



Universiteit
Leiden
The Netherlands

Opleiding Informatica

Finding consensus knowledge in the
Huntington's Disease pathway

Chen Ji Rong Jiang (2636387)

Supervisors:

Katy Wolstencroft & Lu Cao

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)

www.liacs.leidenuniv.nl

19/07/2022

Abstract

Huntington's Disease (HD) is an inherited neurodegenerative disease caused by an expansion of the CAG repeat in the Huntingtin (HTT) gene. Even though it is caused by a single mutation, the underlying mechanism of HD is very complex. Many pathways have been suggested to contribute to the pathogenesis of HD, but a consensus has yet to be established. Protein-Protein Interaction Network (PPIN) tools allow for visualization and analysis of these pathways and provide a great way to consolidate existing information on the HD pathway. By using functional enrichment analysis on these networks the role of HTT can be reaffirmed, which will further the understanding of the affected cellular processes and the role of HTT in different pathways. In this research, knowledge from different databases have been compared in order to find consensus and discrepancies between them and functionally enriched terms within these parts were connected to existing scientific research to give a better image of the current knowledge surrounding the HD pathway.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Huntington's Disease | 1 |
| 1.2 | Protein-Protein Interaction Networks | 1 |
| 1.3 | Related Work | 1 |
| 1.4 | Research | 2 |
| 1.5 | Thesis overview | 2 |
| 2 | Definitions | 2 |
| 3 | Methods | 3 |
| 3.1 | Data gathering | 3 |
| 3.2 | Network creation | 4 |
| 3.3 | Network merging | 5 |
| 3.4 | Network analysis | 6 |
| 3.5 | Clustering | 8 |
| 3.6 | Functional enrichment | 9 |
| 4 | Results | 10 |
| 4.1 | Network creation | 10 |
| 4.2 | Network analysis | 10 |
| 4.3 | Clustering | 12 |
| 4.4 | Functional enrichment of consensus network | 13 |
| 4.4.1 | Cellular response | 17 |
| 4.4.2 | Regulation of cell communication | 17 |
| 4.4.3 | Behavior, movement and memory | 18 |
| 4.4.4 | Oxidative phosphorylation | 18 |
| 4.5 | Functional enrichment of discrepancy clusters | 18 |
| 4.5.1 | 7-methylguanosine mRNA capping | 18 |
| 4.5.2 | Nuclear migration along microtubules | 21 |
| 5 | Conclusions | 21 |
| 6 | Further Research | 21 |
| 7 | Acknowledgements | 22 |
| | References | 26 |

1 Introduction

1.1 Huntington's Disease

Huntington's Disease (HD) is an inherited neuro-degenerative disorder. It is caused by a trinucleotide (*CAG*) -also known as *PolyQ*- repeat expansion in exon 1 of the Huntingtin (*HTT*) gene. The healthy gene contains 17-20 repeats, while in people with HD the sequence is repeated > 35 times. The age of onset averages around 40 years old and patients have an average life expectancy of 20 years after onset. However, cases have been found with the age of onset ranging from infancy to senescence. [vDvdVRB86] The variance is partially due to the length of the repeat, which has an inverse association with the age of onset. [WBC⁺20] The symptoms of the disorder are characterized primarily by involuntary movements, psychiatric problems and cognitive decline. [Roo10, A.10] There is no cure for HD yet and its treatment remains symptomatic.

Even though the HD is caused by a single mutation, the development of the disorder could involve a variety of genes and involve many pathways. Studies have suggested that healthy state *HTT* plays a role in multiple processes including chemical signaling, transcription, autophagy and protection from apoptosis. However, the exact function of the protein is still unknown. [HW03, SL11, MLEH15, VBS20] Similarly, the underlying mechanism that leads to HD is not well understood either. Many studies have suggested different pathways that contribute to HD. [LL06, MNR18, Ros04] Yet no consensus has been made on the pathological pathway. In order to gain a better understanding of HD, the interactions between the affected proteins need to be studied. These protein-protein interactions (PPIs) play a central role in cellular functions and biological processes. By analysing PPIs, a lot of insight could be gained of the healthy state *HTT* functionality and the complex pathological mechanism.

1.2 Protein-Protein Interaction Networks

Protein-Protein Interaction Networks (PPINs) are mathematical representations of the interactions between proteins. In these networks proteins are represented as nodes and the interactions between the proteins as edges. PPINs have a range of applications and play an important role in the understanding of biological systems. These application include the development of precision therapeutics and medicine. Drugs for example have a wide range of effects on different entities within a biological system. The use of PPINs allows for the analysis of the underlying molecular relationships and the network topology within such a system, revealing important entities. [SAB⁺17] shows the application of networks in analysing the action of inhibitory compounds and identifying candidate inhibitors in drug discovery. Another application is using PPIN in the identification of physical structures in biological systems such as protein complexes and signaling pathways. In [DMBM17] such direct contacts between protein complex sub-units were identified using networks.

1.3 Related Work

In [Jan20], a network-based method was used in order to gain a better understanding of polyQ diseases in brain tissue and the workings of regular polyQ proteins. This was done by comparing the differences and commonalities between inter-actors in disease and normal variants. Due to the

lack of information on interactions in the brain in human databases, the data was limited to mouse models. In a similar fashion [Kos21] used functional enrichment analysis to gain information on the functions of polyQ proteins in the human brain. In contrast to the [Jan20], it used human expression data. In this research, a similar approach is used to compare the differences and similarities of HD knowledge in different databases.

1.4 Research

In recent years a lot of information on the HD pathway has accumulated, but it also created a dissemination of pathway information over different databases such as KEGG, Reactome and Wiki pathways. [KG00, COW⁺11, PKvI⁺08] The aim of this thesis is to find consensus and identify differences among these pathway databases and the protein interaction database, STRING. [SGL⁺19] PPINs will be made to visualize the overlapping components and allow for analysis of the processes within the HD pathway. These goals can be summarized in the following research statement (RS) and research question (RQ):

RS: This research focuses primarily on consolidating the existing information surrounding Huntington’s Disease in a protein-protein interaction network to create a consensus of relevant data from different sources such as KEGG, Reactome, Wiki pathways and their annotations.

RQ: Does consolidating knowledge on interactions with the huntingtin protein provide new insights into the HD mechanism? And if so, what are its implications?

1.5 Thesis overview

This chapter contains background information surrounding the topics within this bachelor thesis; Section 2 includes the definitions; Section 3 provides the methodology used; Section 4 describes the results; Section 5 contains the conclusions. This bachelor thesis is written for the bachelor Bioinformatics at LIACS and was supervised by Katy Wolstencroft and Lu Cao.

2 Definitions

The PPINs generated in this research are results of merging, filtering and clustering former networks. In order to distinguish these networks from each other, the following naming conventions will be used:

- Merged networks will start their name with the abbreviations of databases in the network separated by a plus sign. So a network generated by merging the PPINs of STRING (STR), KEGG and Wikipathways (WP) will be called *STR+KEGG+WP*.
- Networks were filtered on a STRING cut-off score. This score can have the value of 0.4 and 0.7 and will be indicated by the score next to the STRING abbreviation. So continuing on the previous example it would result in *STR0.4+KEGG+WP* or *STR0.7+KEGG+WP*.
- Networks were also filtered on a STRING nervous system tissue confidence score range. This range can be between 0 and 5, [0, 5] and between 3 and 5, [3, 5]. This value will be added

after the database names. Previous example network would be called *STR0.4+KEGG+WP[0, 5]* or *STR0.4+KEGG+WP[3, 5]*.

- Clusters from networks will have the same name as the network and have "Cl" added in front and their number in between parentheses. The eighth cluster of the example network would be named *Cl(8)STR0.4+KEGG+WP[0, 5]* or *Cl(8)STR0.4+KEGG+WP[3, 5]*

3 Methods

3.1 Data gathering

In order to find consensus in the HD pathway, the current knowledge on the contributing pathways need to be mapped out. For this the following three biological pathway databases were used to obtain HD related pathways:

- KEGG: Also known as the Kyoto Encyclopedia of Genes and Genomes, KEGG is a database collection that focuses on genomes, biological pathways, diseases, drugs, and chemical substances. It contains the KEGG Pathway database, which contains manually drawn pathway maps that represent the experimental knowledge on cellular processes such as metabolism and membrane transport. [KG00]
- Reactome: Reactome is a manually curated, peer-reviewed pathway database, specifically of human pathways and processes. Their Pathways Browser contains pathways aligned with molecular interaction data from several interaction databases including the Reactome Functional Interaction Network, IntAct, BioGRID, ChEMBL and MINT. [COW⁺11, dTSR⁺21, SBR⁺06, MGB⁺18, CaCP⁺07]
- WikiPathways: In contrast to the curated pathway databases of KEGG and Reactome, WikiPathways is a community resource for contributing and maintaining biological pathways. This means that the responsibility of peer reviewing, editorial curation and maintenance lies with the user community. Contributions, however, are monitored by a group of administrators. [PKvI⁺08]

The pathways from these databases were formed by integrating data curated by several primary databases, meaning that the data is experimental and exclusively from peer-reviewed scientific publications. This ensures the quality of the resulting pathways. The pathways of these databases will be compared to the information in STRING. [SGL⁺19]

- STRING: STRING is an interaction database. In contrast to the aforementioned pathway databases, STRING contains, in addition to experimental data, experimentally inferred data and computational predictions of molecular interactions from various PPI databases such as MINT, BioGRID, KEGG, Reactome, IntAct and NCI-Nature Pathway Interaction Database. [CaCP⁺07, SBR⁺06, dTSR⁺21, KAB⁺07]

The data from STRING will be used as a knowledge space of the information known about the HD pathway and the data from the pathway databases will be able to form consensus or discrepancies with this knowledge space. For the pathway databases, a query search was performed for the

following terms: "Huntington's Disease", "HD", "Huntington", "Huntingtin" and "HTT" in order to find HD related pathways. For the query searches, an organism filter was used by filling in the organism prefix "hsa", referring to Homo sapiens, for KEGG and selecting the species "Homo sapiens" for Reactome and WikiPathways. The query result were then manually classified as relevant or irrelevant. Some nodes in the KEGG pathway were representing a complex of proteins. These proteins were manually divided into there own node as shown in Figure 1.

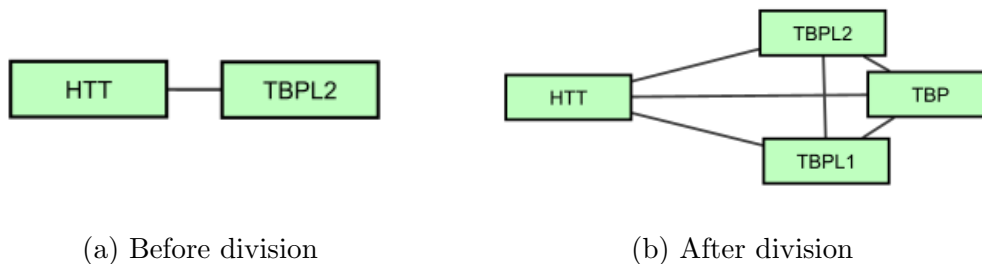


Figure 1: KEGG protein complex division into multiple nodes

3.2 Network creation

Creating networks from the interactions found in biology, allow us to abstract the relationship between genes, proteins, drugs and diseases and discover important associations between them. Together with high-throughput techniques, PPINs can reveal important proteins in complex biological systems. Which will further our understanding of the mechanism in such system and its components.

The discovered HD related pathways were converted into such PPINs. As mentioned before, the networks exist of nodes that represent the proteins that were found in the corresponding database of the network. The interactions between these proteins are represented by the edges between the nodes. Furthermore, the PINNs are directed, unweighted and can be annotated by a relation such as activation and inhibition. The network conversion was done using Cytoscape. [SMO⁺03] Cytoscape is a platform for visualizing molecular interaction networks and biological pathways. These networks can be further integrated with annotations, gene expression profiles and other data. Cytoscape provides a variety of features that allow for smooth data integration, analysis and visualization. Additional features called *Apps* can be installed to support and expand the functionality of Cytoscape. This research uses the version 3.9.1. The Cytoscape Apps, CyKEGGParser (version 1.2.9) and WikiPathways (version 3.3.10), were used to convert the pathways chosen from KEGG and WikiPathways into PPIN. [NSA14, KLEP14] For KEGG, the pathway was downloaded as a KEGG Markup Language (KGML) file and loaded into Cytoscape. For WikiPathways, the pathway could be found within the WikiPathways App by querying "Huntington's Disease" and filtering the species to "Homo sapiens". At last the App, StringApp (version 1.7.1), was used to generate the HD related PPIN of the STRING database. This was done by querying "Huntington's Disease" in the STRING disease search bar in Cytoscape. For this, the Network type parameter was set to *full STRING network*, the *Maximum numbers of proteins* parameter was set to 2000 and the *Load Enrichment Data* parameter was turned on. At last, two confidence cut-off scores were

selected, creating a lenient network with the cut-off at 0.4 and a strict network with the cut-off at 0.7. This cut-off score ranges from 0 to 1 and represents the confidence of an interaction or "how likely STRING judges an interaction to be true, given the available evidence". [SGL⁺19] As example, a score of 0.5 would mean that approximately every second interaction might be erroneous.

