Genetic Commonalities of Autism Spectrum Disorder and Schizophrenia

By:

Amil A. Font Valentin

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTERS OF SCIENCE
in
INDUSTRIAL ENGINEERING

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

2019

Approved by:

_____	_____
Mauricio Cabrera-Rios, Ph.D. President, Graduate Committee	Date
_____	_____
Mayra Méndez, Ph.D. Member, Graduate Committee	Date
_____	_____
Jaime Seguel, Ph.D. Member, Graduate Committee	Date
_____	_____
Gloribell Ortiz, Dr PH Representative of Graduate School	Date
_____	_____
Viviana I. Cesaní Vázquez, Ph.D. Department Chairperson	Date

Abstract of Thesis Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Masters of Science

Genetic Commonalities of Autism Spectrum Disorder and Schizophrenia

By
Amil A. Font Valentin

May 2019

Chair: Dr. Mauricio Cabrera-Rios
Department: Industrial Engineering Department

The characterization of a disease through the detection of differentially expressed genes is of interest for early detection, diagnosis, prognosis, and prediction. Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ) have been correlated in literature for sharing similar symptoms and behaviors, however, not at the genetic level. This work aims to correlate both conditions through the analysis of microarray datasets. Multiple Criteria Optimization was used to find the most significant gene expression changes common to both conditions. Thirty-six genes resulted from this first step. Subsequently, the most correlated structures were identified through the Traveling Salesman Problem and the Minimum Spanning Tree. Mathematical optimization, as a driver for gene selection and structuring, defines an analysis point of view advocated by our group that is different to biostatistics and bioinformatics. ASD and SCZ are analyzed simultaneously for the first time in this work owing to the capabilities of our group's strategy.

Acknowledgments

I want to thank God for giving me the strength to finish this stage of my life as a graduate student. I also want to thank my parents for always supporting me in my short- and long-term goals.

I would like to thank Dr. Mauricio Cabrera because he has guided me throughout this process as a graduate student. Not only that, but also for giving me the tools to develop new skills as a professional and to accomplish one of my goals. I also want to thank Dr. Clara Isaza for guiding through the biology in this work. She helped me learn a lot from this area and that in turn helped me understand my research problem better.

I want to thank my friends who were with me and supported me in this stage of my life. Yesenia Perez, Karen López, Christian Hernández, and Elizabeth Ayala: thank you for being with me in the good and bad moments of this stage.

Also, I want to thank the members of our research group, *The Applied Optimization Group*, for all of their help. I want to thank Yazeli Cruz for helping me in the training process. I also want to thank Diamarys Salomé and Alejandro Marrero for their support in gathering and organizing biological information.

Contents

Table of Figures

Chapter 1 Introduction

1.1 Introduction

This thesis proposes the study of genetic commonalities between two neurological disorders: Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ). Autism Spectrum Disorder (ASD) is a series of neurodevelopmental ailments that can be acquired during childhood. These are characterized by difficulties in social and communication skills and repetitive behavior, noticeable since the first two years of age. ASD's main causes and development are not entirely characterized to date. There could be an explanation that includes either genetic factors or environmental factors, or a combination between them [1]. Another disorder that is similar in symptoms and behavior similarities is Schizophrenia (SCZ). This disease is a non-common chronic neurological brain disorder that tends to alter a person's reality. Its symptoms usually appear between 16 and 30 years of age, but rarely during childhood. Symptoms in SCZ are classified into three categories: positive, negative and cognitive [2]. Positive symptoms (psychotic symptoms) consist of hallucinations, delusions, and formal thought disorder. Negative symptoms relate to social isolation, loss of motivation, difficulty to express emotions, and lack of speech [3]. Cognitive symptoms maintain the person disabled and lead to poor functional outcomes, such as poor concentration [4]. The latter is related to the loss of contact with reality, the lack of ability to function normally and memory and thinking aspects [5]. SCZ's main causes are not characterized to date. It has been noted out that a combination of physical, genetic, psychological and environmental factors are the possible causes of the development of this disease.

While these two disorders present themselves in different life stages, they manifest in behaviors that can be deemed similar. The lack of communications skills, uncommon thinking

and unusual behaviors in ASD, for example, can be similar to negative symptoms and disorganized thought and behavior in SCZ. These characteristics' overlap produces a challenge for doctors in distinguishing between ASD from SCZ to avoid misdiagnosing patients. A study by K. Takara et al. [6] pinpoints that symptoms like unusual fears, thought disorder and unusual anxiety reaction are linked to psychosis in ASD but are also present in SCZ. Woodbury-Smith et al. [7], describe a case where a 24 year- old man was diagnosed with schizophrenia based on persecutory ideation and ten years later was referred and diagnosed with ASD. Studies indicate that ASD in adults is unnoticed or misdiagnosed with other mental disorder. It is imperative to have a record of the childhood history of this patient [8]. Therefore, it is essential to characterize better these two diseases, including their genetic commonalities.

The genesis of this study derived from our previous microarray study based (ASD database GSE22507) on multiple criteria optimization (MCO), where the ASD genes were identified and analyzed, and in which twenty-seven were found. Two of those twenty-seven genes (IFITM2 and IFITM3) are related to SCZ [9]. After obtaining these results, it was proposed to analyze these conditions simultaneously, because if we analyze them individually, the results may be different. However, when performing a simultaneous analysis between them, the outcome would be the genes in common that change their expressions the most.

The analysis of Microarray database provides reliable input to this study. A Microarray database contains relative expression levels of a large number of genes in a tissue and blood samples as compared to a control. This information has been used to find the differentially expressed genes in a specific condition [10]. In literature, it has been found the use of linear regression, general linear effect model, analysis of variance, and statistical packages to identify the differentially expressed genes in a disease [11] [12] [13] [14]. These statistical methods have

in common the requirement of data normalization, parameter estimations, and corrections for the microarray database analysis. On the other hand, our group has utilized methods based on mathematical optimization such as the Multiple Criteria Optimization (MCO), Minimum Spanning Tree (MST) and Traveling Salesman Problem (TSP) that facilitate the microarray database analysis without parameter and data manipulation. In this way, it offers convergence to crucial pieces of information that can be correlated to find genetic commonalities between ailments.

This work aims to use optimization methods previously proposed by our group to find how genetically ASD and SCZ are related. This contribution could lead to findings to enable early detection, diagnosis, prognosis, and prediction of these ailments. This work discusses the issue of finding highly differentiated expressed genes between two diseases simultaneously using the Multiple Criteria Optimization in a list of genes that change their expression the most between the condition samples and the controls. Furthermore, it also discusses finding the most correlated cyclic and noncyclic structure between genes that show their interaction. The Traveling Salesman Problem (TSP) and the Minimum Spanning Tree (MST) were used to find the most correlated structure between genes, as previously proposed by our group.

1.2 Objective

The main objective of this work is to characterize the relation between Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ). Identifying the genetic commonalities between diseases with an operations research approach previously proposed by our group. Beginning with the analysis of both diseases together to identify the differentially expressed genes between ailments with the use of the multiple criteria optimization (MCO). After MCO we continued to build the most correlated structure among the genes that change their expression the most. The cyclic and non-cyclic structures of genes network were obtained using the traveling salesman

problem (TSP) and the minimum spanning tree (MST) respectively. In addition, a cost model for ASD and SCZ to estimate annual expenses for each patient.

1.3 Contribution

Our group had proposed the use of a mathematical optimization model to identify the differentially expressed genes using microarray databases. First Sanchez-Peña [15], proposed to identify the differentially expressed genes from microarray database as an MCO problem with two P-values as performances measures and solved with Data Enveloped Analysis (DEA). Later, Camacho-Cáceres [16] another member of our group, proposed to analyze the databases as an MCO problem with two performance measures in this case with the mean and the median. The MCO was solved applying the Pareto optimality conditions to find the highly differentiated expressed genes following our research group [16].

As part of a continuation of the work of Camacho-Cáceres [16], J. Rosas's work [17] used the Minimum Spanning Tree (MST) to find the relationship between the differentially expressed genes in Alzheimer's disease. Afterwards, Y. Santiago [18] proposed the use of Rank Order Clustering Algorithm to explore more about genes and MiRNAs in Alzheimer. Next N. Ortiz's [19] research used the MCO method and Combinatorial Optimization (include Traveling Salesman Problem (TSP) and MST) to detect differentially expressed genes and signaling structures in common to Alzheimer, Parkinson and Huntington diseases. The analysis was done for each disease at a time while our proposed study analyzes them simultaneously.

This study contributes to analyzing two conditions at the same time to find genetic similarities between them that are not fully understood currently using methods previously proposed by our group. MCO is used to identify the differentially expressed genes among diseases.

The use of TSP and the MST is used to find a cyclic and non-cyclic relation between genes. The principal contribution in the present thesis is the simultaneous microarray analysis obtaining genes that change their expression the most among ASD and SCZ. This work aims to integrate the study of genomics using the relative gene expression obtained in microarray databases. Also, the evaluation of the cost associated to ASD and SCZ through the creation of a cost model and analysis.

Chapter 2 Literature Review/Background

2.1 Literature Review

L.F. Ning et al. [20] proposed a meta-analysis to identify the differentially expressed genes in ASD. To perform the meta-analysis the Gene Expression Omnibus (GEO) dataset was used for ASD. The methodology started with data normalization due to the databases coming from different platforms and different genes nomenclature. To identify the differentially expressed genes the Student t-test statistical value and the failure discovery rate (FDR) was used for the analysis. Finally, Gene Ontology (GO) enrichment analysis was used for the differential expressed genes significance. In this research, they identified a total of 10 expression profiling studies (microarray databases), which included 364 condition and 248 control samples. The value used for the FDR is less than 0.05, and minimal 2-fold changes, 3,105 differentially expressed genes were found between condition and control. Also, the authors identified 7 genes associated with PLA2 that have been related to the PLA2 in red blood cells with schizophrenia [21], and autism [22]

The study described in [20] was similar to the research project proposed in this document because it uses several microarray databases from tissue and blood samples to identify differentially expressed genes in ASD. The differences between this research and the proposed one are that we use a mathematical optimization model that does not require parameter adjustments by the user. Therefore, a more objective set of genes is obtained. The work described previously had limitations due to the heterogeneity and confounding factors of the databases that could lead to distorting their analysis.

The research proposed by S. E. Ellis [23] studies the genetic relationship between ASD, SCZ and bipolar disorder (BPD). ASD database is composed of tissue samples from 40 control

and 32 ASD subjects. Otherwise, SCZ and BPD database are collected from 31 SCZ, 25 BPD and 26 controls subjects from a tissue sample. The data processing was executed using quantile normalization. Subsequently, after the normalization, the outliers were removed. For the genes expressions estimation, it was used the general mixed effect model for ASD and standard linear regression for SCZ and BPD. The differentially expressed genes among disorder were found multiplying the Z-scores across the three disorders. The genes that obtained the highest Z-scores were considered as differentially expressed between the disorders. The genes that obtained a p-value less than 0.05 were considered significant. Then to assess the similarity of genes among disorder, Pearson correlation was obtained for the Z-scores. As a result, the correlation across disorder comparison highlights a similarity of altered transcriptome with a P-value less than 0.001 for ASD-SCZ with 191 differentially expressed genes. Nevertheless, for ASD-BPD and SCZ-BPD correlation was not significant.

The similarity of this research described above [23] and the one proposed in this document, is that both seek to identify genes in common. First, they applied quantile normalization to the data and outliers were eliminated while the proposed one does not manipulate the data. Second, it differs in the way to identify the differentially expressed genes across disease in which they used the multiplication of Z-score while our optimization model identified them using Pareto optimality conditions to analyze all genes expressions of both diseases simultaneously.

Another study found in the literature with the main objective to quantify behavioral phenotypes shared by ASD, and the autistic subgroup of SCZ patients was conducted by A. Kastner et al [24]. The Positive and Negative Syndrome Scale (PANSS) was used to characterize autistic phenotypes in SCZ patients. Due to the high consistency of all individuals with autism variables it was encouraged to add them to create an autism severity score (PANSS autism severity

score, PAUSS); it was achieved and validated by integrating a total of 1,156 schizophrenia patients and 165 autism spectrum disorder adult patients. The evaluations were conducted with observations and questionnaires applying PAUSS. The scores obtained were analyzed through statistical methods. The Spearman rank correlation coefficient was used for the factors evaluated in the questionnaires. To evaluate group differences the use of the Mann Whitney U test and Chi-square/Fisher exact test were applied. Also, Bonferroni multiple corrections were applied for the analysis. As a result, the use of PAUSS could differentiate between ASD, SCZ and control samples. It concluded that the negative symptoms of SCZ overlap with the autistic feature supporting that PAUSS is appropriate for the evaluation of autistic behavior in SCZ patient and with ASD patient.

The previous study's main focus [24] is to identify the symptoms shared between ASD and SCZ while our proposed work identifies the genes in common between them. Another difference is that they applied a statistical method to analyze the data while we used deterministic methods. Besides they had to apply corrections while our method does not require them. Although both studies have different approaches, both seek to identify what these disorders have in common. A similar study, like the one before, was found in the literature proposed by B. St. Pourcain et al. [25]. Their main objective is to evaluate the polygenic influences commonalities in SCZ and ASD through the genetic overlap with phenotypic symptoms. The data used for the analysis was obtained from publicly available genome-wide data. Statistical methods were used to estimate the cumulative effect of common SNP (single-nucleotide polymorphisms). The genetic correlation between clinical ASD and SCZ using linkage disequilibrium (LD) score correlation analysis. The changes in genetic effect over time was determined by a mixed Poisson regression. Analyzing the results,

they found out that ASD and SCZ shared social communication difficulties. Also, both disorders showed to have certain genetic overlap.

In the studies described before, [24] [25] both aim to identify the overlapping symptoms. Kastner [24] and St. Pourcain [25] used questionnaires and genome data respectively to be analyzed by statistical methods. After the analysis, they highlighted that ASD and SCZ share phenotypic symptoms such as negative symptoms and social communication difficulties. Unlike them, the proposed work in this thesis identifies the genes that these diseases have in common through optimization models. Although the focuses of these two studies are different, we have the same objective in common, to be able to identify commonalities among disease.

2.2 Methodology Background

2.2.1 Multiple Criteria Optimization

Multiple Criteria Optimization (MCO) is an area within the field of optimization that deals with multiple objective conflicts [26]. Following the work of Sánchez-Peña et al. [15] and Camacho-Cáceres [16], applying MCO allows us to find the best set of solutions for the performance measures that are going to be analyzed. Figure 1 represents the MCO problem showing the non-dominated solutions having as objective the performance measure (PM). PM's used are the absolute values of difference between control medians and condition medians of relative expression for each disease. These non-dominated solutions are denominated as a Pareto efficient frontier.

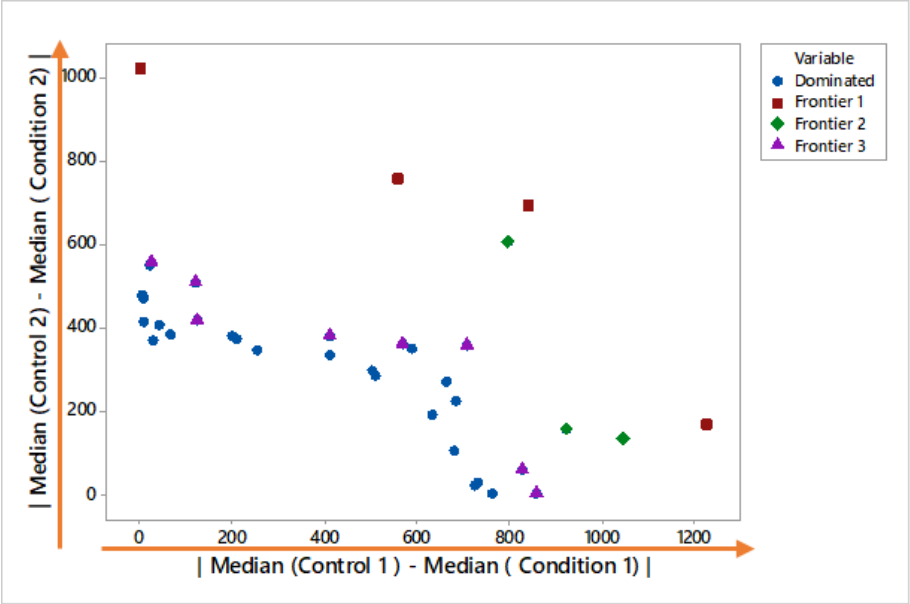

Figure 1: Representation of MCO problem

In MCO, G represents a set of many solutions that contains n genes to be analyzed represented as g_i (i = 1, 2, … n). C is number of PMs that is going to be considered in the analysis where each one is represented as k (k = 1, 2, …, C). The PMs or criteria under analysis are denoted as m_i^k for the *i*-th gene in the *k*-th PM. PMs regularly are related to the difference between genes expressions to look for the highly expressed genes in the ailments that are going to be studied.

To obtain Pareto Efficient solutions (non-dominated solutions), these must meet Pareto optimality conditions. These can be found in as the following: a solution $X^{(1)}$ is said to dominate the other solution $X^{(2)}$ if both condition 1 and 2 are true:

1. The solution $X^{(1)}$ is no worse than $X^{(2)}$ in all PMs.
2. The solution $X^{(1)}$ is strictly better than $X^{(2)}$ in at least one PM

The Pareto optimality conditions mentioned before, are used to evaluate every pair of genes obtaining the ones that are not dominated by others. After doing this, Pareto-efficient genes are found, and these form the Pareto-efficient frontier of the MCO shown in Figure 1.

All gene expressions are measured in two states to look for the differentially expressed genes. In this work, the relative expression of the two states named “control” and “condition” were used to calculate the difference of the medians between them for each gene in each disease. The main objective of this method is to find the Pareto efficient solutions by maximizing the absolute value of the difference of the medians (PMs).

