rule out Alzheimer's. Even more alarming is the fact that in the U.S. alone, every second a baby-boomer reaches their 65th birthday (O'Connor 2010). The implications of which mean there will be an increased burden on the working younger class, who will have to support the fast growing elderly population. For many, a cure would be preferred, however, even if it was possible to offset the emergence of Alzheimer's symptoms in patients for 5 years, trillions of dollars can be saved, lifting some weight off the already burdened American economy (O'Connor 2010).

Existing research already provides why the symptoms of Alzheimer's manifests in its patients. Abnormally high amyloid-  $\beta$  protein levels, as well as abnormally high levels of phosphorylated Tau proteins lead to plaque buildup in the brain, causing the neurodegenerative illness to appear (Selkoe 2001). Recent research also suggests that the reason the plaques are appearing in the brain, is due to an overactive immune system, where the amyloid-  $\beta$  proteins serve as anti-microbial units (Soscia 2010), that end up combining into plaques (Kolata 2010). In addition to the amyloid-  $\beta$  proteins, Tau proteins have also been observed in Alzheimer's patients. There are three versions of Tau proteins, however they can be grouped into two separate states, phosphorylated and non-phosphorylated. A non-phosphorylated Tau protein is important in maintaining the integrity of the brain structure by reinforcing microtubule molecule (Kirschner 1986). However, a phosphorylated Tau protein ends up disrupting microtubule function, causing neurodegenerative symptoms to appear, hastening the onset of Alzheimer's.

In order to properly diagnose and treat Alzheimer's disease, we have to look towards both the Tau and Amyloid proteins as a whole. It has been shown that for Alzheimer's disease to take hold of a patient, both the Tau Proteins, and Amyloid plaques must exist within the brain (Selkoe 2001). Given that information, it is unlikely only one gene controls the expression of Alzheimer's. Tanzi and Bertram (McGreevey 2008), have hypothesized that there are four genes that all may contribute to plaque buildup, and phosphorylated Tau proteins. By looking at various gene expression case control studies, we can identify which genes are either activated or suppressed in Alzheimer's patients, as compared to normal individuals. Further using that information, we can create a statistical model that calculates lifetime risk, as well as the risk of developing Alzheimer's during an individual's lifetime.

## 2. METHODS AND PROCEDURES

We used Bayesian networks to find the key genes in developing Alzheimer's Disease and their possible interactions.

### 2.1 Bayesian Networks

The Bayesian network is represented as a directed acyclic graph where each arc is identified as a direct (causal) influence between the parent node and a child node in relation to all other nodes in a network (Pearl 1998). Figure 1 shows the network that represents the interactions among Alzheimer key genes.

A typical Bayesian network consists of a structure (such as that shown in Figure 1) and a set of probabilities that parameterize the structure (not shown). In general, for each node, there is a conditional probability of that node, given the states of its direct parent. For example, in Figure 1, the probability associated with Alzheimer is  $P(\text{Alzheimer} \mid \text{TGFB11})$ . That is, we give the probability distribution over the values of Alzheimer, which are conditioned by the possible values of TGFB11. For other variables that contain little direct relationships in the network, a prior probability is given. The Markov condition (Spirtes, Glymour et al. 2000) specifies the conditional independence relationships that are shown in a Bayesian network: *Let X and Y be variables. Y is neither a direct nor indirect effect of X. Then X is independent of Y, conditioned on any state of the direct causes of X.* The causal Markov condition permits the joint distribution of the  $n$  variables in a causal Bayesian network to be shown as follows (Pearl 1988)

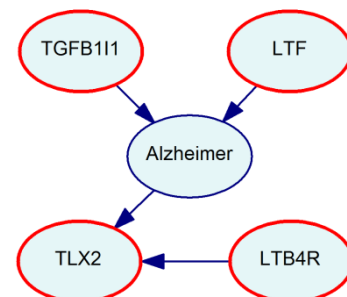
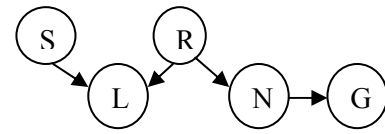


Figure 1 LTF and TGFB111 are related to TLX2 only through Alzheimer

$$P(x_1, x_2, \dots, x_n | K) = \prod_{i=1}^n P(x_i | \pi_i, K) \tag{1}$$

where  $x_i$  denotes a state of the variable  $X_i$ ,  $\pi_i$  denotes a joint state of the parents of  $X_i$  and  $K$  denotes background knowledge.