3.3 Network merging

The proteins of the networks contain identifiers from their respective database. In order to merge the networks and find consensus, the identifiers have to be mapped to each other. For this reason, all identifiers were mapped to Uniprot identifiers. For the KEGG identifier mapping, a Python (version 3.9.0) script was used including the packages Bio.KEGG.REST from Biopython (version 1.76) and Pandas (version 1.4.3). [CAC⁺09, pdt22, WM10] The code and results can be found at: <https://git.liacs.nl/s2636387/protein-interaction-networks-in-huntington-s-disease>. The script (see listing 3.3) uses the KEGG API to generate the list of proteins in the HD pathway and converts these to Uniprot identifiers. Due to the low amount of WikiPathways proteins, the ID mapper on the Uniprot site was used to convert these identifiers. [Con20] Lastly, the STRING proteins already contained Uniprot identifiers in the "stringdb::canonical name" column, so conversion was not necessary. The original identifiers and their respective Uniprot identifiers were saved in a table. These tables were then imported to Cytoscape and joined with the following parameters:

- **Import Data as:** Node Table Columns
- **Key Column for Network:** shared name
- **Case Sensitive Key Values:** True
- **Key Column for Table:** KEGG/WikiPathway identifiers

In addition to the Uniprot identifiers, a *Database* attribute was added to the node tables in order to keep track of which databases contain which proteins. The merging of networks was done by using the Cytoscape merge tool, selecting the networks, opting for "Union" and selecting the Uniprot identifiers as the matching attribute to join on. Additional networks were made by only including proteins up to the second neighbor of HTT. Lastly, the networks were filtered on the *tissue::nervous system* column with values in the ranges [0, 5] and [3, 5]. Note that only STRING proteins were filtered, as the proteins from pathway databases already have a high likelihood of being located in the nervous system. Figure 2 shows the workflow of this process. Note that the Reactome database did not contain any HD related pathway, so no network was be made.

```
1 from Bio.KEGG import REST
2 import pandas as pd
3
4
5 #getting the KEGG IDs of all proteins in the HD pathway
6 def relevant_protein(pathway):
7     request = REST.kegg_get(pathway)
8     full = request.read().split("GENE")[1].split("\nCOMPOUND")[0].split('\n')
9     kegg_id = []
10    for line in full:
11        kegg_id.append(line.strip().split(" ")[0])
```



```

12     return kegg_id
13
14 #converting the KEGG IDs to Uniprot IDs
15 def convert(queries):
16     data = {'kegg'      : [],
17            'uniprot'   : []}
18
19     for query in queries:
20         request = REST.kegg_conv("uniprot", f'hsa:{query}').read()
21         id = request.split("\t")
22
23         kegg = id[0]
24         up = id[1].replace('up:', '').split('\n')[0]
25
26         data['kegg'].append(kegg)
27         data['uniprot'].append(up)
28     return data
29
30 kegg_ids = relevant_protein('hsa05016')
31
32 uniprot_ids = convert(kegg_ids)
33 df = pd.DataFrame(data=uniprot_ids)
34 df.to_excel('kegg_data_hsa.xlsx')

```

Listing 1: Identifier mapping of KEGG protein identifiers to Uniprot identifiers

3.4 Network analysis

As all network, PPINs have topological properties that can be analyzed. These properties allow for the aforementioned identification of important proteins in the network and inferences about the network as a whole. For example, nodes with a high degree, also known as *hubs*, indicate an important role in the network. Furthermore, a network with a higher density and number of average neighbors indicate a more connected network, usually meaning more related to each other. Like this, these network properties provide valuable information about the networks and the protein in them.

By using the *Analyze* feature of Cytoscape, the topological properties of the networks can be quantified. This information provides a better understanding of the networks and their behavior. The feature gives the following information:

- Number of nodes in the network.
- Number of edges in the network.
- Average number of neighbors a node has.
- Network diameter: the shortest distance between the two most distant nodes in a network.
- Network radius: The minimum graph eccentricity of all the nodes in a graph. The graph eccentricity of a single node is the maximum distance of the node to another node. The smallest graph eccentricity is the network radius.

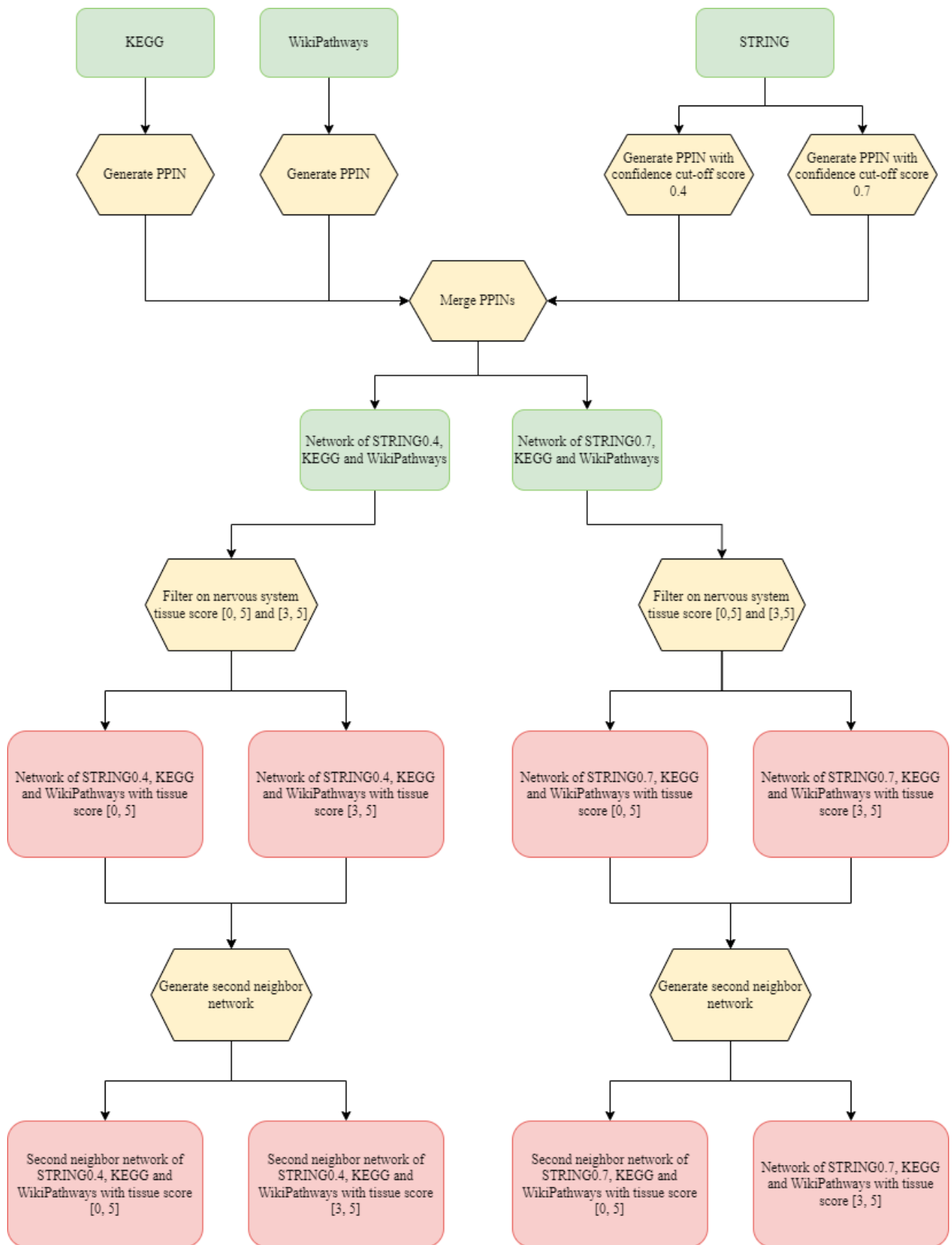


Figure 2: The workflow of how the networks were merged and what filters were used. Green indicates non-final networks, yellow represents actions and red indicates final networks.

- Characteristic path length: The Average shortest path between two nodes.
- Clustering coefficient: The Average of all clustering coefficient of each node. The clustering coefficient of a single node, n , refers to the ratio between the number of edges between the neighbors of n and the the maximum possible number of edges between these neighbors.
- Network density: The proportion of possible interactions between nodes that are actually present.
- Network heterogeneity: Network heterogeneity quantifies the network’s tendency to contain hub nodes. Hub nodes are nodes with a large degree.
- Network centralisation: A measure of the extent to which a network’s connection is concentrated on a single node or group of nodes.
- Connected components: A connected component is a set of nodes that are connected. With other words, there is a path between each pair of nodes in the set.

In addition to these features, the proportion of database consensus and discrepancy were measured.

3.5 Clustering

Clusters refer to highly connected sets of nodes in the network. By clustering the networks in this research, interconnected proteins can be identified. These proteins can indicate protein complexes and functional modules within the network, which can be functionally enriched and analysed to reveal involvement in biological processes and functionalities. The cluster of interest are clusters containing exclusively pathway database proteins. As these represent discrepancies between the databases. The Cytoscape App, MCODE (version 2.0.2)[BH03], was used for the clustering with the default parameter settings as shown in Table 1 on the eight networks resulting from the network merging.

| Parameter | Value |
|--------------------|------------------|
| Find Clusters | In Whole Network |
| Include Loops | False |
| Degree cut-off | 2 |
| Haircut | True |
| Fluff | False |
| Node Score cut-off | 0.2 |
| K-Core | 2 |
| Max. Depth | 100 |

Table 1: MCODE parameter settings

3.6 Functional enrichment

In functional enrichment, the proteins within the PPIN are annotated with information about functions and biological processes that they are involved in. This then allows for functional enrichment analysis, where over-represented classes of proteins are identified. The enriched terms can be plotted in a tree map. Tree maps are used to visualize a large amount hierarchical data using nested rectangles.

A functional enrichment analysis was performed in order to gain insights into the functions of the clusters and other networks of interests such as the consensus between databases. In Cytoscape, this was done by isolating the proteins of interest by clustering or filtering. Then setting the subnetwork as a STRING network and using the *Retrieve functional enrichment* feature of the STRING App with the default background genes, the whole genome. The result will be a table of enrichment categories from various resources. The categories of interest are *GO Biological Process* and *GO Molecular Function*. GeneOntology (GO) is a database containing information on the functions of genes and represents them with ontology terms for specific processes, functions or components. [ABB⁺00, Con21] These kind of databases are valuable resources for the functional characterization of the subnetworks. The over-expressed terms were identified by performing a hyper-geometric test with a p-value of 0.5. This value is the probability of randomly drawing the certain amount of enriched terms that was found in the network of a certain size. Each of the enrichment terms is accompanied by a *False Discovery Rate (FDR)*, which is a p-value adjusted for multiple tests. This value indicates the significance of the term in the network. A built in redundancy filter of STRING was used to filter out terms. This filter is based on the Jaccard Index, which is calculated by taking the intersection divided by the union of -in this case- two text strings. The threshold was set to the default value of 0.5. The resulting terms were visualized in tree maps using Revigo. [SB11] Table 2 shows the default parameter settings of the process that was used. Related terms in the tree maps were coloured. These colours represent overarching process or function 'themes'. These themes were then connected to existing scientific research to validate the network and confirm the contribution of these themes in the HD pathway.

| Parameter | Value |
|---|------------------------|
| Size of resulting list | Medium (0.7) |
| Values associated with GO terms represent | P-value |
| Remove obsolete GO terms | Yes |
| Species | Whole UniProt database |
| Similarity measure | SimRel |

Table 2: Revigo parameter settings

4 Results

4.1 Network creation

The pathways from the KEGG and WikiPathways were converted to PPINs, these networks are shown in Figure 9 and Figure 10. Similarly, the HD related interactions with a cut-off score of 0.4 and 0.7 from STRING were generated and are shown in Figure 11 and Figure 12. Merging and filtering these networks as described in the method section results in the networks $STR0.4+KEGG+WP[0, 5]$, $STR0.4+KEGG+WP[3, 5]$, $STR0.7+KEGG+WP[0, 5]$ and $STR0.7+KEGG+WP[3, 5]$. These networks are shown in Figures 13, 14, 15 and 16, respectively. Each protein is coloured to indicate in which database network they could be found. Table 3 shows the frequency of each of the databases and their combinations within the networks. Each combination of database values is coloured corresponding to the node colours in the networks.

In the first network ($STR0.4+KEGG+WP[0, 5]$), it is shown that of the 256 proteins found in the KEGG network, 103 overlap with the STRING network, 2 proteins overlap with the WikiPathways network and 4 proteins overlap with both, leaving 149 proteins only found in the KEGG network. Of the 16 proteins found in the WikiPathways network, 6 proteins overlap with STRING, again 2 proteins overlap with KEGG and 3 proteins overlap with both.

It also shows that filtering the networks on a more stringent nervous system tissue score, [3, 5], reduces the amount of STRING-only proteins by 207 proteins and the amount of overlap between STRING and KEGG by 7 proteins. Note that there is no difference in the consensus and discrepancies between the networks filtered on a STRING cut-off score of 0.4 and 0.7, because this score only judges the confidence of the interaction and not the proteins.

In Table 4, the database frequencies can be found of the second neighbor networks of the previous networks. It shows that for the network $STR0.4+KEGG+WP[0, 5]$, 20 STRING-only proteins, 51 KEGG-only proteins and 1 WikiPathways-only protein were not within the second neighborhood of the network. All the proteins that could be found in multiple databases remained in the network. Taking the second neighbor of $STR0.4+KEGG+WP[3, 5]$, the same KEGG- and WikiPathways-only proteins were removed. However, only 10 STRING-only proteins were removed. This means that of the 207 removed proteins, due to the stringent tissue score filter, 10 were also not within the second neighbour of the network. Continuing to the second neighbor network of $STR0.7+KEGG+WP[0, 5]$, it is shown that 483 STRING-only proteins were removed, due to the stringent interaction confidence cut-off score. Compared to the second neighbor networks of the more lenient cut-off score, an additional 5 KEGG-only proteins were removed from the network. Lastly, looking at the second neighbor network of $STR0.7+KEGG+WP[3, 5]$, 416 STRING-only proteins and 3 more KEGG-only proteins were removed.

4.2 Network analysis

Table 5 shows the results of the network analyses of the original merged networks. The effect of a more stringent STRING cut-off and tissue filter is shown in Table 7 and 8. It can be seen that the networks with the more stringent STRING confidence cut-off score have fewer edges compared to the networks with the more lenient cut-off score. Similarly, a decrease is seen for the average amount of neighbors, clustering coefficient, network density and network centralisation. Furthermore, an

Table 3: Network database values of the merged networks of STRING (STR), KEGG and Wikipathways (WP), filtered on nervous system tissue scores [0, 5] and [3, 5]

| | STR0.4+KEGG+WP [0, 5] | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------|-----------------------|-----------------------|-----------------------|-----------------------|
| String | 1755 | 1548 | 1755 | 1548 |
| KEGG | 149 | 149 | 149 | 149 |
| WP | 4 | 4 | 4 | 4 |
| String:KEGG | 103 | 96 | 103 | 96 |
| String:WP | 6 | 6 | 6 | 6 |
| KEGG:WP | 2 | 2 | 2 | 2 |
| String:KEGG:WP | 3 | 3 | 3 | 3 |

Table 4: Network database values of second neighbor networks of the merged networks of STRING (STR), KEGG and Wikipathways (WP), filtered on nervous system tissue scores [0, 5] and [3, 5]

| | STR0.4+KEGG+WP [0, 5] | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------|-----------------------|-----------------------|-----------------------|-----------------------|
| String | 1735 | 1537 | 1272 | 1132 |
| KEGG | 98 | 98 | 93 | 90 |
| WP | 3 | 3 | 3 | 3 |
| String:KEGG | 103 | 96 | 103 | 96 |
| String:WP | 6 | 6 | 6 | 6 |
| KEGG:WP | 2 | 2 | 2 | 2 |
| String:KEGG:WP | 3 | 3 | 3 | 3 |

increase is seen for the network diameter, network radius, characteristic path length, network heterogeneity and amount of connected components.

