The Hidden Correlation between Intelligence and Autism

العلاقة الكامنة بين العبقرية والتوحد

by

REEMA TAWFIQ YAHYA

Dissertation submitted in fulfilment

of the requirements for the degree of

MSc INFORMATICS

at

The British University in Dubai

January 2020

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

_____

Signature of the student

COPYRIGHT AND INFORMATION TO USERS

Acknowledgements

Firstly, I would like to express my deepest gratitude to my family for their continuous support. As well, I would like to express my appreciation to my advisor, Dr. Sherief Abdullah, for his guidance.

Abstract

In 2018, Centers for Disease Control and prevention (CDC) [26] determined that approximately 1 in 54 children in US is diagnosed with an autism spectrum disorder (ASD). Within this group of autistic children, 44% had IQ scores in the average to above average range (IQ >85). Moreover, various studies [4,5,6] show that many cases of prodigies or super talented children show some autistic traits, most notably in attention to details and problems in communication or social skills.

The Objective of this study is to discover the connection between autism and extreme talent (i.e. intelligence). In this research, cytoscape was used for visualizing and analyzing autism/intelligence genes' interaction network. It has been found that there are some shared genes as well as a list of pathways demonstrating the strong correlation between Autism and Intelligence. Furthermore, strong contribution of some genes was observed in different pathways. For instance, EP300 was seen in motor delay and autistic behavior pathway. The same gene has a strong relation with Rubinstein Taybi syndrome. Glutamate-dependent receptors group were also seen in different pathways. This research has identified the shared genes and hubs (genes with high degree of connectivity) from statistical point of view. This document is divided into five chapters. Th first chapter of this thesis introduces Autism or Autism spectrum disorder (ASD). The second chapter covers a literature review on the available studies. The third chapter discusses the methodology used to explore the assumed genetic correlation between autism and intelligence. It also introduces and discusses the main tools, applications and databases used in this research. The fourth chapter focuses on analyzing

the genetic pathways that are represented in genetic interaction networks. The last chapter (chapter 5), provides the conclusion for the conducted research.

نبذة مختصرة

في عام 2018، اعلنت مراكز مكافحة الأمراض والوقاية أن طفلاً واحداً من كل 59 طفل تقريباً يتم تشخيصه باضطراب طيف التوحد .(ASD) كما ويعاني 31 في المائة من هؤلاء الأطفال من إعاقة فكرية تتسم بضعف إدراكي كبير، أي أن معدل ذكائهم يقل عن 70%. ومن ناحية أخرى، فقد أظهرت الدراسة السابقة أنّ معدل ذكاء 44% من هؤلاء الأطفال الذين يعانون من التوحد يتراوح بالمدى بين متوسط إلى فوق متوسط .(IQ >85) كما وقد سجلت تلك الدراسة أن هناك مجموعة من الأطفال المصابين بالتوحد يتميزون بالتفوق والموهبة. علاوةً على ذلك، فقد أظهرت العديد من الدراسات أن العديد من الأطفال العباقرة أو الموهوبين للغاية لديهم بعض سمات التوحد، التي تُلاحظ في انتباههم إلى التفاصيل ومواجهتهم لبعض المشاكل في التواصل والمهارات الاجتماعية.

وفي الوقت الحاضر، فإنَّ قواعد البيانات الوراثية متوفرة على نطاق واسع ويجري تعزيزها بسرعة لتيسير عملية استخراج المعلومات واسترجاعها. كما أن البيانات الجينية لعدد كبير من الحالات الطبية متوفرة الآن بأنواع وأشكال مختلفة من البيانات. ففي الآونة الأخيرة، أعلنت مجموعة من العلماء تحديد قائمة طويلة من الجينات التي تتصل بالذكاء. مثل هذه المعلومات يعد هاماً بالنسبة للأبحاث من هذا القبيل لتحديد العلاقة بين بيانات الامراض والحالات الصحية المختلفة. الامر الذي قد يساهم في التوصل الى فهم أعمق للحالات الصحية.

الهدف من هدا البحث هو اكتشاف الصلة بين التوحد والذكاء. حيث يعرض الفصل الأول مقدمة في مرض التوحد أو اضطراب طيف التوحد .(ASD) أما الفصل الثاني فهو مراجعة لدراسات سابقة. فيما يناقش الفصل الثالث المنهجية المستخدمة في هذا البحث لاستكشاف العلاقة الوراثية المُفترضة بين التوحد والذكاء. كما يقدم ويناقش الأدوات والتطبيقات وقواعد البيانات الرئيسية المستخدمة في هذا البحث. ويركز الفصل الرابع على تحليل المسارات الجينية الممثلة في شبكات التفاعل الجيني. كما ويقدم الفصل الخامس والأخير الاستنتاج الذي توصلت إليه البحوث التي أجريت.

Table of Contents

I. Glossary

Number	Term	Definition
1	Gene	A sequence of nucleotides located on a chromosome
2	Alleles	Small differences in the sequence of DNA bases within the same gene. These differences could cause person's unique physical features.
3	SNP	Single nucleotide polymorphisms
4	Locus	Plural loci, the location of a particular gene or genetic marker on a chromosome.

Chapter 1: **Introduction**

Autism or Autism spectrum disorder (ASD) *[1]* is a complex neurodevelopmental disorder that is characterized by impairment in social interaction, repetitive behaviors and difficulties in verbal and nonverbal communication. Usually these symptoms appear during the first three years of age, and vary widely in severity between individuals with autism.

For a long time, Autism has been considered as a black box, and many assumptions were made to explain its root cause. Like the impact of vaccination, certain types of infections, or giving birth to a child from an aged parent(s), or having a family history of autism. Recently, Researchers identified some rare genes or mutations that could be associated with autism.

Despite of the strong facts behind the hypothesis of the autism genetic roots, genetic tests could not explain the majority of autism cases.

In the other hand, Intelligence is another attractive mental condition. There is no clear definition for Intelligence. In dictionary it is: "capacity for learning, reasoning, understanding, and similar forms of mental activity".

For decades Autism was linked to intellectual disability *[2],* which is defined by an IQ below 70. This linkage has been repealed by the appearance of the modern methods of IQ testing, such as the Test of Nonverbal Intelligence (TONI), which can assess the level of intelligence without being affected by Autism symptoms like motor or language difficulties.

The results of applying modern IQ testing methods emphasized that the social and communication difficulties could mask the true intelligence levels of some autism patients.

Furthermore, many recent psychology studies theorize that the correlation could be in the opposite direction and the connection between Intelligence and Autism has much more evidence. Moreover, the most influential scientists, “Isaac Newton” and “Albert Einstein” might have been autistic. Both of these great scientists have shown many signs of Asperger syndrome, a form of Autism that does not cause learning difficulties. This research aims to explore the correlation between the two poles, Autism and intelligence. The contribution of this research can be scoped in the following:

1. Identifying the correlation between Autism and intelligence, and as a result, recognizing the unique genes related to Autism.

2. Shedding the light on the common genes’ interactions and pathways that could lead to common traits.

Chapter 2: **Literature review**

According to the paper "*Autism as a disorder of high intelligence*" *[3]* published in 2016, there is an interesting correlation between Autism and intelligence. This hypothesis could be a step toward getting more insights into the genetic causes of both high and low mental abilities (Intelligence vs Autism). It might also provide a novel implication for general cognitive enhancement and give more clues about the common ground between the poles; Intelligence, autism and other mental disorders such as schizophrenia.

The author of the paper described Autism as a disorder of high intelligence. He elaborated; Autism is a type of intelligence where components are imbalanced. His hypothesis was evaluated based on different criteria such as brain size and growth, neural functions, brain connectivity, sensory functions and attention, decision-making and profession.

In 2011, in a different circumstance, a study *[17]* was conducted on a group of eight prodigies (*definition*: a child prodigy is an individual who rises to professional levels of achievement well before what is observed in the general population) to better understand the basis of extraordinary talents in child prodigies. To better understand the child prodigies, the researched has conducted a comprehensive analysis. Starting with the cognitive development history of the selected children and their family. In addition to that, Stanford-Binet (an individual administered intelligence test) *[18]* was performed on the group of selected children. And finally, the researchers conducted the Autism-Spectrum Quotient (AQ) (a questionnaire published in 2001 to investigate the existence of autism symptoms in adult with average

intelligence) *[19]*.

Analyzing the collected data revealed that each prodigy has at least modest elevation in general intelligence and exceptional working memory. In addition to a general elevation in autistic traits. In fact, the prodigies scored higher than the average in a particular sub category of the AQ, which is attention to details while they did not display many of the other traits associated with autism. The researchers also highlighted the higher incidence of autism in prodigy families than in general population. The study suggests that %56 of participated prodigies had at least one relative (first or second-degree) with ASD.