Also, from a Bayesian network structure, we can learn interactions among the model variables by using arcs orientations (*converging*, *diverging* and *serial*) (Charniak 1991). For example, in the network shown at right, either it rains (R) or you turn on the sprinkler (S), it makes your lawn wet (L) (*converging arcs* that are going into L). Note that, however, if you know your lawn is wet (L) and if there is a low chance that it rained (R) then there is a high chance that your sprinkler was on (S), in other words, *if converging arcs into L, then R and S becomes dependant given L*. Also note that if it rains (R), your lawn (L) and your neighbor lawn get wet (N) (*diverging arcs* that are originated from R). If you know it rained (R), knowing your lawn is wet will not tell you much more about your neighbor lawn being wet (N), in other words, *if diverging arcs from R, then L and N becomes independent given R*. If it rains (R), then your neighbor lawn gets wet (N), and eventually, your neighbor lawn gets green (G) (*serial arcs* that are from R to N to G). In this case, if you know your neighbor lawn is wet (N), then knowing whether it rained (R) or not will not tell you much more about your neighbor lawn getting green (G), in other words, *if serial arcs through G, then R and G is independent given N*.



### 2.2 Dataset Preparation

We collected the data from Gene Expression Omnibus (GEO) (NCBI 2011), which is a database that contains microarray data performed by researches across the globe. We first identified case-control gene expression studies with human subjects in GEO. The studies that were identified are shown in Table 1.

Table 1. GEO Data Studies, with patient size and total patients per study

Study Name	Author	Study Size (Number of Patients)	Number of Diseased Patients	Number of Control Subjects	Platform
Alzheimer's Disease at various stages of Severity (Blalock 2004)	Eric M. Blalock	31	22	9	GPL96
Expression of MRNA's Regulating Synaptic Function in Incipient Alzheimer's Disease (Williams 2008)	Celia Williams	14	6	8	GPL96
mRNA and miRNA expression in parietal lobe cortex in Alzheimer's Disease (Nunez-Iglesias 2009)	Juan Nunez-Iglesias	8	4	4	GPL570
Alzheimer's Disease: Neurofibrillary Tangles (Hamill 2006)	Brandy Hamill	20	10	10	GPL570
Alzheimer's Disease Peripheral Blood Mononuclear Cell Expression (Maes 2006)	Oliver Charles Maes	28	14	14	GPL1211
Transcriptomes in Peripheral Blood Mononuclear Cells of Alzheimer's Patients (Chen 2009)	Kuang-Den Chen	9	6	3	GPL570
	TOTAL PATIENTS	110	62	48	

For each study, we calculated the average expression levels of each gene for both Alzheimer and control subjects. Then we sorted the average gene expression levels from lowest to highest with regards to the control subjects. We split the sorted genes into two groups according to the average gene expression levels: high-expression genes, and low-expression genes. In order to find the high-expression genes, and low-expression genes, we first identify the line of best fit. From the median gene expression value, we then split the genes using the line of best fit, giving us high-expression genes (genes lying above the line) and low-expression genes (genes lying below the line). From there we calculated a percent difference of the average gene expression levels for the high-expression genes, and low-expression genes by using a percent change formula:

$$\Delta A = \frac{A_d - A_c}{A_d} \times 100 \tag{2}$$

Where  $A_d$  represents the averages of gene expression levels in Alzheimer patients, and  $A_c$  represents the averages of gene expression levels in the control patients. The resulting value  $\Delta A$  is the percent change of gene expression from Alzheimer to control patients. By doing so we were able to find which genes had the largest expression level discrepancies between the normal and diseased individuals. The percent change of gene expression is then sorted from the smallest to the largest per group, giving us the highest percent change

of gene expression from genes with low expression and from genes with high expression. In the datasets we collected, low-expression genes with smaller values had higher percent change of gene expression than high-expression genes with larger expression levels. For each study, we selected equal number of genes from high-expression and low-expression groups, approximately 5,000 genes from each and then searched for the common recurring genes among the six studies. The end product left us with 131 high percent difference genes that were common to all six studies. We then discretized each gene's expression level into three states; under-expressed (Low), normal expression (NoChange), and over-expressed (High). With the discretized data file, we looked at the interactions between the 131 genes with Bayesian networks.

