Synaptic Protein-Protein Interaction Networks for insights into Schizophrenia

Nicole Ferguson

4th Year Project Report
Computer Science and Physics
School of Informatics
University of Edinburgh

2024

Abstract

Schizophrenia is a psychiatric disorder affecting 0.45% of the adult population. It can devastate a person and affect their mood and actions through psychosis which can involve delusions and hallucinations. Schizophrenia is associated with brain dysregulation. The brain is a complex organ, which contains billions of neurons and trillions of synapses. The synapses can be modelled using protein-protein interaction networks, this was done for the post-synaptic density network to try and gain insights into schizophrenia.

This dissertation uses a bioinformatics approach to analysing pre-existing synaptic data and comparing it to recently found schizophrenic genes using the novel tool BioNAR. It was found that BioNAR successfully identified useful schizophrenic associations that were consistent with current findings.

A full and consensus post synaptic density protein protein interaction network was compared to a network made only of genes with a SynGO annotation. It was found that by using data with more noise but still confirmed by using consensus genes that a greater number of associations were found than those compared to the reduced SynGO only network, and a greater quality of associations were found than those compared to a full network.

Research Ethics Approval

This project was planned in accordance with the Informatics Research Ethics policy. It did not involve any aspects that required approval from the Informatics Research Ethics committee.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Nicole Ferguson*)

Acknowledgements

First of all, I would like to thank my dissertation supervisors, Professor Douglas Armstrong and Dr Oksana Sorokina, who have made this dissertation possible. They supported me through this project with their insights, patience, and knowledge. They made this project possible.

Secondly, I am grateful for my supportive, encouraging, and caring partner, Dermott, who always stood by my side.

I would also like to thank my friends, who have been with me every step of the way through many long days in Appleton tower. This camaraderie made it possible.

Lastly, I would like to thank my family, for their unending and unwavering belief in me and the encouragement that came with that. To my mum and dad, thank you for your endless support, not just in this project, but throughout university.

Contents

Chapter 1

Introduction

Schizophrenia is a complex psychiatric disorder which strongly affects a person's mood, behaviour and mind. It often manifests in young adulthood, between the ages of 18-25, whilst the pre-frontal cortex is still developing. The brain is a massively complex structure made of a enormous network of around 100 billion neurons. The connections between these neurons are called synapses, of which there are approximately 100 trillion in a brain, and they themselves are equally complex structures [3].

There is evidence that synapses are associated with Schizophrenia. Synapses play a pivotal role in shaping the brains functions and therefore a persons behaviour [20]. These synapses are controlled by protein-protein interactions (PPIs) which in turn determine brain function via neurotransmission regulation, organisation of the Postsynaptic Density, scaffolding proteins and many other roles [66]. By creating PPI graphs, annotating the data and performing bio-informatic analysis, we can gain insights into the enriched genes found and their associated disease.

The main technology used in my dissertation is BioNAR [33]. Since it was recently published in 2023 there have been no published peer-reviewed works yet making use of this software. The primary aim of my dissertation is to look at the influence of recently collected Schizophrenic genes and their impact on the Protein-Protein Interaction Network by making use of a new tool, BioNAR, and seeing how it compares to current research. The next goal is to compare a full network to one with SynGO only genes. There are a few main steps to this:

(1) Create the Protein-Protein Interaction (PPI) Networks and annotate the network with schizophrenic genes.

(2) Analyse the PPI Networks by clustering and finding common clustering community structures.

(3) Run ORA enrichment analysis and consider the results.

(4) Compare the Full and SynGO networks results.

1.1 Ethical Considerations

The research was planned with the Informatics Research Ethics policy, however, I still find it important to consider the ethics of this project further. A significant portion of the data comes from animal research, all of the data was complied as part of the Human Brain Project. The Human Brain Project is a research project focused on mapping the human brain, it has its own ethics board and considerations. Furthermore, the data was originally published in various journals which all have their own ethical standards and considerations, which were verified by the Human Brain Project. Edinburgh University has contributed to the Human Brain Project and as such, follows Edinburgh University's Ethical guidelines.

1.2 Previous Work

There are many studies that already look at the schizophrenic associations of genes, however none that make use of BioNAR. BioNAR was only recently released and as such there is only one paper which is currently published that cites it (excluding the paper that gives the main use cases), however, it is only in the pre-print stage at the moment [52]. The paper makes use of the BioNAR's DYNAMOs algorithm, which is not used in this dissertation.

The main comparison dataset is from Trubetskoy et al., has been cited 1070 times [61]. This large number of cites could indicate the impact it has had on the scientific community, other papers that cite it validate it as useful dataset. This dissertation will allow for a newly developed tool to be properly tested and also help confirm and compare current Schizophrenia research.

Much of the literature published about schizophrenic PPIs focus on the reveal of new PPIs via genome-wide association studies (GWAS) which study the entire genome, looking for small differences (called Single Nucleotide Polymorphisms - SNPs) [61]. The most recent literature about schizophrenic PPI networks focuses on in vivo mapping, however it remains in pre-print [31].

One similar paper created an in silico PPI network, using genes from the the Psychiatric Genomics Consortium (PGC) [50]. This paper has a broader focus involving the entire brain and no specific synapse region, compared to this dissertation which has a narrower scope which focuses on the post synaptic density (PSD) part of the synapse.

Chapter 2

Background

This chapter will provide a literature review of current schizophrenia research, synaptic protein-protein interactions and their networks, and give a brief insight into some of technologies I will be used throughout this dissertation. In order to provide a relevant background to why this research is important I will explain what schizophrenia is. As much of the disease is thought to be influenced by the synapses, I will then cover synapses and the protein protein interactions within them. As I used BioNAR to analyse protein protein interactions via PPI networks, BioNAR and its methods and basic network information is explained. There will be a lot of terms and biological knowledge used throughout, in order to aid explanation there is a table included at the end of this chapter, at section 2.5, it may be useful to reference this throughout.

2.1 Schizophrenia

Schizophrenia is a disability which can greatly impact ones quality of life. It is often characterised by psychosis, meaning the person may experience difficulties separating their own thoughts and ideas from reality. It is estimated that approximately 0.45% of adults world wide suffer from schizophrenia [39].

2.1.1 The Disease

There are many factors that can lead to increased likelihood of developing Schizophrenia. The risk of developing Schizophrenia is 27% for someone with two schizophrenic parents. Contributing factors are not solely genetic, as environmental factors can have a large part. Prenatal environment can have a large effect, if the mother has an infection disease especially in the 2nd trimester of pregnancy it can increase risk of developing Schizophrenia [14]. The use of twins studies showed that if fetal circulation was shared for mono-zygotic twins it could increase risk up to 60% [13]. One model which showed the heritability of Schizophrenia as up to 81% showed that environmental factors could increase risk [57, 60].

2.1.2 Fine-mapping to identify genes

Fine-mapping was used to identify genetics variants that were commonly found in Schizophrenia data. Fine-mapping is the process of separating genetic variants by using a defining trait (phenotype) as a marker. These genetic variants that are identified are likely to have an influence on this defining trait [49]. The genes that are found are SNPs that differ from an ”expected” genome and are used to tag the trait. Its worth noting that just because an SNP is found often with a defining trait, it does not mean it will cause it [61].

2.1.2.1 Schizophrenic Associated Genes

The paper which the schizophrenic genes we will be using throughout are from the paper Trubetskoy et al. (2022) [61]. These genes from now on will be referred to as SCZ genes. The study made use of GWAS studies in order to find SCZ genes, it compared 76,755 individuals with schizophrenia to a control group of 243,649 non-schizophrenic individuals.

There were two types of genes mentioned, broad and prioritised. The broad list also encompassed the prioritised genes. The prioritised genes were those that had a greater confidence that the genes are associated with schizophrenia. The broad genes are genes that whilst not as confident as the prioritised genes, involve a more comprehensive coverage of possible variations found and that could be SCZ associated.

Of the 106 prioritised protein coding genes found with Trubetskoy et al. (2022) study, only 15 have pre-existing SynGO annotations. 7 of them were postsynaptic which follows current schizophrenic research trends of the postsynaptic synapse having a strong significance. However, as mentioned in Trubetskoy et al. (2022) there are also suggestions that the presynaptic may be involved as well.

2.2 Synaptic Protein-Protein Interaction

2.2.1 What is a protein?

Proteins are specialised molecules within the body that perform certain functions. Their functions depend upon their shapes, which are determined by the gene sequence. The genes determine what sequence of amino acids are combined to create a protein. A proteome is the complete set of proteins that is expressed by a particular cell type [42].

2.2.2 Synapses

Synapses pass information from one neuron to the next.

They consist of 5 elements, a presynaptic element, a postsynaptic element a synaptic cleft, astrocytic endfeet, and an extracellular matrix [19]. For the purposes of the this dissertation, the emphasis will be on the pre and post synaptic element and as such the explanations will be reduced to focusing on those elements. Synapses convert the electrical impulse from the neuron with the presynaptic element into a chemical one

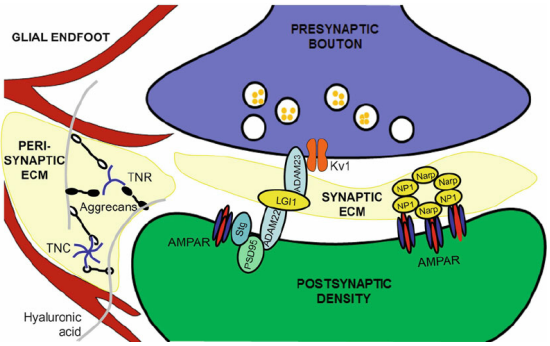


Figure 2.1: Diagram of the synapse that includes the presynaptic, postsynaptic and other synaptic parts. It shows many PPIs that are involved at the tertrapartite synaptic level. The image was taken from Korotchenko et al. (2014) [28].

when passing it on to the next neuron with a postsynaptic element. That chemical impulse is converted back into an electric one again. This is a unidirectional process [66].

2.2.3 Proteins and Protein-Protein Interaction

Protein-Protein Interactions (PPI) involve at least two protein molecules that interact through a biochemical interaction. These PPI interactions are ”Physical Contacts of High Specificity”, which involve region(s) of a protein binding to region(s) of another protein to perform specific functions.

2.2.3.1 Synaptic Protein-Protein Interactions

The brain is a complex mass of neurons which use synapses, the synapses modify signals to pass along information between neurons [3]. The process of a neuron passing along information involves a great number of PPIs.

The process starts at the presynaptic end: a signal is received by an electrical impulse called the action potential. This action potential excites the presynaptic neurons vesicles by causing a depolarisation of the neuron. This depolarisation leads to a change in membrane potential and causes the calcium ion channels to open, allowing calcium into the vesicles. This mobilises the vesicles to prepare them for the neurotransmitter release. There is a protein protein interaction here, where due to the calcium interaction, an enzyme (also a protein) causes a protein called synapsin to separate from the vesicles, which is what mobilises them for release. Once they are mobilised at the post synaptic end, there are more proteins that are involved in a PPI. This family of proteins is called SNARE proteins, they help the vesicles get in contact with the cell membrane allowing the two membranes to fuse together [66, 55, 56].

2.2.4 Protein Interaction Networks

Many proteins work together and are near each other which can be modelled using proteins networks. This makes use of graph theory, having each protein act as a node

and the connections between them being the vertices or edges. These vertices or edges will represent a Protein Protein Interaction. Each vertex between the nodes can have a weight, allowing you to model what proteins actually interact with each other more effectively, however, most commonly, including in this dissertation, unweighted PPI Networks are used. Centrality is very important, as it helps inform what proteins are potentially important. Centrality is measured in numerous ways such as having a higher a score when you add the weights of the neighbours, if it has easy access to other nodes quickly and if its necessary to go through to get to other nodes. More network information will be discussed soon in section 2.3. This is as the more connections a protein has, the greater the likelihood the protein will influence more processes [43, 33].

2.2.5 Clustering

The size of Protein Interaction is massive, often thousands of proteins will be in a network. This can make traversing the graph tricky and time consuming, it is then useful to make use of PPI networks. This involves grouping proteins into clusters based on their connecting architecture, as it is assumed that the shared interconnectedness will imply a shared function. Different algorithms are used to try and identify these communities within these PPI networks, often they are greedy, meaning they make the best decision locally rather than globally. This may be necessary for such large networks [33]. The algorithms used in this dissertation will be discussed further at 3.4.1.

2.3 Networks

The networks generated in this dissertation were Protein-Protein Interaction Networks, each node represents a protein coding gene and a vertex represents an interaction between the proteins that the genes code for. Unlike some other PPIs, the networks used were all undirected and unweighted. The following definitions are useful to know for this dissertation as network analysis will be used throughout [38].

Degree: The number of edges a node is connected to. Given this is an undirected network, there is no in-degree and out-degree for this dissertation, simply a degree.

Distance: The number of vertices that connect two nodes.

Shortest Path: Related to distance, however this is always the minimum number of vertices that connect two nodes.

Centrality: How important a node is, there are many different types of centrality such as Betweenness, Degree, Semi-Local and PageRank.

Betweenness Centrality: It measures the amount a node is on the shortest path between other nodes.

Degree Centrality: The number of vertices that a node is connect to in a network and how important it makes it. The more connections a node has, the higher the degree centrality.

Semi-local Centrality: This very similar to degree centrality, but is limited to a cluster or other nearby nodes.

Page Rank: This ranks a node based on its connections, it ranks how important the node is, but also how important the nodes its connected to.

Bridgeness: This can indicate how much a node acts as a bridge to other clusters within a network, it can indicate if there are genes that perhaps act as a bottle neck between clustering communities.

Clustering Coefficient: How likely nodes are to form interconnected clusters.

2.4 Software

2.4.1 Synaptic Database

The database used was the Synamptosome Proteome Database (Synaptome DB), it is made up ontological information for protein coding genes specific to the mammalian synapse. The data is made up of 58 published synaptic proteomic datasets which were curated by experts [53]. The database is data-driven, the data is gathered from many papers. Whilst it may have more noise than other databases, it is much less biased against under-annotated genes. As these genes were already pre-curated and verified, no further verification was required and no genes were discarded for this paper.

Not all SCZ genes are included with the Synaptome DB. A few of the genes overlap, this overlap will be the schizophrenic associated genes that are used within this dissertation.

2.4.1.1 SynGO

SynGO is collection of synaptic genes, the genes contained are very well documented and annotated [27]. Synaptome DB has a subset of genes that have SynGO annotations, and since SynGO is also curated and maintained by experts, there is an even greater confidence and trust in these genes annotations. Whilst these genes are of a high quality, it can lead to an issue of bias. Researchers more commonly favour less noisy, more highly annotated datasets like SynGO. This can bias research, as under annotated genes may play a larger in some diseases pathology but could be missed due to this bias.

2.4.2 BioNAR

We will be using BioNAR to perform clustering within this dissertation. BioNAR is an analysis package, it is especially useful in proteomic networks as it can predict a proteins impact within multiple complexes [33]. BioNAR is split into seven steps:

1. **Creating a network instance:** Using R data frames a network is constructed, with each row of the data frame representing a vertex interactor pair. Each row makes use of vertex_ID as its key for in its key-value pair.
2. **Adding annotation to the vertices:** Vertices are annotated with metadata, which can contribute to vertex attributes to determine the weight scores in later steps.

3. **Estimating network vertex properties and underlaying structure:** Vertex centrality is measured possibly using some of the metadata from before. The following measurements can be used according to BioNAR, ”degree (DEG), betweenness (BET), clustering coefficient (CC), semilocal centrality (SL), mean shortest path (mnSP), page rank (PR) and standard deviation of the shortest path (sdSP)”.
4. **Clustering:** Algorithms used for clustering are based on Modularity-Maximisation. Multiple clustering algorithms can be used simultaneously and a user can select what specific clustering algorithm they wish to use.
5. **Bridgeness and identifying ”influential” vertices:** As every vertex is placed into a single cluster, and as such, bridgeness must then be used to see if a vertex is likely to belong to more than one community at once. A consensus matrix is used to determine this bridgeness.
6. **Studying the overlap or separation of annotation pairs:** To ensure that points of intersection were not due to random chance, the observed mean shortest path for two distinct annotated networks are compared to a randomly annotated network.
7. **Enrichment Analysis:** Gene enrichment is used to identify what genes are over-represented within a network. If a gene is over-represented that means it can be associated with the desired phenotype that the network is made for. This does not mean it necessarily causes it, just that it is associated with it.

2.5 Terms and Definition Table

Term	Definition
Broad	This is a term specific to schizophrenia genes, in context of this dissertation these are genes that the Trubetskoy et al. (2022) study found during their fine-map [61].
Prioritised	This is another schizophrenic specific term, here these are the most credible genes found in the Trubetskoy et al. (2022) Study [61].
Full	This means all genes. For the purposes of this dissertation, Full can be full with some criteria e.g. Full PSD Network, can be all genes that also meet the PSD criteria.
Consensus	These are genes that were identified at least once in a minimum of two different papers.
GO	Gene Ontology, which is the standard Bioinformatics version of gene data. The GO term can tell us the biological or molecular function that are known to be associated with a gene [4].
SynGO	A synapse specific version of GO, they contain annotations of synapse specific functions [27].
Biological Function	The role this gene or protein is associated with in the body, this will be the task that it fulfills. E.g. Neuron transmission [59].
Molecular Function	The biochemical role that specific molecules perform. E.g. Enezymes [59].
Entrez ID	Unique Identifier for genes, these are integer numbers that are standardised across bioinformatics [30].
SYN	The Synaptosome. An isolated vesicle created in vitro that contains the presynaptic terminal and postsynaptic membrane and density [5].
PSD	Post-synaptic density. A protein network that has many important purposes such as regulating the structural changes of the Post-Synaptic component [10].
PRES	Presynaptic. Similar to the postsynaptic density, the presynaptic region involves protein network that has many purposes such as neurotransmitter release, and regulating signalling [22].
SCZ	Schizophrenic associated genes. For the purpose of this paper these genes are the ones were found in the Trubetskoy et al. (2022) study.

Table 2.1: A small section of terms used throughout and what they mean. While most of these definitions are standard, there can be small differences between papers and these terms are specific to this paper.

Chapter 3

Methodology

In this chapter I will be covering the methods that I used throughout this dissertation, this includes creating the network, annotating it and comparing clusters and algorithms.

To aid reproducibility, the code created and data generated during this dissertation is available on GitHub: https://github.com/NicoleF02/Dissertation

3.1 Programming

Throughout this project a mixture of R and Python were used.

R 4.3.1 was used for the majority of the ”heavy lifting”, this meant calculating things like the network graphing and clustering via BioNAR and the data coming from Synaptome DB [33, 53].

Python 3.9.13 was often used to further refine, process and graph data. Often external libraries were used such as Scikit-Learn, Matplotlib, Pandas and Seaborn which were used for graph creation and data processing [40, 25, 58, 32, 64].

For the Venn Diagrams generated, Python and R were used interchangeably, depending on where the data was last used. For R VennDiagram was used, for Python Matplotlib Venn was used [16, 21].

For visualising the networks and clusters Gephi was used throughout, along with the expansion Circle pack [7].

3.2 Cluster Comparison

Cluster comparison is one of the most important parts of the dissertation: identifying what makes a cluster good and what is consistent between different algorithms can help tell us more about what may be a stronger clustering. A stronger consistent clustering can give us insights into what traits and attributes may be associated with the enriched genes in the cluster and therefore what traits and attributes may be associated with the phenotype being studied.

3.2.1 Jaccard index

In order to compare enriched clusters similarity between different algorithms the jaccard index was used. It only considers the intersection of sets between both sets [15]. The value returned is between 0 and 1, where 0 means it has no common points, and 1 means they are identical. The formula is defined as such:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

This is quite a simple comparison, as it is only a score based on how many they have in common over both datasets, this means other methods to compare clusters should also be used.

3.2.2 Adjusted Rand Index

The adjusted rand index (ARI) was used for overall cluster comparison between algorithms. ARI is often used in machine learning to compare true points and predicted points, it requires lists of the same size so was used for overall algorithm comparison as clusters were different sizes. We were able to arbitrarily choose which algorithm was a proxy for the true one and which is the predicted one as ARI is symmetric, so it would not affect the results. The results given are from -1 to 1, where -1 is disagreement between clusters, 0 is random chance and 1 is agreement [63, 46, 24].

$$ARI = \frac{RI - ExpectedRI}{MaxRI - ExpectedRI}$$

RI is the Rand Index which is given by:

$$RI = \frac{a+b}{\binom{n}{2}}$$

Where a is the number of matching true positives - cluster points exactly agree between algorithms. b is the number of true negatives - data points are not included in either compared clusters, however may still differ.

3.2.2.1 My Modifications

Given the data points are stored as a graph and don't have a direct x-y axis which ARI requires, I created a few heuristic modifications in order to successfully work with this type of data. The steps that I took were as such:

(1) Two algorithms clusters were directly compared using the Jaccard Index, these results were added to a table.

(2) The table was ranked from most to least similar.

(3) Going through the table, we start at the top. Renaming the most similar clusters in both algorithms to n, all rows that contain the clusters that were renamed are removed

from table. This is repeated, renaming the clusters to n+1 and so on, until the table is empty.

(4) If the table is empty but some original cluster names remain, they are renamed in sequence. This occurs when algorithms have different cluster numbers.

Those steps were ran on all combinations of algorithms. This type of modification was required as ARI would purely be ranking them based on cluster naming, rather than how similar each cluster actually is. The Jaccard Index was reused to compare compatibility again as it does not require clusters to be the same size like ARI does.

3.2.2.1.1 Limitations There are a few limitations to this new approach. Given the ARI normally requires an x-y axis and my modifications allow the Entrez ID and Cluster number to act as a pseudo x-y axis, algorithms with vastly different cluster numbers will perform very poorly as for one cluster the gene may be in cluster 12 but the other cluster 300. It also fails to consider when one cluster is similar to two others in another algorithm, and drops whichever is least similar. Due to this, algorithms that may actually be similar but perhaps split one larger cluster into two smaller ones, are seen as much more vastly different than they may be.

3.2.3 Modularity

Modularity is another way of evaluating clusters. It considers how densely connected nodes in a network are and gives them a score. The score is between 0 and 1, with 1 being highly modular and having a network with dense connections within the cluster and few connections outside of its cluster [37].

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Gives the formula for modularity Q, where m is the total number of edges in a network, A is the adjancency matrix where A_{ij} will have 0 or 1 as a value depending on if a node is present, k_i is the degree of node i, and δ is the Kronecker Delta.

BioNAR has inbuilt modularity functionality which is what was used throughout.

3.3 BLAST

The Basic Local Alignment Search Tool (BLAST) is used to analyse how well biological sequences align with each other [**boratyn˙domain 2012**, **altschul˙protein 2005**, 1]. BLASTP is the protein specific BLAST version that was used in this dissertation. Blast returns 3 important values:

E Value: Score that the similarity occurs by chance. The lower the E value, the less likely the proteins are similar by chance.

Query Cover: The coverage of the sequence that the BLAST query covered by the alignment.

Identity: Similar to cover, however, it is the percentage of identical amino acids between the two proteins.

BLASTP is useful in order to see how similar protein sequences are, it can help see if proteins may have a similar function based on their structures.

3.4 BioNAR and Synaptome DB

The backbone of this dissertation and what has allowed for all the networks to be generated and clusters created is BioNAR with Synaptome DB which were facilitated by BioConductor [33, 53]. BioNAR was used to generate the network, annotate it and do enrichment analysis. The data used for the network was taken from Synaptome DB.

The network was generated using the genes from Synaptome DB and then created by BioNAR. Enrichment Analysis was done using ORA through BioNAR.

3.4.1 Algorithms

There were many algorithms that were inherently included with BioNAR through R's iGraph Package [18]. They work in many different ways to cluster the data, this was important to do in order to help find common community structures. This kind of clustering consistency can ensure robustness and help validate the cluster further. Having such a wide variety of clustering algorithms that work in different ways also help discover other data groupings or outliers that may have been missed by others.

Walk-Trap (wt): This algorithm is inspired by random walks, often during these random walks you can get stuck among the same nodes, this helps determine the clusters [44].

Spin-Glass 1,2,5 (sgG): Based on a spin-glass from statistical physics. A spin glass has atoms with magnetisation, they are disordered so the magnetic state is random with atoms having a small amount of magnetisation in different directions. As the temperature of the the spin glass cools, and the energy of the system decreases, there is more order in the magnetisation.

The spin glass algorithm treats each node like the spin in a spin glass. You start at a "high temperature" and gradually lower the temperature. This can be repeated until a solution is found or a stop condition is met. Each node in the network has a spin, and when the solution is found each spin corresponds to the a cluster. [47]

Louvain: This heuristic community detection algorithm works through modularity optimisation. It initially assigns all nodes to their own community, it then considers its neighbours and moves them the neighbour with the highest modularity score. If there is no increase then it stays with its initial community. This is repeated this until modularity can no longer be improved. These clusters are now treated like large nodes, weighted by the sum of the weights of other nodes (though in this dissertation we use unweighted nodes), and the previous steps are applied again until once again modularity can't be optimised further [9].

Spectral: This is the only clustering algorithm that is not included from the iGraph Package and is implemented via BioNAR. Spectral here is a modularity optimiser [34]. First, it creates a modularity matrix between data points, the Laplacian is computed and then a spectrum of eigenvalues and eigenvectors is created. The dimensionality is reduced, then finally, the nodes are clustered based on the eigenvectors.

Fast-Greedy Community (fc): Similar to louvain, fast-greedy community focuses on modularity optimisation and works in a similar way by merging clusters based on modularity gain [17]. However, unlike louvain, fc makes use of a hierarchical approach hence why louvains larger clusters may be similar to many smaller fc clusters.

InfoMAP: This algorithm is similar to walktrap as it is also based on random walks through a network however also utilises information theory to aid in cluster creation. It focuses optimising clustering by trying to minimise the descriptive length of each cluster [48].

There was another algorithm included but was not used in this dissertation, the Leading-Eigenvector (lec) clustering algorithm. This was due to the consistent failure to converge for the larger network. The lec algorithm is also a spectral based algorithm, so considering spectral was already included, and given the wide number of diverse algorithm clustering methods, lec not working was not a large loss.

3.4.2 Enrichment Analysis

Over Representation Analysis (ORA) is another feature of BioNAR. Intially created by Boyle et al. in 2004, ORA was first used only for GO annotations [12]. Here though, it considers whether a set of genes (we also use it for annotations later down the line), are over-represented within the larger network. It calculates the p-value using the hypergeometric distribution mentioned below:

$$P = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

For this dissertation, M is all of the genes in the network, N is the number of genes in the cluster, n are the genes we are looking at (e.g. those with same GO annotations or SCZ broad genes), and k is the number of genes we are looking at within that specific cluster.

3.5 Bugfinding

BioNAR is a very recent development, and as such, various bugs were found and reported during the creation of this dissertation. My reports of these bugs have allowed for them to be fixed and these fixes have been released under the latest version of BioNAR, published on 29-03-2024 [35]. The fixes will be standard when the next version of Bioconductor comes out, though, for the moment the working version would need to be specifically downloaded.

3.5.1 Deprecated Function

The most recent publication on how to use BioNAR with Synaptome DB from McLean (2023) relies upon an older deprecated function from the Synaptome DB package [33]. This function is not available with the most up to date version of BioNAR. Whilst a deprecated function isn't strictly a bug, the fact that this paper was published after it the function was deprecated is relevant. This is important as Bioconductor will use the latest version available, and as the paper is the main tutorial that a new user would use, BioNAR could seem non-function to them.

I successfully identified the function `buildFromSynaptomeGeneTable` as deprecated and found and replaced it with `graphFromSynaptomeByEntrez`, however, there were a few differences in terms of functions. For the function `buildFromSynaptomeGeneTable`, it returned the each of the nodes ids as Entrez IDs, but for `graphFromSynaptomeByEntrez`, it returned each of the nodes with a ID from 0 to n, with the Entrez IDs as a separate "name" column. This lead to some confusion, as the some of the later functions used to annotate the graph applied them to the IDs rather than the "name" column, once this was identified the rest of the functions ran smoothly. This should also emphasise why this was such an important thing to mention, the different functions returning things differently could lead to the wrong genes being annotated and then an entire paper being functionally useless due to this possible user error.

3.5.2 Sigmoid Function

I successfully identified the presence of an important bug for the function `fitSigmoid` from BioNAR. I also identified the exact lines that were wrong.

I identified a bug with it regarding which columns it called and when the correct columns were called it still had an error even when I fixed the original bug. There was a greater bug that occurred about failing to converge once the correct column was called. The bug was reported and accepted and is currently being fixed. Due to this bug, other clustering comparisons were created to compensate.

3.5.3 ORA Enrichment

Another bug I successfully identified was regarding the ORA function from BioNAR. For the function `clusterORA`, it failed when "" was an option. This bug was found as it failed to find any enrichment when I set a column as having two possible values, "TRUE" or "". This was not mentioned in the documentation.

Chapter 4

Preparing the Data

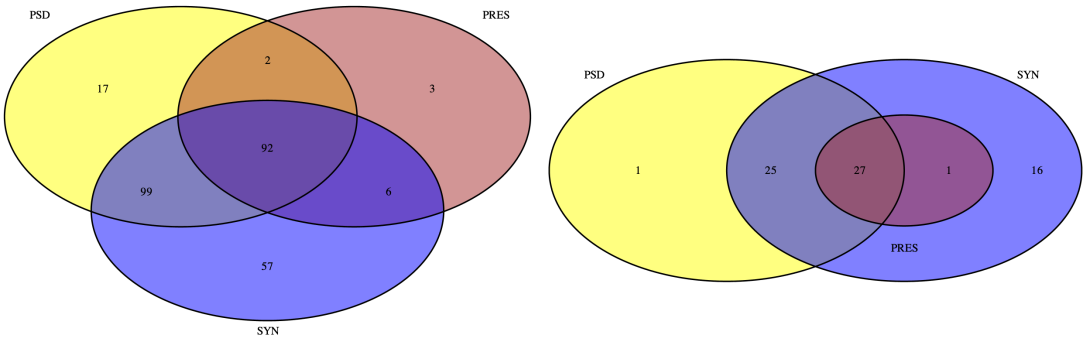
The largest dataset was collected from Synaptome DB, Venn diagrams were generated in order to decide what part of the synapse to look at. The SCZ genes were collected from Trubetskoy et al., these were Broad and Prioritised genes [61].

An undirected and unweighted network graph was created using BioNAR to simulate the post synaptic density part of the synapse. The vertices were informed by the PPI interactions found in Synaptome DB. BioNAR was then used in order to cluster the graphs, using the algorithms mentioned in 3.4.1. Enrichment analysis using ORA was ran on every algorithm for SCZ broad, SCZ prioritised and also all other annotations of the graph including biological and molecular function. These results will be analysed in this chapter below, with the exception of some of the enrichment from annotations, which will be explored later on.

4.1 Synapse Gene Prevalence

In order to get a clearer path for this dissertation, the data from Synaptome DB was processed and compared to the genes from the Trubetskoy et al. study [53, 61]. The Synaptome DB data was split into 3 sections from the different synaptic regions: the Postsynapse Density (PSD); the Pre-synaptic Density (PRES); and the Synaptosome (SYN). All data in Synaptome DB is annotated with at least one of those labels, those that had more than one label were included in all applicable lists.

I compared the SCZ genes to SynaptomeDB genes to see where the largest overlap was, this allowed us to reduce the network size and also focus on a particular part of the synapse. Out of the 469 SCZ Broad genes only 276 were present in the Synaptic Database, 58 of these also were associated with SynGO. For the SCZ Prioritised genes there were 104 total, however 103 were present in the database.



(a) SCZ Broad Coding Genes (b) SCZ Prioritised Coding Genes

Figure 4.1: Venn Diagram of the overlap of the genes found both from the SCZ Gene list and Synaptome DB genes. It shows the gene associations with the synapse areas, PSD is postsynaptic density which is yellow, PRES is presynaptic density which is red and SYN is the synaptosome which is blue. A larger version can be found in the appendices at A.1 and A.2.

The Venn diagrams found in Figure 4.1 were used in order to decide a useful synaptic area to look at. Whilst the SYN region does have the most genes found in the venn diagram, it was disregarded. This is because it is much more complex then either of the other regions, the synaptosome contains parts of the presynaptic synapse and postsynaptic synapse [65]. This kind of broad approach can also be too much to fit into just a dissertation and more can be achieved using a narrower scope.

The next region with the most genes was the Post Synaptic Density, by using this it offers a more focused approach, allowing for a more in depth and precise analysis later on. It was also mentioned before in Trubetskoy et al. (2022) that the data did point more towards a PSD pathology [61].

4.2 Data Clustering

The network was then built using BioNAR. This was done initially on the full Postsynaptic network, then a consensus network was then generated that included genes that were found from more than one paper. The information about each network can be found at table 4.1. From now on, when we refer to the full set of genes or full network, it will be exclusively in reference to the PSD genes list or PSD network unless explicitly stated otherwise.

After these networks were generated, there were even less SCZ genes present. This was due to the function FCC, which found the largest connected component then removed all nodes that were not connected to the most connected part. This brought the SCZ Broad 276 left for the full PSD network down to 191, and the SCZ Prioritised from 103 to 46.

Network	Nodes	SCZ Broad	SCZ Prioritised	Edges	Avg. Degree	Avg. Diameter
Full PSD	4843	191	46	28225	11.66	8
Consensus PSD	3026	132	34	14480	9.57	9

Table 4.1: Graph information for the Full and Consensus Postsynaptic networks. The nodes are all from the largest connected component, those without an edge to the largest connected component were discarded. The number of nodes and edges are given, as well as average degree of the nodes in the network, the average diameter of the network and the number of SCZ genes found in each network. These can be visualised at 4.3 and A.9.

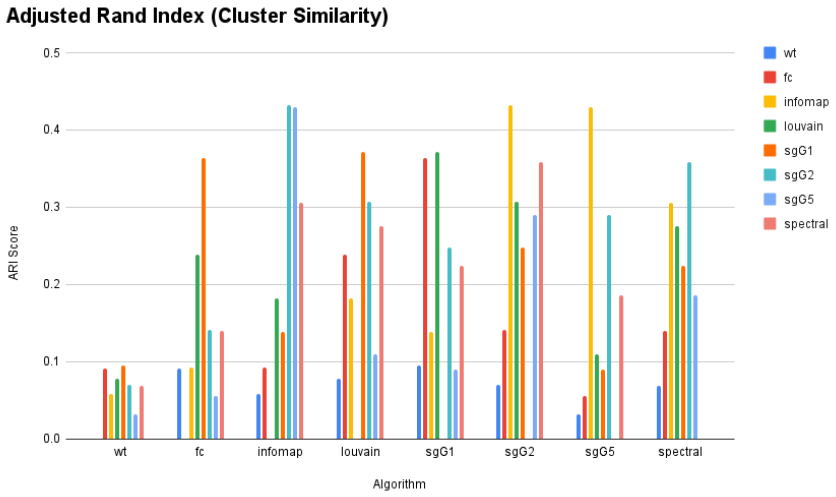


Figure 4.2: The Adjusted Rand Index Score for the Full Network. It gives a score from -1 to 1 for how each algorithm is similar to each other. The Adjusted Rand Index was ran with the modifications mentioned in 3.2.2

The network was then clustered using different kinds of community finding algorithms via BioNAR. A good algorithm must be determined in order to properly analyse the data efficiently. One of the very first things that can be used to determine a good algorithm is just visual inspection as well as considering cluster sizes.