The researchers went further by conducting another related study in The Research Institute at Nationwide Children's Hospital, Ohio, USA *[4]* to test the hypothesis of genetic similarities or shared etiology between the child prodigies, and their relatives with autism.

To test the mentioned hypotheses; family-based genome-wide linkage analysis was performed on 5 nuclear and extended families. The objective of the analysis was to search for genetic loci that may explain the presence of both prodigy and ASD in the same family. The study suggests that a locus on chromosome 1 influences both prodigy and ASD in the participated families.

According to Sikela, J.M., Searles Quick *[20]*, the concept of genomic trade off might apply to brain and cognitive processes. The researchers claimed that the same genomic variants, which led to an enhanced cognitive capacity in the human brain, could also produce cognitive disability. They introduce Olduvai protein domain family, as a protein domain that has been linked before to Autism and schizophrenia. Olduvai protein also has a strong contribution in the rapid expansion of the human brain. The relation was expressed nicely as "the same 'genes' that drive us mad have made us human".

An innovative study done in Melbourne/Australia *[21]* used Network Based Statistics (NBS) to assess the relation between resting-state Functional Magnetic Resonance Imaging (FMRI) connectivity and fluid intelligence ability in both autistic and control children. The two tested groups matched on fluid intelligence performance, gender and age. Interestingly, the results show a significant increase in the association between function connectivity and fluid intelligence in children with autism.

The use of PET and FMRI could provide more insights of the hierarchy of the neural development and the functionality of each region in the brain. In the case of Autism, these methods are useful to understand the mechanism behind the autistic brain. In this study [21], the authors claim that the difference between normal and autistic children is in the number of brain regions allocated to perform a specific task.

Another related study by Pérez Velázquez JL and Galán RF *[24]* , in this study MEG data (*Definition*: MEG is a functional neuroimaging technique that maps brain activity by recording magnetic fields produced naturally in the brain by electrical currents, using sensitive devices called magnetometers *[25]*) was used to record "in resting" state for children with Autism (Asperger syndrome in this case) and non-autistic children. The generated data was analyzed to compute the information gain and relative entropy for both Autistic and non-autistic children. The analysis of MEG signals recorded at rest revealed that the brains of individuals with ASD produce more information than the participated controls (non-autistic) with a 42% increase on average. The researchers suggest that the discovered excessive production of information in the absence of external sensory stimuli could cause the cognitive differences between children with and without autism.

A team of European researchers conducted two consecutive studies about the genetic basis of intelligence *[5][6]*. Both studies aimed to gain better understanding of how human brain works, as well as, identifying genetically related neurological disorders and how the same roots can cause totally different symptoms. In the second study, which was published in 2019, it presents the largest genetic association study of intelligence to date.

According to first paper *[5]*, the researchers analyzed data from 14 different databases; the used databases contain either genetic data or intelligence test scores. By merging the data and using a newly developed statistical method called MAGMA, they could isolate almost 1000 genes associated with high intelligence, in which these genes were never been linked to intelligence before.

Using Generalized Summary statistic-based Mendelian Randomization, the researchers observed a positive correlation between intelligence and autism, while no correlation was observed with other mental disorders such as schizophrenia. In the second study *[6]*, the researchers report finding more genes associated with intelligence by searching for information surrounding neurotic or personality traits like moodiness, anxiety and nervousness.

In 2015, Cambridge University *[7]* conducted a unique study to assess Autism-Spectrum Quotient (AQ) scores in a sample of approximately 450,000 individuals. (AQ is a questionnaire published in Autism Research Centre at the university of Cambridge, to investigate whether normal people with average IQ could experience some symptoms of autism). This huge sample was collected unexpectedly through a live medical TV show, sample was collected on a website with AQ test, and other self-report measures related to psychiatric traits (http://mindchecker.channel4.com).

The researchers aimed to examine the correlation between autism and other variable such as: age, occupation, gender, and geographic region. The result shows negative correlation between autism and geographic region and age. On the other hand, it shows a clear positive correlation between autism, gender and occupation. On average, males scored higher than females and people working in STEM (Science, Technology, Engineering and Math) careers scored higher than people working in non-STEM careers. The results of this study are in favor and supports previous studies conducted on the same field.

Chapter 3: **Methodology**

In this research, we are using two main gene's datasets. The first dataset focuses on Autism genes, while the second set focuses on intelligence related genes. The complete description of the two datasets is illustrated as follow:

Human Gene Module: Lists 1053 genes implicated in Autism. Human Gene Module is available via SFARI Gene *[22]* (an online database designed to facilitate access into the genetics of autism).

Genome-wide Analysis (GWAS) Summary statistics for intelligence: This dataset was published in a recent study done by a group of researchers from different European universities. The study was conducted using data collected from 14 different cohorts (datasets), as illustrated below:

1. UK Biobank: for Verbal and mathematical reasoning.
2. Rotterdam Study (RS).
3. Spit for Science (S4S): SAT test scores.
4. Generation R Study (GENR): spatial visualization and abstract reasoning subsets).
5. Swedish Twin Registry (STR): Logical, verbal, spatial, and technical ability subtests.
6. High-IQ/ Health and Retirement Study (High-IQ cases/unselected population controls).
7. Cognitive Genomics Consortium: One or more neuropsychological tests from three or more domains of cognitive performance.
8. Twins Early Development Study (TEDS).
9. Danish Twin Registry (DTR).
10. IMAGEN.
11. Brisbane Longitudinal Twin Study (BLTS).

12. Netherlands Study of Cognition, Environment, and Genes (NESCOG).
13. Genes for Good (GfG).
14. Swedish Twin Studies of Aging (STSA).

3.1 Data Pre-processing

The approach for the intelligence dataset, GWAS, is to compare the DNA variants between the listed cohorts. A sample of 269,867 individuals participated in the trial in order to find common variants related to intellectual abilities.

GWAS dataset consists of a large number of SNPs (Single-nucleotide polymorphism). "SNP is a single-nucleotide polymorphism, it is defined as a substitution of a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population *[8]*".

The objective of this research is to build genes' interaction network combining both datasets (Human Gene Module, GWAS) in order to identify the correlation between Autism and intelligence. The dataset of the Human Gene Module is gene-oriented, while the GWAS dataset consists of a long list of SNPs. In order to build an interaction network for both datasets, it is mandatory to map the SNPs of GWAS dataset into their related genes.

The mapping process for each SNP in the GWAS dataset with its associated gene is accomplished using SNP2GENE. SNP2GENE is an online tool offered by FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies). FUMA is a platform for annotation and visualization. The output of the mapping process is a subset of genes that are linked to intelligence.

The mapping process is one step toward finding the correlation between the genes of Autism and Intelligence. The next step in the research is to find the genes' interactions between the two datasets illustrated earlier (Human Gene Module and GWAS). This process is done using STRING. STRING *[16]* is an online biological database. It provides information about gene/protein functional interactions. STRING builds the full interaction network between datasets by providing a set of genes as an input. STRING builds the interaction network based on multiple predefined data sources (Experimental Data, Predictions, Computational Methods, and Text Mining).

The latter step is to visualize and analyze the interaction network of genes. This step is accomplished using Cytoscape. Cytoscape is an open source software platform for visualizing and analyzing complex networks. It was originally developed to serve the field of bioinformatics. The initial version of Cytoscape is fully illustrated in a paper done by a research group with biomedical background, the paper was published in 2003 *[9]*. The latest versions of Cytoscape cover a wide range of domains including Semantic Web and Social Networks. As an open source platform, Cytoscape publishes a range of Application-Programming Interfaces (APIs), allowing developers and researchers to extend its capabilities, add more features and integrate it with other applications and platforms.

Figure 1 below illustrates the initial gene's interaction network generated using Cytoscape. The nodes represent the genes (Autism and Intelligence). Edges represent the interactions between the participating genes.