For this work, we adopted the Camacho-Cáceres [16] MCO methodology for microarray analysis. For g_i (m_i^1, m_i^2, …, m_i^k, …, m_i^C) represents the i-th gene in terms of its value of the C PMs. The Pareto efficient solutions (non-dominated solutions) are denoted as g_i^* (m_i^{1*}, m_i^{2*}, …, m_i^{k*}, …, m_i^{C*}) which represent the objective of this analysis. First, we build a matrix δ^k for the k-th PM resulting in C matrices with n x n dimensions for the full pairwise comparison as follows:

$$\delta^k = \begin{bmatrix} & \mathbf{m_1^k} & \mathbf{m_2^k} & \dots & \mathbf{m_j^k} & \dots & \mathbf{m_n^k} \\ \mathbf{m_1^k} & \delta_{11}^k & \delta_{12}^k & \cdots & \delta_{1j}^k & \dots & \delta_{1n}^k \\ \mathbf{m_2^k} & \delta_{21}^k & \delta_{22}^k & \cdots & \delta_{2j}^k & \dots & \delta_{2n}^k \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{m_i^k} & \delta_{i1}^k & \delta_{i2}^k & \dots & \delta_{21}^k & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \mathbf{m_n^k} & \delta_{n1}^k & \delta_{n2}^k & \dots & \delta_{nj}^k & \dots & \delta_{nn}^k \end{bmatrix}$$

Where:

$$\boldsymbol{\delta_{ij}^k} = \begin{cases} 1, & \text{if } m_i^k > m_j^k \\ 0, & \text{if } m_i^k = m_j^k \\ W, & \text{if } m_i^k < m_j^k \end{cases} ; \text{For } i=1, 2, \dots,n;\ j=1,2,\dots,n;\ k=1,2,\dots,C \qquad (2.1)$$

W is defined as a large negative number for this case because it is maximizing as a penalty. For this work, W = - 1000 is used.

The matrix sums are computed as follows:

$$\boldsymbol{\alpha_{ij}} = \sum_{k=1}^{C} \delta_{ij}^{k} \tag{2.2}$$

The matrix γ is built by evaluating α_{ij} values. The general form to obtain the values of γ_{ij} for any value of C is the following:

$$\boldsymbol{\gamma_{ij}} = \begin{cases} \frac{C}{2}W, & \text{if } \alpha_{ij} \in \{0, W, \ldots, (C-1)W\} \\ C\,W, & \text{if } \alpha_{ij} = 2W \\ 0, & \text{Otherwise} \end{cases} \tag{2.3}$$

To find the g_i^*, a β vector is made with the sums of each row of the matrix γ as follows:

$$\boldsymbol{\beta_{ij}} = \sum_{j=1}^{n} \gamma_{ij} \quad i = 1, 2, \ldots n \tag{2.4}$$

The equation (2.5) is used to obtain the Pareto efficient frontier, which contains all non-dominated solutions as following:

$$\boldsymbol{g_i^*} = \{g_i^* \mid \beta_i > CW, i = 1, 2, \ldots n\} \tag{2.5}$$

Through this step, the Pareto efficient solutions are identified as $g_i^* = \{m_i^{1*}, m_i^{2*}, \ldots m_i^{k*}, \ldots, m_i^{C*}\}$. The objective of this algorithm is to identify every solution contained in the Pareto efficient frontier. Applying the MCO in this way, the differentially expressed genes among diseases are obtained.

2.2.2 Pairwise Difference

The pairwise difference is based on the calculation of the difference between one element from the group a and one element from the group b, and this approach allows to consider all possible combination. In this case, group a represents all control samples expressions while group b represents all condition samples expressions. And the difference to be calculated corresponds to the difference of expression of a single gene. To calculate the total number of combinations is by

multiplying the total number of the control sample and the total number of condition sample. For example, if two control sample and three condition sample were obtained the total number of pairs between them will be six pairs. After grouping the six pairs, the difference between them is calculated. An example of this would be if we have a dataset that contains six pairs: (7,2), (7,11), (7,21), (19,2), (19,11) and (19,21). To calculate the difference for the first pair is 7 – 2 = 5, second pair is 7 – 11 = -4 and for the third pair is 7 – 21 = -14. In this way the differences are calculated for the six pairs obtaining the following results: 5, -4, -14, 17, 8 and -2. This method is used to quantify the changes between control and condition expression. Let $\boldsymbol{O_{i,a}}$ represent the relative expression levels of g_i in a-th control sample and $\boldsymbol{T_{i,b}}$ represent the relative expression levels of g_i in b-th condition sample. The pairwise difference is given by $\boldsymbol{D^i_{a,b}}$ which represent the difference for the i-th gene between the a-th control samples and b-th condition samples to be studied. To determine the total number of differences in relative expression levels of $the\ g_i$ is calculated multiplying the total of control and condition samples. The Pairwise Difference is represented as follows:

$$\mathbf{D^i_{a,b}} = O_{i,a} - T_{i,b} \tag{2.6}$$

For i = 1, 2, 3, …, n

For a = 1, 2, 3, …, A

For b = 1, 2, 3, …, B

Figure 2 represents the application of these variables to the previous example.

Gene 1			
Control Samples $O_{1,a}$	Condition Samples $T_{1,b}$	Pairwise Difference Calculation	Pairwise Differences $D^1_{a,b}$
7	2	7 - 2	5
19	11	7 - 11	-4
	21	7 - 21	-14
		19 - 2	17
		19 - 11	8
		19 - 21	-2

Figure 2: Example of Pairwise Difference Calculation

2.2.3 Correlation

The correlation coefficient has been calculated for the expression differences between "control" and "condition" .Let denote that x represents the difference of g_1 which is $\boldsymbol{D^1_{a,b}}$, and **y** represent the difference of $g_2$which is $\boldsymbol{D^2_{a,b}}$. The expression to find the Pearson coefficient for the differences is as follows:

$$r_{xy} = \frac{S_{xy}}{S_x * S_y} = \frac{\frac{\sum x*y}{n-1} - \bar{x}*\bar{y}}{\sqrt{\frac{\sum(x-\bar{x})^2}{n-1} * \frac{\sum(y-\bar{y})^2}{n-1}}} \quad (2.7)$$

Where:

- S_{xy} is the covariance of xy
- S_x is the standard deviation of x
- S_y is the standard deviation of y
- $\bar{x}$ is the mean of x
- $\bar{y}$ is the mean of y
- n is the total of differences.

2.2.4 Travelling Salesman Problem

This work uses the traveling salesman problem method to find the most correlated path optimally. TSP methodology is well known in networks and combinatorial optimization. It is a generic model that has a diversity of application such as planning, logistic, DNA sequencing, and others. This method constructs a cyclic path that starts with a home base node and visiting the remaining nodes once that minimize the total tour cost. Let C_{ij} represent the traveling cost of node i to node j and y_{ij} be a binary variable, which whether the salesman travels from city i to city j. The variable x_{ij} defines the flow on every arc (i,j) and assuming that salesman has n-1 units available at node 1. The source node is selected arbitrary and salesman must deliver 1 unit to each of the other nodes. The model formulation is as following [27]:

$$Minimize \sum_{(i,j)\in A} c_{ij}\, y_{ij} \tag{2.8}$$

Subject to

$$\sum_{1\le j\le n} y_{ij} = 1 \qquad \text{for all i = 1, 2, …, n} \tag{2.9}$$

$$\sum_{1\le j\le n} y_{ij} = 1 \qquad \text{for all j = 1, 2, …, n} \tag{2.10}$$

$$\mathrm{N}x = b \tag{2.11}$$

$$x_{ij} \le (n-1)y_{ij} \qquad \text{for all (i, j)} \in \mathrm{A} \tag{2.12}$$

$$x_{ij} \ge 0 \qquad \text{for all (i, j)} \in \mathrm{A} \tag{2.13}$$

$$y_{ij} = binary \qquad \text{for all (i, j)} \in \mathrm{A} \tag{2.14}$$

The constraint (9) and (10) ensure that one arc enters a node and one arc exits the node. Constraint (11) consists of the incidence matrix that should be equal to the supply/demand ensuring that one unit enters the arc and one unit exits. Constraint (12) represents the flow restriction.

2.2.5 Minimum Spanning Tree

The minimum spanning tree (MST) is a network optimization model previously used by J. Rosas [17].This model seeks to connect all network nodes with minimum possible number of arcs and without cycles, minimizing total arc length. Let A (S) represent the set of arcs contained in the subgraph of G = (N, A) induced by the node set S [i.e., A(S) is the set of arcs of A with both endpoints in S]. Let C_{ij} denote the traveling cost of city i to city j and y_{ij} is the variable that represents the selection of arc (i,j). The formulation for the minimum spanning tree problem is as follow [27]:

$$Minimize \sum_{(i,j)\epsilon A} c_{ij} y_{ij} \tag{2.15}$$

Subject to

$$\sum_{(i,j)\epsilon A} y_{ij} = n - 1 \tag{2.16}$$

$$\sum_{(i,j)\epsilon A(S)} y_{ij} \leq |S| - 1 \text{ for any set S of nodes} \tag{2.17}$$

$$y_{ij} \geq 0 \text{ and integer} \tag{2.18}$$

In this formulation, y_{ij} have values of 0-1indicate whether we select arc (i,j) as part of the chosen spanning tree (note that the second set of constraints with $|S| = 2$ indicates that every $y_{ij} \leq 1$) . The constraint 2.16 is a cardinality constraint implying that we choose exactly n – 1 arc, and the "packing" constraint 2.17 implies that the set of chosen arcs contain no cycles (if the chosen solution contained a cycle and S were the set of nodes on a chose to cycle, the solution would violate this constraint).

Chapter 3 Methodology

This work uses the database from the study “Autism and Increased Paternal Age-Related Changes in Global Levels of Gene Expression Regulation” by Alter et al., published in 2011 [28]. This database of genes expressions were obtained from peripheral blood lymphocytes (PBL) of children with autism (n=82) and controls (n=64). Also, this work used the database from the study "Analysis of gene expression in two large schizophrenia cohorts which identifies multiple changes associated with nerve terminal function" by Maycox et al., published in 2009 [29]. The gene expressions were obtained from postmortem tissue of adults with schizophrenia (n=28) and control (n=23). These two databases have in common that after the extraction of RNA these were labeled using Affymetrix's GeneChip. Figure 3 represents a summary of the databases of each disease.

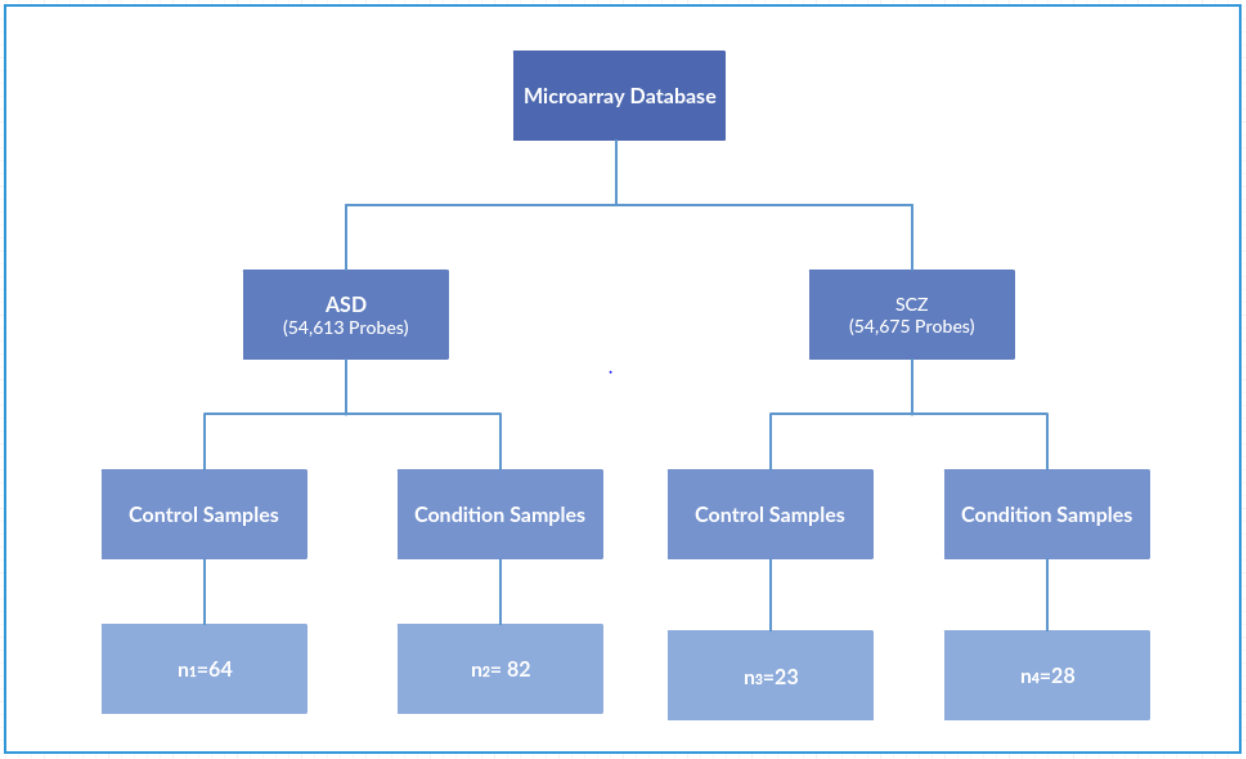

Figure 3: ASD and SCZ database summary

The first step is the pre-processing phase of the data that consists on calculating the median of the relative expressions to consolidate the genes that are repeated. This phase facilitates the analysis of these databases. Figure 4 represents an example of how the database looks like and Figure 5 how it looks after the pre-processing phase. These were made using R software for the ease of calculation.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
Gene 1	56	23	64	48	16	62	19	18
Gene 1	20	26	17	69	35	13	57	31
Gene 2	57	38	67	29	53	55	41	33
Gene 3	74	39	66	35	27	75	44	16
Gene 3	41	58	75	58	34	32	61	67
Gene 4	32	52	14	16	58	27	24	65
Gene 4	70	57	63	40	30	45	73	34
Gene 5	71	13	35	38	23	47	11	28

Figure 4: Representation of the raw database before the pre-processing.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
Gene 1	38	24.5	40.5	58.5	25.5	37.5	38	24.5
Gene 2	57	38	67	29	53	55	41	33
Gene 3	57.5	48.5	70.5	46.5	30.5	53.5	52.5	41.5
Gene 4	51	54.5	38.5	28	44	36	48.5	49.5
Gene 5	71	13	35	38	23	47	11	28

Figure 5: Result of the pre-processing using the median.

As a result, ASD and SCZ database were consolidated to 21,924 genes. After verifying the genes that are overlapping between the databases the overlapping was of 99.9%. The total of genes selected for the simultaneous analysis was of 21,915 genes. After finishing with the pre-processing of the data, it was continued with the identification of the differential expressed genes.

The performance measures used in this work are the absolute value of the median group, which was calculated for ASD (condition and control samples) and SCZ (condition and control

samples). Then the resulting performance measures are graphed and analyzed with the use of MCO problem. This work finds the first five Pareto efficient frontiers for the simultaneous microarray analysis of ASD and SCZ to identify the important genes in common. The databases were divided into 4 groups of 7,000 or less due to the computer memory constraint and iterated until the final efficient frontier. These groups contain the absolute median difference of ASD and SCZ. An example of the MCO is shown in Figure 6.