The modularity score was used to determine the quality of the clusters available for every algorithm, this was done using methods discussed in section 3.2.3. As can be seen from table 4.2, wt has the lowest modularity score overall, this can be seen visually in 4.3d as it has many clusters that have 4 or less with its median cluster size being 1.

A broader general algorithm comparison was ran for the full network, using the ARI with my modifications mentioned at 3.2.2. The figure 4.2 shows the algorithm comparison, the consensus one was similar so was left out of the main text. ARI was ran to get a general algorithm comparison, it can be useful as it allows for a direct comparison between algorithms. If algorithms are consistent without using similar methods, these algorithms can make the findings from them more robust and trustworthy.

Fast-Greedy Community despite its relatively high modularity score is not a ”good” algorithm, this was due to the large difference in mean and median cluster size. It has many small clusters, with the median being 4, however it tends to make a few very large clusters and lots of tiny ones which is not that useful in terms of analysis.

ORA enrichment analysis, in section 4.3, is not very well suited for dealing with smaller clusters. ORA could struggle to identify enrichment of annotations (or even just the SCZ genes) later on due to the smaller clusters providing less information than the bigger ones.

Walk-trap is also not a ”good” algorithm due to similar reasons as fc, it has one significantly bigger cluster and a majority of tiny clusters that are made of only one gene. This can be easily visualised at Figure 4.3d. This is not useful for enrichment analysis as it defeats the purpose of clustering in PPI networks as mentioned in section 2.2.5. The clusters may just be that size however this algorithm and fc skew more to extremes when compared with the others and as such, I don’t believe all these clusters should be so small.

Figure 4.2, also shows how different walk trap is from the rest of the algorithms. Whilst this is probably unfairly biased against it due to the fact that walk trap has so many more communities than the others, walk trap is just different than the others. This can easily be seen visually at 4.3. This confirmation can show that wt is not a good algorithm for this data and should be later disregarded.

Despite having relatively similar clustering methods, louvain and infomap performed quite differently, with $\approx$ 0.05-0.07 difference depending on the network. Infomap performed poorer than louvain, however it has many more clusters. As seen in figure 4.2, infomap is very similar to sgG1, sgG2. Despite infomap having a lower modularity score, it is most similar to the consistently most modular algorithms. Despite it having a lower modularity score, it is not as egregious as wt, and as such, remains an okay algorithm.

Given that infomap is quite similar to some of the best performing algorithms for modularity and has the highest amount of similarity overall, I believe it is a useful algorithm. It also has many enriched clusters that are of a good size as found in 4.4.

4.3 ORA enrichment

ORA enrichment was carried out for both of the networks. This helped identify enriched clusters. These are clusters with SCZ genes that appear more than random chance. A visualisation of the SCZ genes that are found in each cluster can be found at figure A.3

As seen in Table 4.4, the fc algorithm did not have any enrichment, so it was disregarded for the rest of the dissertation and not considered at all in other networks as it would not allow for a comparison later on.

The wt algorithm once again disappoints, with the majority of the enrichment coming from single sized clusters which aren’t statistically significant in both the full and

Algorithm	**No. Clusters**	**Mean Cluster Size**	**Median Cluster Size**	**Modularity Score**
sgG1 (Spin-glass 1)	25	193.72	17.0	0.422611012198729
louvain	14	345.93	294.5	0.41808612715266497
sgG2 (Spin-glass 2)	52	93.13	48.0	0.4088621131753962
spectral	101	47.95	2.0	0.37196844727260453
fc (Fast-Greedy Com-munity)	36	134.53	4.0	0.3688635077641471
sgG5 (Spin-glass 5)	125	38.74	32.0	0.34025541372041235
infomap	289	16.76	9.0	0.35887106346022163
wt (Walk-trap)	862	5.62	1.0	0.28486742165048135

Table 4.2: Information about algorithms generated for the full graph using BioNAR. They can be seen visualised at Figure 4.3. It includes the algorithms shorted and full names, the number of clusters each algorithm has, the mean and median cluster sizes and the modularity score. The table is sorted from highest to lowest modularity score.

consensus networks. This can be seen in Table 4.4. These results are not notable and the wt algorithm will be disregarded from this point.

One thing that was consistent across networks was sgG1 giving exclusively False enrichment. This can be seen Tables 4.4 and B.2. This negative enrichment could indicate that an absence of genes could be statistically significance. It could indicate that a suppression of certain genes could be associated with schizophrenia, rather than strictly an inclusion of others.

Louvain gave only one enriched cluster, this is not much to analyse for comparison so whilst it will be kept in for other analysis of this network, it will not be used for the SynGO network.

The infomap, sgG2, sgG5 and spectral all gave useful positively enriched clusters. The cluster sizes weren't too big, or were at least proportionate to the algorithm. These algorithms were carried forward into the rest of the dissertation.

Algorithm	No. Clusters	Mean Cluster Size	Median Cluster Size	Modularity
sgG1 (Spin-glass 1)	25	121.04	54	0.42597600805836205
sgG2 (Spin-glass 2)	44	68.77	57.5	0.417178528776594
louvain	13	232.77	256	0.4129543202092427
fc (Fast-Greedy Com-munity)	19	159.26	17	0.37896767736409137
infomap	208	14.55	8	0.36616361441576917
spectral	83	36.46	12	0.36084419164555376
sgG5 (Spin-glass 5)	115	26.31	25	0.3481206693934865
wt (Walk-trap)	602	5.03	1	0.28673234276273485

Table 4.3: Information about algorithms generated for the consensus graph using BioNAR. They can be seen visualised at Figure A.9. It includes the algorithms shorted and full names, the number of clusters each algorithm has, the mean and median cluster sizes and the modularity score. The table is sorted from highest to lowest modularity score.

4.4 A Comment on the Consensus Network

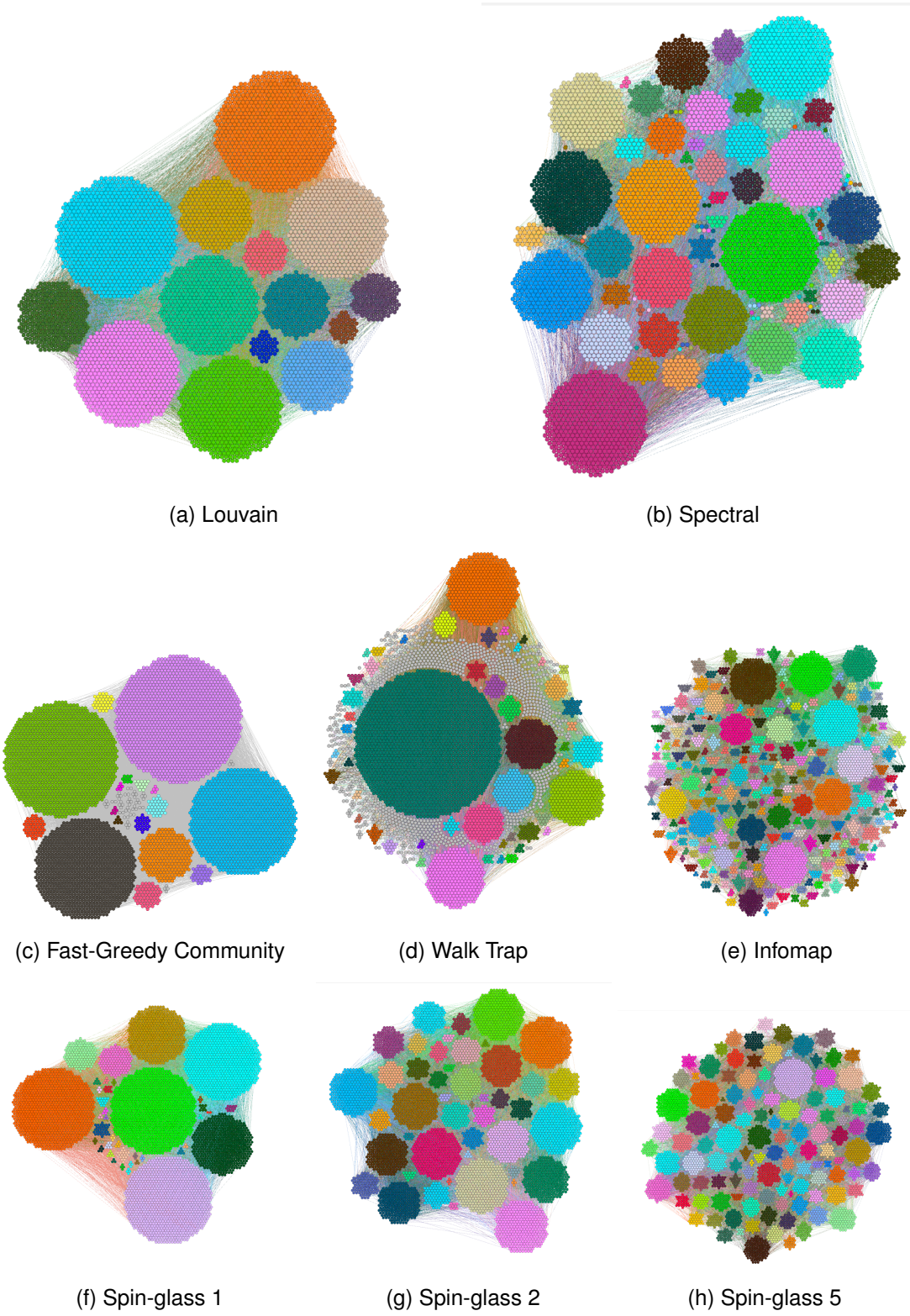
For the rest of the dissertation, due to space constraints, the majority of the graphs and data from the consensus network will be in the appendices. These can be disregarded as they often give consistent results with the other networks. SynGO, which will be talked about extensively later on, can be thought of as a more extreme consensus network. SynGO genes, due to their extensive studying, are arguably much more reliable as only the most extensively validated genes are included in the SynGO set. The use of the consensus network was going to be for a quality control and the SynGO network fills that void whilst also allowing for a more interesting discussion later on about the tendency of researchers to focus solely on SynGO annotated genes.

4.5 Conclusion

Networks were generated and analysed for a full PSD network and consensus PSD network. They were analysed and it was found that fc and wt were not good algorithms and removed from the rest of the dissertation. The remaining algorithms were the sgG1, sgG2, sgG5, infomap, and spectral, due to their better clustering properties and the appearance of enriched clusters.

Algorithm	Cluster Number	Cluster Size	SCZ Broad No.	FL	padj	pval
infomap	222	8	3	True	0.00584	0.00292
infomap	19	83	9	True	0.01006	0.00503
infomap	14	11	3	True	0.01576	0.00788
infomap	188	11	3	True	0.01576	0.00788
infomap	176	14	3	True	0.03187	0.01593
infomap	126	6	2	True	0.04180	0.02090
infomap	93	7	2	True	0.05702	0.02851
infomap	167	7	2	True	0.05702	0.02851
infomap	33	82	7	True	0.08366	0.04183
infomap	5	120	9	True	0.09253	0.04627
louvain	8	44	6	True	0.01407	0.00703
sgG1	30	130	1	False	0.06458	0.03229
sgG2	5	13	3	True	0.02578	0.01289
sgG5	116	33	5	True	0.01748	0.00874
sgG5	112	26	4	True	0.03548	0.01774
spectral	13	11	3	True	0.01576	0.00788
spectral	3	1	1	True	0.07888	0.03944
spectral	76	34	4	True	0.08634	0.04317
wt	141	3	2	True	0.00904	0.00452
wt	52	5	2	True	0.02860	0.01430
wt	15	33	4	True	0.07852	0.0393
wt	225	1	1	True	0.07888	0.03944
wt	257	1	1	True	0.07888	0.03944
wt	278	1	1	True	0.07888	0.03944
wt	291	1	1	True	0.07888	0.03944
wt	295	1	1	True	0.07888	0.03944
wt	310	1	1	True	0.07888	0.03944
wt	313	1	1	True	0.07888	0.03944
wt	347	1	1	True	0.07888	0.03944
wt	470	1	1	True	0.07888	0.03944
wt	480	1	1	True	0.07888	0.03944
wt	509	1	1	True	0.07888	0.03944
wt	574	1	1	True	0.07888	0.03944
wt	582	1	1	True	0.07888	0.03944
wt	601	1	1	True	0.07888	0.03944
wt	613	1	1	True	0.07888	0.03944
wt	730	1	1	True	0.07888	0.03944
wt	735	1	1	True	0.07888	0.03944
wt	761	1	1	True	0.07888	0.03944
wt	767	1	1	True	0.07888	0.03944
wt	797	1	1	True	0.07888	0.03944
wt	817	1	1	True	0.07888	0.03944

Table 4.4: SCZ Broad enriched clusters for the Full Postsynaptic Algorithms and Clusters. It includes the Algorithm, Cluster number, cluster size and number of SCZ Broad genes, the FL indicates if its positive or negative enrichment, the padj and pval. The padj and pval are rounded to 5 decimal places.



(a) Louvain

(b) Spectral

(c) Fast-Greedy Community

(d) Walk Trap

(e) Infomap

(f) Spin-glass 1

(g) Spin-glass 2

(h) Spin-glass 5

Figure 4.3: Visualisation of each clustering algorithm generated for the full postsynaptic network. Colours were generated randomly and chosen arbitrarily. For a some algorithms, those with cluster sizes of 4 or less are visualised in grey. The consensus network algorithm clustering can be found in the appendix at A.9. Information about the algorithms can be found at 4.2.

Chapter 5

Schizophrenic Associations of the Full PSD Network

Now that we have found the enriched genes, we can investigate the specific SCZ enriched genes found in the PSD network. Initially the enriched clusters were compared using the Jaccard Index and the results from that were analysed using various means. A heatmap was generated of the most commonly found enriched SCZ genes and analysed. This analysis will be useful later on for comparison with the SynGO network.

5.1 Cluster Comparison

Clusters were compared using the Jaccard Index mentioned in section 3.2.1, this directly compared the number of similar genes between clusters and can be found in table 5.1. It can easily be seen that the top 3 algorithm clusters from table 5.1 (sgG2 5, infomap 14 and spectral 13) with the most similarity all have the same SCZ genes (with the exception of spectral having one difference). This type of algorithmic consistency gives confidence in their association and in how these particular communities are structured. These genes were SMG6, DGKI, TRPC4 and CUL9. The protein SMG6 appeared 4 times across algorithms, DGKI and TRPC4 across 3 and CUL9 only in 1. These most commonly occurring genes across the most similar clusters were also the most commonly occurring genes overall, once again, this improves confidence in these genes associations with schizophrenia.

A heatmap of the most commonly occurring SCZ genes (occurring $>= 2$ times) that are found in the enriched clusters across different algorithms can be seen at figure 5.1. This is useful for seeing what genes are commonly enriched, and as identified before the SMG6, DGKI, TRPC4 but also the MSANTD2 gene.

One of the most commonly occurring genes, SMG6, is already supported to be associated with schizophrenia [23]. DGKI also already is supported as being linked with schizophrenia [41].

Alg, Cluster No.	Cluster Size	No. SCZ	Alg, Cluster No.	Cluster Size	No. SCZ	Jaccard Index
sgG2 5	13	3	infomap 14	11	3	0.84615
sgG2 5	13	3	spectral 13	11	3	0.41176
spectral 13	11	3	infomap 14	11	3	0.37500
spectral 76	34	4	infomap 176	14	3	0.29729
sgG5 116	33	5	infomap 5	120	9	0.14179
sgG1 30	130	1	sgG5 112	26	4	0.01299
sgG5 112	26	4	infomap 33	82	7	0.00935
sgG5 116	33	5	infomap 33	82	7	0.00877
spectral 76	34	4	sgG1 30	130	1	0.00613

Table 5.1: Table to compare enriched clusters between different algorithms for the Full Synaptic Graph. This compares their simularity using the Jaccard Index, explained in 3.2.1. Only those with a Jaccard Index greater than 0 were included.

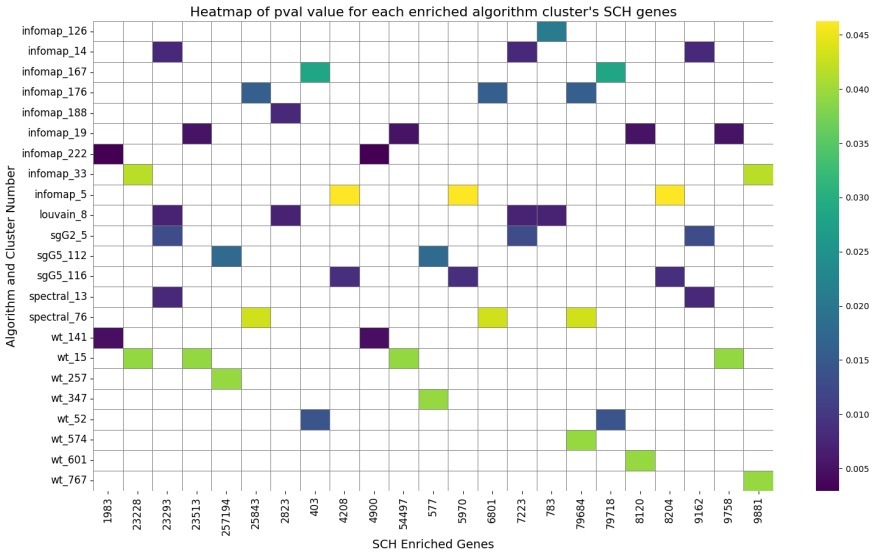


Figure 5.1: Heatmap of the pval for SCZ genes for each enriched algorithm and cluster. SCZ genes with one or less values were disregarded, an expanded version for all SCZ Genes that are found at least once can be found in Appendix A.5

Entrez GeneID	Gene Symbol	Algorithm Associated	Summary
23293	SMG6	infomap 14, louvain 8, sgG2 5, spectral 13	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. Alternatively spliced transcript variants encoding distinct protein isoforms have been described. [provided by RefSeq, Feb 2014]
7223	TRPC4	infomap 14, louvain 8, sgG2 5	This gene encodes a member of the canonical subfamily of transient receptor potential cation channels. The encoded protein forms a non-selective calcium-permeable cation channel that is activated by Gq-coupled receptors and tyrosine kinases, and plays a role in multiple processes including endothelial permeability, vasodilation, neurotransmitter release and cell proliferation. Single nucleotide polymorphisms in this gene may be associated with generalized epilepsy with photosensitivity. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]
79684	MSANTD2	infomap 176, spectral 76, wt 574	N/A
9162	DGKI	infomap 14, sgG2 5, spectral 13	This gene is a member of the type IV diacylglycerol kinase subfamily. Diacylglycerol kinases regulate the intracellular concentration of diacylglycerol through its phosphorylation, producing phosphatidic acid. The specific role of the enzyme encoded by this gene is undetermined, however, it may play a crucial role in the production of phosphatidic acid in the retina or in recessive forms of retinal degeneration. [provided by RefSeq, Jul 2008]

Table 5.2: The top 4 most frequently occurring SCZ Broad genes from the enriched algorithm clusters found from the Full Postsynaptic network.The summaries were scraped directly from PubMed with the exception of 79684 which did not have a summary [45]. A fuller complete table that contains all values can be found in the appendix at B.3.

The top 4 most commonly occurring broad genes were SMG6, TRPC4, MSANTD2 and DGKI which all appeared at least 3 times across different algorithms. These proteins were investigated for similarities to all other SCZ broad coding proteins using PSI-BLAST, this is a standard protein alignment tool in Bioinformatics [1, 11, 2]. Only DGKI showed similarity with proteins other than itself, it aligned with the proteins DGKZ (Entrez ID: 8525) and OSBPL1A (Entrez ID: 114876).

DGKZ appeared enriched in only the infomap algorithm. The results acquired were a 99% query cover, a 64.64% percentage identity, and an e value of 0.00. This is a very high query cover and an amazing e value, these show the proteins are very similar. The percentage identity is also shows some similarity between proteins. This is to be expected as they are both part of the DGK family. They will be explored further on in section 7.2.1.

The OSBPL1A protein did not appear enriched in any clusters. The e value between these two proteins was 8e-04, this can be considered ”good” as it tells us the chance that the alignment was a false positive. The proteins are not that similar though, with a 33% query cover and 9% percentage identity. Due to this poor match, this was not explored further.

Trubetskoy et al. 2022 identifies 10 genes as very strong supporters as schizophrenia [61]. Of these 10 genes, 2 were identified as enriched in my full network, these are GRIN2A and MYT1L. Notably, they both appear in enriched exclusively in the infomap algorithm and does not appear in any of the others. The full table of enriched genes can be found in the appendices at B.3.

5.2 Conclusion

The main proteins that were found consistently across algorithms were SMG6, TRPC4, MSANTD2, and DGKI. Only DGKI was had a similar molecular structure to other enriched proteins. The lack of simularity to other proteins for SMG6, TRPC4 and MSANTD2 could indicate unique roles in the biological or molecular functions. The overall consistent occurrences of all of these proteins do indicate they are strongly associated with schizophrenia.

Chapter 6

SynGO Network

Many papers will make use of genes with SynGO annotations exclusively. Whilst these SynGO genes are very useful as they are well annotated, it can lead to bias. This section will look at how the results will differ when you overlook under annotated genes and use exclusively SynGO genes. The chapter after this one will focus on the differences between the SynGO and the Full Network.

6.1 SynGO Setup and Analysis

The same steps from the previous chapter were repeated for SynGO specific genes. The same Synaptome DB was used, however all non SynGO PSD were disregarded in the creation of this network. This obviously lead to a significantly reduced network.

Network	**Nodes**	**SCZ Broad**	**SCZ Prioritised**	**Edges**	**Average Degree**	**Average Diameter**
SynGO Network	887	48	13	3092	6.972	9

Table 6.1: Graph information for the SynGO network. The nodes are all the connected nodes, those without an edge were discarded.

Immediately, the main difference can be seen with the number of nodes. The Full PSD Network has a total of 4843 nodes compared to 887 of the SynGO PSD Network, this is less than a 1/5 ($\approx$18%) of the size. There are still a significant number of SCZ Broad genes within the network though, 25% of the original 191 from the Full PSD Network remaining. The cluster sizes are naturally smaller, coming from a smaller network size, this can be seen visually when comparing 4.3 for the Full Network and A.4 for the SynGO Network.

The network was clustered using only infomap, sgG1, sgG2 and sgG5. The algorithm information can be found in table 6.2 and the visualisation can be found at figure A.4. When analysing the ARI at figure 6.1, you can see it has a very similar shape to the full

networks ARI at figure 4.2. This is good and expected, as it algorithm behaviour should be consistent across networks.

Unlike the full network, in this instance sgG1 was not similar to the other algorithms. The sgG1 algorithm kept a consistent number of clusters (25 in full, 24 in SynGO), whilst the others reduced the number of clusters for this smaller network. One consistent thing across all networks is the presence of False enrichment within the sgG1 algorithm, seen in table B.2. Despite the bigger difference between algorithms found across networks, sgG1 keeps this False enrichment. One larger constrast though, is the fact that the SynGO network sgG1 actually features some True enrichment, the full network only contains one Falsely enriched cluster. This larger difference in algorithms found in ARI could be why there are more positively enriched clusters found in sgG1 this time.

The jaccard index, found in table 6.4, performed significantly better than the full network, found in table 5.1. In fact, there were 3 identical clusters, these were infomap 10, sgG1 13 and sgG5 69. These clusters were quite small, only containing 4 nodes with 2 SCZ genes, this can be visualised along the right hand side of 6.2. This small size is consistent with the rest those found with a high jaccard index for the SynGO network. Whilst these clusters were smaller, they are about proportional to the network size when compared to the full network and its size.

The identical clusters contained the genes EIF5 and NRGN. The NRGN gene was already found to be associated with schizophrenia [41]. The Pers et al. (2016) paper found that the NRGN gene was associated with three main functions, ”Platelet homeostasis”, ”Signalling by NGF”, and ”MAPK signaling pathway”. These will be mentioned further on in section 7.2.

6.2 Conclusion

The SynGO network does create more consistent clustering than the full network. Naturally due to its reduced size, it provides less enriched genes overall, however, those that it did produce were produced consistently across clusters. Whilst some data is lost with SynGO, only the most robust and consistent genes make it through the enrichment process, making the overall network more trustworthy.

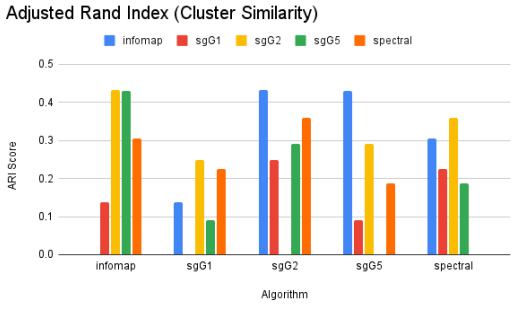


Figure 6.1: The Adjusted Rand Index Score for the SynGO Network. It gives a score from -1 to 1 for how each algorithm is similar to each other. The Adjusted Rand Index was ran with the modifications mentioned in 3.2.2.

Algorithm	**No. Clusters**	**Mean Cluster Size**	**Median Cluster Size**	**Modularity Score**
sgG1 (Spin-glass 1)	24	36.96	15	0.465583
sgG2 (Spin-glass 2)	35	25.34	28	0.439649
infomap	87	10.20	7	0.406317
spectral	49	18.10	16	0.395700
sgG5 (Spin-glass 5)	80	11.09	10	0.383757

Table 6.2: Information about algorithms generated for the SynGO graph using BioNAR. They can be seen visualised at Figure A.4. It includes the algorithms shorted and full names, the number of clusters each algorithm has, the mean and median cluster sizes and the modularity score. The table is sorted from highest to lowest modularity score.

Algorithm	**Cluster Number**	**Cluster size**	**No. SCZ Broad**	**FL**	**padj**	**pval**
infomap	8	53	8	True	0.01114	0.00557
infomap	4	17	4	True	0.02137	0.01068
infomap	10	4	2	True	0.03211	0.01605
infomap	5	52	7	True	0.03584	0.01792
sgG1	23	126	14	True	0.00880	0.00440
sgG1	13	4	2	True	0.03211	0.01605
sgG1	9	88	1	False	0.07592	0.03796
sgG2	13	23	4	True	0.06251	0.03125
sgG2	19	59	7	True	0.06816	0.03480
sgG5	72	15	4	True	0.01328	0.00664
sgG5	69	4	2	True	0.03211	0.01605
sgG5	39	5	2	True	0.05167	0.02583
sgG5	29	13	3	True	0.05802	0.02901
sgG5	22	38	5	True	0.09757	0.04878

Table 6.3: Enrichment values found for each algorithm cluster for the SynGO Postsynaptic Network. Only those with a pval ≤ 0.05 were included. FL tells you if it is positive or negative enrichment, positive means it is enriched for the clusters with SCZ Broad genes present, negative is for the clusters without SCZ Broad genes.

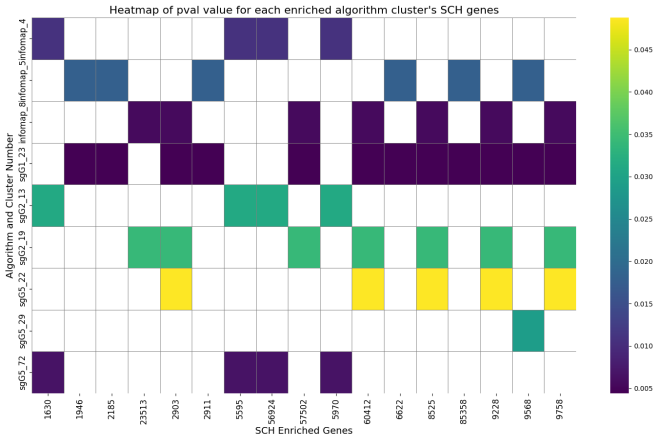


Figure 6.2: Heatmap of the pval for SCZ Broad genes that show up at least twice between the enriched algorithm clusters for the SynGO Network. The x-axis contains the Entrez ID for each SCZ Broad Gene, and the y-axis has the algorithm_clusterNumber formatted as such. The p-val is shown through the heatmap, with purple being the ”best” and lowest p-val and yellow being the highest and ”worst”.

Alg, Cluster No.	**Cluster Size**	**No. SCZ**	**Alg, Cluster No.**	**Cluster Size**	**No. SCZ**	**Jaccard Index**
infomap 10	4	2	sgG1 13	2	4	1.0
sgG1 13	4	2	sgG5 69	2	4	1.0
infomap 10	4	2	sgG5 69	2	4	1.0
infomap, 4	17	4	sgG5 72	4	15	0.77778
infomap 8	53	8	sgG2 19	7	59	0.75
infomap 4	17	4	sgG2 13	4	23	0.66667
infomap 8	53	8	sgG5 22	5	38	0.65455
sgG2 13	23	4	sgG5 72	4	15	0.65217

Table 6.4: Reduced Table to compare enriched clusters between different algorithms for the SynGO Graph. This compares their similarity using the Jaccard Index, explained in 3.2.1. Only those with a Jaccard Index greater than 0 were included. The full table can be found at B.9.

Chapter 7

Full Network vs SynGO Network

In this chapter I will be comparing compare our previous results found in the full database to those of a network made entirely of SynGO genes.

A more direct and useful comparison can be made of the Full and SynGO network. This direct comparison can help justify the trustworthiness of features that appear in both graphs, it can also help highlight the kind of information lost when focusing only on heavily annotated SynGO exclusive genes.

First, I will be comparing the Bridgeness of each network, this can tell us about the differences of what proteins are the most influential between networks.

Next I will be making use of WordClouds to visually aid the analysis of enriched annotations are discovered across the different networks. The Biological function, Molecular function, SynGO annotation, and Disease annotation is analysed here.

7.1 Bridgeness Comparison

Bridgeness is useful when plotted against Semilocal Centrality, both which were mentioned in 2.3. This plotting can tell you about the influence of a protein on the local and global network [33]. The bridgeness plot is split into 4 different regions, they are described below but can easily be visualised in figure 7.1.

Region 1: The upper-left region, these are proteins with a larger global influence, they may not have as strong a local influence.

Region 2: The upper-right region, the proteins here show both a local and global influence.

Region 3: The bottom-left region, these proteins exhibit quite a local influence but also have some influence over other communities.

Region 4: The bottom-right region, proteins tend to have an almost exclusively local influence but may still influence another community.

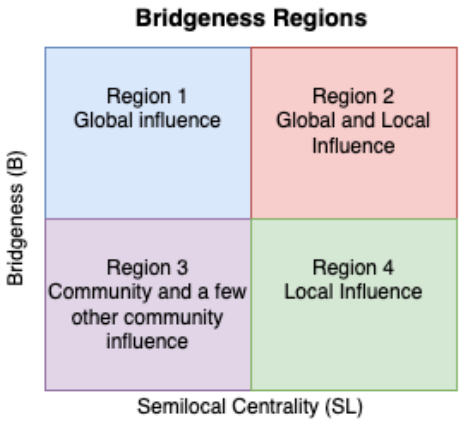


Figure 7.1: Bridgeness sections influence visualised. These section regions can be mapped on the Bridgeness vs Semilocal Centrality plots. Created by me, using draw.io.

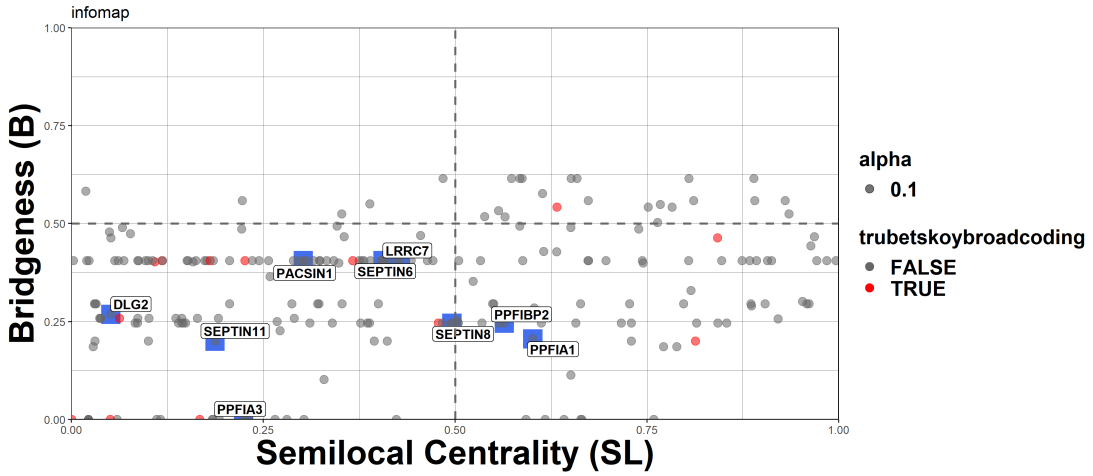


Figure 7.2: Bridgeness vs Semilocal Centrality for the SynGO Network for the Infomap Algorithm. SCZ Broad genes are coloured in red. Alpha was is the transparency of each dot.

The most proteins were found across all clustering algorithms and both networks. These common proteins coding genes can be found in the appendix B.10 but can also be seen in the figures 7.2 and 7.3.

There was two common protein found in both, PACSIN1 and PPFIA3, however neither were SCZ genes. Only one of these proteins were directly in the enriched clusters and a SCZ gene, this was DLGAP2 which is gene that is thought to be associated with signalling [45]. DLGAP2 was actually found in the prioritised list, so there is great confidence that is it not only prioritised, but

DLGAP1 is already associated with Schizophrenia, however not the SCZ genes from the Trubetskoy et al study.

These bridgeness proteins were analysed directly against each other using BLASTP, however other than expected (e.g. SEPTIN6 being similar to SEPTIN11 etc.), no proteins were particularly similar. This direct protein analysis was thus dropped.

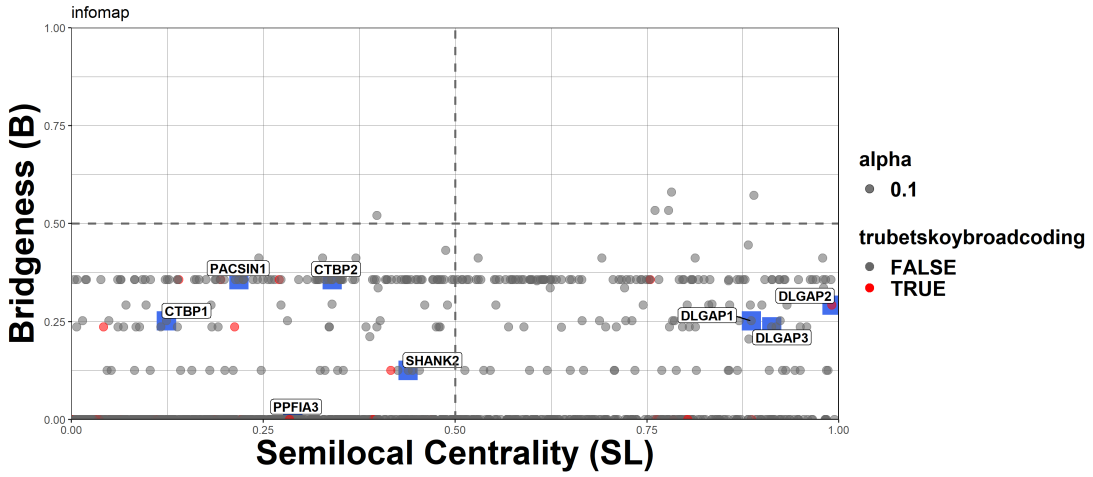


Figure 7.3: Bridgeness vs Semilocal Centrality for the Full Network for the Infomap Algortihm. SCZ Broad genes are coloured in red. Alpha was is the transparency of each dot.