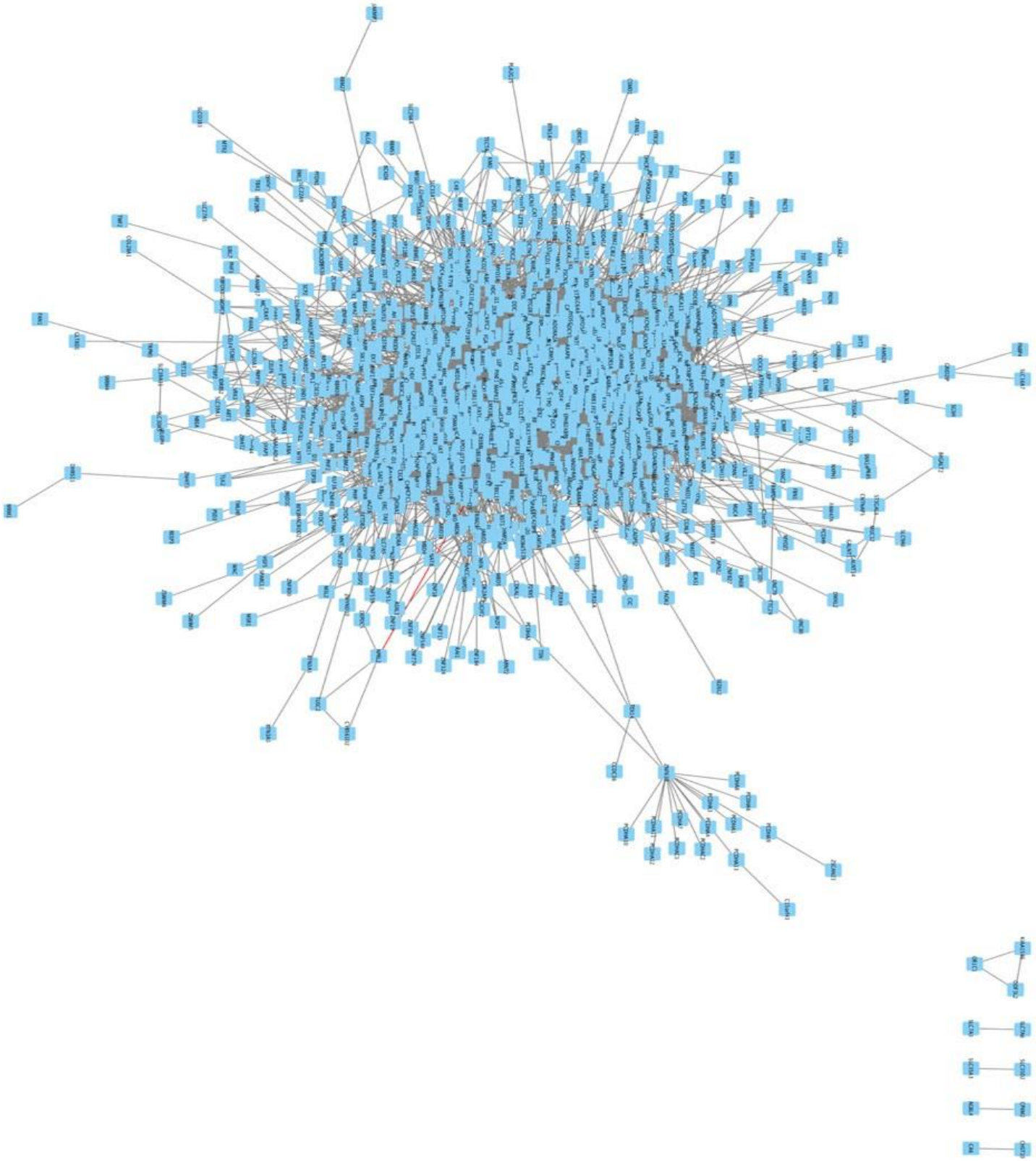

Figure 1: Initial gene network generated by Cytoscape. The network has 1119 nodes and 8304 edges.

Figure 2 below illustrates the Autism/ Intelligence gene interaction network visualized using Grid layout. The network shows Autism gene grouped in red, intelligence genes in green and shared genes in yellow. The size of the node represents the degree of the node (number of edges connected to the node).

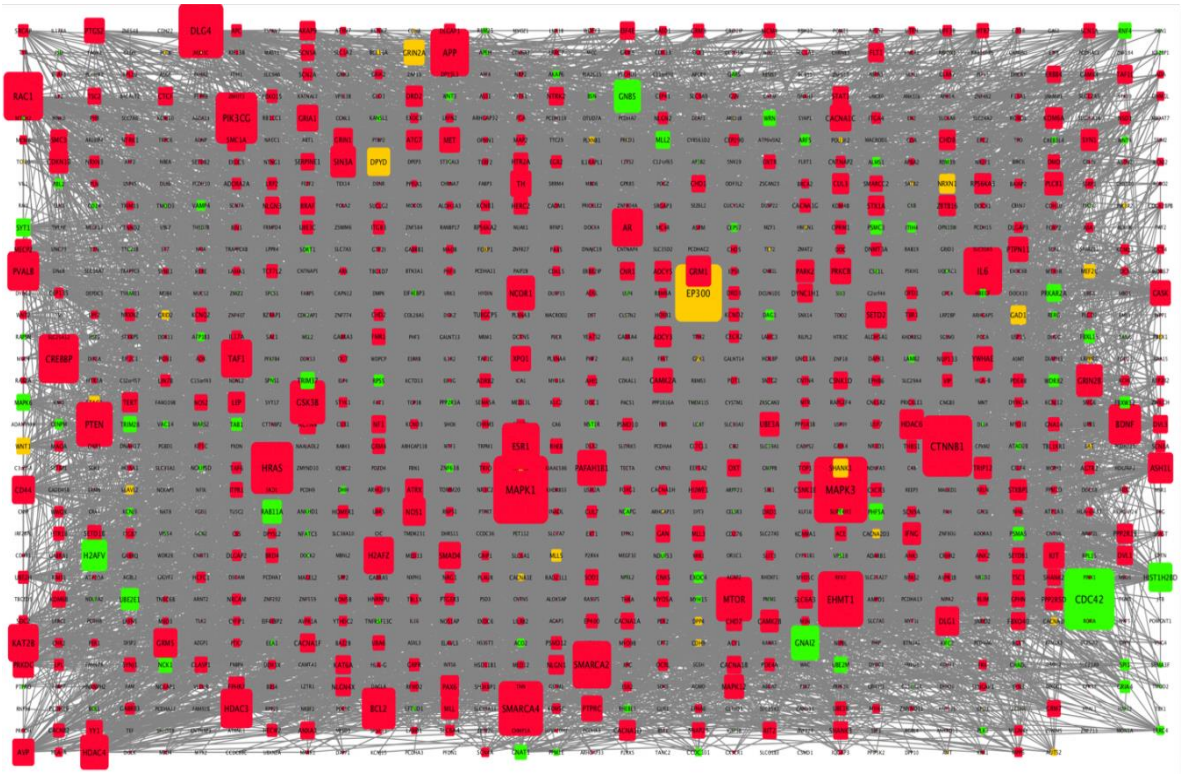

Figure2: Autism/ Intelligence gene interaction network

Figure 3 below illustrates some network statistics generated by Cytoscape.

Clustering coefficient : **0.301**	Number of nodes : **1119**
Connected components : **6**	Network density : **0.013**
Network diameter : **9**	Network heterogeneity : **1.168**
Network radius : **1**	Isolated nodes : **0**
Network centralization : **0.098**	Number of self-loops : **0**
Shortest paths : **1226570 (98%)**	Multi-edge node pairs : **0**
Characteristic path length : **3.310**	Analysis time (sec) : **5.231**
Avg. number of neighbors : **14.842**	

Figure 3: Network Statistics

Figure 4 below illustrates the Degree Distribution chart generated by Cytoscape.