Equation (2.1)

$\delta^1 =$

m1\m2	3	5	2	6	2	3	1	4	5	1
3	0	-1000	1	-1000	1	0	1	-1000	-1000	1
5	1	0	1	-1000	1	1	1	1	0	1
2	-1000	-1000	0	-1000	0	-1000	1	-1000	-1000	1
6	1	1	1	0	1	1	1	1	1	1
2	-1000	-1000	0	-1000	0	-1000	1	-1000	-1000	1
3	0	-1000	1	-1000	1	0	1	-1000	-1000	1
1	-1000	-1000	-1000	-1000	-1000	-1000	0	-1000	-1000	0
4	1	-1000	1	-1000	1	1	1	0	-1000	1
5	1	0	1	-1000	1	1	1	1	0	1
1	-1000	-1000	-1000	-1000	-1000	-1000	0	-1000	-1000	0

Equation (2.2)

$\alpha =$

m1\m2	13	6	11	2	8	3	7	3	2	5
3	0	-999	2	-999	2	1	2	-999	-999	2
5	-999	0	-999	-999	-999	2	-999	2	1	2
2	-2000	-999	0	-999	1	-999	2	-999	-999	2
6	-999	-999	-999	0	-999	-999	-999	-999	1	-999
2	-2000	-999	-1000	-999	0	-999	2	-999	-999	2
3	-1000	-2000	-999	-999	-999	0	-999	-1000	-999	-999
1	-2000	-999	-2000	-999	-2000	-999	0	-999	-999	1
4	-999	-2000	-999	-999	-999	1	-999	0	-999	-999
5	-999	-1000	-999	-1000	-999	-999	-999	-999	0	-999
1	-2000	-2000	-2000	-999	-2000	-999	-1000	-999	-999	0

Equation (2.1)

$\delta^2 =$

m1\m2	13	6	11	2	8	3	7	3	2	5
13	0	1	1	1	1	1	1	1	1	1
6	-1000	0	-1000	1	-1000	1	-1000	1	1	1
11	-1000	1	0	1	1	1	1	1	1	1
2	-1000	-1000	-1000	0	-1000	-1000	-1000	-1000	0	-1000
8	-1000	1	-1000	1	0	1	1	1	1	1
3	-1000	-1000	-1000	1	-1000	0	-1000	0	1	-1000
7	-1000	1	-1000	1	-1000	1	0	1	1	1
3	-1000	-1000	-1000	1	-1000	0	-1000	0	1	-1000
2	-1000	-1000	-1000	0	-1000	-1000	-1000	-1000	0	-1000
5	-1000	-1000	-1000	1	-1000	1	-1000	1	1	0

Equation (2.3)

$\gamma =$

m1\m2	13	6	11	2	8	3	7	3	2	5
3	-1000	0	0	0	0	0	0	0	0	0
5	0	-1000	0	0	0	0	0	0	0	0
2	-2000	0	-1000	0	0	0	0	0	0	0
6	0	0	0	-1000	0	0	0	0	0	0
2	-2000	0	-1000	0	-1000	0	0	0	0	0
3	-1000	-2000	0	0	0	-1000	0	-1000	0	0
1	-2000	0	-2000	0	-2000	0	-1000	0	0	0
4	0	-2000	0	0	0	0	0	-1000	0	0
5	0	-1000	0	-1000	0	0	0	0	-1000	0
1	-2000	-2000	-2000	0	-2000	0	-1000	0	0	-1000

Equation (2.4) and (2.5)

βi	Solution
-1000	Non-dominated
-1000	Non-dominated
-3000	Dominated
-1000	Non-dominated
-4000	Dominated
-5000	Dominated
-7000	Dominated
-3000	Dominated
-3000	Dominated
-10000	Dominated

Figure 6: Example of Multiple Criteria Optimization Method

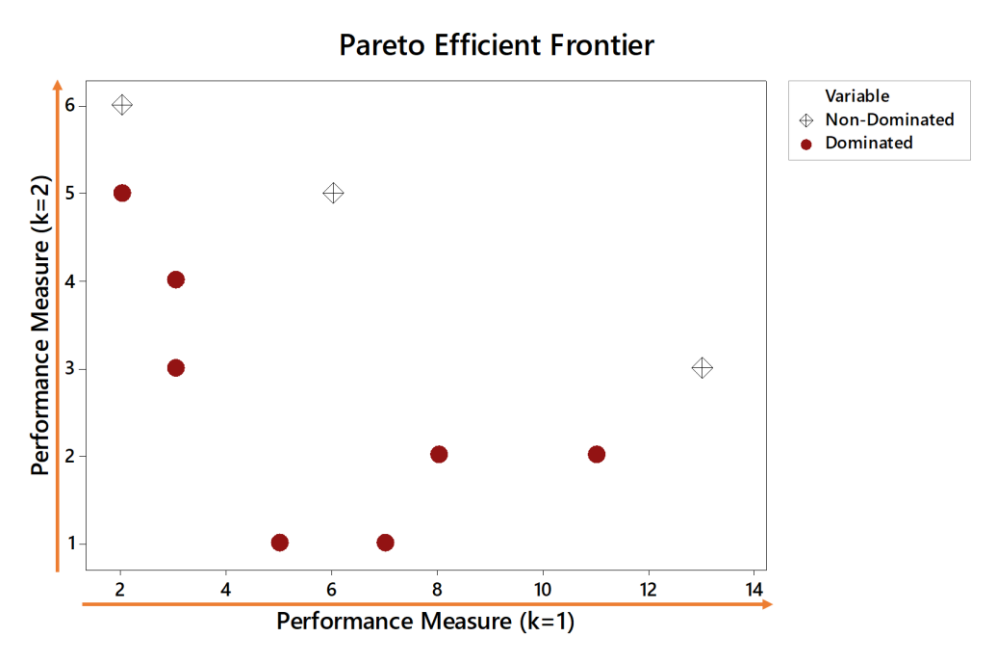

Figure 7: Pareto Efficient Frontier Example Result

With the differentially expressed genes obtained from the MCO, the behavior of the genes among them are measured with the statistical correlation. To measure the statistical correlation first need to calculate the pairwise difference between each gene expressions shown in Section 2.2.2. When the pairwise difference is applied to the genes, a total of 5,248 of differences for a gene in ASD and a total of 552 of differences for a single gene in SCZ . This total can be calculated multiplying the total of control (n = 64) and condition (n = 82) for ASD and for SCZ multiplying the total of control (n = 23) and condition (n=28). After calculating all the differences for all the genes the next step is to calculate the correlation. This correlation is used to measure the coordinated behavior between differences [30]. Figure 8 shows an example of the correlation matrix between genes for each disease.

ASD	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Gene 1	1.0000	-0.0647	-0.1046	-0.1574	-0.0982
Gene 2	-0.0647	1.0000	0.9922	0.0009	0.9348
Gene 3	-0.1046	0.9922	1.0000	0.0120	0.9399
Gene 4	-0.1574	0.0009	0.0120	1.0000	-0.0344
Gene 5	-0.0982	0.9348	0.9399	-0.0344	1.0000

SCZ	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Gene 1	1.0000	0.1297	0.1060	0.5030	0.1333
Gene 2	0.1297	1.0000	0.9822	0.2191	0.9787
Gene 3	0.1060	0.9822	1.0000	0.2118	0.9911
Gene 4	0.5030	0.2191	0.2118	1.0000	0.2354
Gene 5	0.1333	0.9787	0.9911	0.2354	1.0000

Figure 8: Example of correlation matrix between genes for ASD and SCZ

After obtaining the correlation matrix for each disease, the sums of the absolute value the traveling salesman problem (TSP) is followed. The main objective of this method is to find the most correlated cyclic path between genes. Through the TSP it can be seen how the genes interact or the correlated behavior between them in a cyclic way that helps to understand how they contribute to the ailment development. To obtain the cost matrix, the sum of the absolute value of the two matrices shown in Equation 3.1.

$$\rho_{i,j}^{asd.scz} = \left|\rho_{i,j}^{asd}\right| + \left|\rho_{i,j}^{scz}\right| \tag{3.1}$$

Figure 9 represents the cost matrix that is given by c_{ij} explained in Section 2.2.4 . This cost matrix represents the cost of going of gene i to gene j. The selection matrix has the same dimension of the cost matrix and is given by a binary variable y_{ij}. A graphical representation of the network problem can be seen in Figure 10, which contains the genes network and the correlation coefficient as cost. An example of the traveling salesman problem using the correlation matrix between genes to find the most correlated path is shown in Figure 11.

	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Gene 1	2.0000	0.1944	0.2107	0.6605	0.2314
Gene 2	0.1944	2.0000	1.9743	0.2200	1.9135
Gene 3	0.2107	1.9743	2.0000	0.2238	1.9310
Gene 4	0.6605	0.2200	0.2238	2.0000	0.2698
Gene 5	0.2314	1.9135	1.9310	0.2698	2.0000

Figure 9: Example of the Cost Matrix Using the Absolute Value of the Correlation Matrix

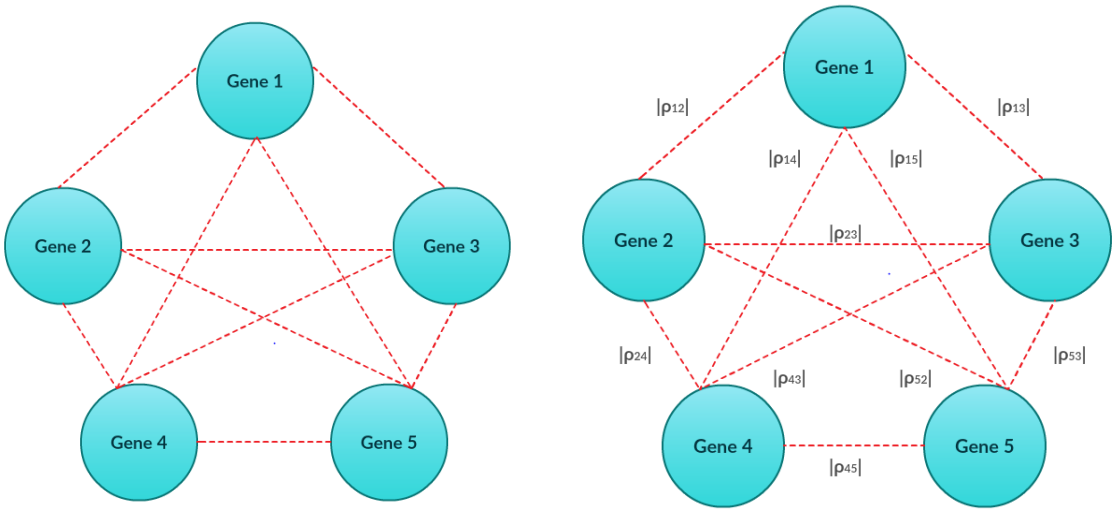

Figure 10: Genes network representation and the correlations as cost between them.

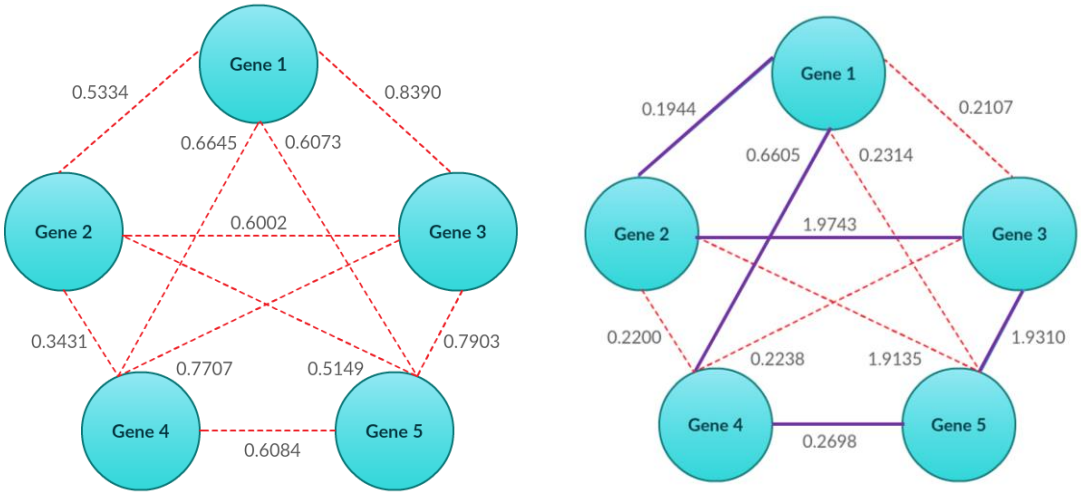

Figure 11: Example of Traveling Salesman Problem Solution

Figure 11 is a graphical representation that shows the interaction between five genes. The purple line represents the most correlated cyclic path. The sum of absolute correlations of this path is of 3.27 in which is maximizing its correlation. Another way to structure the behavior of the genes is using the minimum spanning tree (MST). This method can find the most correlated acyclic path maximizing the correlation between genes. With the example shown before, the solution for

the MST method was obtained. The purple lines represent the path that maximizes its total correlation in Figure 12.

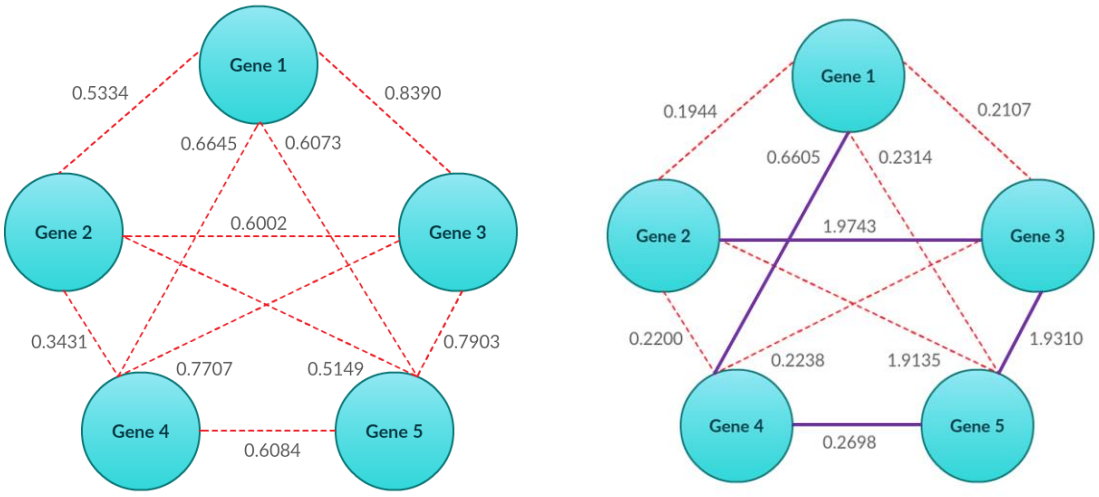

Figure 12: Example of the Minimum Spanning Tree Solution

In addition to the genetic analysis between ASD and SCZ, a cost model was developed to estimate and compare the cost between them. This comparison is made through a cost analysis that allows us to have an idea of the cost associated with misdiagnosis. This way it was possible to identify which condition is more expensive. These results could be as a future reference to raise awareness of the diagnosis process.

The methodology shown above demonstrates its usefulness to characterize two conditions simultaneously using microarray databases. As a result, obtaining the common genetic aspects and most correlated structures between them. These results can be used as guides to scientists for future research to help to understand better ASD and SCZ. In this way, the scientists could transmit the possible findings found to doctors for better diagnosis and treatment.

Chapter 4 Results and Discussion

4.1 Differentially Expressed Genes of Autism Spectrum Disorder and Schizophrenia.

As an outcome of the MCO implementation, it was possible to identify the genes in common that change their expression the most between ASD and SCZ. The total genes selected after data pre-processing for the simultaneous analysis was of 21,915 genes. After solving the MCO with Pareto optimality conditions, a total of 36 differentially expressed genes were identified in the first five Pareto efficient frontiers. These genes represent approximately 0.16% of the total genes selected for the analysis. Table 1 contains genes whose expression differentiates the most between control and condition expression in both diseases obtained with MCO method. The genes obtained in the first frontier are those that change their expression the most.

Table 1 Differentially Expressed Genes of Autism Spectrum Disorder and Schizophrenia.

Frontier 1	Frontier 2	Frontier 3	Frontier 4	Frontier 5
B2M	HBB	EEF1A1	BEX3	CALM2; CALM3; CALM1
HBA1; HBA2	IFITM2	FCGR3B; FCGR3A	COX6C	MT1H
HBA2; HBA1	TXNIP	FTL	HSPA8; SNORD14C; SNORD14D	ND2
PTGDS		IFITM3	MT1HL1	ND5
		MT1X	NRGN	NDRG4
		MT2A	PFN1	RPL13A; SNORD33; RPL13AP5; SNORD34; SNORD35A; SNORD32A
		TMSB4X	RPS11; SNORD35B	RPL34
		TUBA1A	RPS25	RPS6
			S100P	SNORD55; SNORD38B; RPS8
			SNORD14D; HSPA8; SNORD14C	SPP1
				TALDO1

Later, all the genes were classified as upregulated and downregulated compared with controls for each disease shown Table 2. It was observed that 14 genes have the same classification

for ASD-SCZ, where 3 genes were downregulated (TUBA1A, NRGN, and CALM2-CALM3-CALM1) and 11 genes were upregulated (B2M, EEF1A1, TMSB4X, RPS11-SNORD35B, RPS25, SNORD14D-HSPA8-SNORD14C, ND5, RPL34, RPS6, and SNORD55-SNORD38B-RPS8). This consistent behavior across both conditions make these genes especially interesting for follow up.

Table 2 Upregulated and Downregulated Genes

Genes	ASD-SCZ	SCZ	ASD	Genes	ASD-SCZ	SCZ	ASD
B2M	↑			MT1HL1		↑	↓
HBA1; HBA2		↓	↑	NRGN	↓		
HBA2; HBA1		↓	↑	PFN1		↑	↓
PTGDS		↓	↑	RPS11; SNORD35B	↑		
HBB		↓	↑	RPS25	↑		
IFITM2		↑	↓	S100P		↑	↓
TXNIP		↑	↓	SNORD14D; HSPA8; SNORD14C	↑		
EEF1A1	↑			CALM2; CALM3; CALM1	↓		
FCGR3B; FCGR3A		↑	↓	MT1H		↑	↓
FTL		↑	↓	ND2		↓	↑
IFITM3		↑	↓	ND5	↑		
MT1X		↑	↓	NDRG4		↓	↑
MT2A		↑	↓	RPL13A; SNORD33; RPL13AP5; SNORD34; SNORD35A; SNORD32A		↑	↓
TMSB4X	↑			RPL34	↑		
TUBA1A	↓			RPS6	↑		
BEX3		↓	↑	SNORD55; SNORD38B; RPS8	↑		
COX6C		↓	↑	SPP1		↑	↓
HSPA8; SNORD14C; SNORD14D	↑			TALDO1		↑	↓

4.2 Biological Evidence

After identifying the differentially expressed genes with MCO, the next step was to understand the biological role that each gene plays. The information provided in this section contains the genes that belong to the first four Pareto efficient frontiers obtained from the MCO. The biological information was carried out with the collaboration of Diamarys Salome and

Alejandro Marrero, undergraduate students who are part of our research group. This information includes general information of the genes, how these could be related to ASD and/or SCZ, and evidence found in research related to these diseases. These findings help to understand better the genes biological role that leads to the development of the disease.

