Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Camilla Mauroy

A comparative study of gene expression networks associated with Lyme Disease

Master's thesis in Biotechnology

Supervisor: Eivind Almaas

June 2020

NTNU
Norwegian University of
Science and Technology

Camilla Mauroy

A comparative study of gene expression networks associated with Lyme Disease

Master's thesis in Biotechnology
Supervisor: Eivind Almaas
June 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Summary

The desire to identify mechanisms specific to a given biological context, is a key aim in systems biology. The large increase of available data sets of gene expression have made this possible by studying the shared expression patterns between different genes under different conditions. The CSD framework is a newly developed method for conducting differential co-expression analysis, which aims to preserve more information about co-expression patterns by differentiating between different types; conserved (C), specific (S) and differentiated (D). In this work, the CSD method have been applied to perform a comparative study of gene expression from patients with Lyme disease to healthy controls in order to possible identity different co-expression patterns. Lyme disease (also called Lyme Borreliosis) is a tick-borne illness caused by the bacteria *Borrelia Burgdorferi*. The bacteria, itself, is not known to cause any harm to humans, but it initiates biological processes such as inflammation, where most of the tissue damage seems to come from. The CSD method successfully singled out a few known disease related genes while also identifying several interesting genes whose functions may relate to the disease. 21 network hubs were identified: GTF3C2, HKDC1, CCDC157, TNNC2, KHK, LOC440434, CD109, S1PR1, PURA, MMAA, ZNF702P, LRCH1, MBD2, TMEM161B-ASI, NAID1, GSDMA, PARP14, OR6WIP, CCDC85C, PLCD4 AND MMP14. Many of the hubs were, like the rest of the genes in the network, related to cell division, localization and transport of proteins and immune response. These genes are great candidates for further analysis.

While much can be learned about the network topology and its central components by looking at the network as a whole, more can be learned from partitioning the network into smaller sub-graphs with similar topology. In the second part, a set of steps for network analysis that could accompany the CSD-framework was therefore proposed. These steps were as follows; partition the network into modules using community detection tools. Calculate the modularity score for each of the partitions in order to determine which algorithm that works best for that network. Then a gene ontology analysis can be performed, and finally WGCNA can be applied for the RNA-seq dataset. These steps successfully revealed several new genes that were central in their communities, and that could play a part in the pathogenesis of Lyme disease. The Louvain detection algorithm provided the highest quality partition for this network, and revealed eight modules and three new central genes; PSMD13. SAR1A and LBR. The results from the WGCNA analysis successfully identified 159 differentially expressed genes with a suggestive role upon infection by the bacteria. 24 of these genes were of special interest as their function relate to immune responses, inflammation and defense against bacteria.

Sammendrag

Et viktig mål for systembiology er å identifsere mekanismer som er spesifikke for en gitt biologisk kontekst. Den store økningen av tilgjengelig datasett for genuttrykk har gjort det mulig å studere sammenhengen mellom genutrykk for to gener under ulike forhold. CSD rammeverkt er en nyutviklet metode for å gjøre nettop dette, og den gjør det ved å utføre differensiell koekspresjons analyse. Målet med CSD-metoden er å bevare mest mulig informasjon om koeekspresjons-mønstrene ved å differensiere mellom ulike former for differensiell koeskspresjon; konservert (C), spesifikk (S) og differensiert (D). I dette arbeidet er CSD metoden blitt brukt til å utføre et sammenlignende studie av genuttrykk fra pasienter diagnosert med Lyme borreliose mot en kontrollgruppe for å identifisere mulige forskjeller i sameekspresjonsmønster. Lyme Borreliose er en flåttbåren sykdom forårsaket av bakterien *Borrelia Burgdorferi*. Bakterien i seg selv er ikke kjent for å forårsake noen skade hos mennesker, men den inititerer biologiske prosesser som betennelse, hvor det meste av vevsskadene ser ut til å komme fra. CSD-metoden greide å identifiserte noen få sykdomsrelaterte gener, samtidig som den identifiserte flere interessante gener som kan ha tilknytning til sykdommen. 21 nettverksnav ble identifisert: GTF3C2, HKDC1, CCDC157, TNNC2, KHK, LOC440434, CD109, S1PR1, PURA, MMAA, ZNF702P, LRCH1, MBD2, TMEM161B-ASI, NAID1, GSDMA, PARP14, ORCW, OR6W, AND6W, Mange av nettverksnavene var, som resten av genene i nettverket, relatert til celledeling, lokalisering og transport av proteiner og immunrespons. Disse genene er gode kandidater for videre analyse.

Selv om mye kan læres om nettverkstopologien og sentrale komponenter i nettverket ved å se på nettverket i sin helhet, kan man lære enda mer av å dele opp nettverket i mindre grupper med lignende topologi. I den andre delen av oppgave ble det derfor foreslått et sett med trinn for nettverksanalyse som kunne brukes med CSD-rammeverket. Disse trinnene var som følger; del opp nettverket i mindre grupper ved hjelp av eksisterende algoritmer som gjennkjenner fellestrekk i gruppene. Deretter kan man bergene modularitet for de ulike oppdelingene av nettverket for å bestemme hvilken algoritme som fungerer best. Så kan en genontologi-analyse utføres, og til slutt kan WGCNA brukes på RNA-seq datasettet. Disse trinnene klarte å identifisere flere nye gener som var sentrale i deres grupperinger, og som kunne spille en rolle i sykdomsforløpet til Lyme Borreliose. Louvain-deteksjonsalgoritmen ga partisjonen av høyest kvalitet for dette nettverket, og avslørte åtte grupperinger og tre nye sentrale gener; PSMD13. SAR1A og LBR. Resultatene fra WGCNA-analysen lyktes med å identifisere 159 gener med varierende forskjellig genuttrykk ved de to ulike situasjonene, hvorav disse kan antydes å være involvert i kroppens respons til infeksjon av bakterien. 24 av disse genene var av spesiell interesse, da deres proteiners funksjon kan relaterer til immunrespons, betennelse og forsvar mot bakterier.

Preface

The work for this master thesis was carried out at the Department of Biotechnology and Food Sciences at the Norwegian University of Science and Technology (NTNU) in Trondheim. This project concludes my Master of Science degree in Biotechnology where I have specialized in Systems Biology.

Firstly, I would like to express my gratitude to my supervisor, professor Eivind Almaas, for his help and encouragement, providing me with advice and guidance throughout this project. It was professor Eivind Almaas captivating way of lecturing about systems biology and network science that inspired me to specialize in this field. I would also like to give thanks to my co-advisor André Voigt for his input and assistance, especially regarding the technical aspects of this thesis such as computer tricks and debugging abilities. He offered valuable insight into the workings of the CSD method, and answered all sorts of questions I might have. I would also like to thank the other candidates of the Systems Biology group, and especially Martina Hall for her valuable insight in R and the WGCNA method.

A special thanks to my boyfriend, Julian Aaserud, for being the best supporter throughout my five years at NTNU, and especially during the writing of this thesis. Thank you for being there for me at all times, and for helping me with everything from coding in python and LaTeX problems to emotional support and endless care. Thank you to all my friends and classmates for all your support and fun times. You have made my stay in Trondheim unforgettable and I could never have done this without you. Lastly, I want to thank my family for making all of this possible and for the endless encouragement and support.

Camilla Mauroy
June 2020

Table of Contents

List of Tables

List of Figures

Abbreviations

N	=	Size of the network, number of nodes
A_{ij}	=	Adjacency matrix
a_{ij}	=	Elements in the adjacency matrix
k	=	Node degree
BC	=	Betweenness centrality
Q	=	Modularity score
GCN	=	Gene co-expression network
s_{ij}	=	Similarity matrix
ρ	=	Spearman rank-order correlation
C,S,D	=	Conserved, Specific, Differentiated
i, j	=	Gene pair from the two conditions healthy and control
$k_p^{C,S,D}$	=	Threshold values for C, S and D
H	=	Node homogeneity
p	=	Importance level
GO	=	Gene ontology
wTO	=	Weighted topological overlap
TOM	=	Topological overlap matrix
GMM	=	Greedy modularity maximization
LPA	=	Label propagation algorithm
GN	=	Girvan-Newman method
LCA	=	Louvain community detection algorithm
WGCNA	=	Weighted gene co-expression network analysis
DEGs	=	Differentially expressed genes

Chapter 1

Introduction

A change in the way of thinking about biology occured in the 1950's when the Austrian biologist Ludwig Won Bertalanffy set the stage for systems biology. He formulated the notion that all systems share a property of being composed of interlinked components, and that their behaviour is different when studied in isolation [7]. This moved away from the Descartian Reductionism, the idea that higher levels of biological hierarchy can be understood from lower levels, which had dominated the way of thinking since the 1600s [8]. Hence, systems biology is about understanding the biological systems as a whole, rather than piece by piece. It can be defined as the study of biological function emerging from interactions at all levels of the hierarchy. However, this description is brief, and there are many more extensive definitions of it.

A wide variety of systems can be described by complex weblike structures [9]. Take for example the systems encountered in an everyday routine. The bus routes to work are best described as networks of bus stops connected by routes going through the crowded city. The class room or office space can be explained by a complex weblike structure of social relationships where your co-workers or fellow students are connected by friendships, projects or common interest. In these networks we can also find outliers such as a bus stops at the outskirts of the city or a student in the classroom that keeps to himself. On the contrary one can also find central components in these networks who share connections with many others, such as the central bus stop in the town square or the most popular student in the class room. These systems are just a few of the many examples existing and the desire to understand complex systems and to make sense of their behaviour set the stage for *network science*.

Network science emerged at the beginning of the 21 century as a result of the discovery that despite the diverse nature of complex systems, they are all driven by a common set of fundamental laws and principles [3]. The differences in form, size and nature of these complex systems can be disregarded, and instead be presented in their simplest forms. Network science have found many applications across different branches of science and

have enabled the development of network-based algorithms to study the properties and behaviour of systems. In biology, networks science have for example enabled the development of network-based algorithms, which exploit patterns already characterized by networks to predict protein interactions [10] or to identify novel insight in drug discovery [11].

In the same way that there are complex networks everywhere you look, there are those we cannot see but that are vital to us. One of these systems is our body, which in itself is a whole world of large and small complex systems. The cells in your body are organized in complex structures and communicate with each other to transfer signals vital for your body to function. Going even further we can find the complex networks of proteins and genes that decide who you are and what you are able to do based on the way they interact with each other. The analysis of these complex biological networks is the basis of systems biology and there are commonly two types of network studied; protein interaction networks and gene co-expression networks.

In this thesis the focus will be on the latter, which is a way of investigating the relationship between genes based on co-expression: the extent to witch their expression is correlated. More specifically, the focus will be on differentially co-expressed genes, which compare the gene co-expression from two different conditions. This method is widely applicable and can be used on different tissue, species and comparison between healthy and disease states, providing insight into the genetic regulatory mechanism that cause difference in phenotype. The method for differential co-expression analysis used in this thesis is the CSD framework [4]. The CSD framework differs from already existing methods in that it distinguishes between various forms of differential co-expression by assigning them into three categories; conserved (C), specific (S) and differentiated (D) co-expressions, thus providing more information about the co-expression patterns.

The aim of this thesis can be considered in two parts. In the first part of this thesis the goal was to be acquainted and employ the CSD framework to perform differential co-expression analysis. This part focused on obtaining gene expression data, from the GEO database, obtained from whole blood samples of patients diagnosed with Lyme disease, in order to identify differences in gene expression when comparing their data to healthy individuals. The intention was to identify underlying biological processes and the potential role of genes in the network.

The second aim of this thesis was to propose a set of steps for further network analysis, that could accompany the CSD framework, in order to extract more information from the differentially gene co-expression network constructed. The focus was on different community detection algorithms across platforms such as Python and R, and what these different algorithms provided of new information regarding biological processes and important players in the pathogenesis of Lyme disease.

Chapter 2

Theoretical Background

The aim of this chapter is to provide the reader with a basis for understanding the methods employed in this thesis. The following sections is of direct relevance to the following chapters. Most of the theory presented in the network theory section is obtained from Network Science written by Albert Lázló-Barabási [3]. The reader is referred to this book for more information on the subject. The description of the CSD method developed by Voigt et. al [4] is obtained from their article, *"A composite network of conserved and tissue specific gene interactions in glioma"*, where the method was first described.

2.1 Lyme Disease

Lyme disease, also commonly called Lyme borreliosis, is a tick-borne infection caused by the bacteria *Borrelia burgdorferi* (*B. burgdorferi*). *B. burgdorferi* is of a spirochete class of the genus Borrelia and exist in North America and Europe. Common reservoirs for the bacteria are rodents, deer and birds [12].

Ixodes ticks, which are a genus of hard-bodied ticks, are the primary vectors for transmitting the *B. burgdorferi*. *Ixodes* ticks generally have a life cycle lasting for two years where they go from eggs to larvae to nymph and finally to adult. To survive, the ticks must feed on blood at every stage in the cycle and preferably a new host every time. When feeding upon an animal, e.g. deer that contain the bacteria *B. burgdorferi*, the tick will acquire the bacteria in its midgut. The spirochetes increase in number and undergo phenotypic changes, one of them being the expression of outer surface protein C (OspC). Expression of this protein allows the bacteria to invade the ticks salivary glands and thereby deposit onto the skin of a host, e.g. human, that the tick feeds on in the next stage. Then the bacteria spreads from the infection site through the bloodstream to other places of the body [12, 13].

Once bitten by an infected tick it can take from 3 to 32 days for symptoms to occur. If

untreated, a wide range of symptoms may occur depending on stage of infection. In the early stage (3 - 30 days after infection) ”classic” erythema migrans (EM) often occur. These are a red rash that occurs at the site of the tick bite and gradually expands over several days, creating a ”bull's-eye” appearance. Other symptoms occurring at this stage can be fever, chills, headache, fatigue, muscle and joint aches, and swollen lymph nodes. Later occurring symptoms (days to months after tick bite) may include facial palsy which is the loss of muscle tone or droop on one or both sides of the face. One can also experience arthritis with severe joint pain and swelling, episodes of dizziness or shortness of breath or numbness in hands or feet. If the bacteria enters through the heart tissue it can interfere with the normal movement of electrical signals through the heart, a condition called Lyme carditis. This may result in heart palpitations or irregular heartbeat in addition to the other symptoms [1, 13].

However, the bacteria itself is not known to cause any harm to humans, but the infection initiates the innate and adaptive immune response, resulting in inflammatory responses. It is from the inflammatory response most tissue damage seems to result from [12]. In most patients the immune cells first detect the spirochete at the site of the tick bite. This trigger the activation of phagocytes, a type of cells that ingest foreign particles, e.g. bacteria, upon detection. In addition to engulfing the spirochete, phagocytes also produce proinflammatory cytokines, especially TNF-α and IL-1β. These cytokines are signalling molecules of the immune system that promote inflammation, i.e. the activation of T- and B-cells. In addition to killing the spirochetes, the T-cells (especially Th1 cells) and cytokines prime the T-cell dependent B-cell response, the response that produces antibodies. The B-cell response is also directly stimulated by lipoproteins of the spirochete. Within days after infection, most patients will have IgM antibody response to OspC or the flagellar protein of the spirochete [14, 15]. A list of genes previously associated with the disease is presented in table A1.

The Lyme disease have been recognized since the late 20th century and has been an emerging infection ever since [16]. It is the most common tick-borne infection in both the United States and Europe, and table 2.1 present the number of incidence registered in Norway from 2013 to last year (2019) across different age groups [15, 1]. The number of occurrences is obtained from Folkehelseinstituttets (FHI) statistic website: ”Meldingssystem for smittsomme sykdommer (MSIS) [1]. The total amount of occurrences for each year and age group is also presented.

Table 2.1: Occurance of Lyme borreliosis. The number of occurrence of lyme borreliosis in Norway reported to MSIS [1]. The occurrences are ordered by year and age group along with the total of occurrences.

Age group	2013	2014	2015	2016	2017	2018	2019	Total
0-9	86	73	147	104	115	116	120	761
10 - 19	27	27	40	38	28	28	30	218
20 - 29	7	13	10	14	17	14	16	91
30 - 39	22	21	20	25	22	26	32	168
40 - 49	29	25	44	41	41	45	51	276
50 - 59	51	57	53	64	62	56	80	423
60 - 69	56	54	52	60	73	69	69	433
70 - 79	30	40	44	49	60	57	77	357
80 +	8	12	16	14	20	10	18	98
Total	316	322	426	409	438	421	493	2825

2.2 RNA-seq

Our understanding human disease is dependent on the understanding of the genome and its function. The cells constantly needs to regulate their expression of genes to endure the constant change of environment, which they do by initiating the transcription of genes, whose gene product (protein) has the desired function needed by the cell. In more detail, the newly activated gene is transcribed to its corresponding messenger RNA (mRNA) which is then further processed and translated into its corresponding protein. The quantity of RNA transcripts in a cell at given time point is referred to as the *transcriptome* [17]. Looking at the transcriptome, will provide a snapchot of which genes that are expressed in the cell [18].

The number of RNA transcripts provide valuable information about gene expression in human diseases. One way to obtain the number of RNA in a cell at a given time is by using RNA-seq. RNA-seq is the use of high-throughput complementary DNA (cDNA) sequencing to characterize an RNA sample [18]. During RNA-seq the mRNA is first isolated and fragmented. Then, the mRNA is converted to cDNA, creating a cDNA library. Aligning these cDNA sequences to the genome of the organism and counting their relative copy number provides an indication of gene expression levels in the organism. [18]. This relative copy number is referred to as the read count.

Several factors can affect the raw read count, including length of the transcripts and total read count. Therefore, normalization of the read counts is necessary for it to be possible to compare gene expression levels between samples. There are several metrices attempting to perform this normalization, one of them being the FPKM (fragments per kilobase of exon model per million reads mapped) metric. FPKM is able to account for both gene length and number of read counts. The one thing differing FPKM from other metrices is that it accounts for two reads being able to map to one fragment, thus ensuring that the frag-

ments are not counted twice [19]. Once the gene expression levels have been normalized, comparison across samples are possible. By then comparing the gene expression levels between two situations e.g. healthy and sick patients, the results will provide information about which genes that are regulated differently.

2.3 Network Theory

A network is defined as a collection of nodes and connecting edges (also referred to as links) [3]. The nodes can represent everything from locations and people to species, metabolites and genes. Edges are how the nodes are connected, such as interactions, processes or social relationships. Together the nodes and edges form a graph, which is the principle representation of networks [3]. Fig. 2.1 show how a neighborhood can be explained as a network and presented in form of a simple graph. The different houses is represented as circular nodes and the roads between is represented as edges connecting the nodes. In this example, two households are neighbors if there is a direct road between them, not interupted by other households. The graph in Fig. 2.1 is an example of an undirected network, i.e. the direction does not matter. If one household are neighbors with another household the same is true vice versa. For directed networks the opposite is true. Examples of directed networks can be bus routes, or a gene regulating another protein but not the other way around. In these cases edges are often represented as arrows pointing from one node to another. The size of the network (N) correspond to number of components, i.e. nodes in the network. Nodes are labeled with $i = 1, 2, ..., N$. A node pair is denoted i, j, and an edge between a node pair is denoted $E(i, j)$. From Fig. 2.1 the edge between node 1 and 3 is labeled $E(1, 3)$ [3].

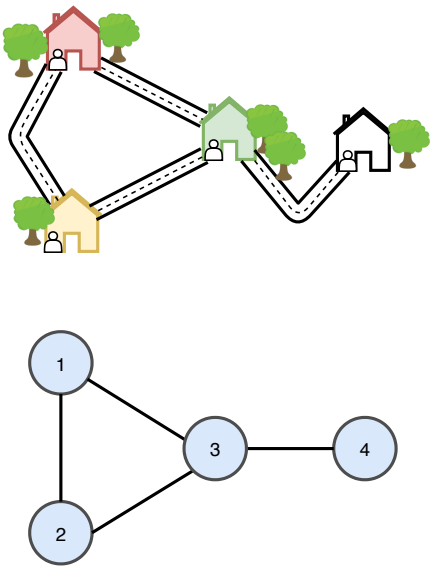


Figure 2.1: Graph representation. The figure illustrates how a neighborhood is a network and can be visualized in a simple graph representation. Four nodes represent the four houses and the edges represent the roads between them.

2.3.1 Adjacency Matrix

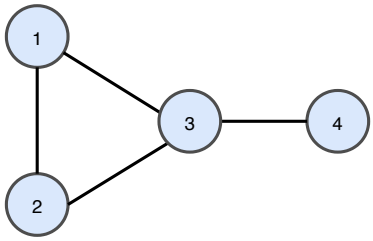
A network can be represented through its *adjacency matrix*, which holds information about all the pairwise connection in the network. The adjacency matrix (A_{ij}) for a network with N nodes will have N rows and N columns [3]. The elements of the matrix, a_{ij}, represent the pairwise connections between nodes i and j as shown below.

$$A_{ij} = \begin{matrix} \mathrm{a}_{11} & \mathrm{a}_{12} & \mathrm{a}_{13} & \mathrm{a}_{14} \\ \mathrm{a}_{21} & \mathrm{a}_{22} & \mathrm{a}_{23} & \mathrm{a}_{24} \\ \mathrm{a}_{31} & \mathrm{a}_{32} & \mathrm{a}_{33} & \mathrm{a}_{34} \\ \mathrm{a}_{41} & \mathrm{a}_{42} & \mathrm{a}_{43} & \mathrm{a}_{44} \end{matrix}$$

For an *unweighted* network the elements of the matrix will either be $a_{ij} = 1$ if there exist an edge between node i and j and $a_{ij} = 0$ if not [3]. An adjacency matrix for the neighborhood network (see Fig. 2.1) is presented in Fig. 2.2. Nodes 1 and 3 are connected by an edge and thus $a_{1,3} = 1$. Nodes 1 and 4 however are not connected and $a_{1,4} = 0$. In an undirected network, like the one presented in Fig. 2.2 an edge will be represented twice; $E(1, 4)$ and $E(4, 1)$. The adjacency matrix of undirected networks will therefore be symmetric; A_{ij} = A_{ji}. The opposite is true for directed networks where edges have

directions.

However, sometimes we need to study *weighted* networks were the edges between node i and j has an unique weight w_{ij}. In the example from Fig. 2.1, instead of just considering which households are neighbor's, one could also consider the strength of their relationships. Household who has stronger relationship with one another will thus have more weighted edges between them. The adjacency matrix for such networks will hence have elements of all values.



$A_{ij} =$

	1	**2**	**3**	**4**
1	0	1	1	0
2	1	0	1	0
3	1	1	0	1
4	0	0	1	0

Figure 2.2: Network and its corresponding adjacency matrix. An undirected network and its corresponding adjacency matrix A_{ij}. The bold numbers represent the nodes and each element represent a pairwise connection in the network (1 indicating an edge between them, and 0 if not).

2.3.2 Degree

A key property for every node in the network is its *degree*. The degree for one node i is denoted k_i and represent the number of edges it has to other nodes [3]. The degree can describe the number of countries bordering other countries, or the number of proteins one protein interacts with. In the neighborhood example (Fig. 2.1) the degree represent the number of neighbors one household has ($k_1 = 2, k_2 = 2, k_3 = 3, k_4 = 1$).

For directed networks the degree is distinguished between in-degree (k_i^{in}) and out-degree (k_i^{out}). In-degree is the number of edges going from other nodes to node i, opposite for out degree. In-degree can describe the number of incoming phone calls while out-degree describe the number of outgoing phone calls. The total degree of node i in a directed network is $k_i = k_i^{in} + k_i^{out}$. [3] Looking at the properties of the whole network, *average degree* is an important parameter. Average degree is defined as

$$\langle k \rangle = \sum_{k=0}^{\infty} k P_k \tag{2.1}$$

where P_k is the probability that a random selected node has degree k [3].

2.3.3 Degree Distribution

The *degree distribution* $P(k)$ for a network, provide the probability P of a randomly selected node having k connecting edges. The degree distribution can be visualized by plotting number of nodes as a function of degree. For a network with N

nodes the degree distribution is defined by

$$P_k = \frac{N_k}{N} \tag{2.2}$$

where N_k is the number of k-degree nodes [3]

Networks generally fall into one of two classes; bell shaped or heavy tailed degree distribution. A bell shaped degree distribution, also called a normal distribution, is characterised by its symmetry and peak in the middle. They are often found in random, social and communication networks. Heavy tailed distributions are more abundant in biological and technological networks [20]. Networks with a heavy tailed distributions are scale free, meaning that their network distribution follows a power law. A power law describes a functional relationship between two quantities where one quantity varies as a power of the other. It can be defined by

$$P(k) \sim k^{-\gamma} \tag{2.3}$$

where γ is the characterizing exponent. In scale free networks few nodes have many connections while the majority of nodes have few. An important difference between bell shaped and heavy tailed distributions is that the average degree of a node in an heavy tailed distribution does not characterize a typical node [3, 20].

These high degree nodes found in scale free networks are called hubs. *Hubs* are nodes with a degree (k) which exceeds the average node degree in the network [3]. In biological networks it is shown that the deletion of these hubs are more likely to have lethal consequences for the functionality of the process described by the network. This phenomenon is believed to reflect the essential and central roles of the hubs, making them some of the most important components in the network [21]. A network with central hubs may be very fragile of robust, depending on the attacks being random or strategic. A random attack will have a lower chance of attacking the hubs as there are so few of them, while a strategical attack can go for the hubs and dissolve the network in few moves [3]. Looking at the degree distribution of a network thus provide valuable insight into the robustness of the network.

2.3.4 Assortativity and Disassortativity

How the hubs connect to other nodes, especially other hubs, are also important aspects to look at when studying networks. Networks can be divided into three categories based on how the hubs connect to each other; neutral networks, assortative networks and disassortative networks. In neutral networks hubs are connected to each other in such a way that

can be expected to happen by chance. In assortative networks, hubs tend to link to each other and avoid linking with low degree nodes. The opposite is true for disassortative networks, where hubs tend to link to low degree nodes and avoid other hubs. In other words, networks display degree correlations if the number of links between high- and low-degree nodes differ from what is expected by chance [3]. While social networks tend to be mostly assortative, biological and technological networks tend to be disassortative [22].

If a network is assortative or disassortative can be identified by quantifying the degree correlations. One way to accomplish this is to measure, for each node *i*, the average degree of its neighbors (*Average neighborhood connectivity*). This gives the degree correlation function and its approximation

$$k_{nn}(k) = \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j \approx a k^{\mu} \tag{2.4}$$

where $\mu < 0$ mean disassortative, $\mu = 0$ mean neutral and $\mu > 0$ mean assortative [3]. A visualization of the degree correlation function for the three types of networks is visualized in Fig 2.3. The green dashed line illustrates the regression line for the data, and the black line is what would be expected if the degree correlation was completely random.

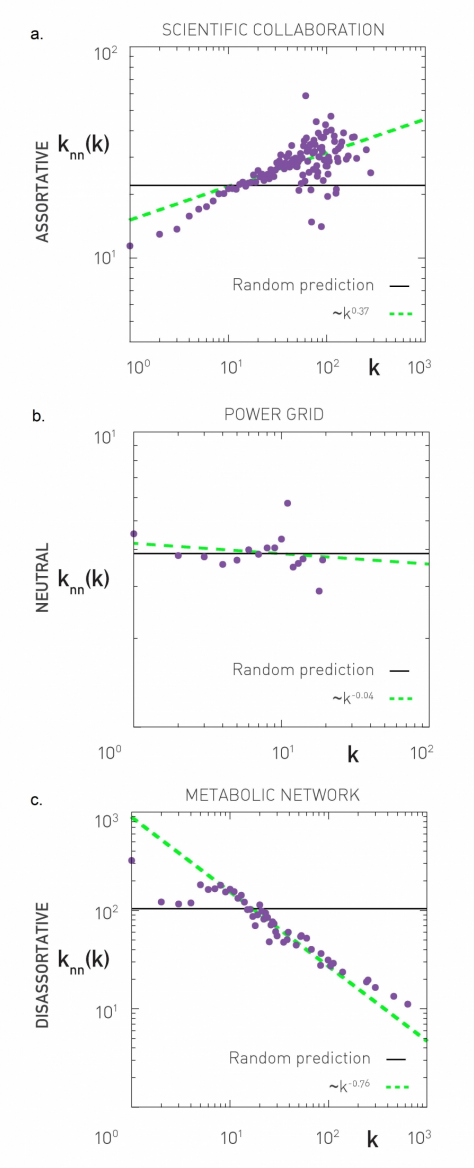


Figure 2.3: Degree correlations. The degree correlation function $k_{nn}(k)$ for three real networks. An assortative scientific collaboration network ($\mu = 0.37$), a neutral power grid network ($\mu = -0.04$) and a disassortative metabolic network ($\mu = -0.76$). Green dashed line gives the regression line for the data points, and black line illustrate the prediction if degree correlation was random. Taken from [3].

2.3.5 Paths

Another interesting aspect in networks is the distance between the components. However, for networks, distance is difficult. How can distance be measured for a relationship or regulatory effect between components of a network. It is not the physical distance that is relevant, but the *path length*.The path of a network is defined as a series of edges connecting two nodes [23]. The length of the path is thus the amount of edges in the path [3]. Examples of different paths between node 1 and node 4 from the graph presented in Fig 2.1 is presented with red and orange if Fig. 2.4. The red path has a path length equal to three and the orange a path length equal to two. The orange path has the shortest amounts of edges in it and is therefore the shortest path. The *shortest path* between node i and j in a network is defined as the path between them with the minimal number of edges.

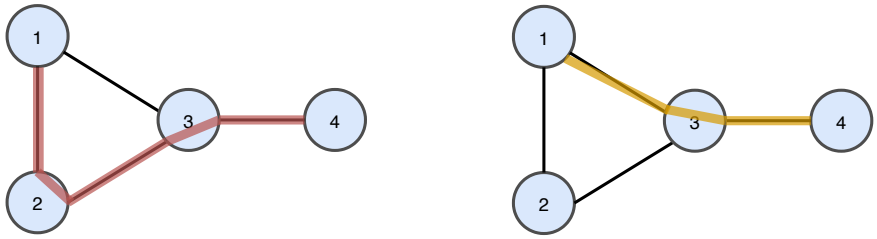


Figure 2.4: Paths. The different non-overlapping paths in the network from node 1 to node 4, shown in red and orange (shortest).

2.3.6 Clustering

A key topological aspect in networks is clustering. Looking back at the neighborhood in Fig. 2.1 one can ask; ”how many of my neighbors are neighbors with each other”. For node 3, the answer will be two (node 1 and 2), and the three nodes can be said to be a cluster in the network. The *clustering coefficient* is a nodes tendency to form clusters and is defined by

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \tag{2.5}$$

where L is the total amount of edges and $k_i(k_i - 1)$ present the possible number of edges. In other words, it captures the degree to which the neighbors of a given node link to each other. In the case of nodes with degrees $k_i = 0, 1$ the clustering coefficient is zero. C_i can have values from 0 to 1, where $C_i = 0$ if none of the neighbors of node i link to each other, $C_i = 1$ if all the neighbors of node i link to each other. The *average clustering coefficient* captures the degree of clustering for the entire network, and is and defined by

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{2.6}$$

and represent the average of C_i for all nodes $i = 1, 2...N$ [3].

2.3.7 Betweenness Centrality

Centrality in a network is used to describe how important a node is. Several ways of calculating centrality exists, one of them being the calculation of *betweenness centrality*. Betweenness centrality (BC) is a measure on how many times node i acts as a bridge along the shortest path between two nodes. It thus captures the nodes role in allowing information passing from one part of the network to another [3, 23]. In Fig. 2.4 node 3 is the node with highest betweenness centrality. BC can be applied on edges as well. Edge betweenness is calculated by finding all the shortest paths between node pairs and counting how many of them that run along each edge [24].

2.3.8 Communities

In network theory, a group of nodes that are more likely to connect to each other than to nodes from other groups are called communities. [3]. However, a community does not have a clear definition, but can be related to a set of hypotheses.

- **H1: Fundamental hypothesis**
  "A network's community structure is uniquely encoded in its wiring diagram."
- **H2: Connectedness and Density Hypothesis**
  "A community is a locally dense connected subgraph in a network."

The first hypothesis state that the way communities are structured in a network can be uncovered by inspecting its adjacency matrix A_{ij}. The second hypothesis states that each community correspond to a connected subgraph. It thus limits the definition by requiring that all nodes in the community must be reachable from other nodes in the same community. However, several community definitions are consistent with these hypotheses.

Three examples of communities are presented in Fig 2.5, all fulfilling the criteria from H2. The three communities are; *cliques*, *strong communities* and *weak communities*. A maximum clique (Fig.2.5 a) is a subgraph where all the nodes are connected to each other. This is a rather restrictive definition of communities and may miss other legit communities. A strong community (Fig 2.5 b) is defined as a connected subgraph whose nodes share more edges with the nodes within its community than with nodes belonging to other communities. Lastly, a weak community (Fig 2.5 c) is a subgraph whose communities total number of internal edges exceeds their total external number of edges. What this figure illustrates is that the identifications of communities depends on the definitions used.