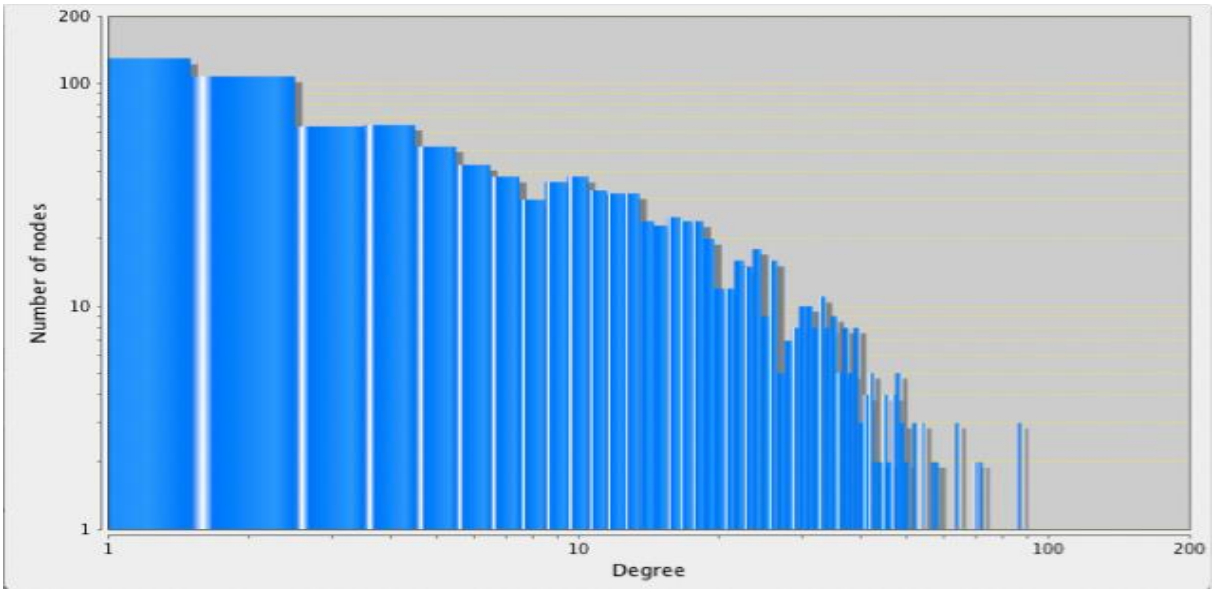

Figure 4: The Degree Distribution Chart.

Figure 5 below illustrates a sample of Autism/ Intelligence database. Degree of connectivity is a measure of the number of in and out links a node has to other nodes.

Name	Lable	Degree
MAPK1	aut	124
MAPK3	aut	121
CDC42	int	113
CTNNB1	aut	112
EP300	aut/int	105
EHMT1	aut	102
DLG4	aut	101
SMARCA4	aut	100
HRAS	aut	95
PIK3CG	aut	94
CREBBP	aut	90
RAC1	aut	89
APP	aut	87
IL6	aut	87
SMARCA2	aut	87
ESR1	aut	85
PTEN	aut	82
MTOR	aut	81
GSK3B	aut	80
BCL2	aut	76
BDNF	aut	73
HDAC3	aut	72
PVALB	aut	72
AR	aut	70
TAF1	aut	70
KAT2B	aut	69
H2AFZ	aut	68
PAFAH1B1	aut	66
GNAI2	int	64
HDAC4	aut	64
NCOR1	aut	64
GNB5	int	61
GRIN2B	aut	60
H2AFV	int	58
SIN3A	aut	58
ASH1L	aut	57

Figure5: Sample of the Database

Cluster Maker *[10]*, is a Cytoscape plugging that implements a wide range of well-known clustering algorithms such as hierarchical clustering, k-means, k-medoid, AutoSOME, SCPS, and native (Java) MCL. Cluster Maker is implemented using Java. It is available along with a friendly GUI that allows the user to specify parameters based on the selected clustering algorithm.

Cluster maker implements three clustering approaches. The first one is hierarchical clustering of genes and gene products. This approach is used to produce dendrogram and heat map in order to facilitate the process of analyzing and visualizing genetic information.

The second clustering approach aims to identify or isolate stable complexes from large sets of interactions. Many built-in clustering algorithms within Cluster maker follow this approach such as Molecular Complex Detection (MCODE), Super Paramagnetic Clustering (SPC), Neighborhood Search Clustering (RNSC) and Neighborhood Search Clustering (RNSC).

The third clustering approach is used to identify similar groups of genes or proteins based on their properties. This approach is needed in the process of inferring the unknown property of a specific protein based on the cluster that it belongs to. For instance, the unknown function of protein X is most likely similar to the function of its neighbor proteins. This approach is implemented via algorithms such as FORCE, TransClust and Affinity Propagation.

Figure 6 below illustrates the result of applying Community clustering (Glay) algorithm on the network. The original network (illustrated in *Figure 1*) consists of almost 27 clusters. The interconnectivity between the genes of both datasets (Autism and Intelligence) is obvious. Only 6 clusters out of 27 clusters do not have mixed genes. Clusters that contain genes from one dataset have stronger link with its group. For instance, cluster number 5 contains four genes, all are strongly linked to Autism specially the one in the center of Cluster 5 (called Gene C3orf58). On the other hand, cluster 12 contains some autism and intelligence genes. The biological process of this group is cell–cell signaling and nervous system development.

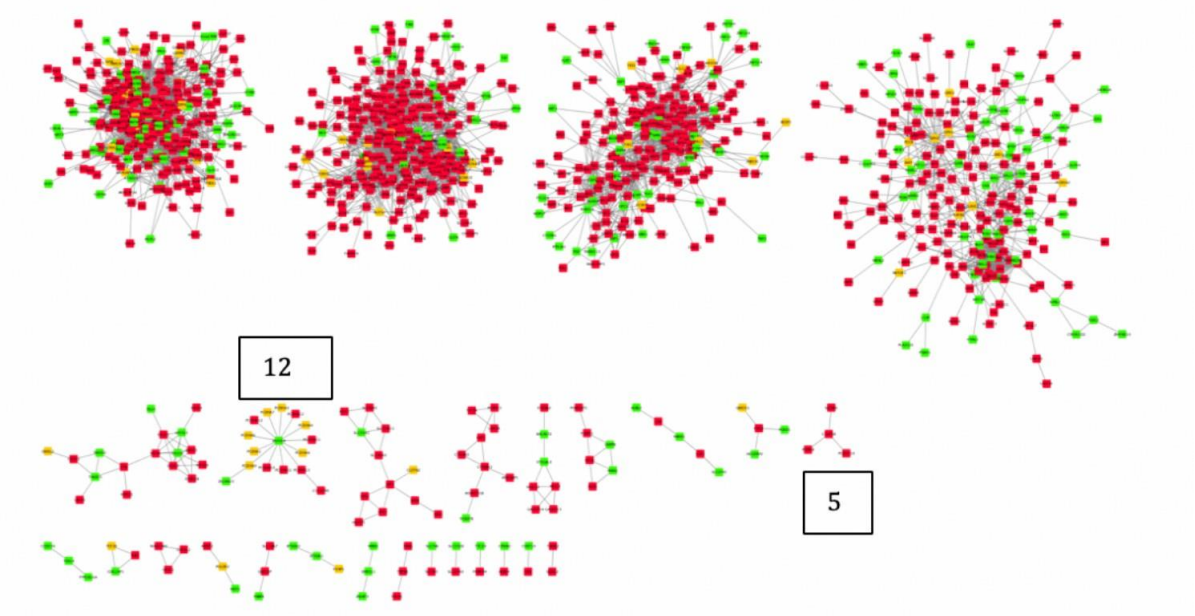

Figure6: Community Clustering

MCL (Markov Cluster Algorithm) *[23]* is an unsupervised graph-clustering algorithm developed by Van Dongen. *Figure 7* below illustrates the Autism/Intelligence clustered network as a result of applying MCL algorithm on the original network.

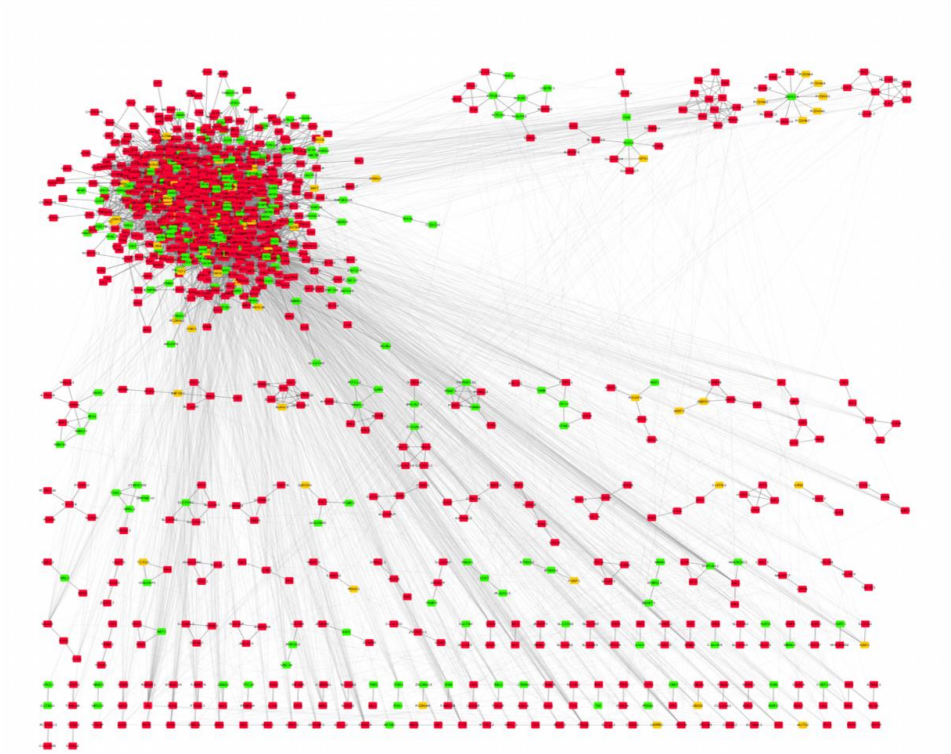

Figure 7: MCL algorithm result on Autism/ Intelligence Cluster

3.2 Enrichment Analysis

In order to gain a deep insight into the biological functions enriched in the Autism/ Intelligence gene network, a functional enrichment analysis, known as, Gene Set Enrichment Analysis (GSEA) is conducted throughout this section. GSEA analysis is considered as a crucial analysis for huge dataset of genes, in order to gain a global perception. Therefore, GSEA is widely used in data interpretation.