The principal adult hemoglobin A is composed of two α-globin and two β-globin polypeptides, in which they are encoded by the genes HBA1, HBA2, and HBB. These hemoglobin proteins' main function is to transport the oxygen through blood to support oxidative metabolism. Hemoglobin Subunit Alpha 1 and 2 are located on chromosome 16. These genes have been linked to the binding of iron ions in the hemoglobin. Iron plays an essential role in brain function, particularly during the early stages of brain development [31]. Without proper brain development, children can develop any one of the numerous neurological disorders like Autism and Schizophrenia. Evidence of iron deficiency was found in a study that showed a very high prevalence of iron deficiency in children with autism [32]. These results suggest that iron deficiency affects the development of these neurodegenerative diseases directly. A study of RNA-Seq evaluation identified risk genes in schizophrenia [33]. Within the list of up-regulated genes HBA1, HBA2, and HBB were found in schizophrenia blood samples. HBB has been identified with a likely pathogenic variant after the evaluation of the heterozygous variant with autosomal dominant inheritance in neurodevelopmental conditions including ASD [34].

Metallothionein (MT) (MT1X, MT2A, MT1H, MT1HL1) is a family of cysteine-rich, low molecular weight proteins located in the membrane of the Golgi apparatus. These proteins act as an antioxidants, and can bind heavy metal for their detoxification [35] and have been reported as oxidative stress-inducing agents [36]. The MTs genes have an essential role in zinc regulation and intracellular space distribution. [35]. Prenatal zinc deficiency can be relevant to SCZ due to lower

levels of maternal zinc or fetal gene variants that affect zinc transportation across cellular membranes [37]. Also, it has been found that children with ASD have lower levels of zinc [38]. Furthermore, having in consideration zinc-mediated up-regulation of MT genes expressions, has been suggested a possible function alteration of the neuroprotective metallothionein system [38]. Research indicates that the frequency of low levels of zinc and copper toxicity can indicate that the function of metallothionein system decreases [39] [40]. A cross-study analysis of psychosis patient has been identified as a group of MT genes including (MT1X, MT2A, MT1H) that are significantly up-regulated in psychosis in the prefrontal cortex [41]. Another study has been reported that MT2A gene expression had increased in prefrontal cortex in SCZ [42].

A potential gene that affects neurodegenerative diseases is TXNIP or Thioredoxin Interacting Protein. According to Al-Gayyar [43]: "thioredoxin-interacting protein (TXNIP) is an endogenous inhibitor of the thioredoxin system". The thioredoxin system is a vital cellular system against oxidative stress. It has been suggested that oxidative stress occurs in some stages of the schizophrenic episode and it could be involved with the pathogenesis and symptomology of SCZ [44] [45] [46]. The normal function of the thioredoxin system proteins is accompanied by cell protection against damage. When these proteins are not interacting correctly, they make cells more vulnerable to death [47] [48]. The role of thioredoxin in brain development has been described previously. [49] A disruption to the regular function of the thioredoxin system has been linked to neurodegenerative diseases [50] [51] [52]. The impact of TXNIP on neuronal apoptosis, neurodevelopment, and neurodegeneration may clarify the fundamental connection between TXNIP and SCZ [53].

IFITM2 and IFITM3 are part of the IFITM family that has been described to have antiviral effects that act against a broad range of RNA viruses in the stage of virus-cell fusion [54] [55]. As

an outcome, they prevent viral replication, the sequelae of virus-associated disease, and inflammation [55]. In the literature, it has been found that prenatal infection could be a risk for SCZ and ASD development [56] [57] [58]. Microarray research with SCZ patient points that IFITM3 could be critical for SCZ pathophysiology. The activation and expression of this gene are essential for the central nervous system (CNS) of maternal immune activation [59]. Research that used RNA-Seq data from the hippocampus, associated IFITM2 and IFITM3 with immune response and inflammation in SCZ [9]. Also, it has been found that high mRNA levels of IFITM genes could be related to neuroinflammatory processes SCZ [9]. It is suggested to study the possible relation of these genes in prenatal infection and ASD.

B2M (Beta 2 Microglobulin) is a protein found in nucleated cells and on most of the biological fluids. This protein is vital for the immune inflection and scrutiny on vertebrate animals. Patients with elevated plasma B2M levels have been found to have a range of inflammatory, hematologic, immunodeficiency, and renal diseases [60]. The abnormality of the B2M is related to multiple diseases and with the progress of cancer cells. Also, B2M is associated with the formation of the MHC (major histocompatibility complex) enclosing from antigen presentation to immune homeostasis [61]. A study [62] found that found that patient with SCZ had higher levels of B2M serum and might play a role in the development of SCZ.

PTGDS (β trace protein) is one of the most common proteins found in the cerebrospinal fluid (CSF) [63]. This protein has a role in the contraction of smooth muscle and platelet aggregation and function as a neuromodulator in the CNS [64]. Also, has been found that PTGDS is a significant apoptotic factor in Alzheimer disease plasma [65]. It has been found that this protein is statistically significant up and down regulations in multiple sclerosis, Parkinson's disease and SCZ when compared with controls [66].

NRGN, known as neurogranin, is a small neuronal protein that binds to calmodulin free form (Ca^{2+}) [67]. Neurogranin is a shared susceptibility gene of SCZ [68]. A study hypothesized that this gene may mediate the risk associated with schizophrenia via intellectual dysfunction [69]. It has been suggested that NRGN is implicated to the pathophysiology of SCZ due to genetic risk variant. [70]

Ferritin family (FTL) is a protein that storages and releases iron in a controlled manner. Even though the iron is critical for the oxygen conversion to cellular energy, it has a high potential of being toxic and ease the free radical formation [71]. The increment of FTL may suggest an oxidative imbalance due to iron worsening the neurotoxicity and mitochondrial failure [72]. It has been found that iron is linked with socio-cognitive and socio-emotional development in ASD [73]. A study identified that 24.1% of children with ASD had lower levels of iron and 15.5% had anemia [74].

EEF1A1 (Eukaryotic Translation Elongation Factor 1 Alpha 1) is a protein in charge of the enzymatic conveyance of aminoacyl tRNAs to the ribosome. Studies point out that can be able to move among the nucleus and cytoplasm [75]. Also, it is a multifunctional gene involved in signaling transduction, cellular apoptosis, heat shock response, cytoskeleton regulation, and RNA virus replication [75] [76] [77].

The FCGR3A and FCGR3B (FC fragment of IgG receptor IIIa and FC fragment of IgG receptor IIIb) are similar genes located on chromosome 1. FCGR3A is expressed on natural killer cells, while FCGR3B is expressed on polymorphonuclear neutrophils. According to the literature that FCGR3A plays a role in the immune complex clearance and is reinforced the importance of these receptors in the development of systemic lupus erythematosus [78].

TMSB4X has a role in the regulation of actin polymerization, cell proliferation, migration, and differentiation. It has been found under-expressed in acute myeloid leukemia. [79].

TUBA1A (tubulin alpha 1a) belongs to the microtubules of the eukaryotic cytoskeleton and is located in the chromosome 12q. Mutation of this gene could lead to severe brain malformation [80]. A study identified abnormalities in cerebellar, brainstem and basal ganglia associated with a mutation in tubulin genes including TUBA1A [81] [82]. Likewise, it has been found that the complex brain malformation is due to the mutation of tubulin genes including TUBA1A [83].

BEX3 belong to the brain-expressed x-linked and is responsible for the regulation of TrkA protein expression improving trkA promoter activity [84]. It has been also reported as a tumor suppressor in numerous cancer [85]. Research indicated that BEX3 is related to cisplatin sensitivity of nasopharyngeal carcinoma [86].

COX6C is part of the enzyme of the mitochondrial respiratory chain and plays a role in the electron transfer catalysis of cytochrome c reduced to oxygen. According to Lomvart V. et al. identified and validated that COX6C is related to brain ischemia and the expression of this protein increases in the ischemic tissue compared to control tissue [87]. COX6C expression was found downregulated in postmortem tissue of patients with bipolar disorder [88].

PFN1 has a role different cellular process and interact with other proteins to develop neurological disease. This gene has been related to neurodegenerative disease including amyotrophic lateral sclerosis [89] [90]. PFN1 was found associated with fragile x syndrome being the most common cause for ASD. This research showed that levels of PFN1 start to reduce after postnatal day [91].

HSPA8 is a heat shock protein that acts as a chaperone, which helps the folding of other proteins newly formed [92]. HSPA8 was detected downregulated across brain tissues samples from Alzheimer disease that includes entorhinal, auditory cortices and hippocampus [93]. Also, wasidentified that methylation affects the function of HSPA8 in Parkinson disease [94].

Calmodulin (CALM1, CALM2, CALM3) is a family of proteins that performs multiple functions through calcium binding. A study suggested that the lack of balance (calcium homeostasis in cells does not occur) of calcium levels in cells leads to neural deterioration in neurodegenerative ailment [95]. Studies indicate that alterations in the calcium signaling systems that regulate many neural activities have been related to psychiatric diseases including SCZ and bipolar disorder [96] [97].

Ribosomes family (RPL13A, RPL13A5, RPL34, RPS6, RPS8, RPS11, RPS25) are cellular organelles the main function of which is to catalyze protein synthesis. Ribosomal proteins play various roles, either increasing immune signaling or helping pathogen production. For example, RPS25 helps in the translation viral transcripts process [98].

SNORD protein family (SNORD14C, SNORD14D, SNORD35B, SNORD33, SNORD34, SNORD35A, SNORD32A) is a small nucleolar RNAs that takes a role in gene expression regulation in the human cell.

In summary, the genes found in the first four Pareto efficient frontiers are related to the immune system, metals, and oxidative stresses as part of the biological processes. Also, the genes that were related to SCZ and ASD in the literature were listed. These findings are summarized in Table 3. The genes found through the MCO method should be studied to see how they could be related to both diseases.

Table 3 Summary of the Findings Through Literature

	Oxidative Stress	Immune System	Metals	SCZ	ASD
HBA1	X		X	X	
HBA2	X		X	X	
HBB	X		X	X	X
MT1X	X		X	X	
MT2A	X		X	X	
MT1H	X		X	X	
MT1HL1	X		X		
TXNIP	X			X	
IFITM2		X			
IFITM3		X			
B2M		X			
PTGDS				X	
NRGN			X	X	
HSPA8					
EEF1A1		X			
FTL	X		X		
FCGR3A FCGR3B		X			
TMSB4X					
TUBA1A					
BEX3					
COX6C					
PFN1					X
HSPA8					

4.3 Most Correlated Structures

This section discusses the results obtained through the Traveling Salesman Problem (TSP) and Minimum Spanning Tree (MST) to build the most correlated structures among genes. This was done using the first four Pareto efficient frontiers obtained with the MCO because there were changes in the MATLAB version, obtaining difficulties to solve the TSP with five frontiers (36 genes). Afterwards, the genes were divided per group to identify to which pathway they are related to using Reactome Pathway Analysis [99].

In order to build the most correlated cyclic and non-cyclic structure among genes, the Pareto efficient solutions obtained in the first four frontiers (25 genes) was used. After obtaining these genes, the TSP and MST methods were applied for the construction of the most correlated

cyclic and non- cyclic structure respectively. There is a total of $(25-1)! \approx 6.20448 \times 10^{23}$ ways in which a cyclic structure can be built among the 25 genes. The optimal cyclic structure is shown in Figure 13. Next, the non-cyclic structure using the MST method with the same matrix used for the TSP. The optimal non-cyclic structure is shown in Figure 14.

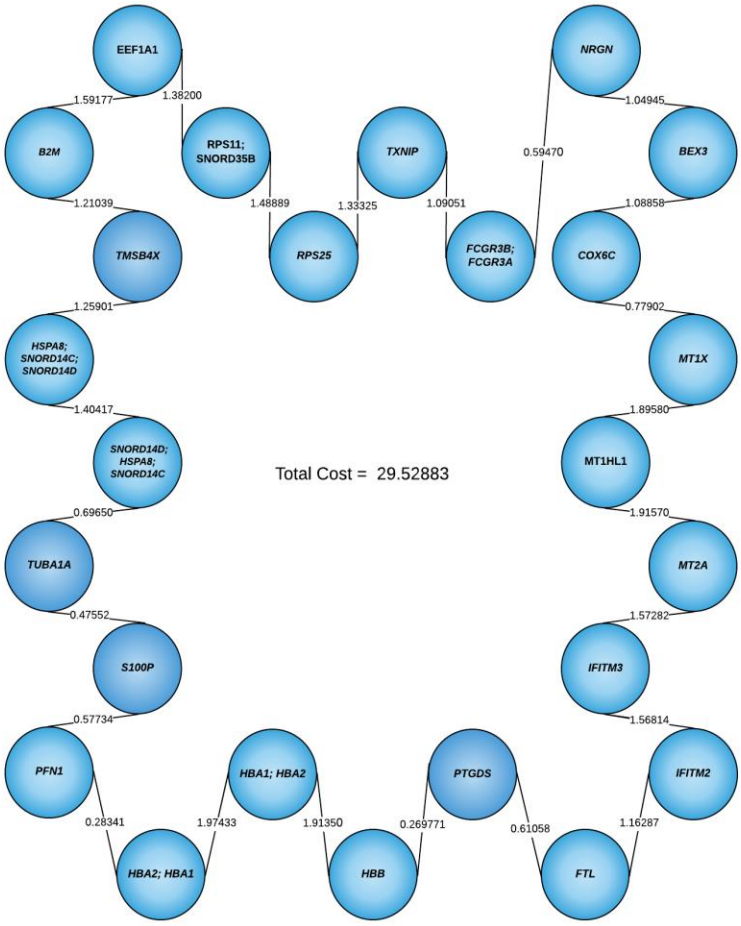

Figure 13 Most Correlated Cyclic Structure

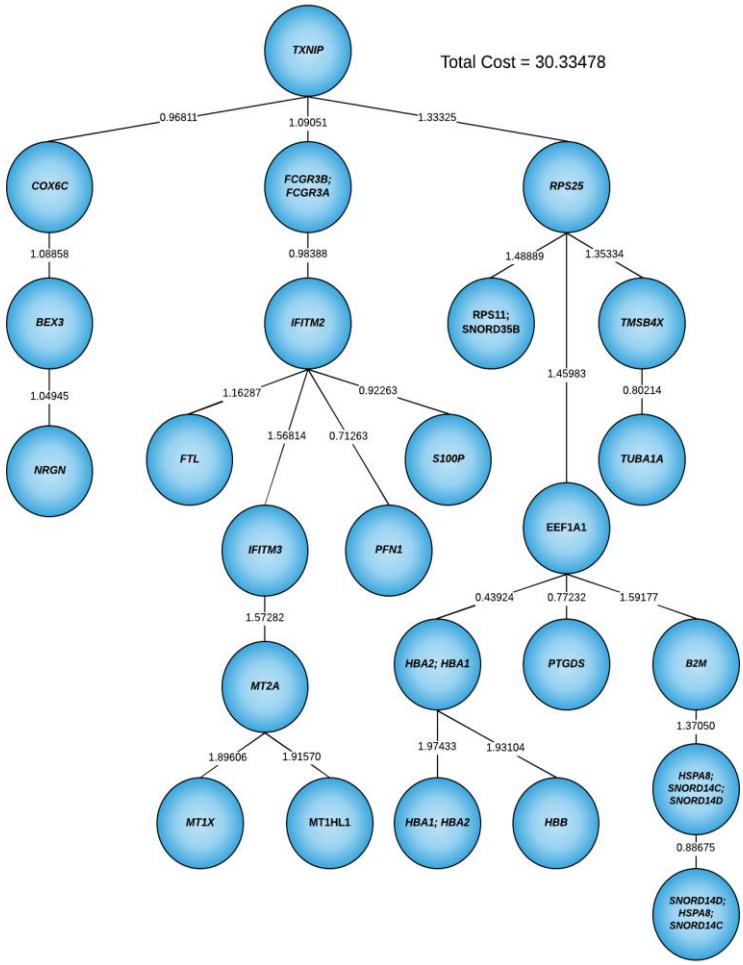

Figure 14 Most Correlated Non-Cyclic Structure

Subsequently, sequential patterns were observed between structures and the genes were divided into six groups shown in Figure 15. The groups are composed of the following genes: Group 1 (RPS11-SNORD35B, RPS25, TXNIP, and FCGR3B-FCGR3A), Group 2 (NRGN, BEX3, and COX6C), Group 3 (MT1X, MT1HL1, MT2A, IFITM3, IFITM2 , and FTL), group 4 (HBB, HBA1-HBA2, AND HBA2-HBA1), Group 5 (SNORD14D-HSPA8-SNORD14C, and HSPA8-SNORD14C-SNORD14D), and Group 6 (B2M and EEF1A1). No pattern was found for TMSB4X, TUBA1A, S100P, PFN1, and PTGDS between the network structures. In the literature it was found that part of group 2 (MT2A, IFITM3, IFITM2) was found overexpressed in SCZ

patients compared to controls. For the other groups, no study was found linking the genes to ASD and/or SCZ.