7.2 WordCloud Comparisons

In order to get a better visualisation of common functions and associations of the enriched clusters, WordClouds were created. They were created for the Full, Consensus and SynGO networks enriched clusters however only the Full and SynGO can be found below and as usual the Consensus can be found in the appendix from A.10 onwards.

Generally, whilst SynGO had less overall annotations included with each word cloud, they had higher frequencies and weights. The full network had a greater number of annotations overall, however, it also had a much lower frequency per annotation. The consensus network, even though it isn't included below it can be found in the appendices at A.10, A.11, A.12, and A.13. The consensus actually has even more frequent annotations than the SynGO network as well as having a greater number of annotations than the SynGO network, whilst it does have less annotations than the full network it still has considerably more than the SynGO network. These reasons make the consensus network a more robust and trustworthy network for ORA and Wordcloud analysis.

The following analysis will continue to use only the full and SynGO networks in order to be consistent with the rest of the dissertation. It may more useful to compare a upper and lower bound for both frequency of annotation, and number of annotations. These next sections will focus on what has been found that is consistent with current data, or previously found analysis in this thesis.

7.2.1 Biological Function

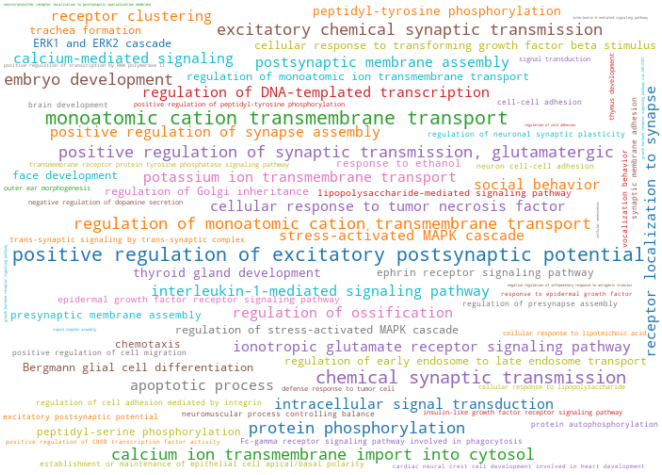


Figure 7.4: Word cloud of the Biological Function for the SCZ enriched clusters of the SynGO only Network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.16.

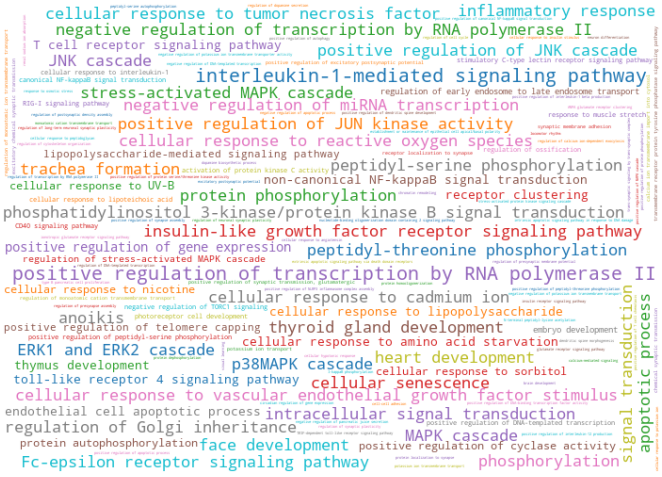


Figure 7.5: Word cloud of the Biological Function for the SCZ enriched clusters of the Full Network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.20.

Thyroid gland development is mentioned in both figures 7.4 and 7.5. Another study that makes use of the same data from the Trubetskoy et al. 2022 (our SCZ genes) found that the schizophrenic genes are associated with Thyroid cancer [68]. However, this study has been recently criticised due the use of older data sets [62].

More recently, a paper was published regarding Thyroid cancer [6]. The paper specifically mentions the DGKI gene specifically as being significant. Earlier in section 5.1, we found that this gene occurred across 3 different algorithms as an enriched SCZ gene. This can further support the link between the thyroid and schizophrenia.

”Dendritic spine development”,”positive regulation of dendritic spine development”, and ”dendritic spine morphogenesis” all appeared in the consensus network consistently

with high frequency. The last two appeared only once in the full network. They did not appear at all in the SynGO network, however, the protein FRMPD4 did appear enriched across 3 different algorithms in the SynGO network and twice in the full network. FRMPD4 is a a positive regulator of dendritic spine morphogenesis. Genes that contribute some some attributes do seem to appear in SynGO however, the attributes themselves don't always seem to appear.

7.2.2 Molecular Function

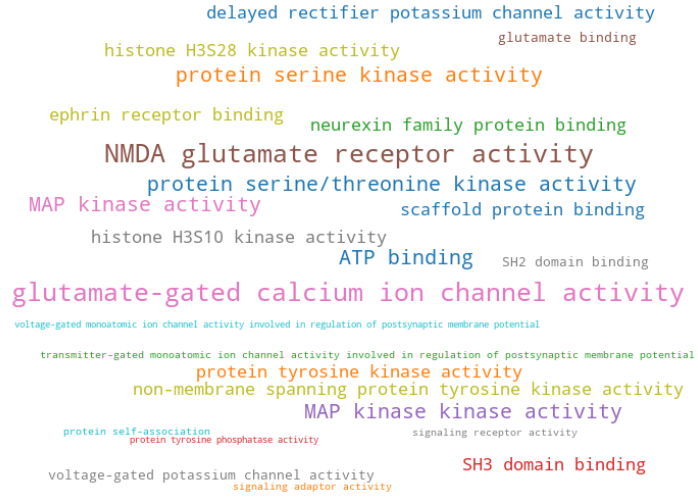


Figure 7.6: Word cloud of the Molecular Function for the SCZ enriched clusters of the SynGO only Network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.15

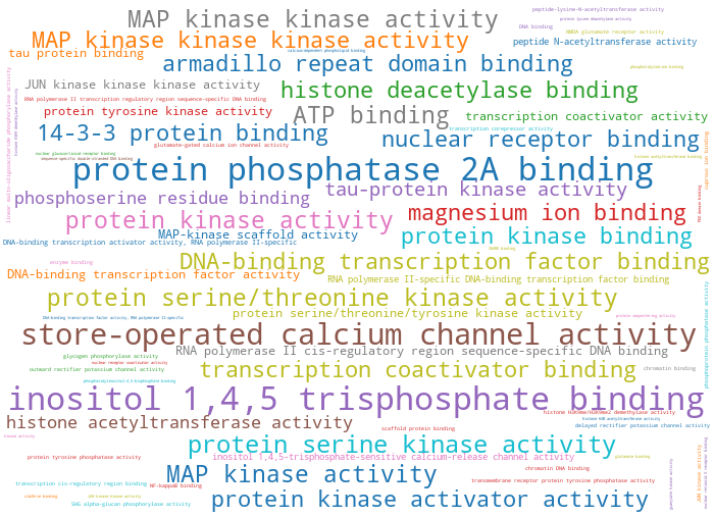


Figure 7.7: Word cloud of the Molecular Function for the SCZ enriched clusters of the Full Network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.19

Visually, it can be seen that there is a significantly reduced number of enriched annotations available in figure 7.6 compared to figure 7.7. This supports the idea of SynGO

leading to lost annotations and information.

MAPK signaling pathway is supported in Pers et al. [41], it has the protein DGKZ which was mentioned earlier in section 5.1 and the protein NRGN, which was mentioned in section 6.1. The MAPK signaling pathway is closely associated with the other MAP functions, here we have MAP kinase kinase activity, which appears in the full, consensus and SynGO network, and MAP kinase kinase kinase activity which appears exclusively in the full and consensus networks.

7.2.3 SynGO Annotations

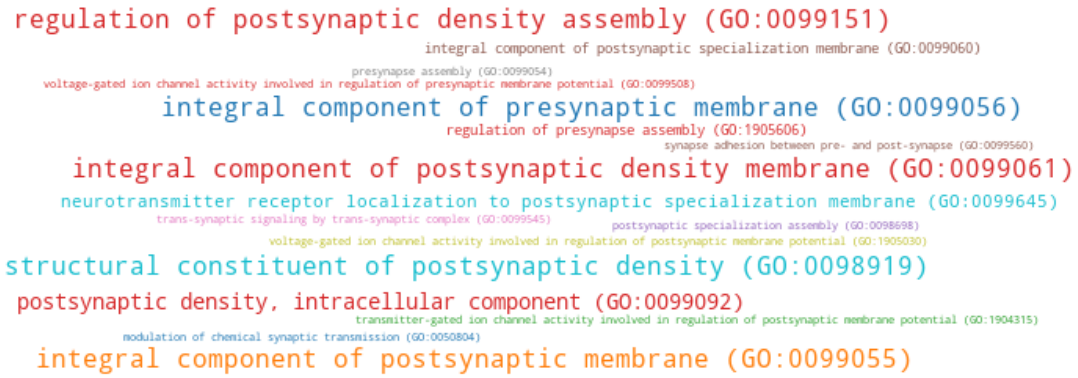


Figure 7.8: Word cloud of the SynGO annotations for the SCZ enriched clusters of the SynGO only network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.18.

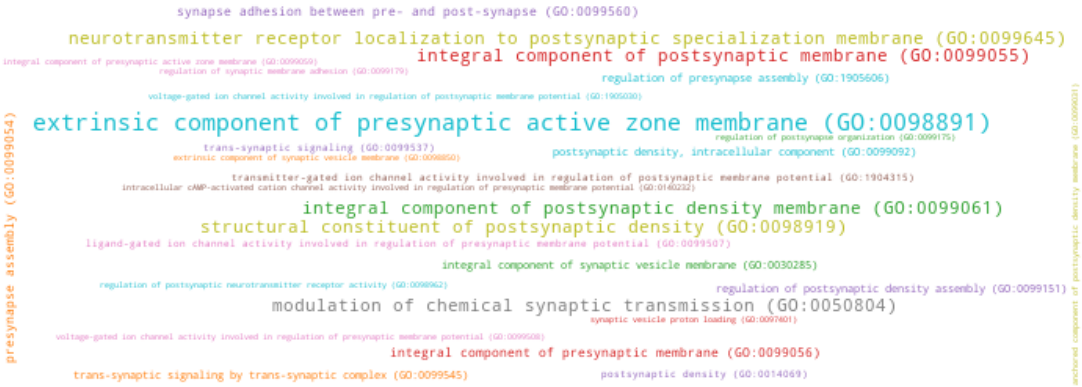


Figure 7.9: Word cloud of the SynGO annotations for the SCZ enriched clusters of the Full Network. The word size is weighted by frequency and 1 - pval. The frequency table can be found at B.22

Given that this a PSD network, much of the SynGO annotations that appear are PSD based or associated. This is expected. Interestingly, as mentioned in Trubetskoy et al. (2022), there does seem to a larger connection and overlap with the presynaptic terminal. There are a large number of presynaptic annotations however there is a decent amount of overlap between them as seen in figure 4.1.

The SynGO annotation ”extrinsic component of presynaptic active zone membrane” occurred 3 times in both the consensus and full network but nowhere in the SynGO only network. It did not appear at all within the SynGO networks enriched SynGO annotations, however the protein SNAP91 which is associated with this SynGO annotation, did appear enriched in the sgG1 algorithm. The SNAP91 did not appear enriched in the full or consensus network.

7.2.4 Diseases

(a) Full Network (b) SynGO Network

Figure 7.10: Wordcloud generated based on enriched disease values for clusters with SCZ Broad enrichment. The size of each word is weighted by the sum of the 1 - pval, every word is weighted on how enriched and how frequently it occurs. Each word cloud represents a different networks. The tables used to generate these can be found in the appendix at B.21, B.17.

The tables for these can be found in the appendix at B.21 for the full network, B.13 for the consensus network, and B.17 for the SynGO network.

One thing to note is that the Schizophrenia that appears in Figure 7.10 is different from the SCZ genes I have been examining in this dissertation. These ones came already pre-associated with the network along with the rest of the disease annotations from the papers used to build Synaptome DB. Genes previously marked as schizophrenic from other studies were to be expected to coincide with the enriched genes found in this paper. As expected, it was one of the top most occurring enriched annotations found.

SynGO has half as many associated diseases than the Full network. Hypertension which was one of the most frequent associations of the Full Network and also appeared frequently in the Consensus network, did not appear at all in SynGO. This speaks volumes about how much data can be lost when you exclusively use data from such a small network. The hypertension is also not a fluke of the full network due to the frequency it appears as well as the similar frequency that was matched in the consensus table.

The link between hypertension and pyschosis (including schizophrenia) has been explored before [54]. Sadarshan and Cheung (2023) pose that the link can come from side effects from antipsychotic medication however also acknowledge there may be a link between irregular nervous system activity. CNTN4, MYT1L, GRM1, MAPT, NLGN4X

are proteins that were enriched and that are associated with the nervous system. CNTN4, MYT1L, GRM1, and NLGN4X are directly involved in the creation or regulation of cells in the nervous system further supporting this hypertension schizophrenia link.

NRGN gene was associated with platelet homeostasis and schizophrenia, it was also found across 3 different clusters in the SynGO network [41]. Platelet homeostasis is associated with hypertension [67]. This further supports the idea that the SynGO network does not lack the specific genes that are associated with certain conditions, but fails specifically to get enriched annotations.

Autistic Disorder and Autism Spectrum Disorder (ASD) appear enriched with Schizophrenia being associated with both diseases within all networks, this concurs with current research [26] [20]. The difference between Autistic disorder and Autism Spectrum Disorder is what version of the DSM it comes from. Autistic disorder comes from the DSM IV, and Autism Spectrum Disorder is from the DSM V, the criteria difference between them are slightly different. ASD is a broader category which moved a few specific disorders, such as Asperger's Disorder, under the one ASD term [29].

The ASD and SCZ link was studied in 2 other studies that used the Trubetskoy et al. paper data. The first one did not find a SCZ ASD link, they performed enrichment analysis using topologically associated domains (TADs), this is focused on the high frequency interactions between DNA sequences [36, 8]. Another paper took a polygenic score (PGS) approach, which focuses on single nucleotide polymorphism (SNP), this paper found a correlation between the SCZ ASD and even ADHD [51]. Shakeshaft et al. [51] paper does find a correlation between specifically in those with late-emerging traits of ASD and ADHD, this is interesting as SCZ tends to develop later as well. These methods are focused much more on the genomics, compared to our PPI networks which have a greater emphasis on molecular interactions.

7.3 Conclusion

The SynGO and Full network had many comparable results. The full network had more varied and a wider number of enriched annotations, whilst the SynGO network had a smaller number of annotations, they occurred with higher frequency across all algorithms and clusters which suggests these annotations are more reliable as schizophrenic associations. The consensus network outperformed both of the networks on both fronts, it had greater quality annotations than the full network, and more annotations than the SynGO network. Overall, it was shown that SynGO may have similar enriched genes that support the full and consensus annotations, SynGO fails to get the full number of enriched annotations.

Chapter 8

Conclusion

At the start of this paper we posed the following questions:

1. How does BioNAR perform when comparing SCZ genes and their impact on protein-protein interaction networks?
2. How does a SynGO only network compare to a full one?

These questions have been successfully answered by this dissertation. Firstly, BioNAR performed well when comparing SCZ genes impact on PPI networks. Many connections that have been found in multiple papers, were verified with BioNAR. It can be a useful tool due to the diversity of features it has, it also finds multiple connections with one tool, that many papers find with many.

Consistently, networks with some noise but verified through consensus outperformed those solely based on highly annotated genes like SynGO. These consensus networks offered better quality and frequency of annotations, enhancing trust despite the SynGO networks more consistent enriched clusters across algorithms.

8.1 Future Work

If there were more time and space, a fully comprehensive analysis that included the presynaptic network and a synaptosome network would have allowed for more comparisons and insights between networks. This further exploration into the presynaptic network is supported by what was found in section 7.2.3.

The steps of this dissertation can easily be repeated with other genes and diseases once an appropriate dataset is found. The style in which the code has been written was in order to generalise very well as the same code was repeatedly used between networks. The code can used immediately with amending only the file paths and what data should be annotated.

It would be very useful to repeat the SynGO vs Full Network vs Consensus Network comparison for other diseases, this could validate some of the findings we found further regarding the consensus network being the strongest network among them all.

Bibliography

[1] S. F. Altschul et al. “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”. eng. In: *Nucleic Acids Research* 25.17 (Sept. 1997), pp. 3389–3402. ISSN: 0305-1048. DOI: 10.1093/nar/25.17.3389.

[2] Stephen F. Altschul et al. “Protein database searches using compositionally adjusted substitution matrices”. eng. In: *The FEBS journal* 272.20 (Oct. 2005), pp. 5101–5109. ISSN: 1742-464X. DOI: 10.1111/j.1742-4658.2005.04945.x.

[3] *An Introduction to Brain Networks - ScienceDirect*. URL: https://www.sciencedirect.com/science/article/pii/B9780124079083000017 (visited on 14/04/2024).

[4] Michael Ashburner et al. “Gene Ontology: tool for the unification of biology”. en. In: *Nature Genetics* 25.1 (May 2000). Publisher: Nature Publishing Group, pp. 25–29. ISSN: 1546-1718. DOI: 10.1038/75556. URL: https://www.nature.com/articles/ng0500_25 (visited on 05/04/2024).

[5] Fengju Bai and Frank A. Witzmann. “Synaptosome Proteomics”. In: *Sub-cellular biochemistry* 43 (2007), pp. 77–98. ISSN: 0306-0225. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853956/ (visited on 01/04/2024).

[6] Miaoyu Bai et al. “Key molecules associated with thyroid carcinoma prognosis: A study based on transcriptome sequencing and GEO datasets”. English. In: *Frontiers in Immunology* 13 (Aug. 2022). Publisher: Frontiers. ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.964891. URL: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.964891/full (visited on 16/04/2024).

[7] Mathieu Bastian, Sebastien Heymann and Mathieu Jacomy. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. 2009. URL: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

[8] Jonathan A. Beagan and Jennifer E. Phillips-Cremins. “On the existence and functionality of topologically associating domains”. en. In: *Nature Genetics* 52.1 (Jan. 2020). Publisher: Nature Publishing Group, pp. 8–16. ISSN: 1546-1718. DOI: 10.1038/s41588-019-0561-1. URL: https://www.nature.com/articles/s41588-019-0561-1 (visited on 10/04/2024).

[9] Vincent D. Blondel et al. “Fast unfolding of communities in large networks”. en. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (Oct. 2008), P10008. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2008/10/P10008. URL: https://dx.doi.org/10.1088/1742-5468/2008/10/P10008 (visited on 10/04/2024).

[10] T. M. Boeckers. “The postsynaptic density”. eng. In: *Cell and Tissue Research* 326.2 (Nov. 2006), pp. 409–422. ISSN: 0302-766X. DOI: 10.1007/s00441-006-0274-5.

[11] Grzegorz M. Boratyn et al. “Domain enhanced lookup time accelerated BLAST”. eng. In: *Biology Direct* 7 (Apr. 2012), p. 12. ISSN: 1745-6150. DOI: 10.1186/1745-6150-7-12.

[12] Elizabeth I. Boyle et al. “GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes”. In: *Bioinformatics* 20.18 (Dec. 2004), pp. 3710–3715. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bth456. URL: https://doi.org/10.1093/bioinformatics/bth456 (visited on 15/04/2024).

[13] Alan S. Brown. “Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism”. In: *National Library of Medicine* (2013). DOI: 10.1002/dneu.22024.

[14] Alan S. Brown and Elena J. Derkits. “Prenatal Infection and Schizophrenia: A Review of Epidemiologic and Translational Studies”. In: *National Library of Medicine* (2010). DOI: 10.1176/appi.ajp.2009.09030361.

[15] Noirrit Kiran Chandra and Sourabh Bhattacharya. “Chapter 3 - Dependent Bayesian multiple hypothesis testing”. In: *Handbook of Statistics*. Ed. by Arni S. R. Srinivasa Rao, G. Alastair Young and C. R. Rao. Vol. 47. Advancements in Bayesian Methods and Implementation. Elsevier, Jan. 2022, pp. 67–81. DOI: 10.1016/bs.host.2022.07.001. URL: https://www.sciencedirect.com/science/article/pii/S0169716122000396 (visited on 14/04/2024).

[16] Hanbo Chen. *Venn: Draw Venn Diagrams*. R package version 1.7.3. 2022-04-12. URL: https://CRAN.R-project.org/package=VennDiagram.

[17] Aaron Clauset, M. E. J. Newman and Cristopher Moore. “Finding community structure in very large networks”. In: *Physical Review E* 70.6 (Dec. 2004). Publisher: American Physical Society, p. 066111. DOI: 10.1103/PhysRevE.70.066111. URL: https://link.aps.org/doi/10.1103/PhysRevE.70.066111 (visited on 15/04/2024).

[18] Gabor Csardi and Tamas Nepusz. *The igraph software package for complex network research*. 2006. URL: https://igraph.org.

[19] Daniela C. Dieterich and Michael R. Kreutz. *Proteomics of the Synapse – A Quantitative Approach to Neuronal Plasticity - PMC*. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4739661/.

[20] Menachem Fromer et al. “De novo mutations in schizophrenia implicate synaptic networks”. eng. In: *Nature* 506.7487 (Feb. 2014), pp. 179–184. ISSN: 1476-4687. DOI: 10.1038/nature12929.

[21] Philippe Gervais. *Matplotlib-Venn: Publication quality Venn diagrams in Python*. Python package version 0.11.10. 2012. URL: https://pypi.org/project/matplotlib-venn/.

[22] Clifton L. Gooch. “Neuromuscular Junction, Normal”. In: *Encyclopedia of the Neurological Sciences*. Ed. by Michael J. Aminoff and Robert B. Daroff. New York: Academic Press, Jan. 2003, pp. 505–507. ISBN: 978-0-12-226870-0. DOI:

10.1016/B0-12-226870-9/00823-6. URL: https://www.sciencedirect.com/science/article/pii/B0122268709008236 (visited on 15/04/2024).

[23] Gabriela Maria Guerra et al. *Cell Type-Specific Role of RNA Nuclease SMG6 in Neurogenesis*. en. Nov. 2021. DOI: 10.3390/cells10123365. URL: http://dx.doi.org/10.3390/cells10123365.

[24] Lawrence Hubert and Phipps Arabie. "Comparing partitions". en. In: *Journal of Classification* 2.1 (Dec. 1985), pp. 193–218. ISSN: 1432-1343. DOI: 10.1007/BF01908075. URL: https://doi.org/10.1007/BF01908075 (visited on 13/04/2024).

[25] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[26] Amandeep Jutla, Jennifer Foss-Feig and Jeremy Veenstra-VanderWeele. "Autism spectrum disorder and schizophrenia: an updated conceptual review". In: *Autism research : official journal of the International Society for Autism Research* 15.3 (Mar. 2022), pp. 384–412. ISSN: 1939-3792. DOI: 10.1002/aur.2659. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8931527/ (visited on 01/04/2024).

[27] Frank Koopmans et al. "SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse". In: *Neuron* 103.2 (July 2019), 217–234.e4. ISSN: 0896-6273. DOI: 10.1016/j.neuron.2019.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0896627319304271 (visited on 13/03/2024).

[28] Svetlana Korotchenko et al. "Zooming in on the (Peri)synaptic Extracellular Matrix". In: vol. 84. Jan. 2014, pp. 187–203. ISBN: 978-1-4614-9178-1. DOI: 10.1007/978-1-4614-9179-8_10.

[29] Matthew J. Maenner et al. "Potential Impact of DSM-5 Criteria on Autism Spectrum Disorder Prevalence Estimates". In: *JAMA psychiatry* 71.3 (Mar. 2014), pp. 292–300. ISSN: 2168-622X. DOI: 10.1001/jamapsychiatry.2013.3893. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041577/ (visited on 01/04/2024).

[30] Donna Maglott et al. "Entrez Gene: gene-centered information at NCBI". In: *Nucleic Acids Research* 35.Database issue (Jan. 2007), pp. D26–D31. ISSN: 0305-1048. DOI: 10.1093/nar/gkl993. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1761442/ (visited on 05/04/2024).

[31] Daniel B. McClatchy, Susan B. Powell and John R. Yates. *In vivo mapping of protein-protein interactions of schizophrenia risk factors generates an interconnected disease network*. en. Pages: 2023.12.12.571320 Section: New Results. Dec. 2023. DOI: 10.1101/2023.12.12.571320. URL: https://www.biorxiv.org/content/10.1101/2023.12.12.571320v1 (visited on 09/04/2024).

[32] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

[33] Colin McLean et al. "BioNAR: An Integrated Biological Network Analysis Package in Bioconductor". In: *BioRxiv* (2023). DOI: 10.1093/bioadv/vbad137.

[34] Colin Mclean et al. "Improved Functional Enrichment Analysis of Biological Networks using Scalable Modularity Based Clustering". English. In: *Journal*

of Proteomics & Bioinformatics 9.1 (Jan. 2016). Publisher: Omics Publishing Group, pp. 9–18. ISSN: 0974-276X. DOI: 10.4172/jpb.1000383. URL: https://www.research.ed.ac.uk/en/publications/improved-functional-enrichment-analysis-of-biological-networks-us (visited on 15/04/2024).

[35] *Multiple platform build/check report for BioC 3.18 - All results for package BioNAR*. URL: https://bioconductor.org/checkResults/release/bioc-LATEST/BioNAR/ (visited on 01/04/2024).

[36] Takumi Nakamura et al. "Topologically associating domains define the impact of de novo promoter variants on autism spectrum disorder risk". en. In: *Cell Genomics* 4.2 (Feb. 2024). Publisher: Elsevier. DOI: 10.1016/j.xgen.2024.100488. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10879036/ (visited on 10/04/2024).

[37] M. E. J. Newman. "Modularity and community structure in networks". In: *Proceedings of the National Academy of Sciences* 103.23 (June 2006). Publisher: Proceedings of the National Academy of Sciences, pp. 8577–8582. DOI: 10.1073/pnas.0601602103. URL: https://www.pnas.org/doi/full/10.1073/pnas.0601602103 (visited on 08/04/2024).

[38] Mark Newman. *Networks*. en. Oxford University Press, July 2018. ISBN: 978-0-19-184323-5. DOI: 10.1093/oso/9780198805090.001.0001. URL: https://academic.oup.com/book/27884 (visited on 01/04/2024).

[39] NHS. "Overview - Schizophrenia". In: *NHS Website* (2023).

[40] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[41] Tune H. Pers et al. "Comprehensive analysis of schizophrenia-associated loci highlights ion channel pathways and biologically plausible candidate causal genes". In: *Human Molecular Genetics* 25.6 (Mar. 2016), pp. 1247–1254. ISSN: 0964-6906. DOI: 10.1093/hmg/ddw007. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4764200/ (visited on 16/04/2024).

[42] Medline Plus. "What are proteins and what do they do?" In: *National Library of Medicine* (2021).

[43] Andrew J Pocklington et al. "The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour". In: *National Library of Medicine* (2006). DOI: 10.1038/msb4100041.

[44] Pascal Pons and Matthieu Latapy. "Computing Communities in Large Networks Using Random Walks". en. In: *Computer and Information Sciences - ISCIS 2005*. Ed. by pInar Yolum et al. Berlin, Heidelberg: Springer, 2005, pp. 284–293. ISBN: 978-3-540-32085-2. DOI: 10.1007/11569596_31.

[45] *PubMed*. en. URL: https://pubmed.ncbi.nlm.nih.gov/ (visited on 30/03/2024).

[46] William M. Rand. "Objective Criteria for the Evaluation of Clustering Methods". In: *Journal of the American Statistical Association* 66.336 (1971). Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 846–850. ISSN: 0162-1459. DOI: 10.2307/2284239. URL: https://www.jstor.org/stable/2284239 (visited on 13/04/2024).

[47] Jörg Reichardt and Stefan Bornholdt. "Statistical mechanics of community detection". In: *Physical Review E* 74.1 (July 2006). Publisher: American Physical Society, p. 016110. DOI: 10.1103/PhysRevE.74.016110. URL: https://link.aps.org/doi/10.1103/PhysRevE.74.016110 (visited on 07/04/2024).

[48] Martin Rosvall and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure". In: *Proceedings of the National Academy of Sciences* 105.4 (Jan. 2008). Publisher: Proceedings of the National Academy of Sciences, pp. 1118–1123. DOI: 10.1073/pnas.0706851105. URL: https://www.pnas.org/doi/10.1073/pnas.0706851105 (visited on 07/04/2024).

[49] Daniel J. Schaid, Wenan Chen and Nicholas B. Larson. "From genome-wide associations to candidate causal variants by statistical fine-mapping". In: *National Library of Medicine* (2018). DOI: 10.1038/s41576-018-0016-z.

[50] Emanuel Schwarz et al. "Protein Interaction Networks Link Schizophrenia Risk Loci to Synaptic Function". In: *Schizophrenia Bulletin* 42.6 (Nov. 2016), pp. 1334–1342. ISSN: 0586-7614. DOI: 10.1093/schbul/sbw035. URL: https://doi.org/10.1093/schbul/sbw035 (visited on 09/04/2024).

[51] Amy Shakeshaft et al. "Co-development of attention deficit hyperactivity disorder and autistic trait trajectories from childhood to early adulthood". In: *Journal of Child Psychology and Psychiatry, and Allied Disciplines* 64.11 (Nov. 2023), pp. 1596–1607. ISSN: 0021-9630. DOI: 10.1111/jcpp.13851. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7615536/ (visited on 08/04/2024).

[52] Anatoly Sorokin et al. *A novel platform for meta-omics analysis. Pilot study on inhaled corticosteroids in asthma patients with side effects*. ISSN: 2693-5015. Dec. 2023. DOI: 10.21203/rs.3.rs-3706647/v1. URL: https://www.researchsquare.com/article/rs-3706647/v1 (visited on 07/04/2024).

[53] Oksana Sorokina et al. "A unified resource and configurable model of the synapse proteome and its role in disease". In: *Scientific Reports* (2021). DOI: 10.7488/ds/3017.

[54] Yauvani Sudarshan and Bernard M Y Cheung. "Hypertension and psychosis". en. In: *Postgraduate Medical Journal* 99.1171 (June 2023), pp. 411–415. ISSN: 0032-5473, 1469-0756. DOI: 10.1136/postgradmedj-2021-141386. URL: https://academic.oup.com/pmj/article/99/1171/411/7192081 (visited on 02/04/2024).

[55] Thomas C Südhof. "A molecular machine for neurotransmitter release: synaptotagmin and beyond". In: *National Library of Medicine* (2013). DOI: 10.1038/nm.3338.

[56] Thomas C Südhof and James E Rothman. "Membrane fusion: grappling with SNARE and SM proteins". In: *National Library of Medicine* (2009). DOI: 10.1126/science.1161748.

[57] Patrick F Sullivan, Kenneth S Kendler and Michael C Neale. "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies". In: *National Library of Medicine* (2003). DOI: 10.1001/archpsyc.60.12.1187.

[58] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[59] Paul D. Thomas. “The Gene Ontology and the meaning of biological function”. In: *Methods in molecular biology (Clifton, N.J.)* 1446 (2017), pp. 15–24. ISSN: 1064-3745. DOI: 10.1007/978-1-4939-3743-1_2. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6438694/ (visited on 05/04/2024).

[60] Simona Corina Trifu et al. “Genetics of schizophrenia (Review)”. In: *National Library of Medicine* (2020). DOI: 10.3892/etm.2020.8973.

[61] Vassily Trubetskoy et al. “Mapping genomic loci implicates genes and synaptic biology in schizophrenia”. In: *Nature* 604.7906 (Apr. 2022), pp. 502–508. ISSN: 0028-0836. DOI: 10.1038/s41586-022-04434-5. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9392466/ (visited on 14/03/2024).

[62] Zhe Wang, Da Li and Cheng Zhang. “Commentary: Causal associations between schizophrenia and cancers risk: a Mendelian randomization study”. In: *Frontiers in Oncology* 14 (Mar. 2024), p. 1374235. ISSN: 2234-943X. DOI: 10.3389/fonc.2024.1374235. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10985143/ (visited on 07/04/2024).

[63] Matthijs J. Warrens and Hanneke van der Hoef. “Understanding the Adjusted Rand Index and Other Partition Comparison Indices Based on Counting Object Pairs”. en. In: *Journal of Classification* 39.3 (Nov. 2022), pp. 487–509. ISSN: 1432-1343. DOI: 10.1007/s00357-022-09413-z. URL: https://doi.org/10.1007/s00357-022-09413-z (visited on 08/04/2024).

[64] Michael Waskom et al. *mwaskom/seaborn: v0.8.1 (September 2017)*. Version v0.8.1. Sept. 2017. DOI: 10.5281/zenodo.883859. URL: https://doi.org/10.5281/zenodo.883859.

[65] I. J. Weiler. “Synaptosomes”. In: *Encyclopedia of Neuroscience*. Ed. by Larry R. Squire. Oxford: Academic Press, Jan. 2009, pp. 815–818. ISBN: 978-0-08-045046-9. DOI: 10.1016/B978-008045046-9.02045-3. URL: https://www.sciencedirect.com/science/article/pii/B9780080450469020453 (visited on 15/04/2024).

[66] Xiaofei Yang et al. “Intercellular protein–protein interactions at synapses”. In: *Protein & Cell* 5.6 (June 2014), pp. 420–444. ISSN: 1674-800X. DOI: 10.1007/s13238-014-0054-z. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4026422/ (visited on 04/04/2024).

[67] Keyhan Sayadpour Zanjani. “Platelets in Pulmonary Hypertension: a Causative Role or a Simple Association?” In: *Iranian Journal of Pediatrics* (June 2012).

[68] Kai Zhou et al. “Causal associations between schizophrenia and cancers risk: a Mendelian randomization study”. In: *Frontiers in Oncology* 13 (Nov. 2023), p. 1258015. ISSN: 2234-943X. DOI: 10.3389/fonc.2023.1258015. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10693432/ (visited on 07/04/2024).