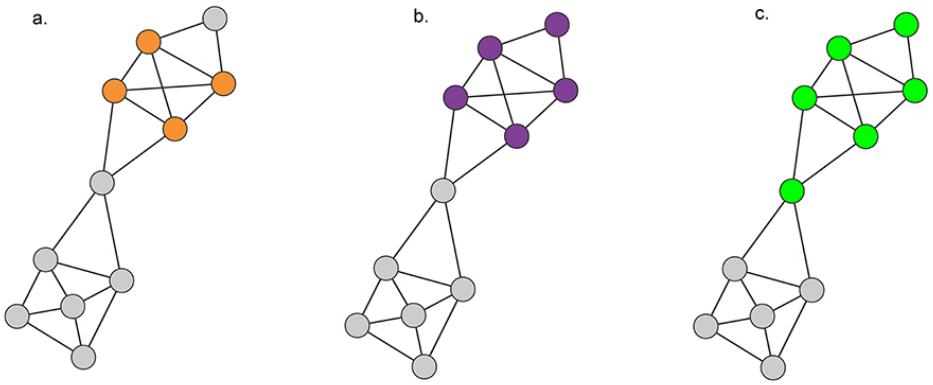


Figure 2.5: Network communities. An illustration of three different definitions of communities. (a) maximum clique (orange), (b) strong community (purple), (c) weak community (green). Taken from [3].

Two more hypothesis exist for communities.

- **H3: Random Hypothesis**
  "Randomly wired networks lack an inherent community structure."
- **H4: Maximal Modularity Hypothesis**
  "For a given network the partition with maximum modularity corresponds to the optimal community structure."

Both the hypotheses are related to the concept of *modularity*. The third hypothesis suggest that by comparing the density of the edges in a community with edge density obtained from the same group of nodes in a random network, it can be identified if the community has occurred by chance or not. This in turn determines the quality of the community, i.e. the modularity (Q_c) which can be defined as

$$Q_c = \frac{1}{2L} \sum_{(i,j) \in C_c} (A_{ij} - P_{ij}) \tag{2.7}$$

where L is the total number of edges in the network. $P_{ij} = \frac{k_i k_j}{2L}$ is the probability of a random connecting edge between nodes of degree k_i and k_j. Modularity can be written on a simpler form

$$Q_c = \frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \tag{2.8}$$

where L_c is the total amount of edges in community C_c and k_c being the total degree of nodes within the community. This equation can then be generalized to calculate the overall modularity score for a full network that has been divided into n_c communities.

$$Q = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \right] \tag{2.9}$$

2.4 Gene Co-expression Analysis

A key goal in biological research is to systemically identify how molecules in a living cell interact. The development of high-throughput technologies such as RNA sequencing (RNA-seq) has provided huge amounts of gene expression data that now can be analyzed. A common way to represent the pairwise similarity between RNA levels in data sets is by constructing gene co-expression networks. In a gene co-expression network (GCN) nodes represent genes and edges represent the significant co-expression that exist between the genes. Genes that are co-expressed will rise and fall together across samples. GCN therefore provide valuable information about which genes that are active simultaneously, which may indicate that they are involved in the same biological processes. It may also indicate that the genes are localized at the same place, or may be involved in the same complex. In a situation where gene co-expression is studied in patients, if some genes are found to be highly co-expressed this may imply that they are central in the disease.

A GCN connects pair of nodes (genes) that are significantly expressed across samples in a study. Several methods have been developed for constructing such networks, but they all follow the same principle. First, a similarity measure is defined and calculated for each pair of genes from the gene expression data. Next, a threshold is determined and the gene pairs with a similarity score higher than the selected threshold are considered to have a significant co-expression relationship. These are thus connected by an edge in the network, and a GCN is fully constructed. Next, the network can be analyzed by detecting communities, which is also referred to as *modules* in relation to gene co-expression analysis. Then gene ontology (GO) enrichment analysis can be performed in order to determine which biological processes that are enriched in these modules. An overview of the construction and analysing steps is presented in Fig. 2.6

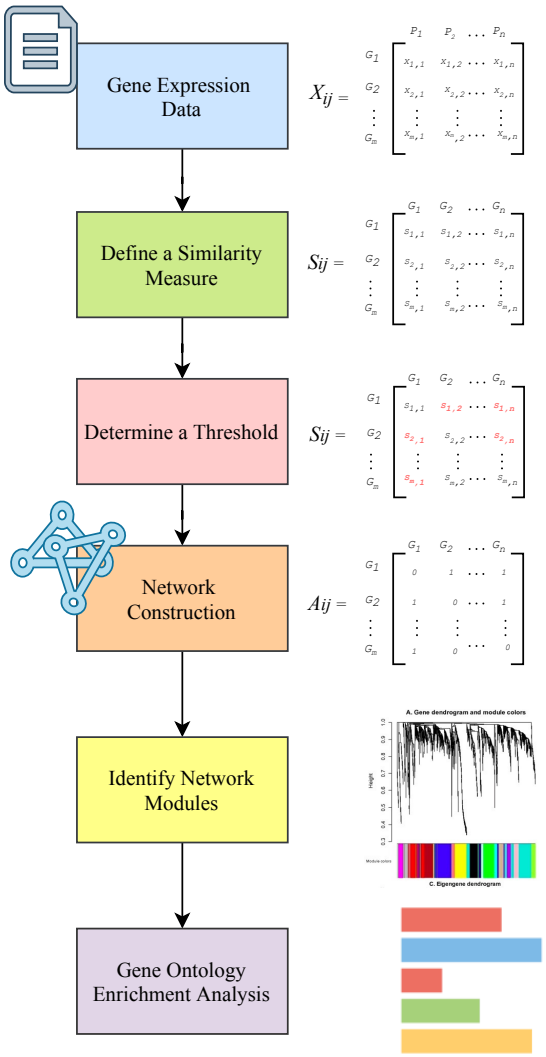


Figure 2.6: Gene co-expression network analysis. A general overview of the steps for constructing and analyzing a gene co-expression network. First, a chosen similarity score is calculated for each possible gene pair in the expression data. Next, a threshold value is defined, and the pairwise genes with a similarity score higher than the threshold are represented with an edge between them in the network. Then, for the analysis, network modules can be detected by for example hierarchical clustering. Gene ontology enrichment analysis can then be applied to the modules. A typical figure is presented next to each step, and the steps are explained in the following sections.

2.4.1 Similarity Measures

The overall goal of gene co-expression analysis is to identify genes having a tendency to show coordinated expression patterns across different samples. This is achieved by identifying co-expressed genes and grouping them together by assigning individual relationships between them. These relationships are based on a measure of similarity found in their expression levels across multiple samples. Defining a similarity measure is therefore the step to begin with in co-expression analysis. There are several similarity measures to choose from, and the chosen similarity measure can have large impact on determining the types and amount of knowledge gained from the analysis. In addition, gene expression data is by nature noisy, and thus requires robust methods for recognition of biological patterns [25]. The chosen similarity measure, denoted s, is calculated for each pair of gene i and j, and arranged in a similarity matrix $S_{ij} = |s_{ij}|$. Often, genetic relationships have been studied using correlation measures, such as Pearson or Spearmans ranks correlation measures. Another common measure is the topological overlap. An explanation of these measure are described next.

Correlation

Correlation refers to the extent of which two variables relate to each other [26]. Examples of such dependent phenomenons are the correlation between product quality and price or obesity and the risk of developing type II diabetes. Correlations are useful as they can be used to indicate predictive relationships that can be exploited in practise. However, it is important to remember that correlation does not mean causation. Also, correlation shows the strength of a linear relationship, but the value itself does not completely characterize their relationship. The two primary types of correlation coefficients are Pearson's correlation and Spearman's rank correlation. While Pearson's correlation is sensitive only to linear relationships between two variables, spearman's correlations is more robust and sensitive to nonlinear relationships as well [26].

The spearman rank-order correlation coefficient is an nonparametric measure of the strength and direction of association that exists between two variables. It is measured on a scale from least to most. There needs to be a monotonic relationship between the two variables, i.e. either the variables increase together, or as one increase in values the other decreases. The spearman's correlation is denoted rho and is defined as:

$$\rho = 1 - \frac{6 \sum d^2}{n^2(n-1)} \tag{2.10}$$

where d is the difference between the two ranks of each observation and n is the number of observations. The variables rank is either higher, lower or equal to the variable it is compared to. The spearman rank coefficient have values from -1 to 1. A correlation score of exactly -1 will indicate a perfect downhill (negative) linear relationship. The opposite is said for a correlation score of exactly 1 which indicates a perfect uphill (positive) linear relationship. A score of zero indicates no correlation (non linear relationship) [26].

Weighted Topological Overlap

Topological overlap as a similarity measure have been found useful in biological networks [27]. Nodes that are similar should have a high similarity score. and in networks, nodes that are similar often connect to each other and share the same neighbors. Therefore, ensuring that similarity is high for these genes can be done by using topological overlap.

Topological overlap is the fraction of number of neighboring nodes shared by two nodes i and j. The weighted topological overlap measure (wTO) is defined by

$$w_{ij}^{TO} = \frac{\sum_k a_{ik}a_{ij} + a_{ij}}{min(k_i, k_j) + 1 - |a_{ij}|} \tag{2.11}$$

where k represent the nodes connectivity and equals the sum of its weighted connections. The node connectivity is defines as

$$k_i = \sum_{j=1}^{n} |a_{ij}|. \tag{2.12}$$

Note that $w_{ij} = 1$ only if two conditions are satisfied. These two conditions being that all neighbors of node i also are neighbors with node j and there is a direct link between node node i and j. Contrarily, $w_{ij} = 0$ only if there is no direct link between i and j are unconnected, and the two nodes do not share any neighbors.

When wTO is used as a similarity measure in gene co-expression analysis, it is calculated using the correlation measure for all pair-wise gene combinations. Correlation scores can have both positive and negative values, while the $1 - |a_{ij}|$ in the denominator in (2.11) makes it impossible for w_{ij} to obtain a negative value. In order to include both positive and negative values, wTO can be multiplied with the signs of the correlation values.

2.4.2 Thresholding

Once the resulting similarity matrix have been constructed the next step is to define which similarity score that are of interest in order to construct the network. The scores of interests is defined by defining a threshold value, were only the scores over a certain threshold value will be included in the network.

The most frequently used threshold value is the signum function which implement a hard threshold value [28]. The hard threshold value involves the threshold parameter τ and is defined by

$$a_{ij} = signum(s_{ij}, \tau) \equiv \begin{cases} 1 \text{ if } s_{ij} \geq \tau \\ 0 \text{ if } s_{ij} < \tau \end{cases} \tag{2.13}$$

Hard thresholding is commonly applied on correlation measures. While hard thresholding is an intuitive method, it also has its drawbacks. By setting $\tau = 0.80$, correlation scores of 0.79 wont be included in the network, which can result in loss of information.

For weighted topological overlap calculations it has instead recently been suggested to use soft thresholding on the correlation data [28]. One way to implement soft thresholding is by using the power adjacency function

$$a_{ij} = power(s_{ij}, \beta) \equiv |s_{ij}|^{\beta}. \tag{2.14}$$

This function includes only one parameter; β which is used to raise the correlation values with the power of its value. By raising the correlation values with a certain power, higher correlation scores become more prominent. Potential drawbacks with soft thresholding is however that it does not make it clear how to define the directly linked neighbors of a node, and it is usually necessary to apply a hard thresholding parameter after doing the soft thresholding [28].

2.4.3 Identify Network Modules

Once a network is constructed from the large amount of expression data, the next step is to extract valuable information from the thousands of interactions present. In addition to looking at degree distribution and assortativity mentioned earlier, modules (or communities) provide valuable information in gene expression networks. By grouping together genes with similar expression data, genes whose expression profiles are highly correlated across samples can be visualized. However, to uncover module structures for large networks, several tools and algorithms must be applied.

On of the most widely used methods for analyzing patterns in networks and assigning genes to modules is *hierarchical clustering*. In hierarchical clustering genes are represented by a tree whose branch length reflects the degree of similarity between the objects. The hierarchical clustering use a pairwise similarity measure function such as those described above; wTO and correlation [29]. In addition to a similarity measure, one has to define group similarity. Several approaches exist, one being the *average cluster similarity*, which defines similarity of two modules as the average of s_{ij} over all node pairs i and j that belong to distinct modules [3].

Once similarity and group similarity measures have been defined the hierarchical clustering can be performed. There are two ways of performing the hierarchical clustering. One option is to use *agglomerative algorithms* which merges nodes with high similarity into the same module. The other option is *divisive algorithms* which isolate modules by removing edges with low similarity that connects to the modules. Both of the algorithms generate a dendrogram [3]. The hierarchical clustering for any set of n genes in an agglomerative algorithm follows these steps:

1. The similarity matrix is scanned to identify the highest value (representing the most similar pair of genes).
2. The most similar pair of genes are merged together into one cluster.

3. The similarity matrix is then updated with the average cluster similarity of these two genes replacing the two joined elements.
4. The process starts over again with step 1 and is repeated $n - 1$ times until only a single element remains.

The pairwise merge of step 3 will eventually merge all nodes into one large module. A dendrogram can therefore be created to extract the underlying organization of these modules. A set of branches will represent each module created during this process. The first module that was created contained the two most similar pairwise genes and is presented as the highest branch. Thus, the highest branches of the dendrogram represent the modules with highest similarity [29, 3]. For a divisive algorithm, the general step-wise approach is the same. However with step 2 being the removal of a low similarity link until the network is split up into several modules.

2.4.4 Gene Ontology Enrichment Analysis

To test whether the identified communities are biologically meaningful, gene ontology information (functional enrichment analysis) can be used [30]. Gene ontology is a bioinformatics initiative which provides a structured, controlled vocabulary that can be applied across all eukaryotes when annotating the function of genes and gene products [31, 32]. Genes annotated with the same GO terms are structured into the same category and share common biological functions.

The ontology covers three independent domains; biological processes, molecular processes and cellular components. [31]. In gene ontology enrichment analysis, the desired domain is chosen and GO terms within this domain is annotated to the genes. The enrichment of the GO terms on genes in the network communities can indicate if they are involved in the same biological functions. It can also indicate the associations of biological function to the variation of co-expression across experimental conditions. All GO ontology files are freely available on the GO website (`http://geneontology.org`), or can be accessed using several online GO browsers.

2.5 Differential Gene Co-expression Analysis

While gene co-expression analysis study the co-expression of genes across different methods, a more recent approach have been developed when analysing co-expression of genes. The recent approach referred to is differential co-expression analysis, which considers the changes in co-expression between different conditions [33, 34]. The comparative conditions can be tissue types, disease state or development stages. By comparing the expression profiles from two experimental conditions, it becomes feasible to to detect expression patterns differing in the two experiments. Thus, by comparing pairwise co-expression between genes in patients with healthy individuals the genes whose change in gene expression cause changes in phenotype can be identified.

The construction of differential gene co-expression networks follows the same approach as for the construction of GCN described earlier. Expression data from two different con-

ditions is used to calculate a similarity score. The pairwise gene combinations with a similarity score above a defined thresholds are then defined as one edges in the network. Finally, different analysing methods can be applied in order to identify significant genes or modules in the network.

2.6 The CSD Method

The CSD framework for differential co-expression network analysis has been developed by André Voigt and Eivind Almaas and is described in the article ”A composite network of conserved and tissue specific gene interactions reveals possible genetic interactions in glioma” [4]. The CSD framework aims to close a gap in the existing methods for differential co-expression analysis, namely the inability to distinguish between various forms of differential co-expression. This is obtained by categorizing the differential co-expression according to mathematically defined scores in three categories; conserved (C), specific (S) and differentiated (D).

- Conserved (C) co-expression is when strong co-expression exists between gene pairs in both condition.
- Specific (S) co-expression is when co-expression between gene pairs is present under only one condition.
- Differentiated (D) co-expression is when there is strong co-expression between gene pairs in both conditions, but with opposite signs.

A differential co-expression network is constructed where nodes represent genes, and the edges represent one of the three interaction types. Figure 2.7 illustrates how the different scores are related to co-expression values in the two different conditions.

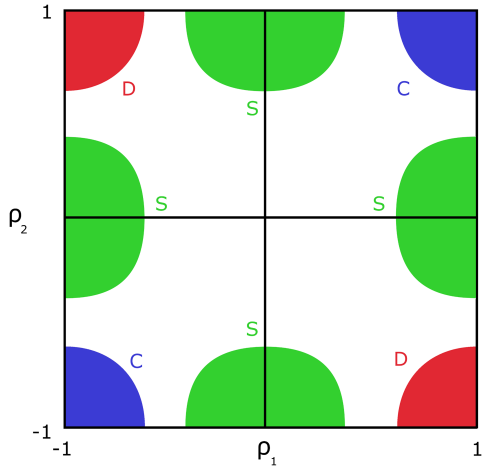


Figure 2.7: CSD diagram: General representation of the differential co-expression of gene pairs obtained with the CSD framework. ρ_1 and ρ_2 denote the Spearman rank-correlation of the expression for a gene pair under condition 1 and 2 respectively. Blue area is for conserved C (Strong co-expression in both conditions), green area is specific S (strong co-expression under only one condition) and red area is differentiated D (Strong co-expression in both conditions, but with opposite signs). White area represent correlations that does not yield an edge in the network. The different co-expression relationships are also presented with their colored letters. Taken from [4].

2.6.1 Gene Relationships Scores

The CSD framework use gene expression data originating from different tissue, species or conditions to construct the differential gene co-expression network. The similarity score s_{ij} for each pairwise combination of genes i and j present in the two datasets (representing the two different conditions) k is calculated separately. This measure is calculated using the Spearman rank-correlation, $\rho_{ij,k}$. The similarity score can have a value of -1 to 1. A score of -1 indicate a strong negative correlation, a score of 0 indicate weak or no correlation and a score of 1 indicate a strong positive correlation.

In the next step, the difference in co-expression between the two conditions are quantified. This is done by introducing three comparative gene co-expression scores to distinguish between the different types of co-expression. The co-expression scores are defined as

$$C_{ij} = \frac{|\rho_{ij,1} + \rho_{ij,2}|}{\sqrt{\sigma^2_{ij,1} + \sigma^2_{ij,2}}} \tag{2.15}$$

$$S_{ij} = \frac{\|\rho_{ij,1}| - |\rho_{ij,2}\|}{\sqrt{\sigma^2_{ij,1} + \sigma^2_{ij,2}}} \tag{2.16}$$

$$D_{ij} = \frac{|\rho_{ij,1}| + |\rho_{ij,2}| - |\rho_{ij,1} + \rho_{ij,1}|}{\sqrt{\sigma^2_{ij,1} + \sigma^2_{ij,2}}} \tag{2.17}$$

and are calculated for each pair of gene i and j in two different conditions, denoted as 1 and 2. $\sigma^2_{ij,k}$ is the estimated variance in the Spearman correlation coefficient for gene i and j in condition k. It is used to correct for confounding factors in the data set. More about how the variance is determined can be found in the article describing the CSD-method [4].

2.6.2 Threshold Values

The calculation of the CSD scores result in three interaction score for each gene pair. The next step is to create the adjacency matrix from the similarity measures by applying a threshold. However, the three co-expression scores follow different distributions, so to make them directly comparable they have to be mapped to a common scale. This is solved by determining threshold values for each interaction score corresponding to an importance level, p. This importance level is not based on a significant level, but on the probability of obtaining a given score from the distribution of the calculated scores. All co-expression scores where $C_{ij} > K^C_p$ will be kept, while those below this threshold will be excluded. The same goes for S_{ij} and D_{ij} with their respective threshold values. The $K^{C,S,D}_p$ and associated $p_{C,S,D}$ is calculated in the following way for M data points (Example shown for K^C_p and p).

1. The C score is calculated for all M gene pairs.
2. m samples, s_i, are drawn from the total set of C scores. Each sample has a size of $L << M$.
3. The threshold value for k^C_p is determined as average of maximal values per sample using the formula:

$$k^C_p = \frac{1}{m} \sum_{i=1}^{m} max_{\{s_i\}} C \tag{2.18}$$

4. The importance level p is then set to $p = 1/L$.

Choosing a common p for all three gene relationship scores makes it possible to obtain comparable threshold values and thus extract C-, S- and D-links that are combined in the final network.

2.6.3 Node Homogeneity

Nodes can be connected to other nodes by the different C,S or D interaction types. In order to distinguish nodes connected by one type of interaction types with nodes connected by multiple interaction types, node homogeneity (H) is used. Node homogeneity is defined by

$$H_i = \sum_{t \in \{C,S,D\}} \left(\frac{k_{t,i}}{k_i} \right)^2 \tag{2.19}$$

where $k_{C,i}$, $k_{S,i}$, $k_{D,i}$ is the number of C, S and D-type interactions node i has. k_i is the node degree i.e. the number of connections it has to other nodes. A node with $H = 1$ indicate that there is only one type of interactions while a node with $H = 1/3$ indicate that there is an even distribution.

Chapter 3

Materials and Methods

This chapter consider the materials and methods used for this thesis. It is divided into two parts where the fist sections 3.1 describe the pre-processing steps of the expression data and the implementation of the CSD-method. Then, in section 3.2 the set of steps for extracting more information from the network is explained. The workings of the different tools investigated are described in detail.

3.1 CSD on Lyme Disease

3.1.1 Data Collection and Processing

The gene expression data for Lyme disease was accessed from Gene Expression Omnibus (GEO) [35]. GEO is a database built by the National Center for Biotechnology (NCBI) [36], which archives and freely distributes several forms of high-throughput function genomics data.

The data was published as part of study done by Bouquet J, Soloski MJ, Swei A, Cheadle C et al: ”Identification of a Molecular Signature for Acute Lyme Disease by Human Transcriptome Profiling” [15]. The article was published February 15, 2016 and the data accessed on GEO (accession number GSE63085). The data consisted of total RNA sequencing measurements from blood from 84 patients and 13 healthy controls. The 84 patients were sampled at three time points: acute Lyme pre-treatment (V1), 3 weeks later, immediately following completion of a standard course of antibiotic (V2) and 6 months following treatment completion (V5). The 13 healthy controls were sampled at one time point. The RNA-seq data was normalised using FPKM.

The expression data from the patients and controls were downloaded in separate files, one file for each patient. In total 84 expression data files were downloaded from the patients, and 13 expression files for healthy controls. These files were merged together into two

complete dataset files, one for the patients and one for the controls. For a successful merge, all files needed an identical gene list. Additionally, the gene list needed to be identical between the control and patient data files. In order to obtain this all gene duplicates and all genes not found in both text files were removed.

Next, all samples from patients sampled at time point 6 months following treatment completion (V5) were removed from the data files. This was done as a result of the study concluding that 15 of the 29 patients at this time point had completely recovered. Also, even though the samples from V2 is taken after treatment with antibiotics three weeks later, the result from the study concluded that the differential gene expression signatures persisted for at least three weeks. These sample points were therefor not excluded from this study.

Finally, for many of the patients and controls, expression data was lacking for some of the genes, and the data therefore included many zeros. As a result from this, all data from patients containing rows with 30 or more zeroes were removed. In the control data, all rows containing 5 or more zeros were removed. Again, genes not found in both data sets after filtering was also excluded. The original data included expression data from 25278 genes, but after data processing and filtering it was reduced to 14805 genes.

3.1.2 Running the CSD Framework

The CSD framework is developed by Voigt and Almaas [4]. It contains three scripts that run separately, one written in C++ and two written in python, all available at GitHub [6]. The CSD framework iterates through all pairs of genes, computing pairwise correlation and estimate variance using each gene's expression vector. Spearman Rank correlation is used as similarity measure. The running of the CSD framework is presented in Fig. 3.1

The processes expression data from blood from patients with Lyme disease and healthy controls were used to generate differential gene-expression networks. In the first step, pairwise Spearman correlations scores and variance was calculated for all possible gene pair combinations. This was done by running the C++ code FindCorrAndVar.cpp on the two text files, one at a time. Some parameters in the code was changed relative to which text file that was run at the given time. Sample size was set to number of data points per gene which in this case was 55 data points for the disease data set and 13 for the control data set. The number of genes was set to 14805 for both text files.

The output data files contained all the correlation and variance values for the gene pairs which was further used when running the script FindCSD.py. This python script compared the correlation values from the two text files and calculated the C, S and D values. This was done by specifying the output data files from the last script as input data in this script. In this case, the sample size was relatively small for the control data, and previous study of the CSD-method have concluded that the sub-sampling can be ignored in such cases. The subsampling was therefore ignored by setting the variance (comboSD) to 1. The output from this script was four text files where three of the files contained either the usable C-, S- or D-values. The fourth text file contained all the correlation and C-, S- and D-values for all gene pairs.

In the third and final step, the C-, S- and D-values were used to generate four networks, one for each interaction type and one combined. This was done by running the python script CreateNetwork.py using the UseableCvalues, UseableSvalues and UseableDvalues text files as input. The interactions included in the networks were exclusively those above the given threshold value for each interaction type; $k^p_{C,S,D}$. The computed threshold was determined by the parameter in the code; SelSize, which was equal to 1/(desired p-value). In order to obtain a network with a size and density that makes it possible for further analysis, several selSizes were tested for the differential co-expression network. The selSize was set to 10^{-4}, $10^{-4.7}$ and 10^{-5} and the different networks were visualized in Cytoscape for inspection [37]. In the resulting network, nodes represented genes and edges types represented interaction types.

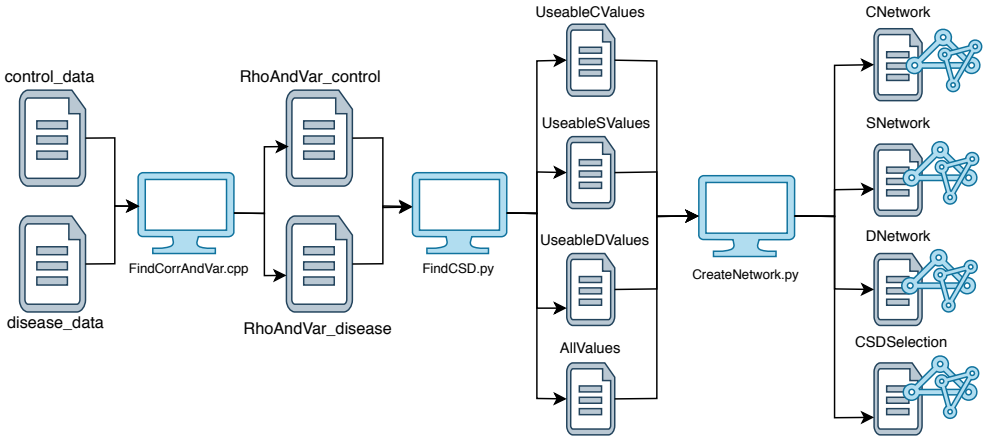


Figure 3.1: Running CSD. Workflow over the process of running the three scripts used in the CSD-method presented in [4]. The files represent text files used as input and obtained as output. The computers represent the three different scripts. Inspired by [5]

3.1.3 Alternative Approach; wTO

The gene expression data from the patients contained 55 sample points (patients) after filtering out samples from time point V5. The control data consisted of 13 sample points. However, 49 sample points or higher is recommended when running the CSD-method [4], and therefore the resulting network contained a lot of noise. In order to remove some of the noise, and to obtain better results, an alternative approach was applied. This approach was using using a different similarity measure, weighted topological overlap (wTO), to calculate the C, S and D interaction types. The wTO was calculated from the correlation data in the existing RhoAndVar files using the R-package *wTO* developed by Gysi et. al. [38]. A soft thresholding was applied to the correlation data, raising the correlations with power

of 5 ($\beta = 5$) in (2.14). The workflow for the alternative approach is presented in Fig 3.2. In order to calculate the wTO from the correlation data, the wTO R-package require the input to be in a matrix format. The RhoAndVar files are organised as edge lists containing the gene pairs and their respective correlation and variance values. A python script was therefore used to convert the the edge lists to matrices where the genes were set as column- and row names and each matrix element had the corresponding correlations. Variance was excluded from the matrix as it was not used further in the CSD-method.

The wTO() function from the R-package was used to calculate the weighted topological overlap. The equation for topological overlap (2.11) computes the connectivity of nodes. In a signed network the node connectivity can be cancelled out if its connected to nodes with positive and negative weights. In order to prevent this canceling out when calculating the node connectivity absolute values are used, ensured by setting the parameter sign = ”abs” in the wTO() function. The equation for topological overlap (2.11) also ensures that the calculated values to be between 0 and 1. Since we are interested in looking at both positive and negative correlations between gene expressions we want to add the signs to the data. This was achieved by multiplying the wTO values with the signs from the correlation data, while preventing the cancellation effect on the node connectivity.

When the wTO values were multiplied with the signs from the correlation data we ended up with two matrices; wTO_control and wTO_disease. These files were converted back to edge list format using another python script, and were then used to calculate the C-, S- and D-values in FindCSD.py. The following steps were performed as previously.

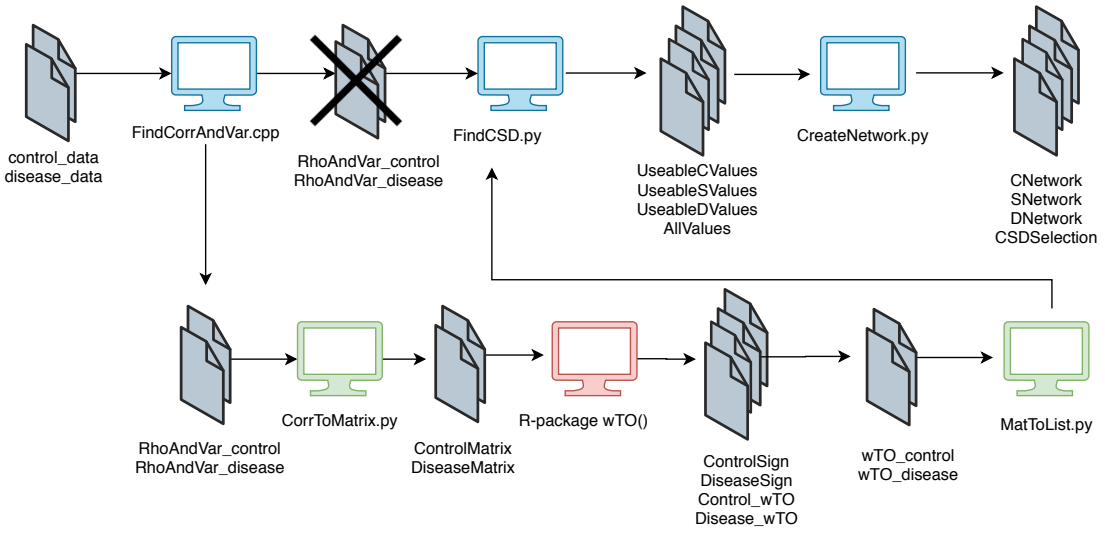


Figure 3.2: Alternative approach. Workflow representing the alternative approach using wTO in the CSD-method. The computers are presented in three different colors: blue for CSD-scripts obtained from [6] , green for self written python scripts and red for R-scripts. Inspired by [5]

3.1.4 Gene Ontology Enrichment Analysis

Investigating whether the genes of the network are related to specific biological processes provides valuable information. In order to to do this, a GO biological process enrichment analysis was performed using the Gene Ontology Consortium [31, 32](`http://geneontology.org/`) on the four networks. The biological process ontology was chosen for the search. The GO Enrichment Analysis tool takes a list of genes as input and present a table of biological processes where the genes related to this given process is either under- or over-represented in the network compared to what would be expected by chance. The results also return the expected number of genes and the fold enrichment score which says by how much the under- or over-representation is. The results were corrected by calculating the false discovery rate (FDR), and only terms with FDR < 0.05 were considered significant.

3.2 Tools for Network Analysis

3.2.1 Module Detection in Python

In order to depict relations among genes in the network, a set of analysing techniques can be applied. A common way is looking at the fundamental structures of the network such as degree, hubs and assortativity. Additionally, functional enrichment analysis provide valuable information about the biological processes enriched in the network. However, as

many biological networks typically are large, complex systems, a frequent approach is to decompose the network into sub-units, which are sets of highly interconnected nodes. The identification of such modules (or communities) are essential as they may uncover clusters of genes related to the same biological processes, that is yet unknown. Therefore, the first proposed step investigated, following network construction, was to partition the network into such modules. However, several methods exist for community detection in a network, and a variety of these were investigated.

The community detection tools investigated were found in the networkX package. NetworkX is a python package that enables exploration and analysis of networks [39]. Many graph algorithms are implemented in the python package for calculating network properties such as degree distribution, clustering, communities and many many more. The community detection algorithms used were; Greedy modularity maximization (GMM), Label propagation algorithm (LPA), Girvan-Newman method (GN) and Louvain community algorithm (LCA). The latter is not a package in NetworkX but it is dependent on it. All of these algorithms used the complete network presented as an unweighted edge-list as input. The different communities generated was written to separate text files. A flowchart over the progress is presented in Fig. 3.3.

The greedy_modularity_communities() function in NetworkX uses Clauset-Newman-Moore's greedy modularity maximization to detect communities in networks. It is an agglomerative algorithm and it is described in detail in: "Finding community structure in very large networks" [40]. In short, the algorithm is based on the greedy optimization of modularity Q. The modularity measure the density of links inside a community compared to links between the communities, and a good community is when there are many edges within and only few between them. The algorithm starts with each node being assigned to an individual community making it as many communities as there is nodes. Then, the algorithm repeatedly join together the two communities whose merging produces the largest increase in M.

The label propagation algorithm (LPA) is also agglomerative. The algorithm generates communities by adding a unique label (in this case colors) to each node in the network, and then for each iteration update each node by choosing the label most frequent among its neighbors. A network is identified as a connected group of vertices having the same final label. The algorithm is a semi–synchronous version of LPA, meaning that it allows for the simultaneously update of all the nodes belonging to the partitions of the networks. A more detailed description of this algorithm is presented in "Community Detection via Semi–Synchronous Label Propagation Algorithms" [41].