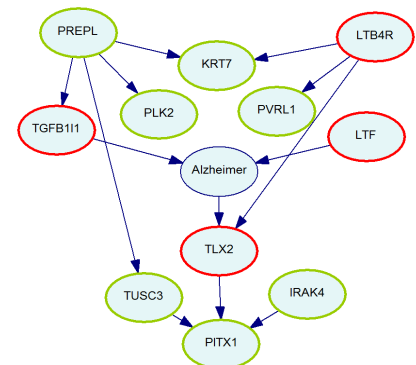
### 2.3 Analysis

We used a program called Bayesian Network Inference with Java Objects (Banjo) (Hartemink 2006) to help analyze the 131 genes. We ran our data through Banjo a total of 12 runs on 3 different machines. Each machine ran Banjo for three hours, six hours, nine hours, and 12 hours. We then collected the most reliable results, and put our results into the Graphical Network Interface (GeNIe) allowing us to look into the interactions among the genes that were found in Alzheimer's studies.

### 3. RESULTS

Based on the highest scoring global Bayesian network (shown in Appendix A), out of 131 genes collected, 108 genes were associated with Alzheimer's Disease. From the Bayesian network, we identified four genes that are Markov Blanket (MB)<sup>1</sup> of Alzheimer's: i.e., TGFB111, LTF, TLX2, and LTB4R (genes marked with red in Figure 2; we refer these four genes as "Alzheimer MB genes").

We further report an additional seven genes, KRT7, PVRL1, PREPL, TUSC3, PITX1, and IRAK4 that are Markov Blanket of Alzheimer MB genes. Additionally we created a local network (from the highest scoring global Bayesian network) with the 11 key genes related to Alzheimer's, shown in Figure 2.



**Figure 2.** Local Network showing genes of Markov Blanket of Alzheimer (Red) with additional genes that are Markov Blanket of genes in red (Green)

As discussed in Section 2.1 (Charniak 1991), from Figure 2, we can infer how other genes are interacting in changing the risk of Alzheimer. For example, LTB4R is independent of Alzheimer, unless we know anything about KRT7 or TLX2. Also, TGFB111 and LTF are associated with LTB4R only if we know the state of TLX2.

In order to find all the possible conditions from these 11 genes – each with three expression states, i.e., Under-Expressed (Low), Normal Expression (NoChange), and Over-Expressed (High) – for calculating conditional probabilities of Alzheimer, we generated a C++ code using SMILE (DecisionSystemsLaboratory 2007) that calculates all the conditional probabilities. The end result gave 4,194,304 (= 4<sup>11</sup>) unique gene conditions among the 11 key genes. From those 4 million conditions, we searched for the conditions of genes that yielded the highest conditional probability of Alzheimer. According to a study done by the Boston University School of Medicine (Seshadri 1997), the lifetime risk of 65 and above male and female are 0.06 and 0.12 respectively. Note that all the conditional probabilities of Alzheimer are calculated using the following Bayes' rule to report the increment of lifetime risk for male and female compare to the baseline:

$$P(\text{Alzheimer} | \text{gene conditions}) = \frac{P(\text{gene conditions} | \text{Alzheimer}) \cdot P(\text{Alzheimer})}{P(\text{gene conditions})} \quad (3)$$

where we used  $P(\text{Alzheimer})$  as 0.06 for male and 0.12 for female. All the other probabilities were calculated from SMILE.

<sup>1</sup> The Markov Blanket of the Alzheimer node, i.e., the parents, children, and parents of children of the Alzheimer node. Nodes in Markov Blanket of the Alzheimer node completely determines Alzheimer node's outcome, i.e., if we know expression level of LTB4R, TGFB111, TLX2, and LTF, then knowing other gene's expression level will not change the probability of a subject having Alzheimer's disease.

In Table 2, we report the odd ratios of three genes, TLX2, TGFB111, and LTF. Of the six odd ratios that were calculated, the top three odd ratios show that the odds of a person with a normally expressed TGFB111 (NoChange) to a person with under-expressed TGFB111 (Low) of developing Alzheimer's is almost six-fold. Furthermore, if a person has a normally expressed TLX2 (NoChange) compared to a person with an over-expressed TLX2 (High) the chance of developing Alzheimer is 3.35 times as likely. Moreover, if a person has a normally expressed LTF (NoChange) compared to a person with an under-expressed LTF (Low) the chance of developing Alzheimer's is approximately 2.8 times as likely.