Functional enrichment analysis involves four main steps:

1. Defining a gene list: Autism /Intelligence gene set.

2. Performing pathway enrichment analysis: Statistically, identifying pathways enriched in a given gene list. There are many well-documented and freely available online software that can perform enrichment analysis, such as g:Profiler, GSEA and DAVID. g:Profiler is used in this research. It is a web server that provides several tools for analysing and mining gene lists. g:Gost tool (part of g:Profiler) is used to retrieve functional enrichment file. The data retrieved by this tool is regularly from KEGG, Reactome and WikiPathways.
3. Visualizing enrichment analysis: To visualize the resulted g:Profiler enrichment analysis in a network, Enrichment Map is used. Enrichment Map is a Cytoscape plugin for functional enrichment visualization. It takes an enrichment file as an input (a g:Profiler file is used in this research. Other enrichment files like GSEA, BINGO or DAVID can be used as well). The output of the Enrichment Map plugin is a network of interacting pathways.
4. Results summarization: After generating the network using enrichment map, the next step is to define clusters of related nodes and annotate them. Clustering or grouping is useful to create a visual summary and highlight main categories within the network. AutoAnnotate is a Cytoscape plugin that perform both clustering and labelling in one interface. To do so, Auto annotation uses functions from two existing Cytoscape plugins: Cluster Maker and Word Cloud.

Within the interface of Auto Annotate, the user is given the option whether to use any cluster maker algorithm or to manually define clusters. In this research, automatic clustering with the default MCL clustering algorithm is chosen.

After that, the user of Auto Annotate is asked to define labelling source or column that could describe the cluster as well as labelling algorithm. As mentioned earlier, labelling algorithms are inherited from Word Cloud plugin, which is a configurable tool for generating visual summaries for Cytoscape sub networks and displaying the summaries on top of each sub network, or cluster as a tag cloud. Given a cluster and a set of attributes (one column or more of type string) Word Cloud calculates word frequencies for the text of the given set of cluster's attributes, after that it computes the best combination of words to describe the given cluster.

Figure 8 below, illustrates the pathways for Autism and Intelligence genes.

Figure 8: Gene Pathways

Figure 9 below, illustrates a tabular visualization of the enriched pathways with the corresponding number of genes.

Pathway Description	no. gene
term synaptic postsynaptic	109
regional abnormality zygomatic	96
projection arborization growth	73
channel gated activity	68
muscle cardiac contraction	47
transmembrane transporter ion	31
actin cytoskeleton filament	26
histone h4 acetyltransferase	26
dopamine uptake catecholamine	24
cerebellar pallium development	22
midbrain proteoglycans cancer	20
secretion action synthesis	20
insulin secretion hormone	18
seizures consciousness focal	16
learning memory behavior	14
transcriptional regulates genes	14
development type eye	13
eeg spike wave	13
cellular homeostasis divalent	12
membrane repolarization inorganic	12
spindle mitotic microtubule	12
demethylase k9 demethylation	11
aggressive hyperactivity attention	10
monocarboxylic acid anion	10
interferon vascular ring	9
unblocking nmda events	9
plasma bounded lamellipodium	8
cerebral atrophy cerebrum	7
circadian clock rhythm	7
modulating coupled messenger	7
ras signal transduction	7
response organonitrogen compound	7
surface wnt canonical	7
dna sequence specific	6
neuroblast proliferation stem	6
regulation mammary gland	6
response ammonium endomembrane	6
systemic arterial vasopressin	6
truncal obesity weight	6
vesicoureteral reflux esophagus	6

Figure 9: Tabular Visualization

Figure 10 below, is an illustration for the pathway network representation clustered and annotated by AutoAnnotate. The font size reflects the number of genes in the pathway.

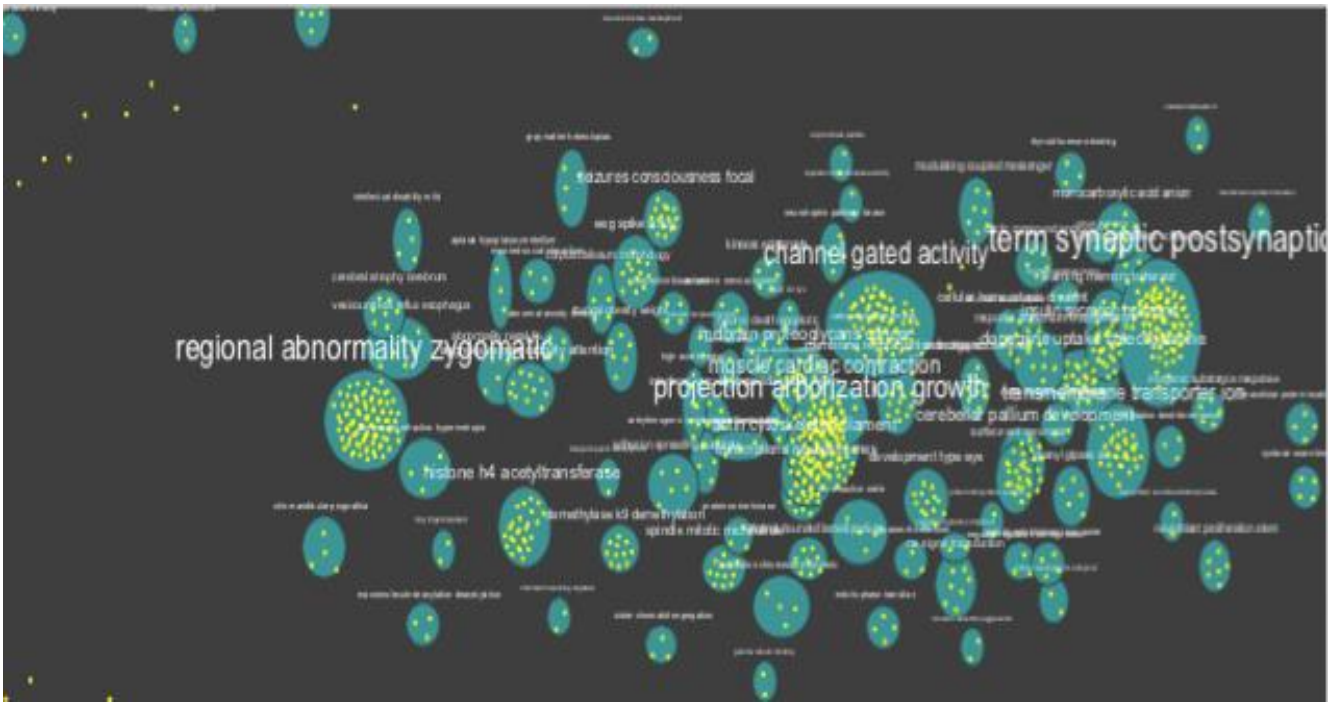

Figure 10: Pathways Network

Note: For enhanced visualization, Auto Annotate allows the user to select the scale of the font size for the tag by the size of its cluster (size reflects number of nodes). Additionally, user has the option to collapse or expand clusters, as well, the layout of the clusters to prevent overlapping.

In summary, we have performed an enrichment analysis, and discovered the main pathways in term of size and number of genes.

Chapter 4: Findings

In order to better understand the primary generic blocks, it is important to get more insights into the discover pathways. In this chapter, genes will be classified based on their categories (Autism, Intelligence or Shared). We will visualize the contribution of every gene category in each pathway.

As previously illustrated in *Figure 10*, "Synaptic Signaling Pathway" contains the lion share of genes. "Synaptic Signaling Pathway" is composed of four pathways:

1. Trans-Synaptic Signaling.
2. Synaptic Signaling.
3. Anterograde Trans-Synaptic Signaling.
4. Chemical Synaptic Transmission.