The Girvan-Newman algorithm detects communities based on centrality measure [42, 24], and differs from the other algorithms in that it is divisive. The algorithm removes edges one at a time based on betweenness centrality. The edges with highest edge betweenness score is removed first, following a new calculation of edge betweenness in each step. The network structure becomes divided into several communities as more and more edges are removed. The algorithm returns all the iterating steps; including first step where the whole network is in one community until the last step where every node is in an individual community. The algorithm does not define a cut value or calculate modularity, so in order

to get communities with an appropriate size for analysis, the iteration was run until the network was divided into 90 partitions.

The final community detection tool investigated was pythons Louvain Community Detection algorithm. Community detection was performed with the package function ”best partition” and the parameter ”randomize” was set to False. The Louvain method is described in detail in the article ”Fast unfolding of communities in large networks” [43]. This algorithm is very similar to the greedy modularity maximization algorithm described previously, in that it detects communities by optimizing modularity. The same definition of modularity (Q) is used here. As with the greedy algorithm the algorithm can be divided into two phases that are repeated iteratively. While the first phase is the same in both algorithms; assigning each node to its individual partition, the second step is different. While the greedy algorithm maximize modularity by merging communities, the Louvain algorithm evaluate the gain of modularity for every node i if the node is taken out of its community and placed in a neighbor nodes j community. The node i is then placed in the community where maximum gain is achieved as long as it is positive.

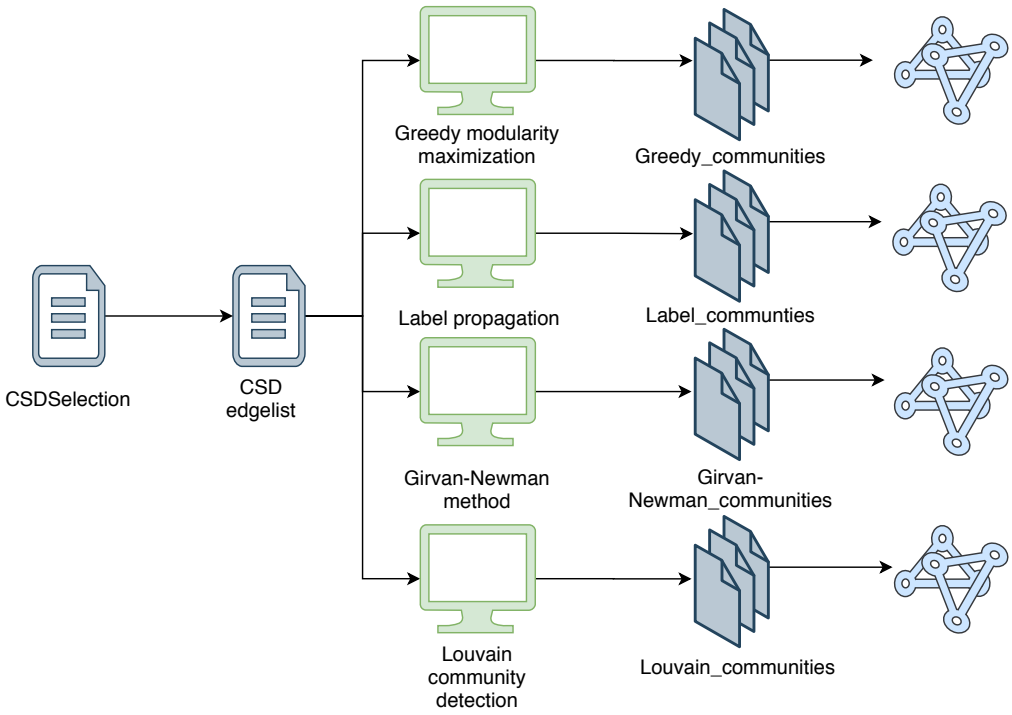


Figure 3.3: Community detection algorithms. The different community detection algorithms investigated, including both input and output files (.txt). The last step represent the communities visualized using Cytoscape. Computers with green color represent python scripts.

After running the community detection algorithms, the detected communities with 30 or more nodes were further investigated. The modules were imported as attribute tables to the CSD-network in Cytoscape. Then, as an additional step, the quality of the partitions detected by the different algorithms was determined by calculating their global modularity score. The score was calculated using the modularity function from NetworkX, which calculates the score from Eq. (2.3.8). Then, the next step performed was a functional gene enrichment analysis was performed on all the different communities using the Gene Ontology Consortium (`http://geneontology.org/`) [31, 32].

3.2.2 WGCNA

The CSD-network is constructed using weighted topological overlap as similarity measure. An interesting approach for extracting more information about the gene expression data from the Lyme disease patients was therefore to perform a weighted correlation network analysis, using the R-package WGCNA. WGCNA stands for weighted gene correlation network analysis and is a method developed by Steve Horvath and Peter Langfelder [30]. The R-package contain several functions and can be used in every step of gene

co-expression analysis. It provides functions for construction of the gene co-expression networks, module identification, and several ways of studying the relations between the detected modules.

Modules are defined as densely interconnected genes and there are several measures that can be used for defining interconnectivity. In WGCNA the default is set to topological overlap and the analysis was therefore applied on the wTO data from patients with Lyme disease. As this topological overlap matrix (TOM) has not been ran through the CSD framework it contains 14805 genes, and only positive values between 0 and 1. TOM was read into R and its corresponding dissimilarity calculated dissTOM = (1 - TOM). The highly co-expressed genes were then grouped by hierarchical clustering based on the dissimilarity measure. The hclust() functions was then used and set to ”average”, which entails that the clusters are compared by the average of each cluster.

The hierarchical clustering resulted in a dendrogram where each branch correspond to modules. The dynamic tree cut algorithm was utilized to cut the branches and to generate modules using the package *cutreeDynamic* [44]. The ’dynamic Tree cut’ first obtained a few large clusters from the dendrogram by defining each branch below a fixed height a separate cluster. Then, by looking at fluctuations patterns (see more detail in article [30]), clusters identified with sub-clusters were split. The function cuttreeDynamic() was then used with deepSplit = 2, minClustersize = 30 and distM = dissTOM. The rest of the settings were set to default.

The Dynamic Tree Cut may, however, identify modules whose expression profiles are very similar. It was therefore reasonable to merge these modules since their genes are highly co-expressed. In order to determine which modules should be merged, the eigengene, i.e. the first principle component for a given module, was defined. The co-expression between the eigengene of each modules were then compared and clustered based on a set cut height. The cut-height was set to 0.25 which correspond to a correlation of 0.75. Finally, the different modules were presented in a dendrogram. A flow chart of the progress is presented in Fig. 3.4.

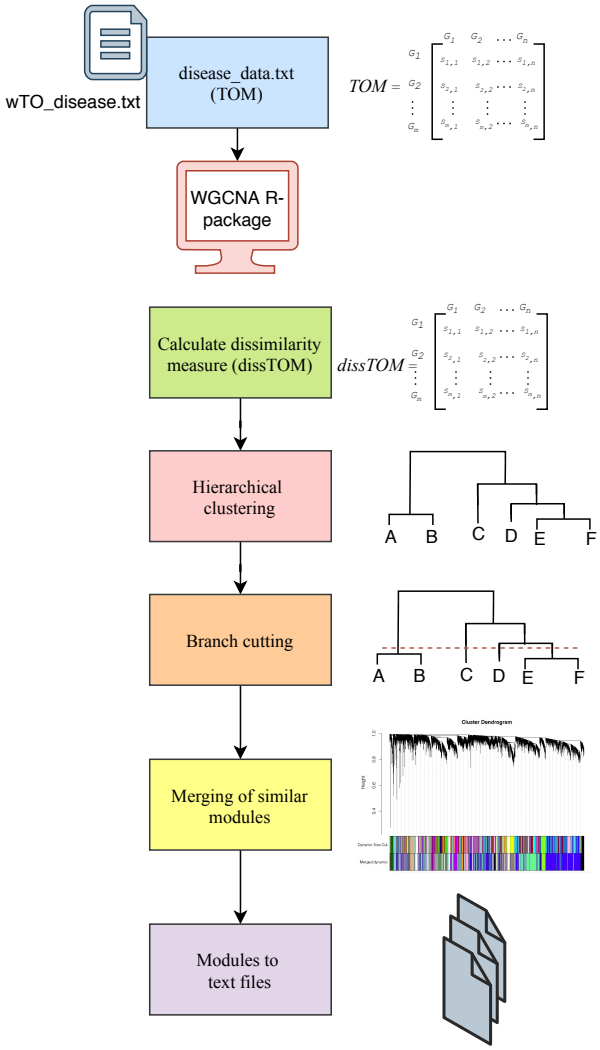


Figure 3.4: WGCNA workflow. A visual representation over the steps of running the WGCNA analysis on the topological overlap matrix from patients with Lyme disease. All the steps below the computer are performed in R using the WGCNA package. TOM is used as input, the dissimilarity measure is calculated and then hierarchical clustering is performed. Then branch cutting is performed to identify the communities, and a merge of similar communities is performed. All communities were written to separate text files.

To find out if the WGCNA analysis provided new insightful information about the network, a systematic investigation of the modules were performed. In the first step, biolog-

ical process enrichment analysis was performed on the modules using the Gene ontology Consoritum (`http://geneontology.org/`) [31, 32]. The modules with relevant biological processes were compared to the differentially co-expression network obtain from the CSD- network, in order to find which genes in the module that were also found to be differentially expressed in the CSD-network.

Chapter 4

Results and Analysis

This chapter regards the results from the analysis performed on the differential co-expression network constructed using transcriptomic data from Lyme disease patients. The results are divided into two sections, with the first section focusing on evaluating network aspects such as degree distribution, hubs, assortativity and functional gene ontology enrichment analysis. The other section regards the results from the analysing steps applied on the CSD-network and gene expression data, focusing on what new insight they provided about the network topology.

4.1 Application on Lyme Disease Expression Data

4.1.1 CSD-Network

The CSD framework was employed to construct a differential co-expression network from transcriptomic data obtained from blood samples from patients with Lyme disease and healthy controls. The processed dataset, used for construction of the network, consisted of gene expression measurements from 14805 genes with 55 samples from patients and 13 control samples. To ensure that the constructed network was suitable for analysis an importance value yielding a proper size and density was chosen. An importance value of $p = 10^{-4}$ yielded a network consisting of 3317 nodes, while an importance value of $p = 10^{-5}$ yielded a smaller network (847 nodes). Using an importance level of $p = 10^{-4.7}$ on the 14805 expressed genes resulted in a network consisting of 1323 nodes (genes) and 3865 edges (interaction types), which was a size suitable for further analysis. The network consisted of 1207 C-links, 1344 S-links and 1314 D-links. The C-link types are less prominent than the other two, indicating that many of the genes have different gene co-expression patterns in the patients compared to healthy controls. The network is visualized in Fig. 4.1 using Cytoscape.

The majority of the nodes and edges are connected in a giant component consisting of

1187 nodes (89.7%) and 3771 edges (97.5%). In addition to the giant component there are 54 other components, everyone consisting of nine nodes or less. A quick look at the network in Fig. 4.1 show that there is a highly interconnected and dense network where all three interaction types are present in the core . However, edges of same interaction types (either C, S or D) show a tendency of clustering together. This is especially prominent for the specific links, which create a highly dense and interconnected cluster. The specific types represent a strong co-expression relations between gene pairs which is completely condition dependent. As prominent as the specific links, but less dense, is the differentiated interaction types. They represent strong co-expression between gene pairs in both conditions where the sign of correlations change as the condition changes. A visualization of the networks for the individual C, S and D interaction types can be found in Fig A1 - A3.

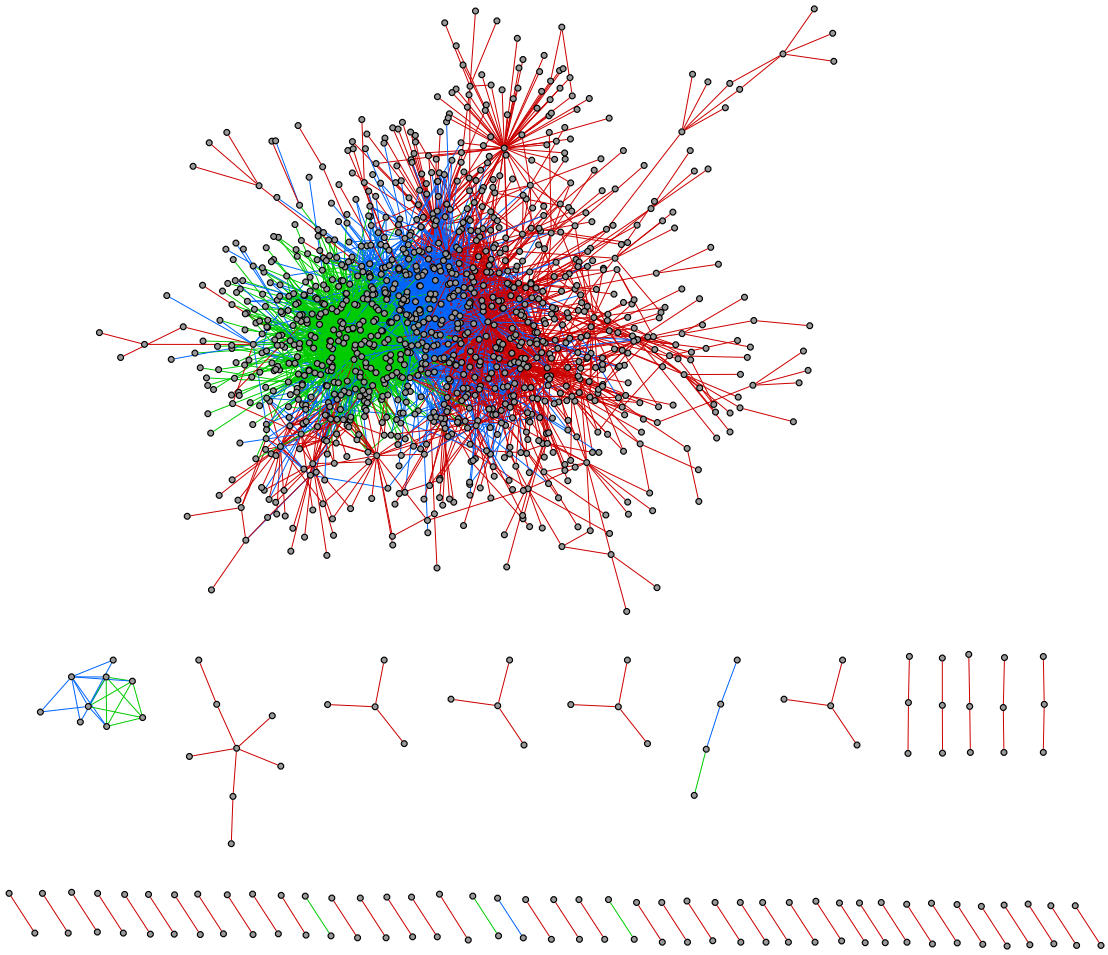


Figure 4.1: CSD-network. Visualization of the CSD-network generated using importance level $10^{-4.7}$. The network consists of 1323 nodes and 3865 edges. The edges are colored according to interaction types; blue is C-type, green i S-type and red is D-type.

4.1.2 Degree Distribution

The degree distributions for the CSD-network were plotted and inspected. The differential gene co-expressed network has a heavy tailed distribution which follows a power law, given by the equation $y = 287x^{-1.329}$. The correlation value of the fitted line is 0.998 ($R^2 = 0.858$). The degree distribution for the complete network on a logarithmic scale is presented in Fig.4.2. The good fit between the fitted power law and the data on a logarithmic scale indicate that the network is scale free, and thus contain a few centralized high degree nodes. This is very different from random networks, suggesting that the relation-

ships between co-expressed genes found in the network is of biological function.

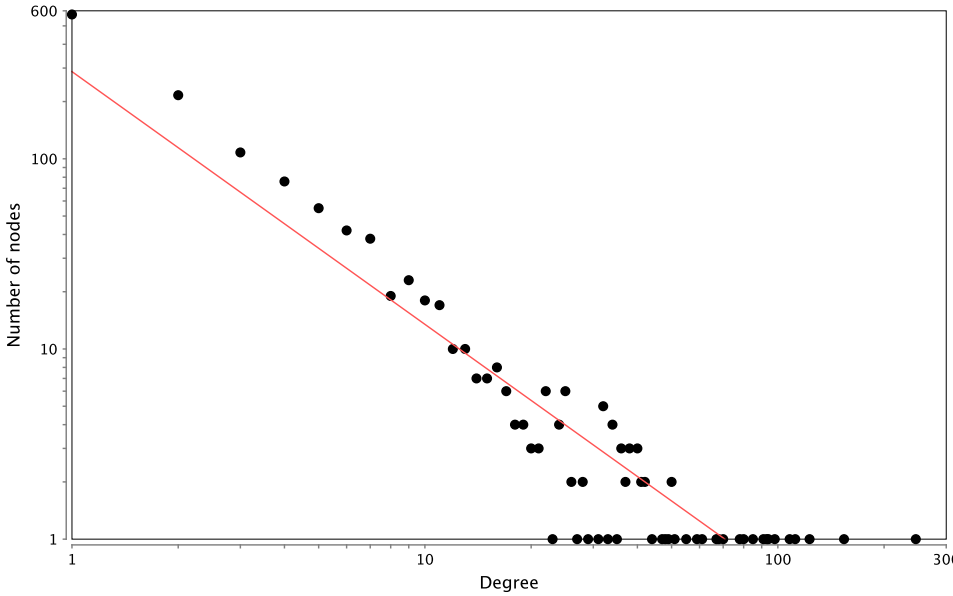


Figure 4.2: Degree distribution for the CSD-network. The distribution is plotted on a log-log scale with number of nodes as a function of node degree. The red line represent the fitted power law to the data points. It's expression is $y = 287x^{-1.329}$. The correlation of the fitted line is 0.998 ($R^2 = 0.858$.

4.1.3 Hubs and Assortativity

The network has an average node degree of $\langle k \rangle = 5.84$. In heavy tailed distribution networks, such as this one, the majority of the nodes are of low degree, while the are a few high degree nodes. The high degree nodes with a minimum of 50 neighbors are considered hubs in this network. The limit of $k \geq 50$ resulted in 21 hubs, which correspond to 1.59% of the nodes in the network.

The hubs are presented in Table 4.1 along with node degree (k), number of edges corresponding to each interaction type (C, S or D) and the node homogeneity score (H), calculated from (2.19). H is color coded based on the interaction type most typical for the hub, were blue is for conserved, green for specific and red for differentiated.

Table 4.1: Network hubs. Genes in the network with a minimum node degree of 50 or higher. k denotes node degree (total number of connections) and k_D, k_S and $k_D$is the degree for each interaction type. H is the node homogeneity score, calculated using (2.19). Colors represent the predominant link type (C = blue, S = green and D = red.)

Gene	k	k_C	k_S	k_D	H
GTF3C2	246	0	246	0	1.0
HKDC1	154	0	154	0	1.0
CCDC157	123	47	0	76	0.53
TNNC2	112	64	0	48	0.51
KHK	108	75	28	5	0.55
LOC440434	98	88	0	10	0.82
CD109	94	74	19	1	0.66
S1PR1	93	13	0	80	0.76
PURA	91	55	0	36	0.52
MMAA	85	0	85	0	1.0
ZNF702P	80	11	69	0	0.76
LRCH1	78	0	0	78	1.0
MBD2	70	54	11	5	0.62
TMEM161B-ASI	68	4	64	0	0.89
NAIF1	67	54	12	1	0.68
GSDMA	61	0	0	61	1.0
PARP14	59	1	58	0	0.97
OR6W1P	55	4	51	0	0.86
CCDC85C	51	36	0	15	0.58
PLCD4	50	20	23	7	0.39
MMP14	50	1	47	2	0.89

A quick look at the table provides information about the degree of the hubs and of what interaction types they are involved in. For nine of the hubs the majority of the interaction types are of the specific (S) type, meaning that the gene co-expression is present only under only one conditions. For four of the hubs, the majority of the interactions are of the differentiated (D) type. This indicates that the genes are strongly co-expressed under both conditions, but with opposite signs. The remaining eighth nodes have the majority of the interactions being of the conserved (C) type. These genes are part of strong co-expression under both conditions.

The hubs in table 4.1 are visualized in Fig. 4.3. The node size and color represent node degree (higher degree = larger node and darker color). Information about the biological functions of the identified hubs is summarized in Table A2, and some are elaborated in the following paragraphs. 5 of the hubs; CD109, S1PR1, LRCH1, GSDMA and PARP14 are involved in the immune system, mainly in activation and migration of T-cells and macrophages and pyroptosis. Several of the hubs are linked to DNA transcription and different regulatory mechanisms. Two of the hubs are a RNA gene and a pseudogene, and

thus do not encode any proteins.

The two largest hubs in the network is the GTF3C2 and HKDC1 gene, both located inside the dense interconnected cluster of specific interaction types. Both the genes are completely homogenous indicating that co-expression with all of their neighbors changes significantly in patients with Lyme disease, compared to the control. GTF3C2 has a degree of $k = 246$, and has almost 100 more neighboring nodes than HKDC1. GTF3C2 encodes a general transcription factor which with a family of GTF3C proteins assemble in a complex on the DNA promoter and recruit RNA polymerase III. These transcription factors are essential for RNA polymerase III to make a number of small nuclear and cytoplasmic RNA, which includes 5S RNA, tRNA and adenovirus-associated (VA) RNA of both cellular and viral origin. Transcription by RNA polymerase III is enhanced during viral infection by the expression of immediate early protein of adenovirus and pseudorabies viruses [45]. HKDC1 encodes one of the five hexokinases that phosphorylates hexose sugars. A study performed on patients with lung cancer have found over-expression of HKDC1 in tumor tissues. They suggest that the up-regulation of HKDC1 helps to overcome the deficiency of energy during stress stimulating the intracellular retention of glucose and driving it to glycolysis [46].

CCDC157 is the third largest hub in the network. It is connected to other genes of the network with a comparable number of C- and S-type links, thus being one of the nodes with the lowest node homogeneity score. The gene encodes a coiled coil domain, however, its function, to my knowledge, is not known. In general, coiled-coils are highly versatile motifs found in wide range of structural and regulatory proteins, and many function in regulation of gene expression [47]. The CCDC157 is not the only hub being a coiled-coil domain. CCDC85C also has a central role in the CSD-network with a comparable number of C- and S-type links. The gene is proposed to be involved in cell-cell adhesion and epithelium development. The latter is of special interest as the tick-bite results in skin damage.

The five hubs encoding proteins with functions associated with the body's immune system are hubs predominantly connected to differentiated and specific interaction types, however with one exception. The largest of these hubs is CD109 which, while being connected with 19 S-type and 1 D-type link, is mostly connected by C-type links. CD109 encodes a cell-surface antigen, meaning that they help identify and classify cells. They localize to the surface of platelets, activated T-cells and endothelial cells, and is found to modulate negatively TGFB1 signaling in keratinocytes (major cell type of the epidermis, the outermost layers of the skin). As TGFB1 signaling negatively regulates keratinocyte proliferation, it has been found that release of CD109 from the cell surface in human keratinocytes may induce molecular changes found in disease such as psoriasis and rheumatoid arthritis [48, 49]. It is also found that CD109 is up-regulated in inflammatory responses [49].

The other hubs involved in the immune system, show both specific and differentiated changes in gene expression patterns in blood infected by the bacteria. Of these, S1PR1 is the largest hub and is almost completely dominated with S-type links. S1PR1 encodes G-coupled receptors highly expressed in endothelial cells. S1PR1 signaling is important in the regulation of lymphocyte maturation, more specifically it is required for normal

egress of mature T-cell from the thymus into the blood stream and into peripheral lymphoid organs [50]. A study done on patients infected with tick-borne encephalitis-virus found significant elevated levels of this gene ([51]. LRCH1 and GSDMA are the only S-type hubs that are completely homogenous. LRCH1 takes part in the regulation of T-cell migration. The gene acts a negative regulator of the GTPase CD42, which restrains CD4(+) T-cell migration [52]. GSDMA however, is proposed to be involved in promotion of pyroptosis; an inflammatory form of programmed cell death that occurs most frequently upon infection with intracellular pathogens. A recent study by Ding et al [53] showed that the gasdermin-N domain of GDSMA can induce pyroptosis in mammalian cell and it is thus suggested that a possible function of GSDMA is to mediate defense in the skin by inducing pyroptosis [54].The gene also binds to bacterial lipids, and exhibits bactericidal activity (kills bacteria) [53].

The smallest of these immune related hubs are PARP14 and PLCD4. While PARP14 is found in the core of the dense specific cluster ($H = 0.97$), PLCD4 is found at the edge of the cluster, with the lowest homogeneity score of all the hubs ($H = 0.39$). PARP14 plays a part in regulating the activation of macrophages. The silencing of the gene cause induction of pro-inflammatory genes by promoting STAT6 phosphorylation in response to IL4 which positively regulates MRC1 expression [55]. PLCD4 encodes a class of phospolipase C enzymes, and the over expression of PLCD4 is found to upregulate the Erk signaling pathway and proliferation [56]. Erk pathway contributes to regulation of cell proliferation such as T-cell activation and also induce inflammation response [57].

While the rest of the hubs do not encode proteins associated with immunity functions, they all have function which can be up- or down-regulated in response to pathogens, as suggested as many of them are connected by S- and D-type links. Especially interesting are the almost completely homogeneous S-type hubs; MMAA, ZNF702P, TMEM161B-ASI, OR6WIP and MMP14. It is difficult based on function to explain why their pairwise gene co-expression with neighboring genes are only identified in one of the conditions, i.e. patients with Lyme disease or healthy individuals. However, their role may be interesting to look further into.

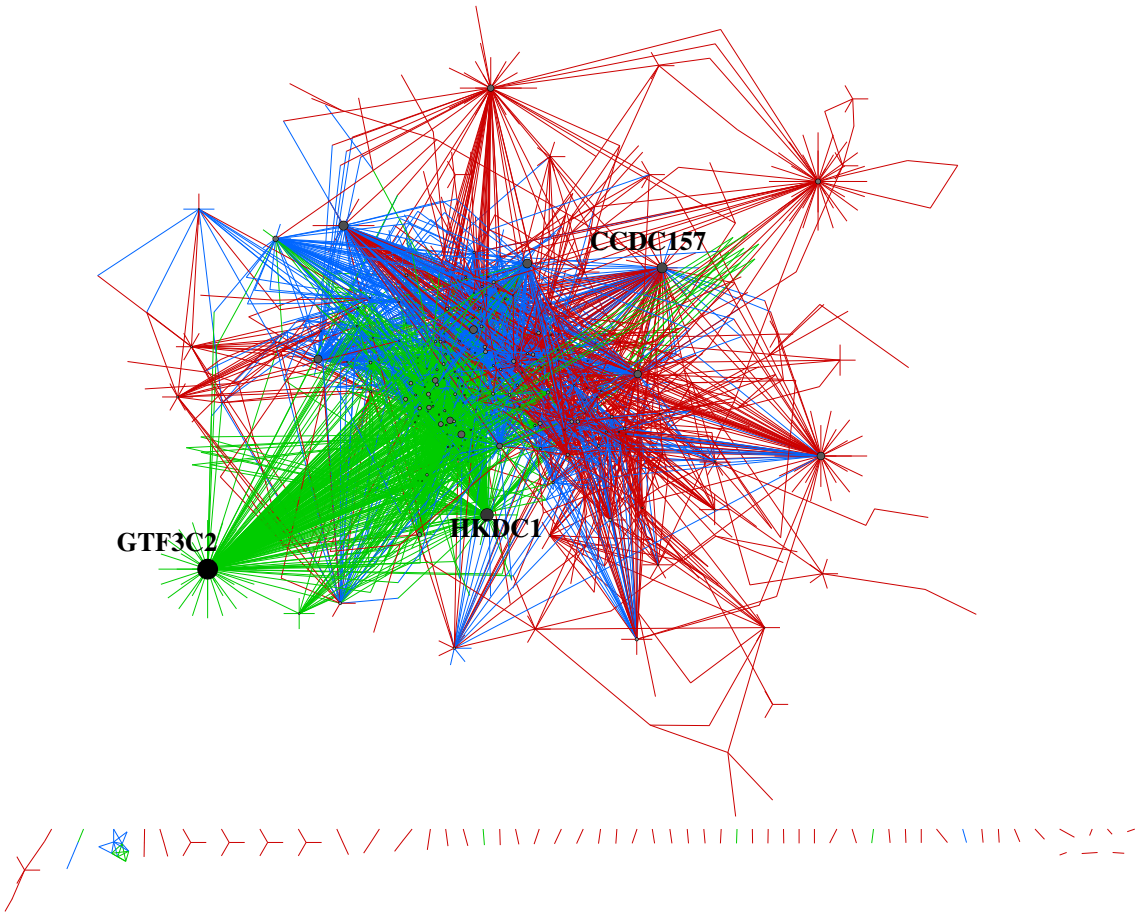


Figure 4.3: Hubs of the network. A visualization of the hubs of the CSD-network.Node size and color is proportional to node degree. Edge color correspond to the different interaction types (C = blue, S = green, D = red). The three largest hubs of the network are highlighted.

How the interaction types are spread in the network is presented in a venn diagram in Fig. 4.4. There are in total 5 of the 21 hubs presented in 4.1 with a homogeneity score equal to 1.0, while the rest mostly range from 0.5 and 0.9. PLCD4 is the only node with a homogeneity score below 0.5 ($H = 0.39$). The majority of interaction types are of the differentiated type (red), followed by specific (green) and then conserved (blue). Of nodes with mixed interaction types the most common combinations are conserved and differentiated, followed closely by conserved and specific. Only 6 of the nodes have a combined interaction type of differentiated and specific. However, 56 of the nodes have all three types.

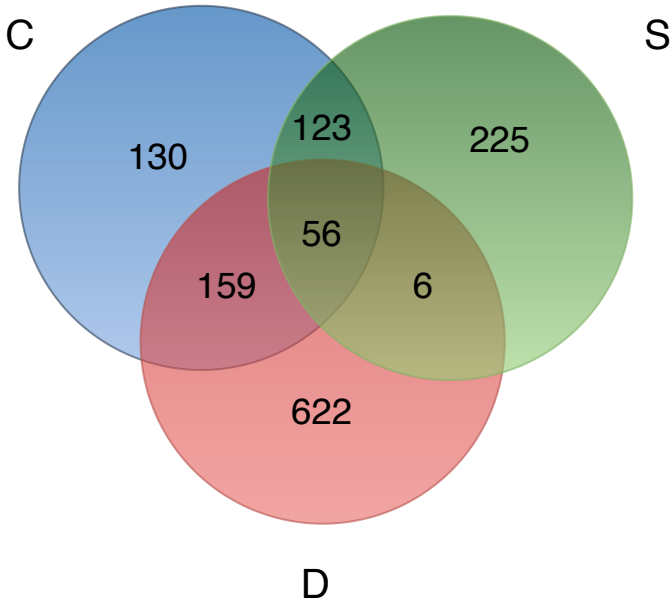


Figure 4.4: Node homogeneity. Venn diagram showing the amount of genes from Fig. 4.1 involved with each interaction type. The colors represent the different interaction types, blue being C-type, green being S-type and red being D-type.

The neighborhood connectivity distribution for the three individual sub-networks is presented in Fig 4.5 - Fig 4.7. All three networks show disassortativity topology, however in various degree. The disassortativity is highest for the specific network (most negative exponent; $\mu = -0.556$) and lowest for the conserved network ($\mu = -0.243$). This tendency can also be revealed form Fig 4.3. While the sub-network of S- and D-type links have a tendency to form "bouquet" like structures this is not as prominent for the sub-network of C-type links. The bouquet like structures are typically found in disassortative networks were central high degree nodes of the network tend to ignore each other and instead connect to low-degree nodes. However, while the networks are disassortative, many of the hubs share interactions with each other. In the sub-network of exclusively S-type links the two largest hubs are connected to each other and several other hubs found in this network. GTF3C2 and HKDC1 is however responsible for the disassortativity as they connect to a large number of low degree nodes as well. The same is seen for the other sub-networks were while they still form bouquet like structures, these structures are connected to each other.

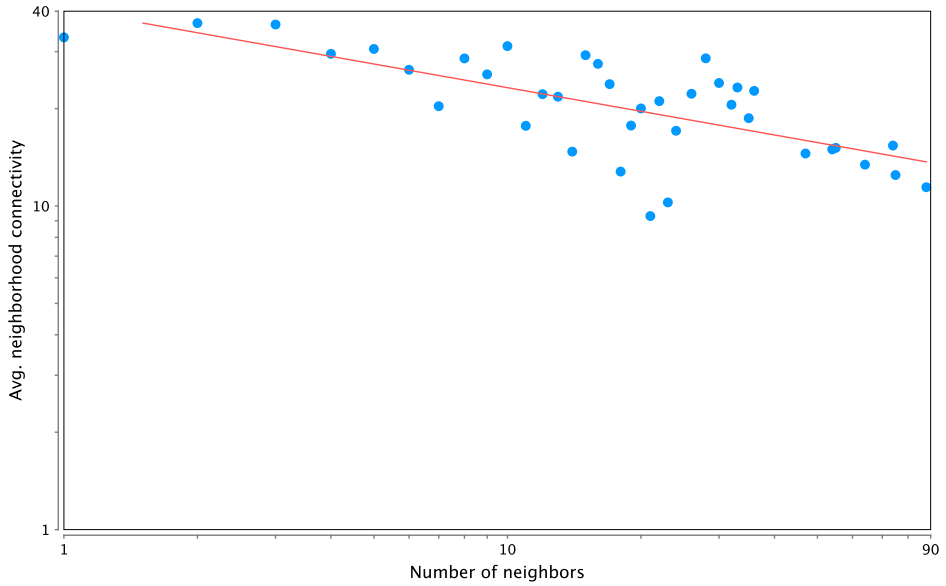


Figure 4.5: Neighborhood connectivity distribution for the C-network. The average number of neighbors plotted as function of node degree (k) on a log-log scale. The regression line for the data points (in red) gives a negative correlation exponent $\mu = -0.2435$.

