

**Investigating the role of Schizophrenia-
associated gene expression in the developing
human brain using Machine Learning**

Katie Kelly

Technological University Dublin

M.X.

2021

Sample Spine of Thesis

M.X.

Katie Kelly

2021

*Investigating the role of schizophrenia-associated
gene expression in the developing human brain
using Machine Learning*

A Thesis Presented for the Award of Masters by Research

Katie Kelly (B.Sc.)



Technological University Dublin – Tallaght Campus

Department of Science

For Research Carried Out Under the Guidance of

Dr Eugene Hickey & Dr Therese Murphy & Kevin Nolan

Submitted to Technological University Dublin

April 2021

DECLARATION

I certify that this thesis which I now submit for examination for the award of _____ is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work.

This thesis was prepared according to the regulations for graduate study by research of the Technological University Dublin (TU Dublin) and has not been submitted in whole or in part for another award in any other third-level institution.

The work reported in this thesis conforms to the principles and requirements of the TU Dublin's guidelines for ethics in research.

(The following sentence may be deleted if access to the thesis is restricted according to Section 4.8 of the TU Dublin Research regulations)

TU Dublin has permission to keep, lend or copy this thesis in whole or in part, on condition that any such use of the material of the thesis is duly acknowledged.

Signature _____ Katie Kelly _____ Date _____

Candidate

Acknowledgements

I would like to thank my supervisors Dr. Eugene Hickey & Dr. Therese Murphy who have been supportive and encouraging throughout this project. In particular I would like to thank Eugene for his positive outlook and assurance which gave me confidence in my research.

Dedicated to

Abstract

Schizophrenia is a debilitating condition which affects 1% of the population and causes significant financial burden. It is a complex mental disorder where some individuals show subclinical cognitive symptoms before psychosis onset in adolescence. Although there are treatments even the best options perform sub optimally and target only a portion of the symptoms. There has been a large amount research into schizophrenia and its underlying mechanisms, and its systems are still unknown. Genome wide association studies have been able to identify 145 loci and through incorporating brain expression, genomic fine mapping, and chromosome conformation data to discover causal genes within 33 loci.

Unlike other neurodevelopmental disorders schizophrenias symptoms do not present themselves until adolescence because of this it is important to study it completely throughout development to see what commences the symptoms. Using ABA's BrainSpan resource of development, schizophrenia-associated genes gene expression across brain areas were able to be studies over development using k-means weighted gene correlation network analysis the genes were separated into modules where gene enrichment was applied.

The aim of this study was to discover underlying mechanisms at developmental stages to identify new therapeutic targets to treat and manage schizophrenia. Throughout the developmental stages positive regulation of macrophage proliferation was enriched as well as cellular response to catecholamine stimulus. These results mirror those of previous studies.

As part of the weighted gene correlation network analysis, hub genes were identified, GRIN2A and CA8 were found to be significant in multiple Developmental Stages. Because of this they should be studied further in the future.

List of Abbreviations

5HT	-	5- hydroxytryptamine receptors
ABA	-	Allen Brain Atlas
AHBA	-	Allen Human brain Atlas
ADHD	-	Attention-Deficit Hyperactivity Disorder
ANPs	-	Antipsychotic Naïve Patients
ASD	-	Autism Spectrum Disorder
BP	-	Bipolar Disorder
CNVs	-	Copy Number Variants
DMAs	-	Dopamine modulating antipsychotics
eQTL	-	Expression quantitative trait loci
FDR	-	False Discovery Rate
GABA	-	Gamma aminobutyric acid
GCN	-	Gene co-expression networks
GMV	-	Grey Matter volume
GO	-	Gene Ontology
GS	-	Gene significance
GWAS	-	Genome Wide Association Study
HCs	-	Healthy controls
HG	-	Hub Genes
IBD	-	Inflammatory bowel disease
ID	-	Intellectual disability
IL1 β	-	Interleukin 1 beta
IL6	-	Interleukin 6
INDELS	-	Insertions and deletions
ISH	-	In Situ Hybridisation
LD	-	Linkage disequilibrium
LoF	-	Loss of Function
LMD	-	Laser Microdissection
LSD	-	Lysergic acid diethylamide
ME	-	Module eigengene
MIA	-	Maternal Immune activation
MK-801	-	Dizocilpine
MM	-	Module membership
MNI	-	Montreal Neurological Institute
MS	-	Module significance
MRI	-	Magnetic Resonance Imaging
mRNA	-	Messenger ribonucleic acid
NS	-	node significance
NMDAR	-	N-methyl-D-aspartate receptors
OTU	-	Operational Taxonomic Units
PCA	-	Principal component analysis
PCP	-	Phenylcyclidine
PEN	-	Polyethylene naphthalene
PFC	-	Prefrontal cortex
PNNs	-	perineuronal nets
PV	-	Parvalbumin
QC	-	Quality control

RNA	-	Ribonucleic acid
SNPs	-	Single Nucleotide Polymorphisms
TO	-	Topological Overlap
TOM	-	Topological Overlap Matrix
TNF- α	-	Tumour necrosis factor alpha
TNF- β	-	Tumour necrosis factor beta
UHR	-	Ultra high risk
VTA	-	Ventral Tegmental Area
WES	-	Whole Exome Sequencing
WGS	-	Whole Genome Sequencing
WGCNA	-	Weighted Correlation Network Analysis

Table of Contents

For Research Carried Out Under the Guidance of.....	3
DECLARATION.....	4
Acknowledgements	5
Abstract	7
List of Abbreviations	8
Table of Contents	10
Chapter 1 – Introduction to schizophrenia.....	13
1.1 Schizophrenia's etiology	14
1.2 Understanding of schizophrenia to date.....	15
1.2.1 Immune System	15
1.2.2 Neurodevelopmental hypothesis.....	18
1.2.3 Dopamine hypothesis	19
1.2.4 Glutamate hypothesis	21
1.2.5 Gamma-aminobutyric Acid (GABA)	23
1.2.6 Serotonin Hypothesis (5-hydroxytryptamine, 5-HT).....	24
1.3 Environmental risks for schizophrenia	25
1.4 MRI Findings.....	27
1.5 Genetics of Schizophrenia	29
1.6 Aim of the project.....	32
Chapter 2 - Methods	34
2.1 Collation of Schizophrenia-associated Genes.....	34
2.2 Allen Brain Atlas	35
2.2.1 Brainspan Atlas of the Developing Human Brain	37
Table 1: Age categories from the developmental stages for ABA's resource BrainSpan available for download in R.....	38
2.2.2 ABA and its application in research	39
2.2.3 Collation of schizophrenia-associated genes from ABA	40
2.3 Machine Learning and Clustering	41
2.3.2 K-means clustering and NbClust	42
2.3.3 Kmeans analysis of schizophrenia-associates genes.....	43
2.4 Co-expression Network Analysis	43
2.5 Weighted correlation network analysis (WGCNA)	44
2.5.1 Network analysis using WGCNA.....	45
2.5.2 Networks and their applications	46
2.5.3 Cytoscape	47
2.5.4 Visualisation of modules using data from WGCNA in Cytoscape	48
2.6 Gene Ontologies (GO).....	49
Chapter 3 -Results	50
3.1 Data Pre-processing	50
Table 2: Wide-format of the schizophrenia-associated genes data frame. Brain areas available from ABAs Brainspan are the columns names in bold and the schizophrenia-associated genes are the row names. Each cell contains the scaled gene expression for each brain area	51
3.2 Unsupervised learning using K-means analysis on the schizophrenia-associated genes	52
3.3 Determining Optimal cluster Number using NbClust.....	53
3.4 WGCNA on schizophrenia-associated genes and Network Visualisation using Cytoscape.....	67
Table 8: The Soft Thresholding power of each developmental stage calculated using WGCNA and shown in Figure 3.4.1	67
3.5 Intramodular Hub Genes and Network Analysis	73
Table 9: Gene Functions and the phenotypes they are involved in for each hub genes identified by the WGCNA function in R when performed on the 104 schizophrenia-associated genes identified by Pardiñas et al. for the five developmental stages available on ABA's Brainspan.	74
3.6 Cytoscape and Network Visualisation	82
3.6.1 Network Visualisation of the schizophrenia-associated genes in Developmental Stage One (<i>in Utero</i>)	83

3.6.2 Network Visualisation of the schizophrenia-associated genes in Developmental Stage Two (0-2 years)	Error! Bookmark not defined.
3.6.3 Network Visualisation of the schizophrenia-associated genes in Developmental Stage Three (3-11 years)	Error! Bookmark not defined.
3.6.4 Network Visualisation of the schizophrenia-associated genes in Developmental Stage Four (12-19 years)	Error! Bookmark not defined.
3.6.5 Network Visualisation of the schizophrenia-associated genes in Developmental Stage Five (>19 years).....	Error! Bookmark not defined.
3.7 Gene Ontologies	137
Table 10: Gene Ontologies of the top enriched gene ontologies in Stage One using anRichment...	138
Table 11: Gene Ontologies for the most enriched ontologies in Stage Two using anRichment.....	139
Table 12: Gene Ontologies for the most enriched ontologies in Stage Three using anRichment....	140
Table 13: Gene Ontologies for the most enriched ontologies in Stage Four using anRichment.....	141
Table 14: Gene Ontologies for the most enriched ontologies in Stage Five using anRichment	143
4.0 Discussion	145
4.1 K-means analysis on the schizophrenia-associated genes	146
4.2 Weighted Gene Correlations Network Analysis on schizophrenia-associated genes	147
4.3 Visualisation of modules using data from WGCNA.	147
4.3.1 Developmental Stage One - Prenatal	148
4.3.2 Developmental Stage Two – Infant (0-2 years)	148
4.3.3 Developmental Stage Three – Child (3- 11 years).....	148
4.3.4 Developmental Stage Four – Adolescent (12-19 years)	149
4.3.5 Developmental Stage Five – Adult (>19 years).....	150
4.3.6 Recurring Hub Genes Across Developmental Stages.....	151
4.4 Gene Ontologies	152
4.5 Limitations of study	155
4.6 Future Directions	156
4.7 Conclusion	158
5.0 Bibliography	159
6.0 Appendix	170
Table 1: Schizophrenia-associated gene set from the 145 loci identified by Pardiñas et al. which are available in ABA’s BrainSpan resource.	170
Table 2: Cluster assignments for each schizophrenia-associated gene over the five stages using the kmeans function available in R.....	171
Table 3: Gene Ontologies identified for the Blue module in developmental stage One using the anRichment function as part of WGCNA in R using the default settings.....	175
Table 4: Gene Ontologies for the Brown module in developmental stage One using the anRichment function of WGCNA on the schizophrenia-associated genes.....	176
Table 5: Gene Ontology Turquoise Module for developmental stage One using the anRichment function as part of WGCNA in R using the default settings.....	180
Table 6: Gene Ontology for Blue Module in Developmental Stage Two One using the anRichment function as part of WGCNA in R using the default settings.....	182
Table 7: Gene Ontologies for Brown Module in Developmental Stage Two One using the anRichment function as part of WGCNA in R using the default settings.....	185
Table 8: Gene Ontologies for Turquoise Module in Developmental Stage Two One using the anRichment function as part of WGCNA in R using the default settings.....	190
Table 9: Gene Ontologies for Green Module in Developmental Stage Two using the anRichment function as part of WGCNA in R using the default settings.....	191
Table 10: Gene Ontology for Blue Module Stage Three One using the anRichment function as part of WGCNA in R using the default settings	193
Table 11: Gene Ontology for Brown Module Stage Three One using the anRichment function as part of WGCNA in R using the default settings	194
Table 12: Gene Ontologies for Blue Module Stage Four using the anRichment function as part of WGCNA in R using the default settings.....	197
Table 13: Gene Ontologies for Turquoise Module Stage Four using the anRichment function as part of WGCNA in R using the default settings	200
Table 14: Gene Ontology for Blue Stage 5 using the anRichment function as part of WGCNA in R using the default settings	202
Table 15: Gene Ontology for Brown Module Stage 5 using the anRichment function as part of WGCNA in R using the default settings.....	209

Table 16: Gene Ontology for Yellow Module Stage 5 One using the anRichment function as part of WGCNA in R using the default settings.....	212
---	-----

Chapter 1 – Introduction to schizophrenia

Schizophrenia is a debilitating psychiatric condition that manifests itself early in adolescence and can last a lifetime, it has a 1% global prevalence and has significant societal and economic costs as well as causing substantial mortality and morbidity. (1) Its complex nature is believed to originate from a mixture of genetic and environmental factors such as prenatal exposure to infection and lack of nutrients which cause disruptions during early brain development *in utero*. (2,3) Even though there are many hypotheses about what causes schizophrenia, there is not yet a clear understanding of its aetiology, this lack of understanding of schizophrenia is evident in its treatments which haven't significantly advanced since the introduction of the first-generation antipsychotic medication such as Chlorpromazine in the 1950s (4). This complex condition presents with three modes of clinical features namely: positive symptoms (psychoses manifesting as delusions and hallucinations, paranoia, hyperactivity and agitation), negative symptoms (social withdrawal, lack of motivation, asociality, avolition, affective flattening, consummatory and anticipatory anhedonia, and alogia), and cognitive symptoms (trouble with critical thinking, working memory and difficulty integrating feelings, thoughts and behaviour, attention and vigilance, verbal learning, reasoning and problem solving, and social cognition) as well as motor disturbances which regularly results in a poor quality of life (5–7). The presentation of symptoms is heterogeneous which makes schizophrenia both difficult to diagnose and treat (8). The negative and cognitive symptoms are chronic and are closely related to functional outcomes, and contribute greatly to illness burden, (6) the positive

symptoms usually relapse and remit. Despite the progress in the understanding of some of the fundamental mechanisms involved in schizophrenia's aetiology, the current treatments available to patients come with serious side effects, inconsistent efficacy, and lack of evidence that treatment substantially improves outcomes (9,10). At present, schizophrenia's treatments consist of antipsychotic drugs, social support, rehabilitation, and psychological therapies (8). Current antipsychotics are associated with serious limitations firstly, around 30% of sufferers are treatment-resistant, secondly they mainly ameliorate positive symptoms only leaving cognitive and negative symptoms untreated and lastly antipsychotics trigger both neurological and metabolic side effects (11). As a result, there is a clear need for more efficient and effective treatments as well as uncovering a model for prediction of efficacy as currently determining the most effective treatment of schizophrenia is a trial and error method (7,12). It is important to study patients at several clinical stages to give insight into the effects of schizophrenia itself, its progression and what alterations are caused by the pharmacological treatments (13).

1.1 Schizophrenia's aetiology

At present, there are no clinical diagnostic tests available for schizophrenia so diagnosis relies on clinical observations as well as self-report (14). Schizophrenia remains incurable and the best outcome continues to be controlling of symptoms and preservation of independence and functionality (9). Until there is a more complete understanding of schizophrenias aetiology, there is little hope for improving diagnosis, predicting susceptibility, management, and treatments for those with schizophrenia.

1.2 Understanding of schizophrenia to date

Schizophrenia's complicated and unknown underlying mechanisms has meant that there has been no fundamental innovation in schizophrenia treatments since the introducing of first-generation antipsychotics in the 1950s (15). As treatments mainly target the positive symptoms there is a clear need for a focus into cognitive and negative symptom domains. These types of studies could lead to novel endophenotypic markers which could promote novel treatment discovery and could initiate concurrent medication strategies with current antipsychotics. (15)

1.2.1 Immune System

Schizophrenia's pathogenesis is not fully understood, and though animal models can be used to understand elements, the human central nervous system (CNS) and immune system are much more complex and intricate (16). These systems share common features in developmental mechanisms, so therefore CNS and immune system dysregulation should be studied in humans (16,17). The immune hypothesis of schizophrenia has been around for a long time and is supported by epidemiological, genetic, imaging and biomarker studies (17). The accumulating evidence that anti-inflammatory and immunosuppressive medications are effective treatments and that autoimmune conditions and immune activation are risk factors for developing schizophrenia provides perhaps the most convincing evidence of immune system involvement (18,19).

Dysregulation of the innate and adaptive immune system have been identified by epidemiological, genetic, postmortem and therapeutic studies and are likely to contribute to some of the symptoms of schizophrenia (20). Though there have been a large number of studies with significant funding devoted to better understand schizophrenia outcomes remain substandard and hope remains in advances in psychoneuroimmunology and other advanced technological research areas to provide more consistent and successful management of schizophrenia (12). Some autoimmune conditions display neuropsychiatric symptoms suspected to be caused by brain reactive antibodies (21). Schizophrenia and autoimmune diseases are often comorbid likely because of their genetic overlap, (22) the genetic overlap affects common underlying pathways which entail inflammatory immune response antibodies which can attack brain tissue (23)(22). A national cohort found if a patient had a prior autoimmune disease they are 29% more likely to develop schizophrenia in adolescence (21). Maternal Immune Activation (MIA) can disrupt normal fetal brain development and has been linked to schizophrenia for over a century, it is estimated that if MIA could be averted 30% of schizophrenia cases would be averted (20). Lower levels of acute-phase proteins in neonates which increases the susceptibility of infection have also been hypothesized to increase the risk of psychosis in adulthood (22). Patients experiencing acute episodes of schizophrenia often have increased levels of Interleukin-1- beta (IL-1 β), Interleukin-6 (IL-6), and transforming growth factor-beta (TGF- β) (23). In unmedicated patients, tumor necrosis factor-alpha (TNF α) protein levels and *IL-1 β* messenger RNA (mRNA) is seen to be elevated (24).

Within the body exists a dynamic population of gut microbes which houses many bacteria approximately 10^{14} cells, the biological biodiversity is established in the first couple of months of existence, and has a continuous role throughout life, and is very susceptible to environmental factors (25). The gut microbiome can control how the brain behaves and functions via the microbiota-gut-brain (MGB) and it has been reported to be related to changes in cognition, anxiety, and memory, as well as development, maturation of immune, neural and endocrine systems in animal models (26). These physiological and behavioural processes are often impaired in people with schizophrenia. A high α -diversity score is usually a sign of good health (27). In a study performed by Zheng et al. medicated and unmedicated patients with schizophrenia it has been observed that they have a decreased α -diversity in their microbiome when compared to healthy controls (HCs) (26). It was also found that Veillonellaceae and Lachnospiraceae found in the microbiome environment were associated with symptom severity in schizophrenia (26). β -diversity analysis of schizophrenia patients and HCs found clear differences in the compositions of each microbiome by looking at operational taxonomic units (OTU) levels (26). In one study when the linear discriminating analysis effect size was applied to 77 differential OTUs it was observed that 23 out of the 77 OTUs saw an increase in patients with schizophrenia patients when compared to controls. The OTUs belonged to the bacterial families Veillonellaceae, Coriobacteriaceae, Bacteroidaceae, and Prevotellaceae, the other 54 OTU levels were seen to be decreased in patients with schizophrenia (Lachnospiraceae, Norank, Ruminococcaceae and Enterobacteriaceae). (26)

1.2.2 Neurodevelopmental hypothesis

Epidemiological, basic, and clinical neuroscience research has presented evidence that schizophrenia is of neurodevelopmental origin (28). This hypothesis is now widely accepted but what differentiates schizophrenia from other neurodevelopmental conditions is its time of onset, in adolescence (29). Autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD) and intellectual disabilities (ID) characteristically present themselves much earlier in childhood (29). Schizophrenia shares many phenotypic and clinical similarities and is often comorbid with these neurodevelopmental disorders but because of its delayed presentation, they were not initially linked (29). Before they were connected it was then hypothesised that schizophrenia may be a neurodegenerative disorder, but when post-mortem studies failed to identify traumatic, neurotoxic, or neurodegenerative mechanisms in the brain this theory was disproven, and the neurodevelopmental hypothesis replaced it.

(29) In neonatal primates and rodents, prenatal cortical lesions were shown to lead to the emergence of abnormalities which mimicked schizophrenia in early adolescence, proving that early developmental abnormalities could have an impact on cortical function in later life, making the neurodevelopmental hypotheses for schizophrenia plausible (30). Instead of each neurodevelopmental disorder being viewed independently an alternative view was proposed, that these neurodevelopmental disorders lie on an etiological continuum with a diverse range of outcomes that follows from early brain development disturbances because of shared genetic variants and environmental factors (29). These neurodevelopmental disorders are diagnosed based on symptoms, the timing of onset, severity/persistence, and abnormal brain

development (29). Before the first psychotic episode, schizophrenia presents itself very similarly to the other neurodevelopmental disorders, but only the negative and cognitive symptoms. There are several rare copy number variants (CNVs), genes affected by loss of function (LoF) mutations, genes enriched with 3 nonsynonymous mutation and alleles that have significant associations with schizophrenia, ASD, ADHD and ID which represent direct outcomes of the rare pathogenic mutations that they share (29). This would also suggest that the risk of developing positive symptoms is not mediated by cognitive impairment.

1.2.3 Dopamine hypothesis

Dopamine is a catecholamine neurotransmitter in the brain which regulates critical neurological processes such as cognition, motor control, reward and learning (31). In the 1950s Chlorpromazine an antipsychotic drug and affective antagonist for the D₂, D₃, and D₅ receptors was released and the treatments dopamine receptor antagonists have remained the most prevalent therapeutic (11). Chlorpromazine controlled the positive symptoms of schizophrenia patients and the theory that dopamine alterations in the mesolimbic pathway caused positive symptoms was strengthened and confirmed (11). Other key evidence supporting the dopamine hypothesis was when amphetamines were administered which increase the extracellular concentrations of dopamine and psychotic symptoms like schizophrenias appeared (32). This evidence was reinforced when treatments that depleted the concentration of dopamine such as alpha-methyl-para-tyrosine and reserpine were shown to reduce psychotic symptoms (32). These antipsychotics target other dopamine receptors,

serotonin, norepinephrine, acetylcholine and histamine as well (32). It is often seen has that in the associative striatum there is an increased dopamine synthesis capacity for people who have psychotic disorders including schizophrenia (33). The increased dopamine synthesis is detectable in ultra-high-risk (UHR) subjects and before early symptoms of people who eventually develop schizophrenia thus are not a consequence of antipsychotic exposure or psychotic episodes (33). Rodent models have been able to replicate this, these models have also shown that increased synthesis and release of striatal dopamine can be a result of acute stressors and inflammatory challenges *in utero* (34). These developmental disruptions cause the dopamine system to become hyper-responsive later in life, in the rodent equivalent of adolescence (34). Recent studies have pinned that part of the cause for schizophrenia to be a combination of an increased spontaneous dopamine release and decreased dopamine release for relevant stimuli (34). Studies using amphetamines have been important for proving this. At moderate doses, amphetamines act as a reward predicting cue by increasing the levels of striatal dopamine appropriately while at larger doses, the amphetamines blunt adaptive responses, which alters the behavioural response and increases the spontaneous transients (spikes in the levels of dopamine) (35). These spontaneous transients may explain the inappropriate phasic firing of dopamine neurons known to be part of schizophrenia (35). All psychostimulants including amphetamines have the effect of increasing spontaneous transients in the striatum which correlate and could explain some of the positive symptoms of schizophrenia. Some of the primary negative symptoms of schizophrenia could be explained by the decreased adaptive transients in the striatum. (35) It is thought that many

antipsychotic drugs perform in the same manner and affect the adaptive and spontaneous transients similarly, where one cannot be fixed without aggravating the other (34).

Around 30% of patients with schizophrenia do not respond to antipsychotics with high D2 occupancy and do not respond to treatments which diminish the levels of presynaptic dopamine concentrations. (32,36) Demjaha et.al found that people who responded to typical antipsychotic treatment had higher dopamine synthesis capacity and that increased synaptic dopamine may be used to predict treatment responsiveness. (36). Treatment-resistant patients did not have this capacity, this demonstrated that there may be a subtype of schizophrenia which is non-dopaminergic. Accumulating evidence has shown that schizophrenias core pathophysiology may also involve dysfunction in glutamatergic, serotonergic and gamma-aminobutyric acid (GABA). (37)

1.2.4 Glutamate hypothesis

The dopamine hypothesis can account for a portion of the psychopathology of schizophrenia, in particular positive symptoms (38). Atypical antipsychotic drugs apart from Clozapine have little to no effect on negative and cognitive symptoms (11). Negative and cognitive symptoms are neglected by antipsychotics and persist causing chronic disability (4). In patients with chronic schizophrenia cortical atrophy correlates with the negative and cognitive symptoms but not with the severity of the psychosis, (39) showing that although some of the cognitive and negative symptoms may be caused by dysregulation in dopamine pathways, not all are.

Glutamatergic pathways are primarily the excitatory neurotransmitters in the brain and glutamatergic neurons utilise between 60-80% of the total brain metabolic activity (32). Glutamate pathways have been linked to the limbic system, cortex, thalamus and are mediated by N-methyl-D-aspartate receptors (NMDARs) (37). Glutamate was originally associated with schizophrenia because it was observed that there were decreased levels of glutamate in cerebrospinal fluid (CSF) of patients with schizophrenia (32). There is now mounting evidence that glutamatergic dysregulation in the prefrontal cortex causes dopamine hyperactivity in the ventral tegmental area (VTA) which causes auditory hallucinations and paranoid delusions (40). Studies using NMDAR antagonists (Ketamine and phencyclidine (PCP), dizocilpine (MK-801)) on HCs induce schizophrenia-like symptoms (negative and cognitive) and increased prefrontal glutamine levels, these can last up to two weeks (41). PCP and Ketamine are non-competitive antagonists that bind at the NMDA subtype of glutamate receptor (39,42). From observing the effects of the NMDAR antagonists on healthy individuals, it has been proposed that certain symptoms of schizophrenia may result from the hypofunction of NMDAR (43). It has also been observed that patients with schizophrenia undergoing long term treatment have increased levels of glutamine in the anterior dorsal cingulate cortex which was linked with the severity of psychotic symptoms (41), suggesting that despite the treatment with antipsychotic treatments there is a basal increase of presynaptic glutamate which is consistent with the NMDAR hypofunction pathophysiological model of schizophrenia (41). It has been seen in patients with schizophrenia that increased synaptic release of glutamine is associated

with psychosis, while glutamate metabolism is related to cognitive impairments (41). In one metanalysis it was observed that glutamate in the frontal region was lower but glutamine is higher in people with schizophrenia when compared with controls, over time the levels of both reduce which could suggest a progressive load of synaptic activity (44). Patients with schizophrenia who don't respond to typical antipsychotic treatment seem to have more marked glutamatergic abnormalities while treatment responders have dopaminergic abnormalities (36).

The role of glutamate in the pathophysiology of schizophrenia has been investigated in Genome-wide association studies (GWAS) they have highlighted several genes associated with glutamatergic neurotransmission or with downstream mediators (*GRM3*, *GRIN2A*, and *GRIA1*) (32).

1.2.5 Gamma-aminobutyric Acid (GABA)

GABA is a major inhibitory neurotransmitter located in the CNS (37). Results from animal models and postmortem studies suggest that part of schizophrenias pathophysiology is caused by both dysfunctions of GABAergic interneurons and NMDARs(45,46). In human postmortem studies in individuals with schizophrenia, alterations were seen in GABA-related epigenetic, transcript, synaptic, and protein markers especially evident was the subpopulation of GABA neurons which encompass calcium-binding protein parvalbumin (PV) (47). GABA interneurons are an important part of the brains rhythm generating network, they are also important in controlling neural oscillations which are fundamental mechanisms for memory, perception and consciousness (5). The

third layer of the prefrontal cortex houses a microcircuit where GABAergic PV cells and glutamatergic cells synchronise neural oscillations (11). The PFC PV neurons have lower levels of PV and proteins and a GABA synthesizing enzyme GAD67 (45). These structural and molecular alterations are hypothesised to alter GABA neurotransmission and weaken the PFC gamma oscillations in people with schizophrenia (47). GABA antagonists have been shown to effective in improving some of the core symptoms of schizophrenia in clinical studies (48). Benzodiazepine which works on the GABA-A receptor allosteric site is used often with antipsychotic medications to treat schizophrenia (5).

1.2.6 Serotonin Hypothesis (5-hydroxytryptamine, 5-HT)

Although the serotonin hypotheses is one of the oldest in regards to schizophrenias pathogenesis it remains highly topical because of the lack of reproducible results (49). Serotonin has been linked to schizophrenia's pathophysiology since studies looking at the interaction between 5-HT and the hallucinogenic drug lysergic acid diethylamide (LSD) which resulted from antagonism of 5-HT in the CNS (37,50). Psychotic symptoms due to dementia and Parkinson's are successfully treated with 5HT2A antagonists without D2 antagonism which halts excess serotonin being released which stops the downstream release of glutamate which can activate the mesolimbic dopamine pathway (40). Sizeable evidence from multiple methods suggests that a subpopulation of patients with schizophrenia display serotonergic function abnormalities (49). It is believed that 5-HT receptors (5-HT_{1A}, 5-HT_{2A}, 5-HT_{2C}) may underlie cognitive symptoms and motivational disabilities and this shown

by atypical antipsychotics which act on 5-HT_{1A} and 5-HT_{2A} receptors and how they ameliorate negative symptoms and mood disturbances (51). In mutant mice who display decreased willingness to work for a reward, there were D₂ receptors overexpression and up-regulation of 5-HT_{2C} receptors in the striatum (51). The extrapyramidal effects of antipsychotics can be ameliorated by serotonin antagonists, (37) though the pathogenesis of schizophrenia has not been explicitly linked with serotonergic dysfunction, 5HT-3 and 5HT-6 continue to be looked into as potential therapeutic targets.

1.3 Environmental risks for schizophrenia

Many epidemiological studies have investigated the impact of the environment on the development of schizophrenia. Several risk factors have been highlighted, such as being raised in an urban environment, early life adversities, and the use of cannabis (especially compounds with high tetrahydrocannabinol levels) early and frequently which has an impact on the developing social brain (2,3,8). Immigration (first and second-generation) has also been studied as well as an increase in the rate of incidence in individuals with young parents or with relatively old parents. (2,8,52,53). Prenatal exposure to infection, preterm births, social disadvantage, and lack of nutrients in the womb have all been linked to heightening the risk of developing schizophrenia (28,54). Prenatal stress increases the basal secretion of glucocorticoid hormones which can reprogram the hypothalamic-pituitary-adrenal axis (3). In rodent models, after prenatal stress malformations in the DNA methylation in GABAergic neurons were observed and connected to schizophrenia-like symptoms (3). Exposure to prenatal infection has been shown to induce epigenetic modifications which can

cause the downregulation of genes essential for synaptic plasticity, transmission, the working memory, and social cognition (55). Obstetric complications such as bleeding during pregnancy, pre-eclampsia, and traumatic births can also increase the risk of developing schizophrenia (3). Severe famine at the time of conception or early in the pregnancy increased the risk of developing schizophrenia two-fold, while mothers with inadequate weight gain increased the risk of psychosis for their offspring by 9-fold (56). Prenatal immune system activation can affect brain development negatively and can slow or alter the neurodevelopmental trajectories which can cause behavioural and cognitive impairments later in life (3). The brain is especially vulnerable in the first and second trimester of pregnancy during critical brain development so maternal stress such as bereavements, unwanted pregnancies and other serious life events are positively linked with development of schizophrenia and other serious mental disorders (3). Particular childhood and adolescent risk factors are capable of predicting the age of manifestation in patients with and without a relative who has schizophrenia (53). Studies have shown that some patients who develop schizophrenia during adolescence experienced delayed developmental milestones in their first year, had hearing impairments, emotional problems, low IQ in childhood, and interpersonal difficulties (2).

Some of the positive symptoms of schizophrenia have been linked to developmental trauma, cannabis use, living an urbanised area and the minority group position in that area, for these studies cultural bias and selective migration were inspected and found to not impact on these association (56). Living in an area densely populated with the same ethnic group and moving

from an urban area to a rural environment decreases the risk of developing any kind psychosis (2,56). The accumulating evidence that environmental exposures occurring preconception through to adolescence and adulthood play a role in the susceptibility of schizophrenia, as well as ample evidence that exposures to environmental factors *in utero* produce brain anomalies as well as phenotypes similar to schizophrenia (56). Though the associations with environmental factors are robust the observational epidemiology cannot distinguish true causation from association as a result of pleiotropy or reverse causation (8).

The role of environment in schizophrenia has been hypothesised for decades but lack of biological models and methodologic limitations has made it a difficult test to what extent they are involved (52). In recent years genetics has dominated the discourse of schizophrenia's etiology (9,52). Twin studies have shown a discordance rate for monozygotic twins who develop schizophrenia of 40-55% because monozygotic twins have identical genomes this illustrates that the risk of developing schizophrenia is not solely genetic but plays an important role (56). The most plausible explanation for this discordance is exposure to environmental factors which are likely to occur as early as *in utero*, or gene-environment interactions during crucial brain development (2,56).

1.4 MRI Findings

The first MRI (magnetic resonance imaging) study focused on schizophrenia was conducted in 1984 and with advancements in the technology over recent years many more followed. These studies have shown that there is not distinct diagnostic neuropathology for schizophrenia, but any of the subtle changes

which are evident is apparent when a patient first becomes symptomatic (49,57). The lack of evidence of distinct neuropathology for schizophrenia could be explained by schizophrenia's diverse presentation, range and severity of the symptoms and if the patient has been treated with antipsychotic medications before the time of MRI scan (48,57). Studies have found that several of the brain abnormalities which can occur are evident before any symptoms appear, hinting at schizophrenia's neurodevelopmental nature and that these abnormalities may change over time (21,26). Understanding these changes in brain structure could prove most valuable in prognosis, treatment, and intervention. Across studies, reduced volume in the intracranial is seen especially with male patients and because 90% of the intracranial volume is usually reached at the age of five this suggests that there is an early developmental cause (59). Reduced total brain volume has also been observed consistently with a marked reduction in grey matter volume (GMV), while cerebrospinal fluid, third and lateral ventricles, and the left side of the planum temporale have increased volumes and are associated with more severe symptoms (60). The levels of GMV reduction is associated with elevated doses of antipsychotics and duration of illness (27). In antipsychotic naive patients (ANPs) most of the brain abnormalities observed are the same as those found in medicated patients but to a lesser extent, in ANPs GMV and total brain volume, the effect size was up to 30% less (24). Conversely thalamic and caudate nucleus volumes are more prominent in ANPs which strengthens the evidence that typical antipsychotic medication enlarges the volume of the basal ganglia (61). The levels of white matter reductions are similar in medicated patients and ANPs which suggests that the levels of white matter do not considerably change

after onset (60). Some post-mortem studies have also uncovered neuroinflammation in the brains of schizophrenia patients which are unrelated to treatments (9,20).

1.5 Genetics of Schizophrenia

Schizophrenia is heritable meaning having a family member with the disorder heightens the odds of developing it during a lifetime (53,62). Offspring with one parent with schizophrenia have a risk rate of 7% while offspring, where neither parent has schizophrenia, is 0.86% (63). Twin studies have been pivotal in furthering our understanding of the role that genetics plays in schizophrenia's etiology (62,63). A study performed using the Danish Wide Twin Register found that the propandwise concordance for schizophrenia is 7% for dizygotic twins (Fraternal) and 33% in monozygotic twins (Identical) in terms of disease liability (8). Although previous studies have found higher rates of concordance, with monozygotic twins achieving 30-40% and heritability estimates for schizophrenia >80% (6,21). Longitudinal twin studies have shown that children of the unaffected monozygotic twin have a similar risk as to the affected twins' children of developing schizophrenia or a schizophrenia-related disease in their lifetimes (9). Taken together, these findings highlighted a clear genetic susceptibility to schizophrenia.

As a result of multiple technological advances and extensive collaboration, there have been remarkable advances in the genetics of schizophrenia in the past decade (64). Genome-wide association studies (GWAS) have been valuable in uncovering many schizophrenia risk loci, including single nucleotide

polymorphisms (SNPs), copy number variants (CNVs) and insertions or deletions in bases in the genome (INDELs) (48). These GWA studies aim to identify areas of the genome that increase an individual's risk for developing schizophrenia (65). Schizophrenia is a complex polygenic psychotic disorder since it is not caused by genetic variation in one gene with a large effect but rather by a combination of multiple genetic variants that each subtly increases the risk of the disease developing (66,67).

Genome-wide association studies (GWAS) are a powerful tool used for studying the genetic architecture of diseases (68). It is an experimental design used to uncover associations between traits of interest and genetic variants, with the aim of better understanding the underlying biology which could lead to better treatments and prevention strategies (69). GWASs have also been successful in uncovering disease-associated biological processes and assisting in risk prediction. (69). GWAS exploits linkage disequilibrium (LD) to measure an association at one genetic variant as a proxy for other genetic variants, the statistical power of these studies depends on the sample size, the distribution of effect size of the causal genetic variants, their frequency in the general population and the LD between genotyped DNA variants and unknown causal variants (65,69). In 2009 the first robustly associated loci linked with schizophrenia was identified using a sample size of 3000, and in 2014 using a sample size of 35,000 cases the number of genetic variants increased to 128 common variant associations across 108 genetic loci (1). In 2018 a schizophrenia meta-analysis identified 179 independent significant SNPs which mapped to 145 loci (1,66).

The 2018 meta-analysis GWA study for schizophrenia had 11,260 people with schizophrenia (cases) and 25,542 healthy controls (HCs). Potential schizophrenia risk genes were generated by taking proximity of a gene to SNPs into account but also the kind of genetic variant, expression quantitative trait loci (eQTLs), chromosome conformation data and genomic finemapping (70). Despite the large sample size, there was not a huge overlap of results from the previous GWAS, which can be expected from studying a complex, polygenic disorder like schizophrenia. (1,66) This study found that associations converge in specific cell types like pyramidal cells, some interneurons and medium spiny neurons (67). It was estimated that from a third to a half of genetic liability derive from common alleles, and a large portion of rare variant architecture comes from mutation intolerant genes which have also consistently been observed in other neurodevelopmental disorder (1,66). It was also noted that in the case of schizophrenia there was an enrichment of common variants associated with loss of function (LoF) and mutation intolerant genes and that these genes accounted for 30% of SNPs based heritability (66). People with schizophrenia have decreased fecundity and early mortality, but the common risk alleles persist in the population which could be because of 1) balance selection that schizophrenia-related alleles have reproductive advantages so are preserved because of their association with positively selected alleles or 2) the effects of gene-environment interaction on these rare variants (8).

Despite this improved knowledge, the understanding of the underlying biological mechanism has not progressed far enough to develop new treatments or cultivate preventative strategies.

1.6 Aim of the project

Understanding how gene expression and regulation differ between individuals has advanced the understanding of healthy tissues and the origins of diseases and complex traits (71). In order to get a better understanding of the control of gene expression it's important to understand the relationship between genotype and phenotype and RNA sequencing which is more quantitatively accurate than absolute transcript (72).

The aims of this study are to examine gene expression of the schizophrenia-associated genetic loci, identified by Pardiñas et al., in the developing human brain using the BrainSpan Atlas of the ABA repository. Of the 145-independent schizophrenia-associated loci provided by Pardiñas et al. 104 genes were available on ABA's resource, BrainSpan for further investigation. The expression profile of the 104 schizophrenia-associated genes was investigated using 1) K-means to identify underlying patterns in the genetic data and 2) network analysis using weight-gene co-expression network analysis (WGCNA) in the developmental stages available in ABA's Brainspan resource. Both steps were performed across the five developmental stages of the ABA dataset. Gene modules identified using WGCNA, were further characterized by identifying hub genes and performing enrichment analysis to identify schizophrenia-related biological pathways. Identifying relevant biological pathways can further our

understanding of disease etiology and present new targets for novel therapeutics which could provide better outcomes for schizophrenia patients.

Chapter 2 - Methods

2.1 Collation of Schizophrenia-associated Genes

Our knowledge of schizophrenias genetics has vastly improved in the past decade, however, identifying gene targets has proven difficult (1). Pardiñas et al. completed a GWAS meta-analysis in 2018 which had 11,260 cases obtained from a CLOZUK sample on genetic information from people with schizophrenia in the UK and 24,542 controls (66). The meta-analysis used cases from the CLOZUK GWAS and combined them with Psychiatric Genetic Consortium (PGC) datasets from the 2014 GWAS excluding any overlapping samples which brought the total number of cases to 40,657 cases and 64,643 controls, the meta-analysis highlighted 179 independent genome-wide SNPs which were significant, which mapped to 145 loci (66). 93 of the 145 loci had been previously identified by the PGC GWAS in 2014 and had shown an increased association in the 2018 meta-analysis (1,66). Summary statistics were added to the CLOZUK genes so a combined analysis could occur. The PGC data was re-examined with a fixed-effects procedure derived from standard errors and polygenic risk scores were calculated for the whole dataset (66). In this study, schizophrenia-associated genes ($n=104$) identified by the GWAS meta-analysis conducted in 2018 by Pardiñas et al. who identified 145 loci total with $P < 5 \times 10^{-8}$. If the loci did not overlap with a gene, the closest gene within 500kb radius was used. Of the 145 loci, 104 genes were available from ABA's BrainSpan resource (66).

2.2 Allen Brain Atlas

The brain is the most complex system in the human body with approximately 86 billion neurons and around a trillion synapses per cubic centimetre of the cortex (73). Its circuitry, cellular and structural diversity and the regulation of its transcriptome are far from being completely understood. There are one billion people suffering worldwide with brain diseases and disorders but a lack of diseased tissue to study. There is a need for another approach to uncovering their aetiologies (74). The challenge of this type of research is the scarcity of high-quality post-mortem human brains, these brains are normally dissected at brain banks and distributed to various research groups and thus data derived from these analyses have diversified hypotheses which are non-parallel as well as different types of research methods, which has hindered the analysis of brain disorders (73). Although other species model systems have been useful, analyses of the human brain itself is essential to get a true understanding (73) .

ABA is a public resource that gives access to gene expression, connectivity, and neuroanatomical data for mouse, primate, adult humans, and developing brains for humans and mice which integrates MRI, genomic and anatomic information, histology, diffusion tensor imaging and gene expression data derived from ISH and microarray methods (75). The original Allen Human Brain Atlas (AHBA) uses high-quality post-mortem brains from males and females between the ages of 18-68 with no known neuropathological or neuropsychiatric history and maps the genes expression to the stereotaxic space. (76) This valuable tool can help researchers trying to comprehend how spatial variation on the molecular scale

associate with macroscopic neuroimaging phenotypes. (76) While there are other human atlases, only Allen Human Brain Atlas (AHBA) possesses high-resolution coverage of the majority of the brain. (76)

The brain tissue underwent several tests including serology, toxicology and tested the RNA quality to determine if it meets the inclusion factors (74). If the brain tissue samples passed, they were then sent to tissue repositories for initial tissue processing. Here the brain tissue was frozen, after and sent to the Allen Institute where a thorough quality control (QC) tests were performed, the brain tissues that passed this threshold have histological data collected from them and the tissues were subdivided and categorised based on if they contain cortical or subcortical substructures. Additional tissue containing subcortical structures were collected and then placed on membrane slides so laser microdissection (LMD) could occur. Both cortical and subcortical tissue samples were collected for microarray analysis. The microarray analysis quantified the expression levels of thousands of genes at once by measuring the hybridisation of Cy3-labelled RNA (cRNA) to a probe on a microarray chip (Agilent 8 X 60K custom design arrays) (74). The probes were mapped to a specific location of DNA which contains single-stranded nucleic acid profiles which recombine with their complementary targets during hybridisation. The gene expression levels in the tissue samples were quantified by measuring the fluorescence at the sequence-specific locations which correlate to the levels of mRNA.(76)

2.2.1 Brainspan Atlas of the Developing Human Brain

The human brain develops following a complex series of histogenic occurrences that depend on differential gene expression and its complex development is not fully understood (77). During the first 6 months of embryonic life the brains general architecture is formed this is driven by strong genetic influences which are silenced in the third trimester allowing for environmental factors to influence the last trimester (78). Mice and non-human primates' models have been useful in developing some knowledge of the brain but the differences between species is a huge limitation. Firstly because of the difference in size, in addition to this the evolutionary differences which are seen in the superficial layers of the neocortex and conjunction the developmental differences in the evolution of GABAergic interneurons (77). The shortage of human prenatal tissue and the use of different species models which has its restrictions has hindered the development of an anatomically comprehensive atlas of the prenatal human brain which could be used for studying the roots of neurodevelopmental and psychiatric disorders (79). The ABA resource BrainSpan transcriptional atlas of developing human brain is a repository of RNA sequences expression profiles of 16 brain structures from 8 weeks post-conception (prenatal) to 40 years of age (80). The stages are outlined in Table 1. The prenatal stage is made up of four high quality mid gestational brains, two from fifteen to sixteen post-conceptual weeks, and two twenty-one post-conceptual weeks specimens. These tissues had no history of maternal drug or alcohol abuse or potential agents that could disturb their development or relations with HIV 1 or 2 or HepB or HepC (77). The specimens were donated from the birth defects research lab at Washington University and the Advanced

Biosciences resource in California (77). The left hemispheres were coronally, serially cytosectioned onto polyethylene naphthalene (PEN) membrane slides for LMD and histologically stained for detailed structure identification, and three hundred regions per specimen were isolated (77). The right hemisphere of two of the specimens was handled similarly and was used further for In situ hybridisation and Nissl staining for structure identification (81). The sample locations were mapped to MRI coordinates and then to the Montreal Neurological Institute (MNI) coordinate space (81). This data was anatomically delineated to create a digital reference atlas which allows for the visualisation of transcriptome data in its exact coordinates. The atlas resources also include MRI, diffusion-weighted MRI from three brains with the approximate same post-conceptual weeks as well as the white matter reconstruction for three additional brains (77).

Table 1: Age categories from the developmental stages for ABA's resource BrainSpan available for download in R.

Stage	Age category
1	Prenatal
2	Infant (0-2 years)
3	Child (3-11 years)
4	Adolescent (12-19 years)
5	Adult (>19 years)

2.2.2 ABA and its application in research

The scarcity of suitable brain tissue available for research led to scientists to develop Allen Brains Atlas human brain resource. ABAs gene expression data being available as high neuroanatomical data makes it possible to identify intricate gene expression patterns for healthy human brains, these profiles for healthy brains can be used as a baseline to identify genes involved in neurological conditions by using machine learning techniques which could be connected with a neurological condition. This approach was successfully applied by Negi et al. where they applied machine learning methods such as hierarchical clustering and weighted co-expression on ABAs gene expression profiles across brain regions (82). From there they were able to build supervised classification models for Autism and Parkinson's with 84% and 81% accuracies respectively (82). Researchers can solely use ABAs resources alone or can apply external data from GWAS or MRI studies to aid their analysis. McCarthy et al. applied the latter technique when investigating Bipolar disorder (BD), they took 58 genes identified to be involved with BD from a previously published GWAS and looked at their expression pattern across 900 areas (83). They also compiled a meta-analysis of MRI studies looking for structural abnormalities across patients diagnosed with BD (83). They aimed to see if they could link unusual gene expression in the BD genes with the brain structural differences (83). Using ABA's Brainspan human brain transcriptome database Mahfouz et al. hypothesised that understanding the functional relationships between ASD candidate genes during normal development could provide insight into ASD's genetic heterogeneity (80). Over human development, the heterogeneous ASD candidate genes share transcriptional networks related to

protein turnover, mitochondrial function and synapse elimination and formation (80).