Just like an increase in stringency of the cut-off score, a more stringent nervous system tissue score results in fewer amount of edges, average amount of neighbors and network heterogeneity. In contrary to the cut-off score, a more stringent tissue score results in an decrease in the amounts of nodes, characteristic path length and connected components and a marginal increase in clustering coefficient, network density and network centralisation. The changes resulting from a more stringent cut-off threshold indicate a less dense and connected network. This is due to the interactions that are filtered out. The changes resulting from a more stringent tissue filter indicate a marginally denser network. The remaining proteins have a higher confidence score of being in the nervous system tissue, making it more likely to have interactions and be interconnected.

Table 6 shows the results of the network analyses of the second neighbor networks. These properties will be compared to the original networks. The changes result from this can be seen in Table 9. For all networks a decrease is seen in the amount of nodes, amount of edges, network diameter, network radius, characteristic path length and network heterogeneity and an increase in average amount of neighbors, network density and network centralisation. Note that the amount of connected components is always 1, because neighbors are always connected. For specifically the second neighbor networks of networks with a STRING cut-off score of 0.4, the clustering coefficient is decreased. While for the networks with a score of 0.7, it is increased. The results indicate that taking the second neighbor networks has similar results on all the merged networks. The networks are all more dense and connected. Networks with a STRING cut-off threshold of 0.7 seem to be impacted more. This could be due to the already low connectivity these networks had before. Interestingly, the networks with a lenient cut-off threshold seem to have a decrease in clustering coefficient. Which could indicate that highly connected components outside of the second neighbor were filtered out.

Table 5: Topological properties of the merged networks of STRING (STR), KEGG and Wikipathways (WP), filtered on nervous system tissue scores [0, 5] and [3, 5].

| | STR0.4+KEGG+WP [0, 5] | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| #nodes | 2022 | 1808 | 2022 | 1808 |
| #edges | 80567 | 69710 | 26549 | 22971 |
| Avg. #neighbors | 79.596 | 76.893 | 26.782 | 25.841 |
| Network diameter | 5 | 5 | 7 | 7 |
| Network radius | 3 | 3 | 4 | 4 |
| Characteristic path length | 2.342 | 2.327 | 3.020 | 3.017 |
| Clustering coefficient | 0.419 | 0.425 | 0.409 | 0.411 |
| Network density | 0.039 | 0.043 | 0.014 | 0.015 |
| Network heterogeneity | 0.960 | 0.943 | 1.069 | 1.048 |
| Network centralisation | 0.327 | 0.332 | 0.141 | 0.140 |
| Connected components | 5 | 2 | 55 | 44 |

Table 6: Topological properties of the second neighbor networks of the merged networks of STRING (STR), KEGG and Wikipathways (WP), filtered on nervous system tissue scores [0, 5] and [3, 5].

| | STR0.4+KEGG+WP [0, 5] | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| #nodes | 1951 | 1743 | 1483 | 1330 |
| #edges | 79993 | 69185 | 22844 | 19797 |
| Avg. #neighbors | 81.783 | 79.159 | 30.552 | 29.510 |
| Network diameter | 4 | 4 | 4 | 4 |
| Network radius | 2 | 2 | 2 | 2 |
| Characteristic path length | 2.294 | 2.275 | 2.739 | 2.730 |
| Clustering coefficient | 0.407 | 0.412 | 0.418 | 0.418 |
| Network density | 0.042 | 0.045 | 0.021 | 0.022 |
| Network heterogeneity | 0.936 | 0.918 | 0.969 | 0.947 |
| Network centralisation | 0.337 | 0.344 | 0.185 | 0.183 |
| Connected components | 1 | 1 | 1 | 1 |

Table 7: Network property changes due to the implementation of a stringent STRING cut-off filter of 0.7.

| Properties | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------------------|-----------------------|-----------------------|
| #nodes | 0 (+0.0%) | 0(+0.0%) |
| #edges | -54018(-67.0%) | -46739 (-67.0%) |
| Avg. #neighbors | -52.814(-66.4%) | -51.052(-66.4%) |
| Network diameter | 2(+40.0%) | 2(+40.0%) |
| Network radius | 1(+33.3%) | 1(+33.3%) |
| Characteristic path length | 0.678(+28.9%) | 0.69(+29.7%) |
| Clustering coefficient | -0.01(-2.4%) | -0.014(-3.3%) |
| Network density | -0.025(-64.1%) | -0.028(-65.1%) |
| Network heterogeneity | 0.109(+11.4%) | 0.105(+11.1%) |
| Network centralisation | -0.186(-56.9%) | -0.192(-57.8%) |
| Connected components | 50(+1000.0%) | 42(+2100%) |

4.3 Clustering

The clustering of the original networks resulted in 8 clusters of interests. Due to the similarities in consensus between the merged networks, many clusters of interest were found in more than one of the merged networks. For example, Cluster 10 from *STRING0.4+KEGG+WikiPathways[0, 5]* is identical to cluster 8 in the second neighbor network of *STRING0.4+KEGG+WikiPathways[3, 5]*. All the clusters that were formed by MCODE contain only KEGG proteins. The clusters can be

Table 8: Network property changes due to the implementation of a stringent STRING tissue filter of [3, 5].

| Properties | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------------------|-----------------------|-----------------------|
| #nodes | -214(-10.6%) | -214(-10.6%) |
| #edges | -10857(-13.5%) | -3578(-13.5%) |
| Avg. #neighbors | -2.703(-3.4%) | -0.941(-3.5%) |
| Network diameter | 0(+0.0%) | 0(+0.0%) |
| Network radius | 0(+0.0%) | 0(+0.0%) |
| Characteristic path length | -0.015(-0.6%) | -0.003(-0.1%) |
| Clustering coefficient | 0.006(+1.4%) | 0.002(+0.5%) |
| Network density | 0.004(+10.3%) | 0.001(+7.1%) |
| Network heterogeneity | -0.017(-1.8%) | -0.021(-2.0%) |
| Network centralisation | 0.005(+1.5%) | -0.001(-0.7%) |
| Connected components | -3(-60.0%) | -11(-20.0%) |

Table 9: Network property changes due to the implementation of second neighbor networks.

| Properties | STR0.4+KEGG+WP [0, 5] | STR0.4+KEGG+WP [3, 5] | STR0.7+KEGG+WP [0, 5] | STR0.7+KEGG+WP [3, 5] |
|----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| #nodes | -71(-3.5%) | -65(-3.6%) | -539(-26.7%) | -478(-26.4%) |
| #edges | -574(-0.7%) | -525(-0.8%) | -3705(-14.0%) | -3174(-13.8%) |
| Avg. #neighbors | 2.187(+2.7%) | 2.266(+2.9%) | 3.77(+14.1%) | 3.669(+14.2%) |
| Network diameter | -1(-20.0%) | -1(-20.0%) | -3(-42.9%) | -3(-42.9%) |
| Network radius | -1(-33.3%) | -1(-33.3%) | -2(-50.0%) | -2(-50.0%) |
| Characteristic path length | -0.048(-2.0%) | -0.052(-2.2%) | -0.281(-9.3%) | -0.287(-9.5%) |
| Clustering coefficient | -0.012(-2.9%) | -0.013(-3.1%) | 0.009(+2.2%) | 0.007(+1.7%) |
| Network density | 0.003(+7.7%) | 0.002(+4.7%) | 0.007(+50.0%) | 0.007(+46.7%) |
| Network heterogeneity | -0.024(-2.5%) | -0.025(-2.7%) | -0.1(-9.4%) | -0.101(-9.6%) |
| Network centralisation | 0.01(+3.1%) | 0.012(+3.6%) | 0.044(+31.2%) | 0.043(+30.7%) |
| Connected components | -4(-80.0%) | -1(-50.0%) | -54(-98.2%) | -43(-97.7%) |

found in the appendix, starting from Figure 21 until Figure 33.

4.4 Functional enrichment of consensus network

Figure 3 and 4 show the top five enriched biological process and molecular function terms of the consensus network respectively. The proteins related to these terms are indicated by a color in a ring around the nodes. The in total 152 Go Biological Process terms and 28 Go Molecular Function terms were visualized in a treemap. Figure 5 and 6 show the tree maps of all the enriched biological process and molecular function terms of the consensus network respectively. In the tree maps the related terms are connected by colour, which represent an overarching theme. Over-represented themes and their connection to existing scientific research in the field of HD will be discussed in the following subsections.

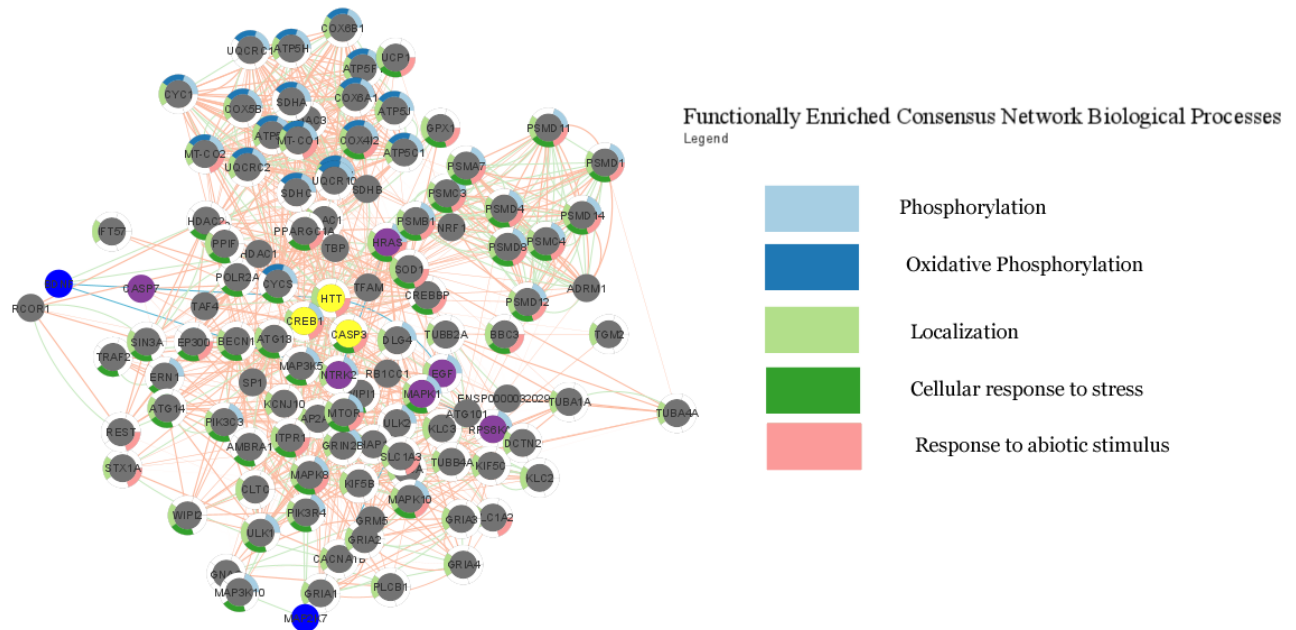


Figure 3: Consensus network enriched with Go Biological Process terms.

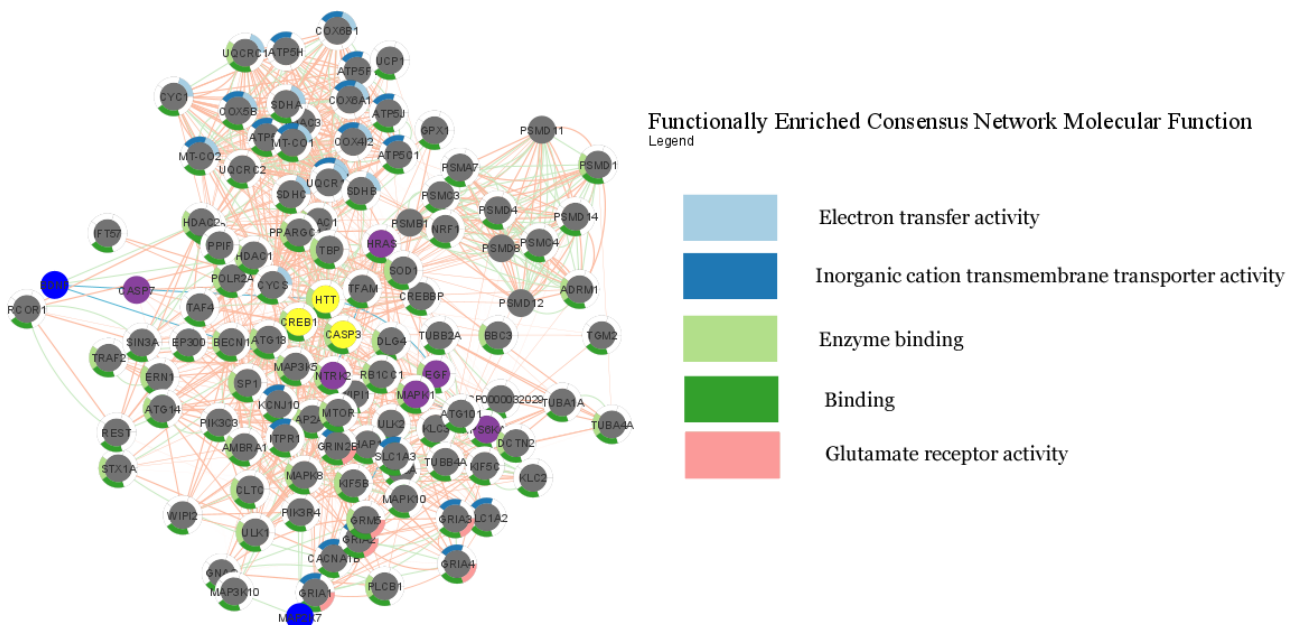


Figure 4: Consensus network enriched with Go Molecular Function terms.