Synaptic Signaling Pathway is illustrated in *Figure 11* below.

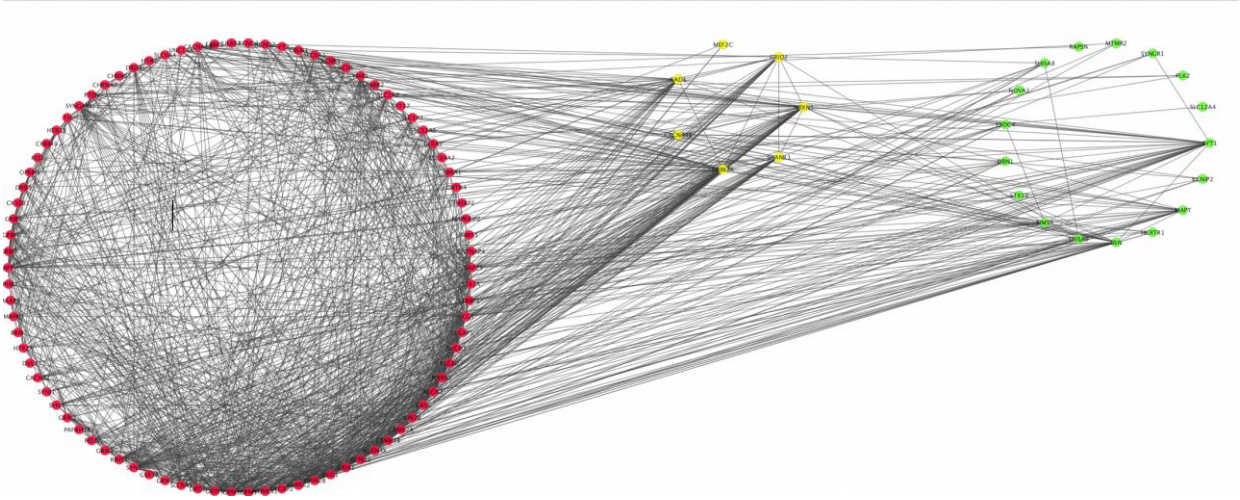

Figure11: Synaptic Signalling Pathway

Genes with highest degree are illustrated in *Figure 12* below.

GRIN2B	aut	0.73648649	71	0.07198011	27.6056338
DLG4	aut	0.73154362	70	0.09447751	26.9285714
GRIN2A	aut/int	0.68553459	60	0.03443251	30.35
GRIN1	aut	0.68987342	60	0.05346174	28.7
GRIA1	aut	0.67701863	58	0.03279648	30.7413793
GRM5	aut	0.67283951	58	0.04095389	30.1551724
SNAP25	aut	0.65662651	54	0.06700511	26.8703704
NRXN1	aut/int	0.65269461	52	0.03214392	30.4615385
SYN1	aut	0.64880952	52	0.03588375	30.5
GRM1	aut	0.63372093	48	0.01944114	32.5
APP	aut	0.62285714	44	0.06207732	26.1363636
CAMK2A	aut	0.61581921	42	0.03208667	32.2142857
NRXN2	aut	0.59562842	39	0.01200238	33.8717949
NLGN1	aut	0.57978723	37	0.01126864	34.1891892
SYT1	int	0.57978723	34	0.01803717	32.4117647
NLGN3	aut	0.56476684	34	0.01061812	32.4411765
DLG1	aut	0.57368421	34	0.01273326	33.1764706
SHANK1	aut/int	0.57068063	34	0.00682041	34.9705882
GAD1	aut/int	0.57368421	32	0.00506628	35.78125
BSN	int	0.57368421	32	0.00704094	36.5625

Figure 12: Genes with Highest Degree

As illustrated in Figure 11 Above, the genes are classified as follow:

1. 24 Intelligence Genes (in green).
2. 93 Autistic Genes (in red).
3. 7 Shared Genes between Intelligence and autism (in yellow).

Having 24 intelligence genes in one pathway reflects the importance of this pathway in the correlation between Autism and Intelligence. This number, 24, is considered a good percentage in reflection to the sample of intelligence genes used in this research.

Figure 11 above shows a high density of connection between the different gene sets. As illustrated, Autistic genes are the most connected. As well, there are representations for Intelligence genes and shared groups.

GRIN2A *[13]* is a shared gene between Autism and Intelligence. GRIN2A gene has a very high degree of connectivity in the signaling pathway (this is connected approximately to 60

genes). The Protein produced by GRIN2A is found in the nerve cells of both the brain and the spinal cord, in particular, brain regions involved in speech and language. Mutations in GRIN2A might result in speech disorders, which might be accompanied with cognitive disability.

GRIN2B *[14]* is an Autism related gene, it is found in the nerve cells of the brain during the brain development before birth. Mutations in this gene could produce intellectual disability and delayed development of speech and motor skills.

Figure 13 below, illustrates the network representation of genes found in Autistic behavior pathway.

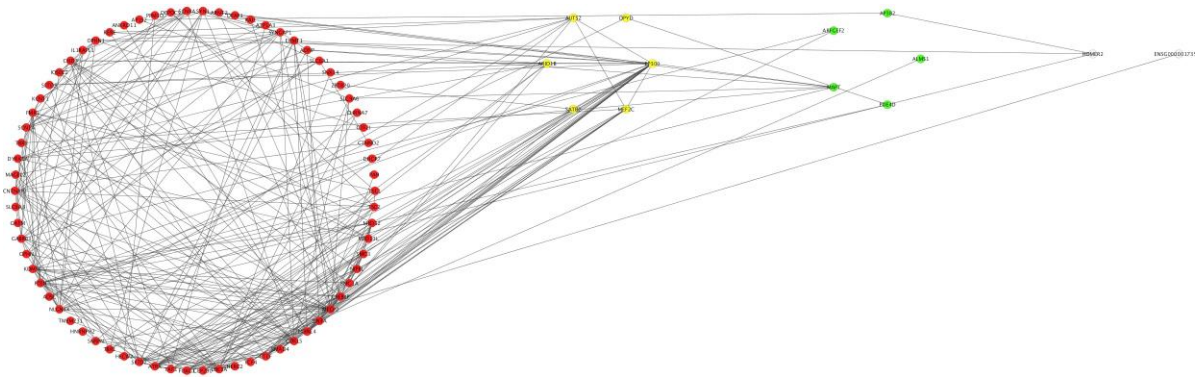

Figure 13: Autistic behavior Pathway

Figure 14 below illustrates the list of genes associated with autistic behavior pathway and its degree of connectivity. EP300 *[15]* gene has the highest value in term of degree. EP300 gene and its products (proteins) are involved in the regulation of cell growth and division. This gene is crucial for normal development before and after birth. Mutations in EP300 were found in

small number of individuals with Rubinstein Taybi syndrome, which can be characterized by Moderate or severe learning difficulties and distinguishing facial features.

EP300	aut/int	28	0.15793682	11.0714286
MECP2	aut	28	0.18727223	12.5
CREBBP	aut	21	0.04938533	12.1904762
UBE3A	aut	19	0.15976226	10.7368421
PTEN	aut	19	0.08573245	13.3157895
ATRX	aut	17	0.02558667	14.1176471
SETD2	aut	16	0.04570338	11.75
KDM5C	aut	16	0.04206775	14.9375
NLGN4X	aut	16	0.06348841	12.4375
CTCF	aut	14	0.02055251	14.2857143
CDKL5	aut	14	0.03410142	12.2142857
CNTNAP2	aut	14	0.04389749	11.7142857
GABRB3	aut	14	0.04152668	11.1428571
SIN3A	aut	14	0.04739707	14
FMR1	aut	14	0.03106745	13.0714286
MED12	aut	13	0.00873154	14.6153846
HDAC4	aut	12	0.05218458	14.0833333
SYNGAP1	aut	11	0.01436608	12.6363636
CHD7	aut	11	0.00809069	12.8181818
MEF2C	aut/int	11	0.0130199	14.0909091
TSC2	aut	10	0.02004105	14.6
SCN1A	aut	10	0.02610135	10.4
FOXG1	aut	10	0.03284564	14.8
SMC3	aut	10	0.03268451	14.6
SMC1A	aut	10	0.03927595	14.6
TAF1	aut	9	0.01445902	14.7777778
SYN1	aut	9	0.01887296	10.8888889
AUTS2	aut/int	9	0.0328539	13.4444444
NIPBL	aut	8	0.01161137	11.625
ARID1B	aut/int	8	0.00303491	13.875
OPHN1	aut	8	0.03389119	7.5
SMAD4	aut	8	0.00288775	15.125
DYRK1A	aut	8	0.01014911	9.375
IQSEC2	aut	7	0.01293124	10.2857143
EHMT1	aut	7	8.65E-04	19.7142857
SNRPN	aut	7	0.03199401	11.8571429
SCN8A	aut	7	0.00229006	10.7142857
MED13L	aut	6	0.00145176	14.1666667
IL1RAPL1	aut	6	0.00285322	11.6666667
MAPT	int	6	0.00876291	12.5
ADNP	aut	6	0.00666842	10.5
CEP290	aut	5	0.05289576	5