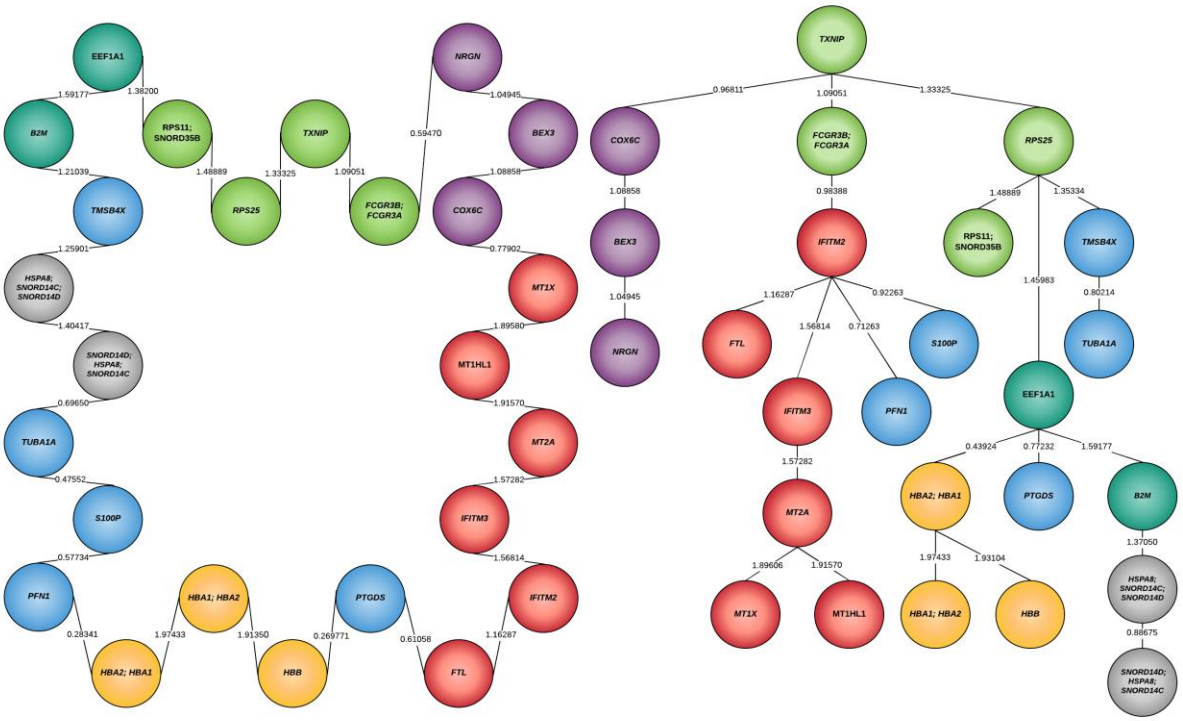

Figure 15 Genes Pattern in the Most Correlated Structures

Later on, the task was to identify to which pathways the groups belonged. For this, the Reactome Pathway Analysis package for R software was used [99]. With this package, it was possible to carry out the gene set enrichment analysis with the groups previously mentioned. As an outcome of the enrichment analysis, the pathway location, p-value, and the adjusted p-value were obtained. For this analysis, the hypergeometric distribution test was used to calculate the p-value and the adjusted p-value the Benjamini-Hochberg procedure (Appendix A). After obtaining the pathway locations, it became the objective to identify the global pathway to which they belonged. These results are summarized from Table 4 to Table 9 for each group.

Table 4 Group 1(RPS11-SNORD35B, RPS25, TXNIP, and FCGR3B-FCGR3A) Pathways Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Developmental Biology (Homo sapiens)	R-HSA-9010553	Regulation of expression of SLITs and ROBOs	2.553E-03	4.441E-03	RPS11/RPS25
	R-HSA-376176	Signaling by ROBO receptors	4.108E-03	6.085E-03	RPS11/RPS25
Disease (Homo sapiens)	R-HSA-192823	Viral mRNA Translation	7.063E-04	2.680E-03	RPS11/RPS25
	R-HSA-168273	Influenza Viral RNA Transcription and Replication	1.558E-03	3.117E-03	RPS11/RPS25
	R-HSA-168255	Influenza Life Cycle	1.797E-03	3.423E-03	RPS11/RPS25
	R-HSA-168254	Influenza Infection	2.079E-03	3.780E-03	RPS11/RPS25
	R-HSA-5663205	Infectious disease	1.220E-02	1.435E-02	RPS11/RPS25
Immune System (Homo sapiens)	R-HSA-2029481	FCGR activation	5.669E-03	7.820E-03	FCGR3A
	R-HSA-844456	The NLRP3 inflammasome	5.669E-03	7.820E-03	TXNIP
	R-HSA-622312	Inflammasomes	8.024E-03	1.035E-02	TXNIP
	R-HSA-2029485	Role of phospholipids in phagocytosis	1.178E-02	1.428E-02	FCGR3A
	R-HSA-168643	Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	2.438E-02	2.787E-02	TXNIP
	R-HSA-2029482	Regulation of actin dynamics for phagocytic cup formation	2.855E-02	3.173E-02	FCGR3A
	R-HSA-2029480	Fcgamma receptor (FCGR) dependent phagocytosis	4.007E-02	4.331E-02	FCGR3A
Metabolism (Homo sapiens)	R-HSA-2408557	Selenocysteine synthesis	7.703E-04	2.680E-03	RPS11/RPS25
	R-HSA-2408522	Selenoamino acid metabolism	1.231E-03	2.680E-03	RPS11/RPS25
	R-HSA-71291	Metabolism of amino acids and derivatives	1.147E-02	1.428E-02	RPS11/RPS25
Metabolism of proteins (Homo sapiens)	R-HSA-72695	Formation of the ternary complex, and subsequently, the 43S complex	2.447E-04	2.680E-03	RPS11/RPS25
	R-HSA-72649	Translation initiation complex formation	3.139E-04	2.680E-03	RPS11/RPS25
	R-HSA-72702	Ribosomal scanning and start codon recognition	3.139E-04	2.680E-03	RPS11/RPS25
	R-HSA-72662	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	3.245E-04	2.680E-03	RPS11/RPS25
	R-HSA-156902	Peptide chain elongation	7.063E-04	2.680E-03	RPS11/RPS25
	R-HSA-156842	Eukaryotic Translation Elongation	7.703E-04	2.680E-03	RPS11/RPS25
	R-HSA-72764	Eukaryotic Translation Termination	7.703E-04	2.680E-03	RPS11/RPS25
	R-HSA-72689	Formation of a pool of free 40S subunits	9.241E-04	2.680E-03	RPS11/RPS25
	R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	1.111E-03	2.680E-03	RPS11/RPS25
	R-HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	1.111E-03	2.680E-03	RPS11/RPS25
	R-HSA-72706	GTP hydrolysis and joining of the 60S ribosomal subunit	1.131E-03	2.680E-03	RPS11/RPS25
	R-HSA-72613	Eukaryotic Translation Initiation	1.273E-03	2.680E-03	RPS11/RPS25
	R-HSA-72737	Cap-dependent Translation Initiation	1.273E-03	2.680E-03	RPS11/RPS25
	R-HSA-72766	Translation	7.257E-03	9.676E-03	RPS11/RPS25
	R-HSA-163125	Post-translational modification: synthesis of GPI-anchored proteins	4.327E-02	4.555E-02	FCGR3B
Metabolism of RNA (Homo sapiens)	R-HSA-975956	Nonsense Mediated Decay (NMD) independent of the Exon Junction Complex (EJC)	8.033E-04	2.680E-03	RPS11/RPS25
	R-HSA-927802	Nonsense-Mediated Decay (NMD)	1.171E-03	2.680E-03	RPS11/RPS25
	R-HSA-975957	Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)	1.171E-03	2.680E-03	RPS11/RPS25
	R-HSA-6791226	Major pathway of rRNA processing in the nucleolus and cytosol	2.979E-03	4.965E-03	RPS11/RPS25
	R-HSA-8868773	rRNA processing in the nucleus and cytosol	3.303E-03	5.285E-03	RPS11/RPS25
	R-HSA-72312	rRNA processing	3.642E-03	5.604E-03	RPS11/RPS25

Table 5 Group 2 (NRGN, BEX3, and COX6C) Pathway Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Gene expression (Transcription) (Homo sapiens)	R-HSA-5628897	TP53 Regulates Metabolic Genes	1.622E-02	3.537E-02	COX6C
Metabolism (Homo sapiens)	R-HSA-611105	Respiratory electron transport	1.885E-02	3.537E-02	COX6C
	R-HSA-163200	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	2.316E-02	3.537E-02	COX6C
	R-HSA-1428517	The citric acid (TCA) cycle and respiratory electron transport	3.250E-02	3.714E-02	COX6C
Signal Transduction (Homo sapiens)	R-HSA-204998	Cell death signalling via NRAGE, NRIF and NADE	1.434E-02	3.537E-02	BEX3
	R-HSA-193704	p75 NTR receptor-mediated signalling	1.829E-02	3.537E-02	BEX3
	R-HSA-73887	Death Receptor Signalling	2.653E-02	3.537E-02	BEX3

Table 6 Group 3 (MT1X, MT1HL1, MT2A, IFITM3, IFITM2, and FTL) Pathway Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Cellular responses to external stimuli (Homo sapiens)	R-HSA-5661231	Metallothioneins bind metals	9.847E-06	9.769E-05	MT2A/MT1X
	R-HSA-5660526	Response to metal ions	1.628E-05	9.769E-05	MT2A/MT1X
Immune System (Homo sapiens)	R-HSA-913531	Interferon Signaling	6.125E-05	2.450E-04	IFITM2/IFITM3/MT2A
	R-HSA-909733	Interferon alpha/beta signaling	4.154E-04	1.246E-03	IFITM2/IFITM3
	R-HSA-877300	Interferon gamma signaling	4.281E-02	4.670E-02	MT2A
Transport of small molecules (Homo sapiens)	R-HSA-917937	Iron uptake and transport	2.716E-02	4.036E-02	FTL
Vesicle-mediated transport (Homo sapiens)	R-HSA-3000480	Scavenging by Class A Receptors	8.965E-03	2.152E-02	FTL
	R-HSA-2173782	Binding and Uptake of Ligands by Scavenger Receptors	1.973E-02	3.946E-02	FTL
	R-HSA-432722	Golgi Associated Vesicle Biogenesis	2.624E-02	4.036E-02	FTL
	R-HSA-199992	trans-Golgi Network Vesicle Budding	3.363E-02	4.036E-02	FTL
	R-HSA-421837	Clathrin derived vesicle budding	3.363E-02	4.036E-02	FTL

Table 7 Group 4 (HBA1, HBA2, HBB) Pathway Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Transport of small molecules (Homo sapiens)	R-HSA-1237044	Erythrocytes take up carbon dioxide and release oxygen	1.457E-09	2.914E-09	HBA1/HBA2/HBB
	R-HSA-1480926	O2/CO2 exchange in erythrocytes	1.457E-09	2.914E-09	HBA1/HBA2/HBB
Vesicle-mediated transport (Homo sapiens)	R-HSA-2168880	Scavenging of heme from plasma	1.457E-09	2.914E-09	HBA1/HBA2/HBB
	R-HSA-2173782	Binding and Uptake of Ligands by Scavenger Receptors	5.849E-08	8.774E-08	HBA1/HBA2/HBB

Table 8 Group 5(SNORD14D-HSPA8-SNORD14C, and HSPA8-SNORD14C-SNORD14D) Pathway Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Cellular responses to external stimuli (Homo sapiens)	R-HSA-2262752	Cellular responses to stress	4.053E-02	4.405E-02	HSPA8
	R-HSA-3371568	Attenuation phase	1.326E-03	1.420E-02	HSPA8
	R-HSA-3371571	HSF1-dependent transactivation	2.273E-03	1.420E-02	HSPA8
	R-HSA-3371497	HSP90 chaperone cycle for steroid hormone receptors (SHR)	5.208E-03	1.420E-02	HSPA8
	R-HSA-3371453	Regulation of HSF1-mediated heat shock response	6.439E-03	1.420E-02	HSPA8
	R-HSA-3371556	Cellular response to heat stress	8.333E-03	1.488E-02	HSPA8
Developmental Biology (Homo sapiens)	R-HSA-373760	L1CAM interactions	1.127E-02	1.761E-02	HSPA8
Immune System (Homo sapiens)	R-HSA-6785807	Interleukin-4 and 13 signaling	1.023E-02	1.704E-02	HSPA8
	R-HSA-449147	Signaling by Interleukins	4.394E-02	4.536E-02	HSPA8
	R-HSA-6798695	Neutrophil degranulation	4.536E-02	4.536E-02	HSPA8
Metabolism of proteins (Homo sapiens)	R-HSA-8876725	Protein methylation	1.799E-03	1.420E-02	HSPA8
Metabolism of RNA (Homo sapiens)	R-HSA-450408	AUF1 (hnRNP D0) binds and destabilizes mRNA	5.303E-03	1.420E-02	HSPA8
	R-HSA-450531	Regulation of mRNA stability by proteins that bind AU-rich elements	8.333E-03	1.488E-02	HSPA8
	R-HSA-72163	mRNA Splicing - Major Pathway	1.733E-02	2.380E-02	HSPA8
	R-HSA-72172	mRNA Splicing	1.809E-02	2.380E-02	HSPA8
	R-HSA-72203	Processing of Capped Intron-Containing Pre-mRNA	2.301E-02	2.739E-02	HSPA8
Neuronal System (Homo sapiens)	R-HSA-888590	GABA synthesis, release, reuptake and degradation	1.799E-03	1.420E-02	HSPA8
	R-HSA-112310	Neurotransmitter release cycle	4.829E-03	1.420E-02	HSPA8
	R-HSA-112315	Transmission across Chemical Synapses	2.140E-02	2.675E-02	HSPA8
	R-HSA-112316	Neuronal System	3.494E-02	3.970E-02	HSPA8
Vesicle-mediated transport (Homo sapiens)	R-HSA-432720	Lysosome Vesicle Biogenesis	3.314E-03	1.420E-02	HSPA8
	R-HSA-432722	Golgi Associated Vesicle Biogenesis	5.303E-03	1.420E-02	HSPA8
	R-HSA-199992	trans-Golgi Network Vesicle Budding	6.818E-03	1.420E-02	HSPA8
	R-HSA-421837	Clathrin derived vesicle budding	6.818E-03	1.420E-02	HSPA8
	R-HSA-8856828	Clathrin-mediated endocytosis	1.326E-02	1.949E-02	HSPA8

Table 9 Group 6(B2M and EEF1A1) Pathway Analysis Results. Source: Reactome PA

Pathway	ID	Pathway Location	p_value	adjusted p_value	geneID
Cellular responses to external stimuli (Homo sapiens)	R-HSA-3371511	HSF1 activation	2.271E-03	1.702E-02	EEF1A1
	R-HSA-3371556	Cellular response to heat stress	1.660E-02	2.827E-02	EEF1A1
Disease (Homo sapiens)	R-HSA-164938	Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters	3.784E-03	1.702E-02	B2M
	R-HSA-164952	The role of Nef in HIV-1 replication and disease pathogenesis	5.107E-03	1.702E-02	B2M
	R-HSA-162909	Host Interactions of HIV factors	2.409E-02	3.312E-02	B2M
Immune System (Homo sapiens)	R-HSA-6798695	Neutrophil degranulation	2.053E-03	1.702E-02	B2M/EEF1A1
	R-HSA-1236977	Endosomal/Vacuolar pathway	2.082E-03	1.702E-02	B2M
	R-HSA-983170	Antigen Presentation: Folding, assembly and peptide loading of class I MHC	4.729E-03	1.702E-02	B2M
	R-HSA-2424491	DAP12 signaling	5.673E-03	1.702E-02	B2M
	R-HSA-2172127	DAP12 interactions	8.127E-03	2.167E-02	B2M
	R-HSA-1236974	ER-Phagosome pathway	1.566E-02	2.827E-02	B2M
	R-HSA-877300	Interferon gamma signaling	1.735E-02	2.827E-02	B2M
	R-HSA-1236975	Antigen processing-Cross presentation	1.866E-02	2.827E-02	B2M
	R-HSA-198933	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	2.484E-02	3.312E-02	B2M
	R-HSA-913531	Interferon Signaling	3.677E-02	4.645E-02	B2M
Metabolism of proteins (Homo sapiens)	R-HSA-8876725	Protein methylation	3.595E-03	1.702E-02	EEF1A1
	R-HSA-156902	Peptide chain elongation	1.697E-02	2.827E-02	EEF1A1
	R-HSA-156842	Eukaryotic Translation Elongation	1.772E-02	2.827E-02	EEF1A1
	R-HSA-977225	Amyloid fiber formation	1.885E-02	2.827E-02	B2M

Subsequently, it was observed that the groups have common pathways shown in Table 10. The pathway represented by most of the groups was the Immune System. On the other hand, the group that had fewer pathways in common with other groups was Group 2. These pathways were added to the correlated structures to see how they are related to each group depending on the structure. In the cyclic structure, there were three groups (Group 1, Group 5, and Group 6) that share three pathways: immune system, disease, and metabolism of proteins. With the information found through the literature provided in Section 4.2, B2M (Group 6) has been related to inflammation, immunological deficiency, and immune homeostasis [60] [61]. The RNA virus replication is one of the EEF1A1 functions [75] [76] (Group 6) while RPS25 helps in the initial phase of virus transcription. FCGR3A (Group 1) has been found associated with systemic lupus disease, which is an immunological disorder [78]. With these details obtained about the genes can be inferred that could be linked with the immune system and at the same time Group 1 and Group 6 had in common Immune System Pathway.