Figure 2.2.2.1 Schizophrenia-associated genes identified by Pardiñas et al and their position on chromosome.

2.2.3 Collation of schizophrenia-associated genes from ABA

The ABA data was downloaded into R (version 4.0.0) using the R packages ABAData and ABAEnrichment. (52,74)The complete ABA dataset has 17,245 genes expressed in 16 distinct areas over five developmental stages from prenatal to adulthood (See Table 1 for more detail). The genes found to be significantly related to schizophrenia identified by Pardiñas et al. determined by their p-values (66) were exported by CSV

file into R. The 17,245 genes available in ABAData were filtered into five dataframes for each of the developmental stages for further analysis. The dataframes were shaped into a wide format using the `pivot_wider` function in R where the brain areas are the column names, and the row names were converted to the schizophrenia-associated gene names using `column_to_rownames` function in R and the gene expression for each gene was scaled. The schizophrenia-associated genes are available in Table 37 in the Appendix.

Figure 2.2.2.1 shows the schizophrenia-associated genes identified by Pardinas et al. which were available on ABA's resource and where they lie on the chromosome.

2.3 Machine Learning and Clustering

Regression analysis, feature selection methods, and classification are elements of the term Machine Learning (84). Classification can be subdivided into supervised, semi-supervised, and unsupervised. Supervised classification deals with objects that are labelled beforehand and build a learning algorithm which is then used to predict the classification of unlabelled data. Semi-supervised uses labelled and unlabelled data to train an algorithm (85). Unsupervised classification defines classes without help from previously known labels (84), clustering is a form of unsupervised clustering.

In genetics, large datasets of genes and their expression are given to a clustering algorithm to cluster genes whose expression are similar to each other. These algorithms can be used for prediction, classification, and identification in DNA sequences but can also be taught to distinguish between phenotypes and identify possible biomarkers (85).

2.3.1 Unsupervised Learning and clustering for gene expression data

Unsupervised learning is a machine learning technique that looks for natural structures in data and groups them without classifying them (85). Gene expression data is massively complex. Clustering is as an unsupervised learning approach capable of discovering subgroups within a dataset, each of these subgroups or clusters have similar observations within them. This type of analysis has been a cornerstone for interpreting biological information from large gene datasets (86). Clustering can group genes based on their similar expression across brain areas and discover patterns in the data. Clustering can suggest regulatory relationships between genes and transcription factors and can further genome annotation by using the principle of guilt by association, as well as give a better understanding of how diseases manifest and can progress over time (87).

2.3.2 K-means clustering and NbClust

One of the most fundamental modes of understanding learning is to organise data into sensible groupings (88). K-means is a numerical, unsupervised, iterative, non-deterministic method which is classified as a partitional clustering algorithm (89). The k-means algorithm finds a split so the squared error between the points in a cluster and the empirical mean is minimized in each of the clusters. To perform k-means the number of clusters (K), distance metric and cluster initialisation must be prespecified before the algorithm can be run (88). The goal of k-means is to produce groupings each with a high degree of similarity and low degree of similarity with the other groupings (90). One of the

issues with k-means is deciding the number of clusters (K) which are suitable for a dataset, there are many different indices to determine this but a package in R called NbClust integrates thirty different indices in one package to determine the optimal number of clusters in a dataset (91) .

2.3.3 Kmeans analysis of schizophrenia-associates genes

In this study, unsupervised machine learning techniques were performed on the schizophrenia-associated gene set to identify underlying patterns. The NbClust package in R was used to determine the optimal number of clusters for each of the developmental stages. NbClust uses thirty different methods of determining cluster number and produces a bar chart to visualise which cluster number fits best (91). The optimal cluster number was put into the kmeans function in R for the centres. Twenty-five was selected to be the optimum number for initial configurations (nstart). Each cluster is filled with genes with similar expression patterns. After the K-means analysis was performed each of the clusters was visualised using the `fvis_cluster` from the factoextra package in R (92). A table to show module assignment for each of the genes is available in table 17 in the appendix.

2.4 Co-expression Network Analysis

The information found in gene expression data can be used to link genes with unknown function to biological processes, identify candidate genes for disease, determine transcriptional regulatory systems, and identify novel targets for therapeutics (93). Co-expression network analysis recognises genes that show coordinated expression patterns, and the networks can be shown as gene-gene similarity matrixes in later analyses. Co-expression looks to identify

relationships between pairs of genes by using mutual background information or correlation (93). These pair-wise correlations between them are then rolled out to the other genes in the dataset until a network is formed where multiple modules are fashioned and each node signifies a gene, and the edges represent the presence and strength of the relationship. Functional enrichment analysis can be applied to the modules formed after applying the co-expression network analysis method, these modules can often represent biological processes (93).

2.5 Weighted correlation network analysis (WGCNA)

Genes do not work alone, and each gene can work with between four to eight genes which in turn could be involved in up to ten biological processes (94). Any dysfunction in these pathways can potentially lead to diseases. There are many ways to analyse complex, multi-dimensional genetic data, and one of the most popular methods are correlation networks. This technique is a useful way of discovering the underlying intrinsic organisation of the transcriptome. Constructing gene co-expression networks (GCN) for complex diseases is an important method of identifying genes involved in disease, highlighting highly connected genes within the networks and modules which can lead to novel therapeutics or biomarkers for diagnosis. WGCNA is an unsupervised learning systems biology network analysis method for associating correlation patterns among genes across gene expression microarray samples. The WGCNA package which is available for download in R can construct gene networks, identify modules, and can detect highly connected genes that are representative of the module using hierarchical clustering (95). When WGCNA is performed the algorithm evaluates the expression for each gene, pairing them based on

topological overlap (TO) and then considering the degree of shared neighbours looking for consistent gene expression patterns and placing them into modules (82). Once the modules are defined the module eigengene (ME) which is the first principal component of the module is isolated and centralised. Highly connected nodes which are most like the ME and that are representative of the modules are specified and these are called hub genes (HG). The module membership (MM) calculates the degree of correlation between the genes within a module and the ME (96). Using module significance (MS) methods it can help detect important modules which contain high average node significance (NS) and the gene significance (GS) which is the correlation between a node and a phenotype of interest (95). The WGCNA algorithm can execute network construction, module detection, gene selection, data simulation, visualisation, and calculate topological properties (95). WGCNA has been applied successfully with cancers, mice and yeast genetics, and brain imaging data.

2.5.1 Network analysis using WGCNA

After the initial unsupervised learning analysis looking for underlying patterns, network analysis was performed using WGCNA in R. WGCNA was performed on the developmental stages because it is an effective way to characterise correlation patterns within the schizophrenia gene set, genes that correlate sometimes are related biologically. Networks were constructed using an adjacency matrix which looked at the co-expression similarity between a pair of genes and constructed a hierarchical graph. Pairwise correlations were used to identify modules where genes with similar gene expression are grouped into

modules. To construct the weighted gene network, a soft threshold power analysis was first performed using the `pickSoftThreshold` function within the WGCNA package to calculate the adjacency by using gene co-expression, the power in the `pickSoftThreshold` function was calculated independently for each developmental stage and verbose was set at five (94,95). Once the power is chosen to calculate adjacencies, the adjacency is transformed into Topological Overlap Matrix (TOM) and used to calculate the dissimilarity. A clustering dendrogram is made from the genes using TOM-based dissimilarity and after a minimum module size of 10 was chosen the genes were assigned to modules with genes of similar expression profiles. HG and ME were identified in each of the modules using the `moduleEigengenes` function and the `chooseTopHubInEachModule` function in WGCNA. Each gene in the module is annotated with respect to its distance from the ME this is the MM.

2.5.2 Networks and their applications

Networks are abstract models made up of nodes, vertices, and a set of edges. The nodes are the entities and the edges are the information that connects them (97). There are different types of networks for different situations that can yield different outputs, directed networks are formed when nodes are asymmetrical and mean one can influence the second, but the second cannot influence the first. Undirected networks are when the relationship between the nodes is symmetrical and is most useful for exploratory analysis of genes (97). Understanding the intricate relationship between diseases or disorders and their underlying mechanism is a subject that continues to challenge the areas of medicine and biology. There is clear evidence that there are disease-disease

associations where two or more conditions can have similar or identical underlying mechanisms and understanding one can further the understanding of the other (80). The advancements of high throughput technologies like DNA microarray and next-generation sequencing have given researchers large scale genomic datasets (98). Constructing new biological pathways is generally achieved by using the interactions found from previous studies with gene regulation information for specific diseases or tissues, using system-level biological data is predicted to improve current knowledge of underlying mechanisms and lead to improvements in diagnosis, prognosis and treatments (99).

2.5.3 Cytoscape

Cytoscape is a free software project which combines expression data with biomolecular networks and aids in visualising, querying, and linking the data to functional annotation databases (100). Functional proteomics and genomics techniques allow for measurements of expression profiles and interactions between cells and tissue to be collected which could potentially map cellular processes and their dynamics. From these expression profiles active biological processes can be identified using enriched gene annotation and by combining expression profiles and cellular network interactions changes in biological activity could be explained (101). Cytoscape allows for protein or gene properties to be associated with the nodes and edges by changing their appearances which allows for numerous types of data to be seen in a network context. It also includes a range of environments that can model gene transcription kinetics, biochemical reactions, and metabolic control, which

advance biological research (100). To gain insight into the structure and organisation of a network Cytoscape's NetworkAnalyzer plugin was developed for visualisation and analysis (102). NetworkAnalyzer computes a set of topological parameters, including, the number of nodes, edges, network diameter, radius, density, centralization, heterogeneity, connected components, clustering coefficient and shortest path lengths (103).

2.5.4 Visualisation of modules using data from WGCNA in Cytoscape

Gene relationship data, weight and direction of the edge was saved to csv files and module membership was saved in csv Node files for each module in the developmental stage, this data was generated using WGCNA analysis in R. These csv files were exported into Cytoscape (104) (version 3.7.1) using the `exportNetworkToCytoscape` function so each of the modules for the developmental stages could be visualised. Firstly, the edge data was exported, source and target node columns were selected, and the p-value for SNP inclusion determined by Pardiñas et al. (66) was marked as the source node attribute and weight which was filtered to 0.7 and above was selected as the edge attribute. To the node file produced by Cytoscape, the node table for each of the modules was loaded in for the module membership of each node. Cytoscape's NetworkAnalyser was used on each of the modules, each of the modules was treated as undirected. Using the visualise parameters function in NetworkAnalyzer the size of the Node (which referred to the genes) used the MM, edge width was mapped to the weight of the schizophrenia genes and the node colours were charted to the p-value determined by Pardiñas et al. (66).

2.6 Gene Ontologies (GO)

A gene ontology defines a gene's function and how the functions of other genes are related to each other (105). GO is described with respects to three features: molecular function (the activities performed at a molecular level by the gene products), cellular components (where the gene product performs a function relative to a cellular structure), and, biological processes (the biological programs which are completed by several molecular activities) (105,106). As the knowledge of gene ontology is expanding so too are the databases which house them.

anRichment a package available in R is used to calculate enrichment for modules provided when compared to known reference gene sets such as KEGG, GO, Reactome , etc. (107). By using the function `enrichmentAnalysis` in `anRichment` and providing a module in the `classes` input and a collection reference gene set, GO was used for this analysis the enrichment analysis is carried out (107). GO enrichment using `anRichment` was performed on each of the modules to identify biological processes which are over-represented in each of the modules (107). To run the analysis using `anRichment` the `enrichmentAnalysis` function was run using `GOcollection` which is built using `org.Hs.eg.db` R package (108) and specifying species as human, the threshold was set at `1e-4`, the threshold type was Bonferroni, `getOverlapEntrez` = TRUE, `getOverlapSymbols` = TRUE and ignoring the grey module.

Chapter 3 -Results

3.1 Data Pre-processing

The schizophrenia-associated genes (genes tagged) from the GWAS meta-analysis by Pardiñas et al. (66) found in the supplementary materials section NIHMS958804-supplement-Supplementary_Table.xlsx on the sheet titled “Supplementary Data Table 4: Independent genome-wide significant association signals from the CLOZUK + PGC meta-analysis, clumped and amalgamated into loci” were downloaded into R (66). The dataset_5_stages function available within the ABAData package was loaded into R, each of the five age categories was separated into their own dataframe and then filtered so they only included the schizophrenia-associated genes identified by Pardiñas et al (66). This gave five dataframes for each developmental age category with 316 schizophrenia-associated genes expression over 16 distinct brain regions. Each of the dataframes was placed in a wide format where the gene names were rows, and the brain regions were columns. This is illustrated below in Table 2.

Table 2: Wide-format of the schizophrenia-associated genes data frame. Brain areas available from ABAs Brainspan are the columns names in bold and the schizophrenia-associated genes are the row names. Each cell contains the scaled gene expression for each brain area

	10163	10173	10185	10194	10209	10225
ABCB1	0.645635	0.193074	0.103555	-1.09892	0.05936	-1.24916
ABCB9	-0.62942	0.871415	1.128378	-0.6889	-0.32854	0.351194
ABCD2	1.410626	0.488282	0.541653	0.88314	1.281896	0.40718
ACO2	-0.66424	-0.09813	-0.02023	-0.50082	-0.88345	-1.04544
ACP2	-0.5259	-0.74472	-0.28589	-0.63244	-0.70708	-0.90644
ACTR1A	0.84579	0.475071	0.904972	0.463709	1.186486	0.717958
ACTR5	-0.97613	0.244193	0.215631	-0.26563	-1.25836	-0.95691

3.2 Unsupervised learning using K-means analysis on the schizophrenia-associated genes

Unsupervised learning looks for patterns in data. During this project, the large amount of gene expression data from the ABA was used to associate schizophrenia-associated genes based on the similarity of their expression profiles across brain areas. Unsupervised learning allows users to group the schizophrenia-associated genes into clusters.

3.3 Determining Optimal cluster Number using NbClust

K-means analysis was used to cluster each developmental stage to show the schizophrenia-associated genes cluster. As k-means requires the number of clusters to be prespecified, the NbClust function in R was used to determine the optimal number of clusters in each developmental stage as is illustrated in Figure 3.3.1.

NbClust uses thirty independent (91) ways of determining cluster number and produces a bar plot to represent how many times each cluster came up. The optimal cluster numbers for developmental stages 1-5 are as seen in Figure 3.3.1. The kmeans function available in R was used to run the analysis and the package factoextra was used to visualise the clusters.(109)

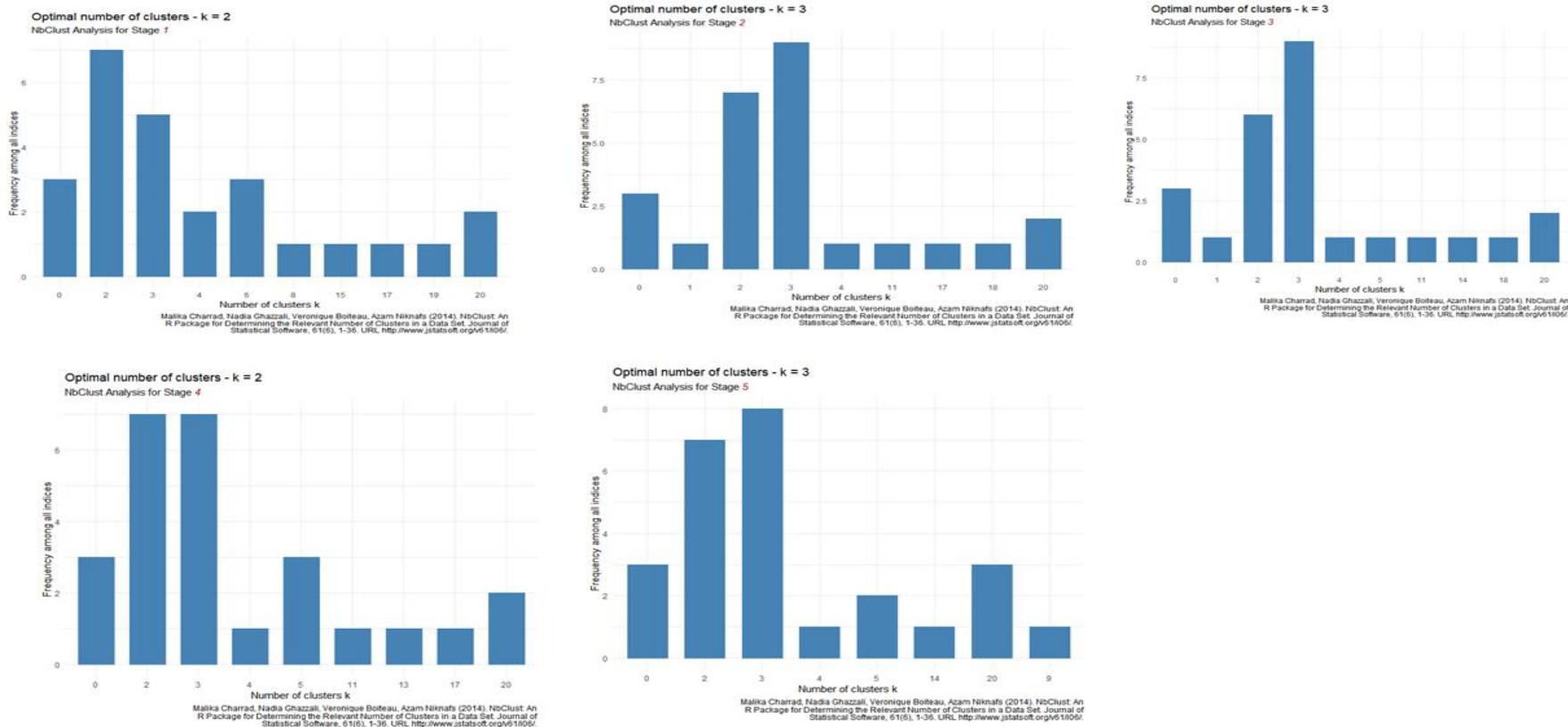


Figure 3.3.1 NbClust analysis performed on the schizophrenia genes identified by Pardiñas et al. to identify the optimal number of clusters (looking between 1-10 clusters) for K-means clustering on each of the five developmental stages available on ABA's BrainSpan resource

From the results in Figure 3.3.1, there is a clear indication of the optimal number for all clusters but Stage Four.

Figure 3.3.2.-3.3.7 visualise the cluster assignments for each of the schizophrenia-associated gene in each developmental stage. Figure 3.3.8 exhibits the clusters of each developmental stage side by side for comparison.

The sum of squares of a cluster measures the total variance within a cluster. A smaller sum of squares means a more compact cluster which means that Cluster Two in this stage houses the schizophrenia-associated genes which are most alike showing internal cohesion in the cluster. A low total variance tells us that the clusters are not that different from each other, in good clustering a total variance would achieve a high percentage where the difference between the groups would explain a majority of the total variance and the within-cluster variance would explain the small fraction left.

Developmental Stage One

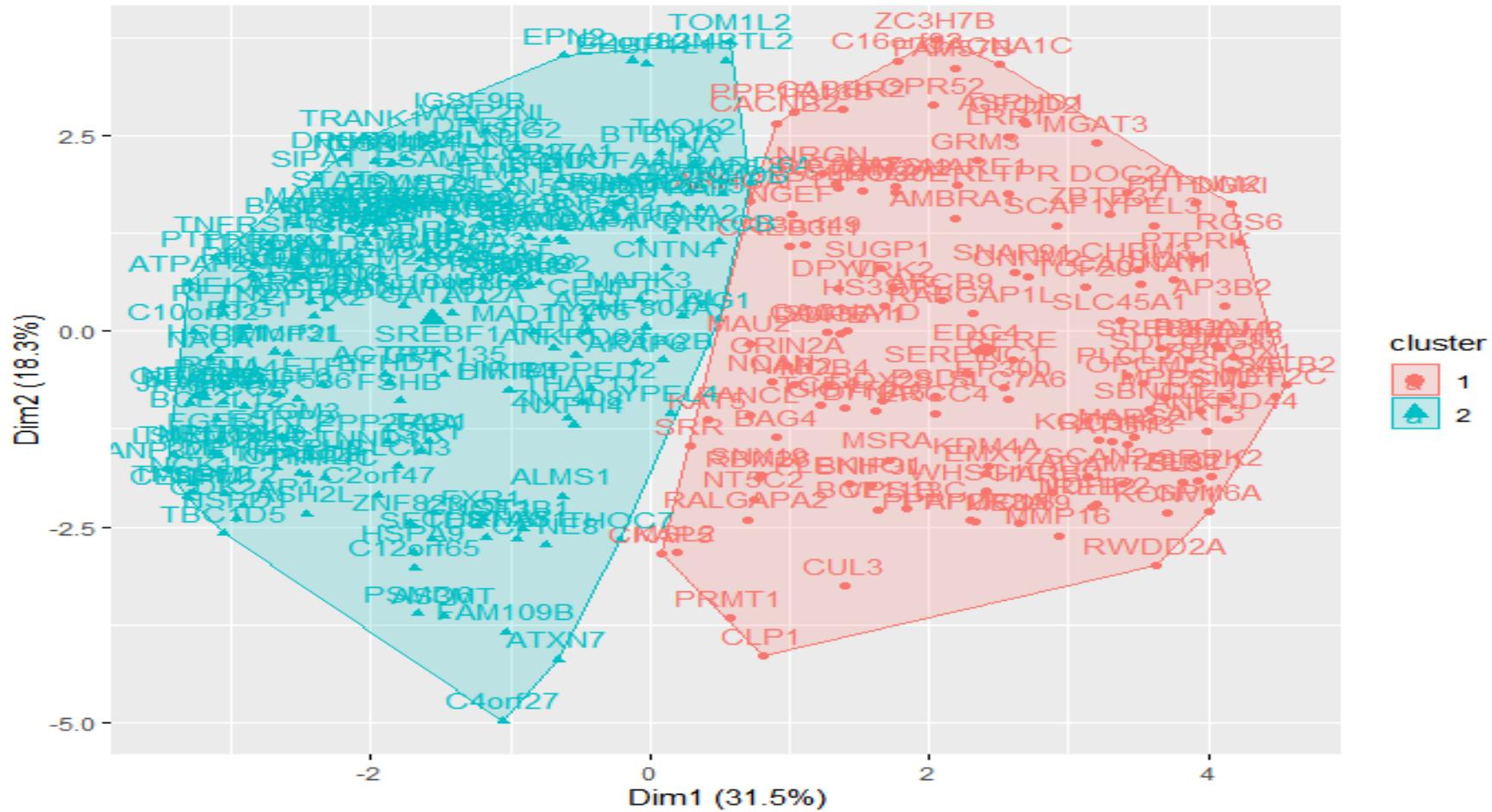


Figure 3.3.2: K-means analysis on the Developmental Stage One genes using the k-means function in R with 2 centres selected as per the NbClust recommendation and nstart= 25

Table 3: Sum of squares for each module in developmental stage One determined using the `k-means` function in R.

Cluster	1	2
Sum of Squares per cluster	1714.970	2115.253

The total variance in the data (Between SS/ Total SS) = 24%

In Figure 3.3.2 it can be seen that cluster One is the most compact cluster, and this is confirmed by it having a smaller sum of squares seen in Table 3. Both cluster One and Two cluster have a very high within cluster variance. A total variance value of 24% shows that there the gene expression data in all of the genes is similar and most of the variance is explained by within cluster variance. A higher total variance is more desirable with only a portion of the variance being explained within the clusters.

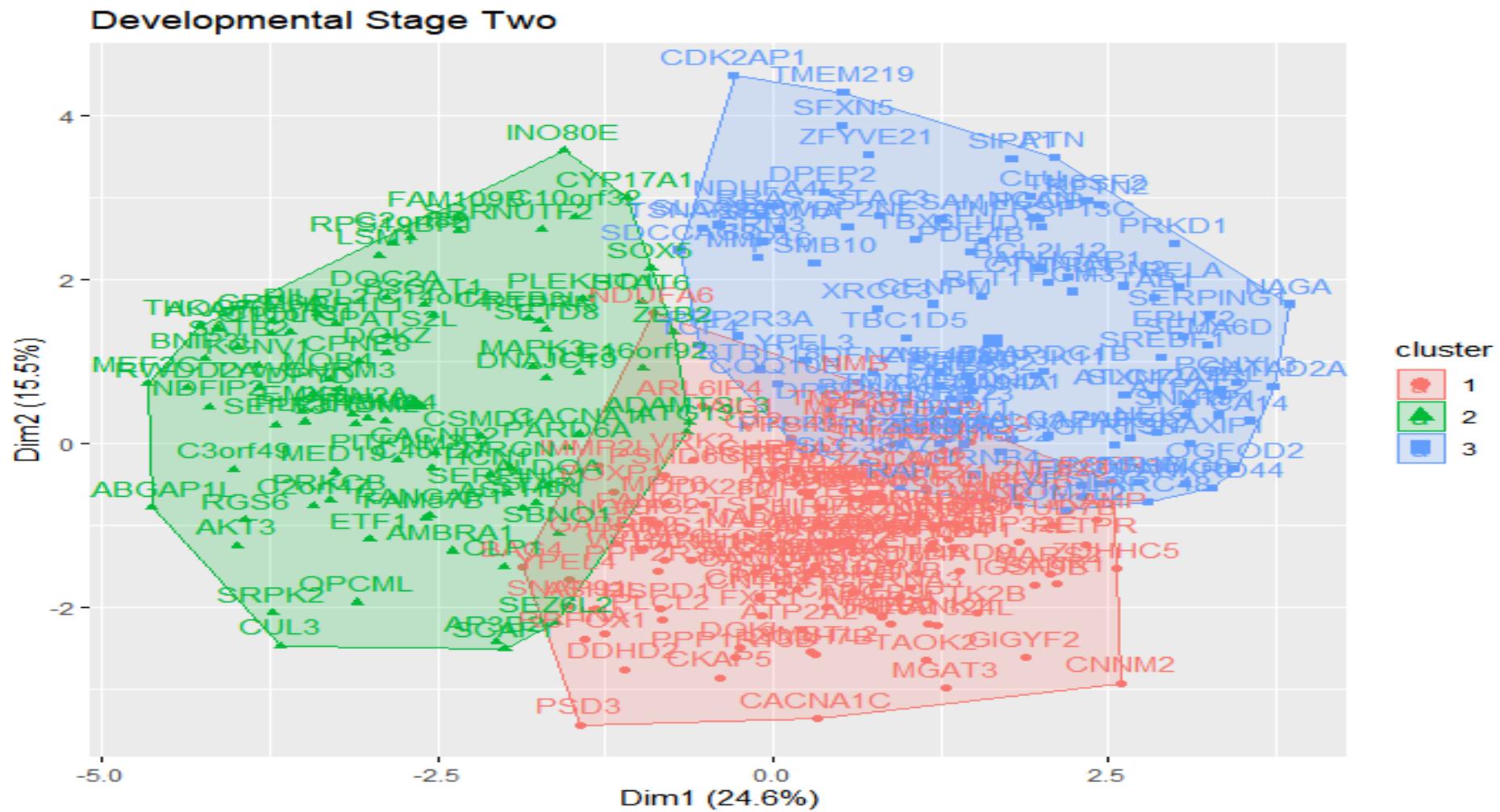


Figure 3.3.3: K-means analysis on the Developmental Stage Two genes using the kmeans function in R with 3 centres selected as per the NbClust recommendation and nstart= 25.

Table 4: Sum of squares for each module in developmental stage Two determined using the kmeans function in R

Cluster	1	2	3
Sum of squares of cluster	1339.484	1128.824	1247.380

Total variance in the data = 26.3%

In the clusters in Figure 3.3.3 each of the sum of squares is high, this tells us that there is a lot of in-cluster variation, the total variance of the data is also low at 26.3%. The high sum of square within the clusters explains a lot of the total variance in the dataset.

Developmental Stage Three

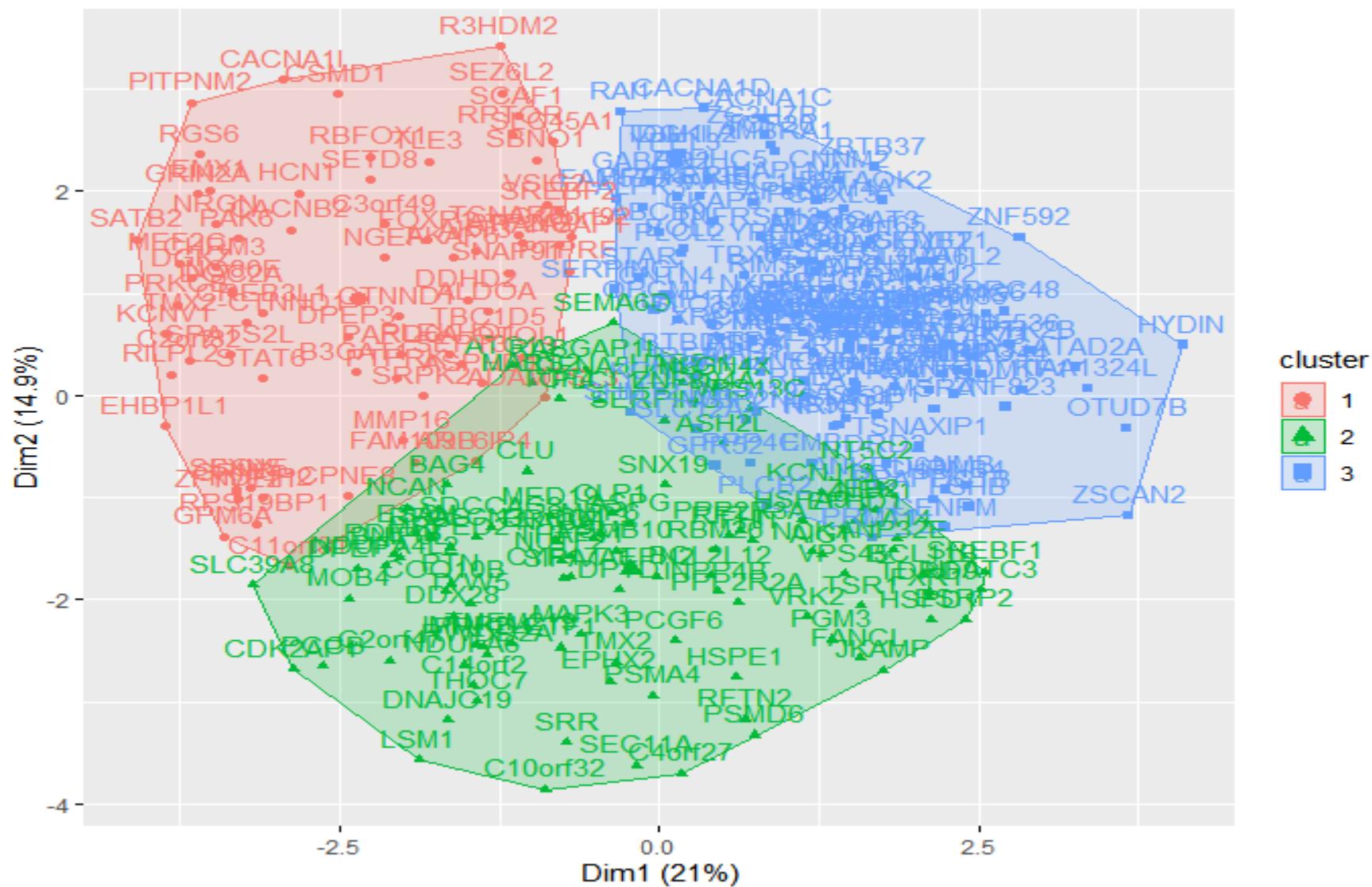


Figure 3.3.4: K-means analysis on the Developmental Stage Three genes using the kmeans function in R with 3 centres selected as per the NbClust recommendation and nstart=25

Table 5: Sum of squares for each module in developmental stage Three determined using the kmeans function in R

Cluster	1	2	3
Sum of squares of cluster	925.8711	1544.4462	1389.3763

The total variance of data in module = 23.4%

The sum of squares of both clusters in Figure 3.3.4 tells us that again there is some variance in the clusters. The low total variance in the data also tells us that the clusters are similar. Ideally, the properties of good clustering would have clusters which are alike, and the other clusters would be very different giving a total variance of the data a percentage closer to 1.

Developmental Stage Four

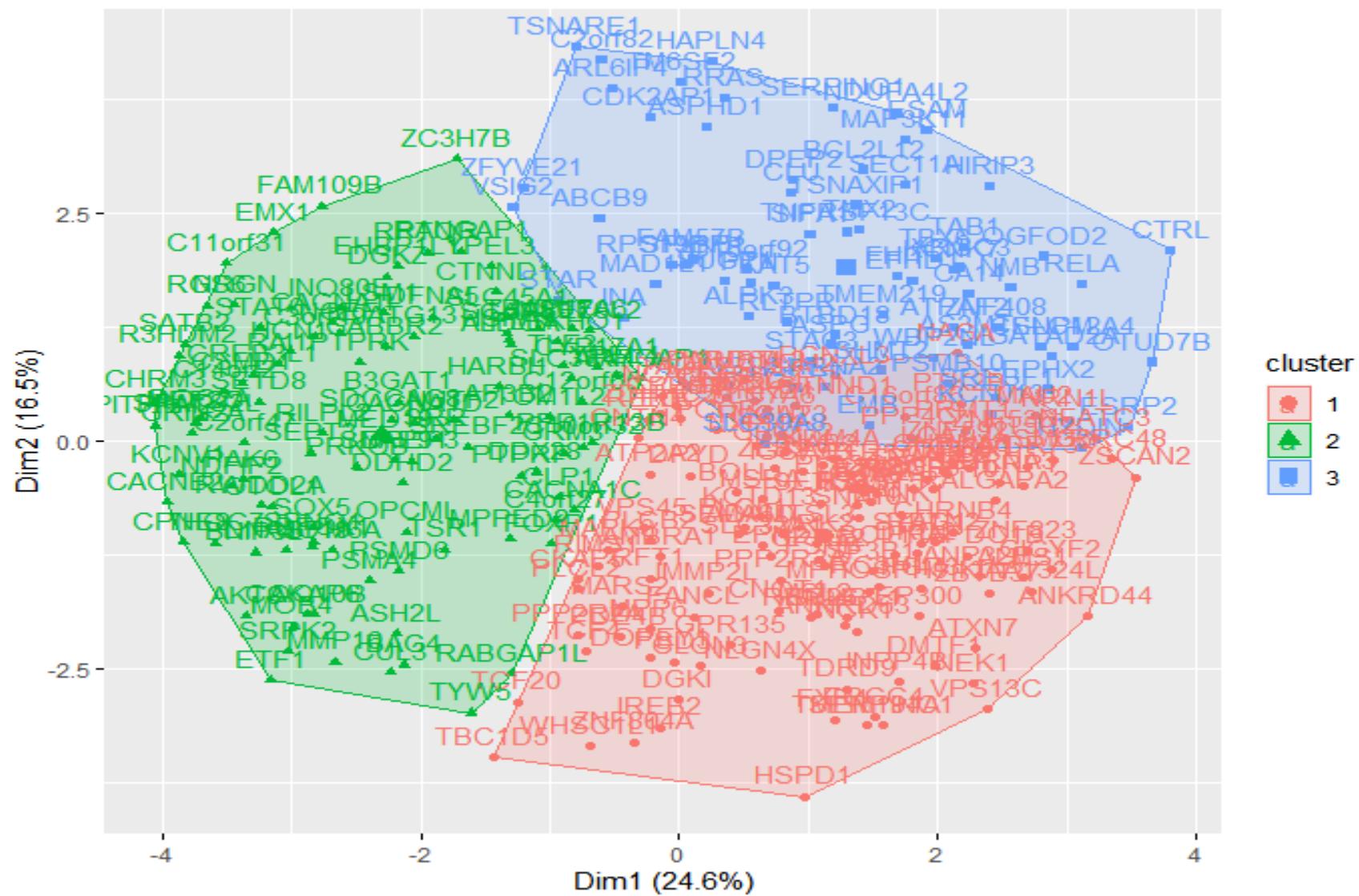


Figure 3.3.5: K-means analysis on the Developmental Stage Four genes using the kmeans function in R with 3 centres selected as per the NbClust recommendation and nstart= 25.

Table 6: Sum of squares for each module in developmental stage Four determined using the kmeans function in R

Cluster	1	2	3
Sum of squares of cluster	1430.5423	1308.7500	993.0729

The total variance of data in module = 25.9%

The in-cluster variation in Figure 3.3.5 is especially high for cluster One whereas cluster Two and Three there have similar in cluster sum of squares. Presumably, a large portion of the total variance in the module would be explained by cluster One.

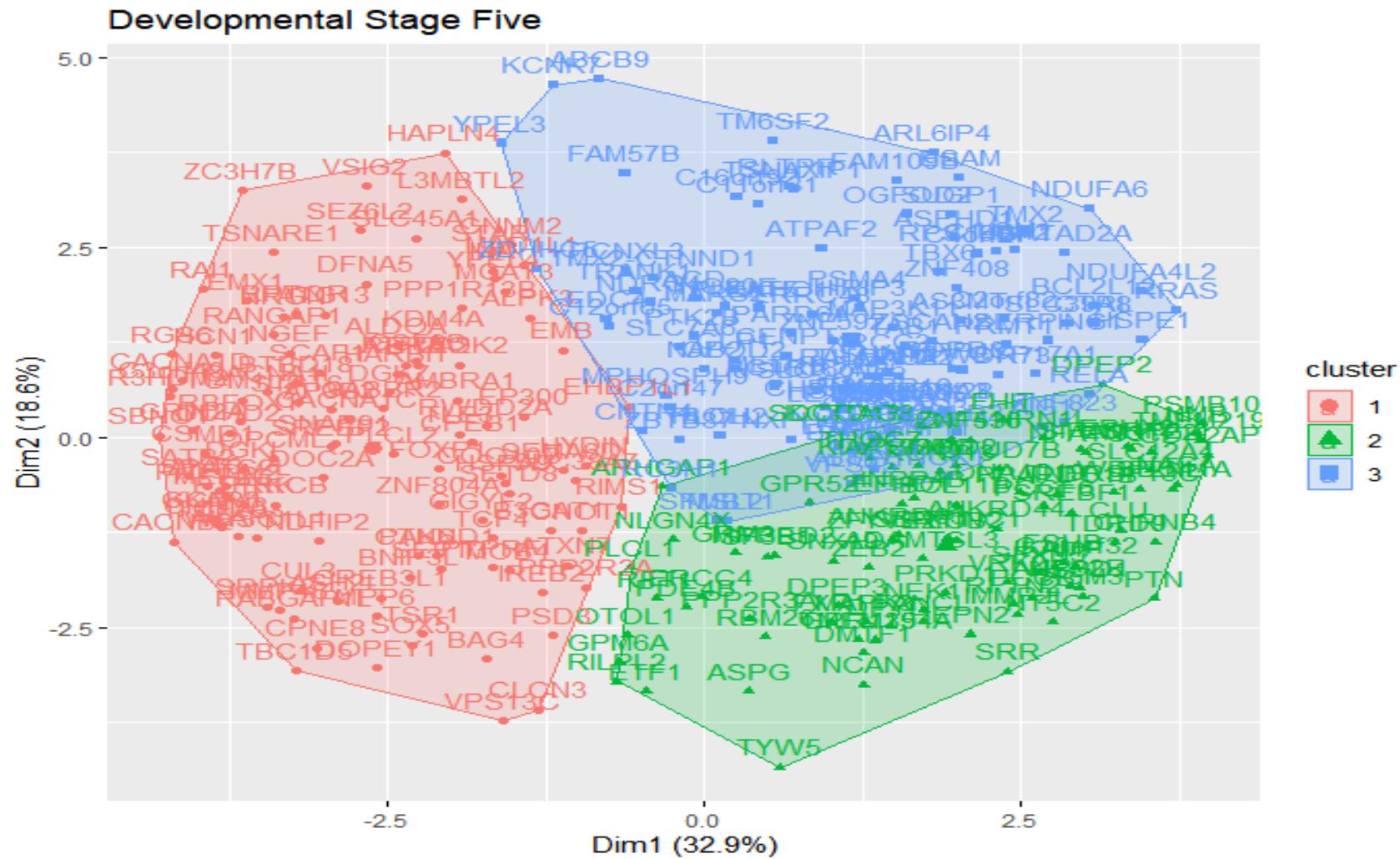


Figure 3.3.6 K-means analysis on the developmental stage Five genes using the kmeans function in R with 3 centres selected as per the NbClust recommendation and nstart=25.

Table 7: Sum of squares for each module in developmental stage Five determined

Cluster	1	2	3
Sum of squares of cluster	1268.6299	952.0207	1120.2446

using the kmeans function in R

The total variance of data in module = 33.7 %

In Figure 3.3.6 all of the clusters have high in cluster variation and a cluster which is not compact, and it can be seen the total variance in the data is low at 33.7% meaning both clusters are similar so a large portion of the 33.7% variation will be explained by the within cluster variation.

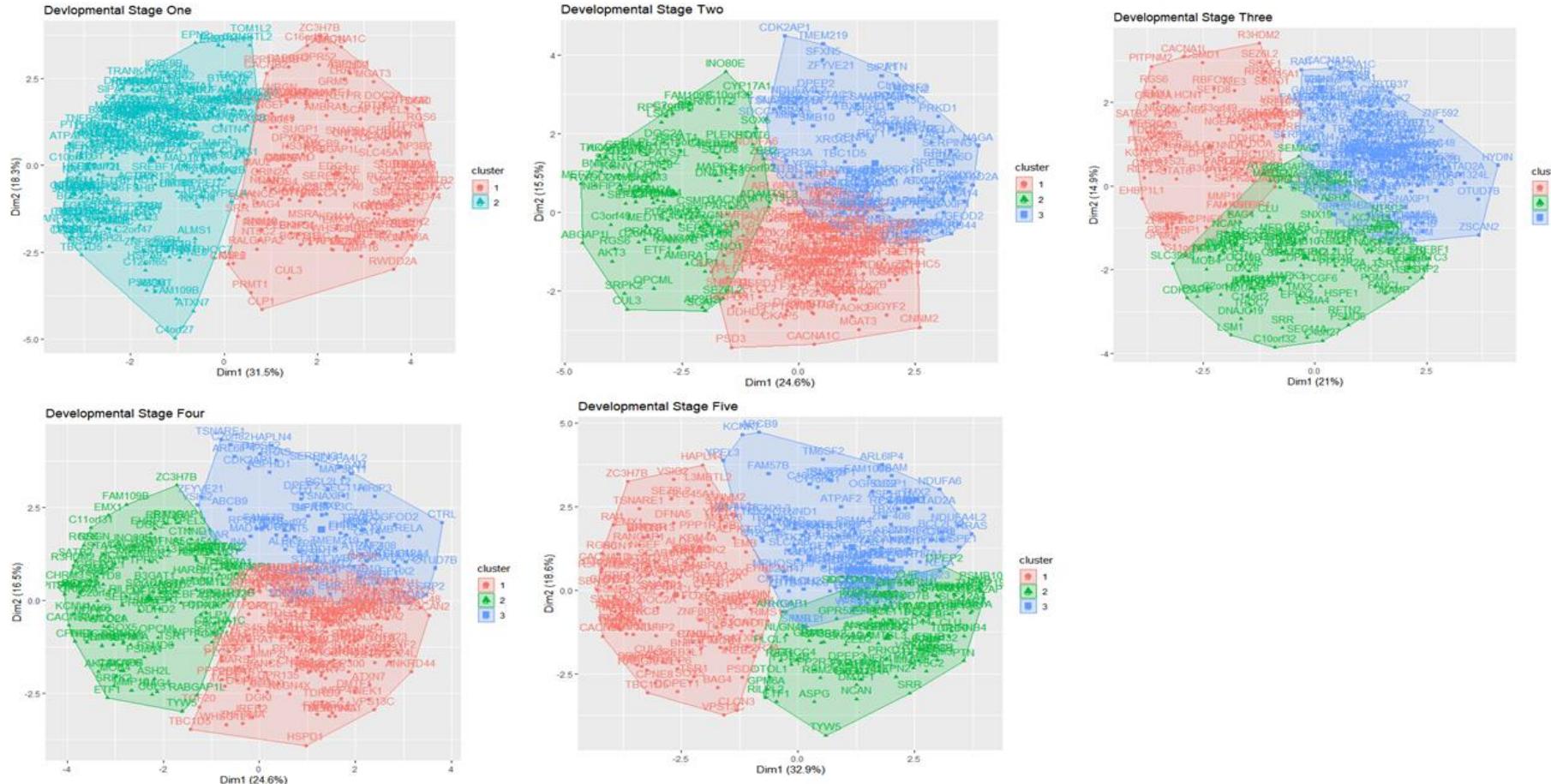


Figure 3.3.7: Kmeans analysis run on schizophrenia genes identified by Pardiñas et al. to determine intrinsic patterns within the genes at each of the developmental stages from ABAs Brainspan resource

3.4 WGCNA on schizophrenia-associated genes and Network Visualisation using Cytoscape

We next employed WGCNA to undertake a system-level approach to identify networks of co-expressed modules of schizophrenia-associated genes. WGCNA is performed to organise highly correlated genes into gene modules.

First, an analysis of the topology of the networks at various soft thresholding powers were performed separately for developmental stages (results are illustrated in Figure 3.4.1). The soft threshold is calculated to identify the power of the gene correlation should be raised to. By raising the correlation to this power, it will reduce the noise of any correlations in the adjacency matrix.

Table 8: The Soft Thresholding power of each developmental stage calculated using WGCNA and shown in Figure 3.4.1

Developmental Stage	Soft Thresholding power
Stage One	7
Stage Two	6
Stage Three	7
Stage Four	9
Stage Five	9

WGCNA is used to organise highly correlated genes into modules. Below a gene co-expression network is constructed which is represented by an adjacency matrix which signifies similar co-expression between a gene pair. Hierarchical clustering is used to identify modules and uses topological overlap to measure dissimilarity. Once the schizophrenia-associated genes are separated into modules these modules were summarised by calculating a module eigengene and defining an intramodular hub

gene. Then the modules were and visualised using Cytoscape's NetworkAnalyzer. Next, GO enrichment was performed on each of the modules using anRICHment an R package (107).