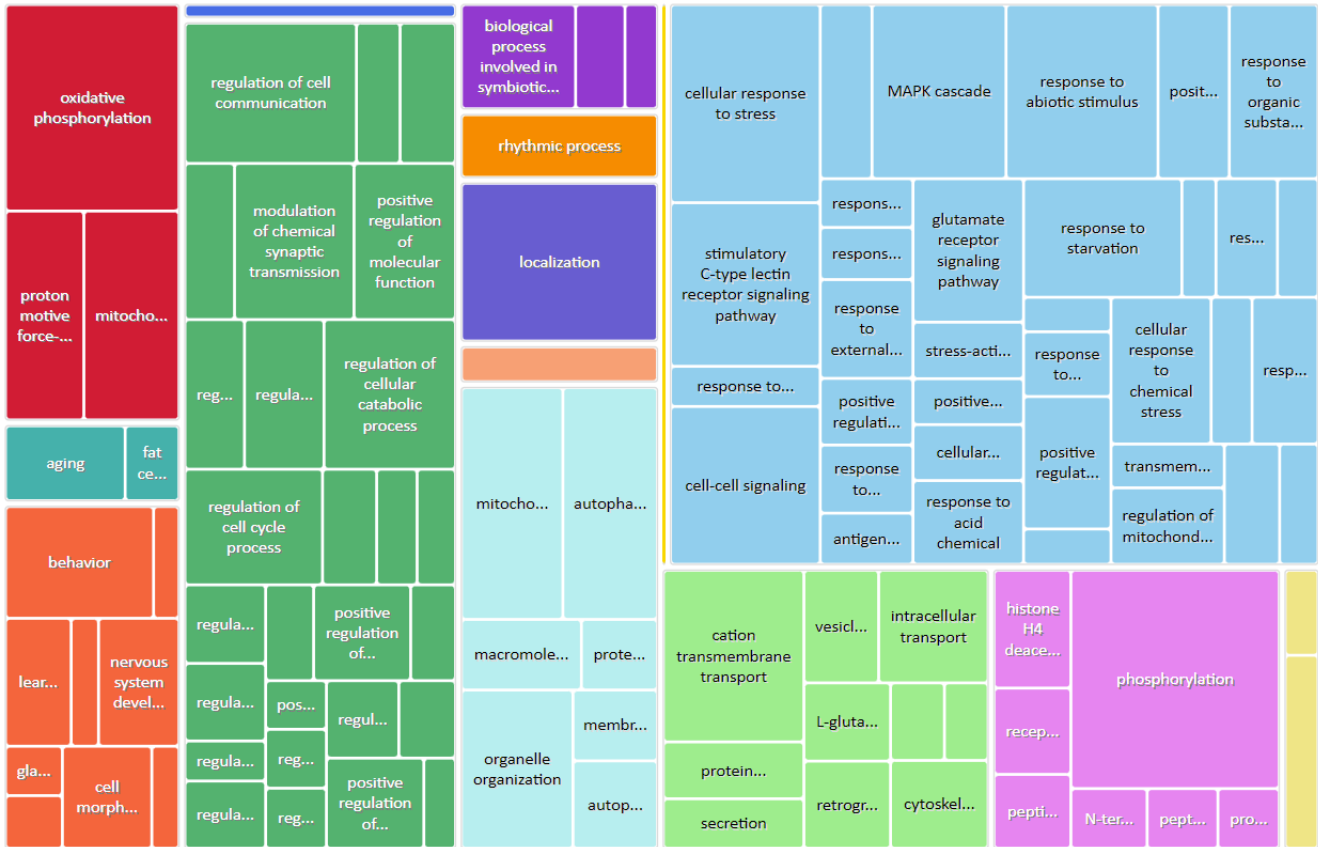


Figure 5: Tree map of the enriched Go Biological Process terms in the consensus network.

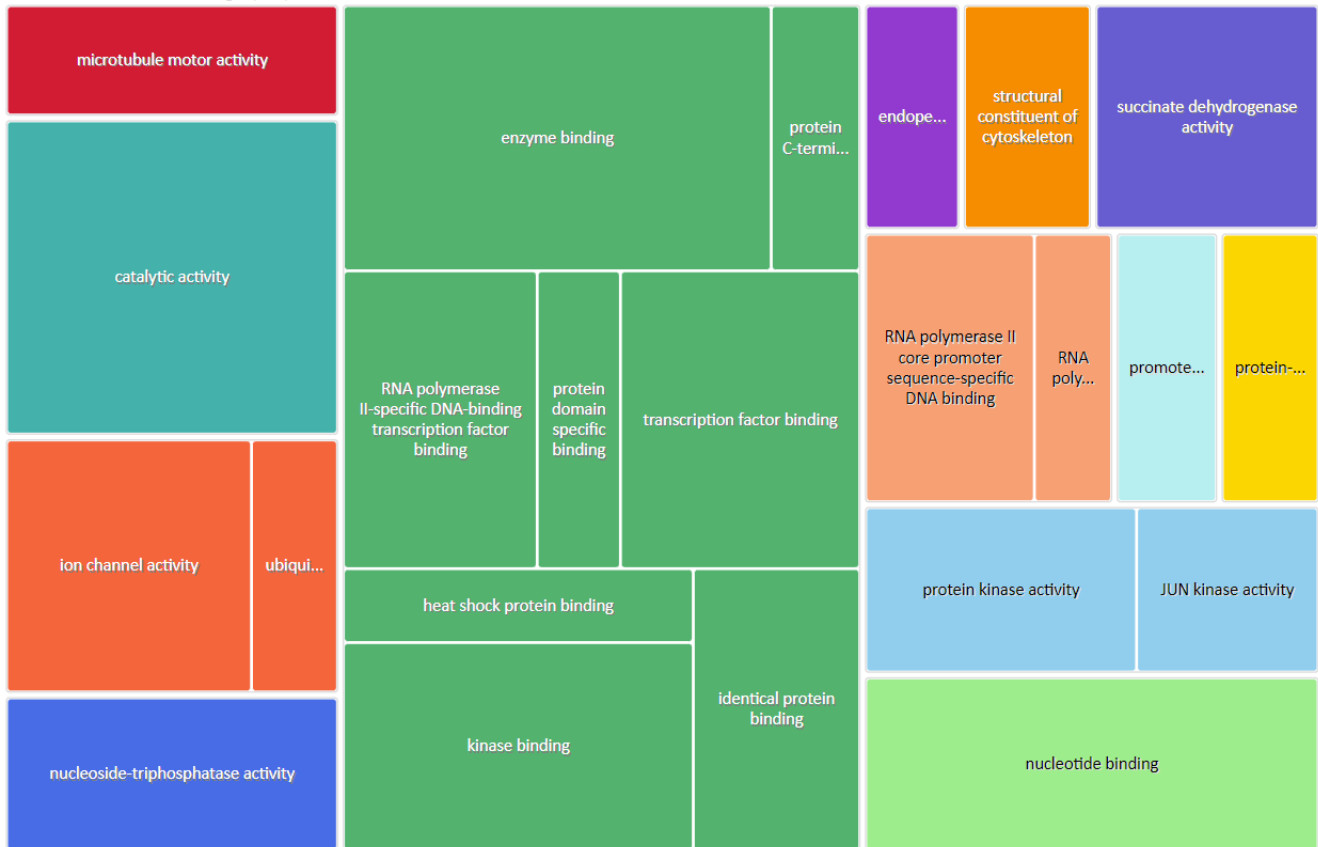


Figure 6: Tree map of the enriched Go Molecular Function terms in the consensus network.

4.4.1 Cellular response

A highly expressed biological process theme is the response to organic and inorganic stimulus. This includes cellular stress, ischemia, staurospine, nicotine, starvation and alkaloids. An important functionality of the endoplasmic reticulum (ER) is its response to stress caused by misfolded proteins. In these cases the ER removes or refolds these proteins. This response is called endoplasmic reticulum associated degradation (ERAD). Mutant HTT depletes the proteins VCP, Npl4 and Ufd1 causing the inhibition of ERAD. The inhibition of ERAD leads to the accumulation of unfolded proteins in the ER, and ER stress, which were both observed in HD models in animal models and postmortem samples of HD patients. [CFA⁺09, CLC⁺09] These proteins were not found in the consensus network. However, VCP and NP14 were both found in the STRING network. A specific term that is highly expressed are mitogen-activated protein kinase (MAPK) cascades. These are central signaling pathways responsible for regulating cellular processes such as stress response, proliferation and apoptosis.[PZPS11] In [AIP⁺06], it is suggested that mHTT alters the MAPK signaling pathways in striatal cells by modulation at upstream points such as the extracellular-signal-regulated kinase (ERK) and c-Jun N-terminal kinases (JNK) pathway. In the past the ERK pathway has been proposed as a potential target for therapeutic intervention for HD. In response to the HD pathogenesis, ERK is activated and directs a protective transcriptional response. However, it has also been established that mutant HTT interferes with the signaling events of the ERK pathway. Furthermore, MHTT downregulates ERK-dependent glutamate transporters causing cells to be vulnerable to excitotoxicity. This seems to be a well-established process in the HD pathway as it is the specific pathway that was extracted from the Wikipathways database. The information found in the scientific research of HD aligns with the higher enriched terms in this theme as well as the GO function terms "kinase binding" and "protein kinase activity". There have also been studies suggesting the involvement of less enriched processes such as response to nicotine and starvation. [TKE⁺05, RSS⁺19]

4.4.2 Regulation of cell communication

Protein phosphorylation is a crucial process in the regulation of cell signaling. Abnormalities in the regulation of this process is known to lead to various diseases. [DSL16] Phosphorylation has been speculated to be involved in the HD pathogenesis by impairing the ULK1-mediated phosphorylation of ATG14, which promotes autophagy. [WLL16] This specific interaction aligns with the enrichment term. According to [Mor09], mutant HTT inhibits fast axonal transport by activating JNK3/MAPK10 and phosphorylating Kinesin-1-heavy chain. Both of these proteins are in the consensus network. Further research emphasizes the importance of the phosphorylation of HTT in regulating structure, toxicity and cellular properties of HTT. [DRF⁺18] These scientific articles describe the involvement of HTT in phosphorylation processes, as inhibitor of phosphorylation, but also as the target of phosphorylation. Recent study has discovered and validated that TANK-binding kinase 1 (TBK1) phosphorylates HTT. [HCP⁺20] This study shows TBK1 expression increases the phosphorylation of mutant HTT and reduces its aggregation and cytotoxicity. TBK1 was not found in the consensus network between KEGG and STRING, but was in the STRING network.

4.4.3 Behavior, movement and memory

HD is generally characterized by behavior, movement and memory impairments. [A.10] Postmortem studies have shown the significant atrophy in HD-affected human brains. Specifically the striatum (caudate-putamen), which is responsible for the coordination of multiple aspects of cognition including motor and action planning, rewards and the conjunction of both movement and reward. [dlMVR88, BMS13] Other affected regions include the motor cortex, which is involved in control and execution of voluntary movements and the cingulate cortex, which is important for emotional regulation. [TOT⁺10] Further research on the HD-affected pathways inside the brain indicate the impairment of long-term potentiation (LTP) in HD mouse models. This process is responsible for strengthening connections between neurons with frequent activation and is part of the underlying mechanism of learning and memory. [DVC⁺11] These studies confirm that HD is related to the processes surrounding behavior, movement and memory as indicated by the functional enrichment terms.

4.4.4 Oxidative phosphorylation

Oxidative phosphorylation (OXPHOS) or electron transport-linked phosphorylation is a pathway related to the synthesis of ATP by an electrochemical transmembrane gradient. This is an essential process for the energy production of the mitochondria and survival of cells, yet its regulation in the brain is still poorly understood. [GM17] Within the field of HD research it has been shown that the cognitive decline as consequence of HD is associated with the preferential loss of striatal medium spiny neurons (MSNs). Even though the specific mechanism that leads to the increased susceptibility of these neurons is not fully understood, there is strong evidence that the loss of medium spiny neurons is related to mitochondrial dysfunction mediated by mutant HTT (mHTT). [BWPk08] Further studies have revealed that the mHTT-induced mitochondrial changes include reduction of Ca²⁺ buffering capacity, loss of membrane potential, and decreased expression of oxidative phosphorylation (OXPHOS) enzymes. [DGDB10] The results of the functional enrichment reaffirm the relation between the mitochondrial dysfunction and mHTT.

4.5 Functional enrichment of discrepancy clusters

Figure 21 until Figure 33 show the top five enriched biological process and molecular function terms in the formed clusters. The in total 17 Go Biological Process terms and 9 Go Molecular Function terms were visualized in a treemap. Figure 7 and 8 show the tree maps of all the enriched biological process and molecular function terms of the clusters respectively. Similarly to the consensus network, a few enriched terms will be connected to existing scientific research. Interestingly, there seem to be many terms related to the previously enriched terms in the consensus network such as "endoplasmic reticulum unfolded protein response" and "oxidative phosphorylation".

4.5.1 7-methylguanosine mRNA capping

The 7-methylguanosine mRNA capping is an essential post transcriptional regulation process and a few papers have mentioned the involvement of this process in the affects of HD. There is a recent paper of 2021 discussing the underlying mechanism of nucleotide repeat expansion disorders such as HD. One of these discussed pathways is translation, which includes 7-methylguanosine mRNA



Figure 7: Tree map of the enriched Go Biological Process terms in the cluster networks.



Figure 8: Tree map of the enriched Go Molecular function terms in the cluster networks.

capping. However, all evidence surrounding this process is more focused towards the larger pathway that 7-methylguanosine mRNA capping is part of.

4.5.2 Nuclear migration along microtubules

The term Nuclear migration along microtubules refers to the transport of the nucleus along microtubules within the cell. An experimental data mining research of 2020 suggests that a significant part of HTT-involved cellular processes is mediated by micro tubules and other cytoskeletal cell structures. In another recent research of 2022, it was found in cell and mouse models that NUMA1 is downregulated in HD. This protein is important in the organization of microtubule and promotes axonal growth. Inhibiting the process causes microtubule and axonal growth defect and consequently disturbs the cytoskeleton. Evidence on the contribution of this pathway to the HD pathology has yet to accumulate, which explains the lack of consensus surrounding this term within the network.