Figure 14: List of genes associated with autistic behavior pathway

An interesting study conducted in 2008, "Face–brain asymmetry in autism spectrum disorders" *[11]*, aimed to identify some phenotypes and biomarkers within ASDs. In this study the researchers used dense surface-modelling techniques to compare the facial morphological features of 72 boys with ASD and 128 first-degree relatives to that of 254 unrelated controls (*control refers to person with no disease*). The used pattern matching algorithms could discriminate between the faces of ASD boys and matched controls. Also, it could discriminate between the faces of unaffected mothers of ASD children and matched female controls. The study found considerable differences in the supraorbital, nasal, zygomatic and perioral regions, especially on the right side.

Despite of the small number of shared genes represented in Regional Abnormality Zygomatic Pathway shown in the *Figure 15* below. These shared genes are highly connected with other genes in the network (example: GRIN2A, illustrated in *Figure 16* below). GRIN2A has been mentioned before in signaling pathway. GRIN2B, autism related gene, has the biggest number of connections in this pathway was also seen in signaling pathway.

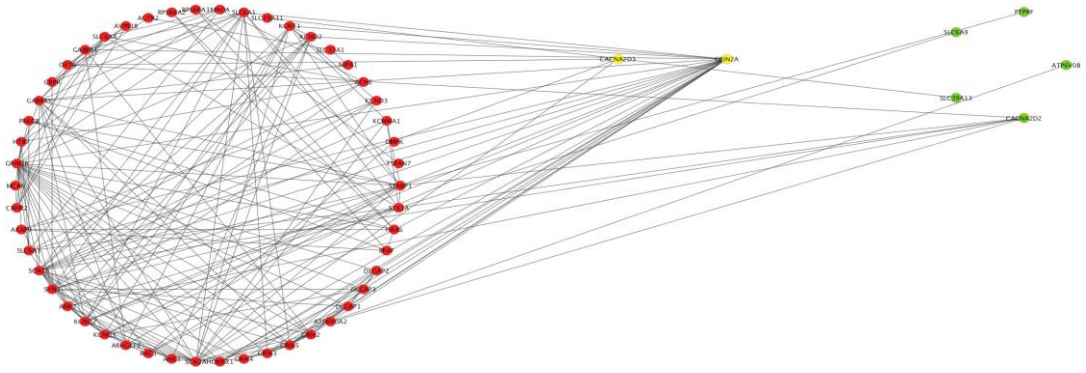

Figure 15: Regional Abnormality Zygomatic

Name	Lable	Degree
GRIN2B	aut	25
GRIN2A	aut/int	22
SCN2A	aut	21
SCN1A	aut	20
STXBP1	aut	16
KCNQ2	aut	15
SYN1	aut	14
GABRB3	aut	13
SLC6A1	aut	13
GRIK5	aut	12
ANK2	aut	11
GRIK2	aut	11
STX1A	aut	11
KCNQ3	aut	10
KCNT1	aut	10
DLGAP1	aut	9
DLGAP3	aut	9
KCND2	aut	9
ARHGEF9	aut	8
HRAS	aut	8
SLC6A3	aut	8
SLC6A4	aut	8
DLGAP2	aut	7
GRIK3	aut	7
GRIK4	aut	7
OXTR	aut	7
AKAP9	aut	6
ANK3	aut	6
CACNA2D2	int	6
HOMER1	aut	6
HTR7	aut	6
ACHE	aut	5
CRHR2	aut	5
GABRB1	aut	5
GRPR	aut	5
KCND3	aut	5
KCNMA1	aut	5
MAOA	aut	5

Figure 16: Regional Abnormality Zygomatic Genes

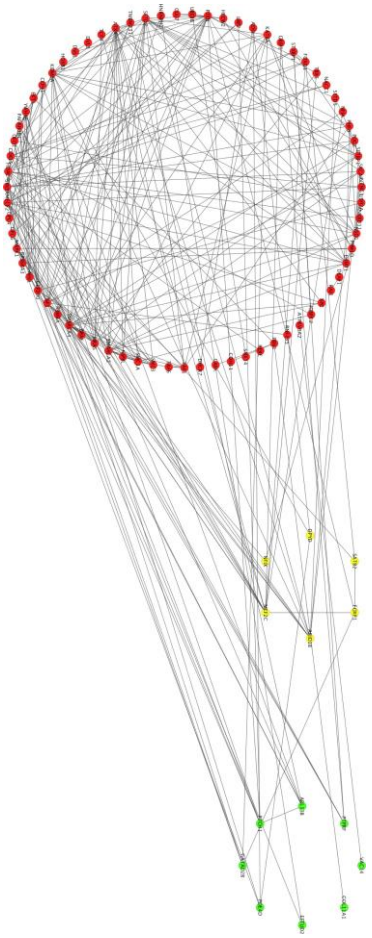

Figure 17: Figure 2 abnormality of the nasal alae pathway

The Figures below (*Figure 18*, *Figure 19*), represent motor delay and language pathways. In both genetic pathways, EP300 gene is a shared gene with a clear role in connecting Autism and Intelligence genes. This gene was found before in Autistic behavior pathway. We have highlighted the proven relation between EP300 and Rubinstein-Taybi syndrome. Despite of the weak score given for this gene by SFARI gene database, another statistic has suggested that around 60% of individuals with Rubinstein-Taybi syndrome show some autistic features especially in term of repetitive behaviors *[12]*.

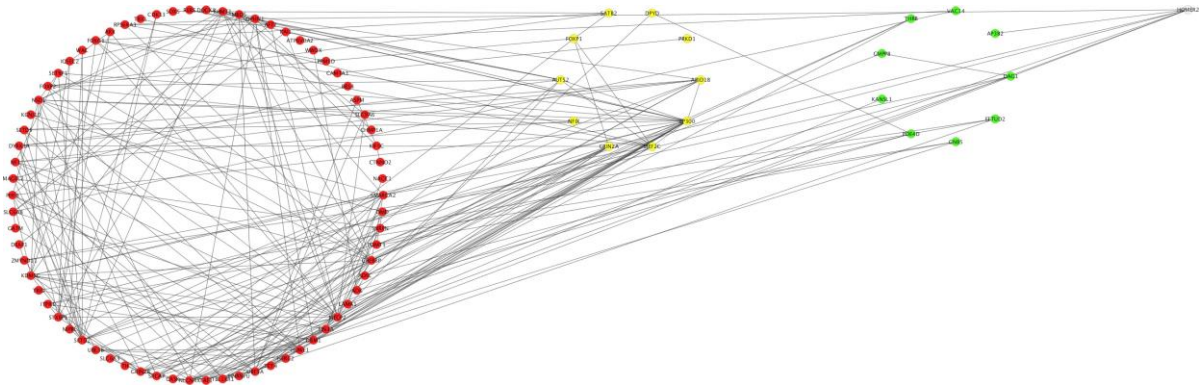

Figure 18: Language pathway

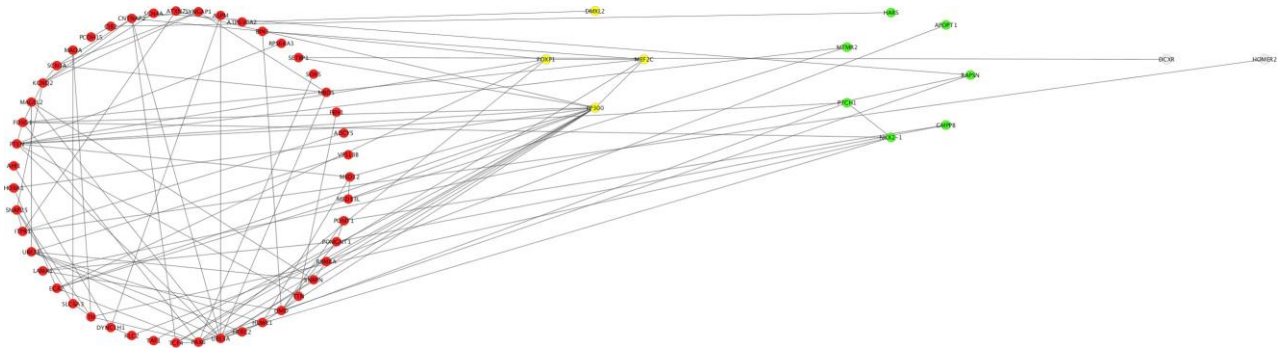

Figure 19: Motor Delay

It is important to note that the pathway for head development has revealed some genes that belong to glutamate-dependent receptors group (examples: GRIN2A, GRIN2B, GRID2). This finding might indicate a relation between head development and Synaptic Signaling. *Figure 20* below illustrates the head development pathway.