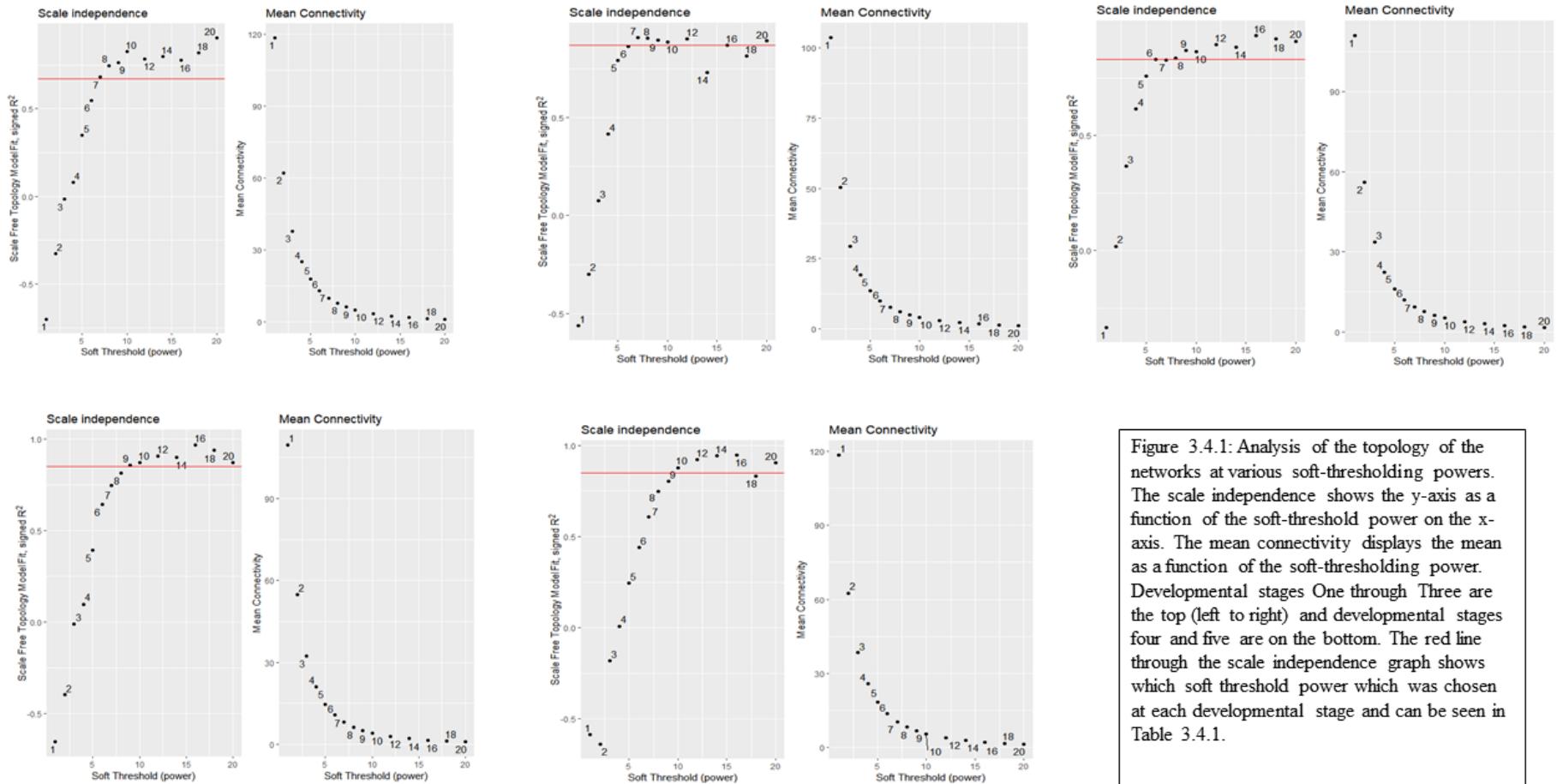


Figure 3.4.1: Analysis of the topology of the networks at various soft-thresholding powers. The scale independence shows the y-axis as a function of the soft-threshold power on the x-axis. The mean connectivity displays the mean as a function of the soft-thresholding power. Developmental stages One through Three are the top (left to right) and developmental stages four and five are on the bottom. The red line through the scale independence graph shows which soft threshold power which was chosen at each developmental stage and can be seen in Table 3.4.1.

After the differential topological matrix is calculated, a gene clustering dendrogram is plotted using the `hclust` function in R. Each leaf of the dendrogram is a gene and after the minimum module is set, the genes may remerge. The dendrogram clusters the branches into coloured modules but some of the modules may need to merge because their genes are highly co-expressed. This is done by calculating eigengenes of each module and re-clustering based on the module eigengene dissimilarity correlations using the `mergeCloseModules` function in WGCNA. Once the modules have merged the module eigengene is re-calculated. WGCNA identified 7 modules in Stage One, 3 modules in Stage Two, 5 modules in Stage Three, 6 modules in Stage Four and 9 modules in Stage Five.

The merged modules and initial modules are illustrated in Figure 3.4.2. After the modules are merged the eigengene is recalculated and in Figure 3.4.3 the adjacency of the eigengene compared to the other eigengenes in the developmental stage is shown.

Network heatmap plot for the developmental stages one through five, using the function `TOMplot` in WGCNA, were created to visualise the topological overlap matrix (illustrated in Figure 3.4.4). This TOM matrix uses the adjacency matrix to build another adjacency matrix which takes topological overlap (the number of shared neighbours of the nodes).

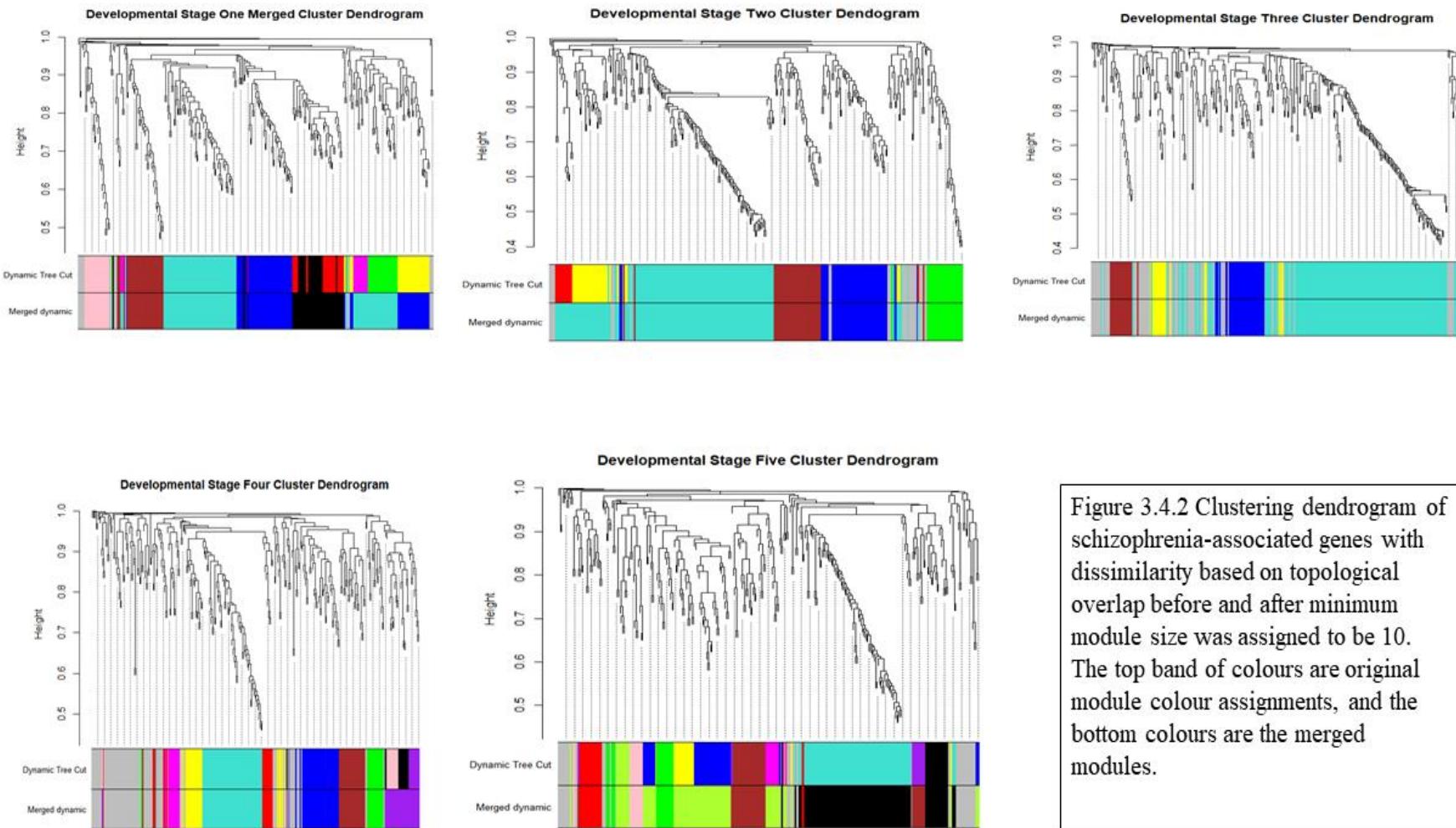


Figure 3.4.2 Clustering dendrogram of schizophrenia-associated genes with dissimilarity based on topological overlap before and after minimum module size was assigned to be 10. The top band of colours are original module colour assignments, and the bottom colours are the merged modules.

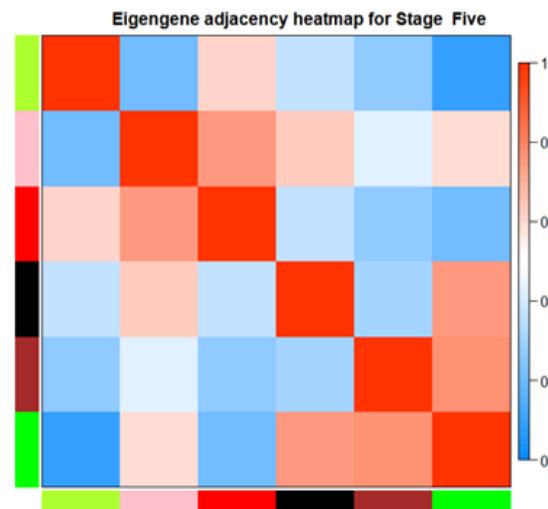
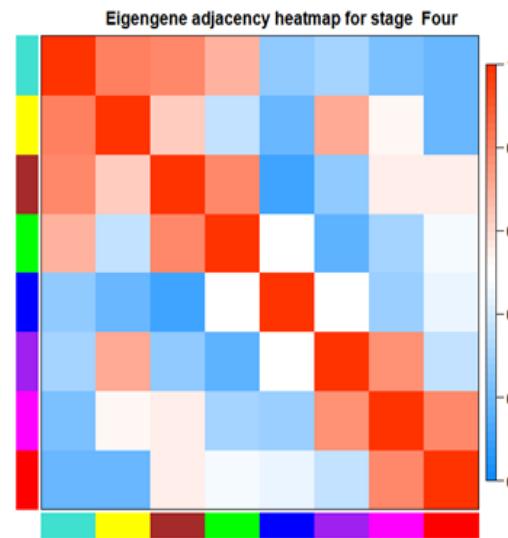
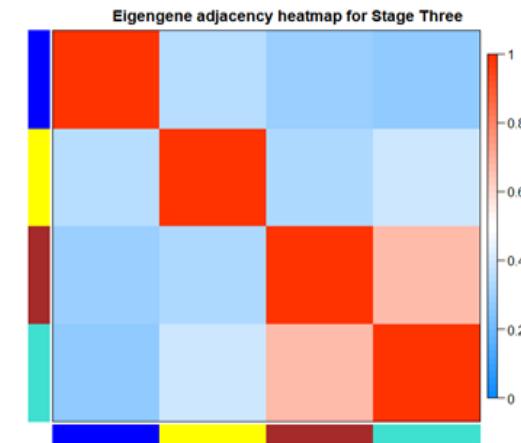
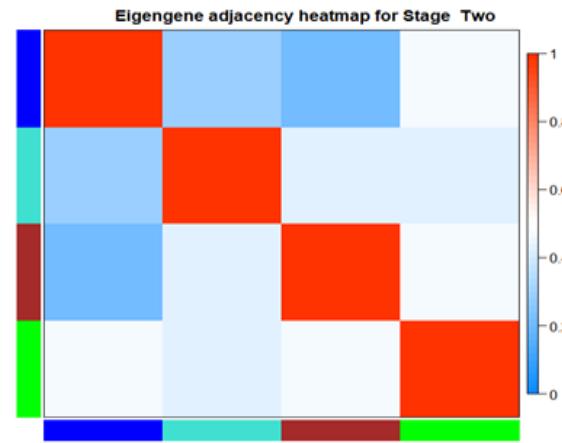
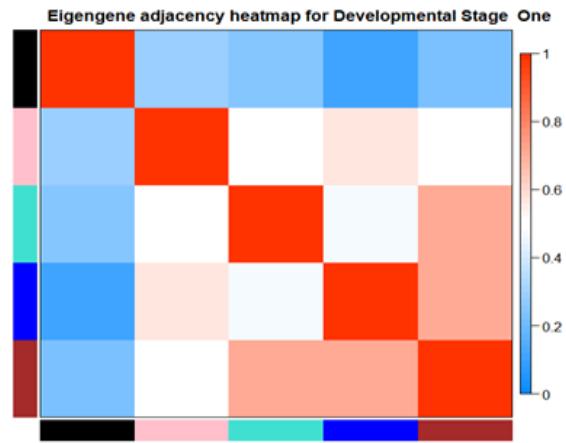


Figure 3.4.3 Eigengene adjacency heatmap for each eigengene network in each developmental stage. Each row and column indicate an eigengene labelled by its module colour. Orange indicates high correlation and blue indicates low correlation.

3.5 Intramodular Hub Genes and Network Analysis

Hub genes (absolute module membership ≥ 0.8) for each module within each stage were identified using the function `chooseTopHubInEachModule` in the WGCNA package. Identified HGs for each module, the function of each gene, as defined by NCBI and their association with disease phenotypes are outlined in Table 9.

Table 9: Gene Functions and the phenotypes they are involved in for each hub genes identified by the WGCNA function in R when performed on the schizophrenia-associated genes identified by Pardiñas et al. for the five developmental stages available on ABA's Brainspan.

Module Colour	Hub Gene	Genomic Location	Function (NCBI gene and Gene ontology)	Association with other Conditions	References (PMID ID)
Stage One					
Black	<i>SOX5</i> -SRY-Box Transcription Factor 5	NC_000012.12	This gene encodes a member of the SOX (SRY-related HMG-box) family of transcription factors involved in the regulation of embryonic development and in the determination of the cell fate. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins. The encoded protein may play a role in chondrogenesis	Lamb-Shaffer Syndrome and Optic Nerve Hypoplasia, Bilateral.	31578471
Blue	<i>FHIT</i> -Fragile Histidine Triad Diadenosine Triphosphatase	NC_000003.12	The protein encoded by this gene is a P1-P3-bis(5'-adenosyl) triphosphate hydrolase involved in purine metabolism. This gene encompasses the common fragile site FRA3B on chromosome 3, where carcinogen-induced damage can lead to translocations and aberrant transcripts. In fact, aberrant transcripts from this gene have been found in about half of all esophageal, stomach, and colon carcinomas. The encoded protein is also a tumor suppressor, as loss of its activity results in replication stress and DNA damage.	Renal Cell Carcinoma, Nonpapillary and Sporadic Breast Cancer.	28404875
Brown	<i>SLC12A4</i> - Solute Carrier Family 12 Member 4	NC_000016.10	The encoded protein controls the movement of potassium and chloride ions across the plasma membrane.	Sickle Cell and Fisheye disease	31792382
Pink	<i>OTOL1</i> -Otolin 1	NC_000003.12	Secreted glycoprotein with a C-terminal complement Cq1-like globular domain that belongs to the C1q/tumor necrosis factor-related protein (CTRP) family. The encoded protein is expressed in the inner ear and forms a multimeric complex called the otoconia, together with cerebellin-1 and otoconin-90, as part of the otoconial membrane. It contains extensive posttranslational modifications including hydroxylated prolines and glycosylated lysines	Benign Paroxysmal Positional Nystagmus and Vestibular Disease	29533337, 31120422

Turquoise					
	<i>TMEM194A</i> - Nuclear Envelope Integral Membrane Protein 1	NC_000012.12	Protein coding gene	Cardiomyopathy, Dilated, 1H	19167377, 31079234, 32923640

Module Colour	Hub Gene	Genomic Location	Function (NCBI gene and Gene ontology)	Association with other Conditions	References (PMID ID)
Stage Two					
Blue	<i>KCNVI</i> - Potassium Voltage-Gated Channel Modifier Subfamily V Member 1	NC_000008.11	Potassium channel subunit that does not form functional channels by itself. Modulates KCNB1 and KCNB2 channel activity by shifting the threshold for inactivation to more negative values and by slowing the rate of inactivation. Can down-regulate the channel activity of KCNB1, KCNB2, KCNC4 and KCND1, possibly by trapping them in intracellular membranes.	Atrial Septal Defect 5 and Familial Adult Myoclonic Epilepsy	
Brown	<i>NDRG4</i> - NDRG Family Member 4	NC_000016.10	This gene is a member of the N-myc downregulated gene family which belongs to the alpha/beta hydrolase superfamily. The protein encoded by this gene is a cytoplasmic protein that is required for cell cycle progression and survival in primary astrocytes and may be involved in the regulation of mitogenic signalling in vascular smooth muscles cells	Infantile Myofibromatosis and Pulmonary Atresia With Ventricular Septal Defect.	31832525, 19711485
Green	<i>GPR52</i> - G Protein-Coupled Receptor 52	NC_000001.11	Members of the G protein-coupled receptor (GPR) family play important roles in signal transduction from the external environment to the inside of the cell		33796846, 24587241
Turquoise	<i>STAG1</i> - Stromal Antigen 1	NC_000003.12	SCC3 family and is expressed in the nucleus. It encodes a component of cohesin, a multisubunit protein complex that provides sister chromatid cohesion along the length of a chromosome from DNA replication through prophase and prometaphase, after which it is dissociated in preparation for segregation during anaphase.	Mental Retardation, Autosomal Dominant 47 and Cornelia De Lange Syndrome.	2467316, 28430577, 28119487, 32778134
Module Colour	Hub Gene	Genomic Location	Function (NCBI gene and Gene ontology)	Association with other Conditions	References (PMID ID)
Stage Three					
Blue	<i>SATB2</i> -SATB homeobox 2	NC_000002.12 (199269500..199471266, complement)	SATB2 encodes for a DNA binding protein that binds specifically at nuclear matrix attachment regions. These regions are involved in chromatin remodelling and transcription regulation.	Glass syndrome (with intellectual disability)	24301056

Brown	<i>INA</i> -Internexin Neuronal Intermediate Filament Protein Alpha	NC_000010.11	Neurofilaments are type IV intermediate filament heteropolymers composed of light, medium, and heavy chains. Neurofilaments comprise the axoskeleton and they functionally maintain the neuronal caliber. They may also play a role in intracellular transport to axons and dendrites. This gene is a member of the intermediate filament family and is involved in the morphogenesis of neurons	Gastroenteropancreatic Neuroendocrine Neoplasm and Medulloepithelioma.	29339073
Turquoise	<i>TRANK1</i> - Tetratricopeptide repeat and ankyrin repeat-containing 1	NC_000003.12 (36826817..369456 62, complement)		Associated with BPD	24309898
Yellow	<i>ANKRD63</i> - Ankyrin repeat domain 63	NC_000015.10 (40278372..402825 86, complement)			24309898

Module Colour	Hub Gene	Genomic Location	Function (NCBI gene and Gene ontology)	Association with other Conditions	References (PMID ID)
Stage Four					
Blue	<i>MEF2C</i> - Myocyte Enhancer Factor 2C	NC_000005.10	Encodes a member of the MADS box transcription enhancer factor 2 (MEF2) family of proteins, which play a role in myogenesis. The encoded protein, MEF2 polypeptide C, has both trans-activating and DNA binding activities. This protein may play a role in maintaining the differentiated state of muscle cells. Mutations and deletions at this locus have been associated with severe cognitive disability, stereotypic movements, epilepsy, and cerebral malformation	Mental Retardation, Autosomal Dominant 20 and Autism Spectrum Disorder.	32418612, 27779093
Brown	<i>SMG6</i> - SMG6 Nonsense Mediated mRNA Decay Factor	NC_000017.11	This gene encodes a component of the telomerase ribonucleoprotein complex responsible for the replication and maintenance of chromosome ends. The encoded protein also plays a role in the nonsense-mediated mRNA decay (NMD) pathway, providing the endonuclease activity near the premature translation termination codon that is needed to initiate NMD	Pancreatic Adenosquamous Carcinoma and Lissencephaly.	25770585
Green	<i>TAOK2</i> -TAO Kinase 2	NC_000016.10	Involved in many different processes, including, cell signalling, microtubule organization and stability, and apoptosis.	Wilson-Turner X-Linked Mental Retardation Syndrome and Syndromic X-Linked Intellectual Disability	29467497
Magenta	<i>OPCML</i> - Opioid Binding Protein/Cell Adhesion Molecule Like	NC_000011.10	Binds opioids in the presence of acidic lipids; probably involved in cell contact.	Ovarian Cancer and Hypogonadotropic Hypogonadism 14 With Or Without Anosmia.	29907679, 33777925, 31577955
Purple	<i>GPR52</i> -G Protein-Coupled Receptor 52	NC_000001.11	Members of the G protein-coupled receptor (GPR) family play important roles in signal transduction from the external environment to the inside of the cell	Psychiatric disorders	33796846, 24587241

Red	<i>WHSC1L1</i> -Nuclear Receptor Binding SET Domain Protein 3	NC_000008.11	Histone methyltransferase. Preferentially dimethylates 'Lys-4' and 'Lys-27' of histone H3 forming H3K2me2 and H3K27me2. H3 'Lys-4' methylation represents a specific tag for epigenetic transcriptional activation, while 'Lys-27' is a mark for transcriptional repression	Wolf-Hirschhorn Syndrome and Nut Midline Carcinoma	31190890, 27285764, 25942451
Turquoise	<i>CA8</i> -Carbonic anhydrase 8	NC_000008.11 (60185412..60281400, complement)	In the carbonic anhydrase family but carbonic anhydrase activity (i.e., the reversible hydration of carbon dioxide) The absence of CA8 gene transcription in the cerebellum of the lurcher mutant in mice with a neurologic defect suggests an important role for this acatalytic form.	Mutations in this gene are associated with cerebellar ataxia, mental retardation, and disequilibrium syndrome 3 (CMARQ3). Polymorphisms in this gene are associated with osteoporosis, and overexpression of this gene in osteosarcoma cells suggests an oncogenic role.	19461874
Yellow	ALMS1-ALMS1 Centrosome And Basal Body Associated Protein	NC_000002.12	Involved in PCM1-dependent intracellular transport. Required, directly or indirectly, for the localization of NCAPD2 to the proximal ends of centrioles. Required for proper formation and/or maintenance of primary cilia (PC), microtubule-based structures that protrude from the surface of epithelial cells.	Alstrom Syndrome and Premature Ovarian Failure 1.	30421101, 32808654

Developmental Stage	Hub Gene	Genomic Location	Function (NCBI gene and Gene ontology)	Association with other Conditions	References (PMID ID)
Stage Five					
Black	<i>C16orf86</i> -Chromosome 16 Open Reading Frame 86	NC_000016.10	Protein Coding Gene		33639916
Brown	<i>RFTN2</i> -Raftlin Family Member 2	NC_000002.12	Upon bacterial lipopolysaccharide stimulation, mediates clathrin-dependent internalization of TLR4 in dendritic cells, resulting in activation of TICAM1-mediated signalling and subsequent IFNB1 production. May regulate B-cell antigen receptor-mediated signalling.	Glass Syndrome.	
Green	<i>NFATC3</i> -Nuclear Factor Of Activated T Cells 3	NC_000016.10	Acts as a regulator of transcriptional activation. Plays a role in the inducible expression of cytokine genes in T-cells, especially in the induction of the IL-2 (PubMed:18815128). Along with NFATC4, involved in embryonic heart development	Crouzon Syndrome with Acanthosis Nigricans and Leukostasis	31249342,33 520407
Greenyellow	<i>SATB2</i> -SATB homeobox 2	NC_000002.12 (199269500..199471266, complement)	SATB2 encodes for a DNA binding protein that binds specifically at nuclear matrix attachment regions. These regions are involved in chromatin remodelling and transcription regulation.	Glass syndrome (with intellectual disability)	24301056
Pink	<i>CHRNA5</i> -Cholinergic Receptor Nicotinic Alpha 5 Subunit	NC_000015.10		Smoking As A Quantitative Trait Locus 3and Tobacco Addiction	33752734,30 366711, 32817066, 33511332
Red	<i>INA</i> -Internexin	NC_000010.11	Neurofilaments are type IV intermediate filament heteropolymers composed of light, medium, and heavy	Gastroenteropancreatic	29339073

	Neuronal Intermediate Filament Protein Alpha		chains. Neurofilaments comprise the axoskeleton and they functionally maintain the neuronal caliber. They may also play a role in intracellular transport to axons and dendrites. This gene is a member of the intermediate filament family and is involved in the morphogenesis of neurons	Neuroendocrine Neoplasm and Medulloepithelioma.	
--	---	--	---	---	--

3.6 Cytoscape and Network Visualisation

`NetworkAnalyzer`, a Cytoscape plugin, aids in visualisation but can also calculate a network's topological properties (103). Using data loaded into Cytoscape including source nodes, target nodes, the p-value for SNP inclusion identified by Pardiñas et al, weights and node attributes, the network could be visualised using `NetworkAnalyzer`. `NetworkAnalyzer` also calculates the properties of a network as shown in Figure 3.6.1, treating the network as undirected. The clustering coefficient is a ratio of closed triangles over the total open and closed triangles (110). Connected components measures the number of separated fragments in the overall network. Network diameter is the largest number of edges to transverse the network. Network centralization measures each nodes centrality in the network, a network which is highly centralized (= 1) contains a few nodes which dominate the network and without these nodes the network would become fragmented and leave unconnected sub networks (111). Network density is a measure of how densely a network is filled with edges, where 0 means there are no edged and 1 means the network is highly populated with edges. Network heterogeneity is a measure of the diversity of connection shown by the node degrees where a homogenous network is equal to 0 and 1 is heterogeneous (112). Networks containing biological date are usually very heterogeneous where most nodes have very few edges apart from HGs which are highly connected (113).

3.6.1 Network Visualisation of the schizophrenia-associated genes in Developmental Stage One (*in Utero*)

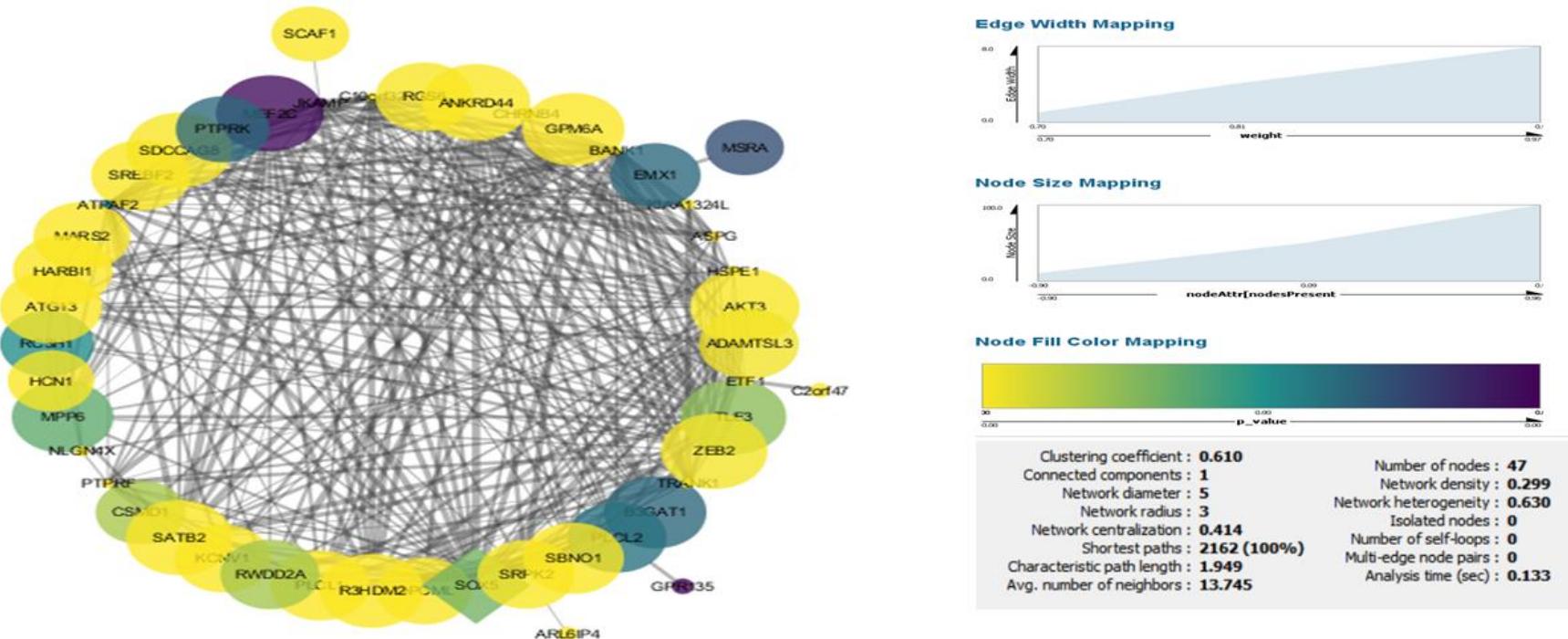


Figure 3.6.1 Black Module for developmental Stage One where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The number of connected components for the Black module in Stage One are One and include 47 genes. The network centralization of the network in Figure 3.6.1 is 0.415, this means it but does not contain essential central nodes. The network heterogeneity is 0.630 which means the network is more heterogenous. Clustering coefficient is the number of edges between neighbours in the module over the maximum number of possible edges between the neighbours, the network clustering coefficient, which is 0.610, this is the average clustering coefficient of all of the nodes this means that 61% of the edges that could exist do. The network density of the module is only 0.299 which means the module is not densely populated with edges and this makes sense because only edges with a weight of 0.8 and above are included. The HG identified by WGCNA and as shown as a diamond shape in Figure 3.6.1 is *SOX5*. SOX protein are made of transcription factors which mediate DNA binding, protein-protein interactions and nuclear trafficking(114). Heterogeneous mutations in *SOX5* are seen in the developmental disorder Lamb Shaffer syndrome which cause developmental delays, motor and language deficits and intellectual disability (114). It has also been shown to be involved in sex determination (115).

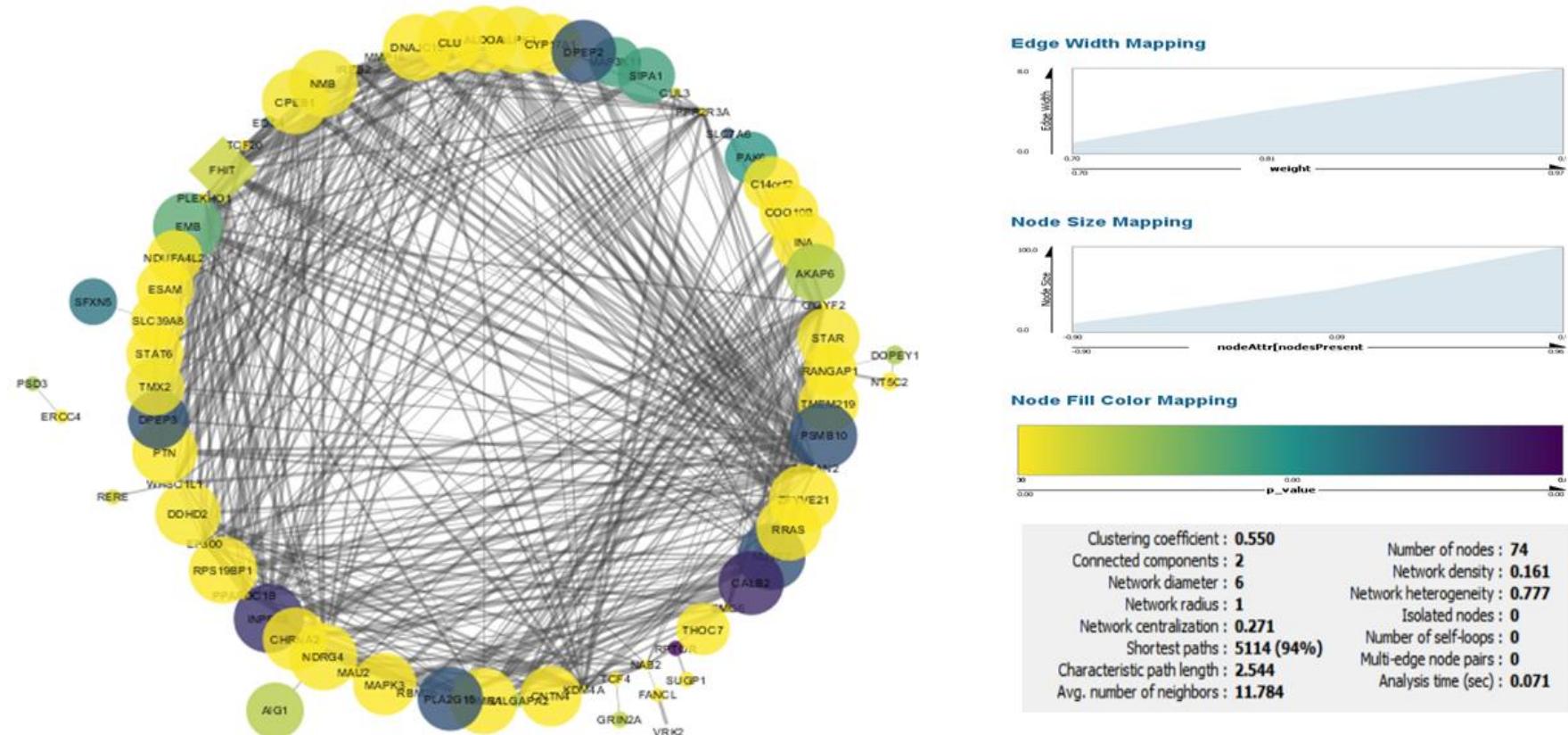


Figure 3.6.2 Blue Module for developmental Stage One where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

Figure 3.6.2 shows the Blue module for Developmental Stage One, as the module has split, they are two connected components. The network centralization is 0.271 meaning that the nodes in the network each have a different number of edges and likely that key players in the network are not central and if they were removed the network would not break into sub networks. The network heterogeneity measures the variance based solely on connectivity, for this module is 0.771 this means if each node has the same number of edges the network heterogeneity is 0 and 1 would be if each of the nodes had a different number of edges. The clustering coefficient looks at how well connected the neighbourhood of a node is. The networks clustering coefficient is 0.550 which means 55% of the potential edges to exist between a node and its neighbours actually exist. The network density is 0.161, the 74 nodes are connected with edges with a high weight seen by the edge width. It is likely because any edge less than 0.8 was excluded that each of the modules will have a lower network density. *PSD3* and *ERCC4* have split from the larger module. The HG in this module as identified by WGCNA and is shown in Figure 3.5.2 as a diamond shape is *FHIT*. Inactivation, deletion, decreased expression of *FHIT* is reported in the majority of human cancers. In mice the restoration of its expression suppresses tumorigenicity by inhibiting apoptosis and inhibiting the proliferation of tumour cells (116).

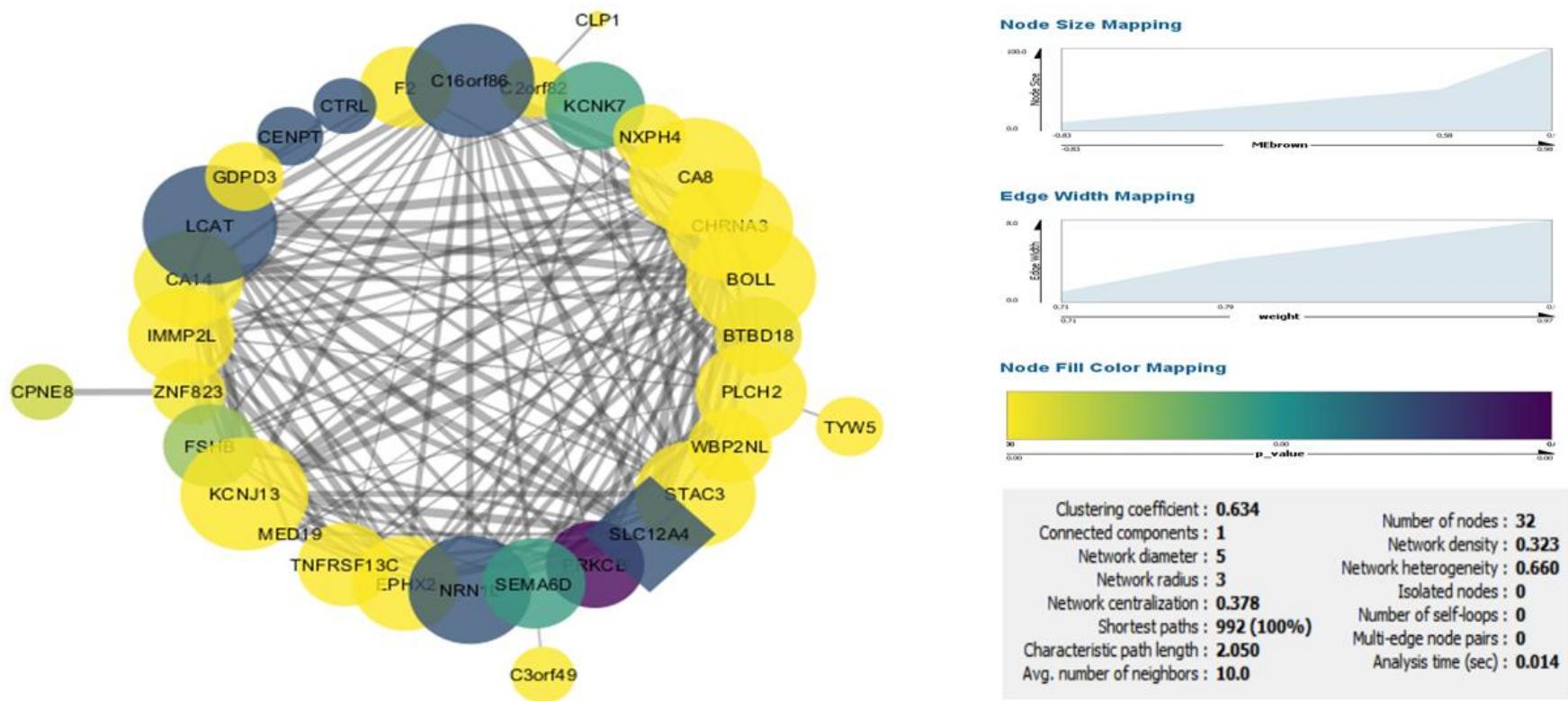


Figure 3.6.3 Brown Module for developmental Stage One where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The schizophrenia-associated genes in the Brown module for stage One are shown in Figure 3.6.3. The connected components for this module are one. The network centralization of 0.378 means there are probably no central key nodes. The network heterogeneity is 0.660 which means that the nodes have variable number of edges. The clustering coefficient of 0.634 measures the degree to which nodes are inclined to cluster together. The network density is likely low at 0.323 because of the weight cut off and the average number of neighbours for the nodes is 10. Most of the schizophrenia-associated genes have a similar module membership shown by them having a similar size except CLP1 and MED19. The HG in this module as calculated by WGCNA and as shown in Figure 3.6.3 is *SLC12A4* also known as *KCC1*. *K⁺ Cl⁻* co-transporter 1 is a membrane protein which facilitates symport of *K⁺* and *Cl⁻* ions through the surface on cells (117). *KCC1* sustains normal erythropoiesis, cancer growth , bone turnover and sickle cell formation (118).

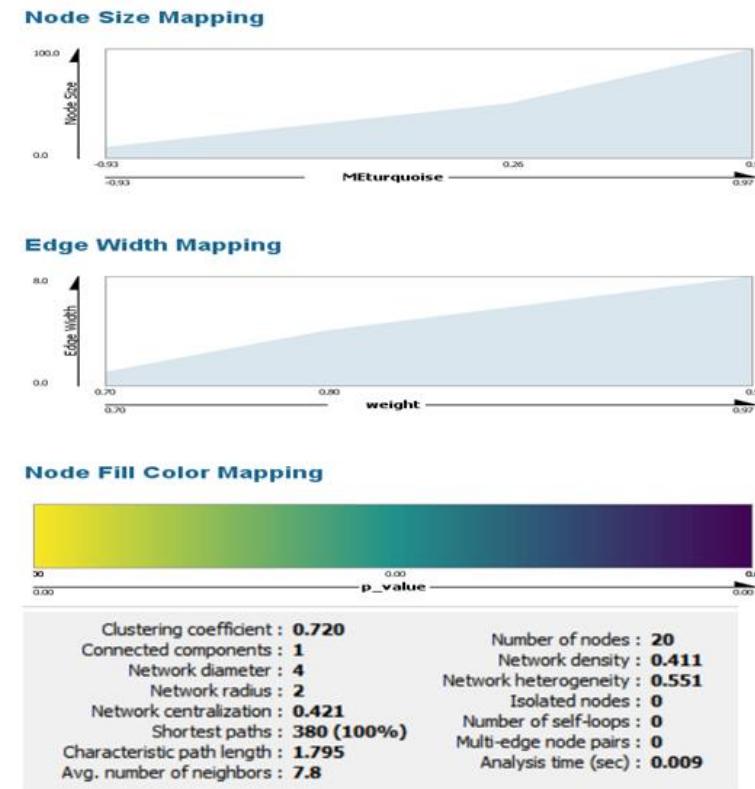
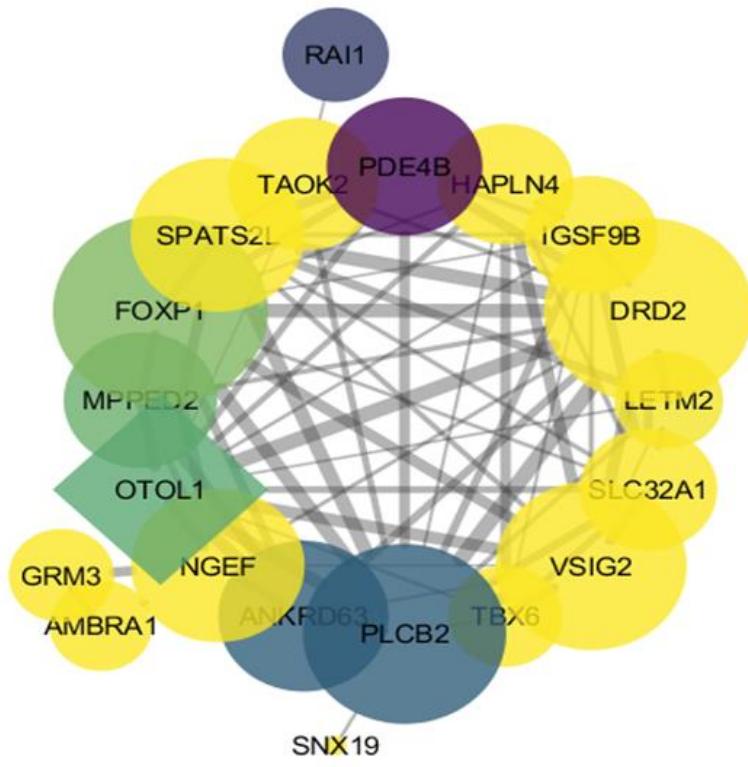


Figure 3.6.4 Pink Module for developmental Stage One where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The connected components for the module shown in Figure 3.6.4 is one. The network centralization which looks at the degree of the nodes and the weight of the edges is 0.421 meaning there are no key central nodes. The network heterogeneity is 0.551 which means the number of edges each node has is different. The clustering coefficient of 0.720 means that 72% of the edges which could exist between neighbours do. The network density is 0.411. The HG of the module is *OTOLI*. Otolin -1 is a secreted glycoprotein who supports vestibular maculae cells, canal cristae and marginal cells by restricting its mRNA expression to the inner ear (119). *OTOLI* is necessary for hearing and vestibular function(120).

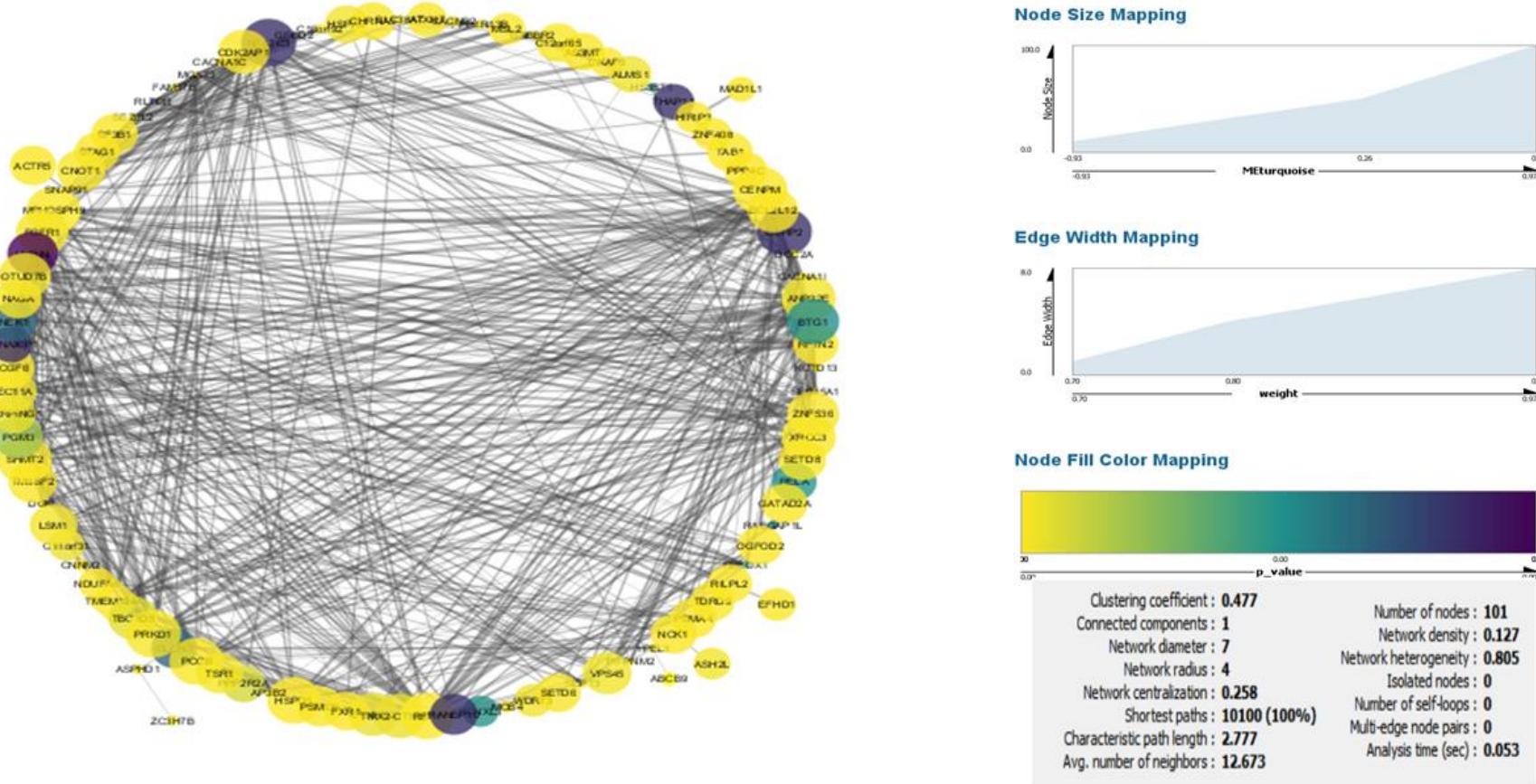


Figure 3.6.5 Turquoise Module for developmental Stage One where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Turquoise module for developmental stage Two seen in Figure 3.6.5 is a large module with 101 genes. The connected component of the module is One as there are no sub structures. The network centralization which looks at the degree of a node and the weight of its edges is 0.258 and the network heterogeneity which 0.805. The clustering coefficient of 0.477 means the genes in the module which means these genes are somewhat likely to cluster together. The network density of 0.127 is low as expected because only edges with a weight above 0.8 are included. The HG for this module is *TMEM194A* or *NEMP1*. Nuclear envelope membrane protein 1 (NEMP1) is highly expressed in various cancer types including breast cancer (121). Cells with high expression of *NEMP1* are resistant to tamoxifen which is one of the main treatments of breast cancer (121). It has also been linked to eye development, early menopause and loss of *NEMP1* homologs in Drosophila , zebrafish and mice leads to early loss of fertility and sterility (122,123).