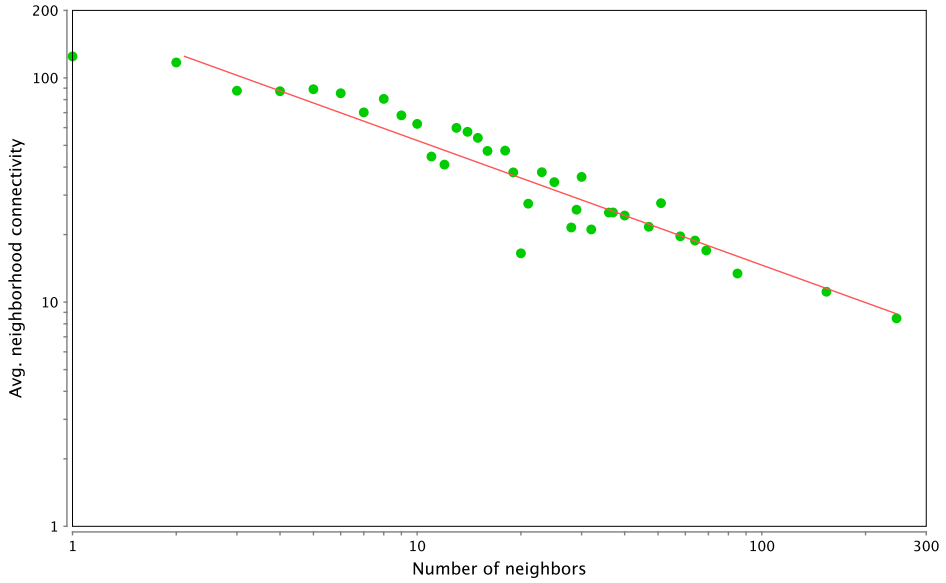


Figure 4.6: Neighborhood connectivity distribution for the S-network. The average number of neighbors plotted as function of node degree (k) on a log-log scale. The regression line for the data points (in red) gives a negative correlation exponent $\mu = -0.556$.

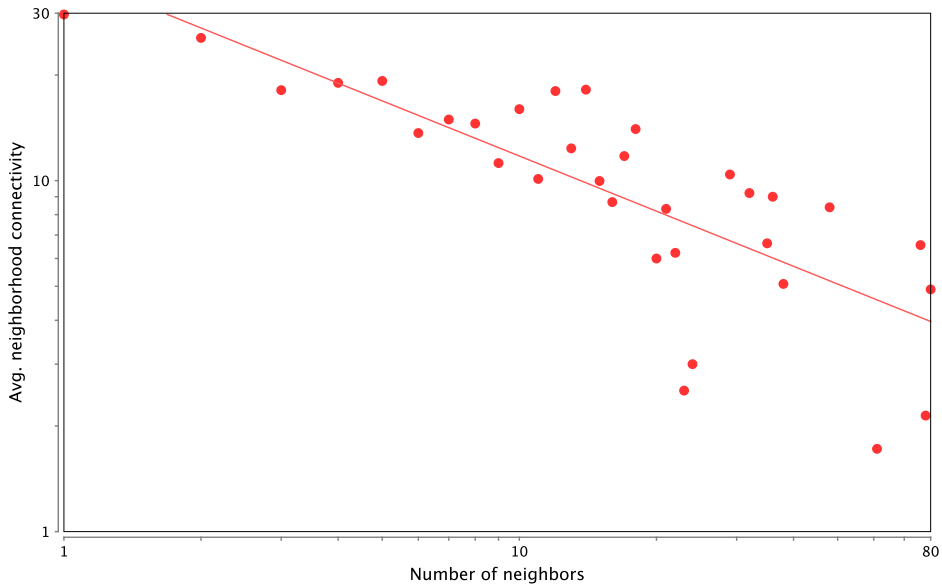


Figure 4.7: Neighborhood connectivity distribution for the D-network. The average number of neighbors plotted as function of node degree (k) on a log-log scale. The regression line for the data points (in red) gives a negative correlation exponent $\mu = -0.522$.

4.1.4 GO Enrichment Analysis on the CSD-Networks

A gene ontology enrichment analysis was performed on the four differential co-expression networks obtained from the CSD-method to provide insight into the characteristic biological processes the differentially expressed genes are involved in. The results from the GO-analysis were corrected by calculating the false discovery rate (FDR), and only the GO terms with FDR < 0.05 were considered significant. Significant results were only found for two of the networks; the complete and the differentiated network. Of the 1323 genes in the complete network, 1179 of these mapped to unique PANTHER gene ID's. In the differentiated network, 748 of 845 genes were mapped to uniquely mapped ID's. The biological processes found to be significantly over-represented for the two networks are presented in Table 4.2 and Table 4.3. The tables are sorted by fold enrichment of the process in reference to healthy humans. Only the most specific terms are presented in the tables. The complete results for the gene ontology enrichment analysis are presented in Table A4 - A3.

Table 4.2: Gene ontology for the complete CSD-network. Biological processes over-represented in GO enrichment analysis based on the complete CSD-network from Fig. 4.1. Sorted by fold enrichment. Only the most specific terms are included. The complete result is presented in A3.

GO	FE
Valyl-tRNA aminoacylation	17.16
Mitotic nuclear division	2.66
Response to ionizing radiation	2.60
chromosome segregation	2.12
Microtubule cytoskeleton organization	1.81
Regulation of mitotic cell cycle	1.71
Intracellular protein transport	1.60
Cellular component assembly	1.41
Regulation of catalytic activity	1.33
Regulation of cellular component organization	1.32
Negative regulation of cellular process	1.21
Regulation of primary metabolic process	1.20
Regulation of nitrogen compound metabolic process	1.19
Regulation of macromolecule metabolic process	1.18

From the table we see that the GO-analysis for the complete CSD-network identified biological processes which are central for whole cell bloods in a normal state, but which also may be an indication of behavior in response to bacterial infection. Most of the biological processes enriched were related to cell division, component assembly and organization, protein transport and regulation of metabolic processes.

Valyl-tRNA aminoacylation is by far the most enriched biological process (17-fold enrichment) in the CSD-network. The high fold enrichment results from the GO-analysis identifying six genes from the network involved in this response. However, the six genes identified were found to be the same one; VARS2, but with different mapping IDs. The GO-analysing tool identified several ensemble and uniprot IDs for the VARS2 gene, and accordingly defined them as six different genes. The high over-representation of the valyl-tRNA aminoacylation is therefore disregarded.

Other enriched processes in the complete CSD-network were associated with cell cycle, component assembly and organization and protein transport. Since these processes are essential in blood cells during normal behaviour, the enrichment analysis of the S- and D-networks is of interest as they are assumed to capture more of the disease related biology. The D-network was the only other network related to significantly enriched biological processes. Genes of this network are highly co-expressed in both conditions but with opposite signs. Since the processes enriched are also very similar to those found in the complete network, this suggests that the genes relating to these processes are differently expressed in patients infected by *B. burgdorferi* compared to healthy individuals.

Table 4.3: Gene ontology for D-type network. Biological processes over-represented in GO enrichment analysis based on the network with only D-type links A3. Only the most specific terms are included. The complete result is presented in A4

GO	FE
Valyl-tRNA aminoacylation	26.87
Maintainance of location in cell	3.48
Response to ionizing radiation	3.15
Regulation of secretory pathway	2.95
Cell cycle process	1.82
Cellular protein localization	1.75
Regulation of cell cycle	1.63
Positive regulation of catalytic activity	1.56
Protein transport	1.56
Cellular component assembly	1.56
Organelle organization	1.47
Establishment of localization in cell	1.44
Regulation of primary metabolic process	1.27
Regulation of nitrogen compound metabolic process	1.27
Regulation of cellular metabolic process	1.26
Regulation of macromolecule metabolic process	1.25

The general trend is that many of the biological processes found in the CSD-network are also found in the D-network, however with the terms being slightly less specific. The GO-terms from the D-network had, in general, lower fold enrichment compared to the complete network. This may, however, be because the requirements for high enrichment scores (number of matching genes), increase with the number of genes assessed. Many of the differentially expressed genes in the network are associated with protein localization and transport. While this occurs in all cells in a normal state, it is also an important part of the inflammatory response upon bacterial infection, as the immune cells is transported and localized to injury site.

Many genes in the D-network are also related to regulation of metabolic processes, which are processes that regulate rate and extent of processes that produce energy. *B. burgdorgeri* is a pathogen which highly depends on the host to survive as it lacks central metabolic pathway to generate its own nutrient. Several studies have provided evidence that the bacteria changes in host immune cell metabolism, and which upon infection, glycolysis is induced to fuel the inflammatory response [58]. Two identified hubs in the CSD-network; HKDC1 and KHK, encode proteins functioning in glycolysis. As HKDC1, is completely homogenous with S-type links, this suggest that this gene is central for this biological process during inflammation. KHK, on the other hand, is mostly dominated by C-type link, but also with a few S- and D-type links. This suggest that its function is somewhat conserved upon infection.

Biological processes related to the cell cycle were also enriched in the D-network. The cell cycle is a four stage cycle the cell passes through to allow them to divide and produce new cells, and the regulation of this process is essential for an organisms survival. In the occurrence of a damaged cell, the cell division of this cell must be stopped, or if growth or wound healing is necessary the cell division need to be started. The regulation is accomplished by cells using chemical stimulus to initiate a signaling cascade that reaches the necessary proteins needed to stop or initiate the cell cycle.

The cell cycle occurs in many types of cells, and epithelial and blood cells are two types of cells that need to be replaced constantly. Therefore, in order to asses if the enrichment was high because of whole-blood samples were being used or because of the disease state, a deeper look into the types of cells enriched was done using the Enrichr website [59, 60]. The enriched cell types with their corresponding p-value, adjusted p-value (q) and odds ratio and combined scores are presented in table 4.4. The adjusted p-value was calculated using the Benjamini-Hockberg method. The odds ratio is the enrichment score and the complete score is a combination of q and the odds ratio. The results from the Enrichr analysis provided some other cell types as well, but they were either not of relevance or the combined score was considered to low.

Natural killer cells (NK cells) and CD33++ Myeloid cells were enriched in both networks. NK cells are lymphocytes (white blood cells) and component of the innate immune system [61]. Myeloid cells, on the other hand, have the potential to develop into mature red blood cells, platelets or one type of white blood cells (granulocytes) [62, 63]. CD33+ Myeloid is a cell surface antigen commonly expressed by these immature myeloid cells which upon stimulation inhibits phagocytosis in the cell. B-lymphoblasts, enriched in the CSD-network, are modified naive lymphocytes which occurs when the lymphocyte is activated by an antigen. The lymphoblast will then go through numerous cell divisions and differentiate into plasma cells that secrete antibodies [64]. Lastly, the CD14+ monocytes are believed to be committed precursors for phagocytes such as macrophages [65]. All of these cell types can therefore be related to the immune system.

Table 4.4: Enriched cell types. The enriched cell types for the complete and D-type network found by using the Enrichr website. The cell types are presented with their corresponding p-value, adjusted p-value, odd ration and combined scores.

Complete network	p	q	odds ratio	combined score
CD56+ NKCells	0.0036000	0.3024	1.39	7.80
721 B lymphoblast	0.004180	0.1756	1.25	6.87
CD33+ Myeloid	0.009210	0.2579	1.36	6.37
CD14+ Monocytes	0.02135	0.3587	1.42	5.47
Differentiated network	p	q	odds ratio	combined score
CD56+ NKCells	0.001067	0.08959	1.57	10.77
CD33+ Myeloid	0.008944	0.3757	1.46	6.91

4.1.5 Genes Previously Associated with Lyme Disease

A list of 19 genes previously associated with Lyme Disease is pesented in table A1. Out of these, only two genes were recognized in the CSD-network: CXCL10, and CXCL2. CXCL10 was located in the specific region of the network with a node degree of $k = 9$. Of the nine neighboring nodes, six of these were hubs in the network; GTF3C2, HKDC1, TNNC2, KHK, MMAA, ZNF70P. CXCL2 however had a modest role in the network, connecting to a single neighbor with a C-type link. Another member of the CXC chemokine family was detected in the network; CXCL3. This gene was also connected by a C-type link to one other node in the network, however with the node being the third largest hub; CCDC157. Also, while non of the member of the CCL chemokine family listed in table A1 were recognized, CCL20 was discovered in the network. It is only connected by two links, however both being D-type links. The same is true for IL-16, a member of the interleukin family. It is connected with D-type link to S1PR1 and a C-type link to CCDC157.

Lastly, MMP14, a family member of MMP9 was detected as a hub in the CSD network, with its majority of the interactions with neighboring nodes being of specific co-expression. The MMP gene family are proteases capable of degrading extracellular matrix proteins, and are thought to play major roles in cell proliferation, migration, apoptosis and host defence. Prior studies have identified that *B. burgodorferi* induces the host protease matrix metalloproteinase 9, suggesting that this may aid the bacteria to produce local tissue damage [66, 67]. MMP14 is found to target CXCL12, resulting in inhibition of T-lymphocyte migration, during inflammatory responses [68].

4.2 Network Analysis

4.2.1 Community Detection

The first suggested step for extracting new information from the constructed network was to partition the network into communities in order to get more insight into the network topology. For this step, different algorithms for community detection found in NetworkX was applied to the CSD-network. Many of theses algorithms detected a large number of communities, however many with negligible size, such as those created by the small components not connected to the giant component of the network. Only those with a community size of minimum 30 nodes were thus considered for further analysis. The total number of communities detected with the four different algorithms is presented in table 4.5 along with the number of relevant communities, and modularity score Q.

Table 4.5: Detected communities. A table over the total amount of communities detected along with the number of communities with the desired community size for the four algorithms and Q.

	Communities	Communities with size ≥ 30	Q
Greedy Modularity Maximization	71	5	0.5171813920312109
Label propagation	105	6	0.4742189918797999
Girvan-Newman method	90	7	0.49035872735792135
Louvain Community detection	64	8	0.531036451786071

From the table we see that the different algorithms detected a large number of communities, where only a few had a desired community size. The modularity score was largest for the Louvain community detection algorithm, thus indicating that the Louvain method detects the communities with highest quality. The relevant communities detected with the different algorithms were visualized using Cytoscape and are presented in Fig 4.8 - 4.11. Each module is represented by color, and the genes with highest number of neighbors for each module are highlighted. The size, average degree, clustering coefficient and number of interaction types for the different modules are presented in Table 4.6.

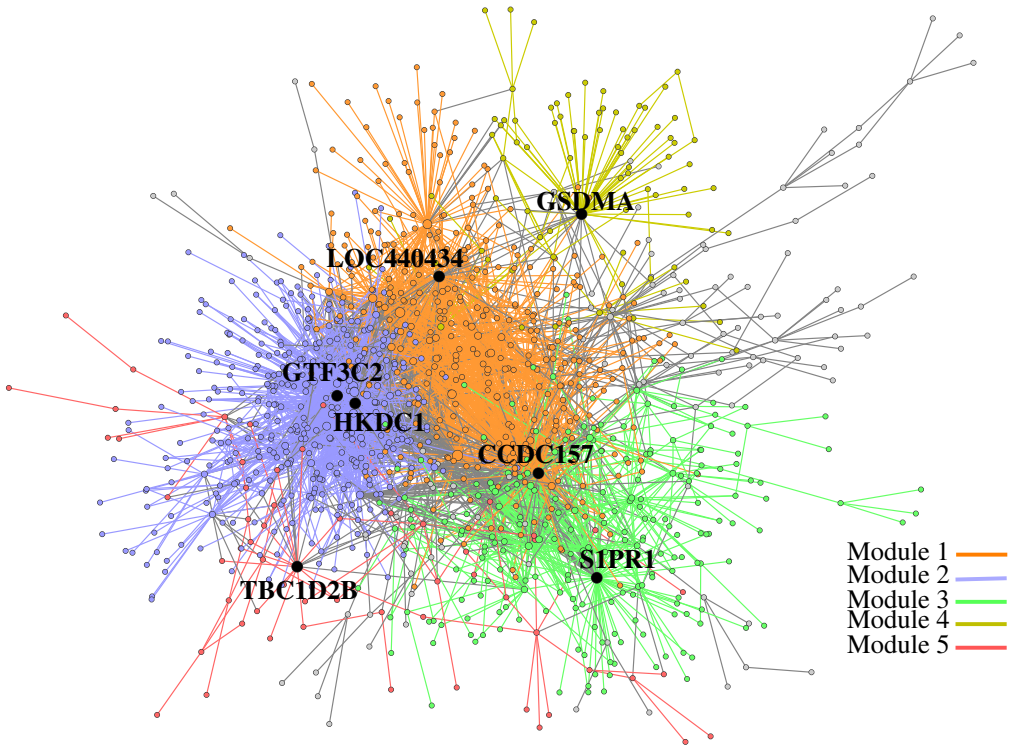


Figure 4.8: Modules detected using greedy modularity maximization algorithm. A visual representation of the most relevant modules detected using the Greedy modularity maximization algorithm from NetworkX. The modules are applied to the giant component of the complete network (Fig. 4.1). The hubs of each module are highlighted. Modules are numbered according to size (large to small).

Five large modules were detected using the greedy modularity maximization algorithm. The modularity score for the partitioning of this network was second largest, meaning that the partitioning is not random. We see, from the figure, that the three largest modules (1, 2 and 3) are locally dense, connected graphs, which relates to the definition of communities. Contradictory, module 4 is a peripheral module, and not densely connected as it has a bouquet like structure, i.e. there is one central high degree node connected by single degree nodes. It does, however, fulfill the criteria of being a connected sub-graph as the nodes share more edges with the nodes within the community than those belonging to other communities. Module 5, which is sparse and spread across the network, is therefore defined as a module as it also fulfills this criteria.

Module 1 (orange) is located in the C and D rich regions of the network. The module is centered in the network and connects to all the other modules. Most represented are the conserved interaction types (Table 4.6). A high level of conserved relationships is less

likely to explain differences in phenotype related to the disease. It is interesting, however, to see which biological processes that are captured in this module as it is surrounded by modules dominated by specific and differentiated interaction types. Most interesting is module 2 (orange) whose average clustering coefficient is the largest and almost exclusively dominated by the specific interaction types. The community also include the largest hubs found in the network. This module is thus expected to consist of gene co-expression relationships directly affected by the disease.

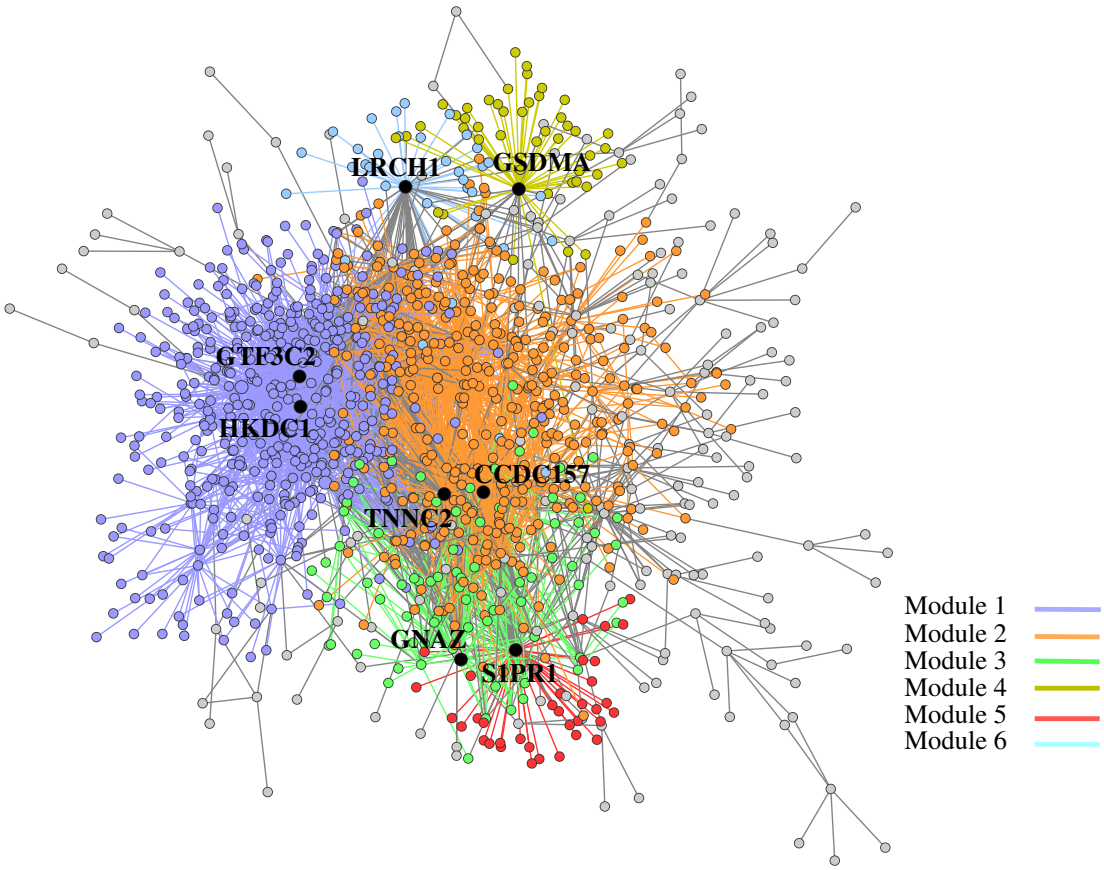


Figure 4.9: Modules detected using LPA. A visual representation of the most relevant modules detected using Label propagation algorithm from NetworkX. The modules are applied to the giant component of the complete network (Fig. 4.1). The hubs of each module are highlighted. Modules are numbered according to size (large to small).

Six large modules were detected using the label propagation algorithm from NetworkX, which is one more than the GMM algorithm. The two large densely clustered modules (purple and orange) are also detected here, however in this network the purple module, dominated by specific interaction types, is larger than the orange module. The most noticeable difference, however, is the increase in number of bouquet-structured modules detected. The hubs; LRCH1 and S1PR1 and the majority of their neighboring nodes are considered individual communities in this algorithm and are completley dominated by D-type links. Also, no sparsely spread community is detected. However, the modularity score for the partitioning of this network is the lowest of all, indicating that the quality of these modules are not as good as the other. The division of the green module from GMM algorithm into two separate ones highlights another high degree node of this community; GNAZ. GNAZ is connected to 40 neighboring nodes in the network by both conserved and differentiated interaction types. GNAZ encodes a membrane bound GTPase that mediates signaling transduction in pertussis toxin-insensitive systems. Pertussis toxin is an AB_5 type exotoxin secreted by some pathogenic bacteria, and may cause harm to the host by disrupting normal cellular metabolismis. A study has also suggested its role in neutrophil migration in inflammatory response [69]. Although, no exotoxins have been identified in *B.burgodorferi* [70], the gene may be interesting to investigate further.

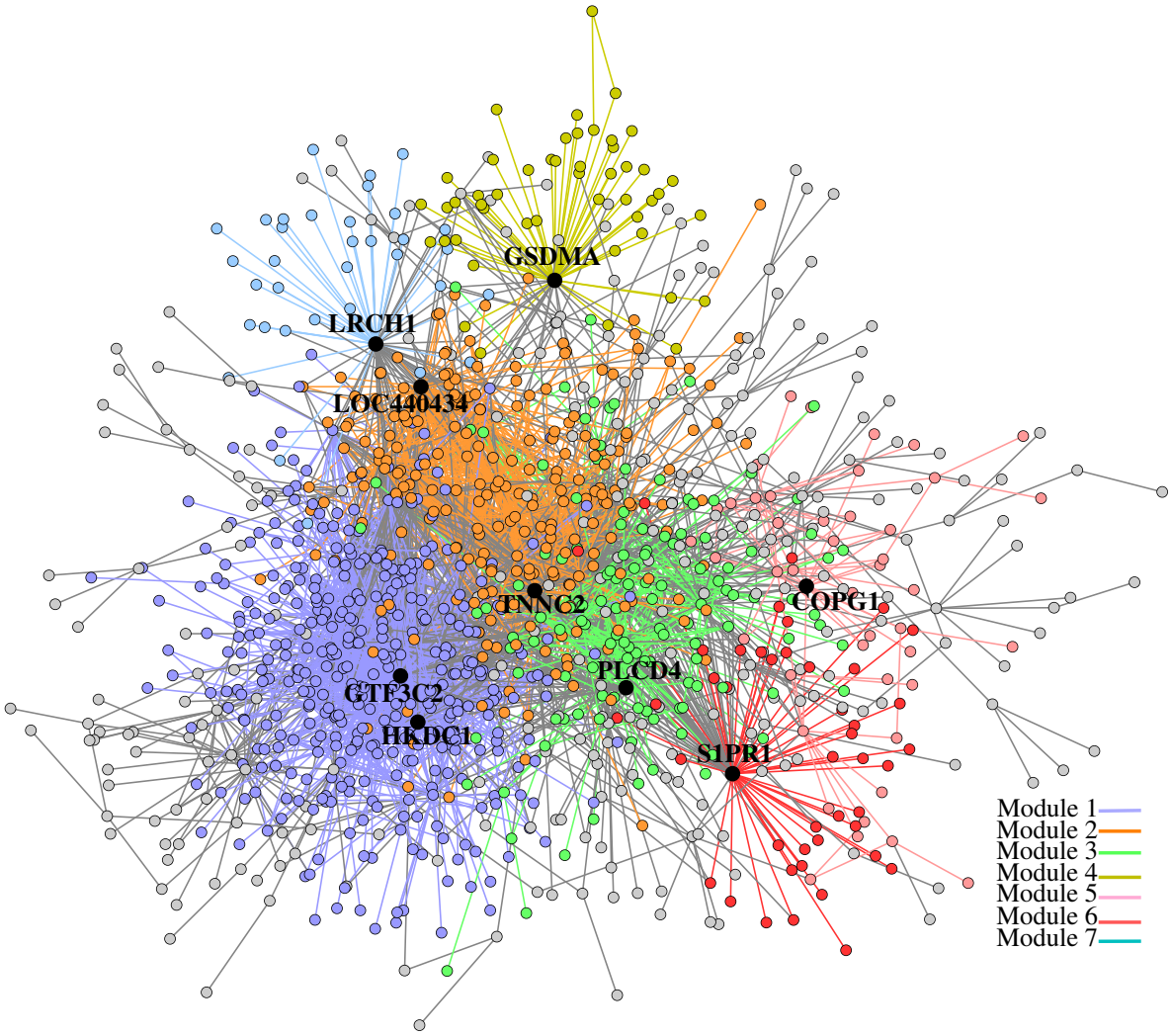


Figure 4.10: Modules detected using Girvan-Newman method. A visual representation of the most relevant modules detected using the Girvan-Newman algorithm from NetworkX. The modules are applied to the giant component of the complete network (Fig. 4.1). Relevant hubs are colored black. Modules are numbered according to size (large to small).

The Girwan-Newman method was the only divisive algorithm investigated. The algorithm detected 7 relevant modules, which is one more than the LPA, and two more than the GMM algorithm. A quick look at Fig 4.10 show that the modules are somewhat identical to the ones found using LPA, but with small differences in size and clustering. The modules from the GN approach generally detected smaller sized modules with higher average degree and clustering. The additional module detected (pink) is sparse and spread across the network,

similar to the red module in Fig. 4.8. However, even though both modules detected one sparsely community, they are not the same one as only five of the genes are detected in both communities. The gene with the highest number of neighbors found in module 5 (pink) is COPG1. COPG1 is connected to 20 other neighbors in the network with only D-type links, suggesting a prominent role in the mechanism involved after infection.

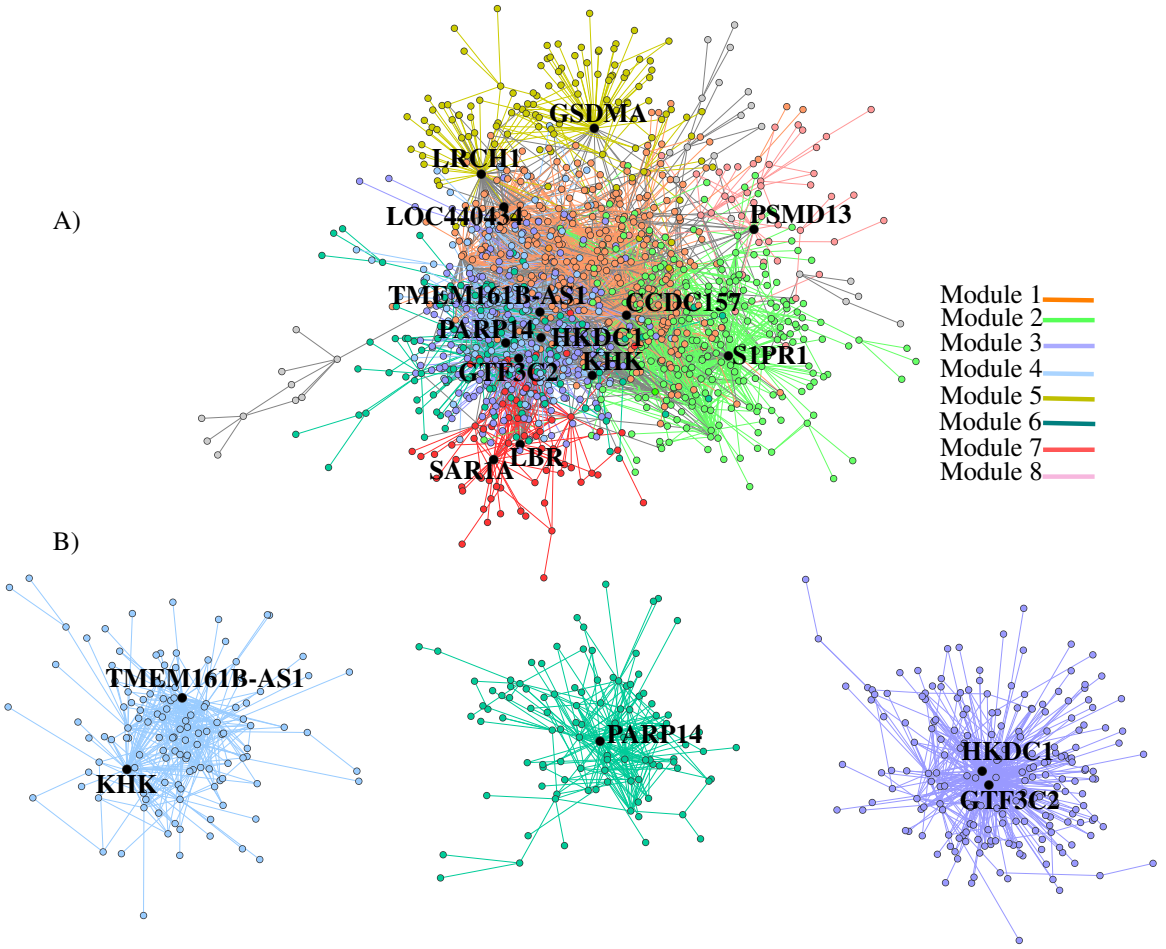


Figure 4.11: Modules detected using Louvain community detection. (A) A visual representation of the most relevant modules detected using Louvain community detection from NetworkX. The modules are applied to the giant component of the complete network (Fig. 4.1). Relevant hubs are colored black. Modules are numbered according to size (large to small) (B) A visual representation of module 3, 4 and 8.

The communities detected using the Louvain community detection algorithm were most different compared to those detected by the other algorithms. The algorithm detected

eight communities with relevant size which are visualized in Fig. 4.11 (A). The most noticeable difference is that this algorithm consider the dense, clustered specific region of the network as three separate modules. From Fig 4.11 (A) it is difficult to separate the three modules, and they are therefore extracted and visualized in 4.11 (B). This algorithm does provide new insight into the dense cluster of the specific co-expressed gene pairs. The other algorithms consider the entire specific region being one large module, and they thereby suggest that all the genes located here can be associated with the same biological processes, complexes or location. However, this new insight suggest that there are three smaller modules, evolving around different hubs, creating three different neighborhoods. While all of these genes have the same expression pattern under both conditions, it can be suggest that there are three different processes involved.

Another difference is that, a part from dividing earlier detected communities into separate ones, LCA also merge together communities such as the communities centered by the hubs LRCH1, GSDMA and S1PR1. Three new high degree nodes have also been identified as it detected two new modules; module 7 (red) and module 8 (pink). While the pink module show the same sparse structure with zero clustering, the red module is more clustered. SAR1A connects to 23 neighbors in the network, all by D-type links, and LBR with 36 neighboring genes, by all three interaction types. LBR is also directly co-expressed with the giant hubs of the network; GTF3C2 and HKDC1. PSMD13 located in module 8 (pink) is connected to 22 neighbors with the majority connected by D-type links. These genes may be good candidates for further studies.