Appendix A

Graphs

A.1 Venn Diagrams Full Network

A.1.1 SCZ Broad Coding Genes

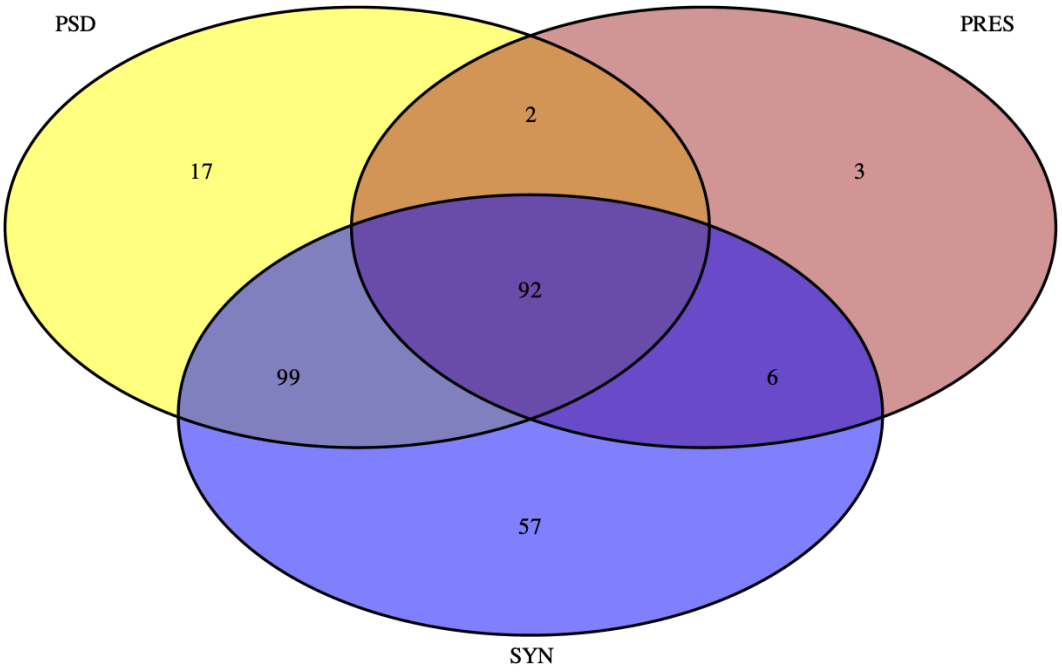


Figure A.1: SCZ Broad Coding Genes. Venn Diagram of the overlap of the genes found both from the SCZ Gene list and Synaptome DB genes. It shows the gene associations with the synapse areas, PSD is postsynaptic density which is yellow, PRES is presynaptic density which is red and SYN is the synaptosome which is blue.

A.1.2 SCZ Prioritised Coding Genes

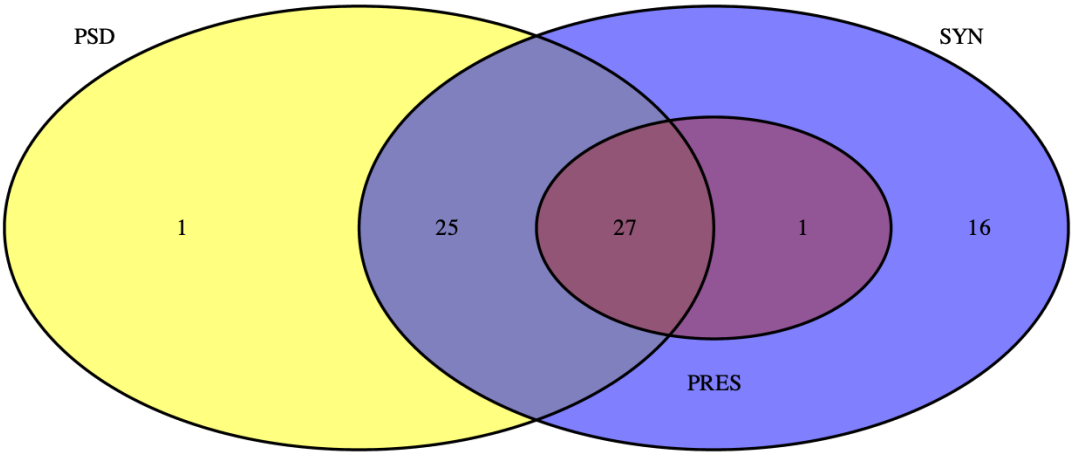


Figure A.2: SCZ Prioritised Coding Genes. Venn Diagram of the overlap of the genes found both from the SCZ Gene list and Synaptome DB genes. It shows the gene associations with the synapse areas, PSD is postsynaptic density which is yellow, PRES is presynaptic density which is red and SYN is the synaptosome which is blue.

A.2 Full network louvain SCZ example

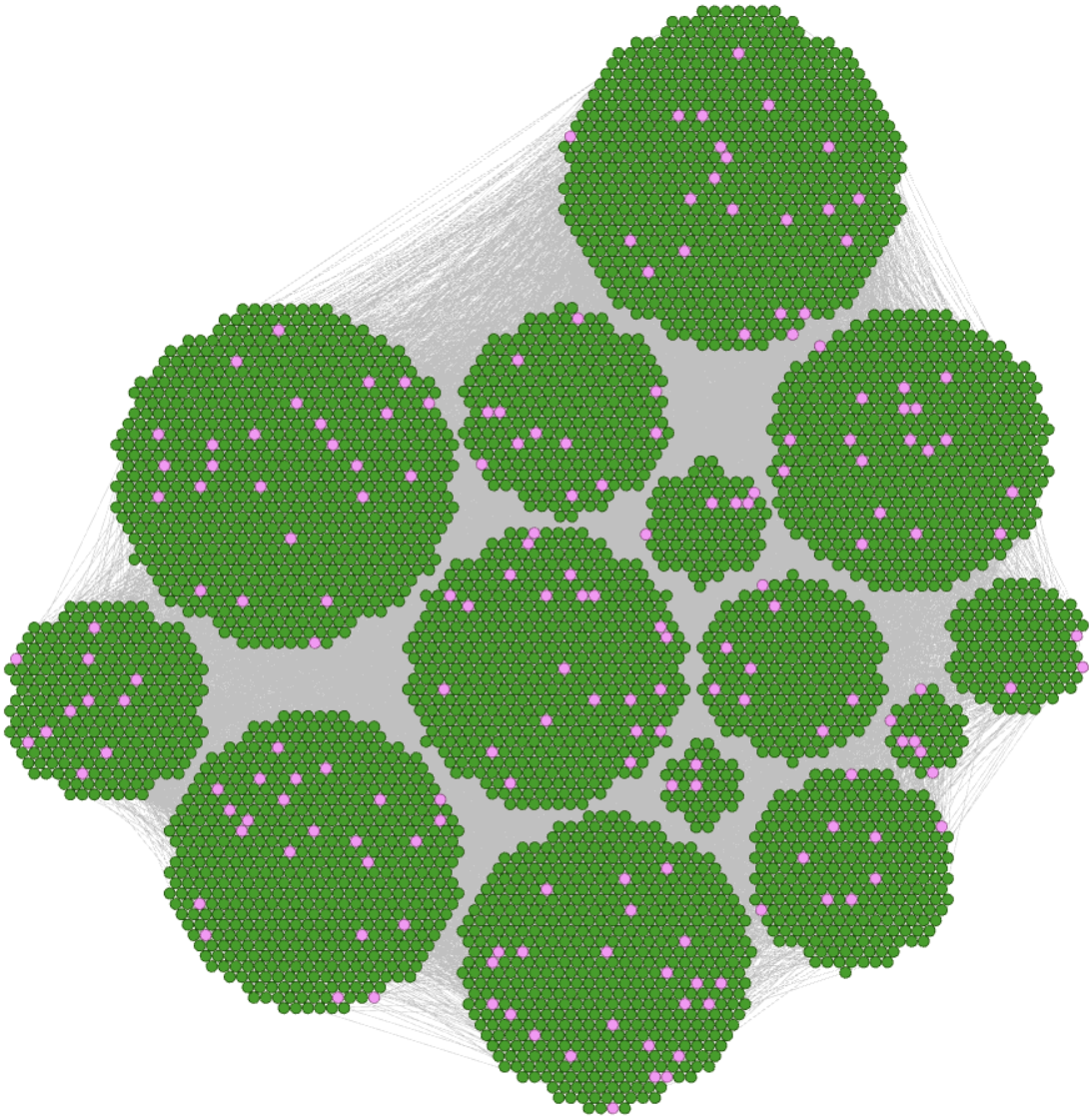
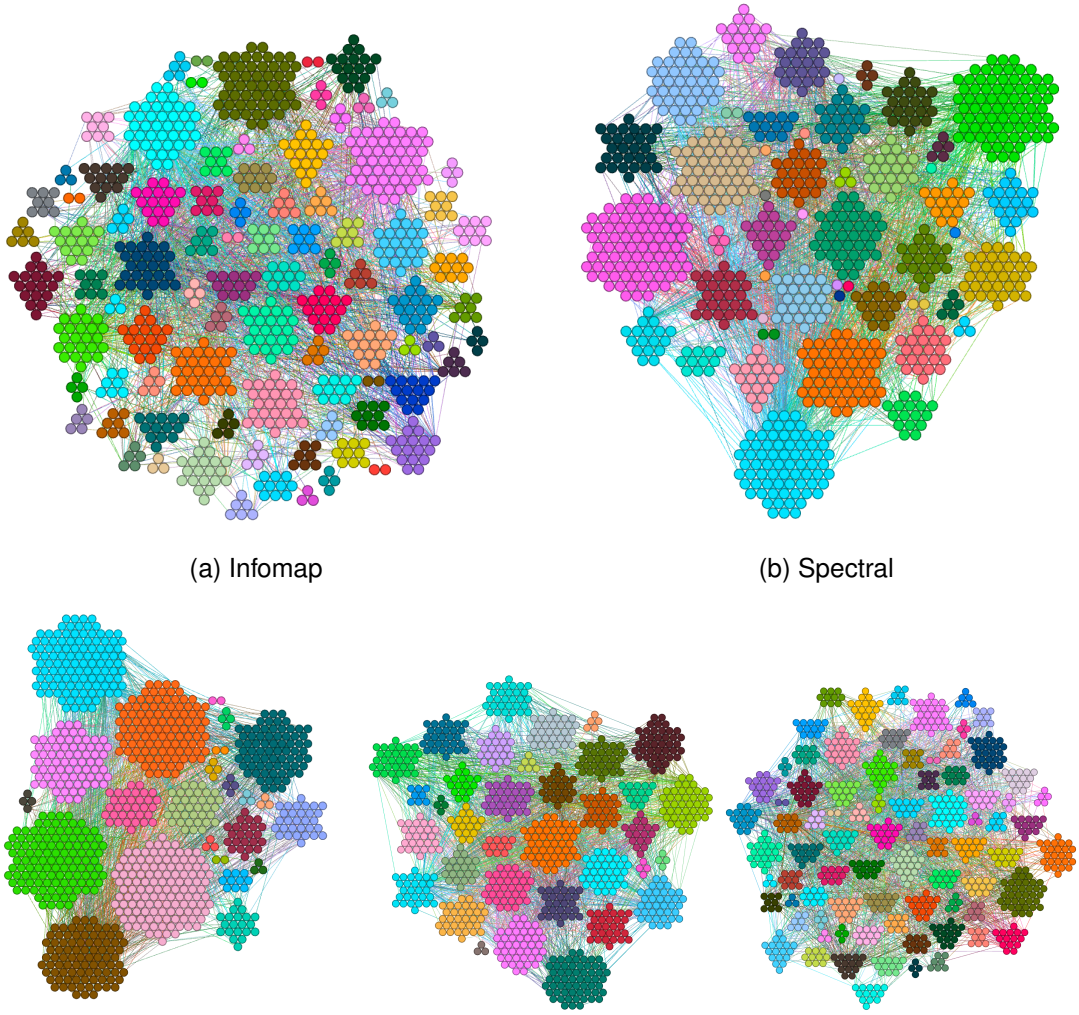


Figure A.3: Example figure of SCZ Broad associated genes for the Louvain algorithm. The green are the non SCZ genes, the pink are the SCZ genes.

A.3 SynGO Networks Visualised



(a) Infomap (b) Spectral

(c) Spin-glass 1 (d) Spin-glass 2 (e) Spin-glass 3

Figure A.4: Visualisation of each clustering algorithm generated for the Postsynaptic Network containing only genes that have SynGO annotations. Colours were generated randomly and chosen arbitrarily. The visualisation was created using Gephi.

A.4 Full Enriched SCZ Genes for the Full Network

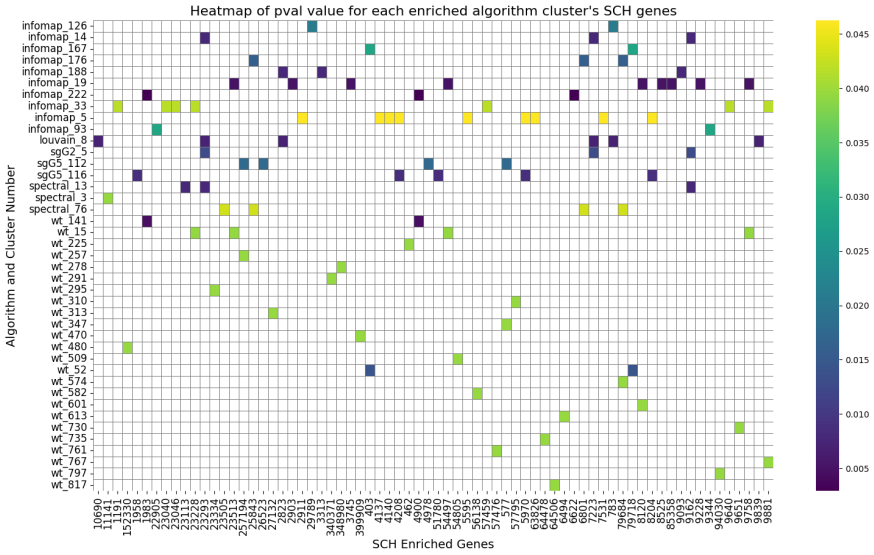


Figure A.5: Heatmap of the pval for SCZ Broad genes that show up at least once between the enriched algorithm clusters for the Full Network. The x-axis contains the Entrez ID for each SCZ Broad Gene, and the y-axis has the algorithm_clusterNumber formatted as such. The p-val is shown through the heatmap, with purple being the ”best” and lowest p-val and yellow being the highest and ”worst”.

A.5 Full Enriched SCZ Genes for the Consensus Network

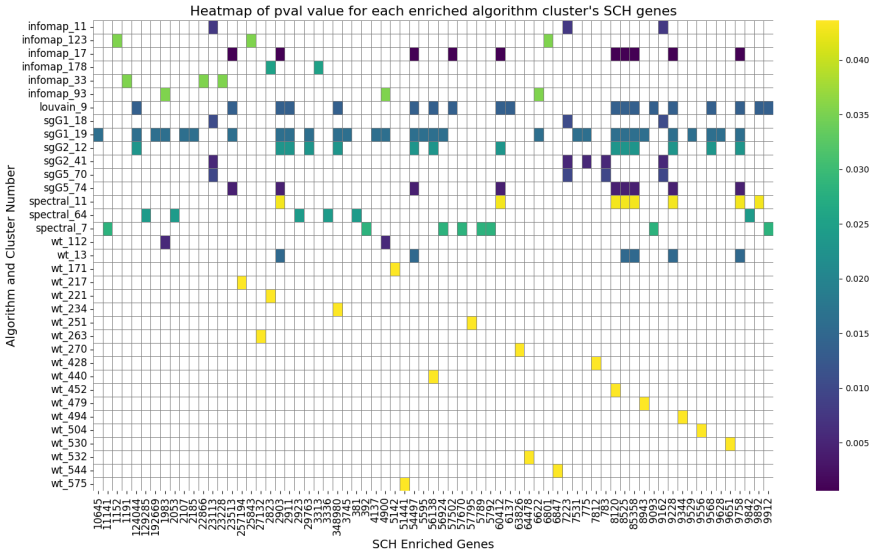


Figure A.6: Heatmap of the pval for SCZ Broad genes that show up at least once between the enriched algorithm clusters for the Consensus Network. The x-axis contains the Entrez ID for each SCZ Broad Gene, and the y-axis has the algorithm_clusterNumber formatted as such. The p-val is shown through the heatmap, with purple being the ”best” and lowest p-val and yellow being the highest and ”worst”.

A.6 Consensus Enriched SCZ Genes for the Consensus Network

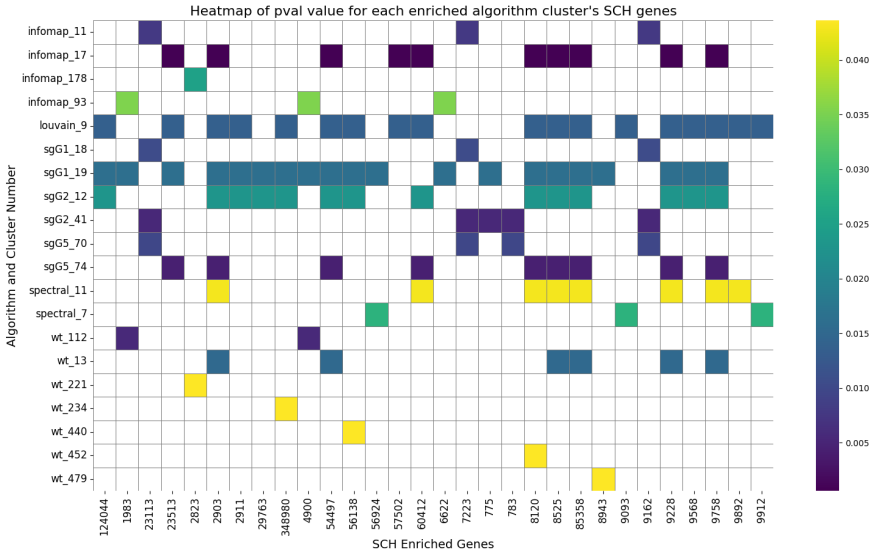


Figure A.7: Heatmap of the pval for SCZ Broad genes that show up at least twice between the enriched algorithm clusters for the Consensus Network. The x-axis contains the Entrez ID for each SCZ Broad Gene, and the y-axis has the algorithm_clusterNumber formatted as such. The p-val is shown through the heatmap, with purple being the ”best” and lowest p-val and yellow being the highest and ”worst”.

A.7 Full Enriched SCZ Genes for the SynGO Network

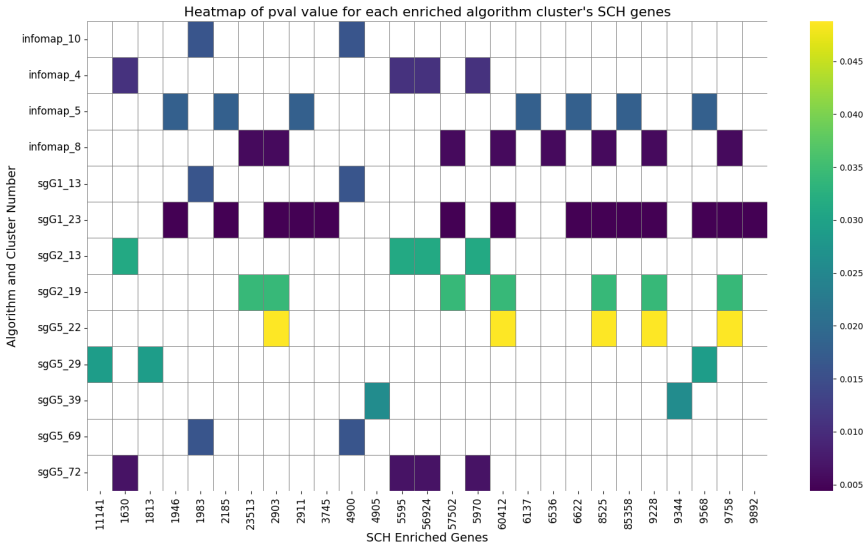
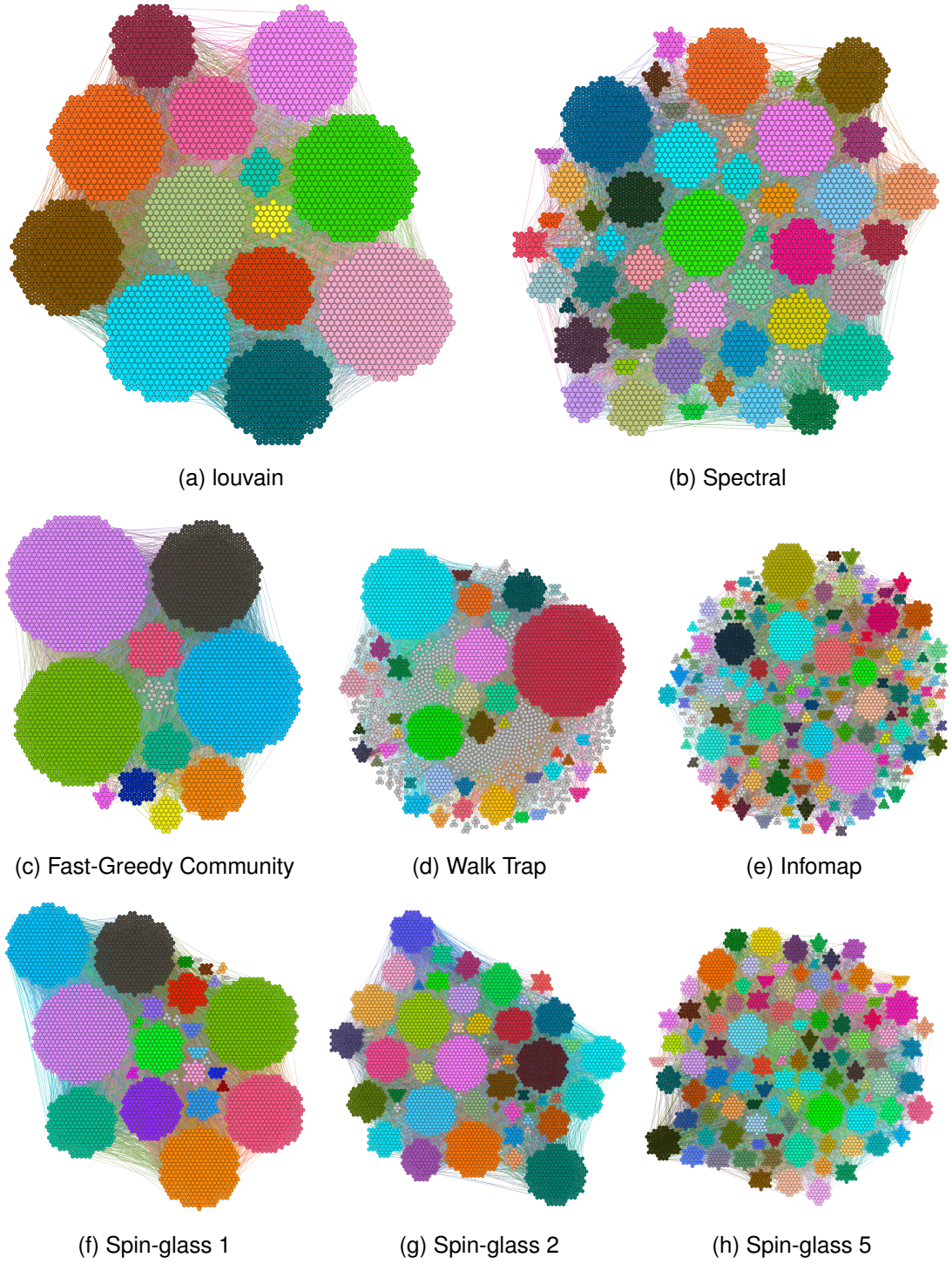


Figure A.8: Heatmap of the pval for SCZ Broad genes that show up at least once between the enriched algorithm clusters for the SynGO Network. The x-axis contains the Entrez ID for each SCZ Broad Gene, and the y-axis has the algorithm_clusterNumber formatted as such. The p-val is shown through the heatmap, with purple being the ”best” and lowest p-val and yellow being the highest and ”worst”.

A.8 Consensus Network Clustering Visualisation



(a) louvain (b) Spectral

(c) Fast-Greedy Community (d) Walk Trap (e) Infomap

(f) Spin-glass 1 (g) Spin-glass 2 (h) Spin-glass 5

Figure A.9: Visualisation of each clustering algorithm generated for the consensus postsynaptic network. Colours were generated randomly and chosen arbitrarily. For a some algorithms, those with cluster sizes of 4 or less are visualised in grey.

A.9 WordClouds

Wordclouds created for the SCZ Broad enriched clusters. The enriched terms of the networks are visualised by word size, which is weighted by frequency and 1 - pval.

A.9.1 Consensus

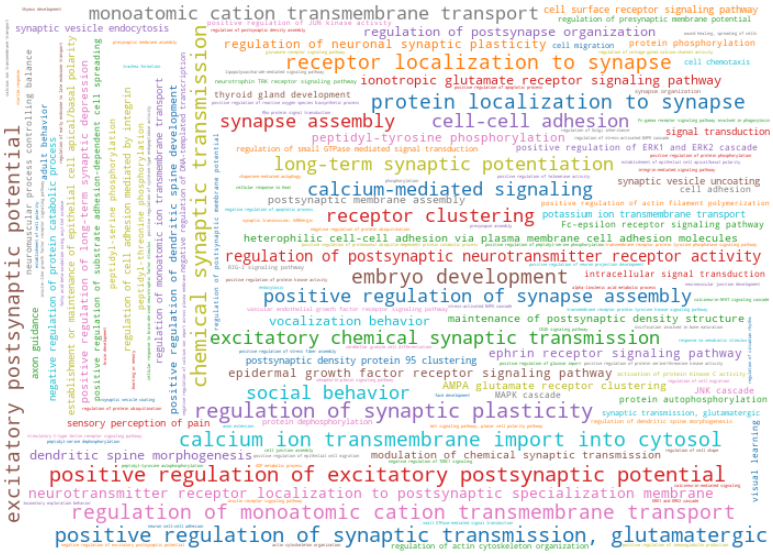


Figure A.10: Wordcloud for the consensus network for Biological function. The data can be found at B.12

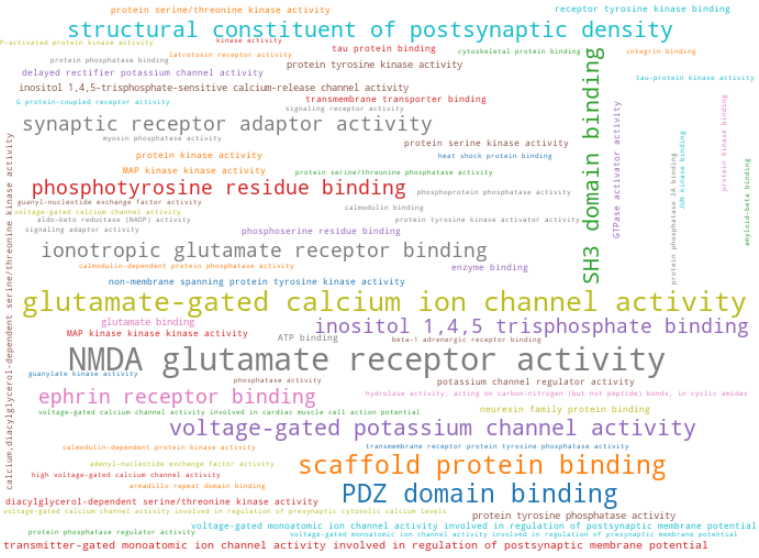


Figure A.11: Wordcloud for the consensus network for Molecular function. The data can be found at B.11.

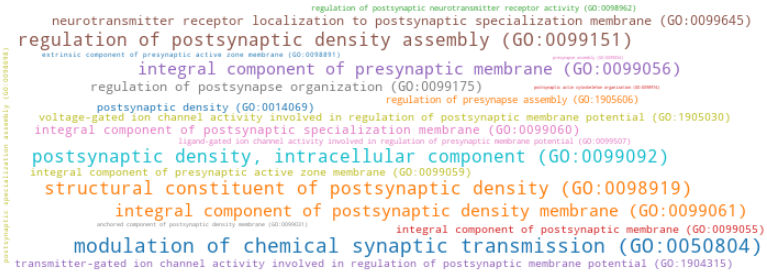


Figure A.12: Wordcloud for the consensus network for SynGO annotations. The data can be found at B.14.

Figure A.13: Wordcloud for the consensus network for Disease annotations. The data can be found at B.13.

Appendix B

Tables

B.1 Venn Diagram Table

	PSD Total	PRES Total	SYN Total	
SCZ Broad	210	103	254	
SCZ Prioritised	53	28	69	

Table B.1: Table containing the number of SCZ genes that overlap with synaptome DB. The genes are categorised by where in the synapse they appear.

B.2 Consensus Enrichment

Algorithm	Cluster Number	Cluster Size	SCZ Broad No.	FL	padj	pval
fc	1	642	0	False	0.03186	0.01593
infomap	17	68	10	True	0.00119	0.00059
infomap	11	10	3	True	0.01554	0.00777
infomap	178	6	2	True	0.05048	0.02524
infomap	33	17	3	True	0.07047	0.03524
infomap	93	17	3	True	0.07047	0.03524
infomap	123	17	3	True	0.07047	0.03524
louvain	9	256	19	True	0.02739	0.01370
sgG1	22	126	0	False	0.00642	0.00321
sgG1	18	11	3	True	0.02070	0.01035
sgG1	19	469	30	True	0.03234	0.01617
sgG2	41	27	5	True	0.01089	0.00544
sgG2	17	92	0	False	0.03097	0.01549
sgG2	12	180	14	True	0.04623	0.02311
sgG5	74	74	9	True	0.00862	0.00431
sgG5	70	20	4	True	0.01945	0.00972
spectral	64	52	6	True	0.04812	0.02406
spectral	7	84	8	True	0.05664	0.02832
spectral	32	76	0	False	0.06455	0.03228
spectral	11	91	8	True	0.08572	0.04286
wt	112	3	2	True	0.01101	0.00550
wt	13	47	6	True	0.03030	0.01515
wt	171	1	1	True	0.08724	0.04362
wt	217	1	1	True	0.08724	0.04362
wt	221	1	1	True	0.08724	0.04362
wt	234	1	1	True	0.08724	0.04362
wt	251	1	1	True	0.08724	0.04362
wt	263	1	1	True	0.08724	0.04362
wt	270	1	1	True	0.08724	0.04362
wt	428	1	1	True	0.08724	0.04362
wt	440	1	1	True	0.08724	0.04362
wt	452	1	1	True	0.08724	0.04362
wt	479	1	1	True	0.08724	0.04362
wt	494	1	1	True	0.08724	0.04362
wt	504	1	1	True	0.08724	0.04362
wt	530	1	1	True	0.08724	0.04362
wt	532	1	1	True	0.08724	0.04362
wt	544	1	1	True	0.08724	0.04362
wt	575	1	1	True	0.08724	0.04362

Table B.2: SCZ Broad enriched clusters for the Consensus Postsynaptic Algorithms and Clusters. It includes the Algorithm, Cluster number, cluster size and number of SCZ Broad genes, the FL which indicates if its positive or negative enrichment, the padj and pval. The padj and pval are rounded to 5 decimal places.