Table 2. One Gene odd ratio with two tailed confidence intervals

Condition	Reference Condition	Odd Ratio	95% Confidence Interval
TGFB111=NoChange	TGFB111=Low	6.533	$2.912 \leq X \leq 14.661$
TLX2=NoChange	TLX2=High	3.351	$1.607 \leq X \leq 6.988$
LTF=NoChange	LTF=Low	2.821	$1.365 \leq X \leq 5.834$
TLX2=Low	TLX2=NoChange	2.214	$1.040 \leq X \leq 4.714$
TGFB111=High	TGFB111=NoChange	1.697	$0.811 \leq X \leq 3.551$
LTF=High	LTF=NoChange	1.215	$0.601 \leq X \leq 2.458$

We also calculated highest odd ratios for the conditions for four genes, three genes, as well as two genes combinations (Table 3). The highest odd ratio of the four genes combinations showed that when an individual who has normally expressed (NoChange) LTF and TLX2, over-expressed (High) LTB4R, as well as under-expressed (Low) TGFB111, showed 25 times more odd of developing Alzheimer's compared to a person who has the same gene expression levels except TLX2 being High. Additionally the highest odd ratio of three genes combinations presented when comparing and individual who has a normally expressed LTF (NoChange) and over-expressed TLX2 and TGFB111(High) with an individual that has and the same gene expression levels except TGFB111 being Low, then the odds of developing Alzheimer's reach close to 23 times higher. The highest odd ratio of two genes combinations were identified when comparing and individual who has under-expressed (Low) TLX2 and LTB4R with an individual that has and the same gene expression levels except TLX2 being High, then the odds of developing Alzheimer's is 11 times higher.

Table 3. Four, three, and two gene odd ratios with two tailed confidence intervals

Condition	Reference Condition	Odd Ratio	95% Confidence Interval
TGFB111=Low LTF=NoChange TLX2=NoChange LTB4R=High	TGFB111=Low LTF=NoChange TLX2=High LTB4R=High	25.17	$3.774 \leq X \leq 167.911$
LTF=NoChange TGFB111=High TLX2=High	LTF=NoChange TGFB111=Low TLX2=High	23	$5.598 \leq X \leq 94.5$
LTF=High TGFB111=Low TLX2=High	LTF=NoChange TGFB111=Low TLX2=High	21	$5.107 \leq X \leq 86.353$
TGFB111=High TLX2=Low LTB4R=Low	TGFB111=High TLX2=High LTB4R=Low	11.11	$3.927 \leq X \leq 31.439$
TGFB111=High LTF=High TLX2=Low LTB4R=Low	TGFB111=High LTF=High TLX2=High LTB4R=Low	11.11	$3.663 \leq X \leq 33.708$
TLX2=Low LTB4R=Low	TLX2=High LTB4R=Low	11.11	$4.672 \leq X \leq 26.424$
TGFB111=High TLX2=Low LTB4R=Low	TGFB111=Low TLX2=Low LTB4R=Low	11.09	$3.919 \leq X \leq 31.376$
TGFB111=High TLX2=Low LTB4R=Low	TGFB111=Low TLX2=Low LTB4R=Low	11.09	$3.919 \leq X \leq 31.376$
TGFB111=High TLX2=High	TGFB111=Low TLX2=High	11.09	$3.748 \leq X \leq 32.805$

Table 4 shows the highest lifetime risk calculated by a single gene, two genes, three genes, and four genes, as well as the lowest lifetime risk calculated by single gene, two genes, three genes, and four genes. Note that because of the property of Markov Blanket, no genes more than the four Alzheimer MB genes will produce higher lifetime risk. From there, we can see that if a male (aged 65 and older) has an over-expressed TGFB111 and LTF, with an under-expressed TLX2 and LTB4R, the probability of developing Alzheimer's is close to 0.44 (more than 7 times than baseline of 0.06 found by Seshadri in 1997), whereas a female (aged 65 and older) with the same genetic expression has 0.66 chance of developing Alzheimer's (more than 5 times than baseline of 0.12, also found by Seshadri in 1997). Inversely we can see that the probability of developing Alzheimer's for a male (aged 65 and older) with a normally expressed LTF and TUSC3 (NoChange), an under-expressed (Low) TGFB111, and an over-expressed (High) KRT7, TLX2, PITX1, PVRL1, and IRAK4 is approximately 0.00104 (less than 0.02 times the baseline of 0.06), whereas a female (aged 65 and older) with the same genetic expression has the probability of approximately 0.0021 (which is also less than 0.02 times the baseline of 0.12).