Table 10 Pathways Summary per Group

Pathway	Group 3	Group 6	Group 5	Group 1	Group 4	Group 2
Immune System	■	■	■	■		
Metabolism of proteins		■	■	■		
Developmental Biology			■	■		
Metabolism of RNA			■	■		
Vesicle-mediated transport	■		■		■	
Metabolism				■		■
Disease		■		■		
Cellular responses to external stimuli	■	■	■			
Transport of small molecules	■				■	
Neuronal System			■			
Gene expression (Transcription)						■
Signal Transduction						■

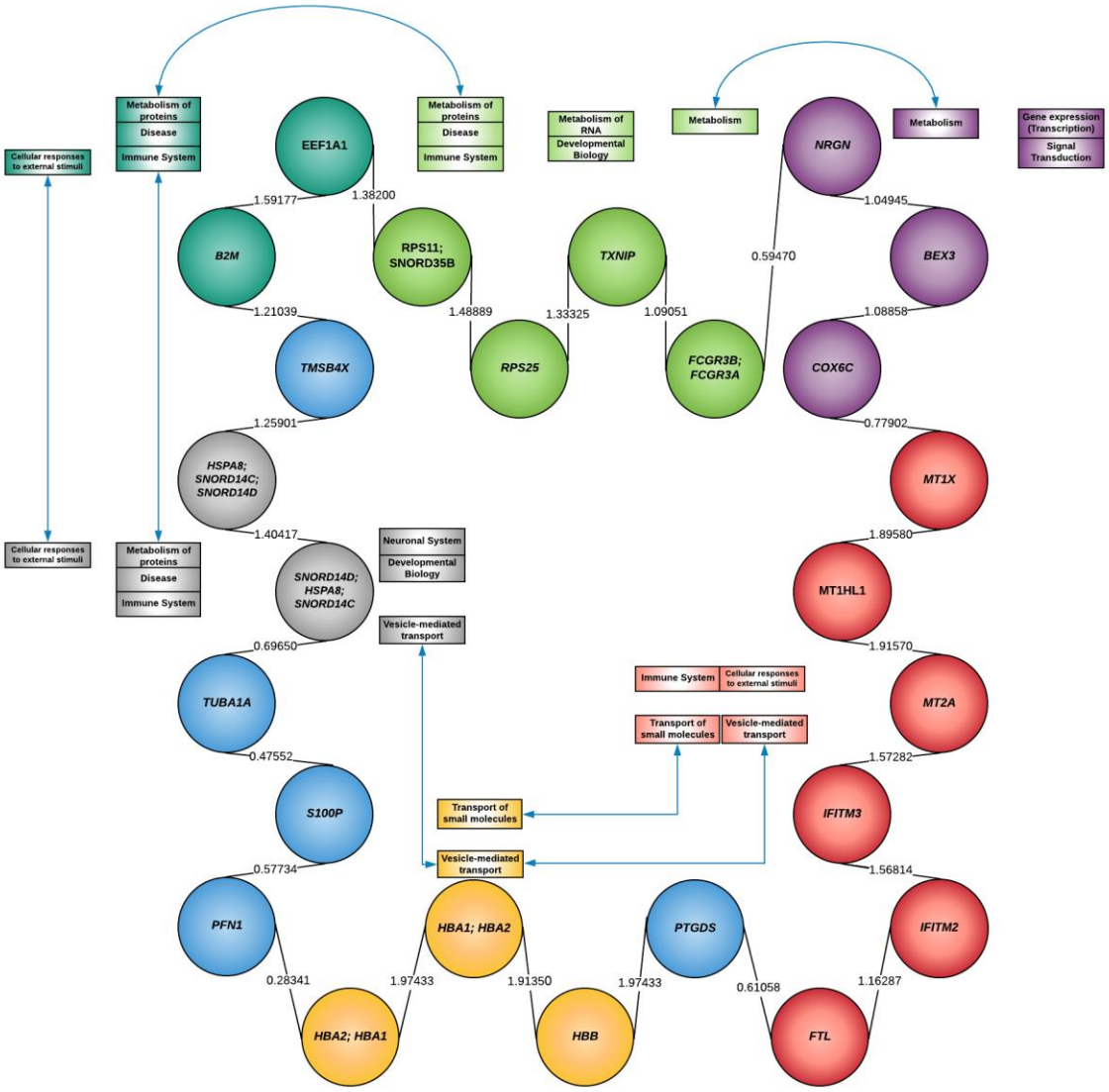

Figure 16 Pathways Relationship Among Group of Genes in Cyclic Structure

Analyzing the non-cycle structure with the pathways obtained for each group shown in Figure 17 it can be seen that three groups (Group 1, Group 6, and Group 5), which are connected, have in common the pathways of Immune System, Disease, and Metabolism of proteins. Otherwise, four groups (Group 1, Group 3, Group 6, and Group 5) have in common the Immune System pathway while in the cycle structure only three groups had in common the Immune System Pathway. IFITM2, IFITM3 (Group 3) has been found in the literature as helping to prevent viral

replication and inflammation [55]. With the literature found in this thesis, Group 1, Group 3, and Group 6 are linked with the immune system as they also belong to the Immune System pathway.

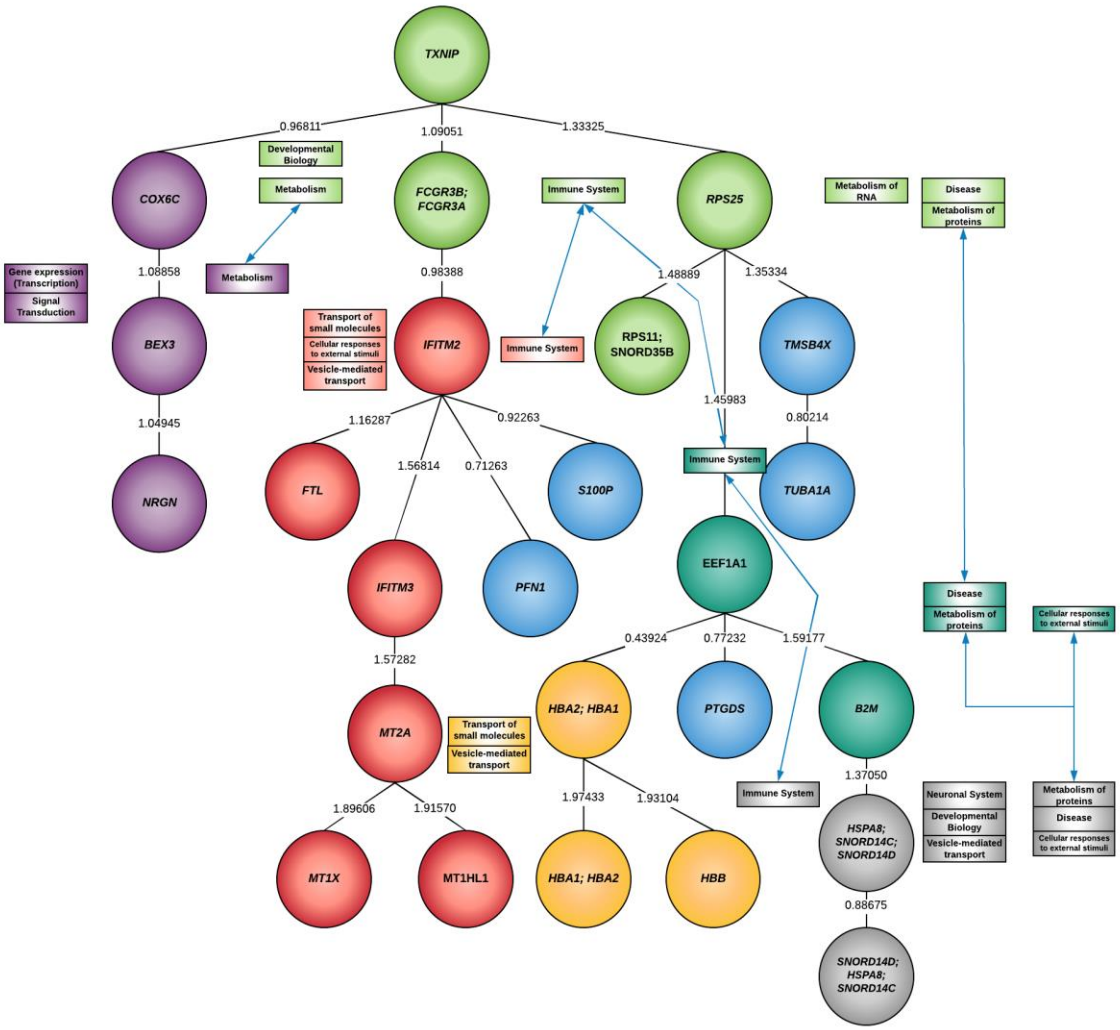

Figure 17 Pathways Relationship Among Group of Genes in Non-Cyclic Structure

As a summary, through this analysis, it was possible to construct the most correlated structures and at the same time identify patterns through them. These patterns were divided into groups and it was observed that Group 1, Group 6, and Group 5 had in common the same pathways (Metabolism of proteins, Immune System, and Disease) in both structures. On the other hand, in the non-cyclic structure, it was observed that Group 1, Group 3, Group 6, and Group 5 have in

common the Immune System pathway. It was concluded that the immune system could play an important role in the development of ASD and SCZ.

.

Chapter 5 Cost Model for ASD and SCZ

This chapter aims to analyze the cost associated with Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ) per patient and create a cost model that estimates the annual cost per patient. To perform the analysis and create the cost models, we used data from the literature where the annual cost for each disease was summarized. Within the literature the cost classification has been identified as follows: Direct Medical Costs, Direct Non-Medical Costs, and Indirect Costs. Table 11 shows the types of costs and the classification in which they belong.

A systematic review by Chong H.Y. [99] was used to analyze the schizophrenia patient cost. Within this review, only three articles were found which estimate schizophrenia cost in the United States [100] [101] [102]. These global estimates were adjusted per patients using the population obtained for the direct and indirect costs (2.9 and 3.03 million patients respectively) [103]. Other SCZ costs were obtained from Miller B. [104], in which 3.5 million of people with SCZ was estimated with a global cost of $37.7 billion of direct medical cost, $9.3 billion of direct non-medical cost and $117.3 billion of indirect cost. With these data, annual costs per patient with SCZ were estimated as shown in Table 12.

In the literature search, two articles were found that estimated the direct (medical and non-medical) and indirect costs for patients with ASD in the United States. Buescher A. V.S et al. estimated the direct and indirect costs per patient with different age, and, with and without intellectual disability [105]. For these costs, the median was calculated for our analysis because is not influenced by extreme values that could affect drastically the cost due to its variability. Gantz M. research estimated ASD cost with a hypothetical incident child born in 2000 and diagnosed 2003 [106]. Table 13 shows the ASD cost obtained in the studies mentioned previously.

Table 11 Type of Cost Found in the Literature

Type of Cost	ASD	SCZ
Direct Medical Cost		
Inpatient and outpatient care (hospital and physician)	X	X
Emergency	X	X
community-based care		X
home care	X	X
long-term institutional care	X	X
rehabilitation care		X
specialist and other professional health care	X	X
diagnostic tests	X	X
pharmacy	X	X
medical supplies	X	X
Direct Non-Medical Cost		
transportation	X	X
food	X	X
lodging incurred during health care visit	X	X
special education	X	
treatment	X	
social services		X
child care	X	
special programs	X	
after school	X	
day care	X	
summer school programs	X	
weekend programs	X	
home care modifications	X	
Indirect Cost		
productivity loss	X	

Table 12 SCZ Annual Cost per Patient

	Annual Global Cost estimates (USD, millions)				Annual Cost estimates per patient (USD)			
Reference	**Direct medical cost**	**Direct nonmedical cost**	**Indirect cost**	**Total**	**Direct medical cost**	**Direct nonmedical cost**	**Indirect cost**	**Total**
[100]	$ 29,279	$ 12,014	$ 41,714	$ 83,007	$ 10,237	$ 4,200	$ 14,585	$ 29,023
[101]	$ 27,745	$ 4,054	$ 70,597	$ 102,396	$ 9,701	$ 1,417	$ 24,684	$ 35,802
[103]	$ 12,078	$ 57	$ 48,200	$ 60,335	$ 4,223	$ 19	$ 16,853	$ 21,096
[104]	$ 37,700	$ 9,300	$ 117,300	$ 164,300	$ 10,771	$ 2,657	$ 33,514	$ 46,942

Table 13 ASD Annual Cost per Patient

Annual Cost Estimates per Patient (USD, millions)				
Reference	**Direct medical cost**	**Direct nonmedical cost**	**Indirect cost**	**Total**
[105]	$ 15,681	$ 8,540	$ 71,989	$ 96,210
[106]	$ 305,956	$ 978,761	$ 1,875,667	$ 3,160,384

After obtaining the cost of ASD and SCZ per patient, a mathematical model to estimate cost using the previous data was created. First the proportions were obtained for each cost classification for ASD and SCZ shown in Table 14 and Table 15. The proportions are the percentages that each cost classification contributes to the total cost of the patient. After obtaining these proportions in this study, coefficients were obtained to be used in the cost model to estimate the annual cost per patient.

Table 14 Proportions of SCZ Cost

Reference	**Direct medical cost**	**Direct nonmedical cost**	**Indirect cost**	**Coefficients** (1+ Indirect cost proportion)
[100]	35.3%	14.5%	50.3%	1.50
[101]	27.1%	4.0%	68.9%	1.69
[103]	20.0%	0.1%	79.9%	1.80
[104]	22.9%	5.7%	71.4%	1.71
Median	25.0%	4.8%	70.2%	1.70

Table 15 Proportions of ASD Cost

References	**Direct medical costs**	**Direct nonmedical costs**	**Indirect costs**	**Coefficients** (1+ Indirect cost proportion)
[105]	9.7%	31.0%	59.3%	1.59
[106]	16.3%	8.9%	74.8%	1.75
Median	13.0%	19.9%	67.1%	1.67

After observing how the proportions of the indirect cost vary, the coefficient selected for the indirect cost is 1.7 for ASD and SCZ. This coefficient means that indirect cost is 1.7 of the direct cost. The coefficient of the indirect cost can vary between 1.59 - 1.8 for SCZ and 1.59 – 1.75 for ASD. The mathematical cost models to estimate the conditions' cost are as follows:

$$SCZCost = DMC_{SCZ} + DNMC_{SCZ} + 1.7\ (\ DMC_{SCZ} + DNMC_{SCZ}) \quad (6.1)$$

$$ASDCost = DMC_{ASD} + DNMC_{ASD} + 1.7\ (\ DMC_{ASD} + DNMC_{ASD}) \quad (6.2)$$

Where DMC_{SCZ} and $DNMC_{SCZ}$ represent the direct medical and direct non-medical cost for SCZ respectively. DMC_{ASD} and $DNMC_{ASD}$ represent direct medical and direct non-medical cost for ASD.

To analyze the cost of misdiagnosing a patient due to the symptom similarities between them, the information provided previously was used. For this analysis, the median of the total cost of SCZ per patient was used to obtain a total annual cost of $32,413. For ASD the total annual cost of $96,210 per patient was used. For this instance, ASD cost is higher than that SCZ due to the indirect cost. For the misdiagnosis cost analysis, assumptions were made for the age of diagnosing and the average time of life. The age of diagnosing ASD and SCZ was of 5 and 16 years old respectively, and the average of time of life used was 75 years. Our objective in this analysis is to evaluate what will be the cost of living with a wrong diagnosis for ten years. Starting with ASD diagnosed at five years old and later having the second diagnose as SCZ at 15 years old. Then for SCZ having the first diagnosed at 16 years old and the second as ASD at 26 years old. Figure 18 shows this information to visualize how cost changes over the years.

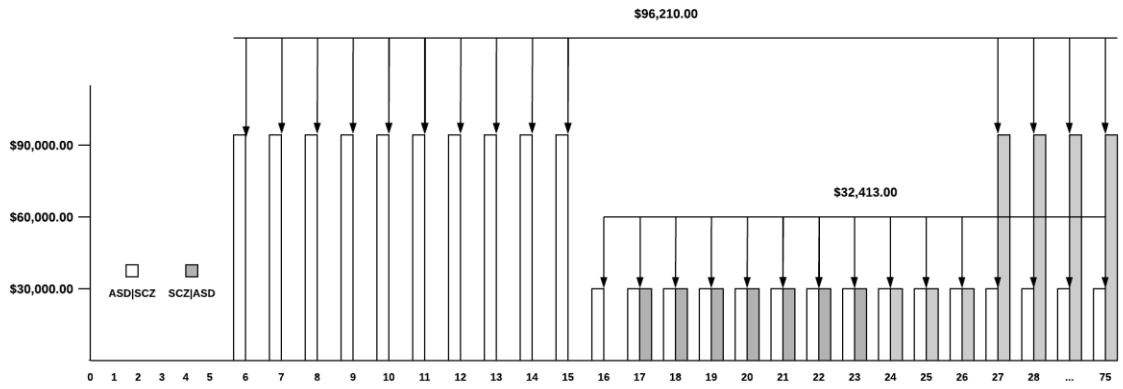

Figure 18 ASD and SCZ Annuity Cash Flow

Subsequently, it was decided to make a cost comparison over time using the present and future value of the cost. The present and future value can be calculated using the annuity, the number of periods, and the interest over time, which in this case, it was the inflation. These values were used to determine which of the cases of misdiagnosis is more expensive. To estimate the present value and future value of the different cases of diagnosis the annual cost, inflation, when it was diagnosed, the average time of life, and the years with the condition were part of the input. The present value represents the cost of diagnosis at the current moment. The future value represents the total cost spent up on the average lifetime. Table 16 shows all the values used to calculate the Future and Present Value. After obtaining the results, it can be seen that Case 1 (SCZ) had a present value of $1.09 million and a future value of $ 3.7 million. Case 2 (ASD) had a present value of $3.5 million and future value of 15.04 million. Evaluating these two cases, we can say that ASD cost is approximately three times higher than SCZ for the present value and four times higher for the future value. Case 3 and Case 4 are the ones that the first diagnosis is wrong and the second one is correct. In other words, we could say it is diagnosed with ASD given it has

SCZ and vice versa. In this Case 4 (SCZ|ASD) cost is higher because the patient will live with ASD for 49 years knowing that ASD is more expensive than SCZ. In Figure 19 it summarizes the cost of misdiagnosing at the present time.