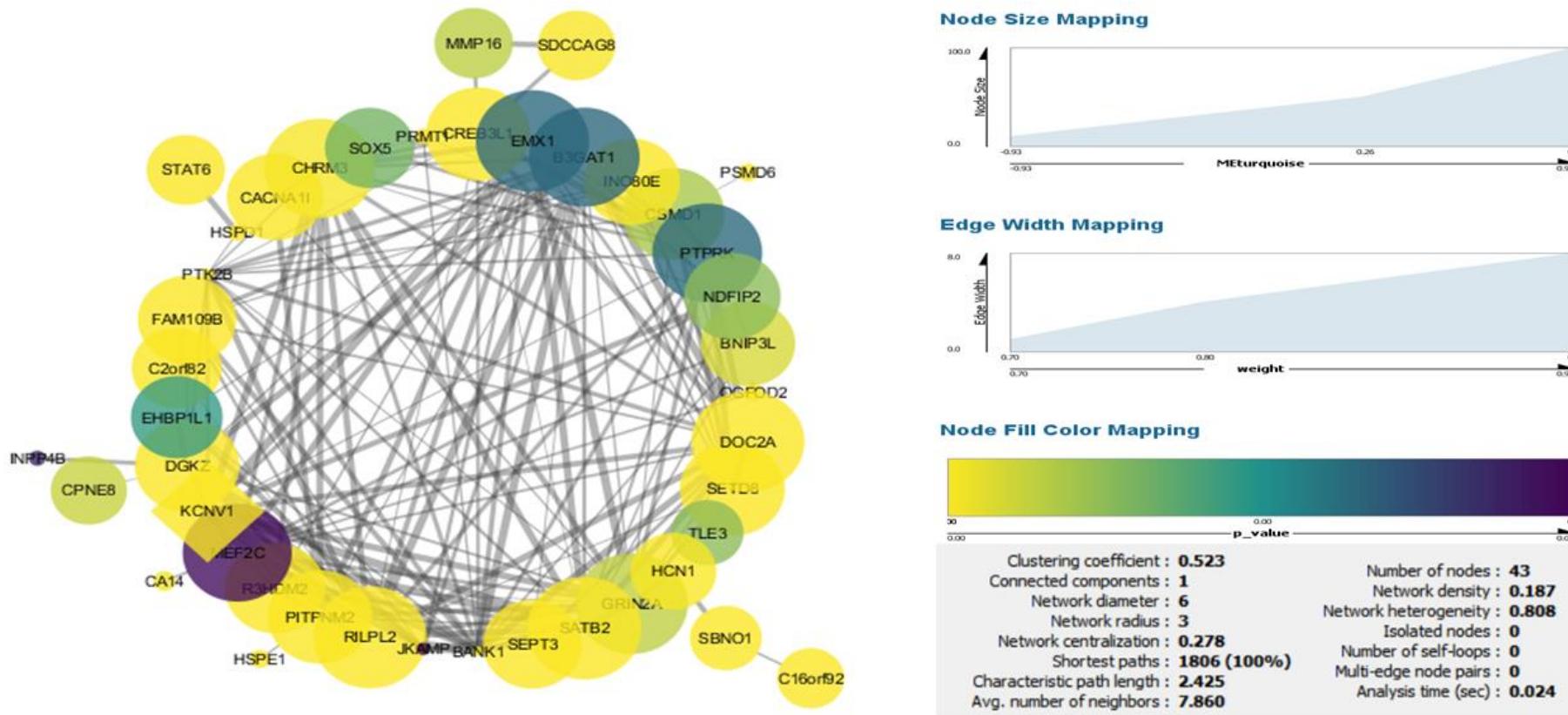


Figure 3.6.6 Blue Module for developmental Stage Two where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

In the Blue module for developmental stage two shown in Figure 3.6.6, the connected components One. The low network centralization of 0.278 meaning it is closer to decentralized network with no key node. The network heterogeneity is 0.808. The network density is low which can probably be explained by only including edges <0.8. The HG *KCNVI*, is yellow which shows it does have a strong link to schizophrenia in the module. *KCNVI*.

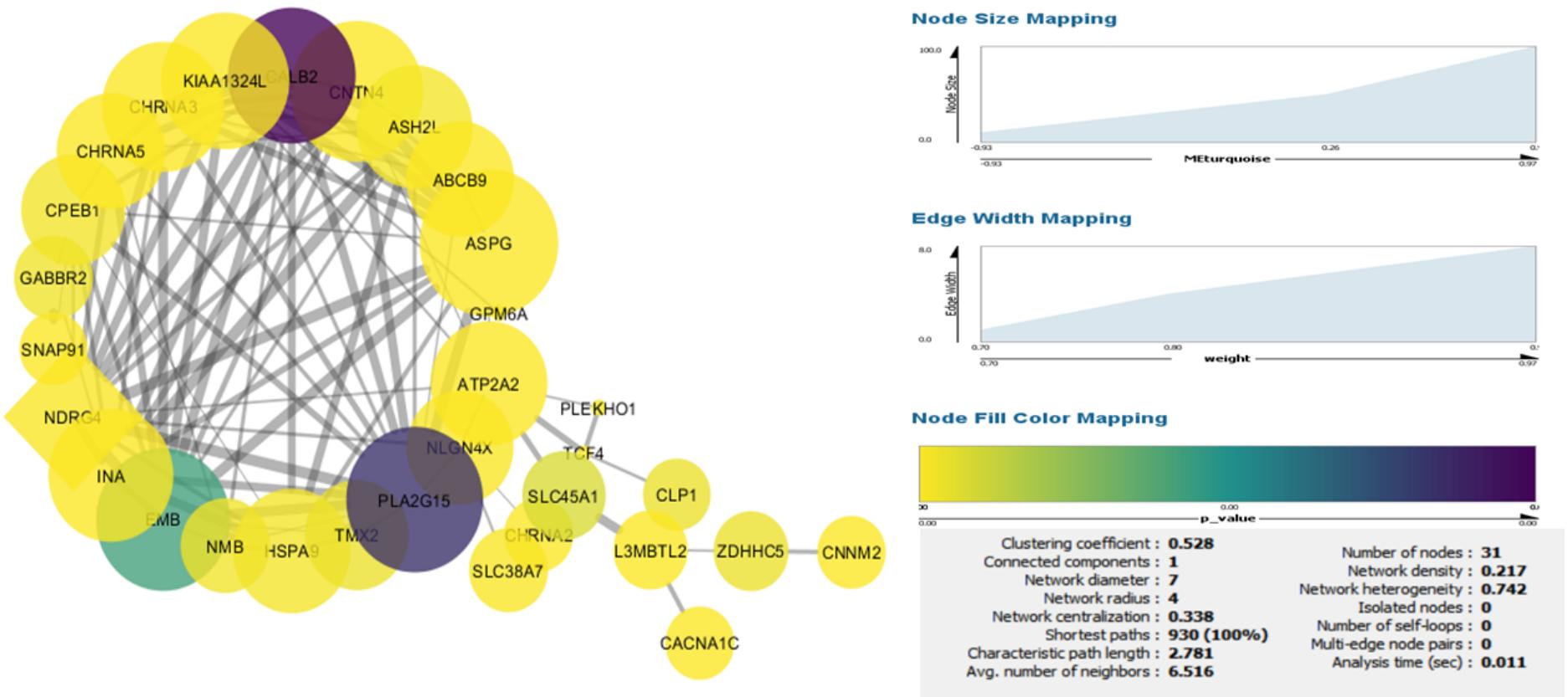


Figure 3.6.7 Brown Module for developmental Stage Two where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Brown module for developmental stage Two is a large module with 31 genes as seen in 3.6.7. The connected components are one as there are no substructures. The network centralization of 0.338 for the network means that the network is more like a decentralized rather a network with no central key players. The network heterogeneity for the module is 0.742 means that the nodes have differing numbers of edges. The clustering coefficient of 0.528 is the average of each of the clustering coefficients per node. This means that a lot of the edges between neighbours are 52.8.% of the edges which could exist. The network looks dense in Figure 3.6.7, but it has only achieved a network density of 0.217. The HG is *NDRG4* is involved with cell differentiation, proliferation , stress and development (124). In rat studies *NDRG4* expression was decreased in brain tissue of ischemia/reperfusion which increased apoptosis in cerebral cells and restoration and increasing *NDRG4* is protective against cerebral ischemia (125).

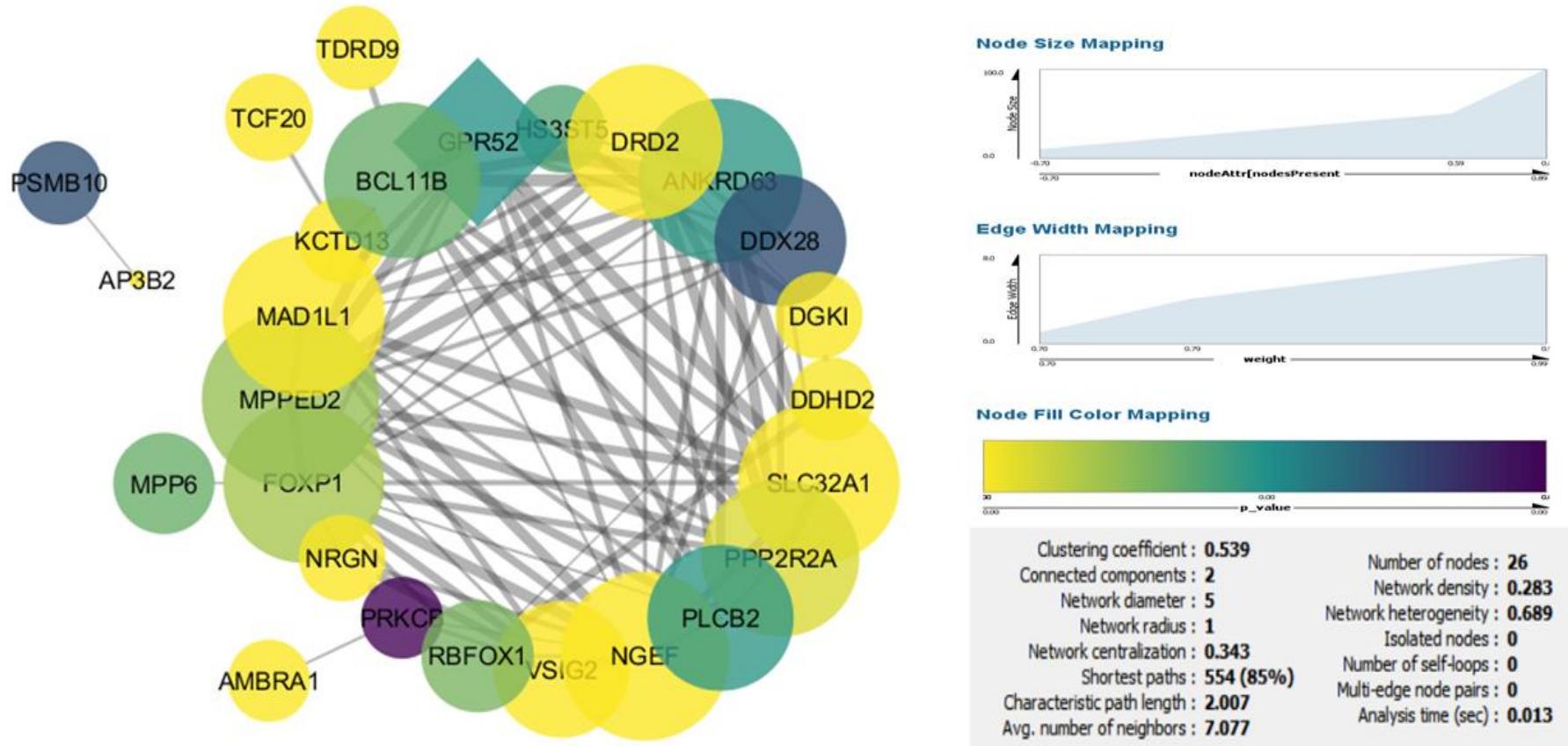


Figure 3.6.8 Green Module for developmental Stage Two where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The schizophrenia-associated genes in the Green module for the developmental stage Two in module 3.6.8 for the most part have two connected components. The genes have a varied module membership from looking at the size of the nodes. The network centralization of 0.343 shows there is not just one key player in this module, but it is not a distributed network where every node is independent and interconnected with each other. The network heterogeneity is 0.689 meaning there is a variance in the number of connections. The clustering coefficient is 0.539 meaning 53.9% of the edges which could exist between the neighbours do in Figure 3.6.8. The network density is 0.285. The edges which are present are found mostly around the HG and the HG's neighbours. The HG for this module is *GPR52*. *GPR52* is an orphan G-protein couple receptor with unknown function (126). Many drugs that are abused regulate G protein coupled receptors in the CNS, *GPR52* may regulate dopaminergic and glutamatergic transmission in neural circuits and could be responsible with cognitive function (127).

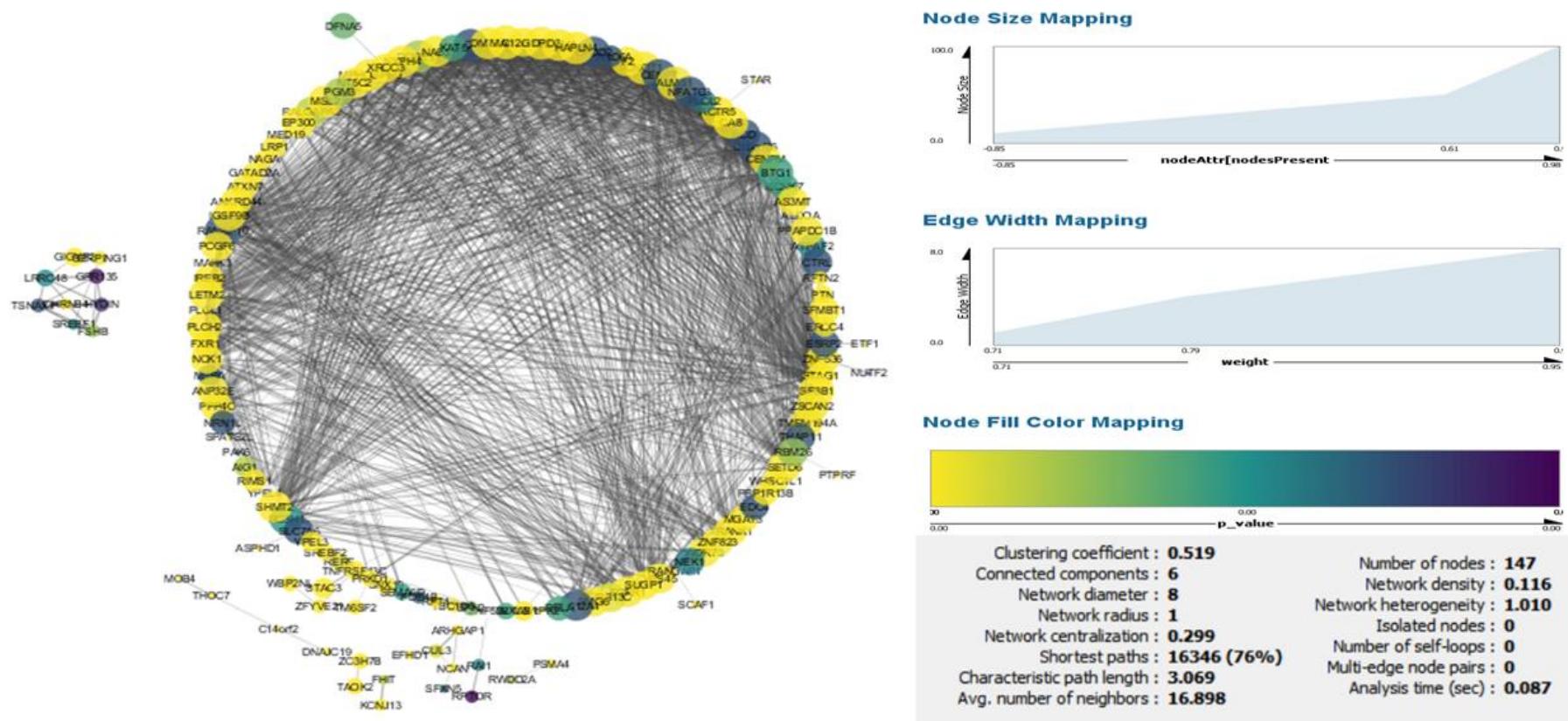


Figure 3.6.9 Turquoise Module for developmental Stage Two where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Turquoise module for developmental stage Two has split into six groupings. The network centralization is low at 0.299 which means it is more like a decentralized network with no central nodes. The network heterogeneity is 1.010 meaning that a lot of the nodes have the same number of edges. The clustering coefficient of 0.519 means 51.9% of the edges which could exist between neighbours do. The network density is 0.116 low, and the top of the module is less populated with edges when compared to the bottom. The HG is *STAG1*, a gene that is frequently mutated in cancers such as leukaemia, glioblastomas and bladder cancers (128,129).*STAG1* regulates gene expression and genome organisation and also has tumour suppressing role(128–130).*STAG1* is required for cohesion at telomeres and is involved in DNA replication (128).

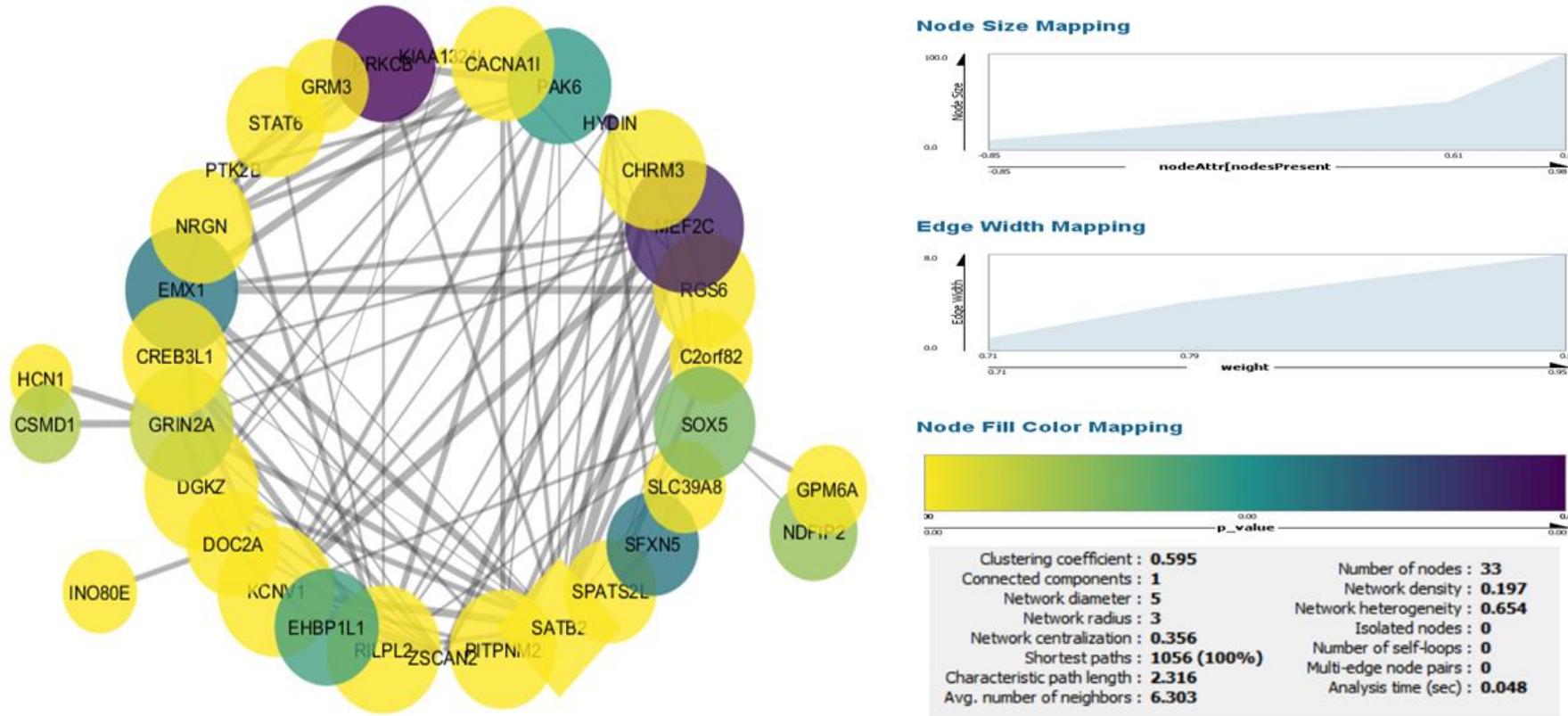


Figure 3.6.10 Blue Module for developmental Stage Three where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Blue module for developmental stage Three seen in Figure 3.6.10 has separated into three submodules. The network centralization is low at 0.356 meaning it is more comparable to a decentralized network. The clustering coefficient of 0.595 means that 59.5% of the edges between neighbours exist. The network heterogeneity is 0.654 meaning that a lot of the nodes have the same number of edges. The HG for this module is *SATB2*, this gene is related to Intelligence, learning and memory (131). *SATB2* is a transcription factor which regulates neocortical circuitry and organisation (131). *SATB2* has been shown to cause *SATB2*-associated syndrome and developmental delays (131,132)

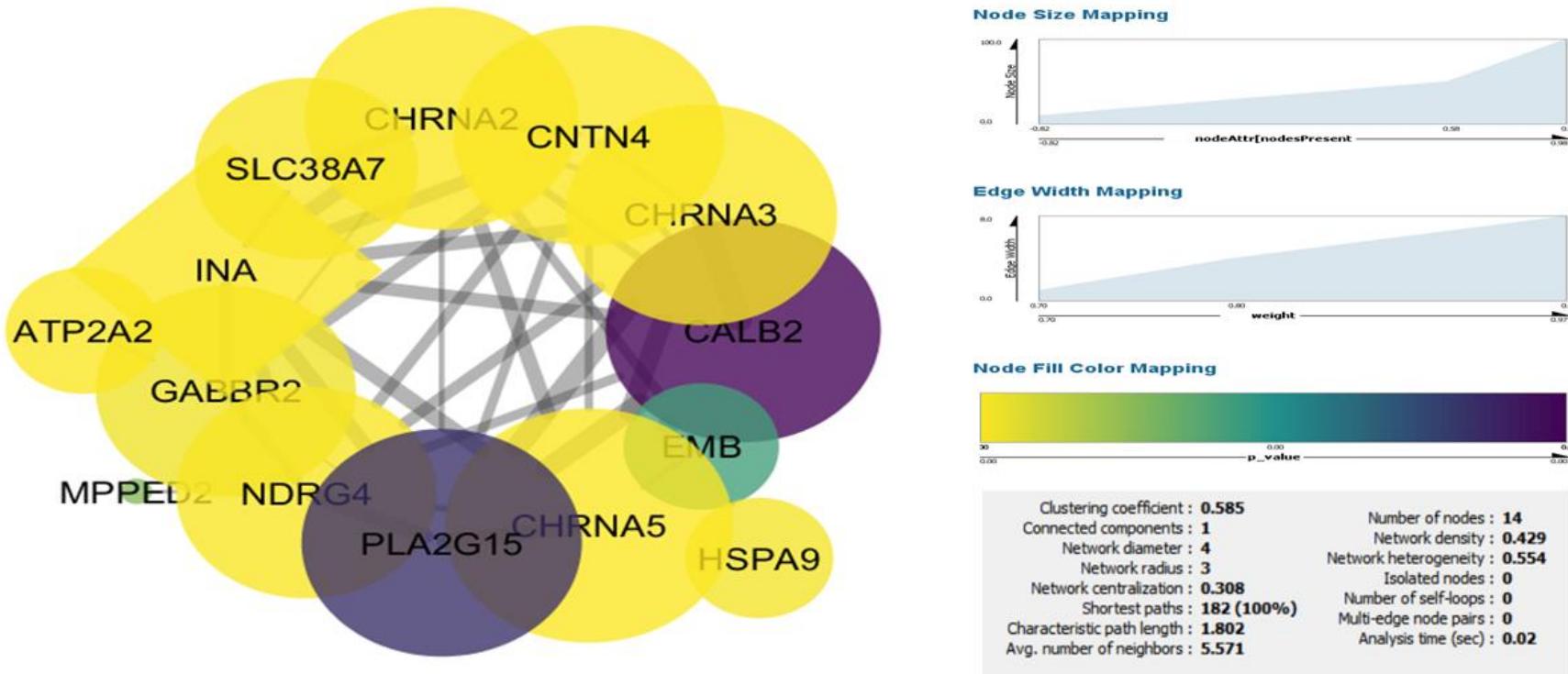


Figure 3.6.11 Brown Module for developmental Stage Three where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Brown module for developmental stage Three seen in Figure 3.6.11 has separated into three submodules. The network centralization is low at 0.308 meaning it is more comparable to a decentralized network. The clustering coefficient of 0.585 means that 58.5% of the edges between neighbours exist. The network heterogeneity is 0.554 meaning that a lot of the nodes have the same number of edges. The HG for this module is *INA*, which is an alpha interleukin which is a class IV neuronal intermediate filament protein which regulates neurons morphogenesis (133). *INA* is expressed in developing neuroblasts and found in the adult CNS in the cytoskeleton in the cerebellar granule cells (133). *INA* has been found to be hypermethylated in CpG islands in the promoter region and is a prognostic marker for large tumours and poor survival rates in colorectal cancer patients (134).

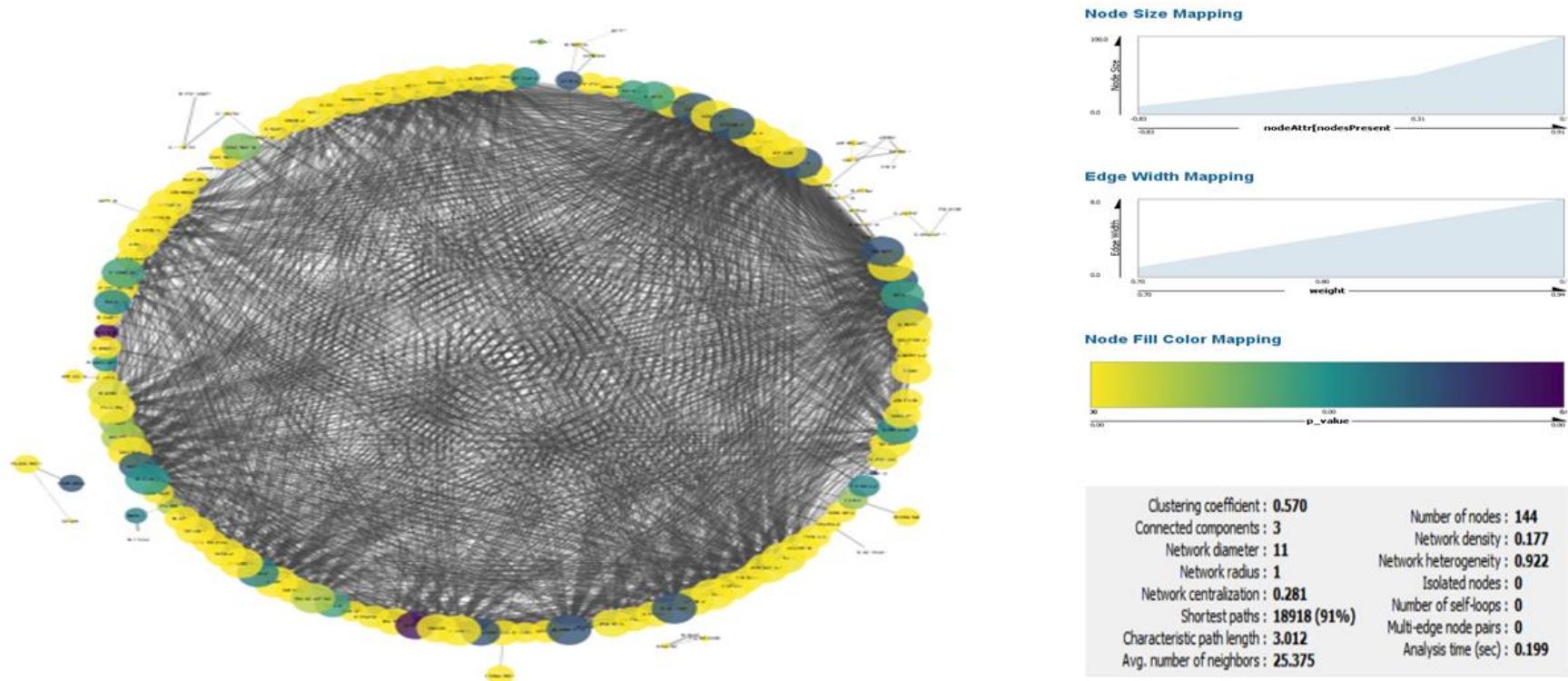


Figure 3.6.12 Turquoise Module for developmental Stage Three where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

This module has three connected components as can be seen in Figure 3.6.12. The network centralization is 0.281 meaning each of the nodes are as important as the rest, like in a decentralized network. The network heterogeneity is 0.922 meaning a lot of the nodes have the same number of edges. The clustering coefficient is 0.614 meaning 61.4% of the edges which could exist between neighbours do. The network density low at 0.177. The HG for this module is *TRANK1*. *TRANK1* is the HG and as well as being linked to schizophrenia it has also been linked to BP. In a study performed by Whelan et al. where circulating IgG antibodies were analysed against 18 target peptide antigens using enzyme linked immunosorbent assays found that IgG levels were increased in *TRANK1* in people with schizophrenia (135). This study also highlighted anti-*TRANK1* IgG as a potential biomarker for a subgroup of schizophrenia (135)

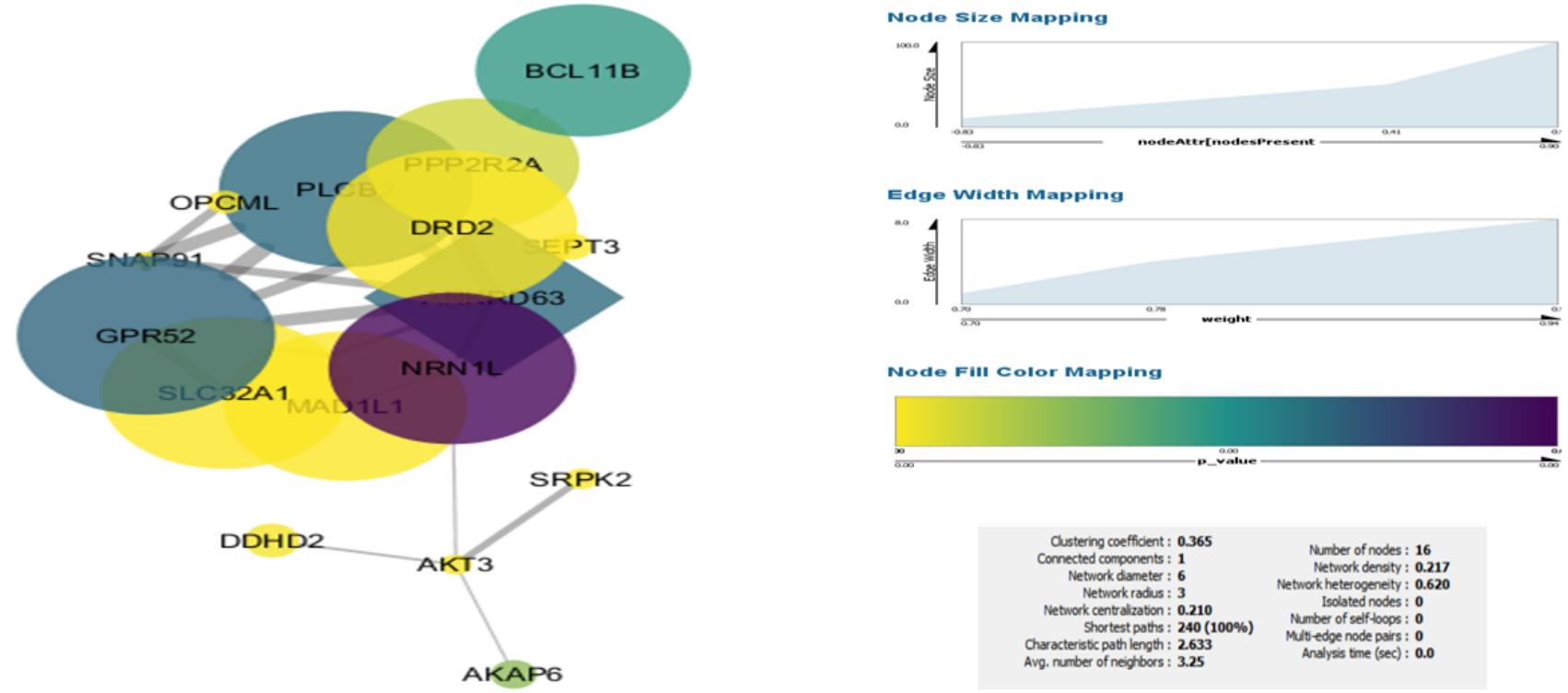


Figure 3.6.13 Yellow Module for developmental Stage Three where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The schizophrenia-associated genes for the Yellow Module in Stage Three The network centralization is low at 0.210, meaning the nodes are of similar importance and are similar to those in a decentralized network. The network heterogeneity is 0.620 meaning there is some diversity regarding the number of edges each node has.

have a similar number of edges. The average clustering coefficient is 0.365 which means 36.5% of the edges which could exist to neighbours do on average. The network density is 0.35 which is low because of the weight threshold. The HG for this module is *ANKRD63*.Little is known about the function of *ANKRD63*.

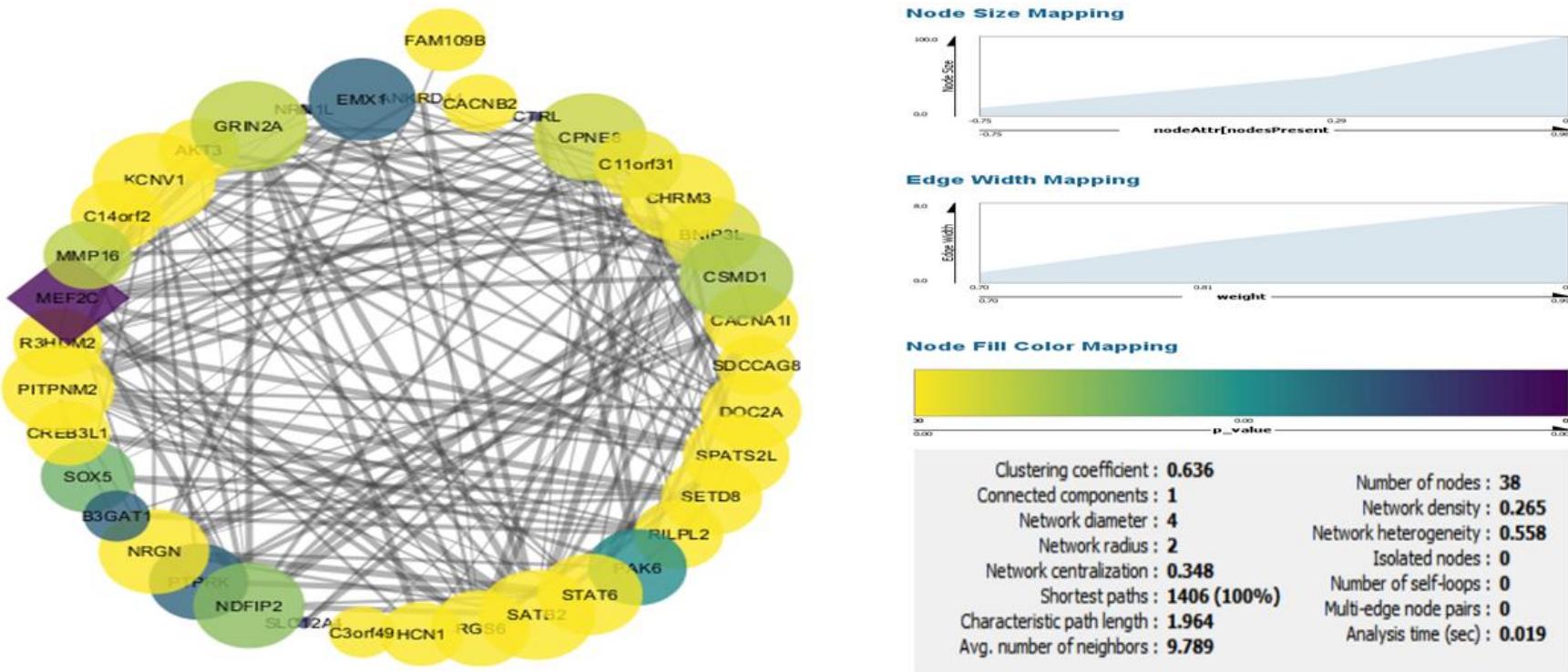


Figure 3.6.14 Blue Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Blue module for developmental stage Four as seen in Figure 3.6.14 is a large module with 38 nodes which all cluster into a group which means the connected component is One. The network centralization is 0.348 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.558 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.636 which is the average clustering coefficient for the module's node, which means on average 63.6% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.265. The HG for this module is *MEF2C*, which encodes a transcription factor and is expressed in microglia (136). It regulates gene expression all through development and regulates genes associated with synapse formation and development, neuronal differentiation, and multiple genes related to neurodevelopmental disorders including ASD (137). *MEF2C* is highly expressed in excitatory cortical neurons during development but its role remains unclear (138).

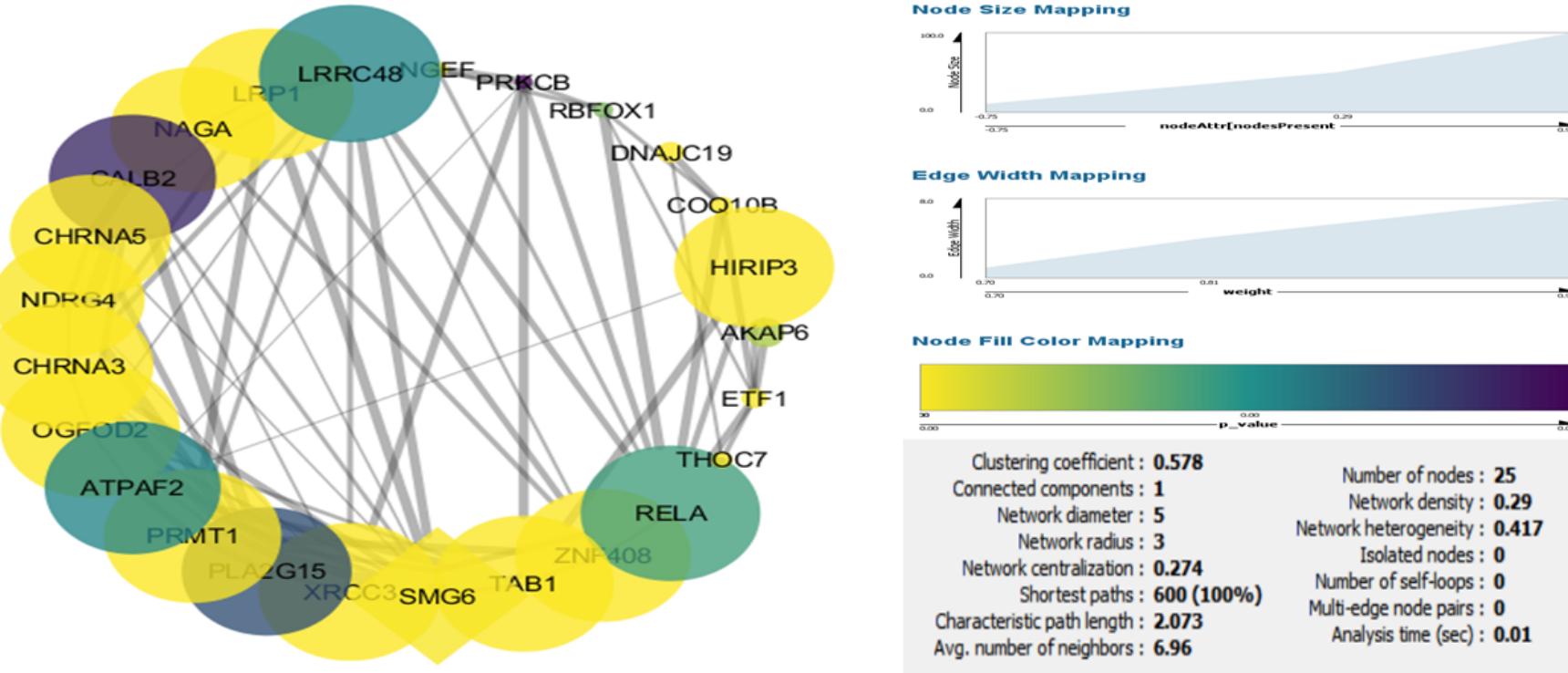


Figure 3.6.15 Brown Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Brown module for developmental stage five as seen in Figure 3.6.14 contains mostly genes with large module membership. No nodes have broken off into sub structure, so the connected components are One. The HG for this module is *SMG6*. Not much is known about *SMG6* but is involved with Pancreatic Adenosquamous Carcinoma and Lissencephaly.

The network centralization is 0.274 meaning the nodes are of a similar importance to the module. The network heterogeneity is 0.417 meaning the degree of most of the nodes is different. The average clustering coefficient is 0.578 which means on average the nodes are connected to 57.8% of the potential edges between its neighbour. The network density is 0.329 which means the portion of potential connections which are actually connections is low. The network centralization is also low which means there are no central players which when removed would collapse the network.

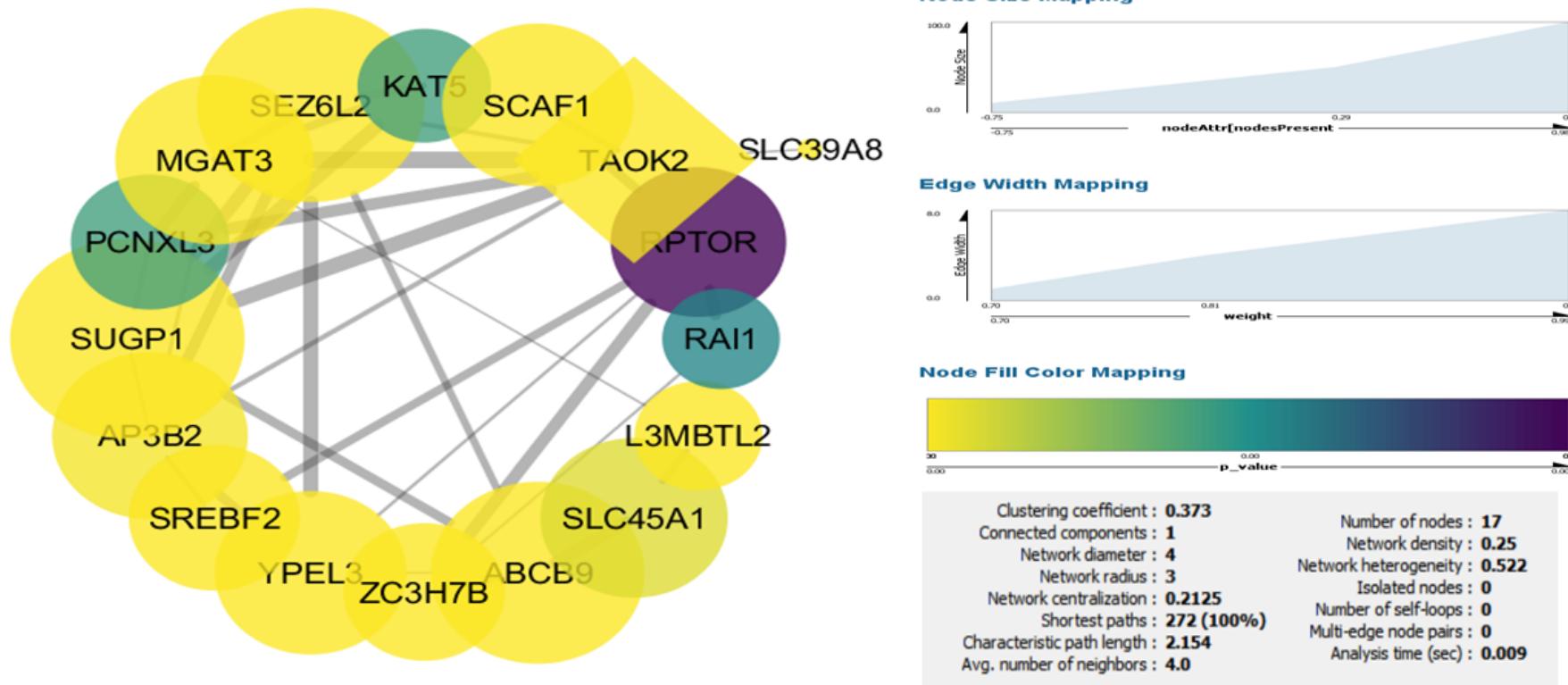


Figure 3.6.16 Green Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Green module for developmental stage Four as seen in Figure 3.6.16 has a connected component on One. The network centralization is 0.2125 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.552 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.373 which is the average clustering coefficient for the module's node, which means on average 37.3% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.25. The HG for this module is *TAOK2* thousand and one kinase 2. This gene has been implicated in neurodevelopmental disorders especially ASD (139). *TAOK2* have been identified to regulate neurodevelopment and synapse formation and regulate the differentiation and development of synapses through modulation of the cytoskeleton (140). It has also been implicated in inflammation and immunity by activating T cells and also can regulate apoptosis (140).