B.3 Full Network SCZ Broad Enriched Genes

Table B.3: SCZ Broad genes from enriched clusters found from the Full Postsynaptic PSD network, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
23293	SMG6	infomap 14, louvain 8, sgG2 5, spectral 13	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. Alternatively spliced transcript variants encoding distinct protein isoforms have been described. [provided by RefSeq, Feb 2014]
7223	TRPC4	infomap 14, louvain 8, sgG2 5	This gene encodes a member of the canonical subfamily of transient receptor potential cation channels. The encoded protein forms a non-selective calcium-permeable cation channel that is activated by Gq-coupled receptors and tyrosine kinases, and plays a role in multiple processes including endothelial permeability, vasodilation, neurotransmitter release and cell proliferation. Single nucleotide polymorphisms in this gene may be associated with generalized epilepsy with photosensitivity. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]
79684	MSANTD2	infomap 176, spectral 76, wt 574	N/A

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9162	DGKI	infomap 14, sgG2 5, spectral 13	This gene is a member of the type IV diacylglycerol kinase subfamily. Diacylglycerol kinases regulate the intracellular concentration of diacylglycerol through its phosphorylation, producing phosphatidic acid. The specific role of the enzyme encoded by this gene is undetermined, however, it may play a crucial role in the production of phosphatidic acid in the retina or in recessive forms of retinal degeneration. [provided by RefSeq, Jul 2008]
1983	EIF5	infomap 222, wt 141	Eukaryotic translation initiation factor-5 (EIF5) interacts with the 40S initiation complex to promote hydrolysis of bound GTP with concomitant joining of the 60S ribosomal subunit to the 40S initiation complex. The resulting functional 80S ribosomal initiation complex is then active in peptidyl transfer and chain elongations (summary by Si et al., 1996 [PubMed 8663286]).[supplied by OMIM, May 2010]
23228	PLCL2	infomap 33, wt 15	Enables GABA receptor binding activity. Predicted to be involved in negative regulation of cold-induced thermogenesis and phosphatidylinositol-mediated signaling. Predicted to act upstream of or within several processes, including B cell activation; gamma-aminobutyric acid signaling pathway; and negative regulation of B cell receptor signaling pathway. Predicted to be located in cytoplasm. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
23513	SCRIB	infomap 19, wt 15	This gene encodes a protein that was identified as being similar to the Drosophila scribble protein. The mammalian protein is involved in tumor suppression pathways. As a scaffold protein involved in cell polarization processes, this protein binds to many other proteins. The encoded protein binds to papillomavirus E6 protein via its PDZ domain and the C-terminus of E6. Two alternatively spliced transcript variants that encode different protein isoforms have been found for this gene. [provided by RefSeq, Nov 2011]
257194	NEGR1	sgG5 112, wt 257	Predicted to act upstream of or within several processes, including feeding behavior; locomotory behavior; and positive regulation of neuron projection development. Predicted to be located in extracellular region and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]
25843	MOB4	infomap 176, spectral 76	This gene was identified based on its similarity with the mouse counterpart. Studies of the mouse counterpart suggest that the expression of this gene may be regulated during oocyte maturation and preimplantation following zygotic gene activation. Alternatively spliced transcript variants encoding distinct isoforms have been observed. Naturally occurring read-through transcription occurs between this locus and the neighboring locus HSPE1.[provided by RefSeq, Feb 2011]
2823	GPM6A	infomap 188, louvain 8	Predicted to enable calcium channel activity. Involved in neuron migration and stem cell differentiation. Located in extracellular exosome. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
403	ARL3	infomap 167, wt 52	Enables GDP binding activity; GTP binding activity; and microtubule binding activity. Involved in several processes, including cilium assembly; protein localization to cilium; and small GTPase mediated signal transduction. Acts upstream of or within post-Golgi vesicle-mediated transport. Located in several cellular components, including microtubule cytoskeleton; midbody; and photoreceptor connecting cilium. Implicated in Joubert syndrome and retinitis pigmentosa 83. [provided by Alliance of Genome Resources, Apr 2022]
4208	MEF2C	infomap 5, sgG5 116	This locus 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. Alternatively spliced transcript variants have been described. [provided by RefSeq, Jul 2010]
4900	NRGN	infomap 222, wt 141	Neurogranin (NRGN) is the human homolog of the neuron-specific rat RC3/neurogranin gene. This gene encodes a postsynaptic protein kinase substrate that binds calmodulin in the absence of calcium. The NRGN gene contains four exons and three introns. The exons 1 and 2 encode the protein and exons 3 and 4 contain untranslated sequences. It is suggested that the NRGN is a direct target for thyroid hormone in human brain, and that control of expression of this gene could underlay many of the consequences of hypothyroidism on mental states during development as well as in adult subjects. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
54497	HEATR5B	infomap 19, wt 15	Predicted to be involved in endocytosis; protein localization; and retrograde transport, endosome to Golgi. Located in membrane. [provided by Alliance of Genome Resources, Apr 2022]
577	ADGRB3	sgG5 112, wt 347	This p53-target gene encodes a brain-specific angiogenesis inhibitor, a seven-span transmembrane protein, and is thought to be a member of the secretin receptor family. Brain-specific angiogenesis proteins BAI2 and BAI3 are similar to BAI1 in structure, have similar tissue specificities, and may also play a role in angiogenesis. [provided by RefSeq, Jul 2008]
5970	RELA	infomap 5, sgG5 116	NF-kappa-B is a ubiquitous transcription factor involved in several biological processes. It is held in the cytoplasm in an inactive state by specific inhibitors. Upon degradation of the inhibitor, NF-kappa-B moves to the nucleus and activates transcription of specific genes. NF-kappa-B is composed of NFKB1 or NFKB2 bound to either REL, RELA, or RELB. The most abundant form of NF-kappa-B is NFKB1 complexed with the product of this gene, RELA. Four transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Sep 2011]
6801	STRN	infomap 176, spectral 76	Enables armadillo repeat domain binding activity; estrogen receptor binding activity; and protein phosphatase 2A binding activity. Involved in Wnt signaling pathway and negative regulation of cell population proliferation. Located in bicellular tight junction. Part of FAR/SIN/STRIPAK complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
783	CACNB2	infomap 126, louvain 8	This gene encodes a subunit of a voltage-dependent calcium channel protein that is a member of the voltage-gated calcium channel superfamily. The gene product was originally identified as an antigen target in Lambert-Eaton myasthenic syndrome, an autoimmune disorder. Mutations in this gene are associated with Brugada syndrome. Alternatively spliced variants encoding different isoforms have been described. [provided by RefSeq, Feb 2013]
79718	TBL1XR1	infomap 167, wt 52	This gene is a member of the WD40 repeat-containing gene family and shares sequence similarity with transducin (beta)-like 1X-linked (TBL1X). The protein encoded by this gene is thought to be a component of both nuclear receptor corepressor (N-CoR) and histone deacetylase 3 (HDAC 3) complexes, and is required for transcriptional activation by a variety of transcription factors. Mutations in these gene have been associated with some autism spectrum disorders, and one finding suggests that haploinsufficiency of this gene may be a cause of intellectual disability with dysmorphism. Mutations in this gene as well as recurrent translocations involving this gene have also been observed in some tumors. [provided by RefSeq, Mar 2016]
8120	AP3B2	infomap 19, wt 601	Adaptor protein complex 3 (AP-3 complex) is a heterotrimeric protein complex involved in the formation of clathrin-coated synaptic vesicles. The protein encoded by this gene represents the beta subunit of the neuron-specific AP-3 complex and was first identified as the target antigen in human paraneoplastic neurologic disorders. The encoded subunit binds clathrin and is phosphorylated by a casein kinase-like protein, which mediates synaptic vesicle coat assembly. Defects in this gene are a cause of early-onset epileptic encephalopathy. [provided by RefSeq, Feb 2017]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
8204	NRIP1	infomap 5, sgG5 116	Nuclear receptor interacting protein 1 (NRIP1) is a nuclear protein that specifically interacts with the hormone-dependent activation domain AF2 of nuclear receptors. Also known as RIP140, this protein modulates transcriptional activity of the estrogen receptor. [provided by RefSeq, Jul 2008]
9758	FRMPD4	infomap 19, wt 15	This gene encodes a multi-domain (WW, PDZ, FERM) containing protein. Through its interaction with other proteins (such as PSD-95), it functions as a positive regulator of dendritic spine morphogenesis and density, and is required for the maintenance of excitatory synaptic transmission. [provided by RefSeq, Jan 2010]
9881	TRANK1	infomap 33, wt 767	Predicted to enable ATP binding activity. [provided by Alliance of Genome Resources, Apr 2022]
10690	FUT9	louvain 8	The protein encoded by this gene belongs to the glycosyltransferase family. It is localized to the golgi, and catalyzes the last step in the biosynthesis of Lewis X (LeX) antigen, the addition of a fucose to precursor polysaccharides. This protein is one of the few fucosyltransferases that synthesizes the LeX oligosaccharide (CD15) expressed in the organ buds progressing in mesenchyma during embryogenesis. It is also responsible for the expression of CD15 in mature granulocytes. A common haplotype of this gene has also been associated with susceptibility to placental malaria infection. [provided by RefSeq, Nov 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
11141	IL1RAPL1	spectral 3	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]
1191	CLU	infomap 33	The protein encoded by this gene is a secreted chaperone that can under some stress conditions also be found in the cell cytosol. It has been suggested to be involved in several basic biological events such as cell death, tumor progression, and neurodegenerative disorders. Alternate splicing results in both coding and non-coding variants.[provided by RefSeq, May 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
152330	CNTN4	wt 480	This gene encodes a member of the contactin family of immunoglobulins. Contactins are axon-associated cell adhesion molecules that function in neuronal network formation and plasticity. The encoded protein is a glycosylphosphatidylinositol-anchored neuronal membrane protein that may play a role in the formation of axon connections in the developing nervous system. Deletion or mutation of this gene may play a role in 3p deletion syndrome and autism spectrum disorders. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2011]
1958	EGR1	sgG5 116	The protein encoded by this gene belongs to the EGR family of C2H2-type zinc-finger proteins. It is a nuclear protein and functions as a transcriptional regulator. The products of target genes it activates are required for differentitation and mitogenesis. Studies suggest this is a cancer suppressor gene. [provided by RefSeq, Dec 2014]
22905	EPN2	infomap 93	This gene encodes a protein which interacts with clathrin and adaptor-related protein complex 2, alpha 1 subunit. The protein is found in a brain-derived clathrin-coated vesicle fraction and localizes to the peri-Golgi region and the cell periphery. The protein is thought to be involved in clathrin-mediated endocytosis. Alternate splicing of this gene results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
23040	MYT1L	infomap 33	This gene encodes a member of the zinc finger superfamily of transcription factors whose expression, thus far, has been found only in neuronal tissues. The encoded protein belongs to a novel class of cystein-cystein-histidine-cystein zinc finger proteins that function in the developing mammalian central nervous system. Forced expression of this gene in combination with the basic helix-loop-helix transcription factor NeuroD1 and the transcription factors POU class 3 homeobox 2 and achaete-scute family basic helix-loop-helix transcription factor 1 can convert fetal and postnatal human fibroblasts into induced neuronal cells, which are able to generate action potentials. Mutations in this gene have been associated with an autosomal dominant form of cognitive disability and with autism spectrum disorder. Alternative splicing results in multiple variants. [provided by RefSeq, Jul 2017]
23046	KIF21B	infomap 33	This gene encodes a member of the kinesin superfamily. Kinesins are ATP-dependent microtubule-based motor proteins that are involved in the intracellular transport of membranous organelles. Single nucleotide polymorphisms in this gene are associated with inflammatory bowel disease and multiple sclerosis. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Nov 2011]
23113	CUL9	spectral 13	Predicted to enable several functions, including ATP binding activity; metal ion binding activity; and ubiquitin protein ligase binding activity. Involved in microtubule cytoskeleton organization; protein ubiquitination; and regulation of mitotic nuclear division. Located in cytosol. Part of cullin-RING ubiquitin ligase complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
23334	SZT2	wt 295	The protein encoded by this gene is expressed in the brain, predominantly in the parietal and frontal cortex as well as in dorsal root ganglia. It is localized to the peroxisome, and is implicated in resistance to oxidative stress. It likely functions by increasing superoxide dismutase (SOD) activity, but itself has no direct SOD activity. Studies in mice show that this gene confers low seizure threshold, and may also enhance epileptogenesis. [provided by RefSeq, Jun 2011]
23505	TMEM131	spectral 76	Predicted to be integral component of membrane. Predicted to be active in membrane. [provided by Alliance of Genome Resources, Apr 2022]
26523	AGO1	sgG5 112	This gene encodes a member of the argonaute family of proteins, which associate with small RNAs and have important roles in RNA interference (RNAi) and RNA silencing. This protein binds to microRNAs (miRNAs) or small interfering RNAs (siRNAs) and represses translation of mRNAs that are complementary to them. It is also involved in transcriptional gene silencing (TGS) of promoter regions that are complementary to bound short antigene RNAs (agRNAs), as well as in the degradation of miRNA-bound mRNA targets. Alternatively spliced transcript variants encoding different isoforms have been found for this gene. A recent study showed this gene to be an authentic stop codon readthrough target, and that its mRNA could give rise to an additional C-terminally extended isoform by use of an alternative in-frame translation termination codon. [provided by RefSeq, Nov 2015]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
27132	CPNE7	wt 313	This gene encodes a member of the copine family, which is composed of calcium-dependent membrane-binding proteins. The gene product contains two N-terminal C2 domains and one von Willebrand factor A domain. The encoded protein may be involved in membrane trafficking. Two alternatively spliced transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Nov 2008]
2903	GRIN2A	infomap 19	This gene encodes a member of the glutamate-gated ion channel protein family. The encoded protein is an N-methyl-D-aspartate (NMDA) receptor subunit. NMDA receptors are both ligand-gated and voltage-dependent, and are involved in long-term potentiation, an activity-dependent increase in the efficiency of synaptic transmission thought to underlie certain kinds of memory and learning. These receptors are permeable to calcium ions, and activation results in a calcium influx into post-synaptic cells, which results in the activation of several signaling cascades. Disruption of this gene is associated with focal epilepsy and speech disorder with or without cognitive disability. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2014]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2911	GRM1	infomap 5	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]
29789	OLA1	infomap 126	This gene encodes a member of the GTPase protein family. The encoded protein interacts with breast cancer-associated gene 1 (BRCA1) and BRCA1-associated RING domain protein (BARD1), and is involved in centrosome regulation. Overexpression of this gene has been observed in multiple types of cancer and may be associated with poor survival. Pseudogenes of this gene have been defined on chromosomes 17 and 22. [provided by RefSeq, Jun 2016]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
3313	HSPA9	infomap 188	This gene encodes a member of the heat shock protein 70 gene family. The encoded protein is primarily localized to the mitochondria but is also found in the endoplasmic reticulum, plasma membrane and cytoplasmic vesicles. This protein is a heat-shock cognate protein. This protein plays a role in cell proliferation, stress response and maintenance of the mitochondria. A pseudogene of this gene is found on chromosome 2.[provided by RefSeq, May 2010]
340371	NRBP2	wt 291	Predicted to enable protein serine/threonine kinase activity. Involved in negative regulation of macroautophagy. Predicted to be active in cytoplasm and endomembrane system. [provided by Alliance of Genome Resources, Apr 2022]
348980	HCN1	wt 278	The membrane protein encoded by this gene is a hyperpolarization-activated cation channel that contributes to the native pacemaker currents in heart and neurons. The encoded protein can homodimerize or heterodimerize with other pore-forming subunits to form a potassium channel. This channel may act as a receptor for sour tastes. [provided by RefSeq, Oct 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
3745	KCNB1	infomap 19	Voltage-gated potassium (Kv) channels represent the most complex class of voltage-gated ion channels from both functional and structural standpoints. Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume. Four sequence-related potassium channel genes - shaker, shaw, shab, and shal - have been identified in Drosophila, and each has been shown to have human homolog(s). This gene encodes a member of the potassium channel, voltage-gated, shab-related subfamily. This member is a delayed rectifier potassium channel and its activity is modulated by some other family members. [provided by RefSeq, Jul 2008]
399909	PCNX3	wt 470	Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]
4137	MAPT	infomap 5	This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. MAPT gene mutations have been associated with several neurodegenerative disorders such as Alzheimer's disease, Pick's disease, frontotemporal dementia, cortico-basal degeneration and progressive supranuclear palsy. [provided by RefSeq, Jul 2008]
4140	MARK3	infomap 5	The protein encoded by this gene is activated by phosphorylation and in turn is involved in the phosphorylation of tau proteins MAP2 and MAP4. Several transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Oct 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
462	SERPINC1	wt 225	The protein encoded by this gene, antithrombin III, is a plasma protease inhibitor and a member of the serpin superfamily. This protein inhibits thrombin as well as other activated serine proteases of the coagulation system, and it regulates the blood coagulation cascade. The protein includes two functional domains: the heparin binding-domain at the N-terminus of the mature protein, and the reactive site domain at the C-terminus. The inhibitory activity is enhanced by the presence of heparin. Numerous mutations have been identified for this gene, many of which are known to cause antithrombin-III deficiency which constitutes a strong risk factor for thrombosis. A reduction in the serum level of this protein is associated with severe cases of Coronavirus Disease 19 (COVID-19). [provided by RefSeq, Sep 2020]
4978	OPCML	sgG5 112	This gene encodes a member of the IgLON subfamily in the immunoglobulin protein superfamily of proteins. The encoded preprotein is proteolytically processed to generate the mature protein. This protein is localized in the plasma membrane and may have an accessory role in opioid receptor function. This gene has an ortholog in rat and bovine. The opioid binding-cell adhesion molecule encoded by the rat gene binds opioid alkaloids in the presence of acidic lipids, exhibits selectivity for mu ligands and acts as a GPI-anchored protein. Since the encoded protein is highly conserved in species during evolution, it may have a fundamental role in mammalian systems. Alternative splicing results in multiple transcript variants, at least one of which encodes an isoform that is proteolytically processed. [provided by RefSeq, Jan 2016]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
51780	KDM3B	sgG5 116	Predicted to enable chromatin DNA binding activity; histone H3-methyl-lysine-9 demethylase activity; and transcription coregulator activity. Predicted to be involved in histone H3-K9 demethylation and regulation of transcription by RNA polymerase II. Located in nucleoplasm. Biomarker of acute lymphoblastic leukemia; breast cancer; colorectal cancer; and lung non-small cell carcinoma. [provided by Alliance of Genome Resources, Apr 2022]
54805	CNNM2	wt 509	This gene encodes a member of the ancient conserved domain containing protein family. Members of this protein family contain a cyclin box motif and have structural similarity to the cyclins. The encoded protein may play an important role in magnesium homeostasis by mediating the epithelial transport and renal reabsorption of Mg2+. Mutations in this gene are associated with renal hypomagnesemia. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Dec 2011]
5595	MAPK3	infomap 5	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
56138	PCDHA11	wt 582	This gene is a member of the protocadherin alpha gene cluster, one of three related gene clusters tandemly linked on chromosome five that demonstrate an unusual genomic organization similar to that of B-cell and T-cell receptor gene clusters. The alpha gene cluster is composed of 15 cadherin superfamily genes related to the mouse CNR genes and consists of 13 highly similar and 2 more distantly related coding sequences. The tandem array of 15 N-terminal exons, or variable exons, are followed by downstream C-terminal exons, or constant exons, which are shared by all genes in the cluster. The large, uninterrupted N-terminal exons each encode six cadherin ectodomains while the C-terminal exons encode the cytoplasmic domain. These neural cadherin-like cell adhesion proteins are integral plasma membrane proteins that most likely play a critical role in the establishment and function of specific cell-cell connections in the brain. Alternative splicing has been observed and additional variants have been suggested but their full-length nature has yet to be determined. [provided by RefSeq, Jul 2008]
57459	GATAD2B	infomap 33	This gene encodes a zinc finger protein transcriptional repressor. The encoded protein is part of the methyl-CpG-binding protein-1 complex, which represses gene expression by deacetylating methylated nucleosomes. Mutations in this gene are linked to intellectual disability and dysmorphic features associated with cognitive disability. [provided by RefSeq, Jun 2016]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
57476	GRAMD1B	wt 761	Predicted to enable cholesterol binding activity; cholesterol transfer activity; and phospholipid binding activity. Predicted to be involved in cellular response to cholesterol and cholesterol homeostasis. Located in endoplasmic reticulum membrane; endoplasmic reticulum-plasma membrane contact site; and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]
57795	BRINP2	wt 310	Predicted to be involved in cellular response to retinoic acid; negative regulation of mitotic cell cycle; and positive regulation of neuron differentiation. Predicted to be located in extracellular region. Predicted to be active in dendrite; endoplasmic reticulum; and neuronal cell body. Implicated in oral squamous cell carcinoma. [provided by Alliance of Genome Resources, Apr 2022]
63826	SRR	infomap 5	Enables several functions, including L-serine ammonia-lyase activity; PDZ domain binding activity; and anion binding activity. Involved in pyruvate biosynthetic process; response to lipopolysaccharide; and serine family amino acid metabolic process. Located in cytoplasm and neuronal cell body. [provided by Alliance of Genome Resources, Apr 2022]
64478	CSMD1	wt 735	Predicted to act upstream of or within several processes, including learning or memory; mammary gland branching involved in pregnancy; and reproductive structure development. Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
64506	CPEB1	wt 817	This gene encodes a member of the cytoplasmic polyadenylation element binding protein family. This highly conserved protein binds to a specific RNA sequence, called the cytoplasmic polyadenylation element, found in the 3' untranslated region of some mRNAs. The encoded protein functions in both the cytoplasm and the nucleus. It is involved in the regulation of mRNA translation, as well as processing of the 3' untranslated region, and may play a role in cell proliferation and tumorigenesis. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jan 2014]
6494	SIPA1	wt 613	The product of this gene is a mitogen induced GTPase activating protein (GAP). It exhibits a specific GAP activity for Ras-related regulatory proteins Rap1 and Rap2, but not for Ran or other small GTPases. This protein may also hamper mitogen-induced cell cycle progression when abnormally or prematurely expressed. It is localized to the perinuclear region. Two alternatively spliced variants encoding the same isoform have been characterized to date. [provided by RefSeq, Jul 2008]
6622	SNCA	infomap 222	Alpha-synuclein is a member of the synuclein family, which also includes beta- and gamma-synuclein. Synucleins are abundantly expressed in the brain and alpha- and beta-synuclein inhibit phospholipase D2 selectively. SNCA may serve to integrate presynaptic signaling and membrane trafficking. Defects in SNCA have been implicated in the pathogenesis of Parkinson disease. SNCA peptides are a major component of amyloid plaques in the brains of patients with Alzheimer's disease. Alternatively spliced transcripts encoding different isoforms have been identified for this gene. [provided by RefSeq, Feb 2016]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
7531	YWHAE	infomap 5	This gene product belongs to the 14-3-3 family of proteins which mediate signal transduction by binding to phosphoserine-containing proteins. This highly conserved protein family is found in both plants and mammals, and this protein is 100% identical to the mouse ortholog. It interacts with CDC25 phosphatases, RAF1 and IRS1 proteins, suggesting its role in diverse biochemical activities related to signal transduction, such as cell division and regulation of insulin sensitivity. It has also been implicated in the pathogenesis of small cell lung cancer. Two transcript variants, one protein-coding and the other non-protein-coding, have been found for this gene. [provided by RefSeq, Aug 2008]
8525	DGKZ	infomap 19	The protein encoded by this gene belongs to the eukaryotic diacylglycerol kinase family. It may attenuate protein kinase C activity by regulating diacylglycerol levels in intracellular signaling cascade and signal transduction. Alternative splicing occurs at this locus and multiple transcript variants encoding distinct isoforms have been identified. [provided by RefSeq, Nov 2010]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
85358	SHANK3	infomap 19	This gene is a member of the Shank gene family. Shank proteins are multidomain scaffold proteins of the postsynaptic density that connect neurotransmitter receptors, ion channels, and other membrane proteins to the actin cytoskeleton and G-protein-coupled signaling pathways. Shank proteins also play a role in synapse formation and dendritic spine maturation. Mutations in this gene are a cause of autism spectrum disorder (ASD), which is characterized by impairments in social interaction and communication, and restricted behavioral patterns and interests. Mutations in this gene also cause schizophrenia type 15, and are a major causative factor in the neurological symptoms of 22q13.3 deletion syndrome, which is also known as Phelan-McDermid syndrome. Additional isoforms have been described for this gene but they have not yet been experimentally verified. [provided by RefSeq, Mar 2012]
9093	DNAJA3	infomap 188	This gene encodes a member of the DNAJ/Hsp40 protein family. DNAJ/Hsp40 proteins stimulate the ATPase activity of Hsp70 chaperones and play critical roles in protein folding, degradation, and multimeric complex assembly. The encoded protein is localized to mitochondria and mediates several cellular processes including proliferation, survival and apoptotic signal transduction. The encoded protein also plays a critical role in tumor suppression through interactions with oncogenic proteins including ErbB2 and the p53 tumor suppressor protein. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9228	DLGAP2	infomap 19	The product of this gene is a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain, however, only the paternal allele is expressed in the testis (PMID:18055845). Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jun 2014]
9344	TAOK2	infomap 93	Enables mitogen-activated protein kinase kinase binding activity; neuropilin binding activity; and protein serine/threonine kinase activity. Involved in several processes, including focal adhesion assembly; intracellular signal transduction; and positive regulation of JNK cascade. Located in cytoplasmic vesicle; cytosol; and nuclear lumen. Part of receptor complex. [provided by Alliance of Genome Resources, Apr 2022]
94030	LRRC4B	wt 797	Predicted to enable signaling receptor binding activity. Predicted to be involved in regulation of synapse assembly and synaptic membrane adhesion. Predicted to be located in cerebellar mossy fiber and presynaptic membrane. Predicted to be active in glutamatergic synapse and plasma membrane. Predicted to be integral component of postsynaptic density membrane. [provided by Alliance of Genome Resources, Apr 2022]
9640	ZNF592	infomap 33	This gene is thought to play a role in a complex developmental pathway and the regulation of genes involved in cerebellar development. Mutations in this gene have been associated with autosomal recessive spinocerebellar ataxia. [provided by RefSeq, Jan 2011]

Continued on next page

Table B.3 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9651	PLCH2	wt 730	PLCH2 is a member of the PLC-eta family of the phosphoinositide-specific phospholipase C (PLC) superfamily of enzymes that cleave PtdIns(4,5) P2 to generate second messengers inositol 1,4,5-trisphosphate and diacylglycerol (Zhou et al., 2005 [PubMed 16107206]).[supplied by OMIM, Jun 2009]
9839	ZEB2	louvain 8	The protein encoded by this gene is a member of the Zfh1 family of 2-handed zinc finger/homeodomain proteins. It is located in the nucleus and functions as a DNA-binding transcriptional repressor that interacts with activated SMADs. Mutations in this gene are associated with Hirschsprung disease/Mowat-Wilson syndrome. Alternatively spliced transcript variants have been found for this gene.[provided by RefSeq, Jan 2010]

B.4 Full Network SCZ Prioritised Enriched Genes

Table B.4: SCZ Prioritised genes from enriched clusters found from the Full Postsynaptic Database, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
11141	IL1RAPL1	infomap 217, sgG5 1, spectral 3, wt 126	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]
257194	NEGR1	fc 9, sgG5 112, wt 257	Predicted to act upstream of or within several processes, including feeding behavior; locomotory behavior; and positive regulation of neuron projection development. Predicted to be located in extracellular region and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
488	ATP2A2	infomap 264, sgG1 16, wt 78	This gene encodes one of the SERCA Ca(2+)-ATPases, which are intracellular pumps located in the sarcoplasmic or endoplasmic reticula of the skeletal muscle. This enzyme catalyzes the hydrolysis of ATP coupled with the translocation of calcium from the cytosol into the sarcoplasmic reticulum lumen, and is involved in regulation of the contraction/relaxation cycle. Mutations in this gene cause Darier-White disease, also known as keratosis follicularis, an autosomal dominant skin disorder characterized by loss of adhesion between epidermal cells and abnormal keratinization. Other types of mutations in this gene have been associated with various forms of muscular dystrophies. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Dec 2019]
2823	GPM6A	infomap 188, wt 90	Predicted to enable calcium channel activity. Involved in neuron migration and stem cell differentiation. Located in extracellular exosome. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2911	GRM1	infomap 5, sgG2 9	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]
4137	MAPT	infomap 5, sgG2 9	This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. MAPT gene mutations have been associated with several neurodegenerative disorders such as Alzheimer's disease, Pick's disease, frontotemporal dementia, cortico-basal degeneration and progressive supranuclear palsy. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
4978	OPCML	fc 9, sgG5 112	This gene encodes a member of the IgLON subfamily in the immunoglobulin protein superfamily of proteins. The encoded preprotein is proteolytically processed to generate the mature protein. This protein is localized in the plasma membrane and may have an accessory role in opioid receptor function. This gene has an ortholog in rat and bovine. The opioid binding-cell adhesion molecule encoded by the rat gene binds opioid alkaloids in the presence of acidic lipids, exhibits selectivity for mu ligands and acts as a GPI-anchored protein. Since the encoded protein is highly conserved in species during evolution, it may have a fundamental role in mammalian systems. Alternative splicing results in multiple transcript variants, at least one of which encodes an isoform that is proteolytically processed. [provided by RefSeq, Jan 2016]
5595	MAPK3	infomap 5, sgG2 9	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
57476	GRAMD1B	infomap 106, wt 761	Predicted to enable cholesterol binding activity; cholesterol transfer activity; and phospholipid binding activity. Predicted to be involved in cellular response to cholesterol and cholesterol homeostasis. Located in endoplasmic reticulum membrane; endoplasmic reticulum-plasma membrane contact site; and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]
94030	LRRC4B	infomap 267, wt 797	Predicted to enable signaling receptor binding activity. Predicted to be involved in regulation of synapse assembly and synaptic membrane adhesion. Predicted to be located in cerebellar mossy fiber and presynaptic membrane. Predicted to be active in glutamatergic synapse and plasma membrane. Predicted to be integral component of postsynaptic density membrane. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
1182	CLCN3	infomap 135	This gene encodes a member of the voltage-gated chloride channel (ClC) family. The encoded protein is present in all cell types and localized in plasma membranes and in intracellular vesicles. It is a multi-pass membrane protein which contains a ClC domain and two additional C-terminal CBS (cystathionine beta-synthase) domains. The ClC domain catalyzes the selective flow of Cl-ions across cell membranes, and the CBS domain may have a regulatory function. This protein plays a role in both acidification and transmitter loading of GABAergic synaptic vesicles, and in smooth muscle cell activation and neointima formation. This protein is required for lysophosphatidic acid (LPA)-activated Cl- current activity and fibroblast-to-myofibroblast differentiation. The protein activity is regulated by Ca(2+)/calmodulin-dependent protein kinase II (CaMKII) in glioma cells. Multiple alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Aug 2011]
152330	CNTN4	wt 480	This gene encodes a member of the contactin family of immunoglobulins. Contactins are axon-associated cell adhesion molecules that function in neuronal network formation and plasticity. The encoded protein is a glycosylphosphatidylinositol-anchored neuronal membrane protein that may play a role in the formation of axon connections in the developing nervous system. Deletion or mutation of this gene may play a role in 3p deletion syndrome and autism spectrum disorders. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2011]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
1630	DCC	sgG2 9	This gene encodes a netrin 1 receptor. The transmembrane protein is a member of the immunoglobulin superfamily of cell adhesion molecules, and mediates axon guidance of neuronal growth cones towards sources of netrin 1 ligand. The cytoplasmic tail interacts with the tyrosine kinases Src and focal adhesion kinase (FAK, also known as PTK2) to mediate axon attraction. The protein partially localizes to lipid rafts, and induces apoptosis in the absence of ligand. The protein functions as a tumor suppressor, and is frequently mutated or downregulated in colorectal cancer and esophageal carcinoma. [provided by RefSeq, Oct 2009]
22864	R3HDM2	infomap 106	Enables RNA binding activity. Predicted to be located in nucleus. [provided by Alliance of Genome Resources, Apr 2022]
399909	PCNX3	wt 470	Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]
56924	PAK6	sgG2 9	This gene encodes a member of a family of p21-stimulated serine/threonine protein kinases, which contain an amino-terminal Cdc42/Rac interactive binding (CRIB) domain and a carboxyl-terminal kinase domain. These kinases function in a number of cellular processes, including cytoskeleton rearrangement, apoptosis, and the mitogen-activated protein (MAP) kinase signaling pathway. The protein encoded by this gene interacts with androgen receptor (AR) and translocates to the nucleus, where it is involved in transcriptional regulation. Changes in expression of this gene have been linked to prostate cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2015]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
57502	NLGN4X	wt 109	This gene encodes a member of the type-B carboxylesterase/lipase protein family. The encoded protein belongs to a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and may be involved in the formation and remodeling of central nervous system synapses. The encoded protein interacts with discs large homolog 4 (DLG4). Mutations in this gene have been associated with autism and Asperger syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Aug 2013]
5789	PTPRD	sgG5 1	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP contains an extracellular region, a single transmembrane segment and two tandem intracytoplasmic catalytic domains, and thus represents a receptor-type PTP. The extracellular region of this protein is composed of three Ig-like and eight fibronectin type III-like domains. Studies of the similar genes in chicken and fly suggest the role of this PTP is in promoting neurite growth, and regulating neurons axon guidance. Multiple alternatively spliced transcript variants of this gene have been reported. A related pseudogene has been identified on chromosome 5. [provided by RefSeq, Jan 2010]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
6444	SGCD	wt 129	The protein encoded by this gene is one of the four known components of the sarcoglycan complex, which is a subcomplex of the dystrophin-glycoprotein complex (DGC). DGC forms a link between the F-actin cytoskeleton and the extracellular matrix. This protein is expressed most abundantly in skeletal and cardiac muscle. Mutations in this gene have been associated with autosomal recessive limb-girdle muscular dystrophy and dilated cardiomyopathy. Alternatively spliced transcript variants encoding distinct isoforms have been observed for this gene. [provided by RefSeq, Jul 2008]
64478	CSMD1	wt 735	Predicted to act upstream of or within several processes, including learning or memory; mammary gland branching involved in pregnancy; and reproductive structure development. Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]
7223	TRPC4	wt 74	This gene encodes a member of the canonical subfamily of transient receptor potential cation channels. The encoded protein forms a non-selective calcium-permeable cation channel that is activated by Gq-coupled receptors and tyrosine kinases, and plays a role in multiple processes including endothelial permeability, vasodilation, neurotransmitter release and cell proliferation. Single nucleotide polymorphisms in this gene may be associated with generalized epilepsy with photosensitivity. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
8204	NRIP1	infomap 5	Nuclear receptor interacting protein 1 (NRIP1) is a nuclear protein that specifically interacts with the hormone-dependent activation domain AF2 of nuclear receptors. Also known as RIP140, this protein modulates transcriptional activity of the estrogen receptor. [provided by RefSeq, Jul 2008]
9093	DNAJA3	infomap 188	This gene encodes a member of the DNAJ/Hsp40 protein family. DNAJ/Hsp40 proteins stimulate the ATPase activity of Hsp70 chaperones and play critical roles in protein folding, degradation, and multimeric complex assembly. The encoded protein is localized to mitochondria and mediates several cellular processes including proliferation, survival and apoptotic signal transduction. The encoded protein also plays a critical role in tumor suppression through interactions with oncogenic proteins including ErbB2 and the p53 tumor suppressor protein. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]
9651	PLCH2	wt 730	PLCH2 is a member of the PLC-eta family of the phosphoinositide-specific phospholipase C (PLC) superfamily of enzymes that cleave PtdIns(4,5) P2 to generate second messengers inositol 1,4,5-trisphosphate and diacylglycerol (Zhou et al., 2005 [PubMed 16107206]).[supplied by OMIM, Jun 2009]

Continued on next page

Table B.4 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9892	SNAP91	infomap 135	Predicted to enable several functions, including SNARE binding activity; clathrin binding activity; and phosphatidylinositol binding activity. Acts upstream of or within regulation of clathrin-dependent endocytosis. Predicted to be located in several cellular components, including postsynaptic density; presynaptic endosome; and presynaptic membrane. Predicted to be extrinsic component of endosome membrane. Predicted to be active in several cellular components, including Schaffer collateral - CA1 synapse; cytoplasmic vesicle; and parallel fiber to Purkinje cell synapse. Predicted to be extrinsic component of presynaptic endocytic zone membrane. Biomarker of Alzheimer's disease. [provided by Alliance of Genome Resources, Apr 2022]