Table 16 Diagnosis Cost Over Time

	Case 1	Case 2		Case 3		Case 4	
	SCZ	**ASD**		**ASD | SCZ**		**SCZ | ASD**	
Annual Cost	$ 32,413	$ 96,210	Annual Cost	$ 96,210	$ 32,413	$ 32,413	$ 96,210
Inflation	2.10%	2.10%	Inflation	2.10%	2.10%	2.10%	2.10%
Diagnosed at	16	5	Diagnosed at	5	15	16	26
Average Life	75	75	2nd Diagnosed at - Average Life	15	75	26	75
Years with the condition	59	70	Years with the condition	10	60	10	49
Future Value	**$ 3,716,923**	**$ 15,043,264**	**Future Value**	$ 1,058,301	$ 3,827,391	$ 356,540	$ 8,102,748
Present Value	**$ 1,090,598**	**$ 3,511,883**	**Present Value**	$ 859,709	$ 1,099,912	$ 289,634	$ 2,926,651
			Total Future Value	**$7,509,984**		**$9,089,867**	
			Total Present Value	**$1,753,221**		**$2,667,110**	

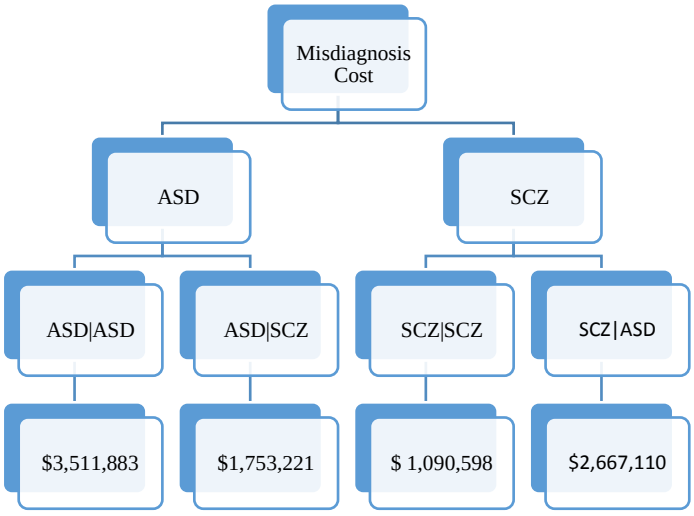

Figure 19 Estimated Cost of Misdiagnosing a Patient at Present Time

Indirect costs tend to have high variability because they can vary depending on the severity of the condition causing productivity loss either to the patient or to the parents that take care of the patient. After analyzing the annual cost per patient for all cases, it can be concluded that ASD is a costly condition due to the indirect cost compared with SCZ. It was proceeded to estimate the misdiagnosis cost due to the similarities that could lead to such wrong assessment. As an outcome

of the misdiagnosis estimation cost it is obtained that the first scenario with higher cost is Case 2 (diagnosed with ASD given he/she has ASD) and the second is Case 4 (diagnosed with SCZ given he/she has ASD). It is very important to be aware of the diagnosis decision because a misdiagnosis causes a financial effect on the patient for an extended period. For future work, it is recommended to do more research to understand better how these conditions affect family economics emphasizing ASD.

Chapter 6 Conclusions

Disease characterization is important because it helps to understand illness behavior and thus it leads to potentially prevent its development and improve treatment. In the literature it has been found that Autism Spectrum Disorder (ASD) and Schizophrenia (SCZ) share symptom similarities that could lead to misdiagnosing [6] [7]. In this direction, our first results of ASD through microarray analysis showed that two genes were also related with SCZ. Because of this, it was proposed to perform a characterization of these diseases simultaneously at the genetic level.

The main contribution of this work is the characterization of ASD and SCZ through simultaneous microarray analysis. This characterization identifies the differentially expressed genes in common in ASD and SCZ, the construction of the most correlated cyclic and non-cyclic structure among the differentially expressed genes, a cost model, a and cost analysis of misdiagnosing a patient. Through the biological evidence, it has been found that metals (iron, zinc, and copper) and the immune system might play a role in the development of ASD and SCZ. Also, it was concluded that through the pathways analysis with most correlated structures that most of the groups were related with Immune System pathway. It is suggested that these findings should be evaluated or validated with biological research to identify the causes that lead to the development of these ailments.

The ASD and SCZ Cost Models and Cost Analyses were done to provide a tool to estimate the cost associated with each disease and at the same time being able to identify the cost of a wrong diagnosis. The cost estimates obtained through the literature for each disease were used to compute the annual cost per patient and misdiagnosis cost. The annual cost obtained for ASD was of $96,210 and for SCZ $32,413 per patient. The cost of misdiagnosing was higher for a patient with SCZ given that ten years later is diagnosed correctly with ASD.

In conclusion, consistency and repeatability in optimization-driven analysis are further evidenced in this work for the analysis of microarrays. The differentially expressed genes identified in this work might help to understand the similarities between ASD and SCZ. The immune system and metals should be investigated in the context of the genesis of these diseases. The TSP and MST helped to build the most correlated structures among the genes previously identified. As a result, most of the groups used to explore such structure were related to the immune system pathway.

For future work, it is proposed to construct the most correlated structures with TSP and MST for additional Pareto efficient frontiers for exploration purposes. With these results, it might be possible to identify other different pathways related to ASD and SCZ. Future recommendations could include the addition of environmental and clinical variables to the analysis to make it more complete.

References

[1] W. J. Boat T, Mental Disorders and Disabilities Among Low-Income Children., Washington (DC): National Academies Press (US), 2015.

[2] The National Institute of Mental Health, "Schizophrenia," [Online]. Available: https://www.nimh.nih.gov/health/topics/schizophrenia/index.shtml. [Accessed 5 October 2017].

[3] A. Jablensky, "The diagnostic concept of schizophrenia: its history, evolution, and future prospects.," *Dialogues in clinical neuroscience,* vol. 12, no. 3, p. 271, 2010.

[4] M. S. Reddy and S. Mythri, "Time to re-focus onto cognitive symptoms in schizophrenia.," *Indian journal of psychological medicine,* vol. 38, no. 2, p. 93, 2016.

[5] M. Clinic, "Mayo Clinic : Schizophrenia Symtom and Cause," [Online]. Available: https://www.mayoclinic.org/diseases-conditions/schizophrenia/symptoms-causes/syc-20354443. [Accessed 5 October 2017].

[6] K. Takara, T. Kondo and T. Kuba, "How and Why is Autism Spectrum Disorder Misdiagnosed in Adult Patients," *Mental Health in Family Medicine,* vol. 11, pp. 73-88, 2015.

[7] M. R. Woodbury-Smith, K. Boyd and P. Szatmari, "Autism spectrum disorders, schizophrenia and diagnostic confusion," *J Psychiatry Neurosci.,* 2010.

[8] P. S. Barneveld, J. Pieterse, S. Hanna and L. M. de Sonneville, "Overlap of autistic and schizotypal traits in adolescents with Autism Spectrum Disorders," *Schizophrenia Research* , vol. 126, 2010.

[9] Y. Hwang, J.-Y. S. J. Kim, J.-S. S. J-II Kim, M. J. Webster, D. Lee and S. Kim, "Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia," *Translational Psychiatry,* vol. 3, no. 10, p. e321, 2013.

[10] R. Govindarajan, D. J., K. K. and M. Palanisamy, "Microarray and its applications," *Journal of pharmacy & bioallied sciences,* vol. 4, pp. S310-2, 2012.

[11] C. Soneson and M. Delorenzi, "A comparison of method for differential expression analysis of RNA-seq data," *BMC bioinformatics,* vol. 14, no. 1, p. 91, 2013.

[12] M. M. H. Mollah, J. R. N. M. Mokhtar, R. Harun and M. N. H. Mollah, "A hybrid one-way ANOVA approach for the robust and efficient estimation of differential gene expression with multiple patterns.," *PloS one,* vol. 10, no. 9, p. e0138810, 2015.

[13] J. S. Myers, A. K. von Lersner, C. J. Robbins and Q. X. A. Sang, "Differentially expressed genes and signature pathways of human prostate cancer.," *PloS one,* vol. 10, no. 12, p. e0145322., 2015.

[14] D. Dembélé and P. Kastner, "Fold change rank ordering statistics: a new method for detecting differentially expressed genes.," *BMC bioinformatics,* vol. 15, no. 1, p. 14, 2014.

[15] M. L. Sánchez- Peña, C. E. Isaza, J. Pérez- Morales, C. Rodríguez- Padilla, J. M. Castro and M. Cabrera- Ríos, "Identification of potential biomarkers from microarray experiments using multiple criteria optimization.," *Cancer medicine,* vol. 2, no. 2, pp. 253-265, 2013.

[16] K. I. Camacho- Cáceres, J. C. Acevedo- Díaz, L. M. Pérez- Marty, M. Ortiz, J. Irizarry, M. Cabrera- Ríos and C. E. Isaza, "Multiple criteria optimization joint analyses of microarray experiments in lung cancer: from existing microarray data to new knowledge.," *Cancer medicine,* vol. 4, no. 12, pp. 1884-1900, 2015.

[17] C.-R. M. I. C. Rosas J, Biological Signaling Pathways and Potential Mathematical Network Representation: Biological Discovery Through Optimization [Dissertation], 2015.

[18] Y. M. Santiago-Correa, The Analysis of '-Omics' to Understand Alzheimer's Disease [Dissertation], 2015.

[19] N. Ortiz, Commonalities in Genetic Signatures and Signaling Pathways in Neurological Disorders [Dissertation], 2017.

[20] L. Ning, Y. Yu, E. GuoJi, C. Kou, Y. Wu, J. Shi, L. Ai and Q. Yu, "Meta-analysis of differentially expressed genes in autism bases on genes expression data," *Genetics and Molecular Research,* vol. 14, no. 1, pp. 2146-2155, 2015.

[21] M. Šakić, D. Karlović, B. Vidrih, V. Peitl, D. Crnković and N. Vrkić, "Increased calcium-independent lipoprotein phospholipase A2 but not protein S100 in patients with schizophrenia.," *Psychiatria Danubina,* vol. 28, no. 1, pp. 0-50, 2016.

[22] H. Qasem, L. Al-Ayadhi, H. Al Dera and A. El-Ansary, "Increase of cytosolic phospholipase A2 as hydrolytic enzyme of phospholipids and autism cognitive, social and sensory dysfunction severity," *Lipids in health and disease,* vol. 16, no. 1, p. 117, 2017.

[23] S. E. Ellis, R. Panitch, A. B. West and D. E. Arking, "Transcriptome analysis of cortical tissue reveals shared sets of downregulated genes in autism and schizophrenia," *Translational Psychiatry,* vol. 6, no. 5, p. e817, 2016.

[24] A. B. M. Kästner, T. M. Michel, S. Everts, B. Stepniak, C. Bach and H. ... & Ehrenreich, "Autism beyond diagnostic categories: characterization of autistic phenotypes in schizophrenia.," *BMC psychiatry,* vol. 15, no. 1, p. 115, 2015.

[25] B. St Pourcain, E. B. Robinson, V. Anttila, B. B. Sullivan, J. Maller, J. Golding and S. E. ... & Fisher, "ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties.," *Molecular Psychiatry,* vol. 23, no. 2, p. 263, 2018.

[26] G. O. Odu and O. E. Charles-Owaba, "Review of Multi-criteria Optimization Methods – Theory and Applications," *IOSR Journal of Engineering (IOSRJEN),* vol. 3, no. 10, pp. PP 01-14, 2013.

[27] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, Network Flows: Theroy, Algorithms and Applications, Upper Saddle River, New Jersey: Prentice-Hall Inc., 1993.

[28] M. D. Alter, R. Kharkar, K. E. Ramsey, D. W. Craig, R. D. Melmed, T. A. Grebe and J. B. Turner, "Autism and increased paternal age related changes in global levels of gene expression regulation," *PloS one,* vol. 6, no. 2, p. e16715, 2011.

[29] P. Maycox and e. al., "Analysis of genes expression in two large schizophreia cohorts identifies multiples changes associated with nerve terminal function," *Molecular Psychiatry,* vol. 14, no. 12, p. 1083, 2009.

[30] C. Isaza, J. F. Rosas, E. Lorenzo, A. Marrero, C. Ortiz, M. R. Ortiz and M. Cabrera- Ríos, "Biological signaling pathways and potential mathematical network representations: biological discovery through optimization.," *Cancer medicine,* 2018.

[31] F. A. Oski, "Iron Deficiency in Infancy and Childhood," *New England Journal of Medicine,* vol. 329, no. 3, pp. 190-193, 1993.

[32] A. Latif, P. Heinz and R. Cook, "Iron deficiency in autism and Asperger syndrome," *Autism* , vol. 6, no. 1, pp. 103-114, 2002.

[33] J. e. a. Xu, ""RNA-Seq analysis implicates dysregulation of the immune system in schizophrenia," *BMC genomics,* vol. 13, no. 8, p. S2, 2012.

[34] A. Prasad, M. A. Sdano, R. J. Vanzo, P. A. Mowery-Rushton, M. A. Serrano, C. H. Hensel and E. R. Wassman, "Clinical utility of exome sequencing in individuals with large homozygous regions detected by chromosomal microarray analysis.," *BMC medical genetics,* vol. 19, no. 1, p. 46, 2018.

[35] B. Ruttkay-Nedecky, L. Nejdl, J. Gumulec, O. Zitka, M. Masarik, T. Eckschlager and e. al., "The Role of Metallothionein in Oxidative Stress," *International journal of molecular sciences,* vol. 14, no. 3, pp. 6044-6066, 2013.

[36] G. K. Andrews, " Regulation of metallothionein gene expression by oxidative stress and metal ions.," *Biochemical pharmacology,* vol. 59, no. 1, pp. 95-104, 2000.

[37] M. A. Petrili, T. M. K. K. Kranz, P. J. P. Joe, M. V. Chao and D. Malaspina, "The emerging role for zinc in depression and psychosis," *Frontiers in pharmacology,* vol. 8, p. 414, 2017.

[38] E. C. Crăciun, G. Bjørklund, A. A. Tinkov, M. A. Urbina, A. V. Skalny, F. Rad and E. & Dronca, "Evaluation of whole blood zinc and copper levels in children with autism spectrum disorder," *Metabolic brain disease,* vol. 31, no. 4, pp. 887-890, 2016.

[39] G. Bjørklund, "The role of zinc and copper in autism spectrum disorders.," *Acta Neurobiol Exp (Wars),* vol. 73, no. 2, pp. 225-236, 2013.

[40] S. Faber, G. M. Zinn, J. C. Kern Ii and H. M. Skip Kingston, "The plasma zinc/serum copper ratio as a biomarker in children with autism spectrum disorders.," *Biomarkers,* vol. 14, no. 3, pp. 171-180, 2009.

[41] K. H. Choi, M. Elashoff, B. W. Higgs, J. Song, S. Kim, S. Sabunciyan and e. al., "Putative psychosis genes in the prefrontal cortex: combined analysis of gene expression microarrays.," *BMC psychiatry,* vol. 8, no. 1, p. 87, 2008.

[42] D. Arion, T. Unger, D. A. Lewis, P. Levitt and K. Mirnics, "Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia.," *Biological psychiatry,* vol. 62, no. 7, pp. 711-721, 2007.

[43] M. M. Al- Gayyar, M. A. Abdelsaid, S. Matragoon, B. A. Pillai and A. B. El- Remessy, "Thioredoxin interacting protein is a novel mediator of retinal inflammation and neurotoxicity," *British journal of pharmacology,* vol. 164, no. 1, pp. 170-180, 2011.

[44] X. Y. Zhang, M. H. Xiu, F. Wang, L. Y. Qi, H. Q. Sun, S. Chen, L. Lu and e. al., "The novel oxidative stress marker thioredoxin is increased in first-episode schizophrenic patients.," *Schizophrenia research,* vol. 113, no. 2-3, pp. 151-157, 2009.

[45] B. Owe-Larsson, K. Ekdahl, T. Edbom, U. Ösby, H. Karlsson, C. Lundberg and M. Lundberg, "Increased plasma levels of thioredoxin-1 in patients with first episode psychosis and long-term schizophrenia.," *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* vol. 35, no. 4, pp. 1117-1121, 2011.

[46] X. Y. Zhang, M. H. Xiu, F. De Yang, Y. L. Tan, S. He, T. A. Kosten and T. R. Kosten, "Thioredoxin, a novel oxidative stress marker and cognitive performance in chronic and

medicated schizophrenia versus healthy controls.," *Schizophrenia research,* vol. 143, no. 2-3, pp. 301-306, 2013.

[47] D. Silva-Adaya, M. E. Gonsebatt and J. Guevara, "Thioredoxin system regulation in the central nervous system: experimental models and clinical evidence.," *Oxidative medicine and cellular longevity,* 2014.

[48] V. Calabrese, E. Guagliano, M. Sapienza, C. Mancuso, D. A. Butterfield and A. M. Stella, "Redox regulation of cellular stress response in neurodegenerative disorders," *The Italian journal of biochemistry,* vol. 55, no. 3-4, pp. 263-282, 2006.

[49] J. J. C. Soerensen, H. Beck, H. Förster, J. Schmidt, W. Schmahl, M. .Brielmeier and e. al., "The role of thioredoxin reductases in brain development.," *PLoS One,* vol. 3, no. 3, p. e1813, 2008.

[50] W. R. Markesbery, "Oxidative stress hypotheis in Alzheimer's disease," *Free Radical Biology and Medicine,* vol. 23, no. 1, pp. 134-147, 1997.

[51] P. Jenner, "Oxidative stress and Parkinson's disease," *Handbook of Clinical Neurology,* vol. 83, pp. 507-520, 2007.

[52] B. S.E., R. Ferrante and B. M.F., "Oxidative stress in Huntington's disease," *Brain Pathology,* vol. 9, no. 1, pp. 147-163, 1999.

[53] Y. Su, W. Ding, M. Xing, D. Qi, Z. Li and D. Cui, " The Interaction of TXNIP and AFq1 Genes Increases the Susceptibility of Schizophrenia.," *Molecular neurobiology,* vol. 54, no. 6, pp. 4806-4812., 2016.