5 Conclusions

From the results it is clear that there is a lot of evidence for the contribution of certain processes in the underlying mechanism of HD such as oxidative phosphorylation, cellular response and regulation of cell communication. Some of these processes seem to be also enriched in non-consensus parts of the networks, which might indicate an expansions on these biological processes that have lower amounts of evidence or just parts that are not as well connected to HD. Newly suggested processes seem to be also enriched in the discrepancy networks, which validates the accuracy of the network as these processes do not have enough evidence to form consensus around. Lastly, to answer the research question, "Does consolidating knowledge on interactions with the huntingtin protein provide new insights into the HD mechanism? And if so, what are its implications?", yes. The consensus that has been made gives insight of the most well researched interactions and processes in the HD pathway and a set of possible interactions that could be targeted for future research.

6 Further Research

Of the three original pathway databases KEGG, Reactome and WikiPathways, as of now, only two contained a HD-related pathways and one of them was a singular ERK pathway. Possible future research could be a similar project, but with more pathway databases to get a more accurate and elaborate consensus of the current knowledge surrounding the HD pathway. This future research could be possibly done in collaboration with HD experts. Because the interpretation of these results and connections to current scientific research would benefit greatly if discussed more thoroughly. Furthermore, this thesis brought consensus and discrepancies of not only processes, but also specific proteins. These specific proteins could be researched in greater detail by enrichment with data from high-throughput methods such as yeast two-hybrid (Y2H) assays and Co-immunoprecipitation (CO-IP) in order to validate their involvement in the HD pathway.

7 Acknowledgements

Major thanks to Katy Wolstencroft for the opportunity to work on this thesis as well as the guidance she has provided over the course of this project.

References

- [A.10] Roos R. A. Huntington's disease: a clinical review. *Orphanet journal of rare diseases*, 5(40), 2010.
- [ABB⁺00] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Jannan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1):25–29, 2000.
- [AIP⁺06] B. L. Apostol, K. Illes, J. Pallos, L. Bodai, J. Wu, A. Strand, J. M. Schweitzer, E. S. ad Olson, A. Kazantsev, J. L. Marsh, and L. M Thompson. Mutant huntingtin alters mapk signaling pathways in pc12 and striatal cells: Erk1/2 protects against mutant huntingtin-associated toxicity. *Human molecular genetics*, 15(2):273–285, 2006.
- [BH03] Gary D. Bader and Christopher W. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, 4:2, 2003.
- [BMS13] R. Báez-Mendoza and W Schultz. The role of the striatum in social behavior. *Frontiers in neuroscience*, 7(233), 2013.
- [BWPK08] Ella Bossy-Wetzel, Alejandra Petrilli, and Andrew B. Knott. Mutant huntingtin and mitochondrial dysfunction. *Trends in Neurosciences*, 31(12):609–616, 2008.
- [CAC⁺09] Peter J.A. Cock, Tiago Antao, Jeffrey T. Chang, Brad A. Chapman, Cymon J. Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, et al. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009.
- [CaCP⁺07] Andrew Chatr-aryamontri, Arnaud Ceol, Luisa M. Palazzi, Giuliano Nardelli, Maria V. Schneider, Luisa Castagnoli, and Gianni Cesareni. Mint: the molecular interaction database. *Nucleic acids research*, 35(Database issue):D572–D574, 2007.
- [CFA⁺09] Alisia Carnemolla, Elisa Fossale, Elena Agostoni, Silvia Michelazzi, Raffaella Caligaris, Luca De Maso, Giannino Del Sal, Marcy E. MacDonald, and Francesca Persichetti. Rrs1 is involved in endoplasmic reticulum stress response in huntington disease. *MOLECULAR BASIS OF CELL AND DEVELOPMENTAL BIOLOGY*, 284(27):18167–18173, 2009.

- [CLC⁺09] K.J. Cho, B.I. Lee, S.Y. Cheon, H.W. Kim, H.J. Kim, and G.W. Kim. Inhibition of apoptosis signal-regulating kinase 1 reduces endoplasmic reticulum stress and nuclear huntingtin fragments in a mouse model of huntington disease. *Neuroscience*, 163(4):1128–1134, 2009.
- [Con20] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489, 2020.
- [Con21] The Gene Ontology Consortium. The Gene Ontology resource: enriching a Gold mine. *Nucleic Acids Research*, 49(Database issue):D325–D334, 2021.
- [COW⁺11] David. Croft, Gavin O’Kelly, Guanming Wu, Robin Haw, Marc Gillespie, Lisa Matthews, Phani Caudy, Michael Garapati, Gopal Gopinath, Bijay Jassal, Steven Jupe, Irina Kalatskaya, Shahana Mahajan, Bruce May, Nelson Ndegwa, Esther Schmidt, Veronica Shamovsky, Christina Yung, Ewan Birney, Henning Hermjakob, Peter D’Eustachio, and Licoln Stein. Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research*, 39(Database issue):D691–D697, 2011.
- [DGDB10] Maria Damiano, Laurie Galvan, Nicole Déglon, and Emmanuel Brouillet. Mitochondria in huntington’s disease. *Biochimica et biophysica acta*, 1802(1):52–61, 2010.
- [dLMVR88] Suzanne M. de la Monte, Jean-Paul Vonsattel, and Jr. Richardson, E. P. Morphometric Demonstration of Atrophic Changes in the Cerebral Cortex, White Matter, and Neostriatum in Huntington’s Disease. *Journal of Neuropathology Experimental Neurology*, 47(5):516–525, 1988.
- [DMBM17] Kevin Drew, Christian L. Müller, Richard Bonneau, and Edward M. Marcotte. Identifying direct contacts between protein complex subunits from their conditional dependence in proteomics datasets. *PLOS Computational Biology*, 2017.
- [DRF⁺18] S. M. DeGuire, F. S. Ruggeri, M. B. Fares, A. Chiki, U. Cendrowska, G. Dietler, and H. A. Lashuel. N-terminal huntingtin (htt) phosphorylation is a molecular switch regulating htt aggregation, helical conformation, internalization, and nuclear targeting. *The Journal of biological chemistry*, 293(48):18540–18558, 2018.
- [DSL16] Evan K Day, Nisha G Sosale, and Matthew J Lazzara. Cell signaling regulation by protein phosphorylation: a multivariate, heterogeneous, and context-dependent process. *Current Opinion in Biotechnology*, 40:185–192, 2016.
- [dTSR⁺21] Noemi del Toro, Anjali Shrivastava, Eliot Ragueneau, Birgit Meldal, Colin Combe, Elisabet Barrera, Livia Perfetto, Karyn How, Prashansa Ratan, Gautam Shirodkar, Odilia Lu, Bálint Mészáros, Xavier Watkins, Sangya Pundir, Luana Licata, Marta Iannuccelli, Matteo Pellegrini, Maria Jesus Martin, Simona Panni, Margaret Duesbury, Sylvain D. Vallet, Juri Rappsilber, Sylvie Ricard-Blum, Gianni Cesareni, Lukasz Salwinski, Sandra Orchard, Pablo Porras, Kalpana Panneerselvam, and

Henning Hermjakob. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Research*, 50(D1):D648–D653, 2021.

- [DVC⁺11] G.M. Dallérac, S.C. Vatsavayai, D.M. Cummings, A.J. Milnerwood, C.J. Peddie, K.A. Evans, S.W. Walters, P. Rezaie, M.C. Hirst, and K.P.S.J. Murphy. Impaired Long-Term Potentiation in the Prefrontal Cortex of Huntington’s Disease Mouse Models: Rescue by D1 Dopamine Receptor Activation. *Neurodegenerative Diseases*, 8:230–239, 2011.
- [GM17] Albert H. Gjedde and Pierre J. Magistretti. *Youmans and Winn Neurological Surgery: Cellular Mechanisms of Brain Energy Metabolism*. Elsevier, 7 edition, 2017.
- [HCP⁺20] Ramanath Narayana Hegde, Anass Chiki, Lara Petricca, Paola Martufi, Nicolas Arbez, Laurent Mouchiroud, Johan Auwerx, Christian Landles, Gillian P. Bates, Malvinder K. Singh-Bains, Mike Dragunow, Maurice A. Curtis, Richard L.M. Faull, Christopher A. Ross, Andrea Caricasole, and Hilal A. Lashuel. Tbk1 phosphorylates mutant huntingtin and suppresses its aggregation and toxicity in huntington’s disease models. *The EMBO Journal*, 39, 2020.
- [HW03] Phoebe Harjes and Erich E. Wanker. The hunt for huntingtin function: interaction partners tell many different stories. *Trends in Biochemical Sciences*, 28(8):425–433, 2003.
- [Jan20] L.M.I. Janssens. Opleiding informatica a network analysis approach to studying polyq protein interactions. *LIACS*, 2020.
- [KAB⁺07] Shiva Krupa, Kira Anthony, Jeffrey Buchoff, Mathey Day, Timo Hannay, and Carl Schaefer. The nci-nature pathway interaction database: A cell signaling resource. *Nature Precedings*, 2007.
- [KG00] Minoru. Kanehisa and Susumu Goto. kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.
- [KLEP14] Martina Kutmon, Samad Lotia, Chris T. Evelo, and Aalexander R. Pico. Wikipathways app for cytoscape: Making biological pathways amenable to network analysis and visualization. *F1000Research*, 3:152, 2014.
- [Kos21] Dylan Kossen. Network analysis of polyq proteins in the human brain. *LIACS*, 2021.
- [LL06] Shihua Li and Xiao-Jiang Li. Multiple pathways contribute to the pathogenesis of Huntington disease. *Molecular Neurodegeneration*, 1(19), 2006.
- [MGB⁺18] David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen DeVeij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, María Gordillo-Marañón, Fiona Hunter, Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis Atkinson, Nicolas Bosc, Chris J Radoux, Aldo Segura-Cabrera, Anne Hersey, and Andrew R Leach. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1):D930–D940, 2018.

- [MLEH15] Dale D.O. Martin, Safia Ladha, Dagmar E. Ehrnhoefer, and Michael R. Hayden. Autophagy in huntington disease and huntingtin in autophagy. *Trends in Neurosciences*, 38(1):26–35, 2015.
- [MNR18] James P. Mackay, Wissam B. Nassrallah, and Lynn A. Raymond. Cause or compensation?—altered neuronal ca²⁺ handling in huntington’s disease. *CNS Neuroscience & Therapeutics*, 24(4):301–310, 2018.
- [Mor09] Gerardo A. et al. Morfini. Pathogenic huntingtin inhibits fast axonal transport by activating jnk3 and phosphorylating kinesin. *Nature neuroscience*, 12(7), 2009.
- [NSA14] Lilit Nersisyan, Ruben Samsonyan, and Arsen Arakelyan. Cykeggparser: tailoring kegg pathways to fit into systems biology analysis workflows. *F1000Research*, 3:145, 2014.
- [pdt22] The pandas development team. pandas-dev/pandas: Pandas 1.4.3, 2022.
- [PKvI+08] Alexander R. Pico, Thomas Kelder, Martijn. P. van Iersel, Kristina Hanspers, Bruce R. Conklin, and Chris Evelo. Wikipathways: pathway editing for the people. *PLoS biology*, 6(7), 2008.
- [PZPS11] Alexander Plotnikov, Eldar Zehorai, Shiri Procaccia, and Rony Seger. The mapk cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1813(9):1619–1633, 2011.
- [Roo10] Raymund A. Roos. Huntington’s disease: a clinical review. *Orphanet J Rare Dis*, 5(40), 2010.
- [Ros04] Christopher A. Ross. Huntington’s disease new paths to pathogenesis. *Cell*, 118(1):4–7, 2004.
- [RSS+19] O. Rudenko, C. Springer, L. J. Skov, A. N. Madsen, L. Hasholt, A. Nørremølle, and B. Holst. Ghrelin-mediated improvements in the metabolic phenotype in the r6/2 mouse model of huntington’s disease. *Journal of neuroendocrinology*, 31(7), 2019.
- [SAB+17] Yao Shen, Mariano J. Alvarez, Brygida Bisikirska, Alexander Lachmann, Ronald Realubit, Sergey Pampou, Jorida Coku, Charles Karan, and Andrea Califano. Systematic, network-based characterization of therapeutic target inhibitors. *PLOS Computational Biology*, 2017.
- [SBR+06] Chris Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. Biogrid: a general repository for interaction datasets. *Nucleic acids research*, 34(Database issue):D535–D539, 2006.
- [SB11] Fran Supek, Matko Bošnjak, Nives Škunca, and Tomislav Šmuc. Revigo summarizes and visualizes long lists of gene ontology terms. *PLOS ONE*, 6:1–9, 2011.

- [SGL⁺19] Damian Szklarczyk, Annika L. Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda H. Doncheva, Nadezhda T. and Morris, P. Bork, and Christian Jensen, Lars J. and von Mering. String v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, 47(D1):D607–D613, 2019.
- [SL11] Joost Schulte and J. Troy Littleton. The biological function of the huntingtin protein and its relevance to huntington’s disease pathology. *Current trends in neurology*, 5:65–78, 2011.
- [SMO⁺03] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.
- [TKE⁺05] Mohammad Tariq, Haseeb Ahmad Khan, Ibrahim Elfaki, Saleh Al Deeb, and Khalaf Al Moutaery. Neuroprotective effect of nicotine against 3-nitropropionic acid (3-np)-induced experimental huntington’s disease in rats. *Brain Research Bulletin*, 67(1):161–168, 2005.
- [TOT⁺10] Doris C. V. Thu, Dorothy E. Oorschot, Lynette J. Tippett, Alissa L. Nana, Virginia M. Hogg, Beth J. Synek, Ruth Luthi-Carter, Henry J. Waldvogel, and Richard L. M. Faull. Cell loss in the motor and cingulate cortex correlates with symptomatology in Huntington’s disease. *Brain*, 133(4):1094–1110, 2010.
- [VBS20] H el ene Vitet, Vicky Brandt, and Fr ed eric Saudou. Traffic signaling: new functions of huntingtin and axonal transport in neurological disease. *Current Opinion in Neurobiology*, 63:122–130, 2020.
- [vDvdVRB86] J. G. van Dijk, E. A. van der Velde, R. A. C. Roos, and G. W. Bruyn. Juvenile huntington disease. *Human Genetics*, 73:235–239, 1986.
- [WBC⁺20] Galen E B Wright, Hailey Findlay Black, Jennifer A Collins, Terence Gall-Duncan, Nicholas S Caron, Christopher E Pearson, and Michael R Hayden. Interrupting sequence variants and age of onset in huntington’s disease: clinical implications and emerging therapies. *The Lancet Neurology*, 19(11):930–939, 2020.
- [WLL16] M.S. Wold, J. Lim, and V. et al. Lachance. Ulk1-mediated phosphorylation of atg14 promotes autophagy and is impaired in huntington’s disease models. *Molecular Neurodegeneration*, 11(76), 2016.
- [WM10] Wes McKinney. Data Structures for Statistical Computing in Python. In St efan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.