B.5 Consensus Network SCZ Broad Enriched Genes

Table B.5: SCZ Broad genes from enriched clusters found from the Consensus Postsynaptic PSD network, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
2903	GRIN2A	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 13	This gene encodes a member of the glutamate-gated ion channel protein family. The encoded protein is an N-methyl-D-aspartate (NMDA) receptor subunit. NMDA receptors are both ligand-gated and voltage-dependent, and are involved in long-term potentiation, an activity-dependent increase in the efficiency of synaptic transmission thought to underlie certain kinds of memory and learning. These receptors are permeable to calcium ions, and activation results in a calcium influx into postsynaptic cells, which results in the activation of several signaling cascades. Disruption of this gene is associated with focal epilepsy and speech disorder with or without cognitive disability. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2014]
8120	AP3B2	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 452	Adaptor protein complex 3 (AP-3 complex) is a heterotrimeric protein complex involved in the formation of clathrin-coated synaptic vesicles. The protein encoded by this gene represents the beta subunit of the neuron-specific AP-3 complex and was first identified as the target antigen in human paraneoplastic neurologic disorders. The encoded subunit binds clathrin and is phosphorylated by a casein kinase-like protein, which mediates synaptic vesicle coat assembly. Defects in this gene are a cause of early-onset epileptic encephalopathy. [provided by RefSeq, Feb 2017]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
8525	DGKZ	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 13	The protein encoded by this gene belongs to the eukaryotic diacylglycerol kinase family. It may attenuate protein kinase C activity by regulating diacylglycerol levels in intracellular signaling cascade and signal transduction. Alternative splicing occurs at this locus and multiple transcript variants encoding distinct isoforms have been identified. [provided by RefSeq, Nov 2010]
85358	SHANK3	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 13	This gene is a member of the Shank gene family. Shank proteins are multidomain scaffold proteins of the postsynaptic density that connect neurotransmitter receptors, ion channels, and other membrane proteins to the actin cytoskeleton and G-protein-coupled signaling pathways. Shank proteins also play a role in synapse formation and dendritic spine maturation. Mutations in this gene are a cause of autism spectrum disorder (ASD), which is characterized by impairments in social interaction and communication, and restricted behavioral patterns and interests. Mutations in this gene also cause schizophrenia type 15, and are a major causative factor in the neurological symptoms of 22q13.3 deletion syndrome, which is also known as Phelan-McDermid syndrome. Additional isoforms have been described for this gene but they have not yet been experimentally verified. [provided by RefSeq, Mar 2012]
9228	DLGAP2	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 13	The product of this gene is a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain, however, only the paternal allele is expressed in the testis (PMID:18055845). Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jun 2014]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9758	FRMPD4	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, spectral 11, wt 13	This gene encodes a multi-domain (WW, PDZ, FERM) containing protein. Through its interaction with other proteins (such as PSD-95), it functions as a positive regulator of dendritic spine morphogenesis and density, and is required for the maintenance of excitatory synaptic transmission. [provided by RefSeq, Jan 2010]
54497	HEATR5B	infomap 17, louvain 9, sgG1 19, sgG2 12, sgG5 74, wt 13	Predicted to be involved in endocytosis; protein localization; and retrograde transport, endosome to Golgi. Located in membrane. [provided by Alliance of Genome Resources, Apr 2022]
60412	EXOC4	infomap 17, louvain 9, sgG2 12, sgG5 74, spectral 11	The protein encoded by this gene is a component of the exocyst complex, a multiple protein complex essential for targeting exocytic vesicles to specific docking sites on the plasma membrane. Though best characterized in yeast, the component proteins and functions of exocyst complex have been demonstrated to be highly conserved in higher eukaryotes. At least eight components of the exocyst complex, including this protein, are found to interact with the actin cytoskeletal remodeling and vesicle transport machinery. The complex is also essential for the biogenesis of epithelial cell surface polarity. Alternate transcriptional splice variants, encoding different isoforms, have been characterized. [provided by RefSeq, Jul 2008]
23113	CUL9	infomap 11, sgG1 18, sgG2 41, sgG5 70	Predicted to enable several functions, including ATP binding activity; metal ion binding activity; and ubiquitin protein ligase binding activity. Involved in microtubule cytoskeleton organization; protein ubiquitination; and regulation of mitotic nuclear division. Located in cytosol. Part of cullin-RING ubiquitin ligase complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
23513	SCRIB	infomap 17, louvain 9, sgG1 19, sgG5 74	This gene encodes a protein that was identified as being similar to the Drosophila scribble protein. The mammalian protein is involved in tumor suppression pathways. As a scaffold protein involved in cell polarization processes, this protein binds to many other proteins. The encoded protein binds to papillomavirus E6 protein via its PDZ domain and the C-terminus of E6. Two alternatively spliced transcript variants that encode different protein isoforms have been found for this gene. [provided by RefSeq, Nov 2011]
348980	HCN1	louvain 9, sgG1 19, sgG2 12, wt 234	The membrane protein encoded by this gene is a hyperpolarization-activated cation channel that contributes to the native pacemaker currents in heart and neurons. The encoded protein can homodimerize or heterodimerize with other pore-forming subunits to form a potassium channel. This channel may act as a receptor for sour tastes. [provided by RefSeq, Oct 2011]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
56138	PCDHA11	louvain 9, sgG1 19, sgG2 12, wt 440	This gene is a member of the protocadherin alpha gene cluster, one of three related gene clusters tandemly linked on chromosome five that demonstrate an unusual genomic organization similar to that of B-cell and T-cell receptor gene clusters. The alpha gene cluster is composed of 15 cadherin superfamily genes related to the mouse CNR genes and consists of 13 highly similar and 2 more distantly related coding sequences. The tandem array of 15 N-terminal exons, or variable exons, are followed by downstream C-terminal exons, or constant exons, which are shared by all genes in the cluster. The large, uninterrupted N-terminal exons each encode six cadherin ectodomains while the C-terminal exons encode the cytoplasmic domain. These neural cadherin-like cell adhesion proteins are integral plasma membrane proteins that most likely play a critical role in the establishment and function of specific cell-cell connections in the brain. Alternative splicing has been observed and additional variants have been suggested but their full-length nature has yet to be determined. [provided by RefSeq, Jul 2008]
7223	TRPC4	infomap 11, sgG1 18, sgG2 41, sgG5 70	This gene encodes a member of the canonical subfamily of transient receptor potential cation channels. The encoded protein forms a non-selective calcium-permeable cation channel that is activated by Gq-coupled receptors and tyrosine kinases, and plays a role in multiple processes including endothelial permeability, vasodilation, neurotransmitter release and cell proliferation. Single nucleotide polymorphisms in this gene may be associated with generalized epilepsy with photosensitivity. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9162	DGKI	infomap 11, sgG1 18, sgG2 41, sgG5 70	This gene is a member of the type IV diacylglycerol kinase subfamily. Diacylglycerol kinases regulate the intracellular concentration of diacylglycerol through its phosphorylation, producing phosphatidic acid. The specific role of the enzyme encoded by this gene is undetermined, however, it may play a crucial role in the production of phosphatidic acid in the retina or in recessive forms of retinal degeneration. [provided by RefSeq, Jul 2008]
124044	SPATA2L	louvain 9, sgG1 19, sgG2 12	Predicted to be active in cytoplasm. [provided by Alliance of Genome Resources, Apr 2022]
1983	EIF5	infomap 93, sgG1 19, wt 112	Eukaryotic translation initiation factor-5 (EIF5) interacts with the 40S initiation complex to promote hydrolysis of bound GTP with concomitant joining of the 60S ribosomal subunit to the 40S initiation complex. The resulting functional 80S ribosomal initiation complex is then active in peptidyl transfer and chain elongations (summary by Si et al., 1996 [PubMed 8663286]).[supplied by OMIM, May 2010]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2911	GRM1	louvain 9, sgG1 19, sgG2 12	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]
4900	NRGN	infomap 93, sgG1 19, wt 112	Neurogranin (NRGN) is the human homolog of the neuron-specific rat RC3/neurogranin gene. This gene encodes a postsynaptic protein kinase substrate that binds calmodulin in the absence of calcium. The NRGN gene contains four exons and three introns. The exons 1 and 2 encode the protein and exons 3 and 4 contain untranslated sequences. It is suggested that the NRGN is a direct target for thyroid hormone in human brain, and that control of expression of this gene could underlay many of the consequences of hypothyroidism on mental states during development as well as in adult subjects. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9568	GABBR2	louvain 9, sgG1 19, sgG2 12	The multi-pass membrane protein encoded by this gene belongs to the G-protein coupled receptor 3 family and GABA-B receptor subfamily. The GABA-B receptors inhibit neuronal activity through G protein-coupled second-messenger systems, which regulate the release of neurotransmitters, and the activity of ion channels and adenylyl cyclase. This receptor subunit forms an active heterodimeric complex with GABA-B receptor subunit 1, neither of which is effective on its own. Allelic variants of this gene have been associated with nicotine dependence.[provided by RefSeq, Jan 2010]
2823	GPM6A	infomap 178, wt 221	Predicted to enable calcium channel activity. Involved in neuron migration and stem cell differentiation. Located in extracellular exosome. [provided by Alliance of Genome Resources, Apr 2022]
29763	PACSIN3	sgG1 19, sgG2 12	This gene is a member of the protein kinase C and casein kinase substrate in neurons family. The encoded protein is involved in linking the actin cytoskeleton with vesicle formation. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2010]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
56924	PAK6	sgG1 19, spectral 7	This gene encodes a member of a family of p21-stimulated serine/threonine protein kinases, which contain an amino-terminal Cdc42/Rac interactive binding (CRIB) domain and a carboxyl-terminal kinase domain. These kinases function in a number of cellular processes, including cytoskeleton rearrangement, apoptosis, and the mitogen-activated protein (MAP) kinase signaling pathway. The protein encoded by this gene interacts with androgen receptor (AR) and translocates to the nucleus, where it is involved in transcriptional regulation. Changes in expression of this gene have been linked to prostate cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2015]
57502	NLGN4X	infomap 17, louvain 9	This gene encodes a member of the type-B carboxylesterase/lipase protein family. The encoded protein belongs to a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and may be involved in the formation and remodeling of central nervous system synapses. The encoded protein interacts with discs large homolog 4 (DLG4). Mutations in this gene have been associated with autism and Asperger syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Aug 2013]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
6622	SNCA	infomap 93, sgG1 19	Alpha-synuclein is a member of the synuclein family, which also includes beta- and gamma-synuclein. Synucleins are abundantly expressed in the brain and alpha- and beta-synuclein inhibit phospholipase D2 selectively. SNCA may serve to integrate presynaptic signaling and membrane trafficking. Defects in SNCA have been implicated in the pathogenesis of Parkinson disease. SNCA peptides are a major component of amyloid plaques in the brains of patients with Alzheimer's disease. Alternatively spliced transcripts encoding different isoforms have been identified for this gene. [provided by RefSeq, Feb 2016]
775	CACNA1C	sgG1 19, sgG2 41	This gene encodes an alpha-1 subunit of a voltage-dependent calcium channel. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization. The alpha-1 subunit consists of 24 transmembrane segments and forms the pore through which ions pass into the cell. The calcium channel consists of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. There are multiple isoforms of each of these proteins, either encoded by different genes or the result of alternative splicing of transcripts. The protein encoded by this gene binds to and is inhibited by dihydropyridine. Alternative splicing results in many transcript variants encoding different proteins. Some of the predicted proteins may not produce functional ion channel subunits. [provided by RefSeq, Oct 2012]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
783	CACNB2	sgG2 41, sgG5 70	This gene encodes a subunit of a voltage-dependent calcium channel protein that is a member of the voltage-gated calcium channel superfamily. The gene product was originally identified as an antigen target in Lambert-Eaton myasthenic syndrome, an autoimmune disorder. Mutations in this gene are associated with Brugada syndrome. Alternatively spliced variants encoding different isoforms have been described. [provided by RefSeq, Feb 2013]
8943	AP3D1	sgG1 19, wt 479	The protein encoded by this gene is a subunit of the AP3 adaptor-like complex, which is not clathrin-associated, but is associated with the golgi region, as well as more peripheral structures. The AP-3 complex facilitates the budding of vesicles from the golgi membrane, and may be directly involved in trafficking to lysosomes. This subunit is implicated in intracellular biogenesis and trafficking of pigment granules, and possibly platelet dense granules and neurotransmitter vesicles. Defects in this gene are a cause of a new type of Hermansky-Pudlak syndrome. [provided by RefSeq, Feb 2017]
9093	DNAJA3	louvain 9, spectral 7	This gene encodes a member of the DNAJ/Hsp40 protein family. DNAJ/Hsp40 proteins stimulate the ATPase activity of Hsp70 chaperones and play critical roles in protein folding, degradation, and multimeric complex assembly. The encoded protein is localized to mitochondria and mediates several cellular processes including proliferation, survival and apoptotic signal transduction. The encoded protein also plays a critical role in tumor suppression through interactions with oncogenic proteins including ErbB2 and the p53 tumor suppressor protein. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9892	SNAP91	louvain 9, spectral 11	Predicted to enable several functions, including SNARE binding activity; clathrin binding activity; and phosphatidylinositol binding activity. Acts upstream of or within regulation of clathrin-dependent endocytosis. Predicted to be located in several cellular components, including postsynaptic density; presynaptic endosome; and presynaptic membrane. Predicted to be extrinsic component of endosome membrane. Predicted to be active in several cellular components, including Schaffer collateral - CA1 synapse; cytoplasmic vesicle; and parallel fiber to Purkinje cell synapse. Predicted to be extrinsic component of presynaptic endocytic zone membrane. Biomarker of Alzheimer's disease. [provided by Alliance of Genome Resources, Apr 2022]
9912	ARHGAP44	louvain 9, spectral 7	Enables phospholipid binding activity. Predicted to be involved in several processes, including modification of dendritic spine; negative regulation of Rac protein signal transduction; and regulation of plasma membrane bounded cell projection organization. Located in leading edge membrane. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
10645	CAMKK2	sgG1 19	The product of this gene belongs to the Serine/Threonine protein kinase family, and to the Ca(2+)/calmodulin-dependent protein kinase subfamily. The major isoform of this gene plays a role in the calcium/calmodulin-dependent (CaM) kinase cascade by phosphorylating the downstream kinases CaMK1 and CaMK4. Protein products of this gene also phosphorylate AMP-activated protein kinase (AMPK). This gene has its strongest expression in the brain and influences signalling cascades involved with learning and memory, neuronal differentiation and migration, neurite outgrowth, and synapse formation. Alternative splicing results in multiple transcript variants encoding distinct isoforms. The identified isoforms differ in their ability to undergo autophosphorylation and to phosphorylate downstream kinases. [provided by RefSeq, Jul 2012]
11141	IL1RAPL1	spectral 7	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
1152	CKB	infomap 123	The protein encoded by this gene is a cytoplasmic enzyme involved in energy homeostasis. The encoded protein reversibly catalyzes the transfer of phosphate between ATP and various phosphogens such as creatine phosphate. It acts as a homodimer in brain as well as in other tissues, and as a heterodimer with a similar muscle isozyme in heart. The encoded protein is a member of the ATP:guanido phosphotransferase protein family. A pseudogene of this gene has been characterized. [provided by RefSeq, Jul 2008]
1191	CLU	infomap 33	The protein encoded by this gene is a secreted chaperone that can under some stress conditions also be found in the cell cytosol. It has been suggested to be involved in several basic biological events such as cell death, tumor progression, and neurodegenerative disorders. Alternate splicing results in both coding and non-coding variants.[provided by RefSeq, May 2011]
129285	PPP1R21	spectral 64	Located in early endosome. [provided by Alliance of Genome Resources, Apr 2022]
192669	AGO3	sgG1 19	This gene encodes a member of the Argonaute family of proteins which play a role in RNA interference. The encoded protein is highly basic, contains a PAZ domain and a PIWI domain, and may play a role in short-interfering-RNA-mediated gene silencing. This gene is located on chromosome 1 in a tandem cluster of closely related family members including argonaute 4 and eukaryotic translation initiation factor 2C, 1. Two transcript variants encoding distinct isoforms have been identified for this gene. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2053	EPHX2	spectral 64	This gene encodes a member of the epoxide hydrolase family. The protein, found in both the cytosol and peroxisomes, binds to specific epoxides and converts them to the corresponding dihydrodiols. Mutations in this gene have been associated with familial hypercholesterolemia. Alternatively spliced transcript variants have been described. [provided by RefSeq, Feb 2012]
2107	ETF1	sgG1 19	This gene encodes a class-1 polypeptide chain release factor. The encoded protein plays an essential role in directing termination of mRNA translation from the termination codons UAA, UAG and UGA. This protein is a component of the SURF complex which promotes degradation of prematurely terminated mRNAs via the mechanism of nonsense-mediated mRNA decay (NMD). Alternate splicing results in multiple transcript variants. Pseudogenes of this gene are found on chromosomes 6, 7, and X. [provided by RefSeq, Aug 2013]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2185	PTK2B	sgG1 19	This gene encodes a cytoplasmic protein tyrosine kinase which is involved in calcium-induced regulation of ion channels and activation of the map kinase signaling pathway. The encoded protein may represent an important signaling intermediate between neuropeptide-activated receptors or neurotransmitters that increase calcium flux and the downstream signals that regulate neuronal activity. The encoded protein undergoes rapid tyrosine phosphorylation and activation in response to increases in the intracellular calcium concentration, nicotinic acetylcholine receptor activation, membrane depolarization, or protein kinase C activation. This protein has been shown to bind CRK-associated substrate, nephrocystin, GTPase regulator associated with FAK, and the SH2 domain of GRB2. The encoded protein is a member of the FAK subfamily of protein tyrosine kinases but lacks significant sequence similarity to kinases from other subfamilies. Four transcript variants encoding two different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
22866	CNKSR2	infomap 33	This gene encodes a multidomain protein that functions as a scaffold protein to mediate the mitogen-activated protein kinase pathways downstream from Ras. This gene product is induced by vitamin D and inhibits apoptosis in certain cancer cells. It may also play a role in ternary complex assembly of synaptic proteins at the postsynaptic membrane and coupling of signal transduction to membrane/cytoskeletal remodeling. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Dec 2009]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
23228	PLCL2	infomap 33	Enables GABA receptor binding activity. Predicted to be involved in negative regulation of cold-induced thermogenesis and phosphatidylinositol-mediated signaling. Predicted to act upstream of or within several processes, including B cell activation; gamma-aminobutyric acid signaling pathway; and negative regulation of B cell receptor signaling pathway. Predicted to be located in cytoplasm. [provided by Alliance of Genome Resources, Apr 2022]
257194	NEGR1	wt 217	Predicted to act upstream of or within several processes, including feeding behavior; locomotory behavior; and positive regulation of neuron projection development. Predicted to be located in extracellular region and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]
25843	MOB4	infomap 123	This gene was identified based on its similarity with the mouse counterpart. Studies of the mouse counterpart suggest that the expression of this gene may be regulated during oocyte maturation and preimplantation following zygotic gene activation. Alternatively spliced transcript variants encoding distinct isoforms have been observed. Naturally occurring read-through transcription occurs between this locus and the neighboring locus HSPE1.[provided by RefSeq, Feb 2011]
27132	CPNE7	wt 263	This gene encodes a member of the copine family, which is composed of calcium-dependent membrane-binding proteins. The gene product contains two N-terminal C2 domains and one von Willebrand factor A domain. The encoded protein may be involved in membrane trafficking. Two alternatively spliced transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Nov 2008]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2923	PDIA3	spectral 64	This gene encodes a protein of the endoplasmic reticulum that interacts with lectin chaperones calreticulin and calnexin to modulate folding of newly synthesized glycoproteins. The protein was once thought to be a phospholipase; however, it has been demonstrated that the protein actually has protein disulfide isomerase activity. It is thought that complexes of lectins and this protein mediate protein folding by promoting formation of disulfide bonds in their glycoprotein substrates. This protein also functions as a molecular chaperone that prevents the formation of protein aggregates. [provided by RefSeq, Dec 2016]
3313	HSPA9	infomap 178	This gene encodes a member of the heat shock protein 70 gene family. The encoded protein is primarily localized to the mitochondria but is also found in the endoplasmic reticulum, plasma membrane and cytoplasmic vesicles. This protein is a heat-shock cognate protein. This protein plays a role in cell proliferation, stress response and maintenance of the mitochondria. A pseudogene of this gene is found on chromosome 2.[provided by RefSeq, May 2010]
3336	HSPE1	spectral 64	This gene encodes a major heat shock protein which functions as a chaperonin. Its structure consists of a heptameric ring which binds to another heat shock protein in order to form a symmetric, functional heterodimer which enhances protein folding in an ATP-dependent manner. This gene and its co-chaperonin, HSPD1, are arranged in a head-to-head orientation on chromosome 2. Naturally occurring read-through transcription occurs between this locus and the neighboring locus MOBKL3.[provided by RefSeq, Feb 2011]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
3745	KCNB1	sgG1 19	Voltage-gated potassium (Kv) channels represent the most complex class of voltage-gated ion channels from both functional and structural standpoints. Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume. Four sequence-related potassium channel genes - shaker, shaw, shab, and shal - have been identified in Drosophila, and each has been shown to have human homolog(s). This gene encodes a member of the potassium channel, voltage-gated, shab-related subfamily. This member is a delayed rectifier potassium channel and its activity is modulated by some other family members. [provided by RefSeq, Jul 2008]
381	ARF5	spectral 64	This gene is a member of the human ADP-ribosylation factor (ARF) gene family. These genes encode small guanine nucleotide-binding proteins that stimulate the ADP-ribosyltransferase activity of cholera toxin and play a role in vesicular trafficking and as activators of phospholipase D. The gene products include 6 ARF proteins and 11 ARF-like proteins and constitute 1 family of the RAS superfamily. The ARF proteins are categorized as class I (ARF1, ARF2,and ARF3), class II (ARF4 and ARF5) and class III (ARF6). The members of each class share a common gene organization. [provided by RefSeq, Dec 2010]
392	ARHGAP1	spectral 7	This gene encodes a member of a large family of proteins that activate Rho-type guanosine triphosphate (GTP) metabolizing enzymes. The encoded protein contains a SRC homology 3 domain and interacts with Bcl-2-associated protein family members. [provided by RefSeq, Aug 2012]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
4137	MAPT	sgG1 19	This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. MAPT gene mutations have been associated with several neurodegenerative disorders such as Alzheimer's disease, Pick's disease, frontotemporal dementia, cortico-basal degeneration and progressive supranuclear palsy. [provided by RefSeq, Jul 2008]
5142	PDE4B	wt 171	This gene is a member of the type IV, cyclic AMP (cAMP)-specific, cyclic nucleotide phosphodiesterase (PDE) family. The encoded protein regulates the cellular concentrations of cyclic nucleotides and thereby play a role in signal transduction. Altered activity of this protein has been associated with schizophrenia and bipolar affective disorder. Alternative splicing and the use of alternative promoters results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Jul 2014]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
51441	YTHDF2	wt 575	This gene encodes a member of the YTH (YT521-B homology) superfamily containing YTH domain. The YTH domain is typical for the eukaryotes and is particularly abundant in plants. The YTH domain is usually located in the middle of the protein sequence and may function in binding to RNA. In addition to a YTH domain, this protein has a proline rich region which may be involved in signal transduction. An Alu-rich domain has been identified in one of the introns of this gene, which is thought to be associated with human longevity. In addition, reciprocal translocations between this gene and the Runx1 (AML1) gene on chromosome 21 has been observed in patients with acute myeloid leukemia. This gene was initially mapped to chromosome 14, which was later turned out to be a pseudogene. Alternatively spliced transcript variants encoding different isoforms have been identified in this gene. [provided by RefSeq, Oct 2012]
5595	MAPK3	sgG1 19	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
57670	KIAA1549	spectral 7	The protein encoded by this gene belongs to the UPF0606 family. This gene has been found to be fused to the BRAF oncogene in many cases of pilocytic astrocytoma. The fusion results from 2Mb tandem duplications at 7q34. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Oct 2012]
57795	BRINP2	wt 251	Predicted to be involved in cellular response to retinoic acid; negative regulation of mitotic cell cycle; and positive regulation of neuron differentiation. Predicted to be located in extracellular region. Predicted to be active in dendrite; endoplasmic reticulum; and neuronal cell body. Implicated in oral squamous cell carcinoma. [provided by Alliance of Genome Resources, Apr 2022]
5789	PTPRD	spectral 7	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP contains an extracellular region, a single transmembrane segment and two tandem intracytoplasmic catalytic domains, and thus represents a receptor-type PTP. The extracellular region of this protein is composed of three Ig-like and eight fibronectin type III-like domains. Studies of the similar genes in chicken and fly suggest the role of this PTP is in promoting neurite growth, and regulating neurons axon guidance. Multiple alternatively spliced transcript variants of this gene have been reported. A related pseudogene has been identified on chromosome 5. [provided by RefSeq, Jan 2010]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
5792	PTPRF	spectral 7	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP possesses an extracellular region, a single transmembrane region, and two tandem intracytoplasmic catalytic domains, and thus represents a receptor-type PTP. The extracellular region contains three Ig-like domains, and nine non-Ig like domains similar to that of neural-cell adhesion molecule. This PTP was shown to function in the regulation of epithelial cell-cell contacts at adherents junctions, as well as in the control of beta-catenin signaling. An increased expression level of this protein was found in the insulin-responsive tissue of obese, insulin-resistant individuals, and may contribute to the pathogenesis of insulin resistance. Two alternatively spliced transcript variants of this gene, which encode distinct proteins, have been reported. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
6137	RPL13	louvain 9	Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L13E family of ribosomal proteins. It is located in the cytoplasm. This gene is expressed at significantly higher levels in benign breast lesions than in breast carcinomas. Alternatively spliced transcript variants encoding distinct isoforms have been found for this gene. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. [provided by RefSeq, Jul 2011]
63826	SRR	wt 270	Enables several functions, including L-serine ammonia-lyase activity; PDZ domain binding activity; and anion binding activity. Involved in pyruvate biosynthetic process; response to lipopolysaccharide; and serine family amino acid metabolic process. Located in cytoplasm and neuronal cell body. [provided by Alliance of Genome Resources, Apr 2022]
64478	CSMD1	wt 532	Predicted to act upstream of or within several processes, including learning or memory; mammary gland branching involved in pregnancy; and reproductive structure development. Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]
6801	STRN	infomap 123	Enables armadillo repeat domain binding activity; estrogen receptor binding activity; and protein phosphatase 2A binding activity. Involved in Wnt signaling pathway and negative regulation of cell population proliferation. Located in bicellular tight junction. Part of FAR/SIN/STRIPAK complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
6847	SYCP1	wt 544	Enables double-stranded DNA binding activity. Involved in protein homotetramerization. Predicted to be located in synaptonemal complex. Predicted to be active in central element; male germ cell nucleus; and transverse filament. [provided by Alliance of Genome Resources, Apr 2022]
7531	YWHAE	sgG1 19	This gene product belongs to the 14-3-3 family of proteins which mediate signal transduction by binding to phosphoserine-containing proteins. This highly conserved protein family is found in both plants and mammals, and this protein is 100% identical to the mouse ortholog. It interacts with CDC25 phosphatases, RAF1 and IRS1 proteins, suggesting its role in diverse biochemical activities related to signal transduction, such as cell division and regulation of insulin sensitivity. It has also been implicated in the pathogenesis of small cell lung cancer. Two transcript variants, one protein-coding and the other non-protein-coding, have been found for this gene. [provided by RefSeq, Aug 2008]
7812	CSDE1	wt 428	Enables RNA stem-loop binding activity. Involved in IRES-dependent viral translational initiation; nuclear-transcribed mRNA catabolic process, no-go decay; and stress granule assembly. Located in Golgi apparatus; cytosol; and plasma membrane. Part of CRD-mediated mRNA stability complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9344	TAOK2	wt 494	Enables mitogen-activated protein kinase kinase binding activity; neuropilin binding activity; and protein serine/threonine kinase activity. Involved in several processes, including focal adhesion assembly; intracellular signal transduction; and positive regulation of JNK cascade. Located in cytoplasmic vesicle; cytosol; and nuclear lumen. Part of receptor complex. [provided by Alliance of Genome Resources, Apr 2022]
9529	BAG5	sgG1 19	The protein encoded by this gene is a member of the BAG1-related protein family. BAG1 is an anti-apoptotic protein that functions through interactions with a variety of cell apoptosis and growth related proteins including BCL-2, Raf-protein kinase, steroid hormone receptors, growth factor receptors and members of the heat shock protein 70 kDa family. This protein contains a BAG domain near the C-terminus, which could bind and inhibit the chaperone activity of Hsc70/Hsp70. Three transcript variants encoding two different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
9556	ATP5MJ	wt 504	Predicted to be located in mitochondrion. Predicted to be integral component of membrane. Predicted to be part of mitochondrial proton-transporting ATP synthase complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.5 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9628	RGS6	sgG1 19	This gene encodes a member of the RGS (regulator of G protein signaling) family of proteins, which are defined by the presence of a RGS domain that confers the GTPase-activating activity of these proteins toward certain G alpha subunits. This protein also belongs to a subfamily of RGS proteins characterized by the presence of DEP and GGL domains, the latter a G beta 5-interacting domain. The RGS proteins negatively regulate G protein signaling, and may modulate neuronal, cardiovascular, lymphocytic activities, and cancer risk. Many alternatively spliced transcript variants encoding different isoforms with long or short N-terminal domains, complete or incomplete GGL domains, and distinct C-terminal domains, have been described for this gene, however, the full-length nature of some of these variants is not known.[provided by RefSeq, Mar 2011]
9651	PLCH2	wt 530	PLCH2 is a member of the PLC-eta family of the phosphoinositide-specific phospholipase C (PLC) superfamily of enzymes that cleave PtdIns(4,5) P2 to generate second messengers inositol 1,4,5-trisphosphate and diacylglycerol (Zhou et al., 2005 [PubMed 16107206]).[supplied by OMIM, Jun 2009]
9842	PLEKHM1	spectral 64	The protein encoded by this gene is essential for bone resorption, and may play a critical role in vesicular transport in the osteoclast. Mutations in this gene are associated with autosomal recessive osteopetrosis type 6 (OPTB6). Alternatively spliced transcript variants have been found for this gene. [provided by RefSeq, Sep 2009]

B.6 Consensus Network SCZ Priortised Enriched Genes

Table B.6: SCZ Priortised genes from enriched clusters found from the Consensus Postsynaptic PSD network, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
23113	CUL9	infomap 11, sgG1 18, sgG2 41, sgG5 70, wt 147	Predicted to enable several functions, including ATP binding activity; metal ion binding activity; and ubiquitin protein ligase binding activity. Involved in microtubule cytoskeleton organization; protein ubiquitination; and regulation of mitotic nuclear division. Located in cytosol. Part of cullin-RING ubiquitin ligase complex. [provided by Alliance of Genome Resources, Apr 2022]
7223	TRPC4	infomap 11, sgG1 18, sgG2 41, sgG5 70, wt 109	This gene encodes a member of the canonical subfamily of transient receptor potential cation channels. The encoded protein forms a non-selective calcium-permeable cation channel that is activated by Gq-coupled receptors and tyrosine kinases, and plays a role in multiple processes including endothelial permeability, vasodilation, neurotransmitter release and cell proliferation. Single nucleotide polymorphisms in this gene may be associated with generalized epilepsy with photosensitivity. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]
57502	NLGN4X	$fc_2, infomap 11, sing gene 9, wt 11$	This gene encodes a member of the type-B carboxylesterase/lipase protein family. The encoded protein belongs to a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and may be involved in the formation and remodeling of central nervous system synapses. The encoded protein interacts with discs large homolog 4 (DLG4). Mutations in this gene have been associated with autism and Asperger syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Aug 2013]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
11141	IL1RAPL1	infomap 99, spectral 7, wt 157	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]
2903	GRIN2A	$fc_2, infomap17$, spectral 9	This gene encodes a member of the glutamate-gated ion channel protein family. The encoded protein is an N-methyl-D-aspartate (NMDA) receptor subunit. NMDA receptors are both ligand-gated and voltage-dependent, and are involved in long-term potentiation, an activity-dependent increase in the efficiency of synaptic transmission thought to underlie certain kinds of memory and learning. These receptors are permeable to calcium ions, and activation results in a calcium influx into postsynaptic cells, which results in the activation of several signaling cascades. Disruption of this gene is associated with focal epilepsy and speech disorder with or without cognitive disability. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2014]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2911	GRM1	$fc_2, infomap157$, cluster 9	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]
488	ATP2A2	$fc_2, infomap185$, cluster 54	This gene encodes one of the SERCA Ca(2+)-ATPases, which are intracellular pumps located in the sarcoplasmic or endoplasmic reticula of the skeletal muscle. This enzyme catalyzes the hydrolysis of ATP coupled with the translocation of calcium from the cytosol into the sarcoplasmic reticulum lumen, and is involved in regulation of the contraction/relaxation cycle. Mutations in this gene cause Darier-White disease, also known as keratosis follicularis, an autosomal dominant skin disorder characterized by loss of adhesion between epidermal cells and abnormal keratinization. Other types of mutations in this gene have been associated with various forms of muscular dystrophies. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Dec 2019]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9093	DNAJA3	$fc_2, louvain9, spis\ra$	This gene encodes a member of the DNAJ/Hsp40 protein family. DNAJ/Hsp40 proteins stimulate the ATPase activity of Hsp70 chaperones and play critical roles in protein folding, degradation, and multimeric complex assembly. The encoded protein is localized to mitochondria and mediates several cellular processes including proliferation, survival and apoptotic signal transduction. The encoded protein also plays a critical role in tumor suppression through interactions with oncogenic proteins including ErbB2 and the p53 tumor suppressor protein. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. [provided by RefSeq, Aug 2011]
9228	DLGAP2	$fc_2, infomap17, louvain9$	The product of this gene is a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain, however, only the paternal allele is expressed in the testis (PMID:18055845). Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jun 2014]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
5789	PTPRD	infomap 99, spectral 7	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP contains an extracellular region, a single transmembrane segment and two tandem intracytoplasmic catalytic domains, and thus represents a receptor-type PTP. The extracellular region of this protein is composed of three Ig-like and eight fibronectin type III-like domains. Studies of the similar genes in chicken and fly suggest the role of this PTP is in promoting neurite growth, and regulating neurons axon guidance. Multiple alternatively spliced transcript variants of this gene have been reported. A related pseudogene has been identified on chromosome 5. [provided by RefSeq, Jan 2010]
9568	GABBR2	$fc_2, louvain9$	The multi-pass membrane protein encoded by this gene belongs to the G-protein coupled receptor 3 family and GABA-B receptor subfamily. The GABA-B receptors inhibit neuronal activity through G protein-coupled second-messenger systems, which regulate the release of neurotransmitters, and the activity of ion channels and adenylyl cyclase. This receptor subunit forms an active heterodimeric complex with GABA-B receptor subunit 1, neither of which is effective on its own. Allelic variants of this gene have been associated with nicotine dependence.[provided by RefSeq, Jan 2010]
9651	PLCH2	infomap 157, wt 530	PLCH2 is a member of the PLC-eta family of the phosphoinositide-specific phospholipase C (PLC) superfamily of enzymes that cleave PtdIns(4,5) P2 to generate second messengers inositol 1,4,5-trisphosphate and diacylglycerol (Zhou et al., 2005 [PubMed 16107206]).[supplied by OMIM, Jun 2009]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9892	SNAP91	$fc_2, louvain9$	Predicted to enable several functions, including SNARE binding activity; clathrin binding activity; and phosphatidylinositol binding activity. Acts upstream of or within regulation of clathrin-dependent endocytosis. Predicted to be located in several cellular components, including postsynaptic density; presynaptic endosome; and presynaptic membrane. Predicted to be extrinsic component of endosome membrane. Predicted to be active in several cellular components, including Schaffer collateral - CA1 synapse; cytoplasmic vesicle; and parallel fiber to Purkinje cell synapse. Predicted to be extrinsic component of presynaptic endocytic zone membrane. Biomarker of Alzheimer's disease. [provided by Alliance of Genome Resources, Apr 2022]
10529	NEBL	fc_2	This gene encodes a nebulin like protein that is abundantly expressed in cardiac muscle. The encoded protein binds actin and interacts with thin filaments and Z-line associated proteins in striated muscle. This protein may be involved in cardiac myofibril assembly. A shorter isoform of this protein termed LIM nebulette is expressed in non-muscle cells and may function as a component of focal adhesion complexes. Alternate splicing results in multiple transcript variants. [provided by RefSeq, Mar 2010]
257194	NEGR1	wt 217	Predicted to act upstream of or within several processes, including feeding behavior; locomotory behavior; and positive regulation of neuron projection development. Predicted to be located in extracellular region and plasma membrane. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
25843	MOB4	fc_2	This gene was identified based on its similarity with the mouse counterpart. Studies of the mouse counterpart suggest that the expression of this gene may be regulated during oocyte maturation and preimplantation following zygotic gene activation. Alternatively spliced transcript variants encoding distinct isoforms have been observed. Naturally occurring read-through transcription occurs between this locus and the neighboring locus HSPE1.[provided by RefSeq, Feb 2011]
2823	GPM6A	wt 221	Predicted to enable calcium channel activity. Involved in neuron migration and stem cell differentiation. Located in extracellular exosome. [provided by Alliance of Genome Resources, Apr 2022]
4137	MAPT	fc_2	This gene encodes the microtubule-associated protein tau (MAPT) whose transcript undergoes complex, regulated alternative splicing, giving rise to several mRNA species. MAPT transcripts are differentially expressed in the nervous system, depending on stage of neuronal maturation and neuron type. MAPT gene mutations have been associated with several neurodegenerative disorders such as Alzheimer's disease, Pick's disease, frontotemporal dementia, cortico-basal degeneration and progressive supranuclear palsy. [provided by RefSeq, Jul 2008]
5142	PDE4B	wt 171	This gene is a member of the type IV, cyclic AMP (cAMP)-specific, cyclic nucleotide phosphodiesterase (PDE) family. The encoded protein regulates the cellular concentrations of cyclic nucleotides and thereby play a role in signal transduction. Altered activity of this protein has been associated with schizophrenia and bipolar affective disorder. Alternative splicing and the use of alternative promoters results in multiple transcript variants encoding different isoforms. [provided by RefSeq, Jul 2014]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
5595	MAPK3	fc_2	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]
56924	PAK6	spectral 7	This gene encodes a member of a family of p21-stimulated serine/threonine protein kinases, which contain an amino-terminal Cdc42/Rac interactive binding (CRIB) domain and a carboxyl-terminal kinase domain. These kinases function in a number of cellular processes, including cytoskeleton rearrangement, apoptosis, and the mitogen-activated protein (MAP) kinase signaling pathway. The protein encoded by this gene interacts with androgen receptor (AR) and translocates to the nucleus, where it is involved in transcriptional regulation. Changes in expression of this gene have been linked to prostate cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2015]
57670	KIAA1549	spectral 7	The protein encoded by this gene belongs to the UPF0606 family. This gene has been found to be fused to the BRAF oncogene in many cases of pilocytic astrocytoma. The fusion results from 2Mb tandem duplications at 7q34. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Oct 2012]