[54] S. K. Narayana, K. J. Helbig, E. M. McCartney, N. S. Eyre, R. A. Bull, A. Eltahla, A. Lloyd and M. R. Beard, "The interferon-induced transmembrane proteins-IFITM1, IFITM2 and IFITM3 inhibit hepatitis C virus entry," *Journal of Biological Chemistry,* vol. 290, no. 43, pp. 25946-25959, 2015.

[55] G. Shi, O. Schwartz and A. A. Compton, "More than meets the I: the diverse antiviral and cellular functions of interferon-induced transmembrane proteins," *Retrovirology,* vol. 14, no. 1, p. 53, 2017.

[56] G. Khandaker, J. Zimbron, G. Lewis and J. P. B., "Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies," *Psychological Medicine,* vol. 43, no. 2, pp. 239-257, 2013.

[57] A. Brown, "Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism," *Developmental neurobiology,* vol. 72, no. 10, pp. 1272-1276, 2012.

[58] A. Brown and E. Derkits, "Prenatal Infection and Schizophrenia: A Review of Epidemiologic and Translational Studies," *The american journal of psychiatry*, vol. 167, no. 3, pp. 261- 280, 2010.

[59] S. Horváth and K. Mirnics, "Immune System Disturbances in Schizophrenia," *Biological Psychiatry,* vol. 75, no. 4, pp. 316-323, 2014.

[60] "Beta-2 microglobulin as an immunological marker to assess the progression of human immunodeficiency virus infected patients on highly active antiretroviral therapy.," *Clinica chimica acta,* vol. 412, no. 11, pp. 1151-1154., 2011.

[61] L. Li, M. Dong and X.-G. Wang, "The Implication and Significance of Beta 2 Microglobulin: A Conservative Multifunctional Regulator.," *Chinese Medical Journal,* vol. 129, no. 4, pp. 448-455, 2016.

[62] F. T. Ali, E. M. A. El-Azeem, M. A. Hamed, M. A. Ali, N. M. A. Al-Kader and E. A. Hassan, "Redox dysregulation, immuno-inflammatory alterations and genetic variants of BDNF and MMP-9 in schizophrenia: pathophysiological and phenotypic implications.," *Schizophrenia research,* vol. 188, pp. 98-109, 2017.

[63] D. N. Melegos, M. S. Freedman and E. P. Diamandis, "Prostaglandin D synthase concentration in cerebrospinal fluid and serum of patients with neurological disorders.," *Prostaglandins,* vol. 54, no. 1, pp. 463-474, 1997.

[64] D. Morell-Garcia, J. M. Bauça, M. P. Sastre, A. Yañez and I. Llompart, "Morell-Garcia, D., Bauça, J. M., Sastre, M. P.,Sample-dependent diagnostic accuracy of prostaglandin D synthase in cerebrospinal fluid leak.," *Clinical Biochemistry,* vol. 50, no. 1-2, pp. 27-31, 2017.

[65] J. K. S. Maesaka, P. B., R. L. T., V. Batuman, N. Miyawaki, S. Sharty, S. Youmans and M. El-Sabban, "Prostaglandin D2 synthase: Apoptotic factor in alzheimer plasma, inducer of reactive oxygen species, inflammatory cytokines and dialysis dementia.," *Journal of nephropathology,* vol. 2, no. 3, p. 166, 2013.

[66] M. G. Harrington, A. N. Fonteh, R. G. Biringer, A. F. Hühmer and R. P. Cowan, "Prostaglandin D synthase isoforms from cerebrospinal fluid vary with brain pathology.," *Disease markers,* vol. 22, no. 1-2, pp. 73-81, 2006.

[67] K. J. Jones, S. Templet, K. Zemoura, B. Kuzniewska, F. X. Pena, H. Hwang, W. Xu and e. al., " Rapid, experience-dependent translation of neurogranin enables memory encoding.," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 115, no. 25, pp. E5805-E5814, 2018.

[68] Z. Wen, J. Chen, R. A. W. Khan, M. Wang, Z. Song and Z. .. &. S. Y. Li, "Polymorphisms in NRGN are associated with schizophrenia, major depressive disorder and bipolar disorder in the Han Chinese population," *Journal of affective disorders*, vol. 194, pp. 180-187, 2016.

[69] K. Ohi, R. Hashimoto, Y. Yasuda, M. Fukumoto, H. Yamamori, S. Umeda-Yano and M. Takeda, "Influence of the NRGN gene on intellectual ability in schizophrenia," *Journal of human genetics*, vol. 58, no. 10, p. 700, 2013.

[70] E. Walton, D. Geisler, J. Hass, J. Liu, J. Turner, A. Yendiki, S. Ehrlich and e. al., "The Impact of Genome-Wide Supported Schizophrenia Risk Variants in the Neurogranin Gene on Brain Structure and Function," *PLos One*, vol. 8, no. 10, p. e76815, 2013.

[71] M. A. Knovich, J. A. Storey, L. G. Coffman and S. V. Torti, " Ferritin for the Clinician.," *Blood Reviews*, vol. 23, no. 3, pp. 95-104, 2009.

[72] A. Datta, H. Akatsu, K. Heese and S. K. Sze, "Quantitative clinical proteomic study of autopsied human infarcted brain specimens to elucidate the deregulated pathways in ischemic stroke pathology.," *Journal of Proteomics*, vol. 91, pp. 556-568, 2013.

[73] M.-H. Chen, T.-P. Su, Y.-S. Chen, J.-W. Hsu, K.-L. Huang, Chang, W.-H., Y.-M. Bai and e. al., " Association between psychiatric disorders and iron deficiency anemia among children and adolescents: a nationwide population-based study.," *BMC Psychiatry*, vol. 13, p. 161, 2013.

[74] S. K. F. M. T. C. &. Ç. M. (. Hergüner, "Ferritin and iron levels in children with autistic disorder.," *European Journal of Pediatrics*, vol. 171 , no. 1, p. 143–146, 2011.

[75] S.-L. Chen, S.-X. Lu, L.-L. Liu, C.-H. Wang, X. Yang, Z. Zhang, J. Yun and e. al., "EEF1A1 Overexpression Enhances Tumor Progression and Indicates Poor Prognosis in Hepatocellular Carcinoma.," *Translatiional Oncology*, vol. 11, no. 1, pp. 125-131-, 2018.

[76] D. Li, T. Wei, C. M. Abbott and D. Harrich, " The Unexpected Roles of Eukaryotic Translation Elongation Factors in RNA Virus Replication and Pathogenesis.," *Microbiology and Molecular Biology Reviews : MMBR,* vol. 77, no. 2, pp. 253-266, 2013.

[77] W. Abbas, A. Kumar and G. Herbein, " The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections.," *Frontiers in Oncology*, vol. 5, p. 75, 2015.

[78] X. Li, T. S. Ptacek, E. E. Brown and J. C. Edberg, " Fcγ Receptors: Structure, Function and Role as Genetic Risk Factors in SLE," *Genes and Immunity,*, vol. 10, no. 5, p. 380–389, 2009.

[79] L. Handschuh, M. Kaźmierczak, M. C. Milewski and M. Ł. M. W. M. …. F. M. (. Góralski, "Gene expression profiling of acute myeloid leukemia samples from adult patients with

AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR.," *International Journal of Oncology,* vol. 52(3), no. 3, pp. 656-678, 2018.

[80] S. Yokoi, N. Ishihara, F. Miya, M. Tsutsumi, I. Yanagihara, N. Fujita, J. Natsume and e. al., "TUBA1A mutation can cause a hydranencephaly-like severe form of cortical dysgenesis," *Scientific Reports,* vol. 5, p. 15165, 2015.

[81] R. Oegema, T. D. Cushion, I. G. Phelps, S.-K. Chung, J. C. Dempsey, S. Collins, D. Doherty and e. al., "Recognizable cerebellar dysplasia associated with mutations in multiple tubulin genes.," *Human Molecular Genetics,* vol. 24, no. 18, p. 5313–5325, 2015.

[82] K. Shimojima, A. Narita, Y. Maegaki, A. Saito, T. Furukawa and T. Yamamoto, "Whole-exome sequencing identifies a de novo TUBA1A mutation in a patient with sporadic malformations of cortical development: a case report," *BMC Research Notes,* vol. 7, p. 465, 2014.

[83] R. Romaniello, F. Arrigoni, A. Cavallini, E. Tenderini, C. Baschirotto, F. Triulzi, R. Borgatti and e. al., "Romaniello, R., Arrigoni, F., Cavallini, A., Tenderini, E., Baschirotto, C., Triulzi, F., … Borgatti, R. (2014). Brain malformations and mutations inα- andβ-tubulin genes: a review of the literature and description of two new cases," *Developmental Medicine & Child Neruology,* vol. 56, no. 4, pp. 354-360, 2014.

[84] L. Calvo, B. Anta, S. López-Benito, C. Martín-Rodriguez, F. S. Lee, P. Pérez and J. C. Arévalo, "Bex3 dimerization regulates NGF-dependent neuronal survival and differentiation by enhancing trkA gene transcription," *Journal of Neuroscience* , vol. 35, no. 18, pp. 7190-7202, 2015.

[85] J. U. Kazi, N. N. Kabir and L. Rönnstrand, " Brain-Expressed X-linked (BEX) proteins in human cancers.," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer,* vol. 1856, no. 2, p. 226–233, 2015.

[86] W. Gao, J. Z. Li, S. Chen, C. Chu, J. Y. Chan and T. Wong, "BEX3 contributes to cisplatin chemoresistance in nasopharyngeal carcinoma.," *Cancer Medicine* , vol. 6, no. 2, pp. 439-451, 2017.

[87] V. Llombart, S. A. Trejo, S. Bronsoms, A. Morancho, M. Feifei, J. Faura and J. Montaner, "Profiling and identification of new proteins involved in brain ischemia using MALDI-imaging-mass-spectrometry.," *Journal of Proteomics,* vol. 152, pp. 243-253, 2017.

[88] X. Sun, J.-F. Wang, M. Tseng and L. T. Young, "Downregulation in components of the mitochondrial electron transport chain in the postmortem frontal cortex of subjects with bipolar disorder.," *Journal of Psychiatry and Neuroscience,* vol. 31, no. 3, pp. 189-196, 2006.

[89] S. Lattante, I. Le Ber, A. Camuzat, A. Brice and E. Kabashi, " Mutations in the PFN1 gene are not a common cause in patients with amyotrophic lateral sclerosis and frontotemporal lobar degeneration in France," *Neurobiology of Aging,* vol. 34, no. 6, pp. 1709.e1-1709e2, 2013.

[90] C. Tiloca, N. Ticozzi, V. Pensato, L. Corrado, R. Del Bo, C. Bertolin and e. al., "Screening of the PFN1 gene in sporadic ALS and in FTD.," *Neurobiology of Aging,* vol. 34, no. 5, p. 1517.e9–1517.e10. , 2013.

[91] K. Michaelsen-Preusse, S. Zessin, G. Grigoryan, F. Scharkowski, J. Feuge, A. Remus and M. Korte, " Neuronal profilins in health and disease: Relevance for spine plasticity and Fragile X syndrome.," *Proceedings of the National Academy of Sciences of the United States of America ,* vol. 113, no. 12, p. 3365–3370, 2016.

[92] F. Stricher, C. Macri, M. Ruff and S. Muller, "HSPA8/HSC70 chaperone protein: structure, function, and chemical targeting," *Autophagy,* vol. 9, no. 12, pp. 1937-1954, 2013.

[93] P. N. Silva, T. K. Furuya, I. L. Braga, L. T. Rasmussen, R. W. Labio, P. H. Bertolucci and J. Mill, "Analysis of HSPA8 and HSPA9 mRNA expression and promoter methylation in the brain and blood of Alzheimer's disease patients," *Journal of Alzheimer's disease,,* vol. 38, no. 1, pp. 165-170., 2014.

[94] M. E. Jakobsson, A. Moen, L. Bousset, W. Egge-Jacobsen, S. Kernstock, R. Melki and P. Ø. Falnes, "Identification and Characterization of a Novel Human Methyltransferase Modulating Hsp70 Protein Function through Lysine Methylation. The Journal," *The Journal of Biological Chemistry,,* vol. 288, no. 39, p. 27752–27763, 2013.

[95] P. Gareri, R. Mattace, F. Nava and G. De Sarro, "Role of calcium in brain aging.," *General Pharmacology: The Vascular System,* vol. 26 , no. 8, p. 1651–1657, 1995.

[96] M. J. Berridge, "Calcium signalling and psychiatric disease: bipolar disorder and schizophrenia.," *2014,* vol. 357, no. 2, p. 477–492, Cell and Tissue Research.

[97] L. Hertzberg, P. Katsel, P. Roussos, V. Haroutunian and E. Domany, "Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia," *Schizophrenia Research,* vol. 164, no. 1-3, p. 92–99. , 2015.

[98] X. Zhou, W. Liao, J. Liao, P. Liao and H. Lu, " Ribosomal proteins: functions beyond the ribosome.," *J Mol Cell Biol,* vol. 7, no. 2, pp. 92-104., 2015.

[99] H. Chong, S. Teoh, D.-C. Wu, S. Kotirum, C.-F. Chiou and C. N., "Global economic burden of schizophrenia: a systematic review.," *Neuropsychiatric Disease and Treatment,* no. 12, pp. 357-373, 2016.

[100] E. Q. Wu, H. G. Birnbaum, L. Shi, D. E. Ball, R. C. Kessler, M. Moulis and J. Aggarwal, "The economic burden of schizophrenia in the United States in 2002.," *Journal of Clinical Psychiatry,* vol. 66, no. 9, pp. 1122-1129, 2005.

[101] R. J. Wyatt, I. Henter, M. C. Leary and E. Taylor, "An economic evaluation of schizophrenia-1991.," *Social psychiatry and psychiatric epidemiology,,* vol. 30, no. 5, pp. 196-205, 1995.

[102] J. Gunderson and L. Mosher, "The cost of schizophrenia.," *Am J Psychiatry,* vol. 132, no. 9, p. 901–906, 1975.

[103] P. R. Desai, K. A. Lawson, J. C. Barner and K. L. Rascati, "Estimating the direct and indirect costs for community- dwelling patients with schizophrenia.," *Journal of Pharmaceutical Health Services Research,* vol. 4 , no. 4, pp. 187-194, 2013.

[104] B. Miller, "The Heavy Economic Burden of Schizophrenia," *Psychiatric Times,* vol. 33, no. 10, 2016.

[105] B. A.V.S, C. Z, M. Knapp and D. Mandell, "Costs of Autism Spectrum Disorders in the United Kingdom and the United States," *JAMA Pediatrics,* vol. 168, no. 8, p. 721–728, 2014.

[106] M. Ganz, "The Lifetime Distribution of the Incremental Societal Costs of Autism," *Arch Pediatr Adolesc Med,* vol. 161, no. 4, p. 343–349, 2007.

[107] Y. a. H. Y. Benjamini, "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society B,* vol. 57, pp. 289 -300, 1995.

Appendix

Appendix A

Hypergeometric Distribution

The pathway enrichment analysis is done by using the hypergeometric distribution. The application of this distribution is as follows:

$$Pvalue = \sum_{k=n}^{\min(m,d)} \frac{\binom{m}{k}\binom{f-m}{d-k}}{\binom{f}{d}}$$

m = is the total number of genes in the pathway of interest

k = the number of genes

n= is number of differentially expressed (DE) genes in the pathway of interest

f= total of genes

d= the set of differentially expressed (DE) genes

Example:

$$Pathway\ 1\ p\_value = \sum_{k=3}^{\min(13,3)} \frac{\binom{13}{k}\binom{10561-13}{3-k}}{\binom{10561}{3}} = 1.457 \times 10^{-9}$$

Pathway	Total Genes	Total Genes in Patway	Total of DE Genes in Pathway	Total of DE Genes of interest	p_value
Pathway 1	10561	13	3	3	1.457E-09
Pathway 2	10561	13	3	3	1.457E-09
Pathway 3	10561	13	3	3	1.457E-09
Pathway 4	10561	42	3	3	5.849E-08
Pathway 5	10561	172	1	3	4.807E-02
Pathway 6	10561	479	1	3	1.300E-01

Benjamini-Hochberg False Discovery Rate

The Benjamini-Hochberg method adjusts the p-value in such a way that it limits the number of false positives that are reported as significant [107]. This adjustment means that the p-values increase their values. The steps to adjust the p-values are the following:

1) Organize the p-values from smallest to largest values.
2) Rank the p-values, where the smallest value has a rank of 1.
3) Start the p-value adjustment from the largest to smallest values
    a. The largest p-value remains the same
    b. The next p-value adjustment is calculated as follows:

$$Adjusted\ pvalue = Min\left(previous\ pvalue, current\ pvalue \times \left(\frac{total\ number\ of\ pvalues}{pvalue\ rank}\right)\right)$$

    c. Continue the process until all the p-values are adjusted

Example:

P-Value	1.46E-09	1.46E-09	1.46E-09	5.85E-08	0.04807	0.13
Rank	1	2	3	4	5	6
Adjusted P-value						0.13
Rank	1	2	3	4	5	6

$$Adjusted\ pvalue = Min\left(0.13\ , 0.04807 * \left(\frac{6}{5}\right)\right) = 0.05769$$

P-Value	1.46E-09	1.46E-09	1.46E-09	5.85E-08	0.04807	0.13
Rank	1	2	3	4	5	6
Adjusted P-value					0.05769	0.13
Rank	1	2	3	4	5	6

$$Adjusted\ pvalue = Min\left(0.04807\ , 5.85E-08 * \left(\frac{6}{4}\right)\right) = 8.77E-08$$

P-Value	1.46E-09	1.46E-09	1.46E-09	5.85E-08	0.04807	0.13
Rank	1	2	3	4	5	6
Adjusted P-value	2.91E-09	2.91E-09	2.91E-09	8.77E-08	0.05769	0.13
Rank	1	2	3	4	5	6