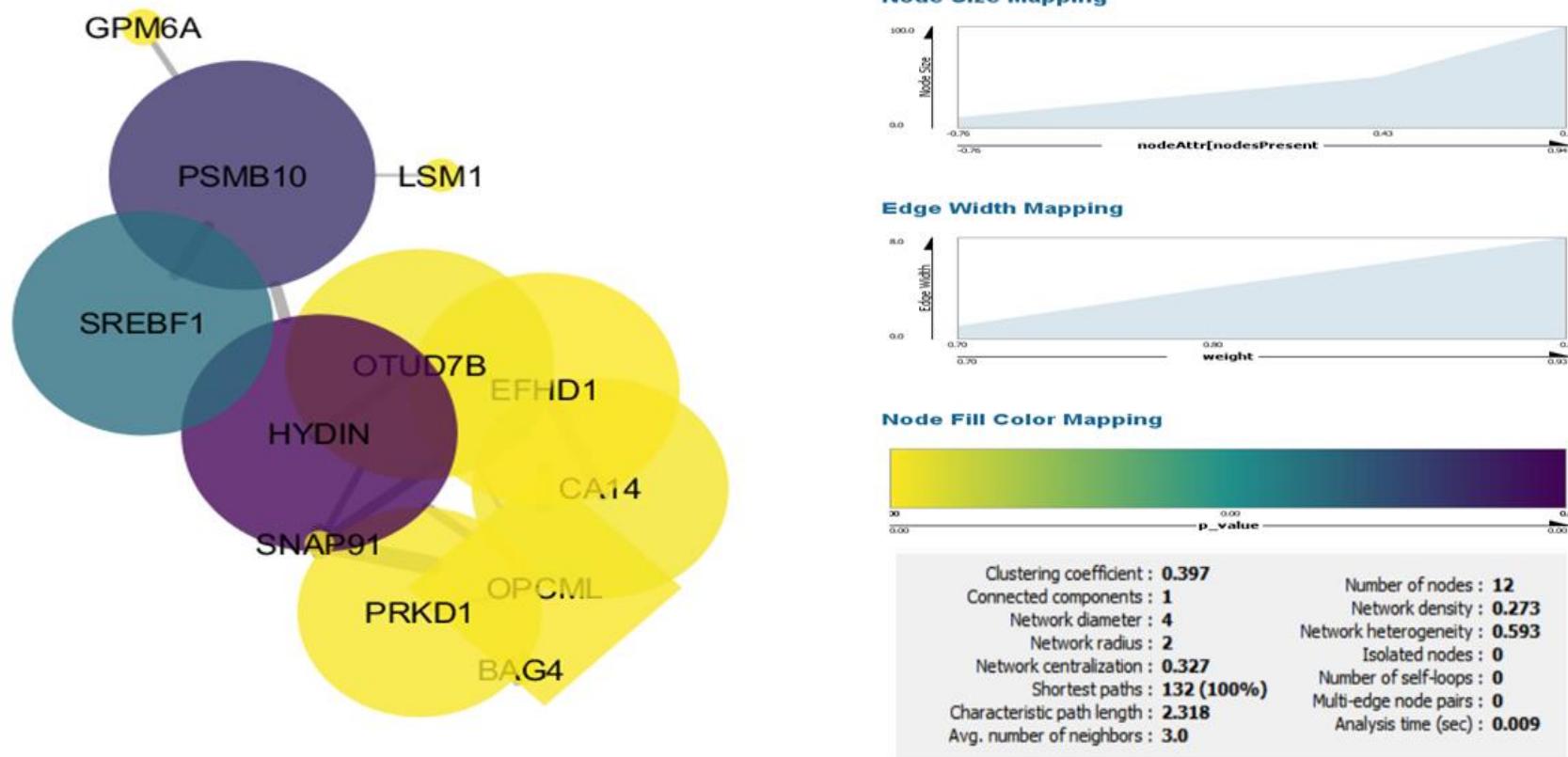


Figure 3.6.17 Magenta Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Magenta module for developmental stage Four as seen in Figure 3.6.17 is a small module with 12 nodes which all cluster into a group which means the connected component is One. The network centralization is 0.327 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.593 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.397 which is the average clustering coefficient for the module's node, which means on average 39.7% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.273. The HG for this module is *OPCML*. Opioid -binding protein/ cell adhesion molecule is a glycosyl-phosphatidylinositol (GPI)-anchored cell adhesion-like molecule which belongs to the IgLON family (141). It is expressed abundantly in the CNS, mainly in the hippocampus and cerebral cortex (142). *OPCML* has tumour suppressor quality and 83% of ovarian cancers the gene is silenced but the hypermethylation of *OPCML* is seen lung, brain and cervical cancer (143). *OPCML* regulates synaptogenesis and synaptic plasticity and the disruption of which is thought to contribute to neurodevelopmental disorders (124). Two SNPs were associated with schizophrenia in Chinese Han population which correlated with reduced *OPCML* expression in the hippocampus (142).

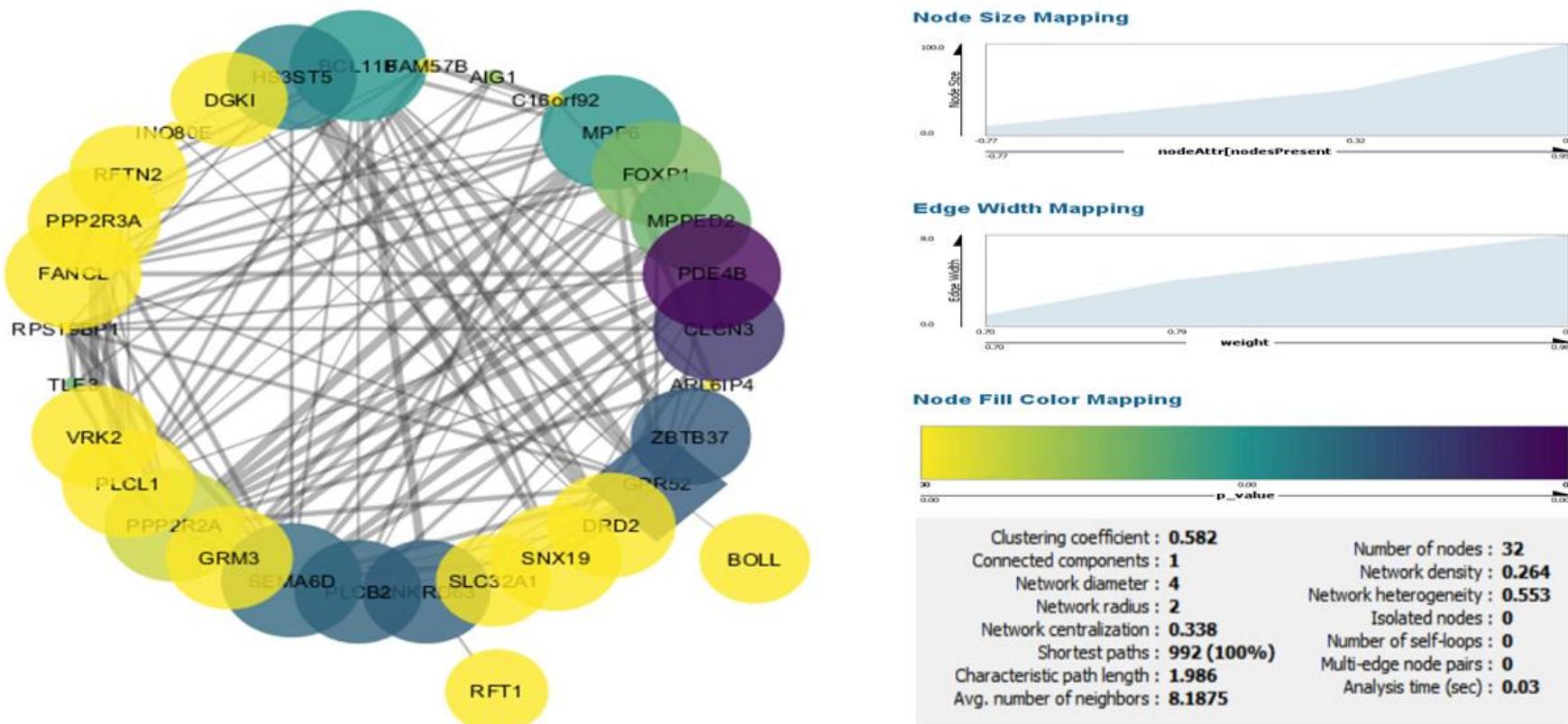


Figure 3.6.18 Purple Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Purple module for developmental stage Four as seen in Figure 3.6.18 is a large module with 32 nodes which all cluster into a group which means the connected component is One. The network centralization is 0.338 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.553 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.582 which is the average clustering coefficient for the module's node, which means on average 58.2% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.264. The HG for this module is *GPR52*. *GPR52* is an orphan G-protein couple receptor with unknown function (126). *GPR52* may regulate dopaminergic and glutamatergic transmission in neural circuits and could be responsible with cognitive function (127).

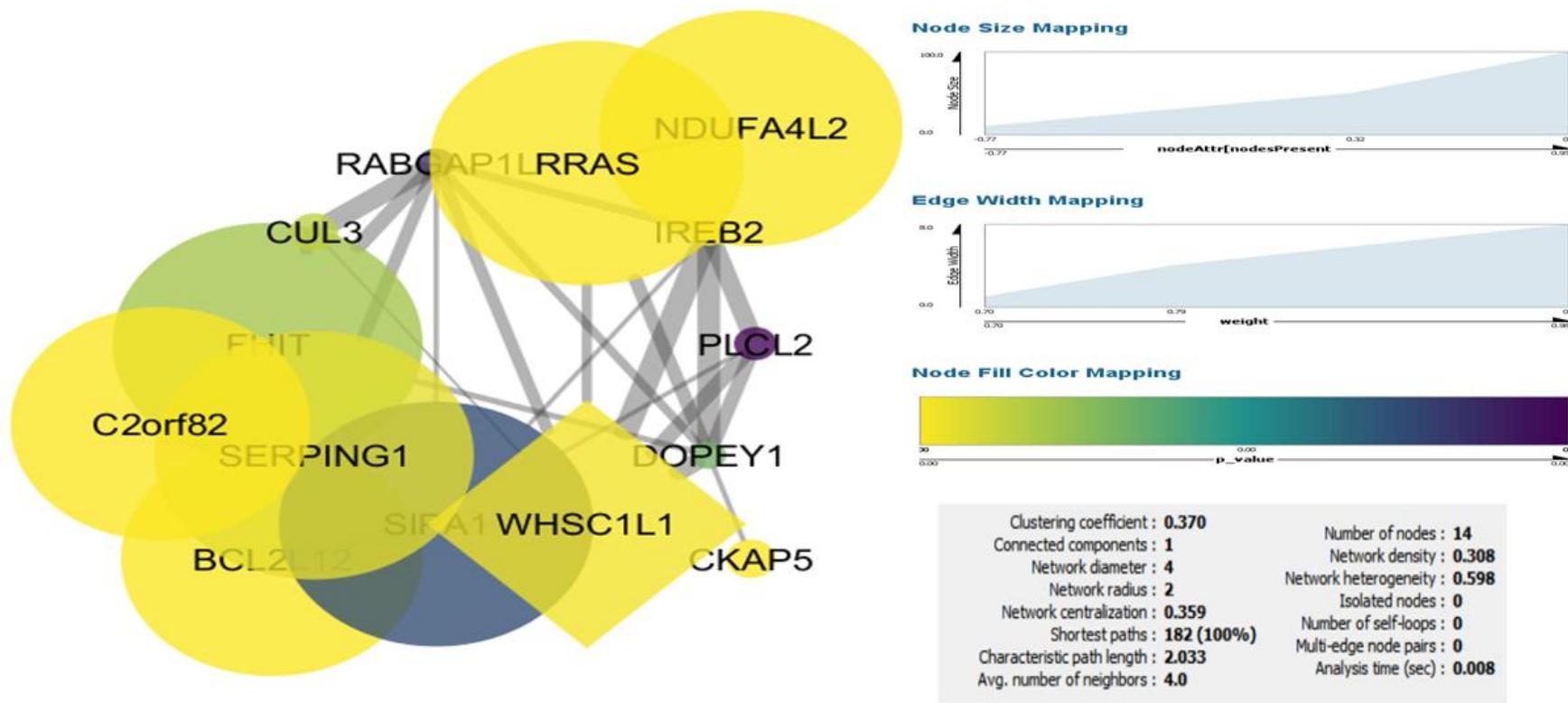


Figure 3.6.19 Red Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Red module for developmental stage Four as seen in Figure 3.6.19 is a large module with 14 nodes and has a connected component of One. The network centralization is 0.359 which means there are no key players. The network heterogeneity is 0.598 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.370 which is the average clustering coefficient for the module's node, which means on average 37% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.308. The HG for this module is *WHSC1LI/NSD3*. This gene is a chromatin modifier by modulating the genes expression through demethylation of lysine 36 on histone H3 (H3K36) (144–146). It is amplified in patients with squamous cell carcinoma of the head and neck (147).

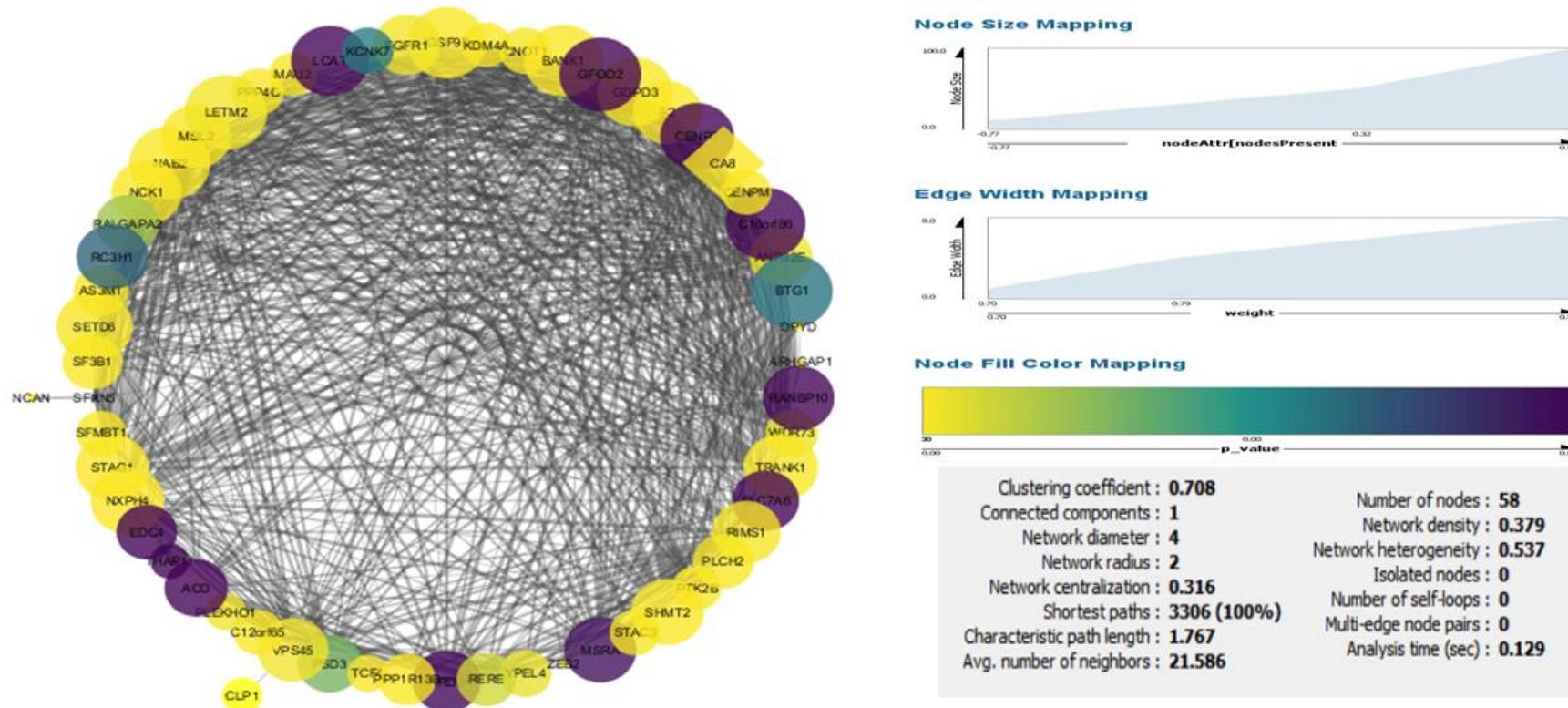


Figure 3.6.20 Turquoise Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Turquoise module for developmental stage Four as seen in Figure 3.6.20 is a large module with 58 nodes which all cluster into a group which means the connected component is One. The network centralization is 0.316 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.537 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.708 which is the average clustering coefficient for the module's node, which means on average 70.8% of the nodes which could exist between a node and its neighbour does exist. The module looks to have a lot of edges, but the network density is only 0.379. The HG for this module is *CA8*, this gene is known as carbonic anhydrase 8 and mutations in this gene can cause mental retardation and cerebellar ataxia (148). *CA8* is still considered a member of the carbonic anhydrase family in spite of the fact it cannot catalyse the hydration of carbon dioxide because of its lack of zinc binding histidine residues (149). It is mostly expressed in Purkinje cells found in the cerebellum but is found in other tissues and has been conserved across tissues which likely means *CA8* has an important function. It is an allosteric inhibitor of IP3R1 which regulates calcium levels in the cell which is linked to key cellular processes such as release of neurotransmitters, nerve processes, neuronal excitability, and mitochondrial energy production (149,150). *CA8* has been linked to epilepsy, MERRF disease, spinal cerebellar ataxia (SCA), mild mental retardation, chronic pain, Alzheimer's and quadrupedal gait (148,150).

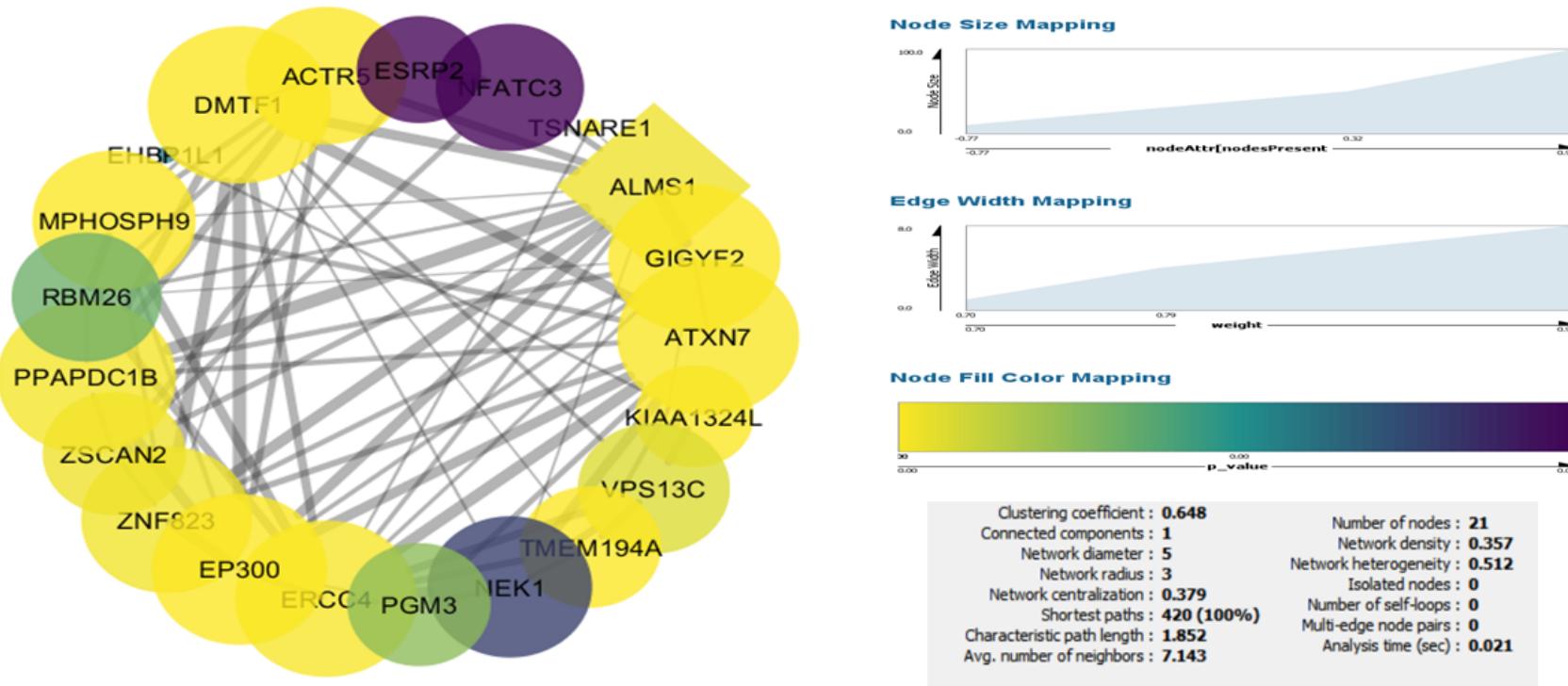


Figure 3.6.21 Yellow Module for developmental Stage Four where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Yellow module for developmental stage Four as seen in Figure 3.6.21 has a connected component of One. The network centralization is 0.379 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.512 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.648 which is the average clustering coefficient for the module's node, which means on average 64.8% of the nodes which could exist between a node and its neighbour does exist. The network density is only 0.358. The HG for this module is *ALMS1*. The *ALMS1* protein is implicated in endosomal trafficking, maintenance of centromere cohesion, transcription and actin organisation (151). *ALMS1* is expressed in most tissues in people who Alstrom syndrome which is a very rare autosomal multisystemic disorder which causes cone-rod dystrophy, obesity, hearing loss, type II diabetes, insulin resistance, progressive hepatic renal dysfunction, and, dilated cardiomyopathy (152).

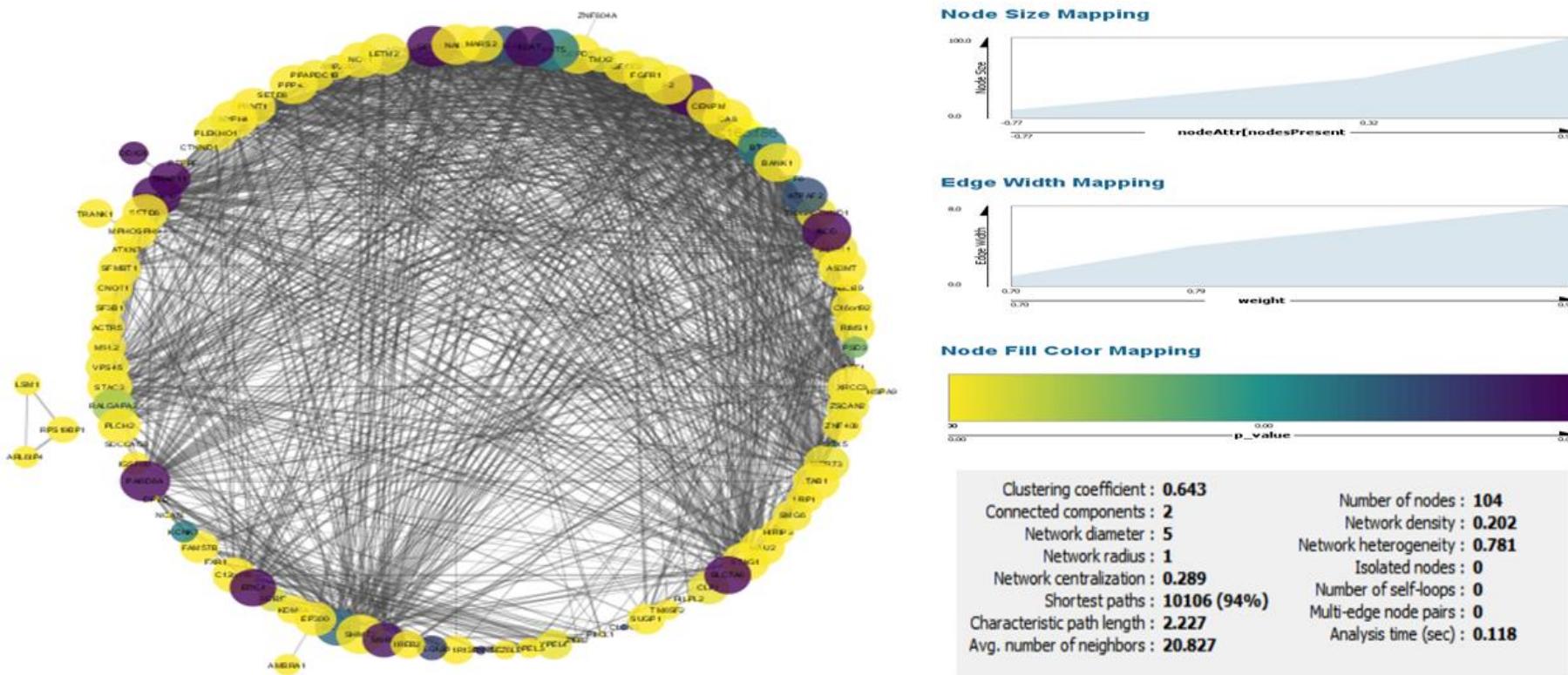


Figure 3.6.22 Black Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Black module for developmental stage Five as seen in Figure 3.6.22 is a large module with 316 and has two connected components. The network centralization is 0.289 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.781 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.643 which is the average clustering coefficient for the module's node, which means on average 64.3% of the nodes which could exist between a node and its neighbour does exist. The module looks like it is densely filled with edges, but the network density is only 0.202. The HG for this module is *C16orf86*. *C16orf86* has been positively correlated with insulin sensitivity in human skeletal muscles (153). Schizophrenia polygenic risk and Insulin resistance are significantly associated with a subgroup of ANP and first episode patients with anti-psychotic resistance (154).

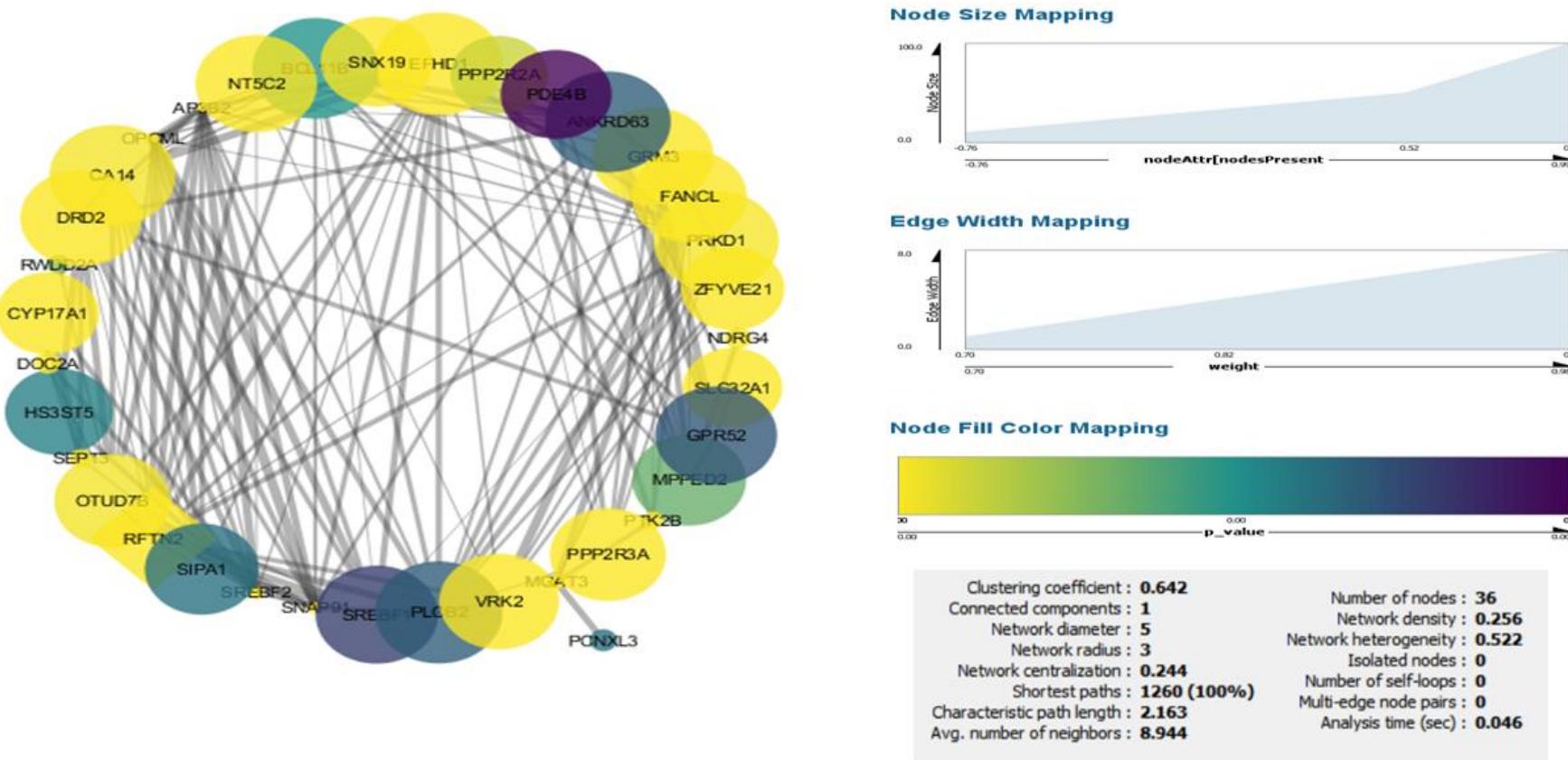


Figure 3.6.23 Brown Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Brown module for developmental stage Five as seen in Figure 3.6.23 and has connected component is One. The network centralization is 0.244 which means there are no key players which when removed would break up the module, each node is of similar importance. The network heterogeneity is 0.552 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.642 which is the average clustering coefficient for the module's node, which means o average 64.2% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.256. The HG for this module is *RFTN2*.*RFTN2* is associated with Glass syndrome/*SATB2* syndrome which is characterised by developmental delay and intellectual disability and limited speech development (132).

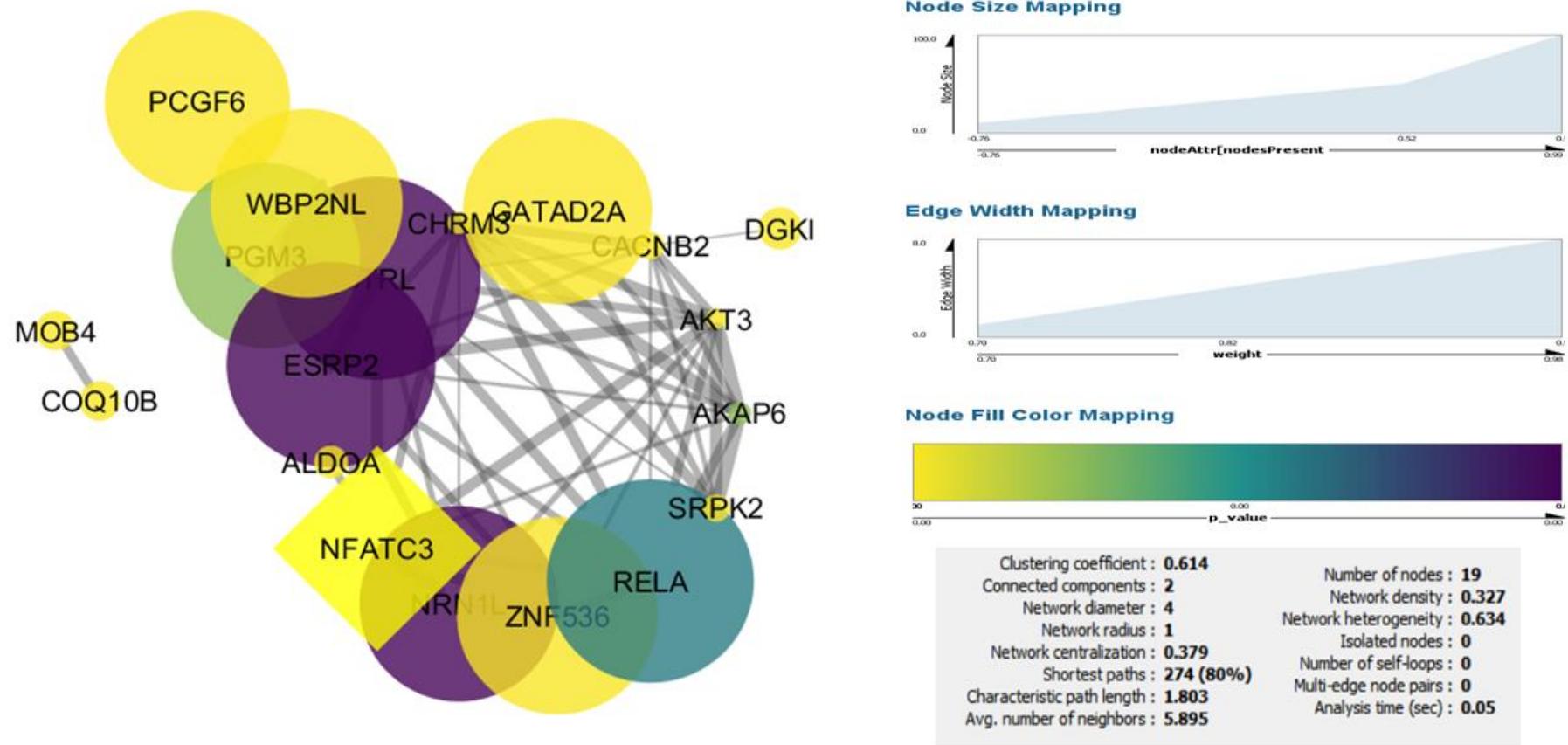


Figure 3.6.24 Green Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Green module for developmental stage Five as seen in Figure 3.6.24 is a small module with 19 nodes and connected component of One. The network centralization is 0.379 which means there are no key players. The network heterogeneity is 0.634 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.614 which is the average clustering coefficient for the module's node, which means on average 61.4% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.327. The HG for this module is *NFATC3*. Nuclear factor of activated T cells 3 upregulates type 1 interferon expression, its abnormal expression is related to tumorigenesis (155). This family of genes is a group of calcium/calcineurin dependant transcription factors which are crucial for homeostasis of brain tissue (155,156). *NFATC3* is involved in development and progression of cancer and is over expressed in tumours, but it has pro and anti-tumour effects depending on the cellular context (155). Members of the NFAT family are expressed in C6, astrocytes and U251 glioma cell lines (156).

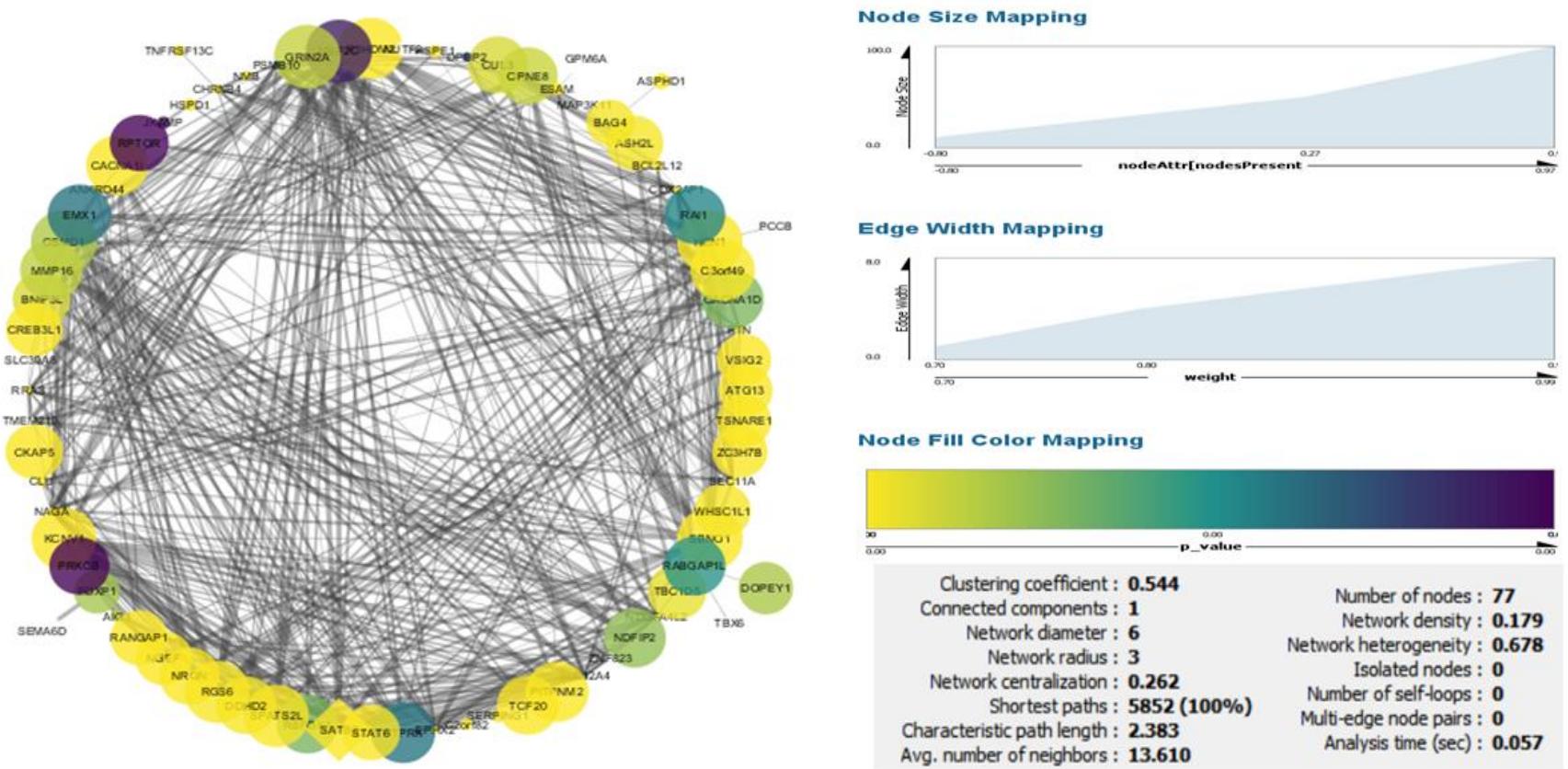


Figure 3.6.25 Greenyellow Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Greenyellow module for developmental stage Five as seen in Figure 3.6.25 is a large module with 77 nodes. The network centralization is 0.262 which means not a lot of the genes have similar importance. The network heterogeneity is 0.678 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.544 which is the average clustering coefficient for the module's node, which means on average 54.4% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.179. The HG for this module is *SATB2*. *SATB2* has been shown to cause SATB2-associated syndrome and developmental delays

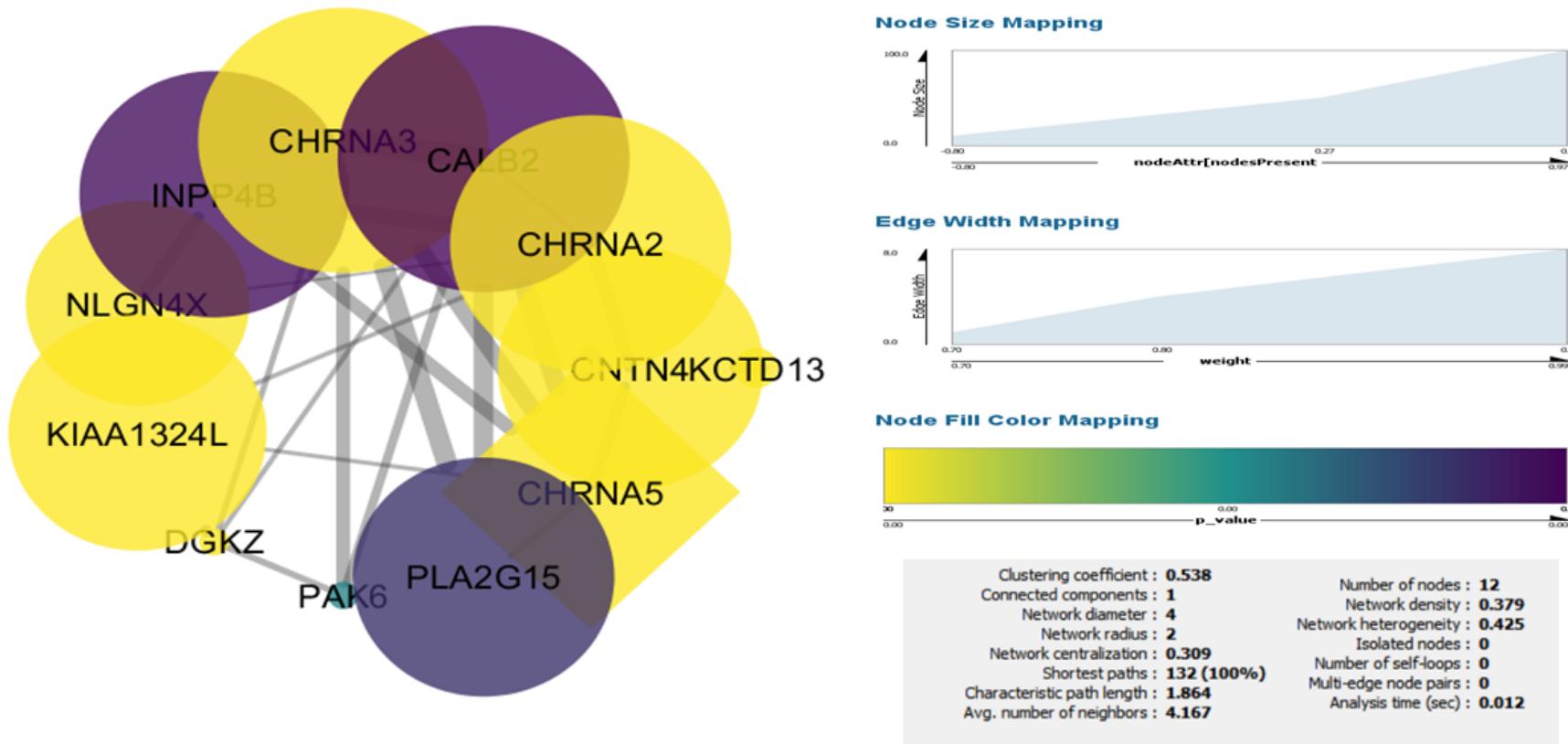


Figure 3.6.26 Pink Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Pink module for developmental stage Five as seen in Figure 3.6.26 is a small module with 12 nodes and one connected component. The network centralization is 0.309 which means there are no key players. The network heterogeneity is 0.425 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.538 which is the average clustering coefficient for the module's node, which means on average 53.8% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.379. The HG for this module is *CHRNA5*. *CHRNA5* a nicotinic acetylcholine receptor alpha 5 subunit is associated with cognitive function in psychiatric disorders and lung cancer(157). Altered cholinergic neural transmission increase the susceptibility in cognitive deficits and the rs16969968 SNP linked to *CHRNA5* was linked with genetic susceptibility, psychotic symptoms and more severe cognitive deficits in early onset schizophrenia in Chinese populations (158).

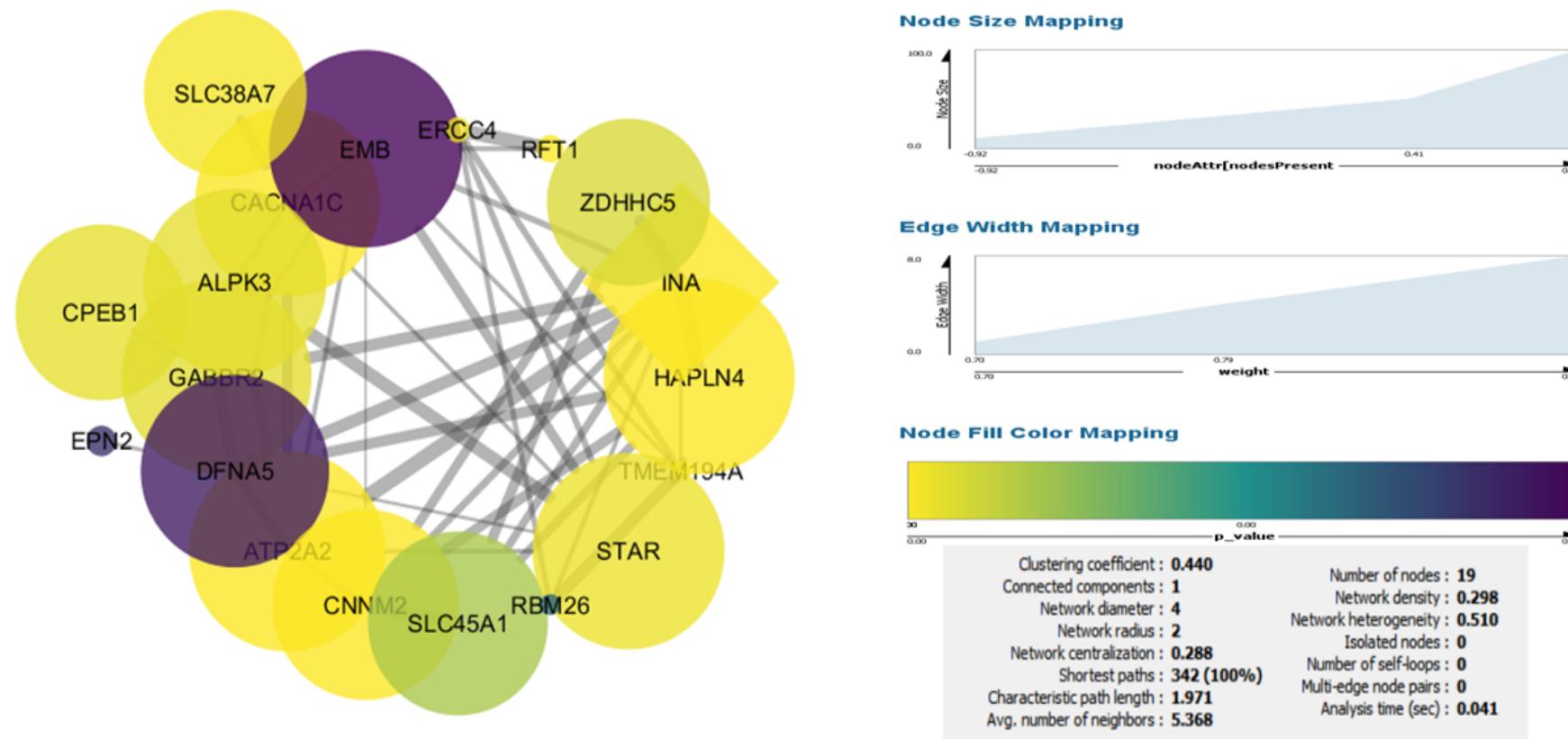


Figure 3.6.27 Red Module for developmental Stage Five where the size of the node is measured using module membership, p-values are based on p-values for SNP inclusion which was determined by Pardiñas et al. and shown by node fill colour and, edge width is measured on weight. The diamond shaped node is the HG identified by WGCNA. The weight of the edges was filtered to only include 0.8 and above.