Table 4.6: Module parameters. The modules detected from the four algorithms with their respective number of genes, average neighborhood connectivity $\langle k \rangle$, average clustering coefficient (C) and number of interaction types.

Greedy modularity	#genes	$\langle k \rangle$	C	#C-links	#S-links	#D-links
module 1	368	6.07	0.344	708	11	398
module 2	360	7.10	0.448	119	1097	62
module 3	215	3.98	0.106	134	-	294
module 4	81	2.07	0.013	7	-	77
module 5	58	2.10	0.058	11	-	50
Label propagation	#genes	$\langle k \rangle$	C	#C-links	#S-links	#D-links
module 1	442	6.89	0.406	239	1192	92
module 2	371	5.52	0.257	549	1	474
module 3	78	5.00	0.238	110	-	85
module 4	47	1.96	0.0	-	-	46
module 5	36	1.94	0.0	-	-	35
module 6	36	1.94	0.0	-	-	35
Girvan-Newman	#genes	$\langle k \rangle$	C	#C-links	#S-links	#D-links
module 1	375	7.36	0.429	187	1154	38
module 2	230	5.60	0.350	461	5	178
module 3	126	4.65	0.161	121	1	171
module 4	49	2.00	0.0	-	-	49
module 5	45	1.96	0.026	2	-	61
module 6	45	2.80	0.0	-	-	45
module 7	34	1.94	0.0	-	-	34
Louvain algorithm	#genes	$\langle k \rangle$	C	#C-links	#S-links	#D-links
module 1	262	5.26	0.313	439	3	249
module 2	237	4.44	0.109	170	-	356
module 3	198	4.41	0.369	39	385	13
module 4	125	5.58	0.322	112	240	9
module 5	116	2.16	0.0	-	-	125
module 6	106	4.23	0.296	22	179	23
module 7	68	4.27	0.249	60	-	85
module 8	42	2.24	0.0	1	-	46

4.2.2 GO Enrichment Analysis on Detected Modules

In the next step, gene ontology enrichment analysis was performed on the different modules in order to investigate whether the genes of the modules related to any biological processes. However, significantly over-represented biological processes was only found in module 1 (purple) detected by the label propagation algorithm. The genes of this module were found to be enriched in cell cycle, and more specifically in mitotic anaphase (Table A5. This may suggest, however, that it is not the common genes for this module, detected

using the different methods, that cause this over-representation.

4.2.3 WGCNA

The final proposed step investigated for extracting more information from the network was to perform weighted gene correlation network analysis (WGCNA). 40 modules with a minimum size of 30 nodes was detected using WGCNA on the TOM created from the gene expression data. Fig 4.12 show the dendrogram for the detected modules, and their respective size i presented in Table 4.7.

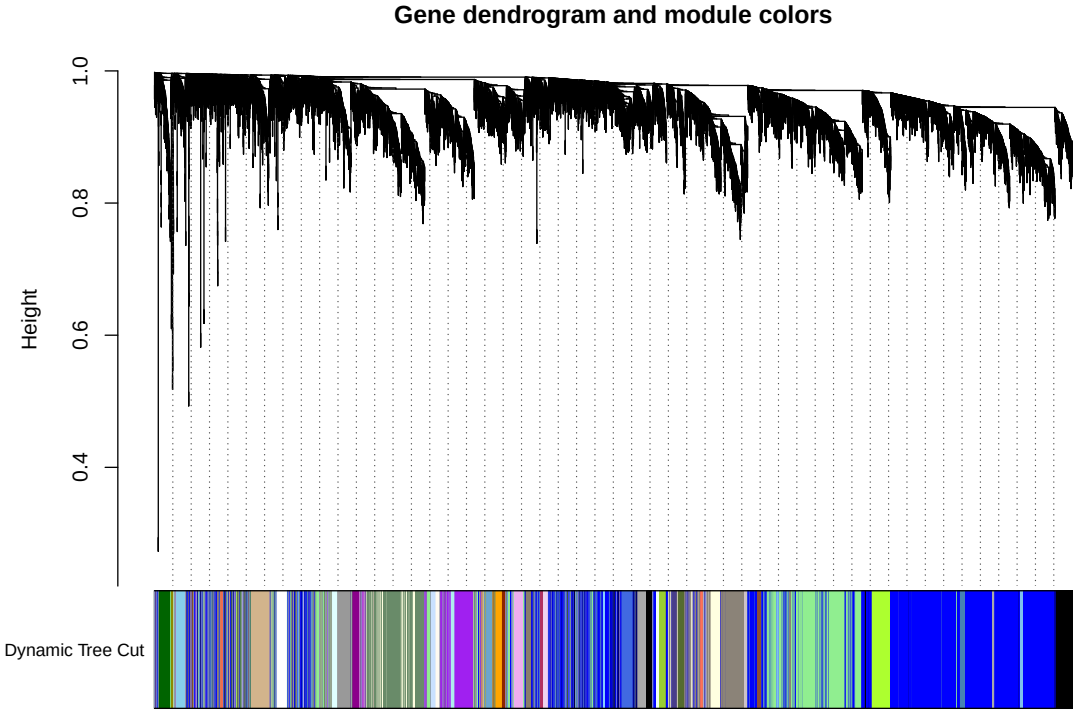


Figure 4.12: Dendrogram of modules. Dendrogram of genes clustered based on a dissimilarity measure (1-TOM) with assigned module colors. The colored rows show the module membership obtained by the dynamic tree cut method after merging the modules.

Table 4.7: Modules detected from WGCNA. The modules labeled by color from the dendrogram and respective module size. Modules investigated further are highlighted.

Module	#genes	Module	#genes
blue	4733	darkgrey	325
lightgreen	2073	floralwhite	79
darkseagreen4	1045	lightcyanin	84
antiquewhite4	**970**	mediumorchid	36
black	651	skyblue2	158
purple	630	white	151
lightyellow	**607**	darkmagenta	114
coral2	36	greenyellow	337
honeydew1	48	orange	174
paleturquoise	126	royalblue	257
skyblue	150	skyblue3	104
tan	299	**yellow4**	**34**
bisque4	78	coral1	40
darkgreen	180	**darkolivegreen**	**123**
darkslateblue	**71**	grey60	219
lavenderblush3	49	lightsteelblue1	88
maroon	57	orangered4	95
plum2	180	salmon4	65
skyblue1	32	steelblue	130
thistle2	71	yellowgreen	106

From the table we see that seven of the modules detected using this approach were larger in size than any of the modules detected earlier were the largest community had a size of 442 genes. Also, 40 modules were detected with a minimum size of 30 nodes compared to eight from the other communities. It is, however, important to emphasise that the community detection method from WGCNA was not applied on the differential co-expression network constructed from the CSD-method, as the other algorithms were, but instead on the similarity matrix obtained from using the wTO R-package described in 3.1.3. This approach thus performed community detection on a data set containing 14805 genes rather than 1323 genes. This explains why so many larger modules were detected using this approach compared to the other community detection algorithms presented earlier.

Gene Ontology Enrichment

To explore the biological relationships of the genes in each module, gene ontology (GO) enrichment analysis was performed. The results were corrected by calculating the false discovery rate (FDR), and only the GO terms with FDR < 0.05 were considered significant. The modules enriched with biological processes most related to the infection with *B. burgdorferi* (immune system responses and bacterial infection) were chosen for further

analysis. Five modules were identified as relevant; antiquewhite4, lightyellow, darkslateblue, yellow4 and darkolivegreen, and are highlighted in table 4.7. A selection of named terms related to each module can be found in Table 4.8 - 4.12 and the full lists are presented in appendix A4.

The genes of the antiquewhite4 module were enriched in apoptotic processes, processes of the immune system, and toll-like receptors. The latter is of special interest as several studies have suggested that *B.burgdorferi* lipoproteins activate inflammatory cells through Toll-like receptor 2 (TLR2), suggesting TLR2 could play a central role in the host response to the infection [71, 72]. The genes are also enriched in germinal center formation; a specialized structure formed in secondary lymphoid tissue which produce antibody secreting plasma cells and memory B-cells in response to infection [73].

Genes of the lightyellow modules were enriched in the response to peptidoglycan, suggesting that some of the genes are involved in the initiating immune response upon detection of the peptidoglycan of the bacterial cell wall. Some studies have found that the peptiodglycan of *B. burgdorferi* is a likely contributor to the inflammatory responses after infection and that they, when detected, stimulates macrophages to produce interleukin 1 which initiates immune responses. [74, 75]. As the other GO terms enriched are involved in regulation of interleukin secretion this suggest that the genes of this module are involved in this mechanism, and could be candidate biomarkers.

Only three biological processes were found to be significantly enriched in the darkslateblue module, all involved in B-cell activation and regulation, suggesting that the genes found in this module may be involved in antigen processing and activation of B cells upon the infection.

The genes of the yellow4 module have large fold enrichment for the process of oxygen transport, membrane disruption in other organisms and iron coordination entity transport. Two of these are related to transport which occurs in blood. This enrichment may be a result of the gene expression data being obtained from whole blood. Also the antibiotic catabolic process and drug process are highly over-represented in this module. As the gene co-expression data is obtained from patient pre-treatment and after completion of antibiotics this suggest that the genes found in this module may be involved in processes targeted by the antibiotics. A part from these processes, many of the GO-terms are related to bacterial responses, and may therefore be of relevance to the disease state.

Lastly, the genes of the darkolivgreen module were enriched in several immune and viral responses. Even though Lyme disease is the cause of bacterial infection it leads to up regulation of interferons which are signaling molecules released by host cells in response to virus [76, 77]. Thus, suggesting that some of the mechanism could be the same for viral response and response to *B. burgdorferi*.

Table 4.8: GO-terms for the antiquewhite4 module. A selection of over-represented biological processes associated with antiquewhite4 module. Sorted by fold enrichment. The complete table is found in table A6.

GO biological process	FE
fibroblast apoptotic process	23.51
positive regulation of cellular response to macrophage colony-stimulating factor stimulus	23.51
positive regulation of response to macrophage colony-stimulating factor	23.51
toll-like receptor 2 signaling pathway	18.81
germinal center formation	11.75
type 2 immune response	10.69

Table 4.9: GO-terms for the lightyellow module. A selection of over-represented biological processes associated with lightyellow module. Sorted by fold enrichment. The complete table is found in table A7.

GO biological process	FE
response to peptidoglycan	13.09
positive regulation of T-helper 1 type immune response	8.72
positive regulation of interleukin-1 beta secretion	8.38
positive regulation of toll-like receptor signaling pathway	7.63
regulation of interleukin-6 biosynthetic process	7.63
regulation of interleukin-1 beta secretion	7.63
cytokine secretion	7.19

Table 4.10: GO-terms for the darkslateblue module. All over-represented biological processes associated with darkslateblue module. Sorted by fold enrichment.

GO biological process	FE
regulation of B cell receptor signaling pathway	66.83
regulation of antigen receptor-mediated signaling pathway	32.58
B cell activation	13.45

Table 4.11: GO-terms for the yellow4 module. A selection of over-represented biological processes associated with yellow4 module. Sorted by fold enrichment. The complete table is found in table A8.

GO biological process complete	FE
oxygen transport	> 100
membrane disruption in other organism	> 100
iron coordination entity transport	> 100
antibiotic catabolic process	48.27
antibacterial humoral response	43.44
drug catabolic process	18.48
antimicrobial humoral response	18.35
neutrophil degranulation	16.22
granulocyte activation	15.64
leukocyte degranulation	15.51
myeloid leukocyte mediated immunity	15.21
cellular response to lipopolysaccharide	13.65
response to molecule of bacterial origin	13.42
defense response to bacterium	9.61

Table 4.12: GO-terms for the darkolivegreen module. A selection of over-represented biological processes associated with darkolivegreen module. Sorted by fold enrichment. The complete table is found in table A9.

GO biological process	FE
regulation of ribonuclease activity	55.16
dicarboxylic acid biosynthetic process	38.19
positive regulation of defense response to virus by host	33.10
type I interferon signaling pathway	32.11
antigen processing and presentation of endogenous peptide antigen	31.52

Differentially Expressed Genes

Genes involved in each module that were also recognized in the CSD-network were identified and are presented in Table 4.13. 159 differentially expressed genes (DEGs) between normal and infected individuals were found in the five different modules, which may suggest their involvement in the disease mechanisms.

The antiquewhite4 module consisted of the highest number of DEGs identified in the CSD-network. Of the 80 identified DEGs, nine were detected outside the giant component, and may, hence, play only a minor role in biological processes occurring upon infection.

13 of the DEGs had functions directly relating to the immune system and inflammatory responses, and are highlighted in Table 4.13. All were connected by differentiated (D) or specific (S) links to other genes in the network. Since a S-type link mean that their co-expression is present under only on conditions, and a D-type link that their is strong co-expression on both conditions, but with opposite signs, this indicates that their co-expression pattern significantly changes in patient infected by the bacterium. Some of the previously associated genes or close family members identified with Lyme disease were found among these DEGs; CXCL10, CCL20 and MMP14. Several of the 13 DEGs related to immune response is directly linked to some of the hubs in the network, further suggesting their involvement in the workings of the disease.

61 genes from the lightyellow module were identified in the network, with six of them not connected to the giant component. Of these, eight DEGs have functions directly linked to immune response, integrin signaling and host defense and are highlighted in 4.13. All except two genes, were connected by D- or S-type links and also share direct co-expression with hubs of the network, indicating that their gene expression pattern changes upon infection. Two genes in this module; ITGAX and LAIR1, where completely dominated with four and eight S-type links respectively. Additionally, they shared these S-type links with several hubs of the network including the two largest ones; GTF3C2 and HKDC1. ITGAX encodes a receptor for fibrinogen which mediates cell-cell interaction during inflammatory response. LAIR1 is leukocyte-associated immunoglobulin-like receptor which plays a fundamental negative regulatory role in the immune system. These genes are therefore great candidates for further research.

From the darkslateblue module, four genes were identified in the CSD-network, and three in the giant component. However, neither of these DEGs had functions relating to the B cell activation pathway. Three of the identified genes (ZNF860, KIAA0125 and MIR600HG) are RNA-genes, i.e. they do not encode proteins. ZNF8760 is however connected to the S1PR1 hub by a D-type link while the two others are connected to one gene each by S-type links. Their expressions patterns are thus different in patients infected by *B. burgdorferi*. The DENND5B gene on the other hand is involved in vesicle mediated transport. Its exclusively dominated by conserved type associations and thus indicating that the co-expression patterns among these genes are similar in patients with Lyme disease and healthy subjects.

Only two genes from the yellow4 module were identified in the CSD-network, both in the giant component. Both genes are one degree nodes, with HBD sharing a S-type link with GTF3C2 and DEFA1 sharing a D-type link with AQP3. Of these two, only DEFA1 had functions relating the gene ontology terms from table 4.11. The gene is part of a family of antimicrobial and cytotoxic peptides thought to be involved in host defense, and are abundant in granules and neutrophils [78].

All, except one DEG in the darkolivegreen module were located in the giant component. Of the identified genes, PARP14 were the only gene with function related to the immune system. PARP14, as mentioned earlier, plays a part in activation of macrophages, and is one of the hubs in the network. Three genes were located in the specific region, three in the conserved region, and the rest in the differentiated region of the network. However, all

DEGs except GNB4 and PDCL showed direct co-expression with the hubs of the network. The giant hubs of the CSD-network; GTF3C2 and HKDC1, were specifically co-expressed with several of the genes in the module; PARP14, LMNB1, IF16 and DDX60L, suggesting their involvement in the disease. All three genes, except DDX60L are located in the specific region of the network. DDX60L is almost exclusively dominated by C-type links, however, with two S-type link connected to GTF3C2 and HKDC1, and one D-type link connected to LRCH1. The gene is an interferon stimulated gene product found to be involved in the innate immune system with functions relating to sensing of viral RNA [79]. It could be interesting to look further into if it has functions relating to sensing of bacteria as well.

Table 4.13: DEGs. Genes involved in each module and recognized in the CSD-network. Genes with function related to immune response, inflammation and host defense are presented in bold.

Module	DEGs
antiquewhite4	**MMP14**, MGC12916, PTPRE, SAR 1A, PREX1 **IL1RL1**, UPP1, EXT1, NINJ1, **THEMIS2**, BTG3, KCNK5, ZNF624, **CXCL10**, NTRK1, LOC100133331 PPP4R2, HRH4, HSH2D, SLC12A7, TXNDC11 **MAP3K8**, PTS, LOX, TOM1, JMJD1C, **SKIL** ELOVL5, PLIN5, P2RX4, ADPGK, IQSEC1, ABCA1 CHMP4B, HIST1H1D, **CCL20**, MN, CYB5R2 **TNFAIP6**, C7orf53, MESDC1, IER5, SCN3B GZF1, BZRAP1, ID2, TRPC5, GMNN, CTH ,PLAA FAM72D, JARID2, **CD44**, **SH2D4A**, TEX14, PODNL1 MS4A3, SRI, CYTIP, LIMK2, PIN4P1, **LYN**, UBXN2A NKX3-1, TFIP11, DHFRL1, MCM6, SMARCA5, LRP2BP GPR84, RASGEF1B, C5orf58, SLU7, **BIRC2**,**CD80**, ZSWIM4, DUSP2APOL3, AZIN1, ROCK1
lightyellow	CCDC85C, USP32, SLFNL1, TXN, NOTCH1, WDR1, PLEKHG6, SPPL2A, LRP10, **LAIR1**, SLC1A3, SLC16A3, PSEN1, **CLEC6A**, C10orf131, **ITGAX**, PPIAL4A, SAMD12, KIAA1161, MAPRE1, PELI2, SMIM3, **ANXA1**, **CCRL1**, AKR7A3, MAP1B, BRI3, ZSCAN12P1, VASH2, PCGF5, SPDYA, SLC6A6, **ADORA3**, SRGAP2D, SPDYE3, VNN1, P2RX1, HECW2, OR2H2, ATOX1, GAB2, NXT2, S100A8 F5, MCL1, SIPA1L1, LACTB, **CXCL3**, DOCK5, GNG5, RAB39A, CSF2RA, CNEP1R1, **CXCL2**, SLC9A7P1 KIF13A, TNFRSF10B, TGFA, H3F3AP4, CELSR1, ACTR3
darkslateblue	DENND5B, KIAA0125, MIR600HG, ZNF860
yellow4	HBD, **DEFA1**
darkolivegreen	**PARP14**, **DDX60L**, CARD6, HAPLN3, VPS26A PDCL, IFI6, GBP3, GNB4, PLSCR1, GLS, LMNB1

Chapter 5

Discussion and Further Work

5.1 CSD Analysis of Lyme Disease

The first aim of this thesis was to apply the CSD method on gene expression data from patients with Lyme disease, in order to investigate changes in regulation of co-expression inflicted by the infection. The study by Bouqet et al. [15], which the gene expression data is obtained from, was the first longitudinal gene co-expression study of patients with Lyme disease. Therefore little is known about the transcriptomics involved in this disease. Additionally, the bacterium causing Lyme borreliosis does not in itself affect those infected, but it initiates inflammatory responses that cause damage to the body. It was therefore expected to be challenging to find any gene related patterns specific for this disease. However, the outcome of performing differential co-expression analysis on the gene expression data provided new insight about the co-expression network and possible genes central to the illness.

The constructed network assessed from applying the CSD method identified significant association between differently expressed genes in those infected with *B. burgdorferi*. The DEGs showing contrasting correlation in expression between the two conditions, this being the specific and differentiated interaction types, were most prominent in the network, thus implying that abnormalities in gene expression patterns occur in patients with Lyme disease. All four networks constructed had a heavy-tailed degree distribution, showing a non-random topology in the network. This non-random topology may reflect that the relationships connecting the DEGs is a result of biological function. However, it is important to emphasize that the presence of a link in this network is not related to any direct biological interaction, but merely an evidence of equal patterns, as stated in [4].

The slight difference in network topology between the different C-, S-, and D-type networks may indicate possible differences regarding the regulatory mechanisms involved. While all the three networks show disassortative topology they do so with varying degree.

The C-type network consist of a densely connected core with two outlying clusters, and is only slightly disassortative. While the S-type networks seems to be very tight and clustered it is the most disassortative sub-network. This is mostly due to the large hubs of the network GTF3C2 and HKDC1 which share connection with the majority of the genes in the cluster, an many of these are of low degree. Lastly, the D-type network show the same disassortative pattern as the S-type Network, but is less dense and tightly connected than the two other networks. The large hubs of the S- and D-networks are mainly responsible for their more extensive disassortative tendencies. These results support the hypothesis suggested by Voigt and Almaas [4], that tightly clustered regions, where high-degree nodes are more prone to sharing links, may be more robust and, hence, more less susceptible to change.

Central genes in the network are likely to be of functional biological importance in the networks. The CSD-method successfully constructed a network with central hubs (Table 4.1) whose association with others could have impact on the response to infection with the *B. burgdorferi* bacteria. Many of the hubs encoded proteins with functions relating to DNA transcription, metabolism, regulatory functions and the immune processes. The hubs relating to the latter function; S1PR1, LRCH1, GSDMA, PARP14 and PLCD4 had many S- and D-type links in the network. Expression of S1PR1 is also found to be elevated in another tick-born disease. There were many homogeneous hubs identified. Most noticeable were the two largest hubs; GTF3C2 and HKDC1, which were completely dominated by specific (S) interaction types. Neither of these genes have been previously associated with the Lyme disease, nor functions related to immune responses. However, GTF3C2 part upon viral detection and HKDC1 suggested part for induced fuel during inflammation may be central in the disease pathogenesis. Also interesting, was that GTF3C2 was found to be strongly co-expressed with the Lyme disease associated gene CXCL10 under one condition. Altogether, CXCL10 were found to be connected to six of the hubs with S-type links. The identified hubs of the CSD-network are thus suggested to be good candidates for further studies.

The results from the gene ontology enrichment analysis (Table 4.2 - 4.3) provided insight into the behaviour of the genes identified in the CSD-network. In addition, it provided insight of the quality of the network. Substantial enrichment of biological processes were identified for the complete CSD-network and the sub-network consisting solely of D-type links. The enriched processes identified in these two networks related to cell cycle, protein transport and localization and regulation of metabolic processes. However, neither of these processes have proven to directly link, that I know of, to the pathogenesis of Lyme disease. The results from the analysis may therefore suggest that the CSD-network failed to represent real, meaningful relationships, relating to the infection of the disease. However, the infection with *B. burgdorferi* doesn't only trigger immune responses, and because the biological processes were found highly enriched in the D-network as well, this may indicate that expression of genes in these processes significantly change in the disease. For example, an extensive number of biological processes enriched in both networks related to metabolic regulation and protein transport and localization. These processes may be enriched because of the transportation and localization of immune cells to the site of infection. The regulation of metabolism has also been found to be important for fueling

of the immune cells, and to fuel *B. burgdorferi* in the host [58]. Several of the biological processes were also related to cell cycle. The transcriptomic data is obtained from whole blood samples and the fact that blood is composed by by 45 % plasma, %96 erythrocytes, 1 % leukocytes and 3 % platelets, all having functions such as transport, immune response and signaling [80], makes it difficult to determine whether these GO-terms are disease related or not. However, the results from the cell-type enrichment analysis using the enrichr website indicated that the biological processes were happening in cells related to the immune responses (Table 4.4). The lack of significantly over-represented biological processes for the two other sub-network may be a result of the limitations of the gene ontology analysis when performing it on a smaller set of nodes. A biological process is only enriched if among the genes investigated, there are more members of that process than what would be expected by chance. This criteria becomes very strict when only a small number of genes are looked into.

The majority of studies performed on the transcriptomics of Lyme disease have identified the up-regulation of toll-like receptors, interferons, cytokines and chemokines. However, the CSD-framework failed to detect the majority of these previously associated genes. This presumed loss of data may be a result of limitations due to the chosen gene expression data set. Some data could be lost due to the high occurrence of zeroes present in the expression data, which was filtered out during the data processing steps. If, for pairwise genes, one gene is expressed while the other gene is silent (i.e. have many zero entities) this pair of genes should be marked as uncorrelated. However, the gene with the many zero entities may have been filtered out. If, for the other situation, the pair of genes are highly correlated an S-type should have been between them, but is instead lost. The smaller the sampling size for the data sets are, the larger the impact of zero entities may be. Since, the expression data for the patients and healthy controls contained 55 and 13 sample points respectively, loss of data is a reasonable assumption. Additionally, some of the sample points from the gene expression data were obtained from patients after completion of antibiotics. While the study, providing the gene expression data, concluded with the persistence of DEGs expression patterns, some loss may have occurred. The method did, however, successfully identify a few of these associated genes, and thus indicate some form of validation for the method employed. Also, the CSD method captured many of the family member of previously associated genes such as CXCL3, CCL20, IL-16 and MMP14. MMP14 was also identified as a hub in the network almost exclusively dominated by specific interaction types. These seem like interesting genes to pay more attention to i further research of the mechanisms involved in Lyme disease.

The CSD-method applies a strict threshold for the construction of the differentially co-expressed network, ensuring that only persistent co-expressed relationships among gene pairs are included in the network. For this reason, all genes, especially those of the S- and D-networks, would be of interest in further studies of Lyme disease. The method successfully identified genes that functions during the inflammatory response initiated as a response to harmful stimuli. For further studies, it would be interesting to compare the results from this analysis with expression data from other infectious diseases causing inflammatory responses. By performing such comparative studies, genes showed to be uniquely expressed in this network could be marked as biomarker specifically for Lyme

borreliosis. Also, a new performance of the CSD-network on a gene expression data set with more sample points, where all the sample points were taken prior to treatment, may yield better results.

5.2 Network Analysis

The second aim of the thesis was to investigate additional steps for network analysis that could accompany the CSD-method in order to extract more information about the shared co-expression patterns in patients with Lyme disease. It is clear from the results that much more can be understood about the network, and its constituents by partitioning the network into modules, making it a valid first step for the further network analysis.

All four algorithms included in the NetworkX package successfully partitioned the network into separate modules revealing dense clusters of genes. Their global modularity score suggest that the detected modules were far from random, and may thus represent clusters of genes related to the same biological processes. An interesting finding was that similar interaction types tend to be in the same modules. This was especially true for the modules containing differentiated or specific interaction types. This may indicate that these genes closely relate to the biological functions of the pathogenesis. A great next step for the analysis would thus be to perform a gene ontology enrichment analysis on the detected modules. However, all except one module, were not significantly enriched with any biological processes. This is likely a result of the limitations of the GO-analysis, stated earlier, that when the analysis is performed on a smaller set of genes the criteria for significant enrichment may become to strict. However, while gene ontology enrichment analysis did not provide significant results, information could be extracted from looking into the central nodes of each module. By doing so, new central nodes were highlighted. Even though they did not meet the criteria for being hubs of the network they are still of high degree, suggesting their involvement in Lyme disease. The genes; TBC1D2B, GNAZ, COPG1, PSMD13, SAR1A and LBR showed pairwise gene correlation of both differentiated and specific interactions, suggesting that they may have a prominent role in the mechanisms of the disease. These genes may therefore be good candidates as well for further analysis.

The different algorithms did, however, partition the network in different ways, because of their different approaches. Which method to prefer may vary from network to network, and thus a good additional step was to calculate the global modularity score, thus providing insight into which of the methods that provided the most quality modules for this network. For this network, this was the Louvain community detection, which divided the network components into modules by calculating the locally modularity score. However, it is important to emphasize that both the LCA and GN methods provided several ways of partitioning the network, and not only the ones portrayed in Fig 4.10 and 4.11. The LCA had the parameter ”randomize” which if set to ”TRUE”, would randomize the node evaluation order and community evaluation in order to get different partitions at each call. Thus, it is not certain that the chosen partition were the best one. The GN-method was the the only divisive algorithm applied to the CSD-network, and detected modules by calculating the edge betweenness for every node pair in the network, and removing the highest

one at each iteration step before recalculating the edge betweenness. These iteration steps would continue until all the nodes were divided into individual modules unless otherwise specified. It was therefore up to the user to determine which partitioning of the network to use. The user could decide to stop the algorithms after a certain amount of iteration or modules detected. For this network the iteration was run until the network was divided into 90 partitions. Thus, choosing a partitioning at another iteration steps would yield a different set of modules. It is therefore an high degree of variance from using these algorithm based and a more optimal approach for these methods would be to calculate the modularity score for each partitioning, selecting the one with highest score. This is though not necessary for the GMM algorithm and LPA as they returned the same partition at each call.

The last suggested step: accompanying the CSD-network analysis with WGCNA, also provided new valuable insight into the co-expression patterns in patients with Lyme disease. Five modules detected in the topological overlap expression data were identified based on the WGCNA which are significantly related to immune responses and defense response to bacteria. Gene ontology enrichment analysis revealed that the antiquewhite4 and lightyellowmodule were implicated in the immune responses through toll-like receptor signaling pathways, macrophages responses and secretion of interleukin's. The genes of the darkslateblue module were involved in B cell processes, the yellow4 module in response to bacteria and the darkolivegreen module in viral defense. It is known that *B. burgdorferi* possess lipoproteins that interacts with toll-like receptors (TLRS) on the surface of cells involving in the innate immune system, to cause them to release inflammatory products that results in the tissue damage of Lyme disease [14, 15, 71, 72, 81]. These lipoproteins have also been found to directly stimulate the B cell responses [14, 15].

Some of the genes found in each module, and also recognized in the CSD-network encoded proteins with functions related the immunity processes (4.13). From the antiquewhite4 module, the Lyme disease associated gene CXCL10 was identified, along with the hub, MMP14. The suggestion of the involvement of MMP14 is thus strengthen from the WGCNA analysis. 60 genes from the lightyellow modules were recognized in the CSD-network. While all of these may play a role in the pathogenesis, the most interesting genes identified were ITGAX and LAIR1. Both of these genes were completely dominated by S-type links and connecting to the two largest hubs; GTF3C2 and HKDC1. While ITGAX may be involved in the cell-cell interactions during the inflammation after infection, LAIR1 may be suppressed during the immune response as it plays fundamental negative regulatory roles in the immune system.

From the three modules, darkslateblue, yellow4 and darkolivegreen, very few genes were also recognized in the CSD-network. Of the genes recognized from the darkslateblue module, three were RNA coding genes and therefore not encoding any protein products. The last of the gene, DENND58, were the only one of the genes encoding a protein. However, it shows exclusively conserved association, and is thus unlikely to be of central involvement in the biological mechanisms of Lyme disease. While the rest of these genes encodes RNA, they have specific and differentiated associations with other genes, and may play a regulatory role in the disease. In the yellow4 module, only two genes were recognized and only DEFA1 had functions relating to the biological processes enriched. This gene,

however, is of special interest as it is part of family of antimicrobial and cytotixic peptides involved in host defense. While its function suggests it's highly relevance to mechanism of antimicrobial infection, its connection to the network is modest, with only one D-type link connected to another gene in the network. Lastly, 12 genes were recognized in the darkolivegreen module, where two could be related to the immune responses, one of them being the hub PARP14. Both PARP14 and DDX60L can be suggested to play a role in Lyme disease through the innate immune system.

Since WGCNA is performed on a much larger data set it didn't become as much affected by the limitations of the gene ontology enrichment analysis. While many of the detected genes in the modules were not recognized in the CSD-network, it did successfully highlight 159 genes that could be specifically related to the clinical mechanisms for Lyme disease. All of these genes, and especially those directly related to immune response and host defense can be possible biomarkers for people infected by the bacteria. The WGCNA analysis is therefore a valid step for network analysis that can reveal a set of central genes and biological processes involved in the disease investigated.

It is necessary to have experience with both Python and R programming to perform these steps. Having to toggle between different languages and formats therefore makes them more time consuming. For further work, it could be an idea to develop a pipeline, in for example python, which implements all of these steps. A menu could be developed, where you have an edge list as input and then you could choose which community algorithm to use. The developed program would perform the necessary steps and calculations, and print the modules out to text files which easily can be implemented into Cytoscape. The program could always calculate the overall modularity score so it would be simple to choose the partitions of highest quality. Also, if possible, it would be interesting if, in addition to perform weighted gene correlation analysis on the topological overlap matrix, it could be done on the constructed CSD-network.

Chapter 6

Conclusion

The aim of this thesis was two-folded. In the first part of the thesis the CSD framework was applied to expression data from patients diagnosed with Lyme disease in order to identify differences in gene expression when comparing the data to those of healthy individuals. This was successfully done by using the CSD framework. The resulting network found suggestive Lyme disease related biological mechanisms such as protein transport and localization, regulation of metabolic processes and cell cycle. It also identified two genes previously associated with the disease. 21 hubs were identified in the network: GTF3C2, HKDC1, CCDC157, TNNC2, KHK, LOC440434, CD109, S1PR1, PURA, MMAA, ZNF702P, LRCH1, MBD2, TMEM161B-ASI, NAID1, GSDMA, PARP14, OR6WIP, CCDC85C, PLCD4 AND MMP14. Six of the green hubs were directly linked to one Lyme disease associated gene, including the two largest hubs, suggesting their prominent role in central pathways of the disease mechanism. The hubs of the network is thus good candidates for further studies. Additionally, related family members of previously disease associated genes, such as CXCL3, CCL20 and IL-16 could be interesting to look into as well. Further study is evidently needed in to determine their possible role in the disease mechanism, but the findings using an CSD approach clearly provide some clues to related genes, and thus providing a basis for further analysis.