4. DISCUSSION

Using the data we have gathered from GEO studies, the analysis we performed with the resulting Bayesian networks provides significant insight into the genetic components of Alzheimer's Disease. We identified 11 key genes, as well as various other gene combinations among the 11 genes that appear to be strongly linked to Alzheimer's Disease. Of the 11 key genes, the four Alzheimer MB genes showed close relationships to Alzheimer's. TLX2, one of the Alzheimer MB genes, is a gene that shows promise in treating neurodegenerative diseases (Bernardo 2006). TLX2, known by its alias, NCX has shown to have strong anti-amyloidogenic properties, which may aid in the clearing, or forestalling of amyloid-β proteins in the brain (a signature earmark of Alzheimer's disease). LTF, another Markov gene, appears to be "of importance in assessment of the Alzheimer's disease pathology" (Berdel 1994). LTF is known to be associated with anti-microbial activity, as well as anti-inflammatory activity, which may have importance in the development of Alzheimer's (Kolata 2010). TGFB11, which codes for a protein called Hic-5 also appears to have connections to Alzheimer's (Caltagaron 2010). TGFB11's encoded protein, Hic-5 is thought to regulate androgen receptor activity, with links to regulating cell growth, proliferation, migration and differentiation (UniProt 2011). LTB4R, currently does not have any supported scientific records tying it to Alzheimer's disease, further requiring study will provide a definite analysis on LTB4R, and whether or not it may be pertinent in the development of the disease.

Our study was designed to try to identify certain genetic interactions that might intermediate the development of Alzheimer's disease. However, the study was not able to completely encompass

all the genetic components of Alzheimer's, due to incomplete data provided through GEO. Since all of the GEO studies did not provide additional important variables, e.g., gender, age, ethnicity, and other basic information, we calculated two different lifetime risks, female (aged 65+) and male (aged 65+), using information provided by the National Institutes of Health (NIH). Furthermore, we had intended to perform a cross-database search with Entrez to see which genes from our Banjo study would overlap with existing Alzheimer's genes in Entrez. The result gave us similar genes, or genes in the same family as one another (i.e. TGFB11 v. TGFB1) but not exactly case-by-case similar genes. Despite being only similar genes, with

Table 4. Lifetime Risk Calculation given Gene conditions

Condition	Lifetime risk (Female)	Lifetime risk (Male)	Condition	Lifetime risk (Female)	Lifetime risk (Male)
TGFB11=High LTF=High TLX2=Low LTB4R=Low	0.6134	0.4389	TGFB11=Low LTF=NC TLX2=High LTB4R=High	2.088xE-03	1.031xE-03
TGFB11=High LTF=High TLX2=Low PREPL=Low KRT7=Low PVRL1=Low	0.5865	0.4116	TGFB11=Low LTF=NC TLX2=High PREPL=NC KRT7=High PVRL1=High	2.092xE-03	1.033xE-03
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PVRL1=Low TUSC3=Low	0.5846	0.3971	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PVRL1=High TUSC3=NC	2.098xE-03	9.83xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PVRL1=Low TUSC3=NC	0.5829	0.3954	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PITX1=NC PVRL1=High	2.098xE-03	9.831xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PITX1=Low PVRL1=Low	0.5824	0.395	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PVRL1=High	2.098xE-03	9.8322xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PITX1=NC PVRL1=Low IRAK4=NC	0.5822	0.3948	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PVRL1=High TUSC3=Low	2.098xE-03	9.833xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PITX1=NC PVRL1=Low	0.5822	0.3948	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PITX1=Low PVRL1=High IRAK4=High	2.098xE-03	9.833xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PVRL1=Low	0.5821	0.3947	LTF=NC PLK2=NC TGFB11=Low KRT7=High TLX2=High PITX1=High PVRL1=High	2.098xE-03	9.833xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PITX1=High PVRL1=Low	0.5818	0.3943	LTF=NC PLK2=NC TGFB11=Low KRT7=Low TLX2=High PITX1=Low PVRL1=High IRAK4=Low	2.099xE-03	9.834xE-04
LTF=High PLK2=Low TGFB11=High KRT7=Low TLX2=Low PITX1=High PVRL1=Low IRAK4=NC	0.5815	0.3941	LTF=NC TGFB11=Low KRT7=High TLX2=High PITX1=High PVRL1=High TUSC3=NC IRAK4=High	2.106xE-03	1.039xE-03
TGFB11=High TLX2=Low LTB4R=Low	0.5682	0.3812	LTF=NC TGFB11=Low TLX2=High	4.091xE-03	1.919xE-03
TLX2=Low LTB4R=Low	0.3937	0.2331	TGFB11=Low TLX2=High	8.227xE-03	3.868xE-03
TLX2=Low	0.2519	0.1361	TGFB11=Low	0.02431	0.0115
TGFB11=High	0.2165	0.1145	TLX2=High	0.04341	0.0208
LTF=High	0.1445	0.0733	LTF=Low	0.04696	0.0225

further research on the similar genes, in the future we can hope to combine the existing Entrez database, with our GEO studies to provide a full picture on the genetic component of Alzheimer's Disease.