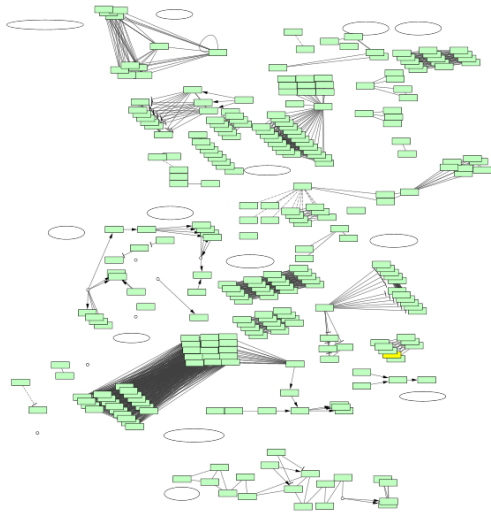


Figure 9: PPIN of the HD pathway from KEGG.

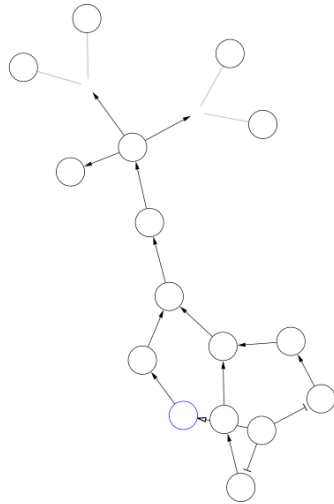


Figure 10: PPIN of the ERK pathway in HD from WikiPathways.

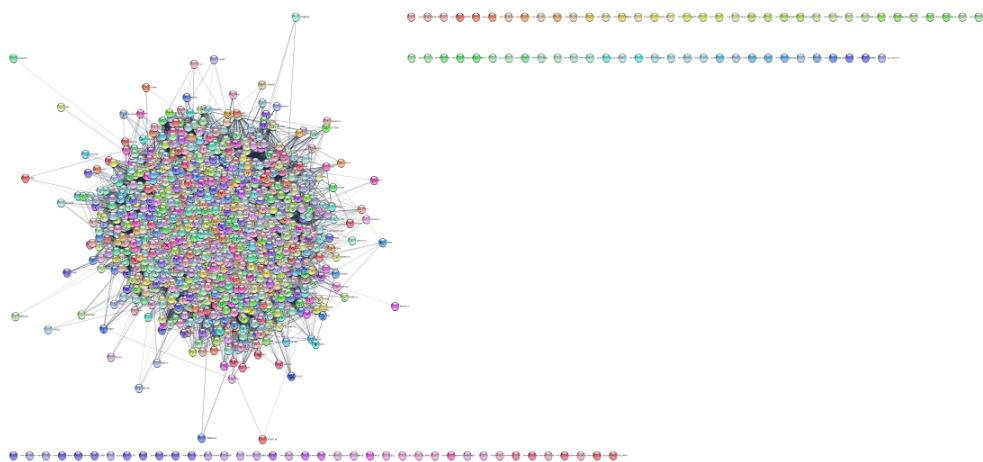


Figure 11: PPIN of HD related interactions with confidence cut-off of 0.4 from STRING.

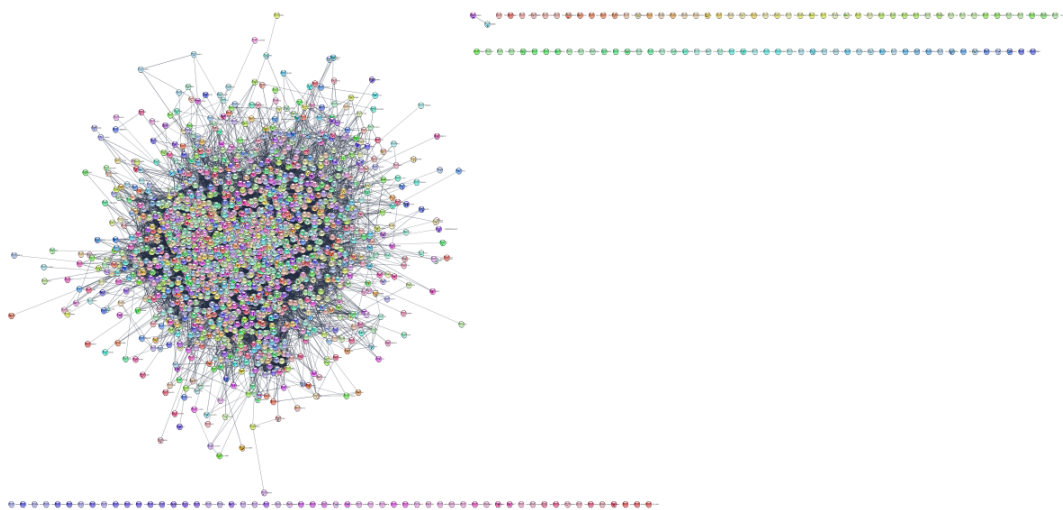


Figure 12: PPIN of HD related interactions with confidence cut-off of 0.7 from STRING.

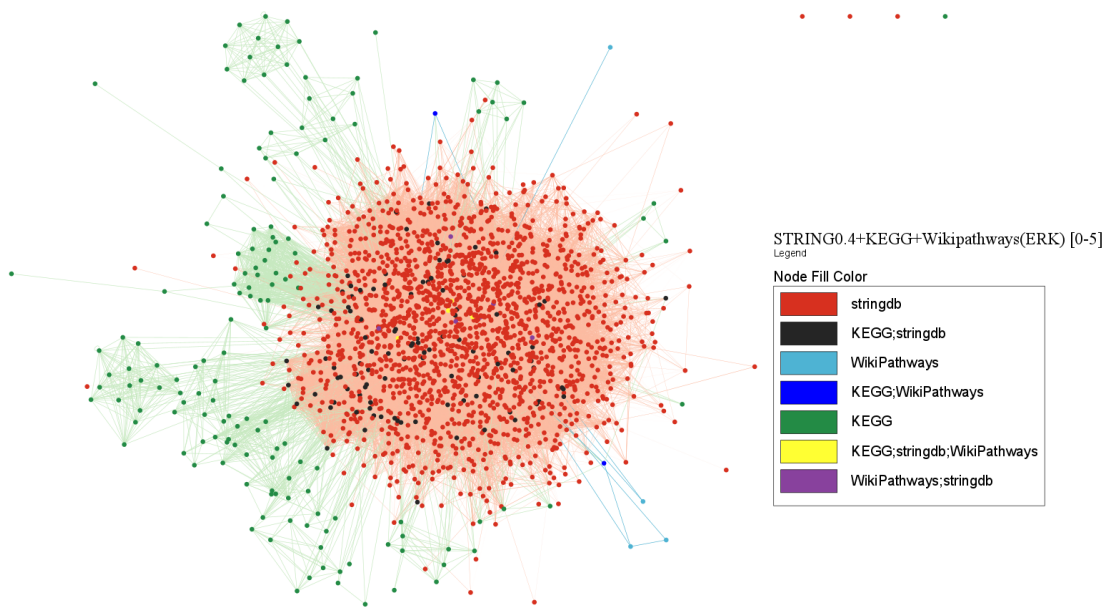


Figure 13: Network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.4 and a STRING nervous system tissue confidence score of [0, 5].

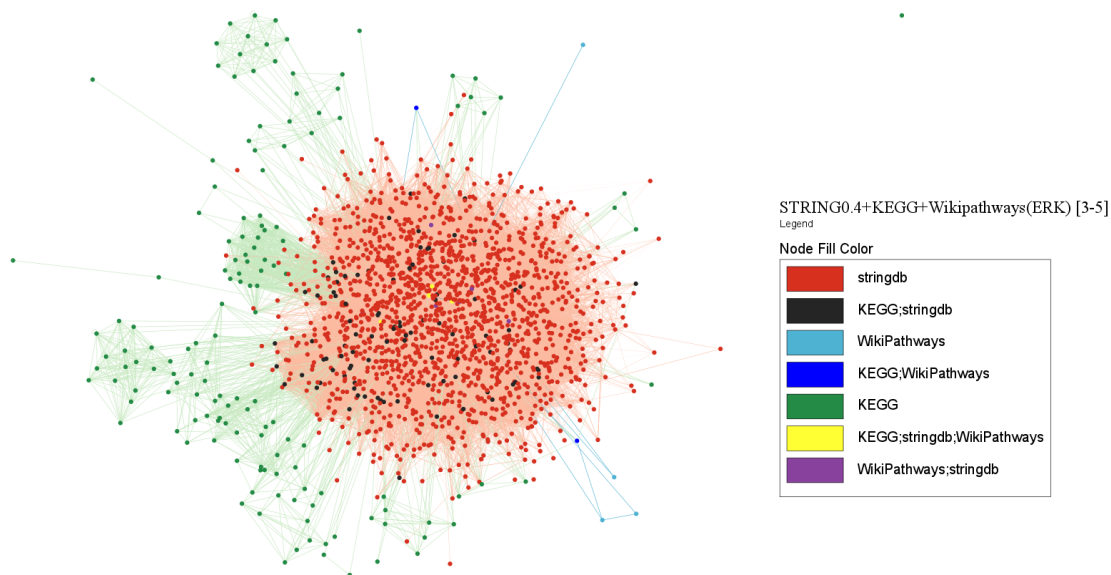


Figure 14: Network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.4 and a STRING nervous system tissue confidence score of [3, 5].

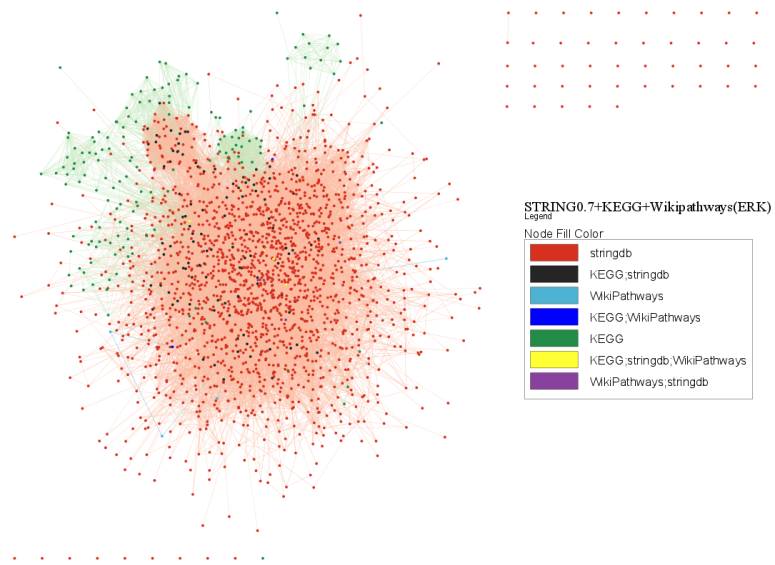


Figure 15: Network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.7 and a STRING nervous system tissue confidence score of [0, 5].

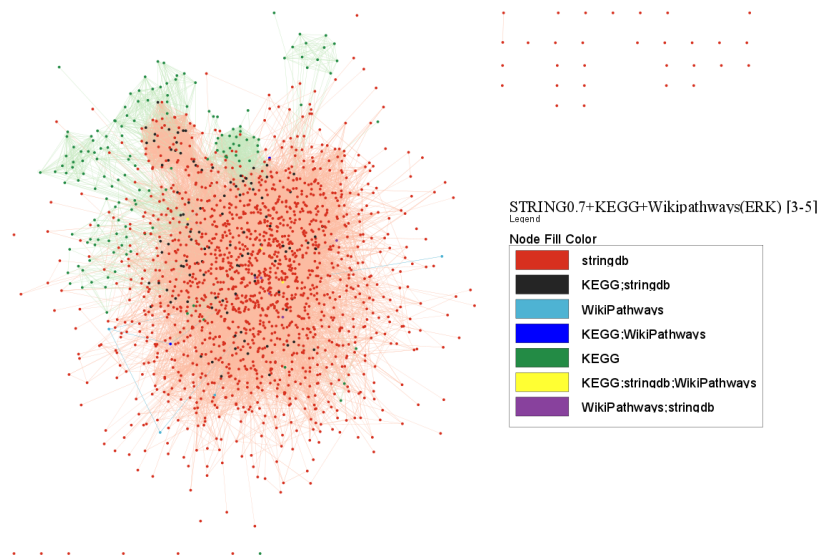


Figure 16: Network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.4 and a STRING nervous system tissue confidence score of [3, 5].