Continued on next page

Table B.6 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
64219	PJA1	spectral 46	This gene encodes an enzyme that has E2-dependent E3 ubiquitin-protein ligase activity. This enzyme belongs to a class of ubiquitin ligases that include a RING finger motif, and it can interact with the E2 ubiquitin-conjugating enzyme UbcH5B. This gene is located in an area of chromosome X where several X-linked cognitive disability disorders have been associated, and it has also been found as part of a contiguous gene deletion associated with craniofrontonasal syndrome, though a direct link to any disorder has yet to be demonstrated. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2010]
64478	CSMD1	wt 532	Predicted to act upstream of or within several processes, including learning or memory; mammary gland branching involved in pregnancy; and reproductive structure development. Predicted to be integral component of membrane. [provided by Alliance of Genome Resources, Apr 2022]
775	CACNA1C	sgG2 41	This gene encodes an alpha-1 subunit of a voltage-dependent calcium channel. Calcium channels mediate the influx of calcium ions into the cell upon membrane polarization. The alpha-1 subunit consists of 24 transmembrane segments and forms the pore through which ions pass into the cell. The calcium channel consists of a complex of alpha-1, alpha-2/delta, beta, and gamma subunits in a 1:1:1:1 ratio. There are multiple isoforms of each of these proteins, either encoded by different genes or the result of alternative splicing of transcripts. The protein encoded by this gene binds to and is inhibited by dihydropyridine. Alternative splicing results in many transcript variants encoding different proteins. Some of the predicted proteins may not produce functional ion channel subunits. [provided by RefSeq, Oct 2012]

B.7 SynGO Network SCZ Broad Enriched Genes

Table B.7: SCZ Broad genes from enriched clusters found from the SynGO PSD network, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
2903	GRIN2A	infomap 8, sgG1 23, sgG2 19, sgG5 22	This gene encodes a member of the glutamate-gated ion channel protein family. The encoded protein is an N-methyl-D-aspartate (NMDA) receptor subunit. NMDA receptors are both ligand-gated and voltage-dependent, and are involved in long-term potentiation, an activity-dependent increase in the efficiency of synaptic transmission thought to underlie certain kinds of memory and learning. These receptors are permeable to calcium ions, and activation results in a calcium influx into post-synaptic cells, which results in the activation of several signaling cascades. Disruption of this gene is associated with focal epilepsy and speech disorder with or without cognitive disability. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2014]
60412	EXOC4	infomap 8, sgG1 23, sgG2 19, sgG5 22	The protein encoded by this gene is a component of the exocyst complex, a multiple protein complex essential for targeting exocytic vesicles to specific docking sites on the plasma membrane. Though best characterized in yeast, the component proteins and functions of exocyst complex have been demonstrated to be highly conserved in higher eukaryotes. At least eight components of the exocyst complex, including this protein, are found to interact with the actin cytoskeletal remodeling and vesicle transport machinery. The complex is also essential for the biogenesis of epithelial cell surface polarity. Alternate transcriptional splice variants, encoding different isoforms, have been characterized. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
8525	DGKZ	infomap 8, sgG1 23, sgG2 19, sgG5 22	The protein encoded by this gene belongs to the eukaryotic diacylglycerol kinase family. It may attenuate protein kinase C activity by regulating diacylglycerol levels in intracellular signaling cascade and signal transduction. Alternative splicing occurs at this locus and multiple transcript variants encoding distinct isoforms have been identified. [provided by RefSeq, Nov 2010]
9228	DLGAP2	infomap 8, sgG1 23, sgG2 19, sgG5 22	The product of this gene is a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain, however, only the paternal allele is expressed in the testis (PMID:18055845). Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jun 2014]
9758	FRMPD4	infomap 8, sgG1 23, sgG2 19, sgG5 22	This gene encodes a multi-domain (WW, PDZ, FERM) containing protein. Through its interaction with other proteins (such as PSD-95), it functions as a positive regulator of dendritic spine morphogenesis and density, and is required for the maintenance of excitatory synaptic transmission. [provided by RefSeq, Jan 2010]
1630	DCC	infomap 4, sgG2 13, sgG5 72	This gene encodes a netrin 1 receptor. The transmembrane protein is a member of the immunoglobulin superfamily of cell adhesion molecules, and mediates axon guidance of neuronal growth cones towards sources of netrin 1 ligand. The cytoplasmic tail interacts with the tyrosine kinases Src and focal adhesion kinase (FAK, also known as PTK2) to mediate axon attraction. The protein partially localizes to lipid rafts, and induces apoptosis in the absence of ligand. The protein functions as a tumor suppressor, and is frequently mutated or downregulated in colorectal cancer and esophageal carcinoma. [provided by RefSeq, Oct 2009]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
1983	EIF5	infomap 10, sgG1 13, sgG5 69	Eukaryotic translation initiation factor-5 (EIF5) interacts with the 40S initiation complex to promote hydrolysis of bound GTP with concomitant joining of the 60S ribosomal subunit to the 40S initiation complex. The resulting functional 80S ribosomal initiation complex is then active in peptidyl transfer and chain elongations (summary by Si et al., 1996 [PubMed 8663286]).[supplied by OMIM, May 2010]
4900	NRGN	infomap 10, sgG1 13, sgG5 69	Neurogranin (NRGN) is the human homolog of the neuron-specific rat RC3/neurogranin gene. This gene encodes a postsynaptic protein kinase substrate that binds calmodulin in the absence of calcium. The NRGN gene contains four exons and three introns. The exons 1 and 2 encode the protein and exons 3 and 4 contain untranslated sequences. It is suggested that the NRGN is a direct target for thyroid hormone in human brain, and that control of expression of this gene could underlay many of the consequences of hypothyroidism on mental states during development as well as in adult subjects. [provided by RefSeq, Jul 2008]
5595	MAPK3	infomap 4, sgG2 13, sgG5 72	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
56924	PAK6	infomap 4, sgG2 13, sgG5 72	This gene encodes a member of a family of p21-stimulated serine/threonine protein kinases, which contain an amino-terminal Cdc42/Rac interactive binding (CRIB) domain and a carboxyl-terminal kinase domain. These kinases function in a number of cellular processes, including cytoskeleton rearrangement, apoptosis, and the mitogen-activated protein (MAP) kinase signaling pathway. The protein encoded by this gene interacts with androgen receptor (AR) and translocates to the nucleus, where it is involved in transcriptional regulation. Changes in expression of this gene have been linked to prostate cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2015]
57502	NLGN4X	infomap 8, sgG1 23, sgG2 19	This gene encodes a member of the type-B carboxylesterase/lipase protein family. The encoded protein belongs to a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and may be involved in the formation and remodeling of central nervous system synapses. The encoded protein interacts with discs large homolog 4 (DLG4). Mutations in this gene have been associated with autism and Asperger syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Aug 2013]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
5970	RELA	infomap 4, sgG2 13, sgG5 72	NF-kappa-B is a ubiquitous transcription factor involved in several biological processes. It is held in the cytoplasm in an inactive state by specific inhibitors. Upon degradation of the inhibitor, NF-kappa-B moves to the nucleus and activates transcription of specific genes. NF-kappa-B is composed of NFKB1 or NFKB2 bound to either REL, RELA, or RELB. The most abundant form of NF-kappa-B is NFKB1 complexed with the product of this gene, RELA. Four transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Sep 2011]
9568	GABBR2	infomap 5, sgG1 23, sgG5 29	The multi-pass membrane protein encoded by this gene belongs to the G-protein coupled receptor 3 family and GABA-B receptor subfamily. The GABA-B receptors inhibit neuronal activity through G protein-coupled second-messenger systems, which regulate the release of neurotransmitters, and the activity of ion channels and adenylyl cyclase. This receptor subunit forms an active heterodimeric complex with GABA-B receptor subunit 1, neither of which is effective on its own. Allelic variants of this gene have been associated with nicotine dependence.[provided by RefSeq, Jan 2010]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
1946	EFNA5	infomap 5, sgG1 23	Ephrin-A5, a member of the ephrin gene family, prevents axon bundling in cocultures of cortical neurons with astrocytes, a model of late stage nervous system development and differentiation. The EPH and EPH-related receptors comprise the largest subfamily of receptor protein-tyrosine kinases and have been implicated in mediating developmental events, particularly in the nervous system. EPH receptors typically have a single kinase domain and an extracellular region containing a Cys-rich domain and 2 fibronectin type III repeats. The ephrin ligands and receptors have been named by the Eph Nomenclature Committee (1997). Based on their structures and sequence relationships, ephrins are divided into the ephrin-A (EFNA) class, which are anchored to the membrane by a glycosylphosphatidylinositol linkage, and the ephrin-B (EFNB) class, which are transmembrane proteins. The Eph family of receptors are similarly divided into 2 groups based on the similarity of their extracellular domain sequences and their affinities for binding ephrin-A and ephrin-B ligands. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2185	PTK2B	infomap 5, sgG1 23	This gene encodes a cytoplasmic protein tyrosine kinase which is involved in calcium-induced regulation of ion channels and activation of the map kinase signaling pathway. The encoded protein may represent an important signaling intermediate between neuropeptide-activated receptors or neurotransmitters that increase calcium flux and the downstream signals that regulate neuronal activity. The encoded protein undergoes rapid tyrosine phosphorylation and activation in response to increases in the intracellular calcium concentration, nicotinic acetylcholine receptor activation, membrane depolarization, or protein kinase C activation. This protein has been shown to bind CRK-associated substrate, nephrocystin, GTPase regulator associated with FAK, and the SH2 domain of GRB2. The encoded protein is a member of the FAK subfamily of protein tyrosine kinases but lacks significant sequence similarity to kinases from other subfamilies. Four transcript variants encoding two different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
23513	SCRIB	infomap 8, sgG2 19	This gene encodes a protein that was identified as being similar to the Drosophila scribble protein. The mammalian protein is involved in tumor suppression pathways. As a scaffold protein involved in cell polarization processes, this protein binds to many other proteins. The encoded protein binds to papillomavirus E6 protein via its PDZ domain and the C-terminus of E6. Two alternatively spliced transcript variants that encode different protein isoforms have been found for this gene. [provided by RefSeq, Nov 2011]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
2911	GRM1	infomap 5, sgG1 23	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]
6622	SNCA	infomap 5, sgG1 23	Alpha-synuclein is a member of the synuclein family, which also includes beta- and gamma-synuclein. Synucleins are abundantly expressed in the brain and alpha- and beta-synuclein inhibit phospholipase D2 selectively. SNCA may serve to integrate presynaptic signaling and membrane trafficking. Defects in SNCA have been implicated in the pathogenesis of Parkinson disease. SNCA peptides are a major component of amyloid plaques in the brains of patients with Alzheimer's disease. Alternatively spliced transcripts encoding different isoforms have been identified for this gene. [provided by RefSeq, Feb 2016]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
85358	SHANK3	infomap 5, sgG1 23	This gene is a member of the Shank gene family. Shank proteins are multidomain scaffold proteins of the postsynaptic density that connect neurotransmitter receptors, ion channels, and other membrane proteins to the actin cytoskeleton and G-protein-coupled signaling pathways. Shank proteins also play a role in synapse formation and dendritic spine maturation. Mutations in this gene are a cause of autism spectrum disorder (ASD), which is characterized by impairments in social interaction and communication, and restricted behavioral patterns and interests. Mutations in this gene also cause schizophrenia type 15, and are a major causative factor in the neurological symptoms of 22q13.3 deletion syndrome, which is also known as Phelan-McDermid syndrome. Additional isoforms have been described for this gene but they have not yet been experimentally verified. [provided by RefSeq, Mar 2012]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
11141	IL1RAPL1	sgG5 29	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]
1813	DRD2	sgG5 29	This gene encodes the D2 subtype of the dopamine receptor. This G-protein coupled receptor inhibits adenylyl cyclase activity. A missense mutation in this gene causes myoclonus dystonia; other mutations have been associated with schizophrenia. Alternative splicing of this gene results in two transcript variants encoding different isoforms. A third variant has been described, but it has not been determined whether this form is normal or due to aberrant splicing. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
3745	KCNB1	sgG1 23	Voltage-gated potassium (Kv) channels represent the most complex class of voltage-gated ion channels from both functional and structural standpoints. Their diverse functions include regulating neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume. Four sequence-related potassium channel genes - shaker, shaw, shab, and shal - have been identified in Drosophila, and each has been shown to have human homolog(s). This gene encodes a member of the potassium channel, voltage-gated, shab-related subfamily. This member is a delayed rectifier potassium channel and its activity is modulated by some other family members. [provided by RefSeq, Jul 2008]
4905	NSF	sgG5 39	Enables PDZ domain binding activity and ionotropic glutamate receptor binding activity. Involved in intracellular protein transport; positive regulation of protein catabolic process; and positive regulation of receptor recycling. Located in Golgi apparatus; cytosol; and plasma membrane. Implicated in developmental and epileptic encephalopathy. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
6137	RPL13	infomap 5	Ribosomes, the organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L13E family of ribosomal proteins. It is located in the cytoplasm. This gene is expressed at significantly higher levels in benign breast lesions than in breast carcinomas. Alternatively spliced transcript variants encoding distinct isoforms have been found for this gene. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudogenes of this gene dispersed through the genome. [provided by RefSeq, Jul 2011]
6536	SLC6A9	infomap 8	The amino acid glycine acts as an inhibitory neurotransmitter in the central nervous system. The protein encoded by this gene is one of two transporters that stop glycine signaling by removing it from the synaptic cleft. [provided by RefSeq, Jun 2016]
9344	TAOK2	sgG5 39	Enables mitogen-activated protein kinase kinase binding activity; neuropilin binding activity; and protein serine/threonine kinase activity. Involved in several processes, including focal adhesion assembly; intracellular signal transduction; and positive regulation of JNK cascade. Located in cytoplasmic vesicle; cytosol; and nuclear lumen. Part of receptor complex. [provided by Alliance of Genome Resources, Apr 2022]

Continued on next page

Table B.7 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9892	SNAP91	sgG1 23	Predicted to enable several functions, including SNARE binding activity; clathrin binding activity; and phosphatidylinositol binding activity. Acts upstream of or within regulation of clathrin-dependent endocytosis. Predicted to be located in several cellular components, including postsynaptic density; presynaptic endosome; and presynaptic membrane. Predicted to be extrinsic component of endosome membrane. Predicted to be active in several cellular components, including Schaffer collateral - CA1 synapse; cytoplasmic vesicle; and parallel fiber to Purkinje cell synapse. Predicted to be extrinsic component of presynaptic endocytic zone membrane. Biomarker of Alzheimer's disease. [provided by Alliance of Genome Resources, Apr 2022]

B.8 SynGO Network SCZ Priortised Enriched Genes

Table B.8: SCZ Priortised genes from enriched clusters found from the SynGO PSD network, it includes the entrez gene ID, the offical Gene Symbol, the specific clusters where it is found and the summary. The summary was scraped from PubMed.

Entrez Gene ID	**Gene Symbol**	**Algorithm & Cluster Associated**	**Summary**
2903	GRIN2A	infomap 8, sgG1 23, sgG2 19, spectral 5	This gene encodes a member of the glutamate-gated ion channel protein family. The encoded protein is an N-methyl-D-aspartate (NMDA) receptor subunit. NMDA receptors are both ligand-gated and voltage-dependent, and are involved in long-term potentiation, an activity-dependent increase in the efficiency of synaptic transmission thought to underlie certain kinds of memory and learning. These receptors are permeable to calcium ions, and activation results in a calcium influx into post-synaptic cells, which results in the activation of several signaling cascades. Disruption of this gene is associated with focal epilepsy and speech disorder with or without cognitive disability. Alternative splicing results in multiple transcript variants. [provided by RefSeq, May 2014]
57502	NLGN4X	infomap 8, sgG1 23, sgG2 19, spectral 5	This gene encodes a member of the type-B carboxylesterase/lipase protein family. The encoded protein belongs to a family of neuronal cell surface proteins. Members of this family may act as splice site-specific ligands for beta-neurexins and may be involved in the formation and remodeling of central nervous system synapses. The encoded protein interacts with discs large homolog 4 (DLG4). Mutations in this gene have been associated with autism and Asperger syndrome. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Aug 2013]

Continued on next page

Table B.8 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
9228	DLGAP2	infomap 8, sgG1 23, sgG2 19, spectral 5	The product of this gene is a membrane-associated protein that may play a role in synapse organization and signalling in neuronal cells. This gene is biallelically expressed in the brain, however, only the paternal allele is expressed in the testis (PMID:18055845). Alternatively spliced transcript variants encoding different isoforms have been identified. [provided by RefSeq, Jun 2014]
1630	DCC	infomap 4, sgG2 13, sgG5 72	This gene encodes a netrin 1 receptor. The transmembrane protein is a member of the immunoglobulin superfamily of cell adhesion molecules, and mediates axon guidance of neuronal growth cones towards sources of netrin 1 ligand. The cytoplasmic tail interacts with the tyrosine kinases Src and focal adhesion kinase (FAK, also known as PTK2) to mediate axon attraction. The protein partially localizes to lipid rafts, and induces apoptosis in the absence of ligand. The protein functions as a tumor suppressor, and is frequently mutated or downregulated in colorectal cancer and esophageal carcinoma. [provided by RefSeq, Oct 2009]
5595	MAPK3	infomap 4, sgG2 13, sgG5 72	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases, also known as extracellular signal-regulated kinases (ERKs), act in a signaling cascade that regulates various cellular processes such as proliferation, differentiation, and cell cycle progression in response to a variety of extracellular signals. This kinase is activated by upstream kinases, resulting in its translocation to the nucleus where it phosphorylates nuclear targets. Alternatively spliced transcript variants encoding different protein isoforms have been described. [provided by RefSeq, Jul 2008]

Continued on next page

Table B.8 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
56924	PAK6	infomap 4, sgG2 13, sgG5 72	This gene encodes a member of a family of p21-stimulated serine/threonine protein kinases, which contain an amino-terminal Cdc42/Rac interactive binding (CRIB) domain and a carboxyl-terminal kinase domain. These kinases function in a number of cellular processes, including cytoskeleton rearrangement, apoptosis, and the mitogen-activated protein (MAP) kinase signaling pathway. The protein encoded by this gene interacts with androgen receptor (AR) and translocates to the nucleus, where it is involved in transcriptional regulation. Changes in expression of this gene have been linked to prostate cancer. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Dec 2015]
9568	GABBR2	sgG1 23, sgG5 29, spectral 5	The multi-pass membrane protein encoded by this gene belongs to the G-protein coupled receptor 3 family and GABA-B receptor subfamily. The GABA-B receptors inhibit neuronal activity through G protein-coupled second-messenger systems, which regulate the release of neurotransmitters, and the activity of ion channels and adenylyl cyclase. This receptor subunit forms an active heterodimeric complex with GABA-B receptor subunit 1, neither of which is effective on its own. Allelic variants of this gene have been associated with nicotine dependence.[provided by RefSeq, Jan 2010]

Continued on next page

Table B.8 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
11141	IL1RAPL1	sgG1 20, sgG5 29	The protein encoded by this gene is a member of the interleukin 1 receptor family and is similar to the interleukin 1 accessory proteins. This protein has an N-terminal signal peptide, three extracellular immunoglobulin Ig-like domains, a transmembrane domain, an intracellular Toll/IL-1R domain, and a long C-terminal tail which interacts with multiple signalling molecules. This gene is located at a region on chromosome X that is associated with a non-syndromic form of X-linked intellectual disability. Deletions and mutations in this gene were found in patients with intellectual disability. This gene is expressed at a high level in post-natal brain structures involved in the hippocampal memory system, which suggests a specialized role in the physiological processes underlying memory and learning abilities, and plays a role in synapse formation and stabilization. [provided by RefSeq, Jul 2017]
2911	GRM1	sgG1 23	This gene encodes a metabotropic glutamate receptor that functions by activating phospholipase C. L-glutamate is the major excitatory neurotransmitter in the central nervous system and activates both ionotropic and metabotropic glutamate receptors. Glutamatergic neurotransmission is involved in most aspects of normal brain function and can be perturbed in many neuropathologic conditions. The canonical alpha isoform of the encoded protein is a disulfide-linked homodimer whose activity is mediated by a G-protein-coupled phosphatidylinositol-calcium second messenger system. This gene may be associated with many disease states, including schizophrenia, bipolar disorder, depression, and breast cancer. Alternative splicing results in multiple transcript variants encoding different isoforms. [provided by RefSeq, May 2013]

Continued on next page

Table B.8 – continued from previous page

Entrez Gene ID	Gene Symbol	Algorithm & Cluster Associated	Summary
5789	PTPRD	sgG1 20	The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. This PTP contains an extracellular region, a single transmembrane segment and two tandem intracytoplasmic catalytic domains, and thus represents a receptor-type PTP. The extracellular region of this protein is composed of three Ig-like and eight fibronectin type III-like domains. Studies of the similar genes in chicken and fly suggest the role of this PTP is in promoting neurite growth, and regulating neurons axon guidance. Multiple alternatively spliced transcript variants of this gene have been reported. A related pseudogene has been identified on chromosome 5. [provided by RefSeq, Jan 2010]
9892	SNAP91	sgG1 23	Predicted to enable several functions, including SNARE binding activity; clathrin binding activity; and phosphatidylinositol binding activity. Acts upstream of or within regulation of clathrin-dependent endocytosis. Predicted to be located in several cellular components, including postsynaptic density; presynaptic endosome; and presynaptic membrane. Predicted to be extrinsic component of endosome membrane. Predicted to be active in several cellular components, including Schaffer collateral - CA1 synapse; cytoplasmic vesicle; and parallel fiber to Purkinje cell synapse. Predicted to be extrinsic component of presynaptic endocytic zone membrane. Biomarker of Alzheimer's disease. [provided by Alliance of Genome Resources, Apr 2022]

B.9 SynGO Full Jaccard Index

Alg, Cluster No.	**Cluster Size**	**No. SCZ**	**Alg, Cluster No.**	**Cluster Size**	**No. SCZ**	**Jaccard Index**
infomap 10	4	2	sgG1 13	2	4	1.0
sgG1 13	4	2	sgG5 69	2	4	1.0
infomap 10	4	2	sgG5 69	2	4	1.0
infomap 4	17	4	sgG5 72	4	15	0.7777777777777778
infomap 8	53	8	sgG2 19	7	59	0.75
infomap 4	17	4	sgG2 13	4	23	0.6666666666666666
infomap 8	53	8	sgG5 22	5	38	0.6545454545454545
sgG2 13	23	4	sgG5 72	4	15	0.6521739130434783
sgG2 19	59	7	sgG5 22	5	38	0.5645161290322581
sgG2 19	59	7	spectral 5	0	82	0.532608695652174
infomap 8	53	8	spectral 5	0	82	0.4673913043478261
sgG1 23	126	14	sgG2 19	7	59	0.3805970149253731
sgG1 23	126	14	spectral 5	0	82	0.35947712418300654
infomap 5	52	7	sgG1 23	14	126	0.35877862595419846
sgG5 22	38	5	spectral 5	0	82	0.34831460674157305
infomap 8	53	8	sgG1 23	14	126	0.32592592592592595
sgG1 23	126	14	sgG5 22	5	38	0.28125

Table B.9: Table to compare enriched clusters between different algorithms for the SynGO Synaptic Graph. This compares their similarity using the Jaccard Index, explained in 3.2.1. Only those with a Jaccard Index greater than 0.1 were included.

Gene	Full Network Regions	SynGO Network Regions
CTBP1	1, 3	N/A
CTBP2	1, 3	N/A
DLG2	N/A	3
DLGAP1	4	N/A
DLGAP2	4	N/A
DLGAP3	4	N/A
LRRC7	N/A	3
PACSIN1	1, 3	3
PPFIA1	N/A	4
PPFIA3	1, 3	3
PPFIBP2	N/A	4
SEPTIN6	N/A	3
SEPTIN8	N/A	3
SEPTIN11	N/A	3
SHANK2	3	N/A

Table B.10: Table of Showing the proteins found in the Bridgeness graph for both the SynGO and the Full Network. It mentions the regions that they appeared in

B.10 Bridgeness Table

B.11 WordCloud Tables: Consensus

B.11.1 Molecular Function

Table B.11: Molecular Function Frequencies and Weights

Molecular Function	Frequency	Weights
NMDA glutamate receptor activity	8	7.968281090857101
glutamate-gated calcium ion channel activity	8	7.9627925213937205
scaffold protein binding	7	6.970118117263481
PDZ domain binding	7	6.9332334761311945
SH3 domain binding	5	4.996938525162547
structural constituent of postsynaptic density	5	4.9517183956470845
voltage-gated potassium channel activity	5	4.937737225311367
phosphotyrosine residue binding	4	3.9885200915117416
inositol 1,4,5 trisphosphate binding	4	3.981379614184542
ephrin receptor binding	4	3.978299556237194
ionotropic glutamate receptor binding	4	3.976585251501493
synaptic receptor adaptor activity	4	3.9527373906770356

transmitter-gated monoatomic ion channel activity involved in regulation of postsynaptic membrane potential	4	3.913829619275053
protein tyrosine kinase activity	3	2.9932587685593
protein tyrosine phosphatase activity	3	2.9555395093417034
glutamate binding	3	2.9498634525951175
inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity	3	2.9330816642749586
protein serine kinase activity	2	1.9999998891251434
protein kinase activity	2	1.999999873547418
protein serine/threonine kinase activity	2	1.9999989129618936
tau protein binding	2	1.999182500350735
GTPase activator activity	2	1.9983314629044329
MAP kinase kinase activity	2	1.997212342552855
neurexin family protein binding	2	1.9923329514388761
transmembrane transporter binding	2	1.991280606620555
MAP kinase kinase kinase activity	2	1.9887929970663907
diacylglycerol-dependent serine/threonine kinase activity	2	1.9887929970663907
calcium,diacylglycerol-dependent serine/threonine kinase activity	2	1.9887929970663907
non-membrane spanning protein tyrosine kinase activity	2	1.9750253081500124
voltage-gated monoatomic ion channel activity involved in regulation of postsynaptic membrane potential	2	1.9659576528169072
ATP binding	2	1.960709024380624
phosphoserine residue binding	2	1.9586859796254974
receptor tyrosine kinase binding	2	1.9548561685254529
enzyme binding	2	1.9544184840284238
potassium channel regulator activity	2	1.9496760686734427
delayed rectifier potassium channel activity	2	1.9496632505163694
calmodulin-dependent protein kinase activity	1	0.999999987170733
protein phosphatase 2A binding	1	0.999992691078995
calmodulin binding	1	0.9999836351085142
myosin phosphatase activity	1	0.9999731447948927
voltage-gated calcium channel activity	1	0.9999533686027647
high voltage-gated calcium channel activity	1	0.9999533686027647
armadillo repeat domain binding	1	0.9994165787790221
hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amides	1	0.9992387614832327

protein serine/threonine phosphatase activity	1	0.997587270560369
protein phosphatase regulator activity	1	0.9975857591453428
adenyl-nucleotide exchange factor activity	1	0.9966658450613574
transmembrane receptor protein tyrosine phosphatase activity	1	0.9915558647830247
G protein-coupled receptor activity	1	0.9909961460345775
voltage-gated monoatomic ion channel activity involved in regulation of presynaptic membrane potential	1	0.9900312826817976
guanyl-nucleotide exchange factor activity	1	0.9859585443510411
integrin binding	1	0.9850235394874555
protein kinase binding	1	0.9840959659985632
phosphoprotein phosphatase activity	1	0.9840959659985632
phosphatase activity	1	0.9840959659985632
AMP-activated protein kinase activity	1	0.9729904520511308
beta-1 adrenergic receptor binding	1	0.9706085740081156
heat shock protein binding	1	0.9703301427735135
kinase activity	1	0.9703301427735135
voltage-gated calcium channel activity involved in cardiac muscle cell action potential	1	0.969538446412087
voltage-gated calcium channel activity involved in regulation of presynaptic cytosolic calcium levels	1	0.969538446412087
JUN kinase binding	1	0.9694344942512043
amyloid-beta binding	1	0.9681122791754
protein tyrosine kinase activator activity	1	0.9645728250338829
calmodulin-dependent protein phosphatase activity	1	0.9645728250338829
aldo-keto reductase (NADP) activity	1	0.9645728250338829
tau-protein kinase activity	1	0.9633335091473111
guanylate kinase activity	1	0.9629961133211397
signaling receptor activity	1	0.9593098557648041
protein phosphatase binding	1	0.9585272061240303
latrotoxin receptor activity	1	0.9582756358493785
cytoskeletal protein binding	1	0.9547419381922811
signaling adaptor activity	1	0.95209878327697

Table B.11: Word FrequenciesConsensus

B.11.2 Biological Function

Table B.12: Biological Function Frequencies and Weights for consensus network.