Further work is necessary to establish the complex mechanisms that underlie Lyme disease. It could be of interest to look further into hubs identified by the CSD-method, and to look more in detail about their functionality in the cells, and if they show similar co-expression patterns in other infectious diseases. It would also be interesting to apply the CSD-method to a dataset including gene expression from patients of other infectious diseases as well. By comparing the result from these two constructed CSD-networks it could be possible to identify genes that are specific for each infection. Also, it could be of interest to further investigate the specific region of the network in more detail, as it is suggested to play in an important role in the mechanisms of the disease. Finally, comparing the CSD-network to a protein-protein interaction network could identify possible direct interacting

links related to biological function between the genes in the network.

In the second part of this thesis the aim was to define a set of steps for network analysis that could accompany the CSD-method in order to extract more information about the shared co-expression patterns. The first suggestive step; partition the network into modules using community detection algorithm, successfully provided valuable insight into the neighborhoods of the CSD-network, and their central genes, that is not easily detected when looking at the network as whole. The modularity score ranked the Louvain community detection to be the best algorithm for this network, revealing eight modules, and three new central genes; PSMD13, SAR1A and LBR. The gene ontology analysis was the next proposed step, and is determined as useful, however with a keeping in mind that it could provide minimal results. Finally, the application of WGCNA on the RNA-seq contributed to large insight into the relevant biological processes enriched without information loss. The WGCNA identified 159 DEGs whose role may be suggested to play part in the pathogenesis of Lyme Disease. Of these 159 DEGs, 24 were of special interest as they encode proteins functioning in the immune response, inflammation and defense against bacteria, making them an interesting focus for further research.

The concluding steps is thus as follows; partitioning the network into modules with different approaches and then calculating the modularity score to determine the partition most fit for the network to be investigated. Then, a gene ontology analysis can be performed for insight into the behaviour of the genes of the detected modules. Finally, WGCNA could be applied on the RNA-seq to get more information about the genes and biological processes involved without the information loss. While these steps provided valuable insight into the network topology and the shared co-expression patterns, the process of applying the steps were troublesome at times. The development of a pipeline, in for example Python, is therefore suggested for further work, as it would make the process more user friendly, and it would be no need to toggle between several programming languages. As the goal was to identify a set of steps for network analysis, the possibilities for further work is unlimited. One possibility is to look into additional steps, which does not involve community detection, such as centrality, robustness and pathway enrichment. Also, only a selection of the community detection algorithms that exist was investigated in this thesis, so more research could be done on other existing methods.

Bibliography

[1] Lyme borreliose: veileder for helsepersonell. 2019 †27. august 2019 [cited 02.mars 2020]. In: Smitteveiledern [Internet]. Oslo: Folkehelseinstituttet, [cited 02.mars 2020]. Available from: https://www.fhi.no/nettpub/smittevernveilederen/sykdommer-a-a/lyme-borreliose—veileder-for-hels/.;.

[2] UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research. 2018 Nov;47(D1):D506–D515. Available from: https://doi.org/10.1093/nar/gky1049.

[3] Barabási AL, Pósfai M. Network science. Cambridge: Cambridge University Press; 2016. Available from: http://barabasi.com/networksciencebook/.

[4] Voigt A, Nowick K, Almaas E. A composite network of conserved and tissue specific gene interactions reveals possible genetic interactions in glioma. PLOS Computational Biology. 2017 09;13(9):1–34. Available from: https://doi.org/10.1371/journal.pcbi.1005739.

[5] Marie Gulla. An integrated systems biology approach to investigate transcriptomic data of thyroid carcinoma. NTNU; 2019.

[6] Voigt A. CSD. GitHub; 2017. https://github.com/andre-voigt/CSD.

[7] Ramage M, Shipp K. In: Ludwig von Bertalanffy. London: Springer London; 2020. p. 53–62. Available from: https://doi.org/10.1007/978-1-4471-7475-2_6.

[8] Trewavas A. A Brief History of Systems Biology. The Plant Cell. 2006 Oct;18(10):2420–2430. Available from: https://doi.org/10.1105/tpc.106.042267.

[9] Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of Modern Physics. 2002 Jan;74(1):47–97. Available from: https://doi.org/10.1103/revmodphys.74.47.

[10] Kovács IA, Luck K, Spirohn K, Wang Y, Pollis C, Schlabach S, et al. Network-based prediction of protein interactions. Nature Communications. 2019 Mar;10(1). Available from: https://doi.org/10.1038/s41467-019-09177-y.

[11] Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. Nature Communications. 2019 Mar;10(1). Available from: https://doi.org/10.1038/s41467-019-09186-x.

[12] Stanek G, Wormser GP, Gray J, Strle F. Lyme borreliosis. The Lancet. 2012;379(9814):461 – 473. Available from: http://www.sciencedirect.com/science/article/pii/S0140673611601037.

[13] Hengge UR, Tannapfel A, Tyring SK, Erbel R, Arendt G, Ruzicka T. Lyme borreliosis. The Lancet Infectious Diseases. 2003 Aug;3(8):489–500. Available from: https://doi.org/10.1016/s1473-3099(03)00722-9.

[14] Steere AC, Coburn J, Glickstein L. The emergence of Lyme disease. Journal of Clinical Investigation. 2004 Apr;113(8):1093–1101. Available from: https://doi.org/10.1172/jci21681.

[15] Bouquet J, Soloski MJ, Swei A, Cheadle C, Federman S, Billaud JN, et al. Longitudinal Transcriptome Analysis Reveals a Sustained Differential Gene Expression Signature in Patients Treated for Acute Lyme Disease. mBio. 2016 Feb;7(1). Available from: https://doi.org/10.1128/mbio.00100-16.

[16] Steere AC. Lyme Disease. New England Journal of Medicine. 2001 Jul;345(2):115–125. Available from: https://doi.org/10.1056/nejm200107123450207.

[17] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009 Jan;10(1):57–63. Available from: https://doi.org/10.1038/nrg2484.

[18] Molecular Biology. Elsevier; 2013. Available from: https://doi.org/10.1016/c2009-0-01986-2.

[19] Guo Y, Li CI, Ye F, Shyr Y. Evaluation of read count based RNAseq analysis methods. BMC Genomics. 2013;14(Suppl 8):S2. Available from: https://doi.org/10.1186/1471-2164-14-s8-s2.

[20] Voit, E. (2018). A First Course in Systems Biology. New York: Garland Science, https://doi.org/10.4324/9780203702260;.

[21] Goymer P. Why do we need hubs? Nature Reviews Genetics. 2008 Sep;9(9):651–651. Available from: https://doi.org/10.1038/nrg2450.

[22] Newman MEJ. Assortative Mixing in Networks. Phys Rev Lett. 2002 Oct;89:208701. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.89.208701.

[23] Golbeck J. Analyzing networks. In: Introduction to Social Media Investigation.

Elsevier; 2015. p. 221–235. Available from: https://doi.org/10.1016/b978-0-12-801656-5.00021-4.

[24] Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 2004 Feb;69:026113. Available from: https://link.aps.org/doi/10.1103/PhysRevE.69.026113.

[25] Kumari S, Nie J, Chen HS, Ma H, Stewart R, Li X, et al. Evaluation of Gene Association Methods for Coexpression Network Construction and Biological Knowledge Discovery. PLOS ONE. 2012 11;7(11):1–17. Available from: https://doi.org/10.1371/journal.pone.0050411.

[26] Akoglu H. Users guide to correlation coefficients. Turkish Journal of Emergency Medicine. 2018 Sep;18(3):91–93. Available from: https://doi.org/10.1016/j.tjem.2018.08.001.

[27] Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical Organization of Modularity in Metabolic Networks. Science. 2002;297(5586):1551–1555. Available from: https://science.sciencemag.org/content/297/5586/1551.

[28] Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. Statistical Applications in Genetics and Molecular Biology. 2005 Jan;4(1). Available from: https://doi.org/10.2202/1544-6115.1128.

[29] Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences. 1998 Dec;95(25):14863–14868. Available from: https://doi.org/10.1073/pnas.95.25.14863.

[30] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec;9(1). Available from: https://doi.org/10.1186/1471-2105-9-559.

[31] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000 May;25(1):25–29. Available from: https://doi.org/10.1038/75556.

[32] The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research. 2018 11;47(D1):D330–D338. Available from: https://doi.org/10.1093/nar/gky1055.

[33] van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene–disease predictions. Briefings in Bioinformatics. 2017 Jan;p. bbw139. Available from: https://doi.org/10.1093/bib/bbw139.

[34] Farahbod M, Pavlidis P. Differential coexpression in human tissues and the confounding effect of mean expression levels. Bioinformatics. 2018 Jul;Available from: https://doi.org/10.1093/bioinformatics/bty538.

[35] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013 Jan;41(Database issue):D991-5;.

[36] National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2020 Feb 04].;Available from: https://www.ncbi.nlm.nih.gov/.

[37] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, Genome research 13, 2498 (2003), ISSN 1088-9051;.

[38] Gysi DM, Voigt A, de Miranda Fragoso T, Almaas E, Nowick K. wTO: an R package for computing weighted topological overlap and a consensus network with integrated visualization tool. BMC Bioinformatics. 2018 Oct;19(1). Available from: https://doi.org/10.1186/s12859-018-2351-7.

[39] Hagberg AA, Schult DA, Swart PJ. Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th Python in Science Conference. Pasadena, CA USA; 2008. p. 11 – 15.

[40] Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys Rev E. 2004 Dec;70:066111. Available from: https://link.aps.org/doi/10.1103/PhysRevE.70.066111.

[41] Cordasco G, Gargano L. Community detection via semi-synchronous label propagation algorithms. In: 2010 IEEE International Workshop on: Business Applications of Social Network Analysis (BASNA); 2010. p. 1–8.

[42] Girvan M, Newman MEJ. Community structure in social and biological networks. Proceedings of the National Academy of Sciences. 2002;99(12):7821–7826. Available from: https://www.pnas.org/content/99/12/7821.

[43] Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008 Oct;2008(10):P10008. Available from: http://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

[44] Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2007 Nov;24(5):719–720. Available from: https://doi.org/10.1093/bioinformatics/btm563.

[45] Sinn E, Wang Z, Kovelman R, Roeder RG. Cloning and characterization of a TFIIIC2 subunit (TFIIIC beta) whose presence correlates with activation of RNA polymerase III-mediated transcription by adenovirus E1A expression and serum factors. Genes & Development. 1995 Mar;9(6):675–685. Available from: https://doi.org/10.1101/gad.9.6.675.

[46] Li GH, Huang JF. Inferring therapeutic targets from heterogeneous data: HKDC1 is a novel potential therapeutic target for cancer. Bioinformatics. 2013 Oct;30(6):748–752. Available from: https://doi.org/10.1093/bioinformatics/btt606.

[47] Delahay RM, Frankel G. Coiled-coil proteins associated with type III secretion systems: a versatile domain revisited. Molecular Microbiology. 2002 Aug;45(4):905–916. Available from: https://doi.org/10.1046/j.1365-2958.2002.03083.x.

[48] Litvinov IV, Bizet AA, Binamer Y, Jones DA, Sasseville D, Philip A. CD109 release from the cell surface in human keratinocytes regulates TGF- receptor expression, TGF- signalling and STAT3 activation: relevance to psoriasis. Experimental Dermatology. 2011 May;20(8):627–632. Available from: https://doi.org/10.1111/j.1600-0625.2011.01288.x.

[49] Song G, Feng T, Zhao R, Lu Q, Diao Y, Guo Q, et al. CD109 regulates the inflammatory response and is required for the pathogenesis of rheumatoid arthritis. Annals of the Rheumatic Diseases. 2019;78(12):1632–1641. Available from: https://ard.bmj.com/content/78/12/1632.

[50] Allende ML, Dreier JL, Mandala S, Proia RL. Expression of the Sphingosine 1-Phosphate Receptor, S1P1, on T-cells Controls Thymic Emigration. Journal of Biological Chemistry. 2004 Jan;279(15):15396–15401. Available from: https://doi.org/10.1074/jbc.m314291200.

[51] Kułakowska A, Byfield FJ, Żendzian-Piotrowska M, Zajkowska JM, Drozdowski W, Mroczko B, et al. Increased levels of sphingosine-1-phosphate in cerebrospinal fluid of patients diagnosed with tick-borne encephalitis. Journal of Neuroinflammation. 2014 Nov;11(1). Available from: https://doi.org/10.1186/s12974-014-0193-4.

[52] Xu X, Han L, Zhao G, Xue S, Gao Y, Xiao J, et al. LRCH1 interferes with DOCK8-Cdc42–induced T cell migration and ameliorates experimental autoimmune encephalomyelitis. Journal of Experimental Medicine. 2016 Dec;214(1):209–226. Available from: https://doi.org/10.1084/jem.20160068.

[53] Ding J, Wang K, Liu W, She Y, Sun Q, Shi J, et al. Pore-forming activity and structural autoinhibition of the gasdermin family. Nature. 2016 Jun;535(7610):111–116. Available from: https://doi.org/10.1038/nature18590.

[54] Alt F. Advances in Immunology. ISSN. Elsevier Science; 2018. Available from: https://books.google.no/books?id=255MDwAAQBAJ.

[55] Iwata H, Goettsch C, Sharma A, Ricchiuto P, Goh WWB, Halu A, et al. PARP9 and PARP14 cross-regulate macrophage activation via STAT1 ADP-ribosylation. Nature Communications. 2016 Oct;7(1). Available from: https://doi.org/10.1038/ncomms12849.

[56] Leung DW, Tompkins C, Brewer J, Ball A, Coon M, Morris V, et al. Molec-

ular Cancer. 2004;3(1):15. Available from: https://doi.org/10.1186/1476-4598-3-15.

[57] Lu N, Malemud CJ. Extracellular Signal-Regulated Kinase: A Regulator of Cell Growth, Inflammation, Chondrocyte and Bone Cell Receptor-Mediated Gene Expression. International Journal of Molecular Sciences. 2019 Aug;20(15):3792. Available from: https://doi.org/10.3390/ijms20153792.

[58] Kerstholt M, Netea MG, Joosten LAB. Borrelia burgdorferi hijacks cellular metabolism of immune cells: Consequences for host defense. Ticks and Tick-borne Diseases. 2020 May;11(3):101386. Available from: https://doi.org/10.1016/j.ttbdis.2020.101386.

[59] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14(1):128. Available from: https://doi.org/10.1186/1471-2105-14-128.

[60] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research. 2016 May;44(W1):W90–W97. Available from: https://doi.org/10.1093/nar/gkw377.

[61] Poli A, Michel T, Thérésine M, Andrès E, Hentges F, Zimmer J. CD56brightnatural killer (NK) cells: an important NK cell subset. Immunology. 2009 Apr;126(4):458–465. Available from: https://doi.org/10.1111/j.1365-2567.2008.03027.x.

[62] Soeno T, Katoh H, Ishii S, Ushiku H, Hosoda K, Hiki N, et al. CD33 Immature Myeloid Cells Critically Predict Recurrence in Advanced Gastric Cancer. Journal of Surgical Research. 2020 Jan;245:552–563. Available from: https://doi.org/10.1016/j.jss.2019.07.095.

[63] Kawamoto H, Minato N. Myeloid cells. The International Journal of Biochemistry & Cell Biology. 2004 Aug;36(8):1374–1379. Available from: https://doi.org/10.1016/j.biocel.2004.01.020.

[64] Murphy K. Janeway's Immunobiology. Garland Science; 2016. Available from: https://www.xarg.org/ref/a/B01D5ARZ90/.

[65] Seta N, Kuwana M. Derivation of multipotent progenitors from human circulating CD14 monocytes. Experimental Hematology. 2010 Jul;38(7):557–563. Available from: https://doi.org/10.1016/j.exphem.2010.03.015.

[66] Heilpern AJ, Wertheim W, He J, Perides G, Bronson RT, Hu LT. Matrix Metalloproteinase 9 Plays a Key Role in Lyme Arthritis but Not in Dissemination of Borrelia burgdorferi. Infection and Immunity. 2009 Apr;77(7):2643–2649. Available from: https://doi.org/10.1128/iai.00214-09.

[67] Gebbia JA, Coleman JL, Benach JL. Selective Induction of Matrix Metalloproteinases byBorrelia burgdorferivia Toll-Like Receptor 2 in Monocytes. The Jour-

nal of Infectious Diseases. 2004 Jan;189(1):113–119. Available from: https://doi.org/10.1086/380414.

[68] Vergnolle N, Chignard M, editors. Proteases and Their Receptors in Inflammation. Springer Basel; 2011. Available from: https://doi.org/10.1007/978-3-0348-0157-7.

[69] Brito GA, Souza MH, Melo-Filho AA, Hewlett EL, Lima AA, Flores CA, et al. Role of pertussis toxin A subunit in neutrophil migration and vascular permeability. Infect Immun. 1997 Mar;65(3):1114–1118.

[70] Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, et al. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature. 1997 Dec;390(6660):580–586. Available from: https://doi.org/10.1038/37551.

[71] Wooten RM, Ma Y, Yoder RA, Brown JP, Weis JH, Zachary JF, et al. Toll-Like Receptor 2 Is Required for Innate, But Not Acquired, Host Defense to Borrelia burgdorferi. The Journal of Immunology. 2002 Jan;168(1):348–355. Available from: https://doi.org/10.4049/jimmunol.168.1.348.

[72] Hirschfeld M, Kirschning CJ, Schwandner R, Wesche H, Weis JH, Wooten RM, et al. Cutting Edge: Inflammatory Signaling by Borrelia burgdorferi Lipoproteins Is Mediated by Toll-Like Receptor 2. The Journal of Immunology. 1999;163(5):2382–2386. Available from: https://www.jimmunol.org/content/163/5/2382.

[73] Stebegg M, Kumar SD, Silva-Cayetano A, Fonseca VR, Linterman MA, Graca L. Regulation of the Germinal Center Response. Frontiers in Immunology. 2018;9:2469. Available from: https://www.frontiersin.org/article/10.3389/fimmu.2018.02469.

[74] Beck G, Benach JL, Habicht GS. Isolation, preliminary chemical characterization, and biological activity of Borrelia burgdorferi peptidoglycan. Biochemical and Biophysical Research Communications. 1990 Feb;167(1):89–95. Available from: https://doi.org/10.1016/0006-291x(90)91734-a.

[75] Jutras BL, Lochhead RB, Kloos ZA, Biboy J, Strle K, Booth CJ, et al. Borrelia burgdorferi peptidoglycan is a persistent antigen in patients with Lyme arthritis. Proceedings of the National Academy of Sciences. 2019 Jun;116(27):13498–13507. Available from: https://doi.org/10.1073/pnas.1904170116.

[76] Mesev EV, LeDesma RA, Ploss A. Decoding type I and III interferon signalling during viral infection. Nature Microbiology. 2019 Apr;4(6):914–924. Available from: https://doi.org/10.1038/s41564-019-0421-x.

[77] Petzke MM, Brooks A, Krupna MA, Mordue D, Schwartz I. Recognition of-Borrelia burgdorferi, the Lyme Disease Spirochete, by TLR7 and TLR9 Induces a Type I IFN Response by Human Immune Cells. The Journal of Immunology. 2009 Sep;183(8):5279–5292. Available from: https://doi.org/10.4049/jimmunol.0901390.

[78] Bruhn O, Cauchard J, Schlusselhuber M, Gelhaus C, Podschun R, Thaller G, et al. Antimicrobial properties of the equine -defensin DEFA1 against bacterial horse pathogens. Veterinary Immunology and Immunopathology. 2009 Jul;130(1-2):102–106. Available from: https://doi.org/10.1016/j.vetimm.2009.01.005.

[79] Grünvogel O, Esser-Nobis K, Reustle A, Schult P, Müller B, Metz P, et al. DDX60L Is an Interferon-Stimulated Gene Product Restricting Hepatitis C Virus Replication in Cell Culture. Journal of Virology. 2015 Aug;89(20):10548–10568. Available from: https://doi.org/10.1128/jvi.01297-15.

[80] Mohr S, Liew CC. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends in Molecular Medicine. 2007 Oct;13(10):422–432. Available from: https://doi.org/10.1016/j.molmed.2007.08.003.

[81] Aliprantis AO, Yang RB, Mark MR, Suggett S, Devaux B, Radolf JD, et al. Cell Activation and Apoptosis by Bacterial Lipoproteins Through Toll-like Receptor-2. Science. 1999;285(5428):736–739. Available from: https://science.sciencemag.org/content/285/5428/736.

[82] Jones KL, Muellegger RR, Means TK, Lee M, Glickstein LJ, Damle N, et al. Higher mRNA Levels of Chemokines and Cytokines Associated with Macrophage Activation in Erythema Migrans Skin Lesions in Patients from the United States than in Patients from Austria with Lyme Borreliosis. Clinical Infectious Diseases. 2008 Jan;46(1):85–92. Available from: https://doi.org/10.1086/524022.

[83] Strle K, Stupica D, Drouin EE, Steere AC, Strle F. Elevated Levels of IL-23 in a Subset of Patients With Post-Lyme Disease Symptoms Following Erythema Migrans. Clinical Infectious Diseases. 2013 Nov;58(3):372–380. Available from: https://doi.org/10.1093/cid/cit735.

[84] Kovalchuka L, Eglite J, Lucenko I, Zalite M, Viksna L, Krumiņa A. Associations of HLA DR and DQ molecules with Lyme borreliosis in Latvian patients. BMC Research Notes. 2012 Aug;5(1). Available from: https://doi.org/10.1186/1756-0500-5-438.

[85] Peng Y, Liang Z, Liu A, Li E, Dai X, Bai R, et al. Borrelia burgdorferi basic membrane protein A stimulates murine macrophage to secrete specific chemokines. International Journal of Medical Sciences. 2018;15(13):1473–1479. Available from: https://doi.org/10.7150/ijms.26657.

[86] Groom JR, Luster AD. CXCR3 in T cell function. Experimental Cell Research. 2011 Mar;317(5):620–631. Available from: https://doi.org/10.1016/j.yexcr.2010.12.017.

Appendices

A.1 Genes previously associated with Lyme Disease

Table A1: Lyme disease associated genes. A list of 19 genes previously associated with Lyme disease.

TLR1 [71, 72]	CXCL1[82]	CXCL11 [82]	IFNB1[83]
TLR2 [71, 72]	CCL3 [82]	IL-6 [82]	IRF7[83]
HLA-DQB1 [84]	CCL4[82]	IL-10[82]	CXCL2 [85]
HLA-DRB1 [84]	CXCL9[82]	IL-23 [83]	CXCR3[86]
MMP-9 [66, 67]	CXCL10[82]	IFNA1[83]	

A.2 The C-, S- and D-networks

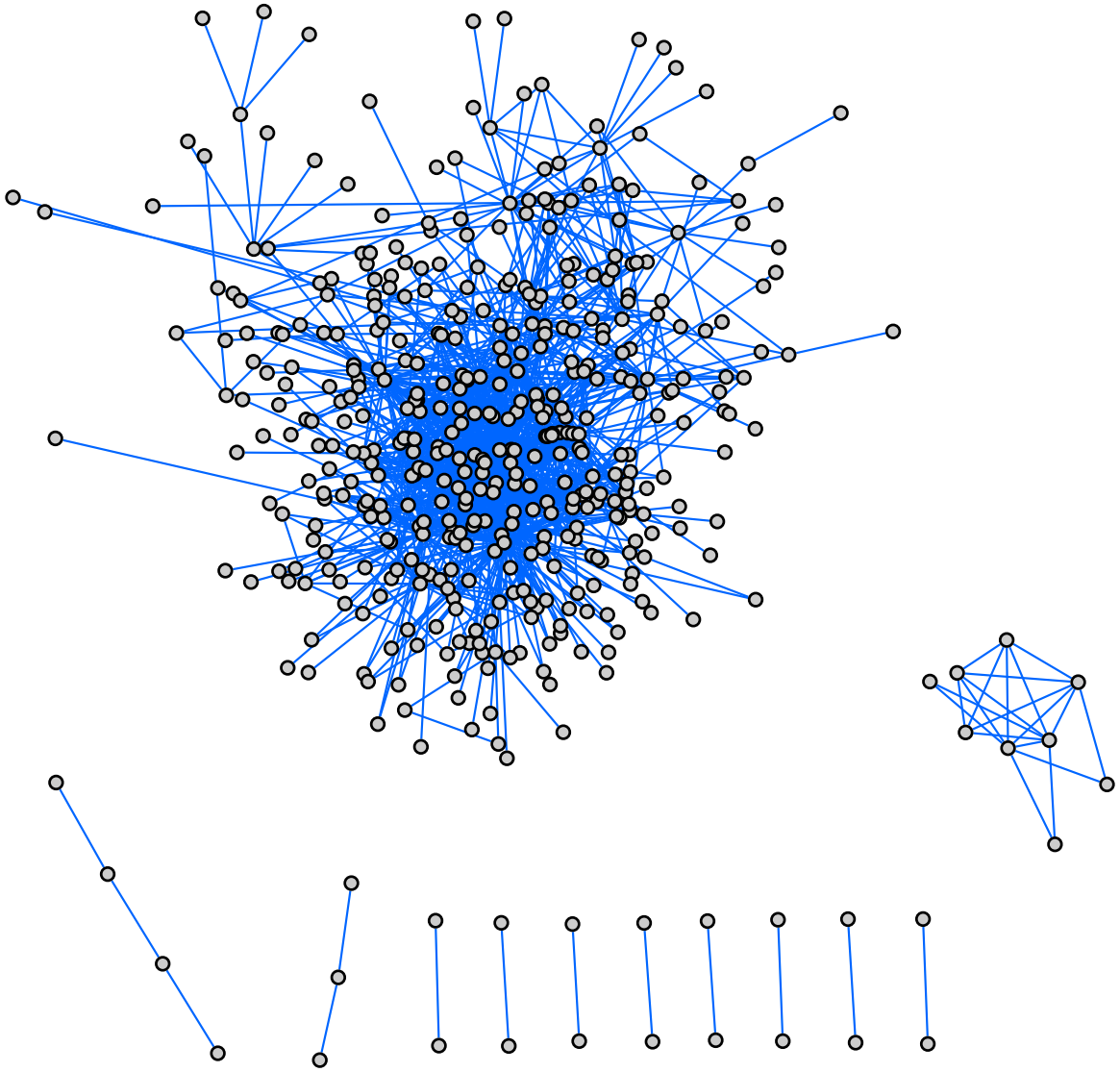


Figure A1: C-network. Visualization of all C-type links and nodes they connect to in the CSD-network.

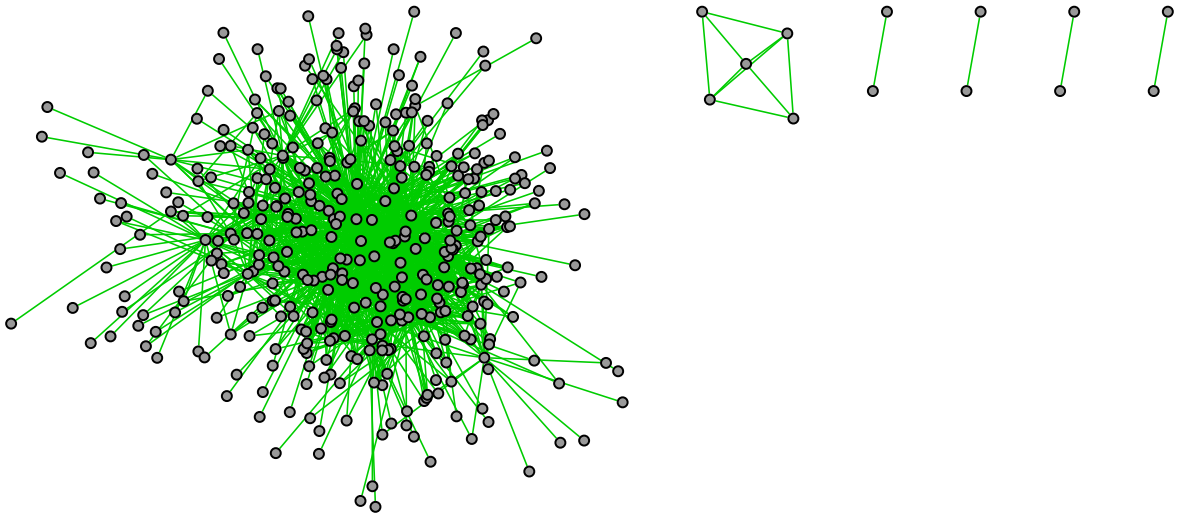


Figure A2: S-network Visualization of all S-type links and nodes they connect to in the CSD-network

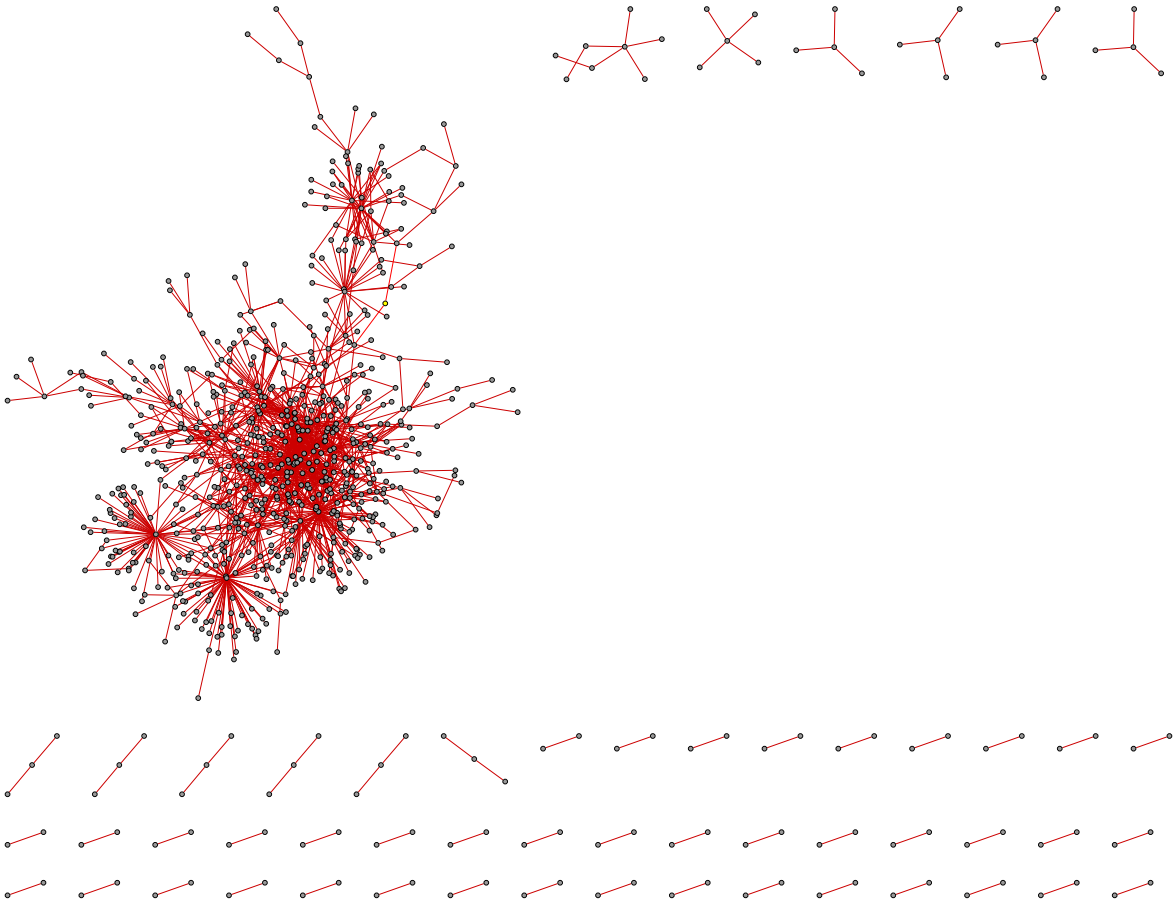


Figure A3: D-network Visualization of all D-type links and nodes they connect to in the CSD-network

Appendix A.3

Table A2: Hubs and their associated biological function. A description of the hubs and the function of their gene product. The source of information about functions is UniProtKB knowlegdebase [2], otherwise citations are provided

Hub	Associated biological function
GTF3C2	Encodes a general transcription factor essential for RNA polymerase III to make 5S RNA, tRNA and adenovirus RNA of both cellular and viral origin [45]
HKDC1	Encodes for one of five hexokinase that phosphorylates hexose sugars. Suggested to be up-regulated during stress to overcome deficiency of energy [46]
CCDC157	coiled coil domain
TNNC2	The central regulatory protein of striated muscle contraction
KHK	Catalyzes the phosphorylation of ketose sugar fructose to fructose-1-phosphate
LOC44034	Non-coding RNA
CD109	A cell surface antigen, attaching to the surfaces of platelets, activated T cells and endothelial cells. Modulates negatively TGFB1 signaling in keratinocytes. Found to be up-regulated in inflammatory responses [49, 48]
S1PR1	G- coupled receptor highly expressed in endothelial cells. S1PR1 signaling is required for normal egress of mature T-cell from thymus. Have been found up-regulated in patients infected with other tick-borne diseases [50, 51]
PURA	A sequence-specific (preferentially purine rich elements (PURA), single stranded DNA binding protein. Involved in the control of DNA replication and transcription
MMAA	GTPase that binds and hydrolyzes GTP. Involved in intracellular vitamin B12 metabolism, mediates the transport of cobalamin (Cbl) into mitochondria for the final steps of adenosylcobalamin (AdoCbl) synthesis.
ZNF702P	A zinc finger protein. May be involved in transcriptional regulation
LRCH1	Involved in the restraining og T-cell migration by negative regulation of the GTPase CDC42
MBD2	Encodes a nuclear protein involved in DNA methylation, capable of binding specifically to methylated DNA. Acts as a transcriptional repressor and is involved in gene silencing
TMEM161B - ASI	Encodes a transmembrane protein
NAIF1	Encodes a nuclear apoptosis inducing factor. Induces apoptosis

GSDMA	A protein coding gene that probably is involved in promotion of pyroptosis. A recent study has proposed its involvement in defending the skin by inducing pyrpotosis. [54] The protein also binds and kills bacteria
PARP14	Encodes an ADP-ribosyltransferase. Plays a part in the activation of macrophages in response to stimulation by IL4
OR6W1P	Pseudogene
CCDC85C	A coiled coil domain. May be involved in cell-cell adhesion and epithelium development.
PLCD4	Encodes a member of phospholipase C enzymes, and plays a critical role in many cellular process. E.g. involved in the Erk signaling pathway and proliferation.
MMP14	Encodes an endopeptidase whose function is to degrade a variety of components of the extracellular matrix. Also acting as a positive regulator for cell growth and migration.