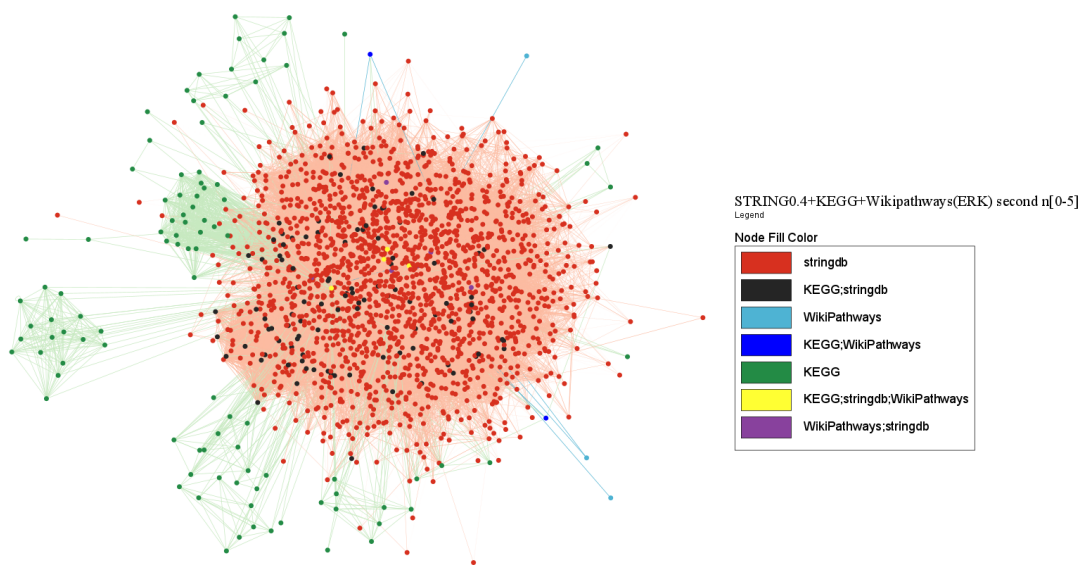


Figure 17: Second neighbor network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.4 and a STRING nervous system tissue confidence score of [0, 5].

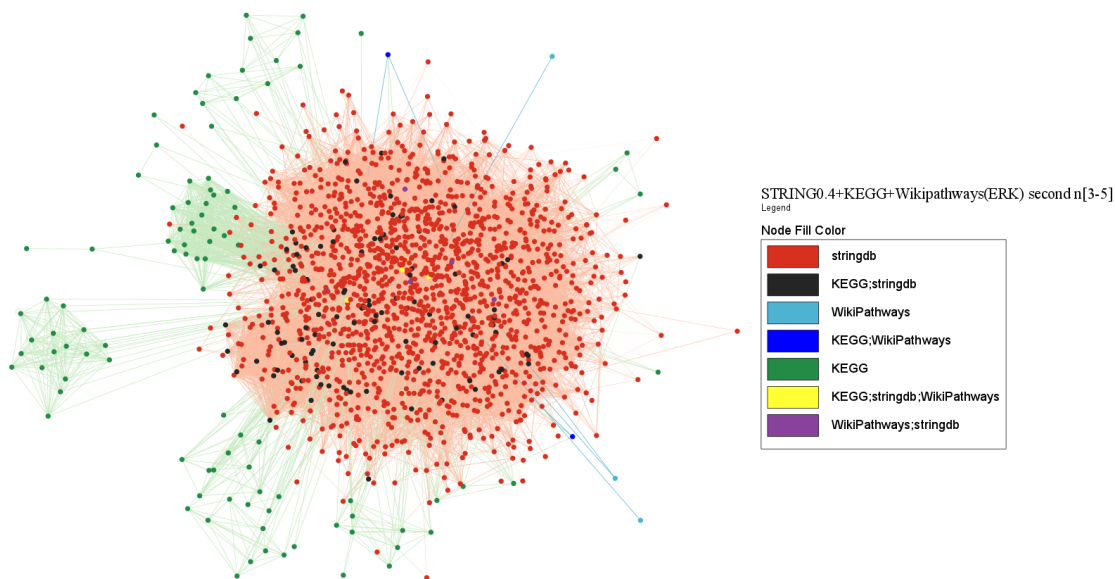


Figure 18: Second neighbor network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.4 and a STRING nervous system tissue confidence score of [3, 5].

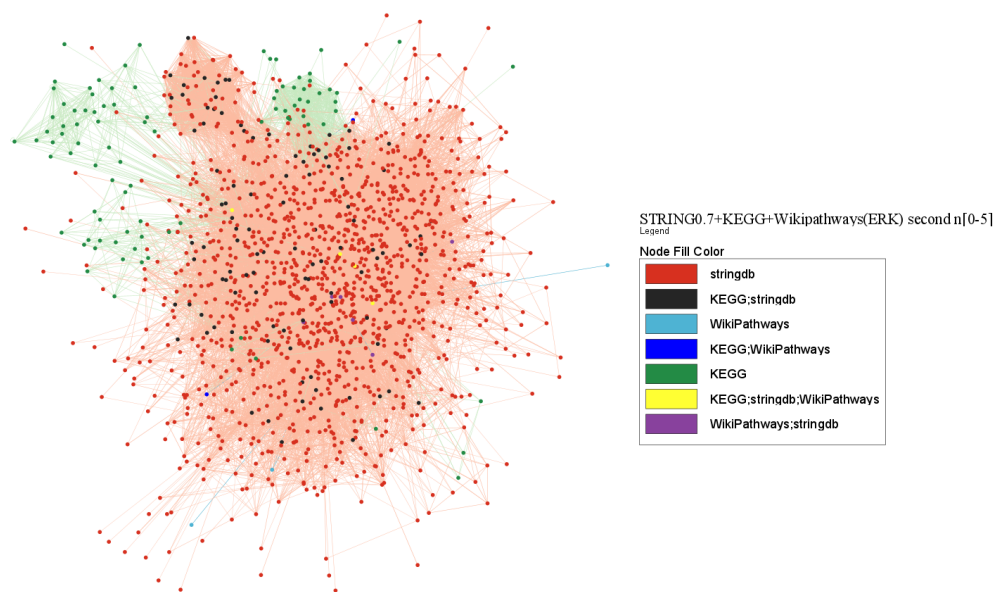


Figure 19: Second neighbor network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.7 and a STRING nervous system tissue confidence score of [0, 5].

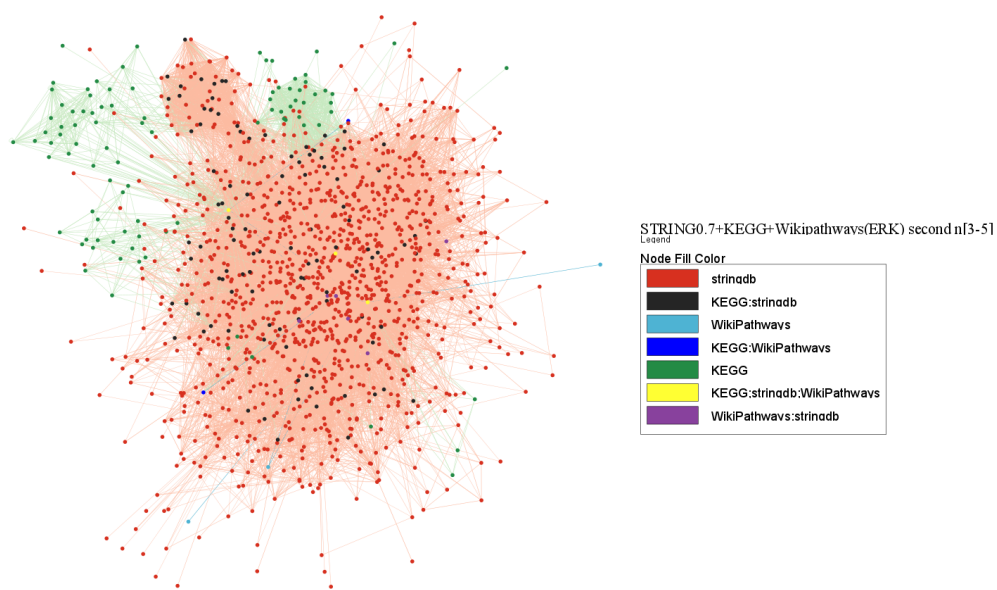


Figure 20: Second neighbor network of the merged network from KEGG, STRING and WikiPathways, with a STRING cut-off score of 0.7 and a STRING nervous system tissue confidence score of [3, 5].

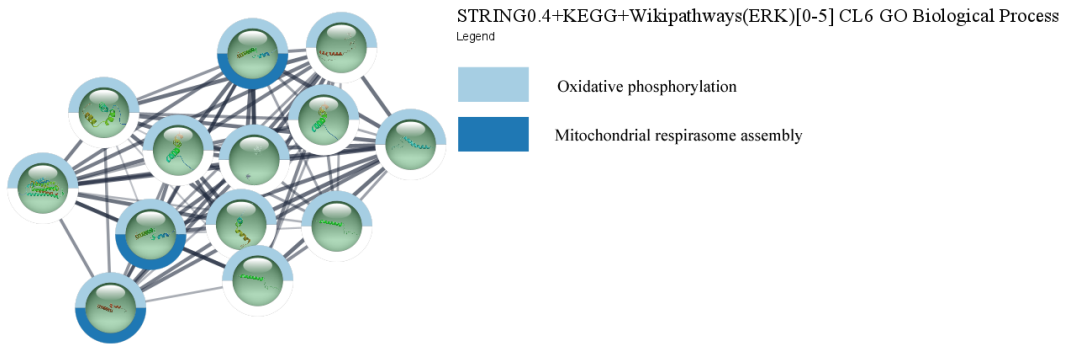


Figure 21: Functional enrichment of Cl(6)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

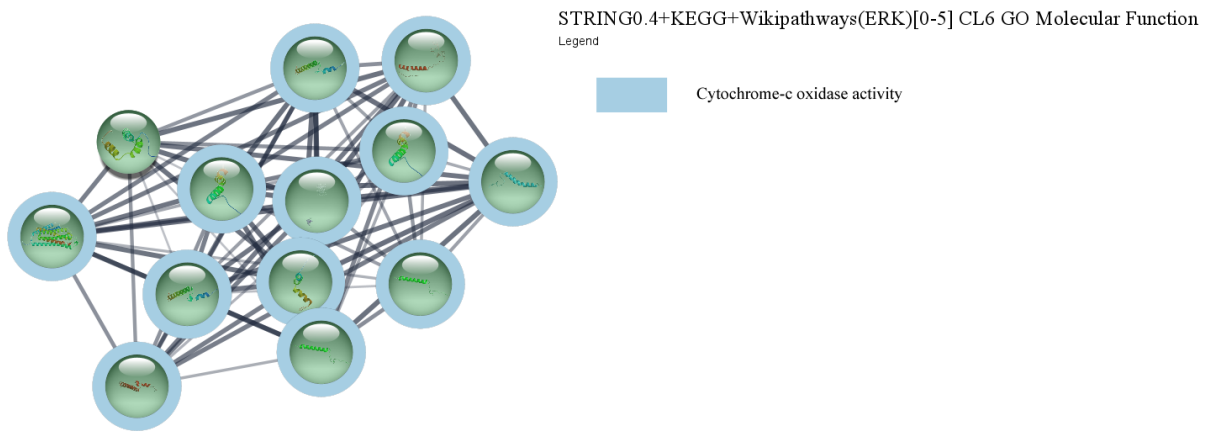


Figure 22: Functional enrichment of Cl(6)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.

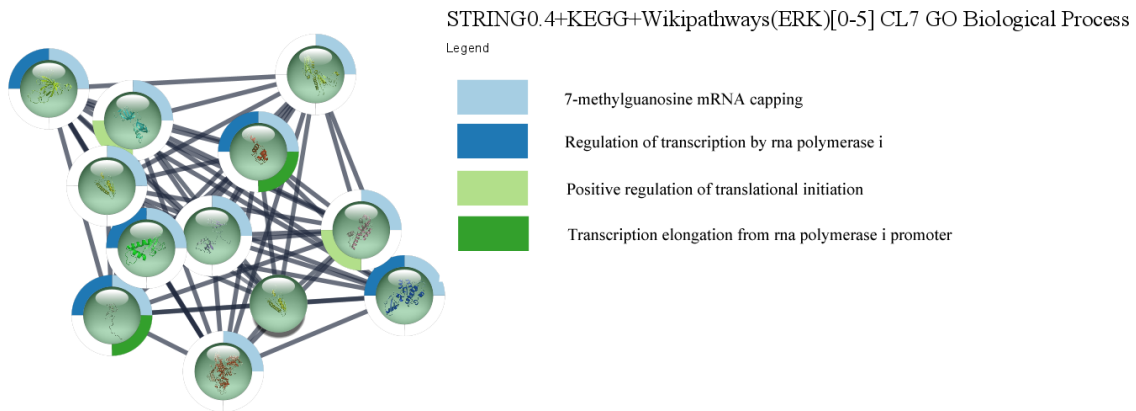


Figure 23: Functional enrichment of Cl(7)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

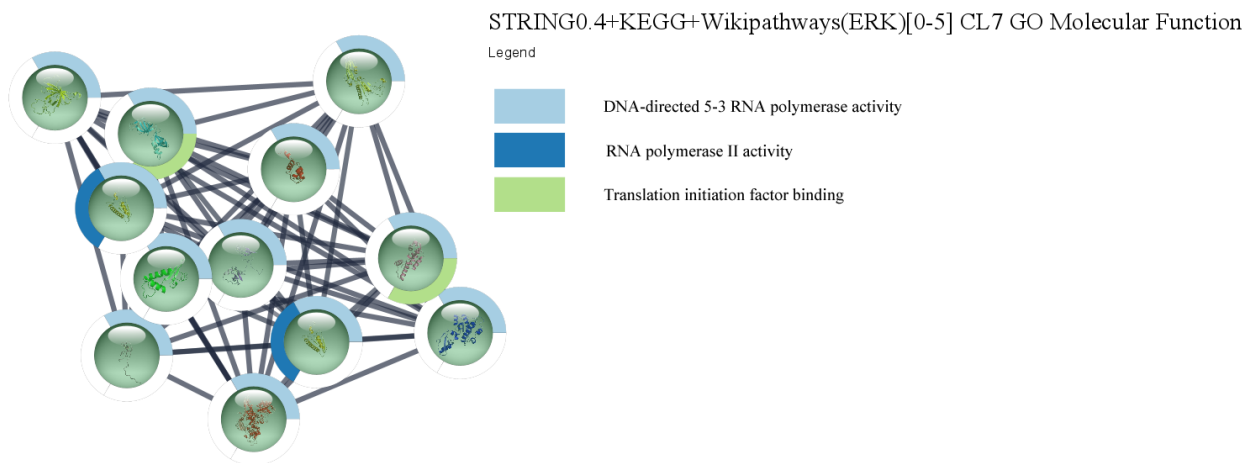


Figure 24: Functional enrichment of Cl(7)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.