Biological Function	**Frequency**	**Weights**
positive regulation of excitatory postsynaptic potential	8	7.999998078151291
positive regulation of synaptic transmission, glutamatergic	8	7.99927157417397
chemical synaptic transmission	8	7.99720082555713
regulation of synaptic plasticity	8	7.9844626374242464
protein localization to synapse	7	6.98960072868879
regulation of monoatomic cation transmembrane transport	7	6.984142309253257
excitatory postsynaptic potential	7	6.9734756248818375
embryo development	7	6.955540810637775
calcium ion transmembrane import into cytosol	7	6.948078067275197
receptor localization to synapse	7	6.93651074007219
social behavior	6	5.986296941703073
receptor clustering	6	5.979862124974069
excitatory chemical synaptic transmission	6	5.976899350950737
positive regulation of synapse assembly	6	5.958497955276778
monoatomic cation transmembrane transport	6	5.951387427023393
long-term synaptic potentiation	6	5.934104810252408
cell-cell adhesion	6	5.933586759546337
calcium-mediated signaling	6	5.9323291486049525
synapse assembly	6	5.924459979092347
regulation of postsynaptic neurotransmitter receptor activity	6	5.895883839147323
neurotransmitter receptor localization to postsynaptic specialization membrane	6	5.8290123778325835
vocalization behavior	5	4.986889864603361
ionotropic glutamate receptor signaling pathway	5	4.9700390079089525
dendritic spine morphogenesis	5	4.93890666748877
regulation of postsynapse organization	5	4.910235748800718
epidermal growth factor receptor signaling pathway	4	3.9966133751021844
peptidyl-tyrosine phosphorylation	4	3.99170495447328
ephrin receptor signaling pathway	4	3.987147824089854
regulation of neuronal synaptic plasticity	4	3.9693144644055227
heterophilic cell-cell adhesion via plasma membrane cell adhesion molecules	4	3.9564835382222516
AMPA glutamate receptor clustering	4	3.95404517255763

maintenance of postsynaptic density structure	4	3.95404517255763
modulation of chemical synaptic transmission	4	3.9531542938551882
positive regulation of dendritic spine development	4	3.9450870907463598
postsynaptic membrane assembly	4	3.9212500955064864
positive regulation of long-term synaptic depression	4	3.9044297156670784
signal transduction	3	2.985147322872259
establishment or maintenance of epithelial cell apical/basal polarity	3	2.981106373763823
potassium ion transmembrane transport	3	2.974715958986814
postsynaptic density protein 95 clustering	3	2.9585887793761856
synaptic vesicle endocytosis	3	2.9559861421674203
synaptic vesicle uncoating	3	2.9473626443738805
sensory perception of pain	3	2.9386104942974867
regulation of monoatomic ion transmembrane transport	3	2.937995511398736
axon guidance	3	2.9335696223975662
regulation of cell adhesion mediated by integrin	3	2.932155855458593
cell surface receptor signaling pathway	3	2.924172304885706
negative regulation of protein catabolic process	3	2.9098584154174723
protein phosphorylation	2	1.9999999781797941
MAPK cascade	2	1.9999631171176266
positive regulation of ERK1 and ERK2 cascade	2	1.999803565493966
peptidyl-serine phosphorylation	2	1.99826367627467
JNK cascade	2	1.9967725524620141
peptidyl-threonine phosphorylation	2	1.9956718201474408
thyroid gland development	2	1.993716564130959
Fc-epsilon receptor signaling pathway	2	1.9917273336577896
intracellular signal transduction	2	1.9863402499051641
visual learning	2	1.9830659581157846
adult behavior	2	1.9799798927544412
cell adhesion	2	1.9785137159473394
protein autophosphorylation	2	1.9768020535234314
positive regulation of substrate adhesion-dependent cell spreading	2	1.968456933333748
neuromuscular process controlling balance	2	1.9640072874512717
regulation of presynaptic membrane potential	2	1.9608745783966814

cell migration	2	1.9603872470228487
regulation of small GTPase mediated signal transduction	2	1.9586963474759742
negative regulation of DNA-templated transcription	2	1.9580170223931685
cell chemotaxis	2	1.9576323174342136
protein dephosphorylation	2	1.9553971879693566
synaptic transmission, glutamatergic	2	1.955206724760688
regulation of actin cytoskeleton organization	2	1.9546685950833151
positive regulation of actin filament polymerization	2	1.9494382892510203
regulation of postsynaptic membrane potential	2	1.9474302853952534
positive regulation of JUN kinase activity	2	1.9446782626053507
regulation of dendritic spine morphogenesis	2	1.9433116912501307
neurotrophin TRK receptor signaling pathway	2	1.9288318897591883
vascular endothelial growth factor receptor signaling pathway	2	1.9288318897591883
RIG-I signaling pathway	2	1.928607713862776
activation of protein kinase C activity	2	1.928607713862776
synapse organization	2	1.9240603363811393
semaphorin-plexin signaling pathway	1	0.9999927485531542
fatty acid beta-oxidation using acyl-CoA oxidase	1	0.9999522688797764
actin cytoskeleton organization	1	0.9999513302407448
phosphorylation	1	0.999880648754647
positive regulation of protein phosphorylation	1	0.9993987396777513
Fc-gamma receptor signaling pathway involved in phagocytosis	1	0.9993241930576289
negative regulation of TORC1 signaling	1	0.9992806126456251
regulation of voltage-gated calcium channel activity	1	0.9988696042670128
Rho protein signal transduction	1	0.998629327072103
negative regulation of protein ubiquitination	1	0.998459279057962
response to xenobiotic stimulus	1	0.9972887776651762
cell junction assembly	1	0.9964600973839992
integrin-mediated signaling pathway	1	0.9955132910272484
establishment of cell polarity	1	0.9953446944547618
thymus development	1	0.9938760086833864
positive regulation of neuron projection development	1	0.9937050794448679

positive regulation of peptidyl-serine phosphorylation	1	0.9897464113647791
face development	1	0.9877872495436953
regulation of protein ubiquitination	1	0.9877872495436953
transmembrane receptor protein tyrosine kinase signaling pathway	1	0.9877872495436953
stimulatory C-type lectin receptor signaling pathway	1	0.9877872495436953
trachea formation	1	0.9877872495436953
cellular response to heat	1	0.9866455009258295
positive regulation of telomerase activity	1	0.9866455009258295
positive regulation of stress fiber assembly	1	0.9858611492422441
regulation of cell migration	1	0.9858611492422441
Wnt signaling pathway, planar cell polarity pathway	1	0.9851613456032348
calcium ion transmembrane transport	1	0.9846827296749125
alpha-linolenic acid metabolic process	1	0.9843475245671364
transmembrane receptor protein tyrosine phosphatase signaling pathway	1	0.9839876721716753
neuron cell-cell adhesion	1	0.9816894616741336
peptidyl-serine dephosphorylation	1	0.9804312351077195
cerebellar granule cell differentiation	1	0.978610637402428
negative regulation of excitatory postsynaptic potential	1	0.978610637402428
regulation of cell shape	1	0.9784672974969504
positive regulation of apoptotic process	1	0.9768020782175404
establishment of epithelial cell apical/basal polarity	1	0.9767600044683256
positive regulation of protein serine/threonine kinase activity	1	0.9756188577431083
positive regulation of protein kinase activity	1	0.9756188577431083
neuromuscular junction development	1	0.9735855710854403
positive regulation of proteasomal ubiquitin-dependent protein catabolic process	1	0.9728079781680927
regulation of postsynaptic density assembly	1	0.9700813109904978
positive regulation of glucose import	1	0.9699714583701466
ERK1 and ERK2 cascade	1	0.9678761843878104
locomotory exploration behavior	1	0.9678761843878104
stress-activated MAPK cascade	1	0.9678761843878104
positive regulation of reactive oxygen species biosynthetic process	1	0.9675273147061709
endocytosis	1	0.9675273147061709
presynaptic membrane assembly	1	0.9661932924405096
presynapse assembly	1	0.9661932924405096

regulation of circadian rhythm	1	0.9647826288824279
small GTPase-mediated signal transduction	1	0.964524791745007
positive regulation of epithelial cell migration	1	0.9634952257297379
axon extension	1	0.9617272690084422
learning or memory	1	0.9617272690084422
synaptic transmission, GABAergic	1	0.9594843145114943
insulin-like growth factor receptor signaling pathway	1	0.9593986885508426
ossification involved in bone maturation	1	0.9593050112062336
wound healing, spreading of cells	1	0.9593050112062336
cellular response to brain-derived neurotrophic factor stimulus	1	0.9588604313890416
calcineurin-NFAT signaling cascade	1	0.9588604313890416
regulation of stress-activated MAPK cascade	1	0.9588604313890416
lipopolysaccharide-mediated signaling pathway	1	0.9588604313890416
regulation of early endosome to late endosome transport	1	0.9588604313890416
glutamate receptor signaling pathway	1	0.9586907750323915
insulin receptor signaling pathway	1	0.9555907373704359
chaperone-mediated autophagy	1	0.9555907373704359
peptidyl-tyrosine autophosphorylation	1	0.9555907373704359
positive regulation of cysteine-type endopeptidase activity	1	0.9555907373704359
regulation of Golgi inheritance	1	0.9555907373704359
calcineurin-mediated signaling	1	0.9555907373704359
negative regulation of calcium ion import across plasma membrane	1	0.9555907373704359
synaptic vesicle coating	1	0.9555907373704359
startle response	1	0.9553533348276035
GDP metabolic process	1	0.9531682548132404
negative regulation of apoptotic process	1	0.9527886426884693
CD40 signaling pathway	1	0.9518951976742039
positive regulation of immunoglobulin production	1	0.9518951976742039
brain development	1	0.9503439066284844

Table B.12: Word FrequenciesConsensus

B.11.3 Diseases

Table B.13: Disease Frequencies and Weights

Molecular Function	**Frequency**	**Weights**
schizophrenia	9	8.962776034966947
intellectual disability	9	8.823617187263215
autism spectrum disorder	8	7.996111152884294
autistic disorder	8	7.996011540987601
bipolar disorder	7	6.868910902907373
epilepsy syndrome	6	5.944837472278476
multiple sclerosis	5	4.874125619692175
Parkinson's disease	5	4.856921204666978
Alzheimer's disease	4	3.962019631709562
Huntington's disease	2	1.9998212043795762
hypertension	2	1.9432526993164894
frontotemporal dementia	1	0.9779033202203808

Table B.13: Word FrequenciesConsensus

B.11.4 SynGO

Table B.14: SynGO Frequencies and Weights

SynGO	Frequency	Weights
modulation of chemical synaptic transmission (GO:0050804)	10	9.948102582816553
postsynaptic density, intracellular component (GO:0099092)	8	7.990435231882806
structural constituent of postsynaptic density (GO:0098919)	8	7.988079909773589
regulation of postsynaptic density assembly (GO:0099151)	8	7.950384525286893
integral component of postsynaptic density membrane (GO:0099061)	7	6.999472594519636
integral component of presynaptic membrane (GO:0099056)	7	6.995920098404641
neurotransmitter receptor localization to postsynaptic specialization membrane (GO:0099645)	7	6.938373826312844
regulation of postsynapse organization (GO:0099175)	7	6.922639669165332
integral component of postsynaptic specialization membrane (GO:0099060)	6	5.9213127038529425
integral component of postsynaptic membrane (GO:0099055)	5	4.998768360484098
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1904315)	5	4.972662727351398
postsynaptic density (GO:0014069)	5	4.941678894078165
integral component of presynaptic active zone membrane (GO:0099059)	5	4.911766837236149
voltage-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1905030)	5	4.898881215737809
regulation of presynapse assembly (GO:1905606)	4	3.9925691306241093
ligand-gated ion channel activity involved in regulation of presynaptic membrane potential (GO:0099507)	4	3.9408232503193488
regulation of postsynaptic neurotransmitter receptor activity (GO:0098962)	4	3.8620187354918625
postsynaptic specialization assembly (GO:0098698)	3	2.9755910878150322
extrinsic component of presynaptic active zone membrane (GO:0098891)	3	2.9581270534028454

anchored component of postsynaptic density membrane (GO:0099031)	3	2.8921418616112815
presynapse assembly (GO:0099054)	2	1.9992687605034505
postsynaptic actin cytoskeleton organization (GO:0098974)	2	1.972077242489263
presynaptic cytosol (GO:0099523)	1	0.996654474159664
synaptic vesicle endocytosis (GO:0048488)	1	0.9959687478697661
voltage-gated ion channel activity involved in regulation of presynaptic membrane potential (GO:0099508)	1	0.9950929659892772
synapse adhesion between pre- and postsynapse (GO:0099560)	1	0.9936749642438447
perisynaptic extracellular matrix (GO:0098966)	1	0.9934330021357601
trans-synaptic signaling by trans-synaptic complex (GO:0099545)	1	0.9931464432741246
modification of postsynaptic actin cytoskeleton (GO:0098885)	1	0.9929279608846068
regulation of neurotransmitter receptor localization to postsynaptic specialization membrane (GO:0098696)	1	0.9912124720866583
postsynaptic modulation of chemical synaptic transmission (GO:0099170)	1	0.9912124720866583
postsynapse (GO:0098794)	1	0.9898657723718011
trans-synaptic signaling (GO:0099537)	1	0.9890778129882247
presynapse (GO:0098793)	1	0.9843382471757733
presynaptic actin cytoskeleton organization (GO:0099140)	1	0.9689817847461139
modification of postsynaptic structure (GO:0099010)	1	0.9689817847461139
regulation of postsynaptic membrane neurotransmitter receptor levels (GO:0099072)	1	0.965790376880625
extrinsic component of postsynaptic density membrane (GO:0099147)	1	0.9608581553961019
postsynaptic density membrane (GO:0098839)	1	0.9604912842789639
presynaptic membrane (GO:0042734)	1	0.9541111215099994
regulation of modification of postsynaptic structure (GO:0099159)	1	0.9538020938983484
synaptic vesicle uncoating (GO:0016191)	1	0.951860359892866
regulation of postsynaptic cytosolic calcium levels (GO:0099566)	1	0.951860359892866

Table B.14: Word FrequenciesConsensus

B.12 WordCloud Tables:SynGO

B.12.1 Molecular Function

Table B.15: Molecular Function Frequencies and Weights

Molecular Function	Frequency	Weights
glutamate-gated calcium ion channel activity	5	4.9861986329595664
NMDA glutamate receptor activity	5	4.979080483329406
protein serine kinase activity	3	2.9999790727821085
MAP kinase kinase activity	3	2.999576856344165
protein serine/threonine kinase activity	3	2.997620207452349
ATP binding	3	2.9968556156356767
MAP kinase activity	3	2.995263212991596
protein tyrosine kinase activity	2	1.9999965153033412
SH3 domain binding	2	1.9995502494710198
ephrin receptor binding	2	1.998148346819758
non-membrane spanning protein tyrosine kinase activity	2	1.99081645816968
scaffold protein binding	2	1.9839518383782504
delayed rectifier potassium channel activity	2	1.980725520401545
neurexin family protein binding	2	1.9616987694127321
histone H3S10 kinase activity	2	1.9224583697952473
histone H3S28 kinase activity	2	1.9224583697952473
SH2 domain binding	1	0.9950309852595552
voltage-gated potassium channel activity	1	0.9882814297732327
glutamate binding	1	0.987550460359991
transmitter-gated monoatomic ion channel activity involved in regulation of postsynaptic membrane potential	1	0.9844564206600133
signaling adaptor activity	1	0.9839553230384052
protein self-association	1	0.9839553230384052
signaling receptor activity	1	0.9831539747142473
voltage-gated monoatomic ion channel activity involved in regulation of postsynaptic membrane potential	1	0.9689753041805258
protein tyrosine phosphatase activity	1	0.9501374344185892

Table B.15: Word FrequenciesSynGO

B.12.2 Biological Function

Table B.16: Molecular Function Frequencies and Weights

Biological Function	Frequency	Weights

positive regulation of excitatory postsynaptic potential	5	4.981904636782143
monoatomic cation transmembrane transport	5	4.967279759461208
chemical synaptic transmission	5	4.908629219279544
positive regulation of synaptic transmission, glutamatergic	4	3.9991893976906407
regulation of monoatomic cation transmembrane transport	4	3.9755421227927896
calcium ion transmembrane import into cytosol	4	3.9755421227927896
excitatory chemical synaptic transmission	4	3.9755421227927896
protein phosphorylation	4	3.97296888775511
embryo development	4	3.957794275159978
apoptotic process	3	2.9876867587148763
interleukin-1-mediated signaling pathway	3	2.9872083123878803
regulation of ossification	3	2.9858987952940956
ionotropic glutamate receptor signaling pathway	3	2.98564967316978
intracellular signal transduction	3	2.9795536444476345
cellular response to tumor necrosis factor	3	2.975278698225889
regulation of DNA-templated transcription	3	2.9723763991379557
social behavior	3	2.9648891587119177
receptor localization to synapse	3	2.961611094155576
postsynaptic membrane assembly	3	2.937718949409792
potassium ion transmembrane transport	3	2.9317747096526605
stress-activated MAPK cascade	3	2.930068208716664
positive regulation of synapse assembly	3	2.9293705917282296
receptor clustering	3	2.921815727976452
calcium-mediated signaling	3	2.904767187098487
peptidyl-tyrosine phosphorylation	2	1.999995400828331
thyroid gland development	2	1.990599997040043
regulation of stress-activated MAPK cascade	2	1.990599997040043
face development	2	1.990599997040043
regulation of Golgi inheritance	2	1.990599997040043
Bergmann glial cell differentiation	2	1.978459783492551
trachea formation	2	1.978459783492551
ERK1 and ERK2 cascade	2	1.978459783492551
regulation of early endosome to late endosome transport	2	1.978459783492551
regulation of monoatomic ion transmembrane transport	2	1.9763623208619727

cellular response to transforming growth factor beta stimulus	2	1.969840901230621
ephrin receptor signaling pathway	2	1.969840901230621
peptidyl-serine phosphorylation	2	1.9692697337636258
response to ethanol	2	1.961528895568437
presynaptic membrane assembly	2	1.9599809567240722
chemotaxis	2	1.9502936284100307
lipopolysaccharide-mediated signaling pathway	2	1.9408225277009357
epidermal growth factor receptor signaling pathway	2	1.9396271724635725
synaptic membrane adhesion	1	0.9999622502439091
protein autophosphorylation	1	0.9964445013631766
neuromuscular process controlling balance	1	0.99370365634578
regulation of cell adhesion mediated by integrin	1	0.9921029085449012
positive regulation of cell migration	1	0.9921029085449012
establishment or maintenance of epithelial cell apical/basal polarity	1	0.9919568049679576
regulation of presynapse assembly	1	0.9868642892392652
trans-synaptic signaling by trans-synaptic complex	1	0.9868642892392652
brain development	1	0.9821400922170009
cell-cell adhesion	1	0.9807781605288621
excitatory postsynaptic potential	1	0.9788870771473469
neuron cell-cell adhesion	1	0.9748221096621892
Fc-gamma receptor signaling pathway involved in phagocytosis	1	0.9744612613653995
regulation of neuronal synaptic plasticity	1	0.9743570548347428
vocalization behavior	1	0.972284370552902
insulin-like growth factor receptor signaling pathway	1	0.9721892177320515
transmembrane receptor protein tyrosine phosphatase signaling pathway	1	0.9689289918544158
signal transduction	1	0.9644379887173694
positive regulation of peptidyl-tyrosine phosphorylation	1	0.9644379887173694
cellular response to lipopolysaccharide	1	0.9644379887173694
negative regulation of dopamine secretion	1	0.9607901067183825
thymus development	1	0.9603432123963196
cellular response to lipoteichoic acid	1	0.9603432123963196
defense response to tumor cell	1	0.9603432123963196
outer ear morphogenesis	1	0.9603432123963196

cardiac neural crest cell development involved in heart development	1	0.9603432123963196
response to epidermal growth factor	1	0.9603432123963196
positive regulation of CREB transcription factor activity	1	0.9603432123963196
positive regulation of transcription by RNA polymerase II	1	0.9587668440995097
negative regulation of inflammatory response to antigenic stimulus	1	0.9584884451427859
neurotransmitter receptor localization to postsynaptic specialization membrane	1	0.9581723268167008
signal complex assembly	1	0.9575585953490878
growth hormone receptor signaling pathway	1	0.9575585953490878
growth hormone receptor signaling pathway via JAK-STAT	1	0.9575585953490878
interleukin-6-mediated signaling pathway	1	0.9575585953490878
regulation of cell adhesion	1	0.9573684756423317
cellular senescence	1	0.9556265623149057

Table B.16: Word FrequenciesSynGO

B.12.3 Diseases

Table B.17: Molecular Function Frequencies and Weights

Disease	**Frequency**	**Weights**
schizophrenia	4	3.853880016349807
autistic disorder	3	2.9814671981022007
autism spectrum disorder	3	2.9814671981022007
intellectual disability	3	2.919036744692884
epilepsy syndrome	2	1.9501567814320135

Table B.17: Word FrequenciesSynGO

B.12.4 SynGO

Table B.18: Molecular Function Frequencies and Weights

Molecular Function	**Frequency**	**Weights**
integral component of postsynaptic density membrane (GO:0099061)	5	4.99735196836674
structural constituent of postsynaptic density (GO:0098919)	5	4.9957012947141655
integral component of presynaptic membrane (GO:0099056)	5	4.96447137600556

regulation of postsynaptic density assembly (GO:0099151)	5	4.946690559341381
integral component of postsynaptic membrane (GO:0099055)	5	4.916073092809541
postsynaptic density, intracellular component (GO:0099092)	4	3.9398819237754856
neurotransmitter receptor localization to postsynaptic specialization membrane (GO:0099645)	3	2.9402769358344814
integral component of postsynaptic specialization membrane (GO:0099060)	2	1.9547025529649726
regulation of presynapse assembly (GO:1905606)	2	1.9520344229125903
synapse adhesion between pre- and postsynapse (GO:0099560)	1	0.9999862496087608
presynapse assembly (GO:0099054)	1	0.9995535256221376
trans-synaptic signaling by trans-synaptic complex (GO:0099545)	1	0.9992742612748858
voltage-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1905030)	1	0.9893896708835846
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1904315)	1	0.9747635432506689
postsynaptic specialization assembly (GO:0098698)	1	0.9698997461037794
modulation of chemical synaptic transmission (GO:0050804)	1	0.9526825450771074
voltage-gated ion channel activity involved in regulation of presynaptic membrane potential (GO:0099508)	1	0.9507738803956713

Table B.18: Word FrequenciesSynGO

For Network Full

B.13 WordCloud Tables:Full

B.13.1 Molecular Function

Table B.19: Molecular Function Frequencies and Weights

Molecular Function	Frequency	Weights
inositol 1,4,5 trisphosphate binding	3	2.9999743877488
protein phosphatase 2A binding	3	2.952237049652479
store-operated calcium channel activity	3	2.9409854646195566
protein serine kinase activity	2	2.0
protein serine/threonine kinase activity	2	2.0
protein kinase activity	2	1.9999999998772122
ATP binding	2	1.9999999992643018
MAP kinase kinase activity	2	1.9999999982942978
MAP kinase activity	2	1.9999999701407893
MAP kinase kinase kinase activity	2	1.9999941994550985
DNA-binding transcription factor binding	2	1.9999928446474098
histone deacetylase binding	2	1.999931760095324
transcription coactivator binding	2	1.9994186779217482
14-3-3 protein binding	2	1.9984130207758166
magnesium ion binding	2	1.9968277270854748
armadillo repeat domain binding	2	1.9967250449772869
protein kinase activator activity	2	1.9959719495815358
nuclear receptor binding	2	1.9951951958970955
protein kinase binding	2	1.995112473999695
tau-protein kinase activity	2	1.992041217285229
phosphoserine residue binding	2	1.9916590248825983
histone acetyltransferase activity	2	1.9896535116285254
RNA polymerase II cis-regulatory region sequence-specific DNA binding	2	1.9880506961023277
protein serine/threonine/tyrosine kinase activity	2	1.9873170312172923
transcription coactivator activity	2	1.9809676689788343
JUN kinase kinase kinase activity	2	1.9754009892255182
tau protein binding	2	1.9751623663361644
DNA-binding transcription factor activity	2	1.972828432793915
MAP-kinase scaffold activity	2	1.970907047880052
protein tyrosine kinase activity	2	1.9608066694069937
inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity	2	1.9571311291436857
peptide N-acetyltransferase activity	2	1.9466960926792194
RNA polymerase II-specific DNA-binding transcription factor binding	2	1.9358469556646059

DNA-binding transcription activator activity, RNA polymerase II-specific	2	1.9276107551041086
glutamate-gated calcium ion channel activity	1	0.9999978515960857
delayed rectifier potassium channel activity	1	0.9999978515960857
transmembrane receptor protein tyrosine phosphatase activity	1	0.9999939057387196
chromatin DNA binding	1	0.9999833665528574
chromatin binding	1	0.9999788404651132
NMDA glutamate receptor activity	1	0.9999774738792073
protein tyrosine phosphatase activity	1	0.999959769546555
transcription cis-regulatory region binding	1	0.9999382823711601
cuprous ion binding	1	0.9999254394822628
glycogen phosphorylase activity	1	0.9997090739836011
linear malto-oligosaccharide phosphorylase activity	1	0.9997090739836011
SHG alpha-glucan phosphorylase activity	1	0.9997090739836011
enzyme binding	1	0.999585508284496
transcription corepressor activity	1	0.9993836152712492
scaffold protein binding	1	0.9972523304784126
outward rectifier potassium channel activity	1	0.9969369090932676
RNA polymerase II transcription regulatory region sequence-specific DNA binding	1	0.9931433383110195
phosphoprotein phosphatase activity	1	0.992747679025817
histone H3K9me/H3K9me2 demethylase activity	1	0.9918845262210577
JUN kinase activity	1	0.9887372293299977
peptide-lysine-N-acetyltransferase activity	1	0.9787878262698564
NF-kappaB binding	1	0.9779727332467696
DNA binding	1	0.9779727332467696
DNA-binding transcription factor activity, RNA polymerase II-specific	1	0.9755503532465155
protein sequestering activity	1	0.9755503532465155
histone H3K9 demethylase activity	1	0.9662053574851347
PDZ domain binding	1	0.9658606972033984
phosphatidylserine binding	1	0.964908834754507
clathrin binding	1	0.964908834754507
calcium-dependent phospholipid binding	1	0.964908834754507
SNARE binding	1	0.9623022583325918
glutamate binding	1	0.9606579272912598
guanylate kinase activity	1	0.9601946598680013
phosphatidylinositol-4,5-bisphosphate binding	1	0.9584756256605445
nuclear receptor coactivator activity	1	0.9576238346223968

protein lysine deacetylase activity	1	0.9568583586303755
nuclear glucocorticoid receptor binding	1	0.9568583586303755
nuclear retinoid X receptor binding	1	0.9568583586303755
histone acetyltransferase binding	1	0.9568583586303755
kinase activity	1	0.9567171042550149
histone H2B acetyltransferase activity	1	0.9548115664581618
JUN kinase kinase activity	1	0.9548115664581618
sequence-specific double-stranded DNA binding	1	0.9505367553240486

Table B.19: Word FrequenciesFull

B.13.2 Biological Function

Table B.20: Biological Function Frequencies and Weights

Molecular Function	**Frequency**	**Weights**
positive regulation of transcription by RNA polymerase II	3	2.996252307524542
interleukin-1-mediated signaling pathway	3	2.960717883095247
MAPK cascade	2	1.999999999999624
protein phosphorylation	2	1.9999999999956677
stress-activated MAPK cascade	2	1.9999999935760702
p38MAPK cascade	2	1.9999951954373705
intracellular signal transduction	2	1.9999887600483603
peptidyl-threonine phosphorylation	2	1.9999778879270922
peptidyl-serine phosphorylation	2	1.999975102613073
face development	2	1.9999749268218385
JNK cascade	2	1.9999694696854404
signal transduction	2	1.9999448389548666
cellular senescence	2	1.9999356233026448
cellular response to tumor necrosis factor	2	1.9999356233026448
thyroid gland development	2	1.9999221479379634
cellular response to cadmium ion	2	1.9999043980634297
insulin-like growth factor receptor signaling pathway	2	1.999862873728445
positive regulation of JUN kinase activity	2	1.9998609124712836
anoikis	2	1.9996944651250579
negative regulation of transcription by RNA polymerase II	2	1.999624286340287
cellular response to reactive oxygen species	2	1.9993185552322834
regulation of Golgi inheritance	2	1.9988818801725607
phosphatidylinositol 3-kinase/protein kinase B signal transduction	2	1.998513216987953

phosphorylation	2	1.998337637670041
ERK1 and ERK2 cascade	2	1.997362863259054
trachea formation	2	1.9959596426040962
inflammatory response	2	1.9957881866912666
Fc-epsilon receptor signaling pathway	2	1.995457993535789
heart development	2	1.9934164522535076
negative regulation of miRNA transcription	2	1.993247563835184
positive regulation of JNK cascade	2	1.9928991257290487
cellular response to vascular endothelial growth factor stimulus	2	1.9916853903734988
apoptotic process	2	1.991576148239844
positive regulation of gene expression	2	1.9901243463343443
non-canonical NF-kappaB signal transduction	2	1.9893670376504207
cellular response to lipopolysaccharide	2	1.9872501986781503
T cell receptor signaling pathway	2	1.9869448554686264
thymus development	2	1.9867857680379875
cellular response to amino acid starvation	2	1.9865940958572084
receptor clustering	2	1.9863519057768058
toll-like receptor 4 signaling pathway	2	1.985456337794778
endothelial cell apoptotic process	2	1.9847786908199199
positive regulation of cyclase activity	2	1.9847786908199199
cellular response to UV-B	2	1.9847786908199199
cellular response to nicotine	2	1.9847786908199199
cellular response to sorbitol	2	1.9847786908199199
protein autophosphorylation	2	1.9847521568547069
lipopolysaccharide-mediated signaling pathway	2	1.9846087017236953
positive regulation of telomere capping	2	1.9828787472425604
regulation of stress-activated MAPK cascade	2	1.9828787472425604
regulation of early endosome to late endosome transport	2	1.9828787472425604
embryo development	2	1.9773317694494106
positive regulation of peptidyl-serine phosphorylation	2	1.9743559429094524
cellular response to interleukin-1	2	1.9725159766530367
negative regulation of TORC1 signaling	2	1.970913434815915
stimulatory C-type lectin receptor signaling pathway	2	1.9694336296154238
canonical NF-kappaB signal transduction	2	1.9685139323311138
regulation of ossification	2	1.9572541787018323
response to muscle stretch	2	1.9567428911774347

positive regulation of DNA-templated transcription	2	1.941950904571569
cellular response to lipoteichoic acid	2	1.9410718492565857
photoreceptor cell development	2	1.9368528927276882
activation of protein kinase C activity	2	1.9253576808556216
CD40 signaling pathway	2	1.9253576808556216
RIG-I signaling pathway	2	1.9253576808556216
positive regulation of excitatory postsynaptic potential	1	0.9999999984202085
positive regulation of synaptic transmission, glutamatergic	1	0.9999999966099253
synaptic membrane adhesion	1	0.9999983717423522
calcium ion transmembrane import into cytosol	1	0.9999845518722282
chemical synaptic transmission	1	0.999966711539099
regulation of monoatomic cation transmembrane transport	1	0.999964438549696
transmembrane receptor protein tyrosine phosphatase signaling pathway	1	0.9999385279610543
excitatory chemical synaptic transmission	1	0.999936057802988
regulation of monoatomic ion transmembrane transport	1	0.999936057802988
receptor localization to synapse	1	0.999936057802988
potassium ion transport	1	0.9997843509469279
establishment or maintenance of epithelial cell apical/basal polarity	1	0.9997791883467259
protein localization to synapse	1	0.9996579181202493
trans-synaptic signaling by trans-synaptic complex	1	0.9994871805318162
potassium ion transmembrane transport	1	0.9992324432750469
circadian regulation of gene expression	1	0.998732246490435
monoatomic cation transmembrane transport	1	0.9981742711035811
chromatin remodeling	1	0.9972196404995264
positive regulation of protein serine/threonine kinase activity	1	0.9969061719670768
locomotor rhythm	1	0.9967487785060807
regulation of neuronal synaptic plasticity	1	0.996566944176069
negative regulation of potassium ion transmembrane transporter activity	1	0.9964987853291034
positive regulation of MAPK cascade	1	0.9955263791024378
regulation of presynaptic membrane potential	1	0.9922254907385226
cellular response to angiotensin	1	0.9921083674921551

AMPA glutamate receptor clustering	1	0.9915098971745135
cell-cell adhesion	1	0.9914427053143852
ionotropic glutamate receptor signaling pathway	1	0.9914427053143852
negative regulation of DNA-templated transcription	1	0.9912775868672083
regulation of presynapse assembly	1	0.9906315263938753
positive regulation of DNA-binding transcription factor activity	1	0.9904997277325658
positive regulation of interleukin-12 production	1	0.9898462393815103
negative regulation of pancreatic juice secretion	1	0.9876626713446655
intrinsic apoptotic signaling pathway in response to DNA damage	1	0.9871476274187142
extrinsic apoptotic signaling pathway via death domain receptors	1	0.9871476274187142
positive regulation of autophagy	1	0.9856200770842948
neuron differentiation	1	0.9855182383307216
regulation of calcium ion-dependent exocytosis	1	0.9848861954742263
insulin receptor signaling pathway	1	0.9848391111130949
regulation of cell cycle	1	0.9848391111130949
brain development	1	0.9843946096360923
calcium-mediated signaling	1	0.9843946096360923
positive regulation of dendritic spine development	1	0.9823371871985437
positive regulation of peptidyl-threonine phosphorylation	1	0.981956335788031
positive regulation of protein phosphorylation	1	0.9819052736350401
stress-activated protein kinase signaling cascade	1	0.9819052736350401
peptidyl-serine autophosphorylation	1	0.9819052736350401
dopamine biosynthetic process	1	0.9819052736350401
regulation of dopamine secretion	1	0.9808595376167839
N-terminal peptidyl-lysine acetylation	1	0.9797127921686298
nucleotide-binding oligomerization domain containing 2 signaling pathway	1	0.9797127921686298
cellular response to peptidoglycan	1	0.9797127921686298
TRIF-dependent toll-like receptor signaling pathway	1	0.9797127921686298
positive regulation of interleukin-1 beta production	1	0.9797127921686298

visual learning	1	0.9793117168304543
excitatory postsynaptic potential	1	0.9793117168304543
protein dephosphorylation	1	0.9789768921638557
regulation of postsynaptic density assembly	1	0.9789768921638557
positive regulation of canonical NF-kappaB signal transduction	1	0.9783740837614656
regulation of transcription by RNA polymerase II	1	0.9783674772495783
positive regulation of synapse assembly	1	0.9781137049803116
regulation of synaptic plasticity	1	0.9762369901831707
positive regulation of NLRP3 inflammasome complex assembly	1	0.9753999571777823
negative regulation of potassium ion transmembrane transport	1	0.9753999571777823
cellular response to insulin stimulus	1	0.9734765772489786
regulation of DNA-templated transcription	1	0.9725582794047883
cellular response to potassium ion	1	0.9723962697672238
positive regulation of apoptotic process	1	0.9704066862045364
negative regulation of apoptotic process	1	0.968524632115749
protein homooligomerization	1	0.9679277142236722
regulation of long-term neuronal synaptic plasticity	1	0.9679277142236722
glutamate receptor signaling pathway	1	0.9679277142236722
dendritic spine morphogenesis	1	0.9679277142236722
type B pancreatic cell proliferation	1	0.9675552413729372
I-kappaB phosphorylation	1	0.9675552413729372
response to osmotic stress	1	0.9675552413729372
regulation of cytoskeleton organization	1	0.9665421333199742
positive regulation of T cell chemotaxis	1	0.9665421333199742
renal sodium ion absorption	1	0.9665421333199742
cellular hypotonic response	1	0.9665421333199742
negative regulation of cell population proliferation	1	0.9659577782106491
learning or memory	1	0.965579551946778
positive regulation of actin filament bundle assembly	1	0.9635545520944938
positive regulation of p38MAPK cascade	1	0.9635545520944938
negative regulation of epithelial to mesenchymal transition	1	0.9635545520944938
tumor necrosis factor-mediated signaling pathway	1	0.9621705400482518
neurotransmitter receptor localization to postsynaptic specialization membrane	1	0.9597456159894631

positive regulation of ubiquitin-dependent protein catabolic process	1	0.9595516946502582
response to activity	1	0.9586425642374123
regulation of axonogenesis	1	0.9586425642374123
GDP metabolic process	1	0.9581068384870522
postsynaptic membrane assembly	1	0.9572267302198184
positive regulation of telomerase activity	1	0.9561754444580816
GMP metabolic process	1	0.9530891270585197

Table B.20: Word FrequenciesFull

B.13.3 Diseases

Disease	Frequency	Weights
hypertension	5	4.90881161733825
intellectual disability	4	3.9254644697989813
schizophrenia	4	3.898976169825788
Alzheimer's disease	2	1.999195432202585
Huntington's disease	2	1.9980415735817225
autistic disorder	2	1.9638122343847804
autism spectrum disorder	2	1.9638122343847804
epilepsy syndrome	1	0.9919705015400118
bipolar disorder	1	0.9914019624780898
multiple sclerosis	1	0.9565867434428734

Table B.21: Word Frequencies Full

B.13.4 SynGO

Table B.22: SynGO Frequencies and Weights

Molecular Function	Frequency	Weights
extrinsic component of presynaptic active zone membrane (GO:0098891)	3	2.988663119938335
integral component of postsynaptic density membrane (GO:0099061)	2	1.9916386850179542

structural constituent of postsynaptic density (GO:0098919)	2	1.9907108406325322
neurotransmitter receptor localization to postsynaptic specialization membrane (GO:0099645)	2	1.98899150223311
modulation of chemical synaptic transmission (GO:0050804)	2	1.9720200755560187
integral component of postsynaptic membrane (GO:0099055)	2	1.9567282164847517
presynapse assembly (GO:0099054)	1	0.9999999895356283
synapse adhesion between pre- and postsynapse (GO:0099560)	1	0.9999998398695149
integral component of presynaptic membrane (GO:0099056)	1	0.999999551566868
integral component of synaptic vesicle membrane (GO:0030285)	1	0.9999976129634156
postsynaptic density, intracellular component (GO:0099092)	1	0.999991951098068
trans-synaptic signaling by trans-synaptic complex (GO:0099545)	1	0.9999848072565288
regulation of postsynaptic density assembly (GO:0099151)	1	0.99997068829
regulation of presynapse assembly (GO:1905606)	1	0.9998216310983132
transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1904315)	1	0.9989210350484116
postsynaptic density (GO:0014069)	1	0.9925583455477984
ligand-gated ion channel activity involved in regulation of presynaptic membrane potential (GO:0099507)	1	0.9925583455477984
trans-synaptic signaling (GO:0099537)	1	0.9918886406481338
voltage-gated ion channel activity involved in regulation of presynaptic membrane potential (GO:0099508)	1	0.9918782105282646
regulation of postsynaptic neurotransmitter receptor activity (GO:0098962)	1	0.9857346933514674
voltage-gated ion channel activity involved in regulation of postsynaptic membrane potential (GO:1905030)	1	0.984369569459572
integral component of presynaptic active zone membrane (GO:0099059)	1	0.9768283395686872
synaptic vesicle proton loading (GO:0097401)	1	0.9650032655183441

regulation of synaptic membrane adhesion (GO:0099179)	1	0.9586049677756836
anchored component of postsynaptic density membrane (GO:0099031)	1	0.9586049677756836
regulation of postsynapse organization (GO:0099175)	1	0.9521752389537201
intracellular cAMP-activated cation channel activity involved in regulation of presynaptic membrane potential (GO:0140232)	1	0.9512698740450212
extrinsic component of synaptic vesicle membrane (GO:0098850)	1	0.9503502909765661

Table B.22: Word FrequenciesFull