The Red module for developmental stage Five as seen in Figure 3.6.27 is a large module with 19 nodes. The network centralization is 0.288, each node has similar importance. The network heterogeneity is 0.510 meaning some of the nodes have the same number edges. The average clustering coefficient is 0.440 which is the average clustering coefficient for the module's node, which means on average 44% of the nodes which could exist between a node and its neighbour does exist. The module has areas of it which are densely filled with edges, but the network density is only 0.298. The HG for this module is *INA*. *INA* is an alpha interleukin which is a class IV neuronal intermediate filament protein which regulates neurons morphogenesis (133). *INA* is expressed in developing neuroblasts and found in the adult CNS in the cytoskeleton in the cerebellar granule cells (133). *INA* has been found to be hypermethylated in CpG islands in the promoter region and is a prognostic marker for large tumours and poor survival rates in colorectal cancer patients (134).

3.7 Gene Ontologies

Gene ontologies which are enriched for each module within each stage were identified using the function `enrichmentAnalysis` using the GO collection database in the `anRichment` Bioconductor package in R. Finding GOs which are enriched in each of the modules can deepen the understanding of schizophrenia and lead to the discovery of novel therapeutic targets.

Table 10: Gene Ontologies of the top enriched gene ontologies in Stage One using anRichment

GOID	DEFINITION	ONTOLOGY	Module	GO Process/ Term	FDR
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Black	protein binding	0.002378
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Black	intracellular organelle lumen	0.004713
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Black	positive regulation of macrophage proliferation	0.006297
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	7.02E-09
GO:0051716	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.	BP	Blue	cellular response to stimulus	8.29E-08
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Blue	protein binding	1.11E-07
GO:0097090	A process which results in the assembly, arrangement of constituent parts, or disassembly of a presynaptic membrane, including any proteins associated with the membrane, but excluding other cellular components. A presynaptic membrane is a specialized area of membrane of the axon terminal that faces the plasma membrane of the neuron or muscle fiber with which the axon terminal establishes a synaptic junction.	BP	Brown	presynaptic membrane organization	7.99E-05
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Brown	positive regulation of macrophage proliferation	8.30E-05
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Brown	intracellular organelle lumen	8.38E-05
GO:0005654	That part of the nuclear content other than the chromosomes or the nucleolus.	CC	Pink	nucleoplasm	0.012863
GO:1905907	Any process that stops, prevents or reduces the frequency, rate or extent of amyloid fibril formation.	BP	Pink	negative regulation of amyloid fibril formation	0.01836
GO:1901398	NA	BP	Pink	regulation of transforming growth factor beta3 activation	0.03198
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Turquoise	intracellular organelle lumen	4.04E-10
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	1.78E-09
GO:0070016	Interacting selectively and non-covalently with the armadillo repeat domain of a protein, an approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity protein armadillo. Arm-repeat proteins are involved in various processes, including intracellular signalling and cytoskeletal regulation.	MF	Turquoise	armadillo repeat domain binding	4.05E-09

Table 11: Gene Ontologies for the most enriched ontologies in Stage Two using anRichment

GOID	DEFINITION	ONTOLOGY	Module	GO Process/ Term	FDR
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Blue	intracellular organelle lumen	0.00018
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Blue	protein binding	0.00048
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	0.000577
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Brown	intracellular organelle lumen	1.85E-05
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Brown	positive regulation of macrophage proliferation	2.91E-05
GO:0005737	All of the contents of a cell excluding the plasma membrane and nucleus but including other subcellular structures.	CC	Brown	cytoplasm	4.47E-05
GO:0005654	That part of the nuclear content other than the chromosomes or the nucleolus.	CC	Green	nucleoplasm	0.021661
GO:0140244	NA	BP	Green	regulation of translation at presynapse	0.039021
GO:0098700	The active transport of neurotransmitters into a synaptic vesicle. This import is fuelled by an electrochemical gradient across the vesicle membrane, established by the action of proton pumps.	BP	Green	neurotransmitter loading into synaptic vesicle	0.040787
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	3.35E-15
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Turquoise	intracellular organelle lumen	9.06E-10
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Turquoise	protein binding	6.74E-09

Table 12: Gene Ontologies for the most enriched ontologies in Stage Three using anRichment

GOID	DEFINITION	ONTOLOGY	Module	GO Process/ Term	FDR
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	0.044306
GO:0072588	A ribonucleoprotein complex that contains an RNA of the box H/ACA type and the four core proteins dyskerin, NOP10, NHP2, and GAR1 (human protein nomenclature). RNA pseudouridylation (isomerization of uridine to pseudouridine) is the major, and most likely the ancestral, function of H/ACA RNPs. Pseudouridylation targets include both large and small ribosomal RNAs (rRNAs), and small nuclear RNA (U2 snRNA). In addition to these catalytic H/ACA RNPs, a less abundant but more diverse class of structural H/ACA RNPs exists, which does not have pseudouridylation activity. These include the vertebrate telomerase RNP complex.	CC	Blue	box H/ACA RNP complex	0.062848
GO:0016787	Catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3.	MF	Blue	hydrolase activity	0.080744
GO:0001659	A homeostatic process in which an organism modulates its internal body temperature.	BP	Brown	temperature homeostasis	0.00224
GO:0070025	Interacting selectively and non-covalently with carbon monoxide (CO).	MF	Brown	carbon monoxide binding	0.002884
GO:0098935	The directed movement of organelles or molecules along microtubules in dendrites.	BP	Brown	dendritic transport	0.003021
GO:0120041	Any process that activates or increases the frequency, rate, or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	1.99E-19
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	BP	Turquoise	cellular component organization	4.26E-15
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl) pyrocatechol [4-(2-aminoethyl) benzene-1,2-diol] and derivatives formed by substitution.	BP	Turquoise	cellular response to catecholamine stimulus	7.52E-15
GO:0044260	The chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass, as carried out by individual cells.	BP	Yellow	cellular macromolecule metabolic process	0.00076
GO:0010468	Any process that modulates the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's coding sequence is converted into a mature gene product or products (proteins or RNA). This includes the production of an RNA transcript as well as any processing to produce a mature RNA product or an mRNA or circRNA (for protein-coding genes) and the translation of that mRNA or circRNA into protein. Protein maturation is included when required to form an active form of a product from an inactive precursor form.	BP	Yellow	regulation of gene expression	0.000894
GO:0051171	Any process that modulates the frequency, rate or extent of the chemical reactions and pathways involving nitrogen or nitrogenous compounds.	BP	Yellow	regulation of nitrogen compound metabolic process	0.001611

Table 13: Gene Ontologies for the most enriched ontologies in Stage Four using anRichment

GOID	DEFINITION	ONTOLOGY	Module	GO Process/ Term	FDR
GO:0051716	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.	BP	Blue	cellular response to stimulus	0.004574
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Blue	intracellular organelle lumen	0.008924
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	0.01012
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Brown	intracellular organelle lumen	0.097398
GO:1901723	NA	BP	Brown	negative regulation of cell proliferation involved in kidney development	0.151629
GO:0018958	The chemical reactions and pathways involving a phenol, any compound containing one or more hydroxyl groups directly attached to an aromatic carbon ring.	BP	Brown	phenol-containing compound metabolic process	0.184619
GO:0045211	A specialized area of membrane facing the presynaptic membrane on the tip of the nerve ending and separated from it by a minute cleft (the synaptic cleft). Neurotransmitters cross the synaptic cleft and transmit the signal to the postsynaptic membrane.	CC	Green	postsynaptic membrane	0.176919
GO:0014069	An electron dense network of proteins within and adjacent to the postsynaptic membrane of an asymmetric, neuron-neuron synapse. Its major components include neurotransmitter receptors and the proteins that spatially and functionally organize them such as anchoring and scaffolding molecules, signalling enzymes and cytoskeletal components.	CC	Green	postsynaptic density	0.176919
GO:0099601	Any process that modulates the frequency, rate, or extent of neurotransmitter receptor activity. Modulation may be via an effect on ligand affinity, or effector function such as ion selectivity or pore opening/closing in ionotropic receptors.	BP	Green	regulation of neurotransmitter receptor activity	0.18328
GO:0072511	The directed movement of inorganic cations with a valency of two into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore. Inorganic cations are atoms or small molecules with a positive charge which do not contain carbon in covalent linkage.	BP	Magenta	divalent inorganic cation transport	0.252996
GO:0010606	Any process that increases the rate, frequency, or extent of the aggregation, arrangement and bonding together of proteins and RNA molecules to form a cytoplasmic mRNA processing body.	BP	Magenta	positive regulation of cytoplasmic mRNA processing body assembly	0.27126
GO:0030015	The core of the CCR4-NOT complex. In <i>Saccharomyces</i> the CCR4-NOT core complex comprises Ccr4p, Caf1p, Caf40p, Caf130p, Not1p, Not2p, Not3p, Not4p, and Not5p.	CC	Magenta	CCR4-NOT core complex	0.27126
GO:0006816	The directed movement of calcium (Ca) ions into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore.	BP	Purple	calcium ion transport	0.011158
GO:0070852	A neuron projection that is found in unipolar neurons and corresponds to the region between the cell body and the point at which the single projection branches.	CC	Purple	cell body fibre	0.014208
GO:0072528	The chemical reactions and pathways resulting in the formation of a pyrimidine-containing	BP	Purple	pyrimidine-containing	0.014208

	compound, i.e. any compound that contains pyrimidine or a formal derivative thereof.			compound biosynthetic process	
GO:0046632	The process in which a precursor cell type acquires the specialized features of an alpha-beta T cell. An alpha-beta T cell is a T cell that expresses an alpha-beta T cell receptor complex.	BP	Red	alpha-beta T cell differentiation	0.151629
GO:0048523	Any process that stops, prevents, or reduces the frequency, rate or extent of a cellular process, any of those that are carried out at the cellular level, but are not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level.	BP	Red	negative regulation of cellular process	0.166883
GO:0032625	The appearance of interleukin-21 due to biosynthesis or secretion following a cellular stimulus, resulting in an increase in its intracellular or extracellular levels.	BP	Red	interleukin-21 production	0.185808
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	0.004574
GO:0016020	A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it.	CC	Turquoise	membrane	0.01012
GO:0003674	A molecular process that can be carried out by the action of a single macromolecular machine, usually via direct physical interactions with other molecular entities. Function in this sense denotes an action, or activity, that a gene product (or a complex) performs. These actions are described from two distinct but related perspectives: (1) biochemical activity, and (2) role as a component in a larger system/process.	MF	Turquoise	Molecular function	0.01012
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Yellow	protein binding	0.021765
GO:0072538	An immune response which is associated with resistance to intracellular bacteria with a key role in inflammation and tissue injury. This immune response is associated with pathological autoimmune conditions such as multiple sclerosis, arthritis and psoriasis which is typically orchestrated by the production of particular cytokines by T-helper 17 cells, most notably interleukin-17, IL-21 and IL-22.	BP	Yellow	T-helper 17 type immune response	0.105736
GO:0030425	A neuron projection that has a short, tapering, morphology. Dendrites receive and integrate signals from other neurons or from sensory stimuli, and conduct nerve impulses towards the axon or the cell body. In most neurons, the impulse is conveyed from dendrites to axon via the cell body, but in some types of unipolar neuron, the impulse does not travel via the cell body.	CC	Yellow	dendrite	0.185808

Table 14: Gene Ontologies for the most enriched ontologies in Stage Five using anRichment

GOID	DEFINITION	ONTOLOGY	Module	GO Process/ Term	FDR
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Black	positive regulation of macrophage proliferation	1.83E-07
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	BP	Black	cellular component organization	1.10E-06
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl) pyrocatechol [4-(2-aminoethyl) benzene-1,2-diol] and derivatives formed by substitution.	BP	Black	cellular response to catecholamine stimulus	1.66E-06
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Brown	intracellular organelle lumen	0.022271
GO:0048522	Any process that activates or increases the frequency, rate or extent of a cellular process, any of those that are carried out at the cellular level but are not necessarily restricted to a single cell. For example, cell communication occurs among more than one cell, but occurs at the cellular level.	BP	Brown	positive regulation of cellular process	0.029478
GO:0051716	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.	BP	Brown	cellular response to stimulus	0.086947
GO:0008104	Any process in which a protein is transported to, or maintained in, a specific location.	BP	Green	protein localization	0.030022
GO:0000228	A chromosome that encodes the nuclear genome and is found in the nucleus of a eukaryotic cell during the cell cycle phases when the nucleus is intact.	CC	Green	nuclear chromosome	0.056322
GO:0060381	Any process that increases the frequency, rate or extent of single-stranded telomeric DNA binding.	BP	Green	positive regulation of single-stranded telomeric DNA binding	0.070819
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Greenyellow	positive regulation of macrophage proliferation	1.09E-08
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Greenyellow	intracellular organelle lumen	4.27E-08
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Greenyellow	protein binding	7.18E-07
GO:1902913	NA	BP	Pink	positive regulation of neuroepithelial cell differentiation	0.057408
GO:0061351	The multiplication or reproduction of neural precursor cells, resulting in the expansion of a cell population. A neural precursor cell is either a nervous system stem cell or a nervous system progenitor cell.	BP	Pink	neural precursor cell proliferation	0.060301
GO:0047710	Catalysis of the reaction: P(1), P(3)-bis(5'-adenosyl) triphosphate + H(2)O = ADP + AMP + 2 H(+).	MF	Pink	bis(5'-adenosyl)-triphosphatase activity	0.086947
GO:0098808	Interacting selectively and non-covalently with a 7-methylguanosine (m7G) group or derivative located at the 5' end of an mRNA molecule.	MF	Red	mRNA cap binding	0.001777
GO:0043197	A small, membranous protrusion from a dendrite that forms a postsynaptic	CC	Red	dendritic spine	0.003828

	compartment - typically receiving input from a single presynapse. They function as partially isolated biochemical and an electrical compartment. Spine morphology is variable including "thin", "stubby", "mushroom", and "branched", with a continuum of intermediate morphologies. They typically terminate in a bulb shape, linked to the dendritic shaft by a restriction. Spine remodelling is thought to be involved in synaptic plasticity.				
GO:0030054	A cellular component that forms a specialized region of connection between two or more cells or between a cell and the extracellular matrix. At a cell junction, anchoring proteins extend through the plasma membrane to link cytoskeletal proteins in one cell to cytoskeletal proteins in neighbouring cells or to proteins in the extracellular matrix.	CC	Red	cell junction	0.008107

4.0 Discussion

Schizophrenia is a chronic and disabling disorder that effects 1% of the general population whose causes remain unclear even with much research into disease etiology (4). Schizophrenia is a neurodevelopmental disorder, therefore, constructing and exploring networks of genes, previously identified as associated with schizophrenia, over key developmental stages could aid our understanding of schizophrenias etiology and identify novel therapeutic targets of disease.

Although schizophrenia is researched extensively from many angles its mechanism still remains elusive. It is difficult to predict who will develop schizophrenia because of the complex genetic and environmental interactions and its neurodevelopmental model. In order to understand schizophrenia it is important to study it at multiple clinical stages (13,53,159). Insults to brain development *in utero* can have an impact on the severity of schizophrenia's symptoms later in life. Changes in gene expression over the course of schizophrenia have been observed, the study performed by Ota et al. looked at participants blood and compared expression levels of schizophrenia genes in clinically high-risk patients through to chronic schizophrenia. They observed changes in gene expression profiles at different clinical stages (13). Studies like these highlight the importance of looking at schizophrenia at multiple stages to captures its heterogeneity.

In this study, genes previously identified in a large-scale schizophrenia GWAS as being significantly associated with schizophrenia were used to filter ABA's Brainspan resource to include gene expression data for these genes across 16 brain areas and five developmental stages. To find underlying patterns in the gene expression data K-means analysis was utilised. A systems biology approach (WGCNA) was used to describe the pairwise relationship between genes at each development stage and to develop gene networks (i.e. modules) of schizophrenia-associated genes which were co-expressed in each developmental stage. Next,

GO enrichment analysis was performed on each of the modules using anRichment to aid biological interpretation of the identified networks in each developmental stage.

4.1 K-means analysis on the schizophrenia-associated genes

K-means is a simple method of Clustering an unsupervised machine learning technique which groups objects but does not classify them. The results can be seen in the form of a bar chart in Figure 3.3.1. In Developmental Stage One the optimal cluster number is Three and Developmental Stage Two to Five the optimal number of clusters is two and. Cluster assignments for each gene in each developmental stage can be found in Table 2 in the appendix. Each schizophrenia-associated gene was assigned to the closest centre (k). Developmental Stage One to Five can be seen in Figure 3.3.2 to 3.3.7 and a comparison of the five stages in 3.3.7.

If a cluster is filled with genes which are very similar the within cluster sum of squares of the cluster will be small and when the cluster is visualised it will appear small and compact. Each of the clusters for the schizophrenia-associated genes for each Developmental Stages have large within cluster sum of squares, this means there is variance within the cluster. The total within variance which is representative of the total variance in the data is low which would usually be expected to be high for good clustering. Usually only a small amount of the total variance is explained by the within cluster sum of squares but as the total variance is low and the within sum of squares is high it is actually explaining a large portion of the variance. This tells us the schizophrenia-associated genes expression profile in the brain are similar to each other. From the k-means results it is clear that these genes do not cluster well together, if you increase the number of clusters for each of the Developmental Stages the sum of squares with each cluster decreases and the total variance increases but this would mean ignoring the NbClust which is not desirable.

4.2 Weighted Gene Correlations Network Analysis on schizophrenia-associated genes

WGCNA is a systems biology method which uses gene transcripts to describe pairwise relationships between the genes. WGCNA was used to identify modules of schizophrenia-associated genes which were co-expressed for each developmental stage. WGCNA identified three modules in developmental stage One and Five, four in developmental stages Two and two in developmental Stages Three and Four. In each a HG was identified and can be seen in Table 9. WGCNA relies strongly on the assumption that gene co-expression networks follow scale-free topology where highly connected genes are essential for a functioning system, meaning HGs or the proteins they code for could be dysfunctional (160). It also assumes that gene products associated with the same phenotype usually participate in the same module (160). When WGCNA was performed on gene expression profiles of 316 schizophrenia-associated genes in the brain, each gene was assigned to a module and each module was given a colour. A hierarchical clustering graph of initial and final (after minimum module size of 10) module assignments for each Developmental Stage can be seen in Figure 3.4.2. After module construction each genes module membership and weight were calculated for downstream visualisation using Cytoscape.

4.3 Visualisation of modules using data from WGCNA.

First genes were assigned to modules. Then ME's and thus HG's were identified. MM and weight were calculated for every gene. This information was collated into csv files, which uploaded to Cytoscape for visualisation. Cytoscape allows for the visualisation of the modules while highlighting important parameters using the attributes of the network graph. In each graph node size is shown by the module membership, edge width signifies weight and node colour displays p-value as calculated by Pardiñas et al. These p-values range from 2.12×10^{-44} –

4.88×10^{-8} the lower the value the greater the association. Each Cytoscape network can be seen in Figure 3.6.1 to 3.6.14. As there was a large number of edges, only the edges width of 0.7 and over were included to capture the most important edges.

4.3.1 Developmental Stage One - Prenatal

BrainSpan's Developmental Stage One includes gene expression for healthy prenatal brains.

4.3.2 Developmental Stage Two – Infant (0-2 years)

BrainSpan's Developmental Stage Two includes gene expression for a healthy infant's brain from 0-2 years old.

.

4.3.3 Developmental Stage Three – Child (3- 11 years)

Developmental Stage Three contains gene expression across brains of children between the ages of 3 and 11 years.

Figure 3.6.9 displays the Brown Module for Developmental Stage Three. This is a large module with 63 genes. The HG for this module is *TRANK1* which encodes a protein in the brain with unknown function. Decreased expression of *TRANK1* affected the expression of several genes which are involved in neural development (180). Low levels of TRANK1 mRNA expression is a BP risk factor (180,181).

Both *GRIN2A* and *TRANK1* are strongly linked to neuropsychiatric disorders. This is a particularly important developmental stage as in some individuals who develop schizophrenia some mild cognitive disturbances may occur before psychosis onset which would typically occur in Developmental Stage Four. *GRIN2A* encodes a NMDA subunit which enables the neurotransmitter to regulate synaptic function in the CNS (177). Synaptic connections are constantly changing (synaptic plasticity) underlie learning, personality and memory (182). Impaired synaptic plasticity has been linked to schizophrenia, synaptic plasticity may interact

with brain maturation over development to progressively see the symptoms associated with schizophrenia (183).

4.3.4 Developmental Stage Four – Adolescent (12-19 years)

Developmental Stage Four captures gene expression in the brain for adolescents. Typically schizophrenia manifests its symptoms during this time frame (184). Figure 3.6.10 illustrates the Blue Module for Developmental Stage Four. This module has two sub-modules connected by one edge. The edge is between *FOXP1*, a transcription factor, and *RBFOX1* which regulates RNA metabolism (185,186). The module shape is similar to that seen in the Blue Module for Developmental Stage Two seen in Figure 3.6.4 which has the same separation in the modules connected by one edge where one of the nodes connected is *FOXP1*. As *FOXP1* is a transcription factor it is believable that it could regulate the sub module. The HG for this module was *GRIN2A* which encodes a subunit of NMDA receptor which is a transmembrane ion channel which is critical for normal development, synaptic plasticity and memory and it has been linked to epilepsy and neuropsychiatric disorders (177–179,187). *GRIN2A* is also the HG for the blue module in Developmental Stage Three, the modules in Figure 3.6.10 and Figure 3.6.8 are well conserved, this could mean that a lot of the processes that are going on during childhood are continued through to adolescence.

Figure 3.6.11 displays the Turquoise Modules for Developmental Stage Four. It is a large module with 65 genes. Of the edges that could exist 15.6% of them actually exist so the density is low. The HG for this module is *CA8* which is a member of the carbonic anhydrase family, it is expressed in Purkinje cells in the cerebellum. It is a IP₃R1 inhibitor which regulates calcium levels which helps key cellular processes (148,188).

CA8 and *GRIN2A* the HG's for this developmental stage have both been linked with ID and epilepsies and ASD.

4.3.5 Developmental Stage Five – Adult (>19 years)

Figure 3.6.12 shows the Blue Module for Developmental Stage Five, *FOXP1* and *SEMA6D* have broken off into their own module. *FOXP1* is a member of the fox family and is a transcriptional factor which when mutation or deletions occur can lead to ID, ASD, psychiatric features and language impairment (185,186,189,190). *SEMA6D* is a member of the semaphorin family which secrete transmembrane and glycosyphosphatidylinositol proteins which are implicated in immune response, neurogenesis, and heart development (191). *SEMA6D* regulates T cells during primary immune response and it has been linked to gastric cancers and tumour angiogenesis (191,192). The HG for this module is *CA8*. *CA8* is also the HG for the turquoise module for Developmental Stage Two, Turquoise Module for Developmental Stage Four, and the Blue Module for Developmental Stage Five. The modules in Figure 3.6.7 and 3.6.12 are most similar, over half of schizophrenia-associated genes overlap between the modules.

Figure 3.6.13 displays the brown for Developmental Stage Five. It is a large module with 58 genes. The HG is *SATB2* which encodes a protein which is involved in transcription regulation and chromatin remodelling (131). *SATB2* is a transcription factor which regulates neocortical circuitry and organisation (131). *SATB2* has been shown to cause *SATB2*-associated syndrome and developmental delays (131,132).

Figure 3.6.14 displays the Yellow Module for Developmental Stage Five, most of the genes have a large module membership shown by the size of their nodes, except *RERE*, *MGAT3* and *IGSF9B*. The HG for this module is *DPYD* which encodes an enzyme which is the rate limiting enzyme for catabolism of uracil and thymidine (193). In individuals with undetectable enzyme activity as a result of homozygous null mutations in *DPYD* have been seen in some individuals with neurodevelopmental disorders such as autism (194).

CA8, *SATB2* and *DPYD* the HG's for Developmental Stage Five are each been strongly linked to developmental delay and neurodevelopmental disorders.

4.3.6 Recurring Hub Genes Across Developmental Stages

Of the 316 schizophrenia-associated genes 24 HGs were identified for the 27 modules. *GRIN2A* was a HG for two modules and *CA8* was a HG for three. *GRIN2A* encodes a subunit of the NMDA receptor which is a transmembrane ligand gated ion channel which has a vital role in normal neuronal development, memory, and, synaptic plasticity (195). NMDA receptors are neurotransmitter ion gated channels which regulate synaptic function in the CNS, they contain two glycine binding NR1 subunits and two glutamate binding NR2 subunits (177). A glutamate binding subunit GluN2 is produced by *GRIN2A*, *GRIN2B*, *GRIN2C* and *GRIN2D* (179,187). GRIN genes have associations with developmental delay and early onset epileptic encephalopathies (DEE) with comorbidities like aphasia, ID, ASD and schizophrenia. It is assumed that *GRIN2A* variants, which are linked to epilepsies, are owing to haploinsufficiency or altered receptor properties (187). *CA8* is still considered a member of the carbonic anhydrase family in spite of the fact it cannot catalyse the hydration of carbon dioxide because of its lack of zinc binding histidine residues (149). It is mostly expressed in Purkinje cells found in the cerebellum but is found in other tissues and has been conserved across tissues which likely means *CA8* has an important function. It is an allosteric inhibitor of IP₃R1 which regulates calcium levels in the cell which is linked to key cellular processes such as release of neurotransmitters, nerve processes, neuronal excitability, and mitochondrial energy production (149,150). *CA8* has been linked to epilepsy, MERRF disease, spinal cerebellar ataxia (SCA), mild mental retardation, chronic pain, Alzheimer's and quadrupedal gait (148,150).

Pardiñas et al. study applied Summary-data-based Mendelians Randomisation analysis to the schizophrenia-associated gene with dorsetal prefrontal cortex expression quantitative trait locus (eQTL) using CommonMind Consortium. The aim of this study was to uncover variant which could be casually linked to expression changes in specific genes. They applied a threshold of 0.05 which highlighted colocalised signals due to a single causal variant. From this they discovered 22 candidate variants at 19 loci with an FDR P <0.05. *EMB* a HG in the Green

Module in Developmental Stage Two and *TRANK1* a HG in the Brown Module of Developmental Stage Three were identified. These genes being HGs suggests that the SNP is causing a change in expression and that these genes would be a good candidate for further research as they may regulate other genes in the region.

4.4 Gene Ontologies

On each of the modules calculated by WGCNA, anRichment was run. In order to determine which GOs were most important in each module, a p-value was calculated and GO's which were lower than the threshold of 0.05 after Bonferroni correction was applied were kept. As anRichment produced many significant GO's, GO parent terms were excluded to focus on more specific pathways. The complete ontologies calculated from anRichment in R including the genes involved can be found in the appendix from Table 3 to Table 16.

Across the developmental stages anRichment has highlighted the immune system and inflammation, specifically in relation to macrophage proliferation (GO:00120041). Inflammation in the CNS is facilitated by astrocytes, microglial cells, proinflammatory cytokines, invading immune cells which includes macrophages, monocytes, and T or B lymphocytes. For appropriate function, a well-regulated inflammatory response is essential, but uncontrolled inflammation caused by infectious agents, genetics or physical trauma can be detrimental (196) . The macrophages of the brain are called microglia and they play a crucial role in the innate immunity of the CNS and represent up to 10% of total brain cells, but their cell density depends on the area of the brain (197,198). Microglial cells are derived from the yolk sac progenitors during embryogenesis and migrate throughout the CNS, they are maintained through adulthood by self-renewal and rapid cell turnover (198,199). Microglia are involved in synaptic organisation, phagocytosis of apoptotic cells during development, maintenance of neuronal excitability, trophic neuronal support in the developing brain and brain protection and repair (199). In post-mortem cortical tissue of patients with schizophrenia,

the synapse density is reduced. This excessive pruning reflects abnormalities in synaptic structures and microglia like cells (200). A lot of this synaptic pruning occurs during adolescence, this is when schizophrenia symptoms occur (200). *In utero* MIA is a risk factor for neurodevelopmental disorders like schizophrenia. O'Loughlin et al. administered lipopolysaccharide to mice on embryonic day 12 to mice to induce MIA. This induced a pro-inflammatory cytokine profile which continued in the amygdala to early adulthood. These alterations in the foetal brain elicited by MIA can lead to alterations to microglia (54). Pre and perinatal activation of the immune system can increase the immune system's sensitivity throughout life (18). Autoimmune disorders and severe infections are linked to schizophrenia risk (201).

Cellular response to catecholamine stimulus was highlighted by enrichment in the at least one module per Developmental Stage (GO:0071870). Catecholamines, which are neurotransmitters in the CNS and peripheral nervous system, include dopamine, norepinephrine and adrenaline (127) . Catecholamine signalling underlies the mesocorticolimbic system and affects executive function and cognition (203). Catecholamine signalling pathways are pharmacological therapy targets for patients with neuropsychiatric disorders because of their relationship with affective, executive and cognitive functions (203). Catecholamines have versatile functions as slow acting neurotransmitters in synaptic neurotransmission, and controlling the effects of fast-acting neurotransmitters (202). Dopamine has been clearly linked with schizophrenia for many years and dopamine receptor antagonists continue to be the leading therapy for schizophrenia (5). Dopamine neurotransmission is altered in a number of neural pathways in schizophrenia (204), these alterations include hyperactive dopaminergic transmission in the striatum, hippocampus and mesolimbic areas and hypoactive transmission in the PFC of patients with schizophrenia (204). Dopamine displays regulatory effects on immune response which depends on dopamine concentration, sub-type of receptors, time of exposure, type of immune cell and immune cell activation (31). This has been shown to have an effect on cognitive functions

(204). Immune cells in particular T-cells, microglial and peripheral monocytes collaborate with the CNS and have cognitive and behavioural function which are seen to be altered in schizophrenia (31,204). Dopamine has been seen to influence the activity of these immune cells since they express dopamine receptors (204). Changes in dopamine concentration and/or receptors in T cells are thought to be the cause of abnormal immune functions in people with schizophrenia and Parkinson's (205). Low levels of dopamine neurotransmission, as well as serotonin and glutamatergic neurotransmission, are seen in people with schizophrenia. These are connected to low levels of neuroinflammation, which has been hypothesised to be the reason for CNS volume loss and low levels of microglial activation in schizophrenia patients in neuroimaging studies (18). Renalase is thought to metabolise dopamine. In a study conducted by Catak et al. which used thirty-three schizophrenia patients it was found that the levels of renalase in these patients was significantly lower than the control group. This could be a potential biomarker for schizophrenia (206).

In Developmental Stage One, chloride ion binding was highlighted as being enriched in the Brown Module (GO:001404). Cl⁻ levels in the CNS are determined by sodium-potassium chloride cotransporters *NKCC1* (Na⁺-K⁺-Cl⁻) and *KCC2* (K⁺-Cl⁻), cation ion transporters which are found in glial cells and neurons (207). GABA and glycine are dependent on the levels of Cl⁻ (208). Neuronal chloride regulation is hugely important component of GABAergic inhibition during and beyond development, in schizophrenia cognitive deficits have been linked to alterations in GABAergic neurotransmission (209). A study which indicated that GABA signalling was altered in schizophrenia patients was performed by Hyde et al. who found a 27% decrease in *KCC2* mRNA in the hippocampus (210). Another study found that expression levels of *NKCC1* and *KCC2*, *WNK3* and *OXSR1* were upregulated in people with schizophrenia with reduced chloride extrusion which altered GABAergic neurotransmission (207).

4.5 Limitations of study

There were several limitations of this study. Firstly, ABA's BrainSpan resource developed in 2014 uses healthy brains for each developmental stage. ABA's BrainSpan contains gene expression data across 16 brain regions from 8 post conception to 40 years, covering the complete development process (79). The atlas contains next generation RNA sequencing data which collected 579 tissue samples from thirty neurologically unremarkable brains over five developmental stages (80). A limitation to the BrainSpan resource is that the brain is divided into just 16 regions. In addition, ~80% of participants had transcriptomes missing from at least one brain region (79). To make this study more reliable there would need to be larger sample size. The genotype tissue expression (GTEx) project was created to enable the study of human genetic regulation and variation of gene expression in multiple tissues (211). The GTEx consortium collected 14,787 transcriptomes from 948 patients including 13 brain regions (79). GTEx does not have a developing brain data. Single cell sequencing (scRNASeq) allows for the dissection of gene expression at single cell resolution, using scRNASeq can lead to findings in cell expression alterations and dynamics (212). This revolutionary tool if applied to schizophrenia could uncover the uniqueness of each brain cell at microscopic resolution (213).

The schizophrenia-associated gene set which was used was taken from a study performed by Pardiñas et al. (66). This study used a previously completed study by the PGC and used their results as a training set to create risk profile scores to identify SNPs at a high confidence (66). This study is the largest schizophrenia GWAS to date with 40,675 cases the study size could be larger and could produce more reproducible SNP's. SNPs identified and mapped and if the locus had no overlapping genes it was mapped to the closest gene within 500 KB radius (66). The Pardiñas study mapped to 747 genes, of these genes only 316 genes were available on ABA's BrainSpan transcriptome atlas. BrainSpan does not contain any open reading frame, locus, or

chromosome location data. This is an enormous limitation as this excludes most of the genes leaving this study with a modest gene set.

K-means as a method has several limitations, firstly because the user has to decide on the number of clusters k before beginning the analysis. Secondly, k-means has a bias of creating modules of a similar size which may not accurately represent the group (90). Lastly, k-means centroids are immensely effected by outliers and can give the outliers disproportionate importance (89,90).

A limitation of this study is that there was no clinical data which could be applied to the WGCNA data. Calculating gene trait significance also allows for the gene expression to be linked to biologically relevant traits. other studies which had this extra data were able to determine gene significance and can set a threshold where if a gene has a module membership and gene significance over a certain threshold, they can be considered HG for a module. In this study module membership was calculated by determining the node distance from the central ME. The MM calculated measured by Pearson correlation between ME and gene expression should only contain genes with the highest correlation but one study concluded that 25% be a better fit in other modules (214). Usually WGCNA analysis uses much more data , more genes could mean more than one HG for a module can give a better insight of the critical underlying pathways within a module.

Gene Ontology enrichment analysis is a powerful tool to give some insight into underlying mechanisms in a gene list. Although it can give some understanding the gene ontology databases are always growing because they are incomplete. As can be seen from the results the gene ontologies provided by anRichment more general than specific.

4.6 Future Directions

To get a more comprehensive understanding of the brain changes which result in severe psychiatric diseases and schizophrenia in particular it is important to have a clear understanding

of mechanism which occur in normal brain development. Once there is a clear understanding of normal brain development it will be easier to observe the changes that occur in conditions like schizophrenia. At present it is difficult to full affected brains at any stage, but as there are no biomarkers or way to predict who will develop schizophrenia studying brains before development becomes more difficult. To complete a project like ABA's BrainSpan you would have to use brains from people who are ultra-high risk but if you will not be able to confirm if they had lived that schizophrenia would develop. A longitudinal study where participants who are more likely to develop schizophrenia than the general population could be studied, the limitation for this kind of study is the inability to study arguably the most important stage during development *in utero*. Thus, incorporation of gene, epigenome and gene expression data and machine learning approaches will allow for a better understanding of abnormal brain development.

GWA studies for schizophrenia have already proved useful, highlighting genes of interest and underlying biological pathways which can be used for developing novel therapeutics. Like ABA's BrainSpan resource scRNA seq could be applied to patients and controls to get robust results at a cellular level. To obtain robust results there would need to be a large cohort in the study.

As Schizophrenia is a heterogeneous condition, a route for future study of schizophrenia is dividing schizophrenia into subgroups where symptoms are more homogeneous. Where the symptoms are more homogeneous it would be likely they have the same underlying mechanism involved.

Gene Ontology databases are ever growing with the continual studies uncovering molecular functions, cellular locations, and biological processes. With more studies utilising WGS there will be a more absolute understanding of the genome and what it does and continuously expand the scope of GO. As can be seen in the GO tables for this study the GOs are not over specific, insight into the definite mechanisms which are different in schizophrenia patients could provide pharmaceutical companies with novel therapeutic targets.

Functional work on the HG's of interest at each of the stages could be performed, looking at risk variants and seeing if they have been fine mapped. These genes could be studied in knock out mice to see the repercussions of not being there. It would be particularly interesting to study EMB and TRANK1 which were highlighted as being interesting in this study and Pardinas et al. In the future a method similar to this could be applied to other neuropsychiatric disorders including ASD, MD and BPD. Studying conditions like these over time could assistance in more accurate prediction of whom will develop, how severe the symptoms could be and which treatments could be effective.

4.7 Conclusion

Schizophrenia is still a long way from being fully understood. From the schizophrenia-associated genes which were highlighted by Pardiñas and available for BrainSpan our gene list contained 316 genes. Once GO enrichment was applied to the modules produced by WGCNA, it was clear that macrophage proliferation and catecholamine dysfunction were important mechanisms underlying schizophrenia in each developmental stage. These results are mirrored by a lot of previous studies which have been conducted and found immune system and catecholamine most notably dopamine. There is interaction between the immune system and dopamine and evidence for this is co-morbidities of schizophrenia and autoimmune diseases. Researching them together may give additional insight to their interdependency. In order to more accurately pinpoint central ontologies for each stage it is important that a study like this is done comparing controls and patients with schizophrenia. Studying them and looking at immune and catecholamine processes specifically throughout development could produce novel therapeutics with better efficacy and less severe side effects.

5.0 Bibliography

1. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological Insights From 108 Schizophrenia-Associated Genetic Loci. *Nature* [Internet]. 2014 [cited 2018 Nov 8];511(7510):421–7. Available from: http://www.nature.com/authors/editorial_policies/license.html#terms
2. Stilo SA, Murray RM. Non-Genetic Factors in Schizophrenia. *Curr Psychiatry Rep.* 2019;21(100).
3. Pugliese V, Bruni A, Carbone EA, Calabro G, Cerminara G, Sampogna G, et al. Maternal stress, prenatal medical illnesses and obstetric complications: Risk factors for schizophrenia spectrum disorder, bipolar disorder and major depressive disorder. *Psychiatry Res.* 2019;271:23–30.
4. Haller CS, Padmanabhan JL, Lizano P, Torous J, Keshavan M. Recent advances in understanding schizophrenia. *F1000Prime Rep.* 2014;6(57):1–11.
5. Amanda Krogmann, Luisa Peters, Laura von Hardenberg, Katja Bödeker, Viktor B. Nöhles and CUC. Keeping up with the therapeutic advances in schizophrenia: a review of novel and emerging pharmacological entities Amanda. *CNS Spectr.* 2019;24:41–68.
6. Carbon M, Correll CU. Thinking and acting beyond the positive: The role of the cognitive and negative symptoms in schizophrenia. *CNS Spectr.* 2014;19(S1):35–53.
7. Gobira PH, Ropke J, Aguiar DC, Crippa JAS, Moreira FA. Animal models for predicting the efficacy and side effects of antipsychotic drugs. *Rev Bras Psiquiatr.* 2013;35:132–9.
8. Owen MJ, Sawa A, Mortensen PB. Schizophrenia. *Lancet.* 2016;388(10039):86–97.
9. Avramopoulos D. Recent Advances in the Genetics of Schizophrenia. *Mol Neuropsychiatry* [Internet]. 2018 [cited 2019 Sep 3];4:35–51. Available from: www.karger.com/mnp
10. Insel TR. Rethinking schizophrenia. *Nature.* 2010. p. 187–93.
11. Stępnicki P, Kondej M, Kaczor AA. Current concepts and treatments of schizophrenia. *Molecules.* 2018;23(2087).
12. Strous RD, Shoenfeld Y. Schizophrenia, autoimmunity and immune system dysregulation: A comprehensive model updated and revisited. *J Autoimmun.* 2006;27:71–80.
13. Ota VK, Moretti PN, Santoro ML, Talarico F, Spindola LM, Xavier G, et al. Gene expression over the course of schizophrenia: from clinical high-risk for psychosis to chronic stages. *npj Schizophr.* 2019;5(5).
14. Gejman P V., Sanders AR, Duan J. The role of genetics in the etiology of schizophrenia. *Psychiatr Clin North Am.* 2010;33(1):35–66.
15. Lally J, MacCabe JH. Antipsychotic medication in schizophrenia: A review. *Br Med Bull.* 2015;114:169–79.
16. Muller N, Schwarz M. The Role of Immune System in Schizophrenia. *Curr Immunol Rev.* 2010;6(3):213–20.
17. Van Kesteren CFMG, Gremmels H, De Witte LD, Hol EM, Van Gool AR, Falkai PG, et al. Immune involvement in the pathogenesis of schizophrenia: A meta-analysis on postmortem brain studies. *Transl Psychiatry* [Internet]. 2017;7(3). Available from: <http://dx.doi.org/10.1038/tp.2017.4>
18. Müller N. Inflammation in schizophrenia: Pathogenetic aspects and therapeutic considerations. *Schizophr Bull.* 2018;44(5):973–82.
19. Müller N, Weidinger E, Leitner B, Schwarz MJ. The role of inflammation in schizophrenia. *Frontiers in Neuroscience.* 2015.
20. Horváth S, Mirnics K. Immune system disturbances in schizophrenia. *Biol Psychiatry* [Internet]. 2014;75(4):316–23. Available from: <http://dx.doi.org/10.1016/j.biopsych.2013.06.010>
21. Benros ME, Nielsen PR, Nordentoft M, Eaton WW, Dalton SO, Mortensen PB. Autoimmune diseases and severe infections as risk factors for schizophrenia: A 30-year population-based register study. *Am J Psychiatry.* 2011;168(12):1303–10.
22. Khandaker GM, Cousins L, Deakin J, Lennox BR, Yolken R, Jones PB. Inflammation and immunity in schizophrenia: Implications for pathophysiology and treatment. *The Lancet Psychiatry.*

- 2015;2(3):258–70.
- 23. Miller BJ, Buckley P, Seabolt W, Mellor A, Kirkpatrick B, Brian J. Miller, Peter Buckley, Wesley Seabolt, Andrew Mellor and BK. Meta-Analysis of Cytokine Alterations in Schizophrenia: Clinical Status and Antipsychotic Effects Brian. *Biol Psychiatry*. 2011;70(7):663–71.
 - 24. Song XQ, Lv LX, Li WQ, Hao YH, Zhao JP. The Interaction of Nuclear Factor-Kappa B and Cytokines Is Associated with Schizophrenia. *Biol Psychiatry*. 2009;65:481–8.
 - 25. Dinan TG, Borre YE, Cryan JF. Genomics of schizophrenia: Time to consider the gut microbiome? *Mol Psychiatry*. 2014;19:1252–7.
 - 26. Zheng P, Zeng B, Liu M, Chen J, Pan J, Han Y, et al. The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Sci Adv*. 2019;5.
 - 27. Kelly JR, Minuto C, Cryan JF, Clarke G, Dinan TG. The role of the gut microbiome in the development of schizophrenia. *Schizophr Res*. 2020;
 - 28. Magdalon J, Mansur F, Teles e Silva AL, de Goes VA, Reiner O, Sertié AL. Complement System in Brain Architecture and Neurodevelopmental Disorders. *Front Neurosci*. 2020;14(23).
 - 29. Owen MJ, O'donovan MC. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. *World Psychiatry*. 2017;16:227–35.
 - 30. Gupta S, Kulhara P. What is schizophrenia: A neurodevelopmental or neurodegenerative disorder or a combination of both A critical analysis. *Indian J Psychiatry*. 2010;52(1):21–7.
 - 31. S.M Matt P. G. Where is Dopamine and How Do Immune Cells See It?: Dopamine-Mediated Immune Cell Function in Health and Disease. *J Neuroimmune Pharmacol*. 2020;15(1):114–64.
 - 32. Howes O, McCutcheon R, Stone J. Glutamate and dopamine in schizophrenia: An update for the 21st century. *J Psychopharmacol*. 2015;29(2):97–115.
 - 33. Kesby JP, Eyles DW, McGrath JJ, Scott JG. Dopamine, psychosis and schizophrenia: The widening gap between basic and clinical neuroscience. *Transl Psychiatry*. 2018;8(30).
 - 34. Murray RM, Bhavsar V, Tripoli G, Howes O. 30 Years on: How the Neurodevelopmental Hypothesis of Schizophrenia Morphed into the Developmental Risk Factor Model of Psychosis. *Schizophr Bull*. 2017;43(6):1190–6.
 - 35. Frank TVM and MJ. An Integrative Perspective on the Role of Dopamine in Schizophrenia. *Biol Psychiatry*. 2017;176(1):52–66.
 - 36. Demjaha A, Egerton A, Murray RM, Kapur S, Howes OD, Stone JM, et al. Antipsychotic treatment resistance in schizophrenia associated with elevated glutamate levels but normal dopamine function. *Biol Psychiatry*. 2014;75(5).
 - 37. Yang AC, Tsai SJ. New targets for schizophrenia treatment beyond the dopamine hypothesis. *Int J Mol Sci*. 2017;18(1689).
 - 38. Brisch R, Saniotis A, Wolf R, Bielau H, Bernstein HG, Steiner J, et al. The role of dopamine in schizophrenia from a neurobiological and evolutionary perspective: Old fashioned, but still in vogue. *Front Psychiatry*. 2014;5(47).
 - 39. Uno Y, Coyle JT. Glutamate hypothesis in schizophrenia. *Psychiatry Clin Neurosci*. 2019;73:204–15.
 - 40. Stahl SM. Beyond the dopamine hypothesis of schizophrenia to three neural networks of psychosis: Dopamine, serotonin, and glutamate. *CNS Spectr*. 2018;23(3):187–91.
 - 41. Bustillo JR, Chen H, Jones T, Lemke N, Abbott C, Qualls C, et al. Increased glutamine in patients undergoing long-term treatment for schizophrenia: A proton magnetic resonance spectroscopy study at 3 T. *JAMA Psychiatry*. 2014;71(3):265–72.
 - 42. Aghajanian GK, Marek GJ. Serotonin model of schizophrenia: emerging role of glutamate mechanisms. In: *Brain Research Reviews*. 2000. p. 302–12.
 - 43. Sánchez-Blázquez P, Rodríguez-Muñoz M, Garzón J. The cannabinoid receptor 1 associates with NMDA receptors to produce glutamatergic hypofunction: Implications in psychosis and schizophrenia. *Front Pharmacol*. 2014;4(169).
 - 44. Marsman A, Van Den Heuvel MP, Klomp DWJ, Kahn RS, Luijten PR, Hulshoff Pol HE. Glutamate in schizophrenia: A focused review and meta-analysis of 1H-MRS studies. *Schizophr Bull*. 2013;39(1):120–9.