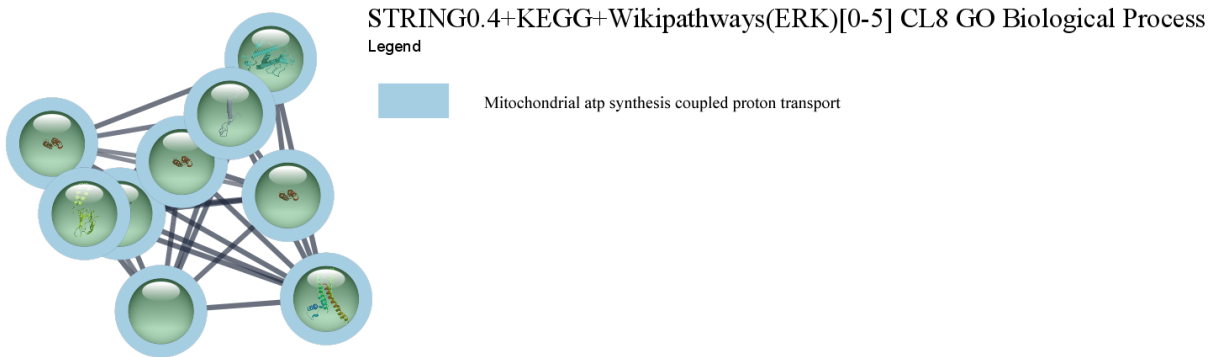


Figure 25: Functional enrichment of Cl(8)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

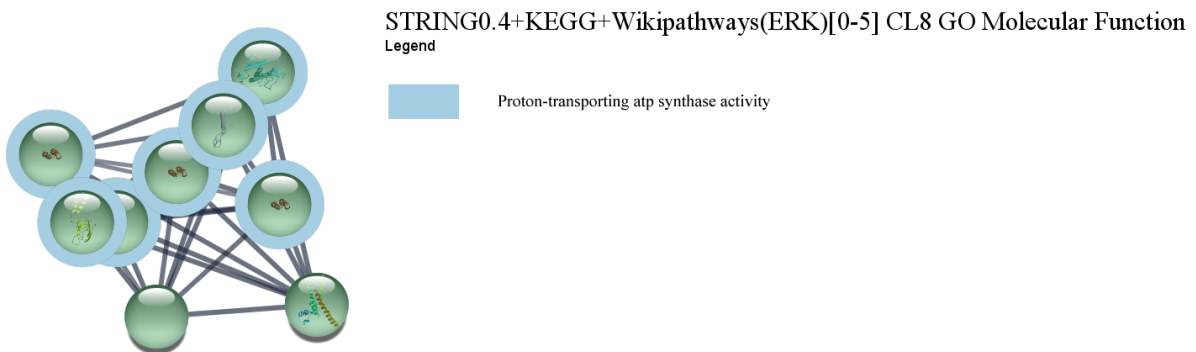


Figure 26: Functional enrichment of Cl(8)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.

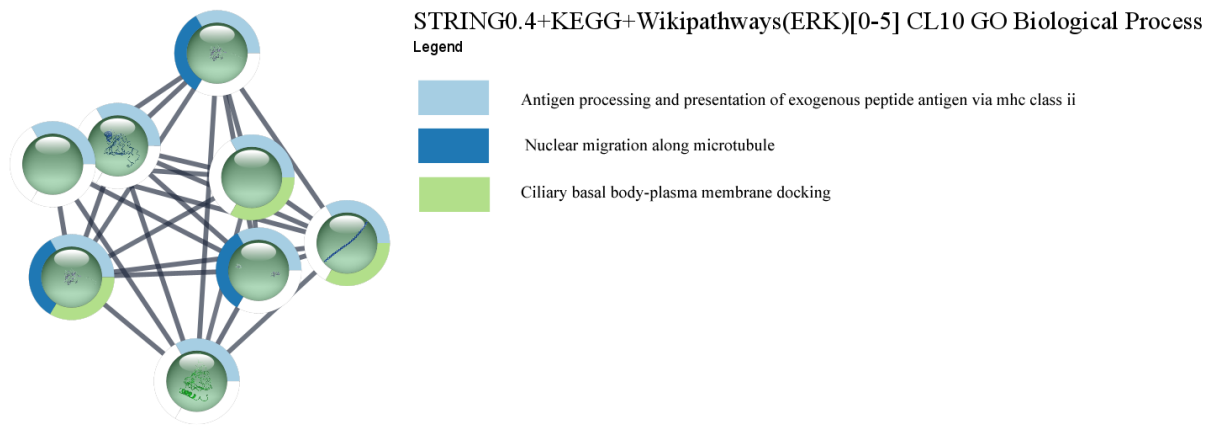


Figure 27: Functional enrichment of Cl(10)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

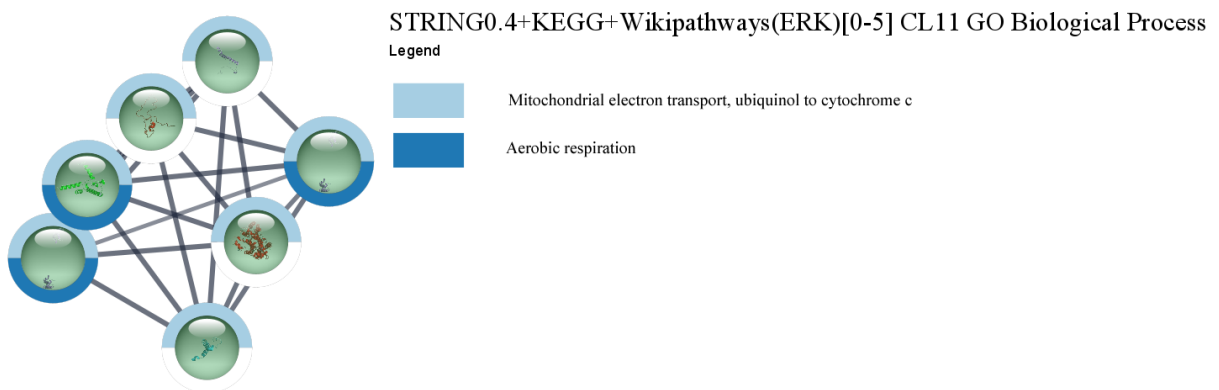


Figure 28: Functional enrichment of Cl(11)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

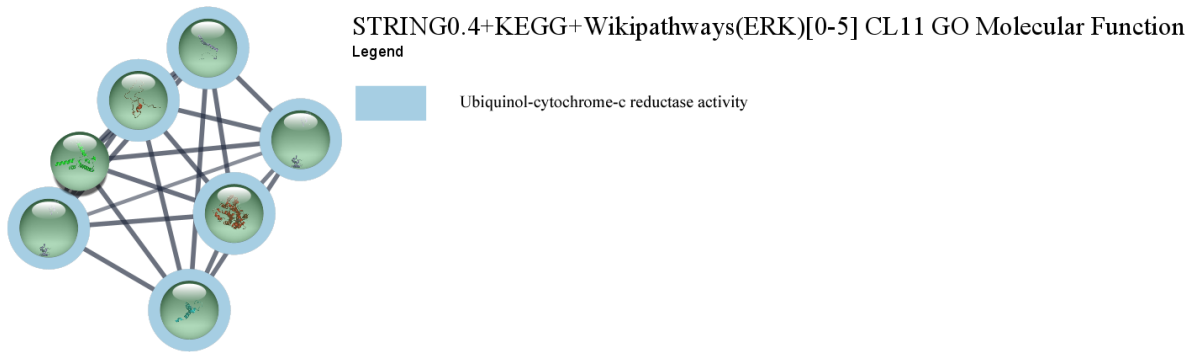


Figure 29: Functional enrichment of Cl(11)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.

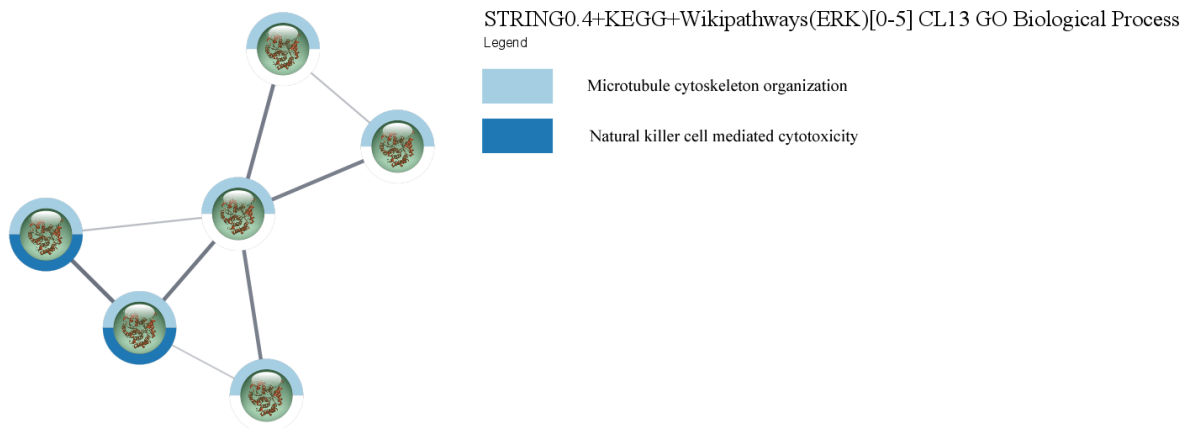


Figure 30: Functional enrichment of Cl(13)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

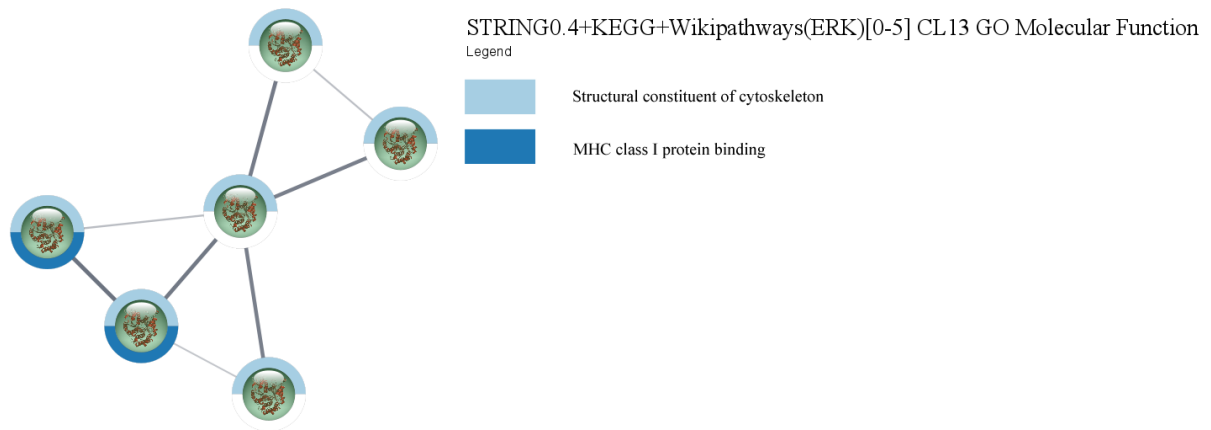


Figure 31: Functional enrichment of Cl(13)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.

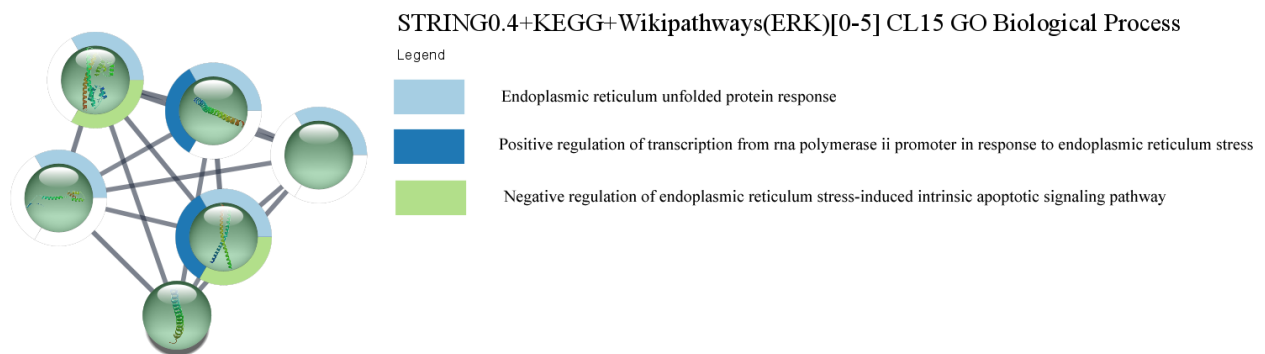


Figure 32: Functional enrichment of Cl(15)STRING0.4+KEGG+WikiPathways[0.5] with GO Biological Process terms.

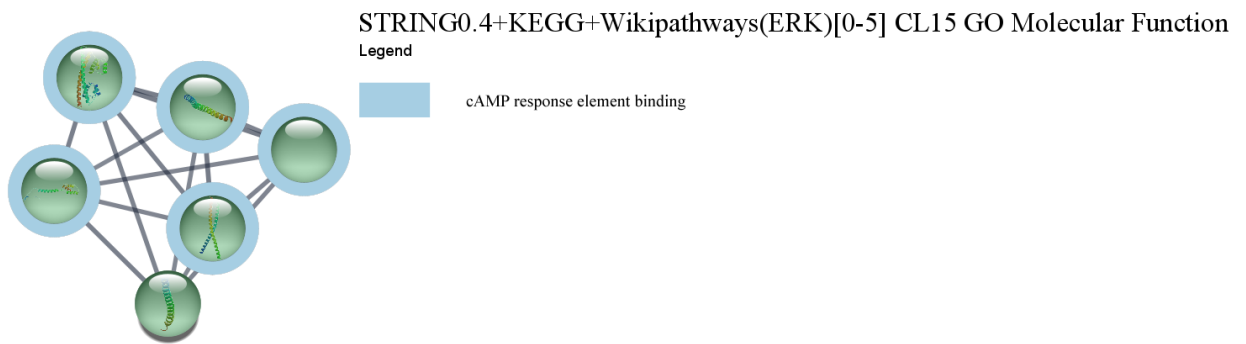


Figure 33: Functional enrichment of Cl(15)STRING0.4+KEGG+WikiPathways[0.5] with GO Molecular Function terms.