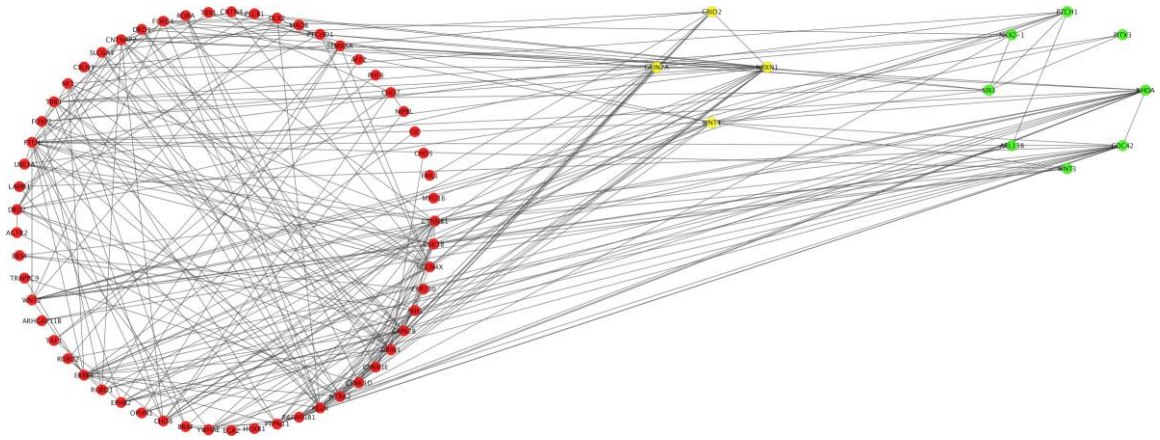

Figure 20: Head Development Pathway

Chapter 5: Conclusion

In this research, a systematic approach was conducted for mining large biological networks, with a focus on finding interesting correlations between the two poles: Autism and Intelligence. We have mainly focused on applying graph theory and data mining techniques on genetic interaction network.

It is important to highlight the following fact: Genes "interact genetically" if they affect each other's function. For instance, a gene could contain a harmless mutation but when it is combined with another mutation, the combination may cause a serious condition. Such interactions form the genetic interaction networks.

In this research, we used Cytoscape for visualizing and analyzing Autism/Intelligence genes' interaction network. Cytoscape provides a powerful toolbox that uses graph theory and data mining techniques to analyze complex biological networks. For instance, Cytoscape can be used effectively in hubs identification as well as determining the role of a gene or protein with unknown function.

In this research, the mining of gene networks has played an important role in identifying causative genes for diseases or syndromes. As well, mining is used to discover hidden relations between different health conditions. We have identified some shared genes between Autism and Intelligence. We also identified a list of pathways that express the strong correlation between these two very different mental conditions (Autism and Intelligence).

We have noticed the strong contribution of some genes in different pathways. For instance, EP300 was seen in motor delay and autistic behavior pathway. The same gene has a strong relation with Rubinstein Taybi syndrome. Glutamate-dependent receptors group were also seen in different pathways.

This research has identified and concluded the shared genes and hubs (genes with high degree of connectivity) from statistical point of view. We believe that, with the advancement of tools and technologies, as well, the contribution of medical experts in the field, this research and others will have a positive impact on the study of genetics. As a future work, spotting the light on SNPs (gene mutations) of the shared genes, will provide more insights on the correlation between genes in Autism and Intelligence, resulting in more specific use cases.

References

[1] Autism spectrum. (2019, September 25). Retrieved from: https://en.wikipedia.org/wiki/Autism_spectrum.

[2] Is Autism Associated with Higher Intelligence? (n.d.). Retrieved from https://www.appliedbehavioranalysisedu.org/is-autism-associated-with-higher-intelligence/.

[3] Crespi, B. J. (2016). Autism As a Disorder of High Intelligence. *Frontiers in Neuroscience*, *10*. doi: 10.3389/fnins.2016.00300.

[4] Ruthsatz, J., Petrill, S. A., Li, N., Wolock, S. L., & Bartlett, C. W. (2015). Molecular Genetic Evidence for Shared Etiology of Autism and Prodigy. *Human Heredity*, *79*(2), 53–59. doi: 10.1159/000373890.

[5] Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., Mcintosh, A. M., Deary, I. J. (2018). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry*, *24*(2), 169–181. doi: 10.1038/s41380-017-0001-5.

[6] Mitchell-Olds, T., Mojica, J., & Wang, B. (2018). Faculty of 1000 evaluation for Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. F1000 - Post-Publication Peer Review of the Biomedical Literature. doi: 10.3410/f.733502533.793549437.

[7] Ruzich E, Allison C, Chakrabarti B, Smith P, Musto H, Ring H, et al. (2015) Sex and STEM Occupation Predict Autism-Spectrum Quotient (AQ) Scores in Half a Million People. PLoS ONE 10(10): e0141229.

[8] January 2020, at 21:23 (UTC)), Available at: https://en.wikipedia.org/wiki/Singlenucleotide_polymorphism.

[9] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks: Cold Spring Harbor Laboratory Press.

[10] Morris, J.H., Apeltsin, L., Newman, A.M. *et al. clusterMaker:* a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics* **12,** 436 (2011) doi:10.1186/1471-2105-12-436.

[11] Hammond, P., Forster-Gibson, C., Chudley, A. et al. Face–brain asymmetry in autism spectrum disorders. Mol Psychiatry 13, 614–623 (2008) doi:10.1038/mp.2008.18.

[12] Available at: https://www.findresources.co.uk/the-syndromes/rubinstein-taybi-syndrome/behavioural-characteristics/autism-spectrum-disorder

[13] Available at https://www.genecards.org/cgi-bin/carddisp.pl?gene=GRIN2A

[14] Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=GRIN2B

[15] Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=EP300

[16] Christian von Mering, Martijn Huynen, Daniel Jaeggi, Steffen Schmidt, Peer Bork, Berend Snel, STRING: a database of predicted functional associations between proteins, Nucleic Acids Research, Volume 31, Issue 1, 1 January 2003, Pages 258 261, https://doi.org/10.1093/nar/gkg034.

[17] Ruthsatz, J., & Urbach, J.B., Child prodigy: A novel cognitive profile places elevated general intelligence, exceptional working memory and attention to detail..., Intelligence (2012), doi:10.1016/j.intell.2012.06.002.

[18] Available: https://stanfordbinettest.com/

[19] Available at: https://en.wikipedia.org/wiki/Autism-spectrum_quotient.

[20] Sikela, J.M., Searles Quick, V.B. Genomic trade-offs: are autism and schizophrenia the steep price of the human brain?. Hum Genet 137, 1–13 (2018). https://doi.org/10.1007/s00439-017-1865-9.

[21] Pua EPK, Malpas CB, Bowden SC, Seal ML. Different brain networks underlying intelligence in autism ML. Different brain networks underlying intelligence in autism spectrum disorders. Hum Brain Mapp. 2018;00:1–10. https:// spectrum disorders. Hum Brain Mapp. 2018;39:3253–3262.
https://doi.org/10.1002/hbm.24074.

[22] Available at: https://www.sfari.org/resource/sfari-gene/.

[23] Stijn van Dongen. *A cluster algorithm for graphs*. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May 2000. Available at: https://micans.org/mcl/.

[24] Pérez Velázquez JL and Galán RF (2013) Information gain in the brain's resting state: A new perspective on autism. Front. Neuroinform. 7:37. doi: 10.3389/ fninf.2013.00037.

[25] Available at: https://en.wikipedia.org/wiki/Magnetoencephalography

[26] Baio J, Wiggins L, Christensen DL, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ 2018;67(No. SS-6):1–23. DOI: http://dx.doi.org/10.15585/mmwr.ss6706a1