45. Egerton A, Modinos G, Ferrera D, McGuire P. Neuroimaging studies of GABA in schizophrenia: A systematic review with meta-analysis. *Transl Psychiatry*. 2017;7:1–10.
46. Cohen SM, Tsien RW, Goff DC, Halassa MM. The impact of NMDA receptor hypofunction on GABAergic neurons in the pathophysiology of schizophrenia. *Schizophr Res*. 2015;167(1–3):98–107.
47. Glausier JR, Lewis DA. GABA and schizophrenia: Where we stand and where we need to go. *Schizophr Res*. 2017;181:2–3.
48. Birnbaum R, Weinberger DR. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat Rev Neurosci*. 2017;18:727–40.
49. Quednow BB, Geyer MA, Halberstadt AL. Serotonin and Schizophrenia. In: *Handbook of Behavioral Neuroscience*. 2010. p. 585–620.
50. Abi-Dargham A. Alterations of Serotonin Transmission in Schizophrenia. *Int Rev Neurobiol*. 2007;78:133–64.
51. Sumiyoshi T, Kunugi H, Nakagome K. Serotonin and dopamine receptors in motivational and cognitive disturbances of schizophrenia. *Front Neurosci*. 2014;8(395).
52. Van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. *Nature* [Internet]. 2010;468(7321):203–12. Available from: <http://dx.doi.org/10.1038/nature09563>
53. Janoutová J, Janáčková P, Šerý O, Zeman T, Ambroz P, Kovalová M, et al. Epidemiology and risk factors of schizophrenia. *Neuroendocr Lett* [Internet]. 2016;37(1):26994378–370116. Available from: www.nel.edu
54. O'Loughlin E, Pakan JMP, Yilmazer-Hanke D, McDermott KW. Acute in utero exposure to lipopolysaccharide induces inflammation in the pre- and postnatal brain and alters the glial cytoarchitecture in the developing amygdala. *J Neuroinflammation*. 2017;14(1):1–12.
55. Hollins SL, Cairns MJ. MicroRNA: Small RNA mediators of the brains genomic response to environmental stress. *Prog Neurobiol* [Internet]. 2016;143:61–81. Available from: <http://dx.doi.org/10.1016/j.pneurobio.2016.06.005>
56. Brown AS. The environment and susceptibility to schizophrenia. *Prog Neurobiol*. 2011;93(1):23–58.
57. Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. *Schizophr Res* [Internet]. 2001;49(1):1–52. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2812015/>
58. Dietsche B, Kircher T, Falkenberg I. Structural brain changes in schizophrenia at different stages of the illness: A selective review of longitudinal magnetic resonance imaging studies. *Aust N Z J Psychiatry*. 2017;1–9.
59. Hajima S V., Van Haren N, Cahn W, Koolschijn PCMP, Hulshoff Pol HE, Kahn RS. Brain volumes in schizophrenia: A meta-analysis in over 18 000 subjects. *Schizophr Bull*. 2013;39(5):1129–38.
60. Fusar-Poli P, Smieskova R, Kempton MJ, Ho BC, Andreasen NC, Borgwardt S. Progressive brain changes in schizophrenia related to antipsychotic treatment? A meta-analysis of longitudinal MRI studies. *Neurosci Biobehav Rev*. 2013;37(8):1680–91.
61. Scheepers FE, De Wied CCG, Pol HEH, Van De Flier W, Van Der Linden JA, Kahn RS. The effect of clozapine on caudate nucleus volume in schizophrenic patients previously treated with typical antipsychotics. *Neuropsychopharmacology*. 2001;24(1):47–54.
62. Hilker R, Helenius D, Fagerlund B, Skytte A, Christensen K, Werge TM, et al. Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. *Biol Psychiatry* [Internet]. 2018;83(6):492–8. Available from: <http://dx.doi.org/10.1016/j.biopsych.2017.08.017>
63. Schwab SG, Wildenauer DB. Genetics of psychiatric disorders in the GWAS era: An update on schizophrenia. *Eur Arch Psychiatry Clin Neurosci*. 2013;263:147–54.
64. Henriksen MG, Nordgaard J, Jansson LB. Genetics of schizophrenia: Overview of methods, findings and limitations. *Front Hum Neurosci*. 2017;11(322).
65. Duncan EL, Brown MA. Genome-Wide Association Studies. In: *Genetics of Bone Biology and Skeletal Disease: Second Edition*. 2017.
66. Antonio F, Pardiñas1, Peter Holmans1, Andrew J, Pocklington1, Valentina Escott- Price1, Stephan Ripke2, 3, Noa Carrera1, Sophie E, Legge1, Sophie Bishop1, Darren Cameron1, Marian L.

- Hamshere1, Jun Han1, Leon Hubbard1, Amy Lynham1, Kiran Mantripragada1 ER. Common schizophrenia alleles are enriched in mutation intolerant genes and in region under strong background selection. *Nat Genet* [Internet]. 2018;50(3):381–9. Available from: file:///C:/Users/Carla Carolina/Desktop/Artigos para acrescentar na qualificação/The impact of birth weight on cardiovascular disease risk in the.pdf
67. Skene NG, Bryois J, Bakken TE, Breen G, James J, Gaspar HA, et al. Genetic identification of brain cell types underlying schizophrenia. 2019;50(6):825–33.
68. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol*. 2012;8(12):1–10.
69. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*. 2017. p. 5–22.
70. van Mierlo HC, Schot A, Boks MPM, de Witte LD. The association between schizophrenia and the immune system: Review of the evidence from unbiased ‘omic-studies.’ *Schizophr Res* [Internet]. 2020;217:114–23. Available from: <https://doi.org/10.1016/j.schres.2019.05.028>
71. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13.
72. Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, et al. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med Sci Monit Basic Res*. 2014;20:138–41.
73. Ball S, Gilbert TL, Overly CC. The Human Brain Online: An Open Resource for Advancing Brain Research. *PLoS Biol*. 2012;10(12).
74. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* [Internet]. 2012 Sep 20 [cited 2018 Dec 1];489(7416):391–9. Available from: <http://www.nature.com/articles/nature11405>
75. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: An integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res*. 2013;41(database issue).
76. Arnatkevičiūtė A, Fulcher BD, Fornito A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage*. 2019;189:353–67.
77. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508:199–206.
78. Vasung L, Abaci Turk E, Ferradal SL, Sutin J, Stout JN, Ahtam B, et al. Exploring early human brain development with structural and physiological neuroimaging. *Neuroimage*. 2019;187:226–54.
79. Pei G, Wang YY, Simon LM, Dai Y, Zhao Z, Jia P. Gene expression imputation and cell-type deconvolution in human brain with spatiotemporal precision and its implications for brain-related disorders. *Genome Res*. 2021;1(1):146–58.
80. Mahfouz A, Ziats MN, Rennert OM, Lelieveldt BPF, Reinders MJT. Shared Pathways Among Autism Candidate Genes Determined by Co-expression Network Analysis of the Developing Human Brain Transcriptome. *J Mol Neurosci*. 2015;57:580–94.
81. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012;489:391–9.
82. Negi SK, Guda C. Global gene expression profiling of healthy human brain and its application in studying neurological disorders OPEN. *Nat Reports* [Internet]. 2017 Dec 18 [cited 2018 Nov 8];7(897):1–12. Available from: www.nature.com/scientificreports/
83. McCarthy MJ, Liang S, Spadoni AD, Kelsoe JR, Simmons AN. Whole Brain Expression of Bipolar Disorder Associated Genes: Structural and Genetic Analyses. *PLoS One* [Internet]. 2014 [cited 2018 Dec 1];9(6):1–13. Available from: www.plosone.org
84. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa L da F, et al. Clustering algorithms: A comparative approach. *PLoS One*. 2019;14(1).
85. Jayanthi K, Mahesh C. A Study on machine learning methods and applications in genetics and genomics. *Int J Eng Technol*. 2018;7(1.7):201–4.
86. Saelens W, Cannoodt R, Saeys Y. A comprehensive evaluation of module detection methods for gene expression data. *Nat Commun*. 2018;9(1090):1–12.

87. Gibbons FD, Roth FP. Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res.* 2002;12:1574–81.
88. Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett.* 2010;31:651–66.
89. Shi N, Liu X, Guan Y. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In: 3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010. IEEE; 2010. p. 63–7.
90. Morissette L, Chartier S. The k-means clustering technique: General considerations and implementation in Mathematica. *Tutor Quant Methods Psychol.* 2013;9(1):15–24.
91. Charrad M, Ghazzali N, Boiteau V, Niknafs A. Nbclust: An R package for determining the relevant number of clusters in a data set. *J Stat Softw.* 2014;61(6):1–36.
92. Kassambara A, Mundt F. Factoextra: extract and visualize the results of multivariate data analyses [Computer software, package]. R package version. 2017.
93. van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform.* 2018;19(4):575–92.
94. Li J, Zhou D, Qiu W, Shi Y, Yang JJ, Chen S, et al. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Sci Rep.* 2018;8(622).
95. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(559).
96. Liu Y, Gu HY, Zhu J, Niu YM, Zhang C, Guo GL. Identification of Hub Genes and Key Pathways Associated With Bipolar Disorder Based on Weighted Gene Co-expression Network Analysis. *Front Physiol.* 2019;10:1–9.
97. Costantini G, Epskamp S, Borsboom D, Perugini M, Mõttus R, Waldorp LJ, et al. State of the aRt personality research: A tutorial on network analysis of personality data in R. *J Res Pers.* 2015;54:13–29.
98. Yu D, Kim M, Xiao G, Hwang TH. Review of Biological Network Data and Its Applications. *Genomics Inform.* 2013;11(4):200–10.
99. Sun K, Gonçalves JP, Larminie C, Pržulj N. Predicting disease associations via biological network analysis. *BMC Bioinformatics.* 2014;15(304):1–13.
100. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
101. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc.* 2007;2(10):2336–82.
102. Gang Su1, John H. Morris2, Barry Demchak3 and GDB. BIOLOGICAL NETWORK EXPLORATION WITH CYTOSCAPE 3. *Curr Protoc Bioinforma.* 2014;74(8).
103. Assenov Y, Ramírez F, Schelhorn SESE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics.* 2008;24(2):282–4.
104. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol.* 2011;
105. Carbon S, Dietze H, Lewis SE, Mungall CJ, Munoz-Torres MC, Basu S, et al. Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Res.* 2017;45(Database issue):D331-D338.
106. Blake JA, Christie KR, Dolan ME, Drabkin HJ, Hill DP, Ni L, et al. Gene ontology consortium: Going forward. *Nucleic Acids Res.* 2015;43(D1):1049–56.
107. Langfelder P. anRichment: Collections and annotation data for use with anRichmentMethods. 2020.
108. Carlson M. org.Hs.eg.db: Genome wide annotation for Human. 2019.
109. Kassambara A, Mundt F, Kassambara, A.; Mundt F. Factoextra: extract and visualize the results of multivariate data analyses. URL <http://www.sthda.com/english/rpkgs/factoextra> BugReports. 2017;
110. Barabási AL, Oltvai ZN. Network biology: Understanding the cell’s functional organization. *Nat Rev Genet.* 2004;5:101–13.
111. Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. Using graph theory to analyze biological networks. *BioData Min.* 2011;4(10).
112. Jacob R, Harikrishnan KP, Misra R, Ambika G. Measure for degree heterogeneity in complex

- networks and its application to recurrence network analysis. *R Soc Open Sci.* 2017;4(1).
113. Dong J, Horvath S. Understanding network concepts in modules. *BMC Syst Biol.* 2007;1(24).
 114. Zawerton, Ash; Mignot, Cyril; Sigafoos, Ashley; Blackburn, Patrick R.; Haseeb, Abdul; McWalter, Kirsty; Ichikawa, Shoji; Nava, Caroline; Keren, Boris; Charles, Perrine; Marey, Isabelle; Tabet, Anne-Claude; Levy, Jonathan; Perrin, Laurence; Hartmann, Andre C. Widening of the genetic and clinical spectrum of Lamb-Shaffer syndrome, a neurodevelopmental disorder due to SOX5 haploinsufficiency. *Genet Med.* 2019;
 115. Manfred Schartl1, 2, 3, Susanne Schories1, Yuko Wakamatsu1, Yusuke Nagao4, Hisashi Hashimoto4, Chloé Bertin5, Brigitte Mourot5, Cornelia Schmidt1, Dagmar Wilhelm6, Lazaro Centanin7 YG and AH. Sox5 is involved in germ-cell regulation and sex determination in medaka following co-option of nested transposable elements. *BMC Biol.* 2018;16(16).
 116. Tae-Gul Lee1, 2 , Eun-Hui Jeong1, Seo Yun Kim1, Hye-Ryoun Kim1 HK, Kim C-H. Fhit, a tumor suppressor protein, induces autophagy via 14-3-3 τ in non-small cell lung cancer cells. *Oncotarget.* 2017;8(19):31923–37.
 117. Garneau AP, Slimani S, Tremblay LE, Fiola MJ, Marcoux AA, Isenring P. K+-Cl⁻ cotransporter 1 (KCC1): a housekeeping membrane protein that plays key supplemental roles in hematopoietic and cancer cells. *J Hematol Oncol [Internet].* 2019 Dec 11;12(1):74. Available from: <https://jhoonline.biomedcentral.com/articles/10.1186/s13045-019-0766-x>
 118. Fiona C. Brown, 1 Ashlee J. Conway, 1 Loretta Cerruti, 1 Janelle E. Collinge, 2 Catriona McLean, 3 James S. Wiley, 4 Ben T. Kile, 2 Stephen M. Jane, 1, 3, 5,* and David J. Curtis1iona C. Brown, 1 Ashlee J. Conway, 1 Loretta Cerruti, 1 Janelle E. Collinge, and DJC. Activation of the erythroid K-Cl cotransporter Kcc1 enhances sickle cell disease pathology in a humanized mouse model. *Blood.* 2015;126(26):2863–70.
 119. Kourosh Parham, M.D., Ph.D.1, George A. Kuchel, M.D.2, Janet E. McElhaney, M.D.4, and Laura Haynes PD. A relationship between blood levels of otolin-1 and vitamin D. *Otol Neurotol.* 2018;39(4):269–73.
 120. Murat Doğan , Mustafa Şahin YK. Otolin-1, as a Potential Marker for Inner Ear Trauma after Mastoidectomy. *J Int Adv Otol.* 2019;15(2):200–3.
 121. Yanyan Liu1 2 · Chuang Tong2 · Jun Cao2 · Maoming Xiong1. NEMP1 Promotes Tamoxifen Resistance in Breast Cancer Cells. *Biochem Genet.* 2019;57(6):813–26.
 122. Yonit Tsatskis1*, Robyn Rosenfeld1, 2*, Joel D. Pearson1*, Curtis Boswell2, 3*, Yi Qu1, Kyunga Kim1, 4, Lacramioara Fabian5, Ariz Mohammad6, Xian Wang7, 8, Michael I. Robson9, Karen Krchma10, Jun Wu10, João Gonçalves1, Didier Hodzic11, Shu Wu1, Daniel Pot HM. The NEMP family supports metazoan fertility and nuclear envelope stiffness. *Sci Adv Genet.* 2020;6.
 123. Hiroshi Mamada, Noriyuki Takahashi MT. Involvement of an inner nuclear membrane protein, Nemp1, in Xenopus neural development through an interaction with the chromatin protein BAF. *Dev Biol.* 2009;327:497–507.
 124. Takashi Hashimoto 1, Shohei Maekawa SM. IgLON cell adhesion molecules regulate synaptogenesis in hippocampal neurons. *Cell Biochem Functopm.* 2009;7:496–8.
 125. Lan Wen a, Lei Liu b, Linyan Tong a, Jinfang Li a, Keming Zhang a, Qinbin Zhang a CL a. NDRG4 prevents cerebral ischemia/reperfusion injury by inhibiting neuronal apoptosis. *Genes Dis.* 2019;6:448–54.
 126. Mitsuo Wada, Kayo Yukawa, Hiroyuki Ogasawara, ..., Takeshi Ohta, Eunyoung Lee TM. GPR52 accelerates fatty acid biosynthesis in a ligand-dependent manner in hepatocytes and in response to excessive fat intake in mice. *iScience.* 2021;24(102260).
 127. Hideyoshi Komatsu1*, Minoru Maruyama2., Shuuhei Yao3, Tokuyuki Shinohara4, Kensuke Sakuma2, Sachiko Imaichi4, Tomoko Chikatsu2, Kanako Kuniyeda5, Foo Kok Siu6, Lam Sock Peng6, Katherine Zhuo7, Lay Sock Mun7, Tan Min Han7, Yoshio Matsumoto3, Tadatoshi Has MM. Anatomical Transcriptome of G Protein-Coupled Receptors Leads to the Identification of a Novel Therapeutic Candidate GPR52 for Psychiatric Disorders. *PLoS One.* 2014;9(2).
 128. Lorena Benedetti1, 2 , Matteo Cereda1, 2 , LeeAnn Monteverde1, 2 , Nikita Desai1 2 and, Ciccarelli FD. Synthetic lethal interaction between the tumour suppressor STAG2 and its paralog STAG1. *Oncotarget.* 2017;8(23):37619–32.

129. Petra van der Lelij^{1,*}, Joseph A Newman^{2,*}, Simone Lieb^{3,*}, Julian Jude^{1,*}, Vittorio Katis², Thomas Hoffmann¹, Matthias Hinterndorfer¹, Gerd Bader³, Norbert Kraut³, Mark A Pearson³, Jan-Michael Peters^{1, 4}, Johannes Zuber^{1, 4}, Opher Gileadi², MP. STAG1 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. *Life Sci Alliance*. 2020;3(7).
130. Daphné Lehalle,^{1, 2} Anne-Laure Mosca-Boidron,^{3, 4} Amber Begtrup,⁵ Odile Boute-Benejean,⁶ Perrine Charles,⁷ Megan T Cho,⁵ Amanda Clarkson,⁸ Orrin Devinsky,⁹ Yannis Duffourd,⁴ Laurence Duplomb-Jego,^{3, 4} Bénédicte Gérard,¹⁰ Aurélia Jacquette,⁷ Pa⁴ Laurence Faivre¹. STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. *BMJ*. 2017;54(7):479–88.
131. Whitton L, Apostolova G, Rieder D, Dechant G, Rea S, Donohoe G, et al. Genes regulated by SATB2 during neurodevelopment contribute to schizophrenia and educational attainment. *PLoS Genet*. 2018;14(7).
132. Qian Y, Liu J, Yang Y, Chen M, Jin C, Chen P, et al. Paternal low-level mosaicism-caused SATB2-associated syndrome. *Front Genet*. 2019;
133. Schult, D., Ho'lsken MB, Schlaffer, S.-M. Siegel, S. Buslei, I. Kreitschmann-Andermahr , R. Fahlbusch R. Expression pattern of neuronal intermediate filament a-internexin in anterior pituitary gland and related tumors. *Pituitary*. 2015;18:465–473.
134. Yingjie Li^{1,*}, Liangliang Bai^{1,*}, Huichuan Yu¹, Du Cai¹, Xiaolin Wang¹, Baoyuan Huang¹, Shaoyong Peng¹, Meijin Huang^{1, 2}, Guangwen Cao³, Andrew M. Kaz^{4, 5, 6}, William M. Grady^{4, 6}, Jianping Wang^{1, 2}, and Yingjie Li^{1,*}, Liangliang Bai^{1,*}, Huichuan Yu¹, Du and YLL. Epigenetic Inactivation of α -Internexin Accelerates Microtubule Polymerization in Colorectal Cancer. *Cancer Res*. 2020;80(23):5203–15.
135. Whelan R, Clair DS, Mustard CJ, Hallford P, Wei J. Study of novel autoantibodies in schizophrenia. *Schizophr Bull*. 2018;44(6):1341–9.
136. Xiaoyu Xu^{1, 2,†} Zhao Zeng^{1,†} Li Huo^{1,†} Hong Liu^{1, 3} Yan Yu¹ Ling Zhang¹ Jiannong Cen¹ Huiying Qiu^{1, 3} Xiaowen Tang^{1, 3} Chengcheng Fu^{1, 3} Yue Han^{1, 2, 3} Miao Miao^{1, 3} Zhengming Jin^{1, 3} Changgeng Ruan^{1, 2, 3} Depei Wu^{1, 2, 3} Suning Chen^{1, 2, 3} Qinrong Wang^{1, 3}, Lingzhi Yan^{1, 3}. High expression of myocyte enhancer factor 2C predicts poor prognosis for adult acute myeloid leukaemia with normal karyotype. *BJH Corresp*. 2019;189(1):23–7.
137. Adam J Harrington¹, Catherine M Bridges², Stefano Berto³, Kayla Blankenship¹, Jennifer Y Cho², Ahlem Assali¹, Benjamin M Siemsen⁴, Hannah W Moore⁵, Evgeny Tsvetkov¹, Acadia Thielking¹, Genevieve Konopka³, David B Everman⁵, Michael D Scofield⁴ CWC⁶. MEF2C Hypofunction in Neuronal and Neuroimmune Populations Produces MEF2C Haploinsufficiency Syndrome-like Behaviors in Mic. *Biol Psychiatry*. 2020;15(88):488–99.
138. Adam J Harrington^{1, 2†}, Aram Raissi^{2†}, Kacey Rajkovich³ SB, Jaswinder Kumar² 4, Molinaro³ G, Raduazzo² J, Guo² Y, et al. MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *Elife*. 2016;5.
139. Melanie Richter¹, Nadeem Murtaza^{2, 3} Robin Scharrenberg¹ Sean H. White^{2, 3} Ole Johanns¹ Susan Walker^{4, 5}, Ryan K. C. Yuen^{4, 5} Birgit Schwanke¹ Bianca Bedürftig¹ Melad Henis^{1, 6} Sarah Scharf⁷ Vanessa Kraus⁷, Ronja Dörk⁷ Jakob Hellmann⁷ Zsuzsa Lindenmaie³ Froylan Calderon de Anda. Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. *Mol Psychiatry*. 2019;24:1329–50.
140. Chih-Yeu Fang¹, Tsung-Ching Lai^{2, 3}, Michael Hsiao^{1, 4} and Y-CC. The Diverse Roles of TAO Kinases in Health and Diseases. *J Mol Sci*. 2020;27(7463).
141. Yang Shao^{1, 2}, Jing Kong^{1†} HX, , Xiaoli Wu¹, YuePeng Cao¹, Weijian Li¹, Jing Han⁴, Dake Li¹ KX and JW. OPCML Methylation and the Risk of Ovarian Cancer: A Meta and Bioinformatics Analysis. *Front Cell Dev Biol*. 2021;9(570908).
142. Zhengrong Zhang, Maoqing Ye, Qiongwei Li, Rena Li, Jun Li DZ. The Schizophrenia Susceptibility Gene OPCML Regulates Spine Maturation and Cognitive Behaviors through Eph-Cofilin Signaling. *Cell Rep*. 2019;29:49–61.
143. Jane Antony^{1, 2, 3}, Elisa Zanini¹, Zoe Kelly¹, Tuan Zea Tan², Evdoxia Karali¹, Mohammad Alomary¹, Youngrock Jung¹, Katherine Nixon¹, Paula Cunneal¹, Christina Fotopoulou¹, Andrew Paterson¹, Sushmita Roy-Nawathe¹, Gordon B Mills⁴, Ruby Yun-Ju Huang^{2, 5}, Je & CR. The

- tumour suppressor OPCML promotes AXL inactivation by the phosphatase PTPRG in ovarian cancer. *EMBO J.* 2081;19(e45670).
144. Vougiouklakis1 T, Ryuji Hamamoto1 Y, Vassiliki N&, Saloura* 1. The NSD family of protein methyltransferases in human cancer. *Epigenomics.* 2015;7(5).
 145. Li LYLYQLC. Downregulation of NSD3 (WHSC1L1) inhibits cell proliferation and migration via ERK1/2 deactivation and decreasing CAPG expression in colorectal cancer cells. *Onco Targets Ther.* 2019;12:3933–43.
 146. Chunmei Wang, 1, 2 Qinlan Wang, 3 Xiaoqing Xu, 1 Bin Xie, 3 Yong Zhao, 4 Nan Li, 2 and Xuetao Cao1, 2 3. The methyltransferase NSD3 promotes antiviral innate immunity via direct lysine methylation of IRF3. *J Exp Med.* 2017;214(12):3597–610.
 147. Vassiliki Saloura1 , Theodore Vougiouklakis1, Makda Zewde1, Kiyotani1 Kazuma, Jae-Hyun Park1, Guimin Gao2, Theodore Garrison2, Mark Lingen3 , Yusuke Nakamura1, 4 RH. WHSC1L1 drives cell cycle progression through transcriptional regulation of CDC6 and CDK2 in squamous cell carcinoma of the head and neck. *Oncotarget.* 2016;7(27):42527–38.
 148. Paternoster L, Soblet J, Aeby A, De Tiège X, Goldman S, Yue WW, et al. Novel homozygous variant of carbonic anhydrase 8 gene expanding the phenotype of cerebellar ataxia, mental retardation, and disequilibrium syndrome subtype 3. *Am J Med Genet Part A.* 2020;97(3):516–20.
 149. Ma HL, Yu SJ, Chen J, Ding XF, Chen G, Liang Y, et al. CA8 promotes RCC proliferation and migration though its expression level is lower in tumor compared to adjacent normal tissue. *Biomed Pharmacother.* 2020;121(109578).
 150. Zhuang GZ, Keeler B, Grant J, Bianchi L, Fu ES, Zhang YP, et al. Carbonic anhydrase-8 regulates inflammatory pain by inhibiting the ITPR1-cytosolic free calcium pathway. *PLoS One.* 2015;10(3).
 151. Tom Hearn. ALMS1 and Alström syndrome: a recessive form of metabolic, neurosensory and cardiac deficits. *J Mol Med.* 2019;97:1–17.
 152. Shao-Yan Zhang, Chao Xuan, Yi Wang, Shao-Qiang Zhang, Hui Li, Guo-Wei He and Q-WT. Association between ALMS 1 variants and early-onset coronary artery disease: a case–control study in Chinese population. *Biosci Rep.* 2020;40(BSR20193637).
 153. Hemang M. Parikh1 2*, , Targ Elgzyri2, Amra Alibegovic3, Natalie Hiscock4, Ola Ekström2, Karl-Fredrik Eriksson2, Allan Vaag3, Leif C. Groop2, 5, Kristoffer Ström2 6† and Ola Hansson. Relationship between insulin sensitivity and gene expression in human skeletal muscle. *BMC Endocr Disord.* 2021;21(32).
 154. Jakub Tomasik, PhD1; Santiago G. Lago, PhD1; Javier Vázquez-Bourgon, MD P. Association of Insulin Resistance With Schizophrenia Polygenic Risk Score and Response to Antipsychotic Treatment. *JAMA Psychiatry.* 2019;76(8).
 155. Zaoa X, *, Jin Chengb *, Congle Shena , Guiwen Guana, Xiaoyu Fenga, Jun Zoua, Jing Zhangb TL, Huiling Zhenga, Ting Zhangc, Jie Wangd, Jia Liua, Deyao Lia, Fengmin Lu, Fuping You and XC. NFATc3 inhibits hepatocarcinogenesis and HBV replication via positively regulating RIG-I-mediated interferon transcription. *Oncoimmunology.* 2021;10(1).
 156. Katia Urso, Andrés Fernández, Patricia Velasco, Javier Cotrina, Belén de Andrés, Pilar Sánchez-Gómez , Aurelio Hernández-Laín, Sonsoles Hortelano JMR& EC. NFATc3 controls tumour growth by regulating proliferation and migration of human astrogloma cells. *Nat Sci Reports.* 2019;9(9361).
 157. Sridevi Venkatesan EKL. Chrna5 is essential for a rapid and protected response to optogenetic release of endogenous acetylcholine in prefrontal cortex. *bioRxiv.* 2020;40(38):7255–68.
 158. Thaka'a K. Al-Omoush1, Karem H. Alzoubi, Omar F. Khabour, Fawzi M. Alsheyab, Ahmed Abu-siniyeh, Nour A. Al-Sawalha, Fadia A. Mayyas, Caroline O. Cobb TE. The CHRNA5 Polymorphism (rs16969968) and its Association with Waterpipe Smoking Addiction among Jordanians. *Arab J Basix Appl Sci.* 2020;27(1):450–5.
 159. Thyme SB, Pieper LM, Li EH, Pandey S, Wang Y, Morris NS, et al. Phenotypic Landscape of Schizophrenia-Associated Genes Defines Candidates and Their Shared Functions. *Cell [Internet].* 2019;177(2):478–91. Available from: <https://doi.org/10.1016/j.cell.2019.01.048>
 160. Sant'Anna Feltrin A, Tahira AC, Simões SN, Brentani H, Martins DC. Assessment of complementarity of WGCNA and NERI results for identification of modules associated to schizophrenia spectrum disorders. *PLoS One [Internet].* 2019 [cited 2019 Jul 3];14(1):1–22.

- Available from: <https://doi.org/10.1371/journal.pone.0210431>
161. Sisodiya SM, Thompson PJ, Need A, Harris SE, Weale ME, Wilkie SE, et al. Genetic enhancement of cognition in a kindred with cone-rod dystrophy due to RIMS1 mutation. *J Med Genet.* 2007;44:373–80.
 162. Klasić M, Markulin D, Vojta A, Samaržija I, Biruš I, Dobrinić P, et al. Promoter methylation of the MGAT3 and BACH2 genes correlates with the composition of the immunoglobulin G glycome in inflammatory bowel disease. *Clin Epigenetics.* 2018;10(1):1–14.
 163. Xie Y, Huang D, Wei L, Luo XJ. Further evidence for the genetic association between CACNA1I and schizophrenia. *Hereditas.* 2018;155(16).
 164. Andrade A, Hope J, Allen A, Yorgan V, Lipscombe D, Pan JQ. A rare schizophrenia risk variant of CACNA1I disrupts CaV3.3 channel activity. *Sci Rep [Internet].* 2016;6:1–13. Available from: <http://dx.doi.org/10.1038/srep34233>
 165. Howell KR, Floyd K, Law AJ. PKB γ /AKT3 loss-of-function causes learning and memory deficits and deregulation of AKT/mTORC2 signaling: Relevance for schizophrenia. *PLoS One.* 2017;12(5):1–21.
 166. Ding L, Biswas S, Morton RE, Smith JD, Hay N, Byzova T V., et al. Akt3 deficiency in macrophages promotes foam cell formation and atherosclerosis in mice. *Cell Metab [Internet].* 2012;15(6):861–72. Available from: <http://dx.doi.org/10.1016/j.cmet.2012.04.020>
 167. Jiang NM, Cowan M, Moonah SN, Petri WA. The Impact of Systemic Inflammation on Neurodevelopment. *Trends Mol Med.* 2018;24(9):794–804.
 168. Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH, Hariri AR, et al. Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc Natl Acad Sci U S A.* 2004;101(34).
 169. Saini SM, Mancuso SG, Mostaid MS, Liu C, Pantelis C, Everall IP, et al. Meta-analysis supports GWAS-implicated link between GRM3 and schizophrenia risk. *Transl Psychiatry.* 2017;7(1196).
 170. Georg I, Díaz-Barreiro A, Morell M, Pey AL, Alarcón-Riquelme ME. BANK1 interacts with TRAF6 and MyD88 in innate immune signaling in B cells. *Cell Mol Immunol [Internet].* 2020;17(9):954–65. Available from: <http://dx.doi.org/10.1038/s41423-019-0254-9>
 171. Dieudé P, Wipff J, Guedj M, Ruiz B, Melchers I, Hachulla E, et al. BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. *Arthritis Rheum.* 2009;60(11):3447–54.
 172. Jiang SH, Athanasopoulos V, Ellyard JI, Chuah A, Cappello J, Cook A, et al. Functional rare and low frequency variants in BLK and BANK1 contribute to human lupus. *Nat Commun [Internet].* 2019;10(1):1–12. Available from: <http://dx.doi.org/10.1038/s41467-019-10242-9>
 173. Zhao Z, Zhang KN, Chai RC, Wang KY, Huang RY, Li GZ, et al. ADAMTSL4, a secreted glycoprotein, is a novel immune-related biomarker for primary glioblastoma multiforme. *Dis Markers.* 2019;1802620.
 174. Gabriel LAR, Wang LW, Bader H, Ho JC, Majors AK, Hollyfield JG, et al. ADAMTSL4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis. *Investig Ophthalmol Vis Sci.* 2012;53(1).
 175. Kobeissy FH, Hansen K, Neumann M, Fu S, Jin K, Liu J. Deciphering the role of Emx1 in neurogenesis: A neuroproteomics approach. *Front Mol Neurosci.* 2016;9(98).
 176. Zhou J, Ma C, Wang K, Li X, Zhang H, Chen J, et al. Rare and common variants analysis of the EMB gene in patients with schizophrenia. *BMC Psychiatry.* 2020;20(1):1–11.
 177. Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, et al. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat Genet.* 2010;42(11):1021–8.
 178. Demontis D, Nyegaard M, Buttenschøn HN, Hedemand A, Pedersen CB, Grove J, et al. Association of GRIN1 and GRIN2A-D With schizophrenia and genetic interaction with maternal herpes simplex virus-2 infection affecting disease risk. *Am J Med Genet Part B Neuropsychiatr Genet.* 2011;156:913–22.
 179. Myers SJ, Yuan H, Kang JQ, Tan FCK, Traynelis SF, Low CM. Distinct roles of GRIN2A and GRIN2B variants in neurological conditions [version 1; peer review: 2 approved]. *F1000Research.*

- 2019;8:1–14.
180. Jiang X, Detera-Wadleigh SD, Akula N, Mallon BS, Hou L, Xiao T, et al. Sodium valproate rescues expression of TRANK1 in iPSC-derived neural cells that carry a genetic variant associated with serious mental illness. *Mol Psychiatry* [Internet]. 2019;24(4):613–24. Available from: <http://dx.doi.org/10.1038/s41380-018-0207-1>
181. Li W, Cai X, Li HJ, Song M, Zhang CY, Yang Y, et al. Independent replications and integrative analyses confirm TRANK1 as a susceptibility gene for bipolar disorder. *Neuropsychopharmacology*. 2020;0:1–10.
182. Obi-Nagata K, Temma Y, Hayashi-Takagi A. Synaptic functions and their disruption in schizophrenia: From clinical evidence to synaptic optogenetics in an animal model. *Proc Japan Acad Ser B Phys Biol Sci*. 2019;95(5):179–97.
183. Forsyth JK, Lewis DA. Mapping the Consequences of Impaired Synaptic Plasticity in Schizophrenia through Development: An Integrative Model for Diverse Clinical Features. *Trends Cogn Sci*. 2017;21(10):760–78.
184. Laskaris LE, Di Biase MA, Everall I, Chana G, Christopoulos A, Skafidas E, et al. Microglial activation and progressive brain changes in schizophrenia. *Br J Pharmacol*. 2016;173(4):666–80.
185. Bacon C, Rappold GA. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum Genet*. 2012;131:1687–98.
186. Ayhan, Fatma Konopka G. Regulatory Genes and Pathways Disrupted in Autism Spectrum Disorders. *Prog Neuro-Psychopharmacology Biol Psychiatry*. 2019;08(87):57–64.
187. Amador A, Bostick CD, Olson H, Peters J, Camp CR, Krizay D, et al. Modeling and treating GRIN2A developmental and epileptic encephalopathy in mice. *bioRxiv*. 2019;143(1):2039–57.
188. Chi TH, Hsieh BY, Liang PS, Han TH, Hsieh M. Expression and Functional Study of Single Mutations of Carbonic Anhydrase 8 in Neuronal Cells. *Cell Mol Neurobiol* [Internet]. 2020;(0123456789). Available from: <https://doi.org/10.1007/s10571-020-00907-1>
189. Siper PM, De Rubeis S, Trelles MDP, Durkin A, Di Marino D, Muratet F, et al. Prospective investigation of FOXP1 syndrome. *Mol Autism*. 2017;8(57).
190. Meerschaut I, Rochefort D, Revençu N, Pêtre J, Corsello C, Rouleau GA, et al. FOXP1-related intellectual disability syndrome: A recognisable entity. *J Med Genet*. 2017;54:613–23.
191. Lu Y, Xu Q, Chen L, Zuo Y, Liu S, Hu Y, et al. Expression of semaphorin 6D and its receptor plexin-A1 in gastric cancer and their association with tumor angiogenesis. *Oncol Lett*. 2016;12:3967–74.
192. Chen D, Li Y, Wang L, Jiao K. SEMA6D expression and patient survival in breast invasive carcinoma. *Int J Breast Cancer*. 2015;539721.
193. Mukherji D, Massih SA, Tfayli A, Kanso M, Faraj W. Three different polymorphisms of the DPYD gene associated with severe toxicity following administration of 5-FU: A case report. *J Med Case Rep*. 2019;13(76).
194. Carter M, Nikkel S, Fernandez B, Marshall C, Noor A, Lionel A, et al. Hemizygous deletions on chromosome 1p21.3 involving the DPYD gene in individuals with autism spectrum disorder. *Clin Genet*. 2011;80(5):435–43.
195. Addis L, Virdee JK, Vidler LR, Collier DA, Pal DK, Ursu D. Epilepsy-associated GRIN2A mutations reduce NMDA receptor trafficking and agonist potency-molecular profiling and functional rescue. *Sci Rep* [Internet]. 2017;7(1):1–14. Available from: <http://dx.doi.org/10.1038/s41598-017-00115-w>
196. Müller N, Weidinger E, Leitner B, Schwarz MJ. The role of inflammation in schizophrenia. *Front Neurosci*. 2015;9(372).
197. Monji A, Kato TA, Mizoguchi Y, Horikawa H, Seki Y, Kasai M, et al. Neuroinflammation in schizophrenia especially focused on the role of microglia. *Prog Neuro-Psychopharmacology Biol Psychiatry* [Internet]. 2013;42:115–21. Available from: <http://dx.doi.org/10.1016/j.pnpbp.2011.12.002>
198. Pepe G, De Maglie M, Minoli L, Villa A, Maggi A, Vegeto E. Selective proliferative response of microglia to alternative polarization signals. *J Neuroinflammation*. 2017;14(236).
199. Bachiller S, Jiménez-Ferrer I, Paulus A, Yang Y, Swanberg M, Deierborg T, et al. Microglia in

- neurological diseases: A road map to brain-disease dependent-inflammatory response. *Front Cell Neurosci.* 2018;12(488).
- 200. Sellgren CM, Gracias J, Watmuff B, Biag JD, Thanos JM, Whittredge PB, et al. Increased synapse elimination by microglia in schizophrenia patient-derived models of synaptic pruning. *Nat Neurosci [Internet]*. 2019;22(3):374–85. Available from: <http://dx.doi.org/10.1038/s41593-018-0334-7>
 - 201. Miller BJ, Goldsmith DR. Towards an Immunophenotype of Schizophrenia: Progress, Potential Mechanisms, and Future Directions. *Neuropsychopharmacology*. 2017;42:299–317.
 - 202. Nagatsu T. The catecholamine system in health and disease - Relation to tyrosine 3-monoxygenase and other catecholamine-synthesizing enzymes. *Proc Japan Acad Ser B Phys Biol Sci.* 2006;82:388–415.
 - 203. Fujita H, Carlson ES. Catecholamine signaling that modulates cerebellar operations in cognition. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2021.
 - 204. Vidal PM, Pacheco R. The Cross-Talk Between the Dopaminergic and the Immune System Involved in Schizophrenia. *Front Pharmacol.* 2020;11(394).
 - 205. Sarkar C, Basu B, Chakroborty D, Dasgupta PS, Basu S. The immunoregulatory role of dopamine: An update. *Brain Behav Immun [Internet]*. 2010;24(4):525–8. Available from: <http://dx.doi.org/10.1016/j.bbi.2009.10.015>
 - 206. Catak Z, Kocdemir E, Ugur K, Yardim M, Sahin I, Kaya H, et al. A novel biomarker renalase and its relationship with its substrates in schizophrenia. *J Med Biochem.* 2019;38(3):299–305.
 - 207. Arion, Dominique, Lewis DA. Altered expression of regulators of the cortical chloride transporters NKCC and KCC in schizophrenia. *Arch Neuropsychiatry.* 2011;68(1):21–31.
 - 208. Ben-Ari Y. NKCC1 Chloride Importer Antagonists Attenuate Many Neurological and Psychiatric Disorders. *Trends Neurosci.* 2017;40(9).
 - 209. Schulte JT, Wierenga CJ, Bruining H. Chloride transporters and GABA polarity in developmental, neurological and psychiatric conditions. *Neuroscience and Biobehavioral Reviews.* 2018.
 - 210. Hyde TM, Lipska BK, Ali T, Mathew S V., Law AJ, Metitiri OE, et al. Expression of GABA signaling molecules KCC2, NKCC1, and GAD1 in cortical development and schizophrenia. *J Neurosci.* 2011;31(30):11088–95.
 - 211. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
 - 212. Chen G, Ning B, Shi T. Single-cell RNA-seq technologies and related computational data analysis. *Front Genet.* 2019;10(317).
 - 213. Hedlund E, Deng Q. Single-cell RNA sequencing : Technical advancements and biological applications. *Mol Aspects Med [Internet]*. 2018;59:36–46. Available from: <https://doi.org/10.1016/j.mam.2017.07.003>
 - 214. Botía JA, Vandrovčová J, Forabosco P, Guelfi S, D’Sa K, Hardy J, et al. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol.* 2017;

6.0 Appendix

Table 1: Schizophrenia-associated gene set from the 145 loci identified by Pardiñas et al. which are available in ABA's BrainSpan resource.

ABCB9	BTBD18	CNOT1	ERCC4	IGSF9B	MPPED2	PITPNM2	RELA	SLC39A8	TNFRSF13C
ACD	BTG1	CNTN4	ESAM	IMMP2L	MSL2	PLA2G15	RERE	SLC45A1	TOM1L2
ACTR5	C10orf32	COQ10B	ESRP2	INA	MSRA	PLCB2	RFT1	SLC7A6	TRANK1
ADAMTSL3	C11orf31	CPEB1	ETF1	INO80E	NAB2	PLCH2	RFTN2	SMG6	TSNARE1
AIG1	C12orf65	CPNE8	F2	INPP4B	NAGA	PLCL1	RGS6	SNAP91	TSNAXIP1
AKAP6	C14orf2	CREB3L1	FAM109B	IREB2	NCAN	PLCL2	RILPL2	SNX19	TSR1
AKT3	C16orf86	CSMD1	FAM57B	JKAMP	NCK1	PLEKHO1	RIMS1	SOX5	TYW5
ALDOA	C16orf92	CTNND1	FANCL	KAT5	NDFIP2	PPAPDC1B	RLTPR	SPATS2L	VPS13C
ALMS1	C2orf47	CTRL	FGFR1	KCNJ13	NDRG4	PPP1R13B	RPS19BP1	SREBF1	VPS45
ALPK3	C2orf82	CUL3	FHIT	KCNK7	NDUFA4L2	PPP2R2A	RPTOR	SREBF2	VRK2
AMBRA1	C3orf49	CYP17A1	FOXP1	KCNV1	NDUFA6	PPP2R3A	RRAS	SRPK2	VSIG2
ANKRD44	C4orf27	DDHD2	FSHB	KCTD13	NEK1	PPP4C	RWDD2A	SRR	WBP2NL
ANKRD63	CA14	DDX28	FXR1	KDM4A	NFATC3	PRKCB	SATB2	STAC3	WDR73
ANP32E	CA8	DFNA5	GABBR2	KIAA1324L	NGEF	PRKD1	SBNO1	STAG1	WHSC1L1
AP3B2	CACNA1C	DGKI	GATAD2A	L3MBTL2	NLGN4X	PRMT1	SCAF1	STAR	XRCC3
ARHGAP1	CACNA1D	DGKZ	GDPD3	LCAT	NMB	PSD3	SDCCAG8	STAT6	YPEL3
ARL6IP4	CACNA1I	DMTF1	GFOD2	LETM2	NRGN	PSMA4	SEC11A	SUGP1	YPEL4
AS3MT	CACNB2	DNAJC19	GIGYF2	LRP1	NRN1L	PSMB10	SEMA6D	TAB1	ZBTB37
ASH2L	CALB2	DOC2A	GPM6A	LRRC48	NT5C2	PSMD6	Sep-03	TAOK2	ZC3H7B
ASPG	CDK2AP1	DOPEY1	GPR135	LSM1	NUTF2	PTK2B	SERPINC1	TBC1D5	ZDHHC5
ASPHD1	CENPM	DPEP2	GPR52	MAD1L1	NXPH4	PTN	SERPING1	TBX6	ZEB2
ATG13	CENPT	DPEP3	GRIN2A	MAP3K11	OGFOD2	PTPRF	SETD6	TCF20	ZFYVE21
ATP2A2	CHRM3	DPYD	GRM3	MAPK3	OPCML	PTPRK	SETD8	TCF4	ZNF408
ATPAF2	CHRMA2	DRD2	HAPLN4	MARS2	OTOL1	R3HDM2	SEZ6L2	TDRD9	ZNF536
ATXN7	CHRMA3	EDC4	HARBI1	MAU2	OTUD7B	RABGAP1L	SF3B1	THAP11	ZNF592

B3GAT1	CHRNA5	EFHD1	HCN1	MED19	PAK6	RAI1	SFMBT1	THOC7	ZNF804A
BAG4	CHRN B4	EHB P1L1	HIRIP3	MEF2C	PARD6A	RALGAPA2	SFXN5	TLE3	ZNF823
BANK1	CKAP5	EMB	HS3ST5	MGAT3	PCCB	RANBP10	SHMT2	TM6SF2	ZSCAN2
BCL11B	CLCN3	EMX1	HSPA9	MMP16	PCGF6	RANGAP1	SIPA1	TMEM194A	
BCL2L12	CLP1	EP300	HSPD1	MOB4	PCNXL3	RBFOX1	SLC12A4	TMEM219	
BNIP3L	CLU	EPHX2	HSPE1	MPHOSPH9	PDE4B	RBM26	SLC32A1	TMX2	
BOLL	CNNM2	EPN2	HYDIN	MPP6	PGM3	RC3H1	SLC38A7	TMX2-CTNND1	

Table 2: Cluster assignments for each schizophrenia-associated gene over the five stages using the kmeans function available in R.