We can also try to combine our study, with the existing databases, such as Entrez, on Alzheimer's Disease, in order to help provide more insight into Alzheimer's Disease. In addition, we intend on collaborating with existing Alzheimer's Research Centers in order to strengthen our existing data. We will also further look into the relationships between the Entrez genes that were found to be similar to the genes identified from the analysis presented here to identify which genes play a crucial role in the development of Alzheimer's.

## REFERENCES

- Berdel (1994). "Variability of the Parahippocampal surface of the transverse fissure in the human brain." 77-83.
- Bernardo (2006). Dynamic regulation of microglial functions by the non-steroidal anti-inflammatory drug NCX 2216: implications for chronic treatments of neurodegenerative diseases, National Institutes of Health: 25-32.
- Blalock, E. M. (2004) "Incipient Alzheimer's Disease: Microarray Correlation Analyses."
- Caltagarone, H., Murdoch, Jinq, DeFranco, Bowser (2010). Paxillin and hydrogen peroxide-inducible clone 5 expression and distribution in control and Alzheimer disease hippocampi: 356-371.
- Charniak, E. (1991). Bayesian Network without Tears. AI Magazine, American Association for Artificial Intelligence: 3.
- Chen, K.-D. (2009) "Transcriptomes in Peripheral Blood Mononuclear cells of Dementia and Alzheimer Patients." DecisionSystemsLaboratory. (2007, 2010). "Structural Modeling, Inference, and Learning Engine." from <http://genie.sis.pitt.edu/>.
- Hamill, B. (2006) "Alzheimer's disease: neurofibrillary tangles."
- Hartemink, A. J. (2006). Bayesian Network Inference with Java Objects. Durham, Duke University: Banjo is a software application and framework for structure learning of static and dynamic Bayesian networks, developed under the direction of Alexander J. Hartemink in the Department of Computer Science at Duke University. Banjo was designed from the ground up to provide efficient structure inference when analyzing large, research-oriented data sets, while at the same time being accessible enough for students and researchers to explore and experiment with the algorithms. Because it is implemented in Java, the framework is easy to maintain and extend. .
- Ihara, Y. (1986). "Phosphorylated Tau Protein is Integrated into Paired Helical Filaments in Alzheimer's Disease." The Journal of Biochemistry **99**(6): 3.
- Kirschner, D. G. D. a. M. W. (1986). "Tau Protein Function in Living Cells." Journal of Cell Biology **103**(6): 7.
- Kolata, G. (2010) "Infection Defense may Spur Alzheimer's." New York Times.
- Kolata, G. (2010). Infection Defense May Supr Alzheimer's. New York Times. New York.
- Maes, O. C. (2006) "Alzheimer Disease peripheral blood mononuclear cell expression."
- McGreevey, S. (2008). Gene scan of Alzheimer's families identifies four new suspect genes. Harvard Gazette. Harvard.
- NCBI (2011). Gene Expression Omnibus.
- Nunez-Iglesias, J. (2009) "mRNA and miRNA expression in parietal lobe cortex in Alzheimer's disease."
- O'Connor, S. D. (2010). The Age of Alzheimer's. New York Times. New York.
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. San Mateo, CA, Morgan Kaufmann.
- Selkoe, D. J. (2001). "Alzheimer's Disease: Genes, Proteins, and Therapy." Physiological Reviews **81** (2): 741-745.
- Seshadri (1997). Lifetime risk of dementia and Alzheimer's disease. The impact of mortality on risk estimates in the Framingham Study: 1498-1504.
- Soscia, K., Washicosky, Tucker, Ingelsson, Hyman, Burton, Goldstein, Duong, Tanzi, Moir (2010). The Alzheimer's disease-associated amyloid beta protein is an anti-microbial peptide.
- Spirtes, P., C. Glymour, et al. (2000). Causation, prediction, and search. Cambridge, MA, MIT Press.
- UniProt (2011). "Transforming growth factor beta 1-induced transcript 1 protein."
- Williams, C. (2008) "Expression of mRNA Regulating Synaptic Function and Neuroplasticity in Incipient AD."