Gene Name	Developmental Stage One	Developmental Stage Two	Developmental Stage Three	Developmental Stage Four	Developmental Stage Five
ABCB1	3	2	2	2	2
ABCB9	1	2	1	2	1
ABCD2	2	1	2	2	2
ACO2	3	2	2	2	2
ACP2	3	2	1	2	1
ACTR1A	2	1	1	2	1
ACTR5	3	2	1	2	2
ADAMTSL3	2	2	2	2	2
ADAMTSL4	3	2	1	2	2
AIG1	3	2	1	2	2
AKAP6	3	1	2	1	1
AKT3	2	1	2	1	1
ALAS1	2	1	1	2	2
ALDOA	3	1	2	2	1
ALMS1	3	2	1	2	2
ANAPC7	3	2	1	2	2
ANKRD44	2	2	1	2	2
ANKRD45	2	1	1	2	2

ANKRD63	3	2	1	2	2
APOPT1	3	1	2	1	2
ARL5B	2	1	2	1	1
ARTN	3	2	1	2	2
AS3MT	3	2	1	2	2
ATF4	1	1	1	1	2
ATP13A1	1	2	1	2	1
ATPAF2	3	2	1	2	2
ATXN7	2	2	1	2	1
B9D1	3	1	2	2	2
BAG4	2	1	2	1	1
BANK1	3	2	1	2	2
BCL11B	2	2	1	2	2
BCL2L12	3	2	1	2	2
BNIP3L	2	1	2	1	1
BRD8	3	2	1	2	1
BTBD18	1	2	1	2	1
C2orf47	3	1	2	1	2
C2orf82	1	1	2	1	2
CA8	3	2	1	2	2
CACNA1C	1	2	1	2	1
CACNA1D	2	2	1	1	1
CACNA1I	1	1	2	1	1
CACNB2	1	1	2	1	1
CALB2	3	2	1	2	2
CENPM	3	2	1	2	2
CENPT	1	2	1	2	2
CEP170	2	1	1	2	1
CHRNA2	3	2	1	2	2
CHRNA3	3	2	1	2	2
CLCN3	3	2	2	2	1

CLDN23	2	2	2	2	2
CNOT1	3	2	1	2	1
CNTN4	3	2	1	2	2
CSMD1	2	1	2	1	1
CUL3	2	1	2	1	1
DFNA5	2	2	2	1	1
DGKI	1	2	1	2	1
DNAJC19	3	1	2	1	2
DOPEY1	2	2	1	2	1
DPYD	1	1	2	2	2
DRD2	3	2	1	2	2
EMB	3	2	1	2	1
EMX1	1	1	2	1	1
ESAM	3	2	2	2	2
FANCL	1	1	1	2	2
FHIT	3	2	1	2	2
FOXP1	3	1	2	1	1
FTSJ2	3	1	2	1	1
GABBR2	1	1	1	1	1
GPM6A	2	1	2	1	2
GRIA1	1	2	1	2	2
GRIN2A	1	1	2	1	1
GRM3	1	2	2	2	2
HCN1	1	1	2	1	1
IGSF9B	3	2	1	2	1
IL20RB	3	2	1	2	2
IMMP2L	3	1	2	2	2
INHBC	3	2	1	2	2
INPP4B	3	2	1	2	2
ME1	3	2	2	1	2
MGAT3	1	2	1	2	1

MMP16	2	2	2	1	1
NDFIP2	2	1	2	1	1
NLGN4X	3	2	1	2	2
OPCML	2	1	1	1	1
PDE4B	3	2	2	2	2
PSD3	1	1	1	2	1
PTPRK	1	1	2	1	1
RBFOX1	2	1	2	1	1
RERE	2	2	1	2	1
RGS6	1	1	2	1	1
RIMS1	3	2	1	2	2
RPTOR	1	2	2	1	1
SATB2	2	1	2	1	1
SEMA6D	3	2	2	2	1
SNX19	2	2	2	2	2
SPATS2L	3	1	2	1	1
TBC1D5	3	2	2	2	1
TCF4	1	2	2	2	1
TRANK1	3	2	1	2	1
TSNARE1	1	2	2	1	1
ZEB2	2	2	1	2	2
ZNF440	2	2	1	2	2
ZNF536	3	2	1	2	2
ZNF804A	3	2	1	2	1

Table 3: Gene Ontologies identified for the Black module in developmental stage One using the anRichment function as part of WGCNA in R using the default settings.

GOID	DEFINITION	ONT OLO GY	Mod ule	GO Process	FDR	Genes
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Black	positive regulation of macrophage proliferation	0.002378	ATP2A2,NAB2,ATXN7,INPP4B,TBC1D5,SNAP91,HPF1,BANK1,ADAMTSL3,ANKRD44,DNAJC19,CNTN4
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Black	protein binding	0.004713	ATP2A2,NAB2,INPP4B,TBC1D5,SNAP91,KCNV1,HPF1,BANK1,CSMD1,RFT1,DNAJC19,CNTN4
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Black	cellular metabolic process	0.006297	ATP2A2,NAB2,ATXN7,INPP4B,TBC1D5,SNAP9,KCNV1,HPF1,ADAMTSL3,AS3MT,CSMD1,RFT1,DNAJC19,CNTN4
GO:0051049	Any process that modulates the frequency, rate or extent of the directed movement of substances (such as macromolecules, small molecules, ions) into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore.	BP	Black	membrane-bounded organelle	0.025576	ATP2A2,TBC1D5,SNAP91,KCNV1,BANK1
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	BP	Black		0.029282	ATP2A2,ATXN7,TBC1D5,SNAP91,KCNV1,RFT1,DNAJC19,CNTN4

Table 4: Gene Ontologies for the Brown module in developmental stage One using the anRichment function of WGCNA on the schizophrenia-associated genes.

GOID	DEFINITION	ONTOLOGY	Module	GO Process	FDR	Genes
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	7.02E-09	CHRNA2, FHIT, GRIN2A, MMP16, MSRA, PDE4B, PTPRK, SOX5, ALMS1, CUL3, ZEB2, EPN2, PLCL2, SATB2, PSD3, FOXP1, B3GAT1, NDFIP2, BCL11B, MAIP1, ACTR5, SEMA6D, ESAM, EMB, HS3ST5, HCN1, SNORC
GO:0051716	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.	BP	Blue	cellular response to stimulus	8.29E-08	CHRNA2, FHIT, GRIN2A, MSRA, PDE4B, PTPRK, SOX5, CUL3, ZEB2, EPN2, PLCL2, SATB2, PSD3, FOXP1, B3GAT1, NDFIP2, ACTR5, SEMA6D, HCN1

GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Blue	protein binding	1.11E-07	FHIT,GRIN2A, PDE4B, PTPRK, SOX5, ALMS1,CUL3, ZEB2,EPN2 PLCL2,SATB2 ,PSD3,FOXP1,NDFIP2,BCL11B,MAIP1,ACTR5, SEMA6D,ESAM, EMB, HS3ST5, HCN1
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Blue	intracellular organelle lumen	4.63E-07	CHRNA2,FHIT,GRIN2A,MMP16,PDE4B,PTPRK, SOX5, ALMS1, CUL3,ZEB2,EPN2 ,PLCL2,SATB2,PSD3,FOXP1,NDFIP2,BCL11B,MAIP1,ACTR5, SEMA6D, HS3ST5,HCN1
GO:0051179	Any process in which a cell, a substance, or a cellular entity, such as a protein complex or organelle, is transported, tethered to or otherwise maintained in a specific	BP	Blue	localization	1.18E-06	CHRNA2,GRIN2A,PDE4B,PTPRK,ALMS1,CUL3,ZEB2 ,EPN2 ,SATB2,FOXP1,NDFIP2, BCL11B,MAIP1,SEMA6D,ESAM,EMB,HCN1

	location. In the case of substances, localization may also be achieved via selective degradation.						
GO:0040011	Self-propelled movement of a cell or organism from one location to another.	BP	Blue	locomotion	1.33E-06	GRIN2A,PDE4B,PTPRK,CUL3,ZEB2,SATB2,FOXP1, BCL11B,SEMA6D, ESAM,EMB	
GO:0048731	The process whose specific outcome is the progression of an organismal system over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together	BP	Blue	system development	1.48E-06	GRIN2A,MMP16,PTPRK,SOX5,CUL3,ZEB2,EPN2,PLCL2,SATB2,FOXP1,BCL11B,SEMA6D,EMB,HCN1,SNORC	

	to carry out a given biological process.					
GO:0016020	A lipid bilayer along with all the proteins and protein complexes embedded in it and attached to it.	CC	Blue	membrane	1.80E-06	CHRNA2,FHIT,GRIN2A,MMP16,MSRA,PDE4B,PTPRK,CUL3,EPN2,PSD3,B3GAT1,NDFIP2,MAIP1,SEMA6D,ESAM,EMB),HS3ST5,HCN1,SNORC

Table 5: Gene Ontology Brown Module for developmental stage One using the anEnrichment function as part of WGCNA in R using the default settings

GOID	DEFINITION	ONTOLOGY	Module	GO Process	FDR	Genes
GO:0097090	A process which results in the assembly, arrangement of constituent parts, or disassembly of a presynaptic membrane, including any proteins associated with the membrane, but excluding other cellular components. A presynaptic membrane is a specialized area of membrane of the axon terminal that faces the plasma membrane of the neuron or muscle fiber with which the axon terminal establishes a synaptic junction.	B P	Brown	presynaptic membrane organization	7.99E-05	CALB2,GPM6A,DGKI,GABBR2,NLGN4X
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	B P	Brown	positive regulation of macrophage proliferation	8.30E-05	BNIP3L,CALB2,EMX1,GPM6A,OPCML,TLE3,DGKI,GABBR2,RGS6,KDM4A,CNOT1,DOP1A,RBFOX1,ZNF823,RALGAPA2,NLGN4X
GO:0070013	An organelle lumen that is part of an intracellular organelle.	C C	Brown	intracellular organelle lumen	8.38E-05	BNIP3L,CALB2,EMX1,GPM6A,TLE3,DGKI,GABBR2,RGS6,KDM4A,CNOT1,RBFOX1,ZN

						F823,RALGAPA2,NLGN4X
GO:0060076	A synapse in which an action potential in the presynaptic cell increases the probability of an action potential occurring in the postsynaptic cell.	C C	Brown	excitatory synapse	0.00022	CALB2,DGKI,NLGN4X

Table 6: Gene Ontology for Greenyellow Module in Developmental Stage One using the anRichment function as part of WGCNA in R using the default settings

GOID	DEFINITION	ONTOLOGY	Module	GO Process	FDR	Genes
GO:0005654	That part of the nuclear content other than the chromosomes or the nucleolus.	CC	Pink	nucleoplasm	0.012863	BTG1,ERCC4,DMTF1,CENPM
GO:1905907	Any process that stops, prevents or reduces the frequency, rate or extent of amyloid fibril formation.	BP	Pink	negative regulation of amyloid fibril formation	0.01836	ERCC4
GO:1901398	NA	BP	Pink	regulation of transforming growth factor beta3 activation	0.03198	ERCC4
GO:1904844	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a L-glutamine stimulus.	BP	Pink	response to L-glutamine	0.03198	ERCC4
GO:0000110	One of several protein complexes involved in nucleotide-excision repair; possesses DNA	CC	Pink	nucleotide-excision repair factor 1 complex	0.042679	ERCC4

	damage recognition and endodeoxynuclease activities. In <i>S. cerevisiae</i> , it is composed of Rad1p, Rad10p, and Rad14p; in human the subunits are ERCC4/XPF, ERCC1 and XPA, respectively.					
GO:0061819	A telomere maintenance process that results in the formation of small fragments of circular extrachromosomal DNA elements which contain telomeric DNA. It is speculated that telomeric DNA-containing double minutes are formed through a recombination event between the telomere and chromosome-internal TTAGGG-like sequences. Telomeric DNA-containing double minutes appear as two closely positioned dots in	BP	Pink	telomeric DNA-containing double minutes formation	0.042679	ERCC4

	metaphase.					
GO:1905903	Any process that stops, prevents or reduces the frequency, rate or extent of mesoderm formation.	BP	Pink	negative regulation of mesoderm formation	0.042679	ERCC4
GO:1905904	Any process that activates or increases the frequency, rate or extent of mesoderm formation.	BP	Pink	positive regulation of mesoderm formation	0.042679	ERCC4

Table 7: Gene Ontologies for Turquoise Module in Developmental Stage One using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Turquoise	intracellular organelle lumen	4.04E-10	ALDOA ,RERE, CA8, CACNA1C, CACNA1D, CACNB2, CHRM3, CHRNA3, GSDME,DPYD, DRD2, ETF1, PRKCB, TCF4, ASH2L, GPR52, AKAP6, PLCH2, IGSF9B, RIMS1, VPS13C, FANCL, RPTOR, ACD, ANP32E ,ZNF804A,SNX19,BTBD18
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	1.78E-09	ALDOA, RERE, CA8, CACNA1C, CACNA1D, CACNB2, CHRM3, CHRNA3, GSDME, DPYD, DRD2, ETF1, PRKCB, TCF4, ASH2L, GPR52, AKAP6, PLCH2, IGSF9B, RIMS1, ABCB9, VPS13C, FANCL,RPTOR,ACD,ANP32E ,IMMP2L,ZNF804A, CPNE8, SNX19, BTBD18
GO:0070016	Interacting selectively and non-covalently with the armadillo repeat domain of a protein, an approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity protein armadillo. Arm-repeat proteins are involved in various processes, including intracellular signalling and cytoskeletal regulation.	MF	Turquoise	armadillo repeat domain binding	4.05E-09	ALDOA, CACNA1C, CACNA1D, CACNB2, CHRM3, CHRNA3, DPYD, DRD2, PRKCB, AKAP6, PLCH2, IGSF9B, RIMS1, VPS13C, RPTOR, ACD, ZNF804A, SNX19
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or	BP	Turquoise	cellular component organization	2.76E-08	ALDOA,RERE,CACNB2,CHRNA3,DRD2,ETF1),PRKCB, TCF4,ASH2L,AKAP6 ,IGSF9B, RIMS1, VPS13C, RPTOR, ACD, ANP32E, IMMP2L, ZNF804A, SNX19, BTBD18

	disassembly of a cellular component.					
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl)pyrocatechol [4-(2-aminoethyl)benzene-1,2-diol] and derivatives formed by substitution.	BP	Turquoise	cellular response to catecholamine stimulus	4.40E-08	ALDOA,RERE,CACNB2,CHRNA3,DRD2,ETF1,PRKCB,TCF4,ASH2L,AKAP6,IGSF9B,RIMS1,VPS13C,RPTOR,ACD,ANP32E,IMMP2L,ZNF804A,SNX19,BTBD18
GO:0046928	Any process that modulates the frequency, rate or extent of the regulated release of a neurotransmitter from a cell.	BP	Turquoise	regulation of neurotransmitter secretion	1.11E-07	CACNA1D,CACNB2,CHRNA3,DRD2,PRKCB,RIMS1
GO:0044057	Any process that modulates the frequency, rate or extent of a system process, a multicellular organismal process carried out by any of the organs or	BP	Turquoise	regulation of system process	1.46E-07	CACNA1C,CACNA1D,CACNB2,CHRM3,CHRNA3,DRD2,AKAP6,IGSF9B,RIMS1

	tissues in an organ system.					
GO:0042391	Any process that modulates the establishment or extent of a membrane potential, the electric potential existing across any membrane arising from charges in the membrane itself and from the charges present in the media on either side of the membrane.	BP	Turquoise	regulation of membrane potential	3.30E-07	CACNA1C,CACNA1D,CACNB2,CHRNA3,DRD2,AKAP6,IGSF9B,RIMS1
GO:0003012	A organ system process carried out at the level of a muscle. Muscle tissue is composed of contractile cells or fibers.	BP	Turquoise	muscle system process	5.33E-07	ALDOA,CACNA1C,CACNA1D,CACNB2,CHRM3, CHRNA3,DRD2,AKAP6
GO:0043167	Interacting selectively and non-covalently with ions, charged atoms or groups of atoms.	MF	Turquoise	ion binding	1.13E-06	RERE,CA8,CACNA1C,CACNA1D,CHRM3,CHRNA3,GSDME,DPYD,DRD2, PRKCB,ASH2L, PLCH2, RIMS1,ABCB9,FANCL,ZNF804A,CPNE8,SNX19
GO:0045202	The junction between a nerve fiber of one neuron and another neuron, muscle fiber or glial cell. As the nerve fiber approaches the synapse it enlarges into	CC	Turquoise	synapse	1.31E-06	CACNA1C,CACNA1D,CACNB2,CHRM3,CHRNA3,DRD2,PRKCB,IGSF9B, RIMS1,ZNF804A

	a specialized structure, the presynaptic nerve ending, which contains mitochondria and synaptic vesicles. At the tip of the nerve ending is the presynaptic membrane; facing it, and separated from it by a minute cleft (the synaptic cleft) is a specialized area of membrane on the receiving cell, known as the postsynaptic membrane. In response to the arrival of nerve impulses, the presynaptic nerve ending secretes molecules of neurotransmitters into the synaptic cleft. These diffuse across the cleft and transmit the signal to the postsynaptic membrane.					
GO:0101020	Catalysis of the reaction: estrogen + donor-H2 + O2 = 16-alpha-hydroxyestrogen + H2O.	MF	Turquoise	estrogen 16-alpha-hydroxylase activity	1.33E-06	CACNA1D,CACNB2,CHRNA3,DRD2,PRKCB,RIMS1
GO:0016020	A lipid bilayer along with	CC	Turquoise	membrane	1.51E-	ALDOA,CACNA1C,CACNA1D,CACNB2,CHRM3,CHRNA3,GSDME,DRD2,

	all the proteins and protein complexes embedded in it an attached to it.				06	PRKCB,GPR52, AKAP6,PLCH2,IGSF9B,RIMS1,ABCB9,VPS13C,RPTOR,IMMP2L,ZNF804A,CPNE8,SNX19
GO:2000374	Any process that modulates the frequency, rate or extent of oxygen metabolic process.	BP	Turquoise	regulation of oxygen metabolic process	1.80E-06	CACNA1D,CACNB2,DRD2,PRKCB,RIMS1

Table 8: Gene Ontologies for Blue Module in Developmental Stage Two using the anEnrichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Blue	intracellular organelle lumen	0.00018	CHRM3,CHRNA2,GSDME ,EMX1 ,SOX5,RGS6,CNOT1,PSD3,KCNV1, HPF1, BANK1,RALGAPA2,NLGN4X, RPTOR,CSMD1,DNAJC19
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Blue	protein binding	0.00048	CHRM3,EMX1,SOX5,RGS6,CNOT1,PSD3,AIG1,HPF1,BANK 1,RALGAPA2,ADAMTSL3, NLGN4X,RPTOR,ATPAF2,DNAJC19
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	0.000577	CHRM3,CHRNA2,GSDME, EMX1, SOX5, RGS6,CNOT1,PSD3,KCNV1, AIG1, HPF1,RALGAPA2,ADAMTSL3, NLGN4X,RPTOR,CSMD1, ATPAF2,DNAJC19
GO:0051716	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus. The process begins with detection of the stimulus by a cell and ends with a change in state or activity of the cell.	BP	Blue	cellular response to stimulus	0.001382	CHRM3,CHRNA2,GSDME,SOX5,RGS6,CNOT1,PSD3,HPF1, BANK1,RALGAPA2, NLGN4X,RPTOR

Table 9: Gene Ontologies for Brown Module in Developmental Stage Two using the anRICHment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Brown	intracellular organelle lumen	1.85E-05	ALDOA,BNIP3L,CACNA1C,DRD2,FHIT,MMP16,NAB2,ALMS1,GABBR2,ZEB2,SNAP91,RIMS1,PLCL2,ZNF804A,CNTN4,BTBD18
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Brown	positive regulation of macrophage proliferation	2.91E-05	ALDOA,BNIP3L,CACNA1C,DRD2,FHIT,MMP16,NAB2,ALMS1,GABBR2,ZEB2,SNAP91,RIMS1,PLCL2,AS3MT,ZNF804A,CNTN4,EHBP1L1,BTBD18
GO:0005737	All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures.	CC	Brown	cytoplasm	4.47E-05	ALDOA,BNIP3L,CACNA1C,DRD2,FHIT,MMP16,ALMS1,GABBR2,ZEB2,SNAP91,RIMS1,PLCL2,AS3MT,ZNF804A,EHBP1L1
GO:0003674	A molecular process that can be carried out by the action of a single macromolecular machine, usually via direct physical interactions with other molecular entities. Function in this sense denotes an action, or activity, that a gene product (or a complex) performs. These actions are described from two distinct but related perspectives: (1)	MF	Brown	molecular_function	0.000133	ALDOA,BNIP3L,CACNA1C,DRD2,FHIT,MMP16,NAB2,ALMS1,GABBR2,ZEB2,SNAP91,RIMS1,PLCL2,AS3MT,ZNF804A,CNTN4,EHBP1L1

	biochemical activity, and (2) role as a component in a larger system/process.					
GO:0097090	A process which results in the assembly, arrangement of constituent parts, or disassembly of a presynaptic membrane, including any proteins associated with the membrane, but excluding other cellular components. A presynaptic membrane is a specialized area of membrane of the axon terminal that faces the plasma membrane of the neuron or muscle fiber with which the axon terminal establishes a synaptic junction.	BP	Brown	presynaptic membrane organization	0.000307	CACNA1C,DRD2,GABBR2,SNAP91,RIMS1

Table 10: Gene Ontology for Green Module Stage Two using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0005654	That part of the nuclear content other than the chromosomes or the nucleolus.	CC	Green	nucleoplasm	0.021661	BTG1,TLE3,DMTF1,SATB2,CENPM
GO:0140244	NA	BP	Green	regulation of translation at presynapse	0.039021	BTG1,TLE3,DMTF1,SATB2
GO:0098700	The active transport of neurotransmitters into a synaptic vesicle. This import is fuelled by an electrochemical gradient across the vesicle membrane, established by the action of proton pumps.	BP	Green	neurotransmitter loading into synaptic vesicle	0.040787	CACNA1D
GO:0050806	Any process that activates or increases the frequency, rate or extent of synaptic transmission, the process of communication from a neuron to a target (neuron, muscle, or secretory cell) across a synapse.	BP	Green	positive regulation of synaptic transmission	0.063569	CACNA1D,GSF9B

Table 11: Gene Ontology for Turquoise Module Stage Two using the anRICHment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	3.35E-15	RERE,ATP2A2,CA8,CACNB2,CALB2,CHRNA3,DPYD,ETF1,GRIN2A,OPCML,PDE4B,PTPRK,TCF4,MAD1L1,INPP4B,ASH2L,DGKI),GPR52,PLCH2,TBC1D5,EPN2,DOP1A,ABCB9,FOXP1,B3GAT1,RBFOX1,FANCL,MSL2,ZNF823,AMBRA1,BCL11B,ACD,MAIP1,ACTR5,SEMA6D,IMMP2L,ESAM,RFT1,OTOL1,CPNE8,HCN1,SNORC,SNX19
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Turquoise	intracellular organelle lumen	9.06E-10	RERE,ATP2A2,CA8,CACNB2,CALB2,CHRNA3,DPYD,ETF1,GRIN2A,PDE4B,PTPRK,TCF4,MAD1L1,INPP4B,ASH2L,DGKI,GPR52,PLCH2,TBC1D5,EPN2,FOXP1,RBFOX1,FANCL,ZNF823,AMBRA1,BCL11B,ACD,MAIP1,ACTR5,SEMA6D,RFT1, HCN1,SNX19
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Turquoise	protein binding	6.74E-09	RERE,ATP2A2,CA8,CACNB2,CHRNA3,DPYD,ETF1,GRIN2A,PDE4B,PTPRK,TCF4,MAD1L1,INPP4B,ASH2L,DGKI,TBC1D5,EPN2,ABCB9,FOXP1,RBFOX1,FANCL, AMBRA1,BCL11B,ACD,MAIP1,ACTR5,SEMA6D,ESAM,ANKRD44,HCN1,SNX19
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	BP	Turquoise	cellular component organization	1.00E-08	RERE,ATP2A2,CACNB2,CHRNA3,ETF1,PTPRK,TCF4,MAD1L1,ASH2L,TBC1D5,EPN2,MSL2,AMBRA1,BCL11B,ACD,MAIP1,ACTR5,SEMA6D,IMMP2L,ESAM,RFT1,OTOL1, HCN1,SNX19
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a	BP	Turquoise	cellular response to catecholamine stimulus	1.71E-08	RERE,ATP2A2,CACNB2,CHRNA3,ETF1,PTPRK,TCF4,MAD1L1,ASH2L,TBC1D5, EPN2,MSL2,AMBRA1,BCL11B,ACD,MAIP1,ACTR5,SEMA6D,IMMP2L,ESAM,RFT1,OTOL1, HCN1,SNX19

	catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl)pyrocatechol [4-(2-aminoethyl)benzene-1,2-diol] and derivatives formed by substitution.					
GO:0043227	Organized structure of distinctive morphology and function, bounded by a single or double lipid bilayer membrane. Includes the nucleus, mitochondria, plastids, vacuoles, and vesicles. Excludes the plasma membrane.	CC	Turquoise	membrane-bounded organelle	4.26E-07	RERE,ATP2A2,CALB2,GRIN2A,PDE4B,PTPRK,TCF4,MAD1L1,ASH2L,DGKI, TBC1D5,EPN2, DOP1A,ABCB9,FOXP1,B3GAT1,RBFOX1, FANCL,MSL2,ZNF823,AMBRA1, BCL11B,ACD, MAIP1,ACTR5,SEMA6D, IMMP2L,RFT1,CPNE8),SNX19
GO:0016020	A lipid bilayer along with all the proteins and protein complexes embedded in it and attached to it.	CC	Turquoise	membrane	5.12E-07	ATP2A2,CACNB2,CALB2,CHRNA3,GRIN2A,OPCML,PDE4B,PTPRK,DGKI),GPR52, PLCH2, TBC1D5,EPN2,DOP1A,ABCB9,B3GAT1,AMBRA1, MAIP1,SEMA6D,IMMP2L,ESAM,RFT1,CPNE8,HCN1,SNORC,SNX19
GO:0032991	A stable assembly of two or more macromolecules, i.e. proteins, nucleic acids, carbohydrates or lipids, in which at least one component is a protein	CC	Turquoise	protein-containing complex	8.57E-07	RERE,ATP2A2,CACNB2,CHRNA3,ETF1,GRIN2A,PDE4B,TCF4,MAD1L1, ASH2L,DGKI, TBC1D5,EPN2,FANCL,MSL2,ACD, ACTR5, IMMP2L,ESAM,OTOL1,HCN1

	and the constituent parts function together.					
GO:0051234	Any process that localizes a substance or cellular component. This may occur via movement, tethering or selective degradation.	BP	Turquoise	establishment of localization	1.06E-06	ATP2A2,CACNB2,CHRNA3,GRIN2A,PDE4B,MAD1L1,DGKI,PLCH2,TBC1D5,EPN2,DOP1A,ABCB9,FOXP1,RBFOX1,ACD,MAIP1,IMMP2L,RFT1,HCN1,SNX19
GO:0043169	Interacting selectively and non-covalently with cations, charged atoms or groups of atoms with a net positive charge.	MF	Turquoise	cation binding	1.81E-06	RERE,ATP2A2,CA8,CALB2,CHRNA3,DPYD,GRIN2A,PDE4B,ASH2L,DGKI,PLCH2,FOXP1,B3GAT1,FANCL,MSL2,ZNF823,BCL11B,OTOL1
GO:0048731	The process whose specific outcome is the progression of an organismal system over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together to carry out a given biological process.	BP	Turquoise	system development	1.97E-06	RERE,CHRNA3,GRIN2A,OPCML),PTPRK,TCF4,MAD1L1,ASH2L,EPN2,FOXP1,RBFOX1,AMBRA1,BCL11B,SEMA6D,IMMP2L,OTOL1,HCN1,SNORC,SNX19

Table 12: Gene Ontologies for Blue Module Stage Three using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Blue	positive regulation of macrophage proliferation	0.044306	CHRM3,GSDME,SOX5,RGS6,AIG1,HPF1,ADAMTSL3,CSMD1,ATPAF2
GO:0072588	A ribonucleoprotein complex that contains an RNA of the box H/ACA type and the four core proteins dyskerin, NOP10, NHP2, and GAR1 (human protein nomenclature). RNA pseudouridylation (isomerization of uridine to pseudouridine) is the major, and most likely the ancestral, function of H/ACA RNPs.	CC	Blue	box H/ACA RNP complex	0.062848	HPF1

	targets include both large and small ribosomal RNAs (rRNAs), and small nuclear RNA (U2 snRNA). In addition to these catalytic H/ACA RNPs, a less abundant but more diverse class of structural H/ACA RNPs exists, which does not have pseudouridylation activity. These include the vertebrate telomerase RNP complex.					
GO:0016787	Catalysis of the hydrolysis of various bonds, e.g. C-O, C-N, C-C, phosphoric anhydride bonds, etc. Hydrolase is the systematic name for any enzyme of EC class 3.	MF	Blue	hydrolase activity	0.080744	CHRM3,RGS6,AIG1,ADAMTSL3

GO:0018312	The transfer, from NAD, of ADP-ribose to peptidyl-serine to form peptidyl-O-(ADP-ribosyl)-L-serine.	BP	Blue	peptidyl-serine ADP-ribosylation	0.082715	HPF1
GO:1904796	Any process that modulates the frequency, rate or extent of core promoter binding.	BP	Blue	regulation of core promoter binding	0.082715	CHRM3
GO:2000830	Any process that activates or increases the frequency, rate or extent of parathyroid hormone secretion.	BP	Blue	positive regulation of parathyroid hormone secretion	0.103383	SOX5
GO:0005515	Interacting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	MF	Blue	protein binding	0.108449	CHRM3, SOX5, RGS6, AIG1, HPF1, ADAMTSL3, ATPAF2

Table 13: Gene Ontologies for Brown Module Stage Three using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0001659	A homeostatic process in which an organism modulates its internal body temperature.	BP	Brown	temperature homeostasis	0.00224	DRD2,ALMS1,PLCL2
GO:0070025	Interacting selectively and non-covalently with carbon monoxide (CO).	MF	Brown	carbon monoxide binding	0.002884	CACNA1D,DRD2,MMP16,GABBR2,ZEB2,PLC L2
GO:0098935	The directed movement of organelles or molecules along microtubules in dendrites.	BP	Brown	dendritic transport	0.003021	CACNA1D,DRD2,GABBR2,PLCL2
GO:0099552	Cell-cell signaling between presynapse and postsynapse, via the release and reception of lipid molecules, that modulates the synaptic transmission properties of the synapse.	BP	Brown	trans-synaptic signaling by lipid, modulating synaptic transmission	0.003101	CACNA1D,DRD2,GABBR2,PLCL2
GO:0099551	Cell-cell signaling between presynapse and postsynapse, via the vesicular release and reception of neuropeptide molecules, that modulates the synaptic transmission properties of the synapse.	BP	Brown	trans-synaptic signaling by neuropeptide, modulating synaptic transmission	0.003292	CACNA1D,DRD2,GABBR2,PLCL2
GO:0044093	Any process that activates or increases the rate or extent of a molecular function, an elemental biological activity occurring at the molecular level, such as catalysis or binding.	BP	Brown	positive regulation of molecular function	0.003876	CACNA1D,DRD2,MMP16,ZEB2,PLCL2
GO:0007214	The series of molecular signals	BP	Brown	gamma-	0.004111	GABBR2,PLCL2

	generated by the binding of gamma-aminobutyric acid (GABA, 4-aminobutyrate), an amino acid which acts as a neurotransmitter in some organisms, to a cell surface receptor.			aminobutyric acid signaling pathway		
GO:0005737	All of the contents of a cell excluding the plasma membrane and nucleus, but including other subcellular structures.	CC	Brown	cytoplasm	0.00445	CACNA1D,DRD2,FHIT,MMP16,ALMS1,GABBR2,ZEB2,PLCL2,AS3MT

Table 14: Gene Ontology for Turquoise Stage three using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Turquoise	positive regulation of macrophage proliferation	1.99E-19	(More than 50 overlapping genes)
GO:0016043	A process that results in the assembly, arrangement of constituent parts, or disassembly of a cellular component.	BP	Turquoise	cellular component organization	4.26E-15	ALDOA,ATP2A2,BNIP3L,CACNB2,CHRNA2,CHRNA3,ERCC4,ETF1,PRKCB,PTPRK,ATXN7,TCF4,MAD1L1),EPN2,IGSF9B,RIMS1,CNOT1,KCNV1,VPS13C,AMBRA1,RPTOR,BCL11B,CENPM,MAIP1,SEMA6D,ANP32E,IMMP2L,ZNF804A,RFT1,DNAJC19,EMB,HCN1,SNX19
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl)pyrocatechol [4-(2-aminoethyl)benzene-1,2-diol] and derivatives formed by substitution.	BP	Turquoise	cellular response to catecholamine stimulus	7.52E-15	ALDOA,ATP2A2,BNIP3L,CACNB2,CHRNA2,CHRNA3,ERCC4,ETF1,PRKCB,PTPRK,ATXN7,TCF4,MAD1L1,EPN2,IGSF9B,RIMS1,CNOT1,KCNV1,VPS13C,AMBRA1,RPTOR,BCL11B,CENPM,MAIP1,SEMA6D,ANP32E,IMMP2L)ZNF804A,RFT1,DNAJC19,EMB,HCN1,SNX19
GO:0070013	An organelle lumen that	CC	Turquoise	intracellular	7.66E-	ALDOA,ATP2A2,BNIP3L,CA8,CACNA1C,CACNB2,CALB2,CHRNA2,CHRNA3,

	is part of an intracellular organelle.			organelle lumen	14	DPYD,ERCC4,ETF1,NAB2, PDE4B, PRKCB,PTPRK,TCF4,MAD1L1,INPP4B,DGKI,GPR52,EPN2,I GSF9B,RIMS1,CNOT1,KCNV1,RBFOX1,VPS13C,ZNF823, AMBRA1,RALGAPA2,RPTOR,BCL11B,MAIP1,SEMA6D,ANP32E, ZNF804A,RFT1,DNAJC19,HCN1,SNX19
GO:0070016	Interacting selectively and non-covalently with the armadillo repeat domain of a protein, an approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity protein armadillo. Arm-repeat proteins are involved in various processes, including intracellular signalling and cytoskeletal regulation.	MF	Turquoise	armadillo repeat domain binding	6.19E-10	ALDOA,ATP2A2,BNIP3L,CACNA1C,CACNB2,CALB2,CHRNA2,CHRNA3, DPYD,ERCC4,PDE4B,PRKCB,DGKI,IGSF9B,RIMS1,VPS13C,RPTOR,MAIP1, SEMA6D,ZNF804A,RFT1,HCN1,SNX19
GO:0042391	Any process that modulates the establishment or extent of a membrane potential, the electric potential existing across any membrane arising from charges in the membrane itself and from the charges present in the media on	BP	Turquoise	regulation of membrane potential	2.04E-08	ATP2A2,BNIP3L,CACNA1C,CACNB2,CHRNA2,CHRNA3,DGKI, IGSF9B,RIMS1,HCN1

	either side of the membrane.						
GO:0120060	Any process that modulates the frequency, rate or extent of any gastric emptying process, the process in which the liquid and liquid-suspended solid contents of the stomach exit through the pylorus into the duodenum.	BP	Turquoise	regulation of gastric emptying	7.63E-08	ATP2A2,CACNA1C,CALB2,CHRNA2,CHRNA3,PDE4B,PRKCB,PTPRK,DGKI,IGSF9B,AMBRA1,RPTOR,BCL11B,ZNF804A,EMB,HCN1	
GO:0045202	The junction between a nerve fiber of one neuron and another neuron, muscle fiber or glial cell. As the nerve fiber approaches the synapse it enlarges into a specialized structure, the presynaptic nerve ending, which contains mitochondria and synaptic vesicles. At the tip of the nerve ending is the presynaptic membrane; facing it, and separated from it by a minute cleft (the synaptic cleft) is a specialized area of membrane on the	CC	Turquoise	synapse	7.73E-08	ATP2A2,CACNA1C,CACNB2,CALB2,CHRNA2,CHRNA3,PDE4B,PRKCB,DGKI,IGSF9B,RIMS1,ZNF804A,EMB	

	receiving cell, known as the postsynaptic membrane. In response to the arrival of nerve impulses, the presynaptic nerve ending secretes molecules of neurotransmitters into the synaptic cleft. These diffuse across the cleft and transmit the signal to the postsynaptic membrane.					
GO:0048731	The process whose specific outcome is the progression of an organismal system over time, from its formation to the mature structure. A system is a regularly interacting or interdependent group of organs or tissues that work together to carry out a given biological process.	BP	Turquoise	system development	1.48E-07	CACNA1C,CHRNA3,NAB2,OPCML,PRKCB,PTPRK,TCF4,MAD1L1,EPN2,GSF9B,RIMS1,RBFOX1,AMBRA1,BCL11B,SEMA6D,IMMP2L,ZNF804A,DNAJC19,EMB,HCN1,SNORC,SNX19
GO:0030425	A neuron projection that has a short, tapering, morphology. Dendrites receive and integrate signals from other	CC	Turquoise	dendrite	2.97E-07	CACNA1C,CALB2,CHRNA3,PDE4B,PTPRK,DGKI,IGSF9B,RPTOR,ZNF804A,HCN1

	neurons or from sensory stimuli, and conduct nerve impulses towards the axon or the cell body. In most neurons, the impulse is conveyed from dendrites to axon via the cell body, but in some types of unipolar neuron, the impulse does not travel via the cell body.					
GO:0097473	Any apoptotic process in a retinal rod cell, one of the two photoreceptor cell types of the vertebrate retina.	BP	Turquoise	retinal rod cell apoptotic process	2.98E-07	CACNA1C,CALB2,CHRNA3,PDE4B,PTPRK,DGKI,IGSF9B,RPTOR,ZNF804A,HCN1
GO:0043231	Organized structure of distinctive morphology and function, bounded by a single or double lipid bilayer membrane and occurring within the cell. Includes the nucleus, mitochondria, plastids, vacuoles, and vesicles. Excludes the plasma membrane.	CC	Turquoise	intracellular membrane-bounded organelle	4.83E-07	ALDOA,ATP2A2,BNIP3L,CALB2,ERCC4,MSRA,NAB2,PRKCB,PTPRK,ATXN7,TCF4,MAD1L1,DGKI,EPN2,CNOT1,DOP1A,ABCB9,RBOFOX1,VPS13C,ZNF823,AMBRA1,RALGAPA2,RPTOR,BCL11B,CENPM,MAIP1,SEMA6D ,ANP32E,IMMP2L,ZNF804A,DNAJC19
GO:0046928	Any process that modulates the frequency, rate or extent of the regulated release	BP	Turquoise	regulation of neurotransmitter secretion	1.22E-06	ATP2A2,CACNB2,CHRNA3,PRKCB,DGKI,RIMS1

	of a neurotransmitter from a cell.					
GO:0030054	A cellular component that forms a specialized region of connection between two or more cells or between a cell and the extracellular matrix. At a cell junction, anchoring proteins extend through the plasma membrane to link cytoskeletal proteins in one cell to cytoskeletal proteins in neighboring cells or to proteins in the extracellular matrix.	CC	Turquoise	cell junction	1.39E-06	ATP2A2,CACNA1C,CACNB2,CALB2,CHRNA2,CHRNA3,PDE4B,PRKCB,PTPRK,DGKI,IGSF9B,RIMS1,ZNF804A,EMB
GO:0005829	The part of the cytoplasm that does not contain organelles but which does contain other particulate matter, such as protein complexes.	CC	Turquoise	cytosol	1.62E-06	ALDOA,BNIP3L,CALB2,DPYD,ETF1,MSRA,PDE4B,PRKCB,ATXN7,MAD1L1,INPP4B,DGKI,EPN2,RIMS1,CNOT1,DOP1A,VPS13C,AMBRA1,RALGAPA2,RPTOR,CENPM
GO:0055085	The process in which a solute is transported across a lipid bilayer, from one side of a membrane to the other.	BP	Turquoise	transmembrane transport	2.25E-06	ATP2A2,CACNA1C,CACNB2,CHRNA2,CHRNA3,PDE4B,PRKCB,ABCB9,KCNV1,MAIP1,DNAJC19,EMB,HCN1
GO:0071986	A eukaryotically conserved protein	CC	Turquoise	Ragulator complex	2.25E-06	ATP2A2,CACNA1C,CACNB2,CALB2,CHRNA2,CHRNA3,OPCML,PDE4B,PRKCB,PTPRK,DGKI,GPR52,IGSF9B,RIMS1,

<p>complex; in humans, it is comprised of LAMTOR1, LAMTOR2, LAMTOR3, LAMTOR4, and LAMTOR5. The complex is anchored to lipid rafts in late endosome membranes via LAMTOR1, constitutes a guanine nucleotide exchange factor (GEF) for the Rag GTPases.</p>				<p>KCNV1, RALGAPA2, SEMA6D, ZNF804A, EMB, CPNE8, HCN1, SNORC</p>
---	--	--	--	--

Table 15: Gene Ontology for Yellow Module Stage Three using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	GO Process	FDR	Genes
GO:0120041	Any process that activates or increases the frequency, rate or extent of macrophage proliferation.	BP	Yellow	positive regulation of macrophage proliferation	8.68E-23	More than 50 overlapping genes
GO:0070013	An organelle lumen that is part of an intracellular organelle.	CC	Yellow	intracellular organelle lumen	2.81E-17	RERE, BNIP3L, CA8, CACNA1C, CACNA1D, CHRNA2, GSDME, EMX1, FHIT, GPM6A, GRM3, ME1, MGAT3, PDE4B, ALMS1, CUL3, CACNA1I, DGKI, AKAP6, BAG4, GABBR2, RGS6, TBC1D5, ZEB2, AKT3, BRD8, RIMS1, CNOT1, SATB2, PSD3, B9D1, FOXP1, ANAPC7, NDFIP2, ATP13A1, NLGN4X, RPTOR, MAIP1, ACTR5, BCL2L12, ZNF804A, ZNF440, DNAJC19, CNTN4, HCN1, SNX19, BTBD18
GO:0071870	Any process that results in a change in state or activity of a cell (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a catecholamine stimulus. A catecholamine is any of a group of biogenic amines that includes 4-(2-aminoethyl) pyrocatechol [4-(2-aminoethyl) benzene-1,2-diol] and derivatives formed by substitution.	BP	Yellow	cellular response to catecholamine stimulus	2.16E-13	RERE, BNIP3L, CHRNA2, EMX1, GPM6A, ME1, ALMS1, CUL3, AKAP6, BAG4, TBC1D5, ZEB2, AKT3, BRD8, RIMS1, CNOT1, SATB2, B9D1, NLGN4X, RPTOR, CENPM, MAIP1, ACTR5, IMMP2L, ESAM, ATPAF2, ZNF804A, DNAJC19, EMB, CNTN4, HCN1, SNX19, BTBD18
GO:0007165	The cellular process in which a signal is conveyed to trigger a	BP	Yellow	signal transduction	1.30E-09	BNIP3L, CA8, CACNA1C, CACNA1D, CHRNA2, GSDME, FHIT, GRM3, PDE4B, CUL3, CACNA1I, DGKI, AKAP6, BAG4, GABBR2, RGS6, ZEB2, AKT3,

	change in the activity or state of a cell. Signal transduction begins with reception of a signal (e.g. a ligand binding to a receptor or receptor activation by a stimulus such as light), or for signal transduction in the absence of ligand, signal-withdrawal or the activity of a constitutively active receptor. Signal transduction ends with regulation of a downstream cellular process, e.g. regulation of transcription or regulation of a metabolic process. Signal transduction covers signalling from receptors located on the surface of the cell and signalling via molecules located within the cell. For signalling between cells, signal transduction is restricted to events at and within the receiving cell.					BRD8, RIMS1, CNOT1, PSD3, B9D1, FOXP1, NDFIP2, NLGN4X, RPTOR, BCL2L12
GO:0070016	Interacting selectively and non-covalently with the armadillo repeat domain of a protein, an approximately 40 amino	MF	Yellow	armadillo repeat domain binding	4.16E-09	BNIP3L, CACNA1C, CACNA1D, CHRNA2, EMX1, GPM6A, PDE4B, ALMS1, CUL3, CACNA1I, DGKI, AKAP6, BAG4, AKT3, RIMS1, ATP13A1, NLGN4X, RPTOR, MAIP1, ZNF804A, CNTN4, HCN1, SNX19

acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity protein armadillo. Arm-repeat proteins are involved in various processes, including intracellular signalling and cytoskeletal regulation.					
--	--	--	--	--	--

Table 16: Gene Ontology for Yellow Module Stage 5 One using the anRichment function as part of WGCNA in R using the default settings

GOID	Definition	Ontology	Module	Go Process	FDR	Genes
GO:0044260	The chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass, as carried out by individual cells.	BP	yellow	cellular macromolecule metabolic process	0.00076	BTG1,DMTF1,SATB2,FOXP1,NDFIP2,BANK1,FANCL,ACTR5
GO:0010468	Any process that modulates the frequency, rate or extent of gene expression. Gene expression is the process in which a gene's	BP	yellow	regulation of gene expression	0.000894	BTG1,DMTF1,SATB2,FOXP1,NDFIP2,BANK1,ACTR5

	<p>coding sequence is converted into a mature gene product or products (proteins or RNA). This includes the production of an RNA transcript as well as any processing to produce a mature RNA product or an mRNA or circRNA (for protein-coding genes) and the translation of that mRNA or circRNA into protein. Protein maturation is included when required to form an active form of a product from an inactive precursor form.</p>					
GO:0051171	Any process that modulates the	BP	yellow	regulation of nitrogen	0.001611	BTG1,DMTF1,SATB2,FOXP1,NDFIP2,BANK1,ACTR5

	frequency, rate or extent of the chemical reactions and pathways involving nitrogen or nitrogenous compounds.			compound metabolic process		
GO:0005654	That part of the nuclear content other than the chromosomes or the nucleolus.	CC	yellow	nucleoplasm	0.001721	BTG1,DMTF1,SATB2,FOXP1,FANCL,ACTR5
GO:0080154	Any process that modulates the rate, frequency or extent of fertilization. Fertilization is the union of gametes of opposite sexes during the process of sexual reproduction to form a zygote. It involves the fusion of the gametic nuclei (karyogamy) and cytoplasm	BP	yellow	regulation of fertilization	0.001817	BTG1,DMTF1,SATB2, FOXP1,NDFIP2,BANK1,ACTR5

	(plasmogamy).					
GO:2000225	Any process that stops, prevents, or reduces the frequency, rate or extent of testosterone biosynthetic process.	BP	Yellow	negative regulation of testosterone biosynthetic process	0.002856	BTG1,DMTF1,SATB2,FOXP1,BANK1,ACTR5