Appendix A.4

Table A3: All GO-terms from CSD-network module. The number of related genes in the reference list (ref), number of relevant genes from the network (genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process	#ref	#genes	expected	+/-	FE	raw P-value	FDR
valyl-tRNA aminoacylation	6	6	.35	+	17.22	1.87E-05	6.50E-03
mitotic nuclear division	142	22	8.25	+	2.67	9.52E-05	2.62E-02
response to ionizing radiation	145	22	8.42	+	2.61	2.02E-04	4.61E-02
Chromosome segregation	275	34	15.97	+	2.13	1.21E-04	3.02E-02
mitotic cell cycle process	668	73	38.80	+	1.88	1.48E-06	8.72E-04
cell cycle process	1050	113	60.98	+	1.85	2.68E-09	6.12E-06
response to radiation	437	47	25.38	+	1.85	1.60E-04	3.93E-02
microtubule-based process	727	78	42.22	+	1.85	1.17E-06	7.75E-04
mitotic cell cylce	746	79	43.33	+	1.82	1.57E-06	8.97E-04
microtubule cytoskeleton organization	540	57	31.36	+	1.82	5.14E-05	1.55E-02
regulation of mitotic cell cycle	683	68	39.67	+	1.71	5.64E-05	1.67E-02
cell cycle	1371	133	79.63	+	1.67	3.93E-08	5.71E-05
regulation of cell cycle	1240	117	72.02	+	1.62	1.01E-06	7.01E-04
DNA metabolic process	756	71	43.91	+	1.62	1.76E-04	4.25E-02
intracellular protein transport	1011	94	58.72	+	1.60	2.38E-05	7.90E-03
cellular macromolecule localization	1651	153	95.89	+	1.60	4.35E-08	5.78E-05
cellular protein localization	1643	152	95.42	+	1.59	5.48E-08	5.83E-05
protein localization	2208	194	128.24	+	1.51	2.43E-08	3.87E-05
peptide transport	1566	135	90.95	+	1.48	1.13E-05	4.50E-03
protein transport	1535	132	89.15	+	1.48	1.58E-05	5.59E-03
cytoskeleton organization	1187	102	68.94	+	1.48	1.90E-04	4.38E-02
amide transport	1600	137	92.93	+	1.47	1.36E-05	5.17E-03
establishment of protein localization	1620	138	94.09	+	1.47	1.54E-05	5.57E-03
nitrogen compound transport	1879	160	109.13	+	1.47	3.14E-06	1.35E-03
cellular localization	2999	255	174.18	+	1.46	1.28E-09	4.09E-06
macromolecule localization	2549	215	148.04	+	1.45	7.50E-08	7.04E-05
intracellular transport	1546	128	89.79	+	1.43	1.13E-04	2.91E-02
organelle organization	3481	287	202.17	+	1.42	1.62E-09	4.31E-06
cellular component assembly	2331	192	135.38	+	1.42	2.01E-06	9.74E-04
establishment of localization in cell	2387	195	138.63	+	1.41	2.58E-06	1.21E-03
cellular component biogenesis	2590	211	150.42	+	1.40	1.21E-06	7.40E-04
organic substance transport	2234	178	129.75	+	1.37	3.60E-05	1.15E-02
regulation of catalytic activity	2393	185	138.98	+	1.33	1.08E-04	2.88E-02
regulation of cellular component organization	2428	187	141.02	+	1.33	1.19E-04	3.03E-02
cellular component organization	5536	424	321.52	+	1.32	3.47E-10	1.85E-06

cellular component organization or biogenesis	5754	437	334.19	+	1.31	4.53E-10	1.81E-06
regulation of molecular function	3073	230	178.48	+	1.29	9.58E-05	2.59E-02
cellular nitrogen compound metabolic proces	3464	259	201.18	+	1.29	2.79E-05	9.09E-03
cellular macromolecule metabolic process	5146	375	298.87	+	1.25	1.70E-06	9.07E-04
localization	5802	420	336.97	+	1.25	4.79E-07	3.82E-04
establishment of localization	4680	337	271.81	+	1.24	2.26E-05	7.67E-03
transport	4550	324	264.26	+	1.23	8.71E-05	2.44E-02
regulation of cellular metabolic process	6284	445	364.97	+	1.22	1.95E-06	9.73E-04
negative regulation of cellular process	4888	344	283.89	+	1.21	1.13E-04	2.95E-02
nitrogen compound metabolic process	7072	495	410.73	+	1.21	9.85E-07	7.15E-04
regulation of primary metabolic process	6077	425	352.94	+	1.20	1.37E-05	5.09E-03
regulation of metabolic process	6772	473	393.31	+	1.20	3.10E-06	1.37E-03
macromolecule metabolic process	6248	436	362.88	+	1.20	1.19E-05	4.65E-03
regulation of nitrogen compound metabolic process	5891	409	342.14	+	1.20	4.86E-05	1.49E-02
primary metabolic process	7537	521	437.74	+	1.19	1.89E-06	9.75E-04
cellular metabolic process	7742	535	449.65	+	1.19	1.17E-06	7.47E-04
organic substance metabolic process	7917	546	459.81	+	1.19	9.58E-07	7.28E-04
negative regulation of biological process	5541	382	321.81	+	1.19	1.85E-04	4.35E-02
metabolic process	8492	585	493.20	+	1.19	2.30E-07	1.93E-04
regulation of macromolecule metabolic process	6215	427	360.96	+	1.18	7.30E-05	2.12E-02
regulation of cellular process	11288	734	655.59	+	1.12	1.10E-05	4.49E-03
regulation of biological process	11825	766	686.78	+	1.12	7.34E-06	3.08E-03
biological regulation	12525	809	727.44	+	1.11	2.73E-06	1.24E-03
cellular process	15458	987	897.78	+	1.10	4.02E-09	8.01E-06
biological_process	17810	1111	1034.38	+	1.07	7.94E-11	6.34E-07
nervous system process	1403	50	81.48	-	.61	2.12E-04	4.75E-02
Unclassified	3041	100	176.62	-	.57	7.94E-11	1.27E-06
detection of stimulus	704	17	40.89	-	.42	4.20E-05	1.31E-02
detection of stimulus involved in sensory perception	546	8	31.71	-	.25	1.63E-06	8.95E-04
sensory perception of chemical stimulus	540	6	31.36	-	.19	1.01E-07	8.96E-05
detection of chemical stimulus	519	5	30.14	-	.17	5.28E-08	6.02E-05
detection of chemicals stimulus involved in sensory perception	483	4	28.05	-	.14	5.22E-08	6.41E-05
detection of chemical stimulus involved in sensory perception of smell	439	3	25.50	-	.12	6.90E-08	6.88E-05
sensory perception of smell	468	3	27.18	-	.11	1.58E-08	2.80E-05
complement activation, classical pathway	161	0	9.35	-	<0.01	1.81E-04	4.32E-02
complement activation	176	0	10.22	-	<0.01	8.22E-05	2.34E-02

Table A4: All GO-terms from D-network. The number of related genes in the reference list (ref), number of relevant genes from the network (genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process complete	#ref	#genes	exp	+/-	FE	raw P-value	FDR
valyl-tRNA aminoacylation (GO:0006438)	6	6	.22	+	26.87	1.61E-06	1.60E-03
maintenance of location in cell (GO:0051651)	108	14	4.02	+	3.48	1.24E-04	4.49E-02
response to ionizing radiation (GO:0010212)	145	17	5.40	+	3.15	7.59E-05	3.28E-02
regulation of regulated secretory pathway (GO:1903305)	164	18	6.10	+	2.95	1.02E-04	3.88E-02
maintenance of location (GO:0051235)	177	19	6.59	+	2.88	8.67E-05	3.38E-02
cell cycle process (GO:0022402)	1050	71	39.08	+	1.82	3.15E-06	2.65E-03
cellular protein localization (GO:0034613)	1643	107	61.15	+	1.75	4.33E-08	7.68E-05
cellular macromolecule localization (GO:0070727)	1651	107	61.44	+	1.74	6.57E-08	1.05E-04
protein localization (GO:0008104)	2208	137	82.17	+	1.67	6.51E-09	2.60E-05
cell cycle (GO:0007049)	1371	83	51.02	+	1.63	2.79E-05	1.71E-02
regulation of cell cycle (GO:0051726)	1240	75	46.15	+	1.63	6.44E-05	3.21E-02
cellular localization (GO:0051641)	2999	177	111.61	+	1.59	8.50E-10	1.36E-05
macromolecule localization (GO:0033036)	2549	150	94.86	+	1.58	3.54E-08	8.06E-05
positive regulation of catalytic activity (GO:0043085)	1477	86	54.97	+	1.56	7.07E-05	3.22E-02
peptide transport (GO:0015833)	1566	91	58.28	+	1.56	4.57E-05	2.51E-02
establishment of protein localization (GO:0045184)	1620	94	60.29	+	1.56	3.36E-05	1.99E-02
protein transport (GO:0015031)	1535	89	57.13	+	1.56	6.98E-05	3.28E-02
cellular component assembly (GO:0022607)	2331	135	86.75	+	1.56	4.82E-07	5.92E-04
amide transport (GO:0042886)	1600	92	59.55	+	1.55	5.55E-05	2.86E-02
positive regulation of molecular function (GO:0044093)	1832	102	68.18	+	1.50	8.61E-05	3.44E-02
cellular component biogenesis (GO:0044085)	2590	144	96.39	+	1.49	1.55E-06	1.65E-03
nitrogen compound transport (GO:0071705)	1879	104	69.93	+	1.49	8.01E-05	3.37E-02
organelle organization (GO:0006996)	3481	191	129.55	+	1.47	4.23E-08	8.43E-05
establishment of localization in cell (GO:0051649)	2387	128	88.84	+	1.44	3.82E-05	2.18E-02
regulation of molecular function (GO:0065009)	3073	161	114.37	+	1.41	9.77E-06	6.78E-03
cellular component organization (GO:0016043)	5536	280	206.03	+	1.36	1.13E-08	3.60E-05

cellular component organization or biogenesis (GO:0071840)	5754	287	214.14	+	1.34	2.68E-08	7.14E-05
localization (GO:0051179)	5802	284	215.93	+	1.32	2.23E-07	3.23E-04
establishment of localization (GO:0051234)	4680	224	174.17	+	1.29	4.72E-05	2.51E-02
transport (GO:0006810)	4550	217	169.33	+	1.28	8.39E-05	3.43E-02
regulation of primary metabolic process (GO:0080090)	6077	288	226.16	+	1.27	3.09E-06	2.74E-03
regulation of nitrogen compound metabolic process (GO:0051171)	5891	278	219.24	+	1.27	8.05E-06	5.84E-03
regulation of cellular metabolic process (GO:0031323)	6284	295	233.87	+	1.26	4.65E-06	3.54E-03
cellular macromolecule metabolic process (GO:0044260)	5146	240	191.52	+	1.25	1.16E-04	4.32E-02
regulation of macromolecule metabolic process (GO:0060255)	6215	288	231.30	+	1.25	1.93E-05	1.23E-02
regulation of metabolic process (GO:0019222)	6772	310	252.03	+	1.23	1.78E-05	1.18E-02
regulation of cellular process (GO:0050794)	11288	487	420.10	+	1.16	2.09E-06	1.96E-03
biological regulation (GO:0065007)	12525	535	466.14	+	1.15	5.20E-07	5.93E-04
regulation of biological process (GO:0050789)	11825	504	440.08	+	1.15	4.54E-06	3.62E-03
cellular process (GO:0009987)	15458	637	575.29	+	1.11	2.94E-07	3.91E-04
biological_process (GO:0008150)	17810	717	662.82	+	1.08	5.69E-09	4.54E-05
Unclassified (UNCLASSIFIED)	3041	59	113.18	-	.52	5.69E-09	3.03E-05
detection of chemical stimulus (GO:0009593)	519	4	19.32	-	.21	6.53E-05	3.16E-02
sensory perception of smell (GO:0007608)	468	3	17.42	-	.17	7.45E-05	3.30E-02

Table A5: GO-terms for module 1 detected using LPA. All over-represented biological processes associated with module 1 detected using LPA. Sorted by fold enrichment. The number of related genes in the reference list (ref), number of relevant genes from the network (genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process	ref	gene	expected	+/-	FE	raw P-value)	FDR
anaphase	156	14	2.94	+	4.76	3.57E-06	5.70E-02
mitotic anaphase	156	14	2.94	+	4.76	3.57E-06	2.85E-02
M phase	170	14	3.20	+	4.37	8.93E-06	4.75E-02
mitotic M phase	170	14	3.20	+	4.37	8.93E-06	3.56E-02
biological phase	269	17	5.07	+	3.35	2.64E-05	6.03E-02
mitotic cell cycle phase	269	17	5.07	+	3.35	2.64E-05	5.28E-02
cell cycle phase	269	17	5.07	+	3.35	2.64E-05	4.69E-02
cell cycle	1371	49	25.84	+	1.90	2.36E-05	6.28E-02
cellular metabolic process	7742	187	145.92	+	1.28	3.02E-05	4.83E-02
metabolic process	8492	204	160.06	+	1.27	1.01E-05	3.22E-02

Table A6: All GO-terms from Antiquewhite4 module. The number of relevant genes from the network (genes), the expected number genes found,over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process complete	#genes	exp	+/-	FE	raw P-value	FDR
fibroblast apoptotic process (GO:0044346)	3	.13	+	23.51	1.23E-03	2.99E-02
positive regulation of cellular response to macrophage colony-stimulating factor stimulus (GO:1903974)	3	.13	+	23.51	1.23E-03	2.99E-02
positive regulation of response to macrophage colony-stimulating factor (GO:1903971)	3	.13	+	23.51	1.23E-03	2.98E-02
toll-like receptor 2 signaling pathway (GO:0034134)	4	.21	+	18.81	2.94E-04	9.10E-03
germinal center formation (GO:0002467)	4	.34	+	11.75	1.05E-03	2.61E-02
establishment of endothelial intestinal barrier (GO:0090557)	5	.47	+	10.69	3.35E-04	1.02E-02
type 2 immune response (GO:0042092)	5	.47	+	10.69	3.35E-04	1.02E-02
regulation of germinal center formation (GO:0002634)	4	.38	+	10.45	1.47E-03	3.46E-02
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-independent (GO:0002480)	4	.38	+	10.45	1.47E-03	3.45E-02
regulation of tolerance induction (GO:0002643)	7	.72	+	9.68	3.46E-05	1.43E-03
regulation of interleukin-12 biosynthetic process (GO:0045075)	4	.43	+	9.40	1.99E-03	4.47E-02

positive regulation of prostaglandin secretion (GO:0032308)	4	.43	+	9.40	1.99E-03	4.46E-02
tolerance induction (GO:0002507)	5	.55	+	9.04	6.15E-04	1.69E-02
nucleoside triphosphate catabolic process (GO:0009143)	5	.55	+	9.04	6.15E-04	1.69E-02
positive regulation of leukocyte apoptotic process (GO:2000108)	11	1.28	+	8.62	5.04E-07	3.37E-05
TRIF-dependent toll-like receptor signaling pathway (GO:0035666)	10	1.23	+	8.11	2.63E-06	1.51E-04
relaxation of muscle (GO:0090075)	7	.89	+	7.84	1.03E-04	3.66E-03
regulation of B cell apoptotic process (GO:0002902)	6	.77	+	7.84	3.25E-04	9.89E-03
positive regulation of nitric-oxide synthase biosynthetic process (GO:0051770)	5	.64	+	7.84	1.04E-03	2.60E-02
positive regulation of T cell apoptotic process (GO:0070234)	5	.64	+	7.84	1.04E-03	2.60E-02
regulation of nitric-oxide synthase biosynthetic process (GO:0051769)	6	.81	+	7.42	4.13E-04	1.21E-02
positive regulation of lymphocyte apoptotic process (GO:0070230)	6	.81	+	7.42	4.13E-04	1.21E-02
lipopolysaccharide-mediated signaling pathway (GO:0031663)	11	1.49	+	7.39	1.77E-06	1.05E-04
MyD88-independent toll-like receptor signaling pathway (GO:0002756)	10	1.36	+	7.35	5.45E-06	2.82E-04
toll-like receptor 9 signaling pathway (GO:0034162)	5	.72	+	6.91	1.65E-03	3.83E-02
regulation of necroptotic process (GO:0060544)	5	.72	+	6.91	1.65E-03	3.83E-02
response to angiotensin (GO:1990776)	8	1.19	+	6.72	8.14E-05	2.98E-03
defense response to protozoan (GO:0042832)	7	1.06	+	6.58	2.53E-04	7.98E-03
positive regulation of oxidative stress-induced cell death (GO:1903209)	5	.77	+	6.53	2.04E-03	4.56E-02
positive regulation of tumor necrosis factor secretion (GO:1904469)	6	.94	+	6.41	7.91E-04	2.08E-02
leukocyte apoptotic process (GO:0071887)	7	1.11	+	6.33	3.10E-04	9.55E-03
response to protozoan (GO:0001562)	7	1.11	+	6.33	3.10E-04	9.53E-03
regulation of lymphocyte chemotaxis (GO:1901623)	7	1.11	+	6.33	3.10E-04	9.51E-03
positive regulation of p38MAPK cascade (GO:1900745)	7	1.11	+	6.33	3.10E-04	9.50E-03
positive regulation of acute inflammatory response (GO:0002675)	8	1.28	+	6.27	1.22E-04	4.25E-03
monocyte chemotaxis (GO:0002548)	11	1.79	+	6.16	7.77E-06	3.86E-04
positive regulation of macrophage migration (GO:1905523)	6	.98	+	6.13	9.63E-04	2.45E-02
positive regulation of protein processing (GO:0010954)	6	.98	+	6.13	9.63E-04	2.44E-02
cellular response to angiotensin (GO:1904385)	6	.98	+	6.13	9.63E-04	2.44E-02
I-kappaB kinase/NF-kappaB signaling (GO:0007249)	19	3.11	+	6.12	4.55E-09	4.59E-07

mononuclear cell migration (GO:0071674)	12	2.04	+	5.88	4.54E-06	2.43E-04
positive regulation of vascular endothelial growth factor production (GO:0010575)	7	1.19	+	5.88	4.54E-04	1.31E-02
positive regulation of chemokine production (GO:0032722)	15	2.59	+	5.78	3.56E-07	2.51E-05
positive regulation of macrophage activation (GO:0043032)	6	1.06	+	5.64	1.39E-03	3.32E-02
positive regulation of protein maturation (GO:1903319)	6	1.06	+	5.64	1.39E-03	3.31E-02
type I interferon signaling pathway (GO:0060337)	16	2.85	+	5.61	2.05E-07	1.54E-05
cellular response to type I interferon (GO:0071357)	16	2.85	+	5.61	2.05E-07	1.53E-05
positive regulation of interleukin-6 secretion (GO:2000778)	9	1.62	+	5.57	1.02E-04	3.65E-03
positive regulation of lymphocyte migration (GO:2000403)	9	1.62	+	5.57	1.02E-04	3.64E-03
response to type I interferon (GO:0034340)	17	3.06	+	5.55	9.90E-08	7.86E-06
T cell migration (GO:0072678)	7	1.28	+	5.49	6.48E-04	1.77E-02
positive regulation of response to cytokine stimulus (GO:0060760)	13	2.42	+	5.36	4.37E-06	2.35E-04
positive regulation of T cell migration (GO:2000406)	7	1.32	+	5.31	7.67E-04	2.04E-02
negative regulation of insulin receptor signaling pathway (GO:0046627)	7	1.32	+	5.31	7.67E-04	2.03E-02
cellular response to lipopolysaccharide (GO:0071222)	43	8.13	+	5.29	8.21E-17	3.05E-14
lymphocyte chemotaxis (GO:0048247)	11	2.08	+	5.28	2.69E-05	1.17E-03
regulation of tumor necrosis factor secretion (GO:1904467)	7	1.36	+	5.14	9.02E-04	2.33E-02
regulation of vascular endothelial growth factor production (GO:0010574)	7	1.36	+	5.14	9.02E-04	2.33E-02
regulation of T cell cytokine production (GO:0002724)	7	1.36	+	5.14	9.02E-04	2.32E-02
inflammatory response to antigenic stimulus (GO:0002437)	8	1.57	+	5.08	4.19E-04	1.22E-02
regulation of interferon-beta production (GO:0032648)	11	2.17	+	5.07	3.71E-05	1.51E-03
regulation of lymphocyte apoptotic process (GO:0070228)	12	2.38	+	5.04	1.75E-05	7.90E-04
cellular response to molecule of bacterial origin (GO:0071219)	43	8.55	+	5.03	4.03E-16	1.24E-13
positive regulation of interferon-beta production (GO:0032728)	7	1.40	+	4.99	1.06E-03	2.62E-02
negative regulation of cellular response to insulin stimulus (GO:1900077)	7	1.40	+	4.99	1.06E-03	2.62E-02
T cell mediated immunity (GO:0002456)	8	1.62	+	4.95	4.90E-04	1.39E-02

regulation of p38MAPK cascade (GO:1900744)	8	1.62	+	4.95	4.90E-04	1.39E-02
negative regulation of viral genome replication (GO:0045071)	13	2.64	+	4.93	9.67E-06	4.63E-04
positive regulation of nitric oxide biosynthetic process (GO:0045429)	9	1.83	+	4.92	2.28E-04	7.32E-03
regulation of leukocyte apoptotic process (GO:2000106)	18	3.70	+	4.86	2.34E-07	1.73E-05
pattern recognition receptor signaling pathway (GO:0002221)	25	5.15	+	4.86	1.13E-09	1.27E-07
establishment of endothelial barrier (GO:0061028)	7	1.45	+	4.84	1.23E-03	2.98E-02
positive regulation of nitric oxide metabolic process (GO:1904407)	9	1.87	+	4.81	2.65E-04	8.30E-03
cellular response to biotic stimulus (GO:0071216)	46	9.57	+	4.81	1.67E-16	5.67E-14
response to interferon-gamma (GO:0034341)	38	7.91	+	4.80	7.45E-14	1.42E-11
toll-like receptor signaling pathway (GO:0002224)	20	4.21	+	4.75	7.12E-08	5.80E-06
cellular response to interferon-gamma (GO:0071346)	33	6.98	+	4.73	4.70E-12	7.22E-10
positive regulation of cytokine-mediated signaling pathway (GO:0001961)	10	2.13	+	4.70	1.44E-04	4.86E-03
regulation of endothelial cell apoptotic process (GO:2000351)	10	2.13	+	4.70	1.44E-04	4.85E-03
interferon-gamma-mediated signaling pathway (GO:0060333)	14	3.06	+	4.57	9.44E-06	4.58E-04
regulation of T cell apoptotic process (GO:0070232)	7	1.53	+	4.57	1.64E-03	3.82E-02
positive regulation of cytokine biosynthetic process (GO:0042108)	14	3.06	+	4.57	9.44E-06	4.57E-04
cytoplasmic pattern recognition receptor signaling pathway (GO:0002753)	7	1.53	+	4.57	1.64E-03	3.82E-02
cellular response to interleukin-1 (GO:0071347)	34	7.49	+	4.54	5.99E-12	9.11E-10
cellular response to chemokine (GO:1990869)	17	3.79	+	4.49	1.33E-06	8.10E-05
response to chemokine (GO:1990868)	17	3.79	+	4.49	1.33E-06	8.07E-05
regulation of chemokine production (GO:0032642)	16	3.57	+	4.48	2.80E-06	1.59E-04
cellular response to vascular endothelial growth factor stimulus (GO:0035924)	8	1.79	+	4.48	8.73E-04	2.26E-02
positive regulation of B cell proliferation (GO:0030890)	8	1.79	+	4.48	8.73E-04	2.26E-02
autophagosome assembly (GO:0000045)	12	2.68	+	4.48	4.84E-05	1.91E-03
chemokine-mediated signaling pathway (GO:0070098)	15	3.36	+	4.46	5.89E-06	3.02E-04
positive regulation of fatty acid metabolic process (GO:0045923)	7	1.57	+	4.45	1.88E-03	4.27E-02
regulation of interleukin-12 production (GO:0032655)	11	2.51	+	4.38	1.17E-04	4.09E-03
positive regulation of cytokine production involved in immune response (GO:0002720)	11	2.51	+	4.38	1.17E-04	4.08E-03
response to lipopolysaccharide (GO:0032496)	60	13.74	+	4.37	2.23E-19	1.42E-16
regulation of macrophage activation (GO:0043030)	10	2.30	+	4.35	2.49E-04	7.93E-03

positive regulation of reactive oxygen species biosynthetic process (GO:1903428)	10	2.30	+	4.35	2.49E-04	7.91E-03
positive regulation of T cell mediated immunity (GO:0002711)	10	2.30	+	4.35	2.49E-04	7.90E-03
ER-nucleus signaling pathway (GO:0006984)	7	1.62	+	4.33	2.15E-03	4.79E-02
positive regulation of interleukin-10 production (GO:0032733)	7	1.62	+	4.33	2.15E-03	4.78E-02
regulation of response to cytokine stimulus (GO:0060759)	32	7.40	+	4.32	7.64E-11	1.03E-08
regulation of nitric oxide biosynthetic process (GO:0045428)	11	2.55	+	4.31	1.34E-04	4.59E-03
autophagosome organization (GO:1905037)	12	2.81	+	4.27	7.19E-05	2.68E-03
lymphocyte migration (GO:0072676)	13	3.06	+	4.24	3.87E-05	1.58E-03
regulation of cytokine-mediated signaling pathway (GO:0001959)	29	6.93	+	4.18	1.16E-09	1.28E-07
cytokine secretion (GO:0050663)	9	2.17	+	4.15	6.86E-04	1.86E-02
response to molecule of bacterial origin (GO:0002237)	60	14.46	+	4.15	1.95E-18	1.07E-15
negative regulation of viral life cycle (GO:1903901)	15	3.62	+	4.15	1.28E-05	6.00E-04
regulation of lymphocyte migration (GO:2000401)	11	2.68	+	4.10	1.95E-04	6.40E-03
response to interleukin-1 (GO:0070555)	35	8.55	+	4.09	3.77E-11	5.28E-09
regulation of acute inflammatory response (GO:0002673)	8	1.96	+	4.09	1.47E-03	3.45E-02
regulation of CD4-positive, alpha-beta T cell differentiation (GO:0043370)	8	1.96	+	4.09	1.47E-03	3.44E-02
purinergic nucleotide receptor signaling pathway (GO:0035590)	8	1.96	+	4.09	1.47E-03	3.44E-02
positive regulation of tumor necrosis factor superfamily cytokine production (GO:1903557)	16	3.91	+	4.09	7.80E-06	3.87E-04
neutrophil chemotaxis (GO:0030593)	13	3.19	+	4.07	5.63E-05	2.17E-03
negative regulation of viral process (GO:0048525)	18	4.42	+	4.07	2.28E-06	1.33E-04
macrophage activation (GO:0042116)	9	2.21	+	4.07	7.76E-04	2.05E-02
negative regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043154)	15	3.70	+	4.05	1.63E-05	7.44E-04
regulation of interleukin-2 production (GO:0032663)	10	2.47	+	4.05	4.13E-04	1.21E-02
regulation of extrinsic apoptotic signaling pathway via death domain receptors (GO:1902041)	10	2.47	+	4.05	4.13E-04	1.21E-02
negative regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043124)	8	2.00	+	4.00	1.65E-03	3.83E-02
cellular response to leukemia inhibitory factor (GO:1990830)	17	4.25	+	4.00	5.36E-06	2.79E-04
response to leukemia inhibitory factor (GO:1990823)	17	4.25	+	4.00	5.36E-06	2.78E-04

positive regulation of reactive oxygen species metabolic process (GO:2000379)	17	4.30	+	3.96	6.02E-06	3.07E-04
negative regulation of fat cell differentiation (GO:0045599)	8	2.04	+	3.92	1.86E-03	4.24E-02
positive regulation of tumor necrosis factor production (GO:0032760)	15	3.83	+	3.92	2.32E-05	1.01E-03
positive regulation of interleukin-8 production (GO:0032757)	9	2.34	+	3.85	1.11E-03	2.72E-02
regulation of interleukin-10 production (GO:0032653)	9	2.34	+	3.85	1.11E-03	2.72E-02
negative regulation of intrinsic apoptotic signaling pathway (GO:2001243)	16	4.17	+	3.84	1.57E-05	7.22E-04
regulation of cytokine biosynthetic process (GO:0042035)	18	4.72	+	3.81	5.13E-06	2.70E-04
inflammatory response (GO:0006954)	83	21.78	+	3.81	5.26E-23	8.40E-20
regulation of viral genome replication (GO:0045069)	16	4.21	+	3.80	1.76E-05	7.91E-04
positive regulation of blood vessel endothelial cell migration (GO:0043536)	9	2.38	+	3.78	1.24E-03	2.99E-02
positive regulation of type I interferon production (GO:0032481)	13	3.45	+	3.77	1.13E-04	3.95E-03
DNA damage response, signal transduction by p53 class mediator (GO:0030330)	13	3.45	+	3.77	1.13E-04	3.95E-03
positive regulation of gliogenesis (GO:0014015)	12	3.19	+	3.76	2.10E-04	6.80E-03
positive regulation of interferon-gamma production (GO:0032729)	11	2.94	+	3.75	3.92E-04	1.16E-02
negative regulation of ERK1 and ERK2 cascade (GO:0070373)	11	2.94	+	3.75	3.92E-04	1.16E-02
positive regulation of cytokine secretion (GO:0050715)	24	6.42	+	3.74	2.01E-07	1.53E-05
regulation of B cell proliferation (GO:0030888)	10	2.68	+	3.73	7.35E-04	1.98E-02
granulocyte chemotaxis (GO:0071621)	13	3.49	+	3.73	1.26E-04	4.35E-03
regulation of T cell mediated immunity (GO:0002709)	12	3.23	+	3.71	2.34E-04	7.48E-03
negative regulation of cysteine-type endopeptidase activity (GO:2000117)	15	4.04	+	3.71	4.06E-05	1.65E-03
regulation of adaptive immune response (GO:0002819)	27	7.27	+	3.71	3.98E-08	3.42E-06
regulation of cytokine production involved in immune response (GO:0002718)	14	3.79	+	3.70	7.54E-05	2.79E-03
positive regulation of fat cell differentiation (GO:0045600)	11	2.98	+	3.69	4.38E-04	1.27E-02
neutrophil migration (GO:1990266)	13	3.53	+	3.68	1.40E-04	4.76E-03
positive regulation of inflammatory response (GO:0050729)	22	6.00	+	3.67	8.47E-07	5.38E-05
myeloid leukocyte migration (GO:0097529)	19	5.19	+	3.66	4.78E-06	2.55E-04
positive regulation of T cell differentiation (GO:0045582)	14	3.83	+	3.66	8.39E-05	3.06E-03

regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (GO:0002822)	24	6.64	+	3.62	3.42E-07	2.41E-05
demethylation (GO:0070988)	10	2.77	+	3.62	9.12E-04	2.34E-02
regulation of type I interferon production (GO:0032479)	20	5.53	+	3.62	3.20E-06	1.78E-04
cellular response to hydrogen peroxide (GO:0070301)	11	3.06	+	3.59	5.41E-04	1.51E-02
positive regulation of vasculature development (GO:1904018)	29	8.08	+	3.59	2.45E-08	2.17E-06
regulation of interferon-gamma production (GO:0032649)	16	4.47	+	3.58	3.35E-05	1.39E-03
positive regulation of leukocyte chemotaxis (GO:0002690)	14	3.91	+	3.58	1.03E-04	3.68E-03
positive regulation of leukocyte migration (GO:0002687)	21	5.91	+	3.55	2.36E-06	1.37E-04
regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043122)	37	10.42	+	3.55	3.83E-10	4.66E-08
regulation of reactive oxygen species biosynthetic process (GO:1903426)	14	3.96	+	3.54	1.15E-04	4.00E-03
regulation of CD4-positive, alpha-beta T cell activation (GO:2000514)	9	2.55	+	3.53	1.90E-03	4.30E-02
regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0000079)	15	4.25	+	3.53	6.85E-05	2.57E-03
regulation of epithelial cell apoptotic process (GO:1904035)	12	3.40	+	3.53	3.57E-04	1.07E-02
negative regulation of innate immune response (GO:0045824)	9	2.55	+	3.53	1.90E-03	4.29E-02
positive regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043123)	28	7.95	+	3.52	6.11E-08	5.06E-06
regulation of oxidative stress-induced cell death (GO:1903201)	10	2.85	+	3.51	1.12E-03	2.75E-02
regulation of fat cell differentiation (GO:0045598)	20	5.70	+	3.51	4.81E-06	2.56E-04
positive regulation of lymphocyte differentiation (GO:0045621)	15	4.30	+	3.49	7.58E-05	2.79E-03
stress-activated MAPK cascade (GO:0051403)	16	4.59	+	3.48	4.54E-05	1.81E-03
positive regulation of leukocyte proliferation (GO:0070665)	21	6.04	+	3.48	3.19E-06	1.79E-04
positive regulation of NF-kappaB transcription factor activity (GO:0051092)	23	6.64	+	3.47	1.15E-06	7.17E-05
regulation of alpha-beta T cell activation (GO:0046634)	14	4.04	+	3.46	1.40E-04	4.75E-03
positive regulation of leukocyte cell-cell adhesion (GO:1903039)	34	9.83	+	3.46	3.54E-09	3.64E-07
regulation of cytokine secretion (GO:0050707)	32	9.27	+	3.45	1.08E-08	1.03E-06

positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains (GO:0002824)	16	4.64	+	3.45	5.01E-05	1.96E-03
regulation of blood coagulation (GO:0030193)	12	3.49	+	3.44	4.36E-04	1.27E-02
interleukin-1-mediated signaling pathway (GO:0070498)	14	4.08	+	3.43	1.55E-04	5.19E-03
regulation of interleukin-6 production (GO:0032675)	21	6.13	+	3.43	3.89E-06	2.12E-04
negative regulation of leukocyte differentiation (GO:1902106)	15	4.38	+	3.42	9.24E-05	3.37E-03
regulation of toll-like receptor signaling pathway (GO:0034121)	10	2.94	+	3.41	1.37E-03	3.27E-02
regulation of hemostasis (GO:1900046)	12	3.53	+	3.40	4.81E-04	1.37E-02
regulation of cyclin-dependent protein kinase activity (GO:1904029)	15	4.42	+	3.39	1.02E-04	3.64E-03
response to hydrogen peroxide (GO:0042542)	17	5.06	+	3.36	4.01E-05	1.63E-03
regulation of viral life cycle (GO:1903900)	21	6.25	+	3.36	5.18E-06	2.71E-04
negative regulation of extrinsic apoptotic signaling pathway (GO:2001237)	15	4.47	+	3.36	1.12E-04	3.95E-03
signal transduction in response to DNA damage (GO:0042770)	15	4.47	+	3.36	1.12E-04	3.94E-03
leukocyte chemotaxis (GO:0030595)	20	6.00	+	3.33	9.49E-06	4.57E-04
extrinsic apoptotic signaling pathway (GO:0097191)	14	4.21	+	3.32	2.06E-04	6.71E-03
positive regulation of angiogenesis (GO:0045766)	24	7.23	+	3.32	1.36E-06	8.19E-05
regulation of tumor necrosis factor superfamily cytokine production (GO:1903555)	22	6.64	+	3.32	3.75E-06	2.05E-04
leukocyte homeostasis (GO:0001776)	10	3.02	+	3.31	1.66E-03	3.84E-02
signal transduction by p53 class mediator (GO:0072331)	17	5.15	+	3.30	4.84E-05	1.91E-03
positive regulation of adaptive immune response (GO:0002821)	16	4.85	+	3.30	8.08E-05	2.97E-03
regulation of blood vessel endothelial cell migration (GO:0043535)	13	3.96	+	3.29	3.80E-04	1.13E-02
granulocyte migration (GO:0097530)	13	3.96	+	3.29	3.80E-04	1.13E-02
positive regulation of mononuclear cell proliferation (GO:0032946)	19	5.79	+	3.28	1.90E-05	8.47E-04
G1/S transition of mitotic cell cycle (GO:0000082)	17	5.19	+	3.28	5.30E-05	2.05E-03
acute inflammatory response (GO:0002526)	10	3.06	+	3.26	1.82E-03	4.16E-02
positive regulation of cytokine production (GO:0001819)	64	19.65	+	3.26	5.14E-15	1.26E-12
cellular response to antibiotic (GO:0071236)	17	5.23	+	3.25	5.81E-05	2.23E-03
endothelial cell differentiation (GO:0045446)	11	3.40	+	3.23	1.18E-03	2.88E-02
regulation of leukocyte proliferation (GO:0070663)	32	9.91	+	3.23	4.39E-08	3.75E-06
cell cycle G1/S phase transition (GO:0044843)	17	5.27	+	3.22	6.36E-05	2.42E-03
regulation of gliogenesis (GO:0014013)	17	5.27	+	3.22	6.36E-05	2.42E-03
positive regulation of T cell activation (GO:0050870)	29	9.02	+	3.22	2.02E-07	1.52E-05
negative regulation of T cell activation (GO:0050868)	16	4.98	+	3.21	1.06E-04	3.76E-03

cytokine production (GO:0001816)	22	6.85	+	3.21	5.92E-06	3.03E-04
positive regulation of lipid localization (GO:1905954)	12	3.74	+	3.21	7.68E-04	2.03E-02
negative regulation of leukocyte proliferation (GO:0070664)	12	3.74	+	3.21	7.68E-04	2.03E-02
regulation of coagulation (GO:0050818)	12	3.74	+	3.21	7.68E-04	2.03E-02
cellular response to mechanical stimulus (GO:0071260)	11	3.45	+	3.19	1.29E-03	3.11E-02
regulation of cellular response to oxidative stress (GO:1900407)	11	3.45	+	3.19	1.29E-03	3.11E-02
regulation of leukocyte cell-cell adhesion (GO:1903037)	43	13.49	+	3.19	3.02E-10	3.77E-08
positive regulation of interleukin-6 production (GO:0032755)	13	4.08	+	3.18	4.99E-04	1.41E-02
regulation of leukocyte migration (GO:0002685)	28	8.81	+	3.18	4.00E-07	2.75E-05
regulation of intrinsic apoptotic signaling pathway (GO:2001242)	22	6.93	+	3.17	7.07E-06	3.56E-04
cellular response to alcohol (GO:0097306)	12	3.79	+	3.17	8.40E-04	2.19E-02
regulation of leukocyte chemotaxis (GO:0002688)	16	5.06	+	3.16	1.27E-04	4.38E-03
negative regulation of apoptotic signaling pathway (GO:2001234)	31	9.83	+	3.15	1.13E-07	8.87E-06
positive regulation of animal organ morphogenesis (GO:0110110)	12	3.83	+	3.13	9.17E-04	2.35E-02
positive regulation of lymphocyte proliferation (GO:0050671)	18	5.74	+	3.13	5.43E-05	2.10E-03
regulation of extrinsic apoptotic signaling pathway (GO:2001236)	21	6.72	+	3.12	1.40E-05	6.51E-04
response to tumor necrosis factor (GO:0034612)	36	11.53	+	3.12	1.38E-08	1.30E-06
positive regulation of hemopoiesis (GO:1903708)	26	8.38	+	3.10	1.56E-06	9.31E-05
regulation of blood vessel diameter (GO:0097746)	18	5.83	+	3.09	6.44E-05	2.44E-03
regulation of tube diameter (GO:0035296)	18	5.83	+	3.09	6.44E-05	2.43E-03
regulation of tumor necrosis factor production (GO:0032680)	20	6.51	+	3.07	2.76E-05	1.19E-03
negative regulation of lymphocyte activation (GO:0051250)	20	6.51	+	3.07	2.76E-05	1.19E-03
regulation of mononuclear cell proliferation (GO:0032944)	29	9.44	+	3.07	4.78E-07	3.25E-05
regulation of tube size (GO:0035150)	18	5.87	+	3.07	7.01E-05	2.63E-03
positive regulation of autophagy (GO:0010508)	16	5.23	+	3.06	1.79E-04	5.91E-03
regulation of T cell activation (GO:0050863)	43	14.08	+	3.05	1.00E-09	1.15E-07
positive regulation of leukocyte differentiation (GO:1902107)	20	6.55	+	3.05	3.00E-05	1.28E-03
regulation of NIK/NF-kappaB signaling (GO:1901222)	14	4.59	+	3.05	4.59E-04	1.32E-02
regulation of reactive oxygen species metabolic process (GO:2000377)	24	7.91	+	3.03	5.51E-06	2.84E-04
response to virus (GO:0009615)	37	12.25	+	3.02	1.92E-08	1.74E-06
defense response to virus (GO:0051607)	26	8.64	+	3.01	2.57E-06	1.48E-04

chromatin organization involved in regulation of transcription (GO:0034401)	11	3.66	+	3.01	1.99E-03	4.47E-02
cell chemotaxis (GO:0060326)	26	8.68	+	3.00	2.78E-06	1.59E-04
regulation of lymphocyte proliferation (GO:0050670)	28	9.36	+	2.99	1.19E-06	7.39E-05
positive regulation of cell-cell adhesion (GO:0022409)	35	11.70	+	2.99	5.71E-08	4.75E-06
cytokine-mediated signaling pathway (GO:0019221)	88	29.52	+	2.98	4.37E-18	2.11E-15
response to cytokine (GO:0034097)	142	47.69	+	2.98	5.22E-29	4.17E-25
cellular response to tumor necrosis factor (GO:0071356)	31	10.42	+	2.97	3.64E-07	2.54E-05
regulation of cytokine production (GO:0001817)	92	31.22	+	2.95	1.40E-18	8.01E-16
negative regulation of leukocyte cell-cell adhesion (GO:1903038)	16	5.45	+	2.94	2.69E-04	8.41E-03
cellular response to hypoxia (GO:0071456)	24	8.17	+	2.94	8.94E-06	4.36E-04
regulation of heart rate (GO:0002027)	12	4.08	+	2.94	1.52E-03	3.54E-02
modulation of process of other organism involved in symbiotic interaction (GO:0051817)	14	4.76	+	2.94	6.37E-04	1.74E-02
positive regulation of production of molecular mediator of immune response (GO:0002702)	13	4.42	+	2.94	9.82E-04	2.48E-02
cellular response to lipid (GO:0071396)	66	22.50	+	2.93	1.62E-13	2.93E-11
regulation of viral process (GO:0050792)	26	8.89	+	2.92	4.14E-06	2.23E-04
vascular process in circulatory system (GO:0003018)	21	7.19	+	2.92	3.45E-05	1.43E-03
cellular response to decreased oxygen levels (GO:0036294)	25	8.59	+	2.91	6.84E-06	3.47E-04
cellular response to cytokine stimulus (GO:0071345)	127	43.86	+	2.90	6.68E-25	1.52E-21
response to reactive oxygen species (GO:0000302)	24	8.34	+	2.88	1.22E-05	5.74E-04
regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043281)	27	9.40	+	2.87	3.70E-06	2.03E-04
positive regulation of programmed cell death (GO:0043068)	80	28.03	+	2.85	1.46E-15	4.00E-13
activation of GTPase activity (GO:0090630)	12	4.21	+	2.85	1.92E-03	4.33E-02
negative regulation of leukocyte activation (GO:0002695)	22	7.74	+	2.84	3.32E-05	1.39E-03
regulation of T cell differentiation (GO:0045580)	18	6.34	+	2.84	1.69E-04	5.63E-03
positive regulation of chemotaxis (GO:0050921)	17	6.00	+	2.83	2.59E-04	8.12E-03
intrinsic apoptotic signaling pathway (GO:0097193)	18	6.38	+	2.82	1.82E-04	6.00E-03
tissue remodeling (GO:0048771)	12	4.25	+	2.82	2.07E-03	4.62E-02
regulation of leukocyte differentiation (GO:1902105)	34	12.08	+	2.81	3.24E-07	2.33E-05
positive regulation of apoptotic process (GO:0043065)	78	27.78	+	2.81	7.40E-15	1.79E-12
negative regulation of hemopoiesis (GO:1903707)	17	6.08	+	2.79	3.00E-04	9.27E-03
cellular response to oxygen levels (GO:0071453)	26	9.32	+	2.79	8.84E-06	4.33E-04
positive regulation of DNA-binding transcription factor activity (GO:0051091)	32	11.49	+	2.79	8.74E-07	5.52E-05

regulation of leukocyte mediated immunity (GO:0002703)	25	8.98	+	2.79	1.35E-05	6.31E-04
negative regulation of cell activation (GO:0050866)	24	8.64	+	2.78	2.06E-05	9.05E-04
cellular response to toxic substance (GO:0097237)	26	9.36	+	2.78	9.51E-06	4.57E-04
stress-activated protein kinase signaling cascade (GO:0031098)	17	6.13	+	2.78	3.23E-04	9.88E-03
apoptotic signaling pathway (GO:0097190)	34	12.25	+	2.78	4.35E-07	2.98E-05
negative regulation of protein serine/threonine kinase activity (GO:0071901)	16	5.79	+	2.77	4.96E-04	1.40E-02
regulation of smooth muscle cell proliferation (GO:0048660)	16	5.79	+	2.77	4.96E-04	1.40E-02
positive regulation of small molecule metabolic process (GO:0062013)	17	6.17	+	2.76	3.48E-04	1.04E-02
regulation of muscle contraction (GO:0006937)	19	6.93	+	2.74	1.71E-04	5.69E-03
regulation of receptor signaling pathway via STAT (GO:1904892)	16	5.87	+	2.73	5.73E-04	1.59E-02
regulation of vasculature development (GO:1901342)	39	14.34	+	2.72	1.00E-07	7.89E-06
regulation of symbiosis, encompassing mutualism through parasitism (GO:0043903)	26	9.57	+	2.72	1.36E-05	6.35E-04
positive regulation of cell death (GO:0010942)	82	30.25	+	2.71	1.09E-14	2.48E-12
regulation of innate immune response (GO:0045088)	35	12.93	+	2.71	5.04E-07	3.35E-05
negative regulation of cytokine production (GO:0001818)	32	11.83	+	2.71	1.55E-06	9.29E-05
positive regulation of immune effector process (GO:0002699)	26	9.61	+	2.70	1.46E-05	6.75E-04
regulation of cell-cell adhesion (GO:0022407)	49	18.12	+	2.70	2.77E-09	2.88E-07
epithelial cell development (GO:0002064)	23	8.51	+	2.70	4.49E-05	1.80E-03
negative regulation of cell-cell adhesion (GO:0022408)	21	7.78	+	2.70	9.75E-05	3.52E-03
regulation of myeloid leukocyte differentiation (GO:0002761)	14	5.19	+	2.70	1.35E-03	3.23E-02
response to mechanical stimulus (GO:0009612)	25	9.27	+	2.70	2.22E-05	9.75E-04
response to decreased oxygen levels (GO:0036293)	41	15.23	+	2.69	6.13E-08	5.04E-06
cellular response to chemical stress (GO:0062197)	33	12.29	+	2.68	1.25E-06	7.72E-05
regulation of protein modification by small protein conjugation or removal (GO:1903320)	27	10.08	+	2.68	1.18E-05	5.59E-04
regulation of T cell proliferation (GO:0042129)	19	7.10	+	2.67	2.27E-04	7.30E-03
regulation of production of molecular mediator of immune response (GO:0002700)	17	6.38	+	2.66	4.96E-04	1.41E-02
regulation of apoptotic signaling pathway (GO:2001233)	46	17.27	+	2.66	1.84E-08	1.67E-06
response to hypoxia (GO:0001666)	39	14.68	+	2.66	2.72E-07	1.98E-05
regulation of lymphocyte differentiation (GO:0045619)	20	7.53	+	2.66	1.71E-04	5.69E-03

cellular response to reactive oxygen species (GO:0034614)	15	5.66	+	2.65	1.09E-03	2.68E-02
positive regulation of protein modification by small protein conjugation or removal (GO:1903322)	16	6.04	+	2.65	7.59E-04	2.02E-02
response to oxygen levels (GO:0070482)	43	16.25	+	2.65	6.28E-08	5.14E-06
regulation of DNA-binding transcription factor activity (GO:0051090)	49	18.55	+	2.64	6.82E-09	6.68E-07
negative regulation of transmembrane transport (GO:0034763)	14	5.32	+	2.63	1.67E-03	3.85E-02
regulation of receptor signaling pathway via JAK-STAT (GO:0046425)	15	5.70	+	2.63	1.16E-03	2.84E-02
regulation of angiogenesis (GO:0045765)	34	12.93	+	2.63	2.01E-06	1.19E-04
positive regulation of cell activation (GO:0050867)	48	18.29	+	2.62	1.14E-08	1.08E-06
regulation of wound healing (GO:0061041)	17	6.51	+	2.61	6.08E-04	1.67E-02
positive regulation of leukocyte activation (GO:0002696)	46	17.70	+	2.60	2.88E-08	2.53E-06
negative regulation of cell population proliferation (GO:0008285)	77	29.69	+	2.59	6.51E-13	1.13E-10
positive regulation of MAPK cascade (GO:0043410)	60	23.14	+	2.59	2.08E-10	2.64E-08
regulation of cysteine-type endopeptidase activity (GO:2000116)	27	10.42	+	2.59	3.14E-05	1.32E-03
regulation of lymphocyte mediated immunity (GO:0002706)	17	6.59	+	2.58	6.95E-04	1.87E-02
cell cycle arrest (GO:0007050)	16	6.21	+	2.58	9.93E-04	2.50E-02
positive regulation of protein serine/threonine kinase activity (GO:0071902)	38	14.76	+	2.57	6.09E-07	4.00E-05
peptide secretion (GO:0002790)	20	7.78	+	2.57	2.55E-04	8.04E-03
positive regulation of cell migration (GO:0030335)	57	22.21	+	2.57	8.31E-10	9.68E-08
positive regulation of stress-activated protein kinase signaling cascade (GO:0070304)	18	7.02	+	2.56	5.19E-04	1.46E-02
regulation of G1/S transition of mitotic cell cycle (GO:2000045)	17	6.64	+	2.56	7.42E-04	1.99E-02
regulation of leukocyte activation (GO:0002694)	67	26.16	+	2.56	3.63E-11	5.13E-09
defense response (GO:0006952)	149	58.24	+	2.56	4.77E-24	8.46E-21
positive regulation of protein secretion (GO:0050714)	31	12.12	+	2.56	8.17E-06	4.02E-04
modulation of process of other organism (GO:0035821)	14	5.49	+	2.55	2.18E-03	4.83E-02
regulation of cell activation (GO:0050865)	71	27.91	+	2.54	1.03E-11	1.50E-09
positive regulation of defense response (GO:0031349)	40	15.74	+	2.54	3.84E-07	2.66E-05
negative regulation of DNA-binding transcription factor activity (GO:0043433)	18	7.10	+	2.53	5.91E-04	1.63E-02
cAMP-mediated signaling (GO:0019933)	16	6.34	+	2.52	1.21E-03	2.94E-02
regulation of protein serine/threonine kinase activity (GO:0071900)	56	22.21	+	2.52	2.70E-09	2.84E-07

response to bacterium (GO:0009617)	76	30.16	+	2.52	2.56E-12	4.13E-10
positive regulation of peptide secretion (GO:0002793)	33	13.10	+	2.52	4.98E-06	2.63E-04
response to lipid (GO:0033993)	93	36.97	+	2.52	9.10E-15	2.11E-12
cellular response to oxidative stress (GO:0034599)	26	10.34	+	2.52	5.78E-05	2.22E-03
negative regulation of cell adhesion (GO:0007162)	30	11.95	+	2.51	1.47E-05	6.78E-04
positive regulation of leukocyte mediated immunity (GO:0002705)	15	6.00	+	2.50	1.84E-03	4.19E-02
regulation of protein ubiquitination (GO:0031396)	22	8.81	+	2.50	2.53E-04	7.99E-03
regulation of ERK1 and ERK2 cascade (GO:0070372)	31	12.42	+	2.50	1.13E-05	5.38E-04
positive regulation of response to external stimulus (GO:0032103)	54	21.65	+	2.49	6.90E-09	6.71E-07
negative regulation of MAPK cascade (GO:0043409)	18	7.23	+	2.49	1.10E-03	2.71E-02
response to other organism (GO:0051707)	142	57.22	+	2.48	6.95E-22	9.25E-19
innate immune response (GO:0045087)	82	33.05	+	2.48	7.68E-13	1.32E-10
response to external biotic stimulus (GO:0043207)	142	57.30	+	2.48	7.63E-22	9.37E-19
regulation of response to biotic stimulus (GO:0002831)	43	17.36	+	2.48	3.60E-07	2.52E-05
positive regulation of cell adhesion (GO:0045785)	45	18.21	+	2.47	1.84E-07	1.41E-05
positive regulation of cellular component movement (GO:0051272)	59	23.95	+	2.46	1.78E-09	1.92E-07
regulation of hemopoiesis (GO:1903706)	49	19.91	+	2.46	5.04E-08	4.28E-06
response to biotic stimulus (GO:0009607)	144	58.66	+	2.45	1.09E-21	1.09E-18
response to oxidative stress (GO:0006979)	41	16.72	+	2.45	8.20E-07	5.25E-05
positive regulation of stress-activated MAPK cascade (GO:0032874)	17	6.93	+	2.45	1.69E-03	3.90E-02
protein secretion (GO:0009306)	17	6.93	+	2.45	1.69E-03	3.90E-02
positive regulation of cell motility (GO:2000147)	57	23.27	+	2.45	4.06E-09	4.13E-07
regulation of lipid biosynthetic process (GO:0046890)	20	8.17	+	2.45	5.90E-04	1.63E-02
regulation of response to wounding (GO:1903034)	19	7.78	+	2.44	8.54E-04	2.22E-02
establishment of protein localization to extracellular region (GO:0035592)	17	6.98	+	2.44	1.74E-03	4.01E-02
regulation of lymphocyte activation (GO:0051249)	55	22.63	+	2.43	1.54E-08	1.42E-06
positive regulation of locomotion (GO:0040017)	58	23.91	+	2.43	5.49E-09	5.44E-07
positive regulation of intracellular signal transduction (GO:1902533)	106	43.99	+	2.41	1.37E-15	3.83E-13
negative regulation of secretion by cell (GO:1903531)	21	8.72	+	2.41	4.79E-04	1.37E-02
positive regulation of MAP kinase activity (GO:0043406)	27	11.23	+	2.40	7.08E-05	2.65E-03
regulation of cell cycle G1/S phase transition (GO:1902806)	18	7.49	+	2.40	1.33E-03	3.18E-02
negative regulation of intracellular signal transduction (GO:1902532)	52	21.65	+	2.40	5.07E-08	4.28E-06
positive regulation of lymphocyte activation (GO:0051251)	38	15.87	+	2.39	3.12E-06	1.75E-04
regulation of myeloid cell differentiation (GO:0045637)	24	10.04	+	2.39	1.96E-04	6.39E-03

apoptotic process (GO:0006915)	92	38.67	+	2.38	2.65E-13	4.75E-11
positive regulation of apoptotic signaling pathway (GO:2001235)	18	7.57	+	2.38	1.43E-03	3.39E-02
regulation of chemotaxis (GO:0050920)	22	9.27	+	2.37	4.00E-04	1.18E-02
myeloid cell differentiation (GO:0030099)	23	9.70	+	2.37	2.94E-04	9.10E-03
regulation of inflammatory response (GO:0050727)	54	22.84	+	2.36	3.50E-08	3.05E-06
response to topologically incorrect protein (GO:0035966)	19	8.04	+	2.36	1.08E-03	2.67E-02
aging (GO:0007568)	28	11.87	+	2.36	1.02E-04	3.64E-03
cellular response to external stimulus (GO:0071496)	37	15.70	+	2.36	5.50E-06	2.84E-04
regulation of MAPK cascade (GO:0043408)	73	31.01	+	2.35	1.33E-10	1.72E-08
defense response to other organism (GO:0098542)	97	41.26	+	2.35	1.22E-13	2.27E-11
positive regulation of ERK1 and ERK2 cascade (GO:0070374)	21	8.93	+	2.35	6.00E-04	1.65E-02
regulation of stress-activated protein kinase signaling cascade (GO:0070302)	23	9.78	+	2.35	3.24E-04	9.88E-03
lymphocyte differentiation (GO:0030098)	24	10.21	+	2.35	2.39E-04	7.60E-03
response to ketone (GO:1901654)	20	8.55	+	2.34	8.59E-04	2.23E-02
myeloid leukocyte activation (GO:0002274)	58	24.80	+	2.34	1.98E-08	1.78E-06
protein localization to extracellular r egion (GO:0071692)	17	7.27	+	2.34	2.23E-03	4.95E-02
regulation of MAP kinase activity (GO:0043405)	34	14.55	+	2.34	1.59E-05	7.28E-04
positive regulation of innate immune response (GO:0045089)	21	9.02	+	2.33	6.59E-04	1.79E-02
positive regulation of secretion by cell (GO:1903532)	42	18.08	+	2.32	2.21E-06	1.30E-04
regulation of muscle system process (GO:0090257)	23	9.91	+	2.32	3.76E-04	1.12E-02
leukocyte migration (GO:0050900)	37	15.95	+	2.32	7.25E-06	3.63E-04
leukocyte differentiation (GO:0002521)	34	14.68	+	2.32	1.81E-05	8.08E-04
positive regulation of protein phosphorylation (GO:0001934)	100	43.22	+	2.31	1.29E-13	2.37E-11
cell activation (GO:0001775)	105	45.43	+	2.31	2.50E-14	5.40E-12
angiogenesis (GO:0001525)	31	13.44	+	2.31	4.80E-05	1.90E-03
adenylate cyclase-modulating G protein-coupled receptor signaling pathway (GO:0007188)	21	9.15	+	2.30	7.61E-04	2.02E-02
intracellular receptor signaling pathway (GO:0030522)	20	8.72	+	2.29	1.03E-03	2.60E-02
regulation of cell adhesion (GO:0030155)	70	30.54	+	2.29	1.04E-09	1.18E-07
regulation of epithelial cell migration (GO:0010632)	22	9.61	+	2.29	9.18E-04	2.35E-02
regulation of stress-activated MAPK cascade (GO:0032872)	22	9.66	+	2.28	9.35E-04	2.38E-02
regulation of immune effector process (GO:0002697)	46	20.29	+	2.27	1.52E-06	9.16E-05
response to toxic substance (GO:0009636)	50	22.12	+	2.26	4.41E-07	3.01E-05

regulation of protein secretion (GO:0050708)	47	20.80	+	2.26	1.14E-06	7.09E-05
negative regulation of secretion (GO:0051048)	22	9.74	+	2.26	9.77E-04	2.47E-02
positive regulation of GTPase activity (GO:0043547)	39	17.31	+	2.25	8.58E-06	4.22E-04
positive regulation of protein modification process (GO:0031401)	119	53.05	+	2.24	3.30E-15	8.64E-13
positive regulation of phosphorylation (GO:0042327)	104	46.50	+	2.24	2.85E-13	5.05E-11
cellular response to extracellular stimulus (GO:0031668)	28	12.59	+	2.22	1.83E-04	6.04E-03
circulatory system process (GO:0003013)	38	17.10	+	2.22	1.42E-05	6.58E-04
regulation of peptide secretion (GO:0002791)	49	22.08	+	2.22	8.35E-07	5.33E-05
blood circulation (GO:0008015)	37	16.68	+	2.22	1.93E-05	8.53E-04
hemostasis (GO:0007599)	28	12.63	+	2.22	1.91E-04	6.29E-03
programmed cell death (GO:0012501)	98	44.24	+	2.22	2.56E-12	4.17E-10
negative regulation of apoptotic process (GO:0043066)	85	38.46	+	2.21	9.76E-11	1.29E-08
cell death (GO:0008219)	101	45.77	+	2.21	1.12E-12	1.90E-10
regulation of GTPase activity (GO:0043087)	45	20.46	+	2.20	3.09E-06	1.74E-04
regulation of defense response (GO:0031347)	80	36.41	+	2.20	3.49E-10	4.32E-08
positive regulation of proteolysis (GO:0045862)	34	15.48	+	2.20	5.25E-05	2.04E-03
regulation of intracellular signal transduction (GO:1902531)	168	76.53	+	2.20	8.73E-21	7.33E-18
leukocyte activation (GO:0045321)	86	39.18	+	2.20	8.43E-11	1.13E-08
negative regulation of programmed cell death (GO:0043069)	86	39.18	+	2.20	8.43E-11	1.12E-08
regulation of cell migration (GO:0030334)	82	37.39	+	2.19	2.95E-10	3.71E-08
negative regulation of immune system process (GO:0002683)	58	26.50	+	2.19	1.77E-07	1.37E-05
positive regulation of secretion (GO:0051047)	42	19.23	+	2.18	8.15E-06	4.03E-04
cellular response to peptide (GO:1901653)	31	14.21	+	2.18	1.35E-04	4.61E-03
cellular response to oxygen-containing compound (GO:1901701)	100	45.94	+	2.18	3.64E-12	5.69E-10
positive regulation of immune system process (GO:0002684)	101	46.41	+	2.18	2.67E-12	4.26E-10
regulation of cellular component movement (GO:0051270)	95	43.69	+	2.17	1.14E-11	1.64E-09
cellular response to organic substance (GO:0071310)	222	102.10	+	2.17	1.43E-27	4.57E-24
negative regulation of immune response (GO:0050777)	32	14.72	+	2.17	1.05E-04	3.72E-03
blood coagulation (GO:0007596)	27	12.42	+	2.17	4.66E-04	1.34E-02
positive regulation of protein kinase activity (GO:0045860)	50	23.01	+	2.17	1.30E-06	7.98E-05
response to antibiotic (GO:0046677)	29	13.40	+	2.16	2.61E-04	8.20E-03
T cell activation (GO:0042110)	22	10.17	+	2.16	1.32E-03	3.16E-02
coagulation (GO:0050817)	27	12.51	+	2.16	4.85E-04	1.38E-02

positive regulation of response to biotic stimulus (GO:0002833)	23	10.68	+	2.15	1.04E-03	2.60E-02
cellular response to peptide hormone stimulus (GO:0071375)	25	11.61	+	2.15	9.18E-04	2.35E-02
regulation of apoptotic process (GO:0042981)	142	66.23	+	2.14	1.25E-16	4.44E-14
regulation of programmed cell death (GO:0043067)	144	67.17	+	2.14	6.66E-17	2.53E-14
positive regulation of phosphorus metabolic process (GO:0010562)	106	49.47	+	2.14	1.50E-12	2.50E-10
positive regulation of phosphate metabolic process (GO:0045937)	106	49.47	+	2.14	1.50E-12	2.47E-10
regulation of binding (GO:0051098)	34	15.91	+	2.14	1.10E-04	3.87E-03
lymphocyte activation (GO:0046649)	35	16.38	+	2.14	8.11E-05	2.98E-03
negative regulation of cell death (GO:0060548)	92	43.09	+	2.13	7.00E-11	9.63E-09
cellular response to nutrient levels (GO:0031669)	24	11.27	+	2.13	1.32E-03	3.16E-02
negative regulation of protein kinase activity (GO:0006469)	21	9.87	+	2.13	2.00E-03	4.48E-02
positive regulation of protein transport (GO:0051222)	42	19.74	+	2.13	1.79E-05	8.03E-04
leukocyte activation involved in immune response (GO:0002366)	56	26.33	+	2.13	5.03E-07	3.38E-05
regulation of protein phosphorylation (GO:0001932)	132	62.15	+	2.12	4.27E-15	1.08E-12
regulation of peptidyl-tyrosine phosphorylation (GO:0050730)	23	10.85	+	2.12	1.83E-03	4.18E-02
cell activation involved in immune response (GO:0002263)	56	26.50	+	2.11	5.79E-07	3.82E-05
regulation of protein kinase activity (GO:0045859)	72	34.12	+	2.11	1.39E-08	1.30E-06
positive regulation of multicellular organismal process (GO:0051240)	160	75.85	+	2.11	3.31E-18	1.70E-15
regulation of immune system process (GO:0002682)	154	73.08	+	2.11	1.58E-17	6.65E-15
positive regulation of cell cycle (GO:0045787)	35	16.63	+	2.10	9.48E-05	3.44E-03
immune response (GO:0006955)	167	79.63	+	2.10	7.70E-19	4.55E-16
regulation of cell motility (GO:2000145)	84	40.07	+	2.10	1.15E-09	1.29E-07
regulation of cell death (GO:0010941)	152	72.57	+	2.09	5.55E-17	2.16E-14
negative regulation of multicellular organismal process (GO:0051241)	109	52.07	+	2.09	2.91E-12	4.60E-10
positive regulation of kinase activity (GO:0033674)	55	26.29	+	2.09	1.32E-06	8.07E-05
positive regulation of transferase activity (GO:0051347)	62	29.69	+	2.09	2.92E-07	2.11E-05
wound healing (GO:0042060)	43	20.63	+	2.08	1.74E-05	7.88E-04
regulation of response to external stimulus (GO:0032101)	106	50.88	+	2.08	8.09E-12	1.22E-09
positive regulation of cellular protein metabolic process (GO:0032270)	144	69.13	+	2.08	5.57E-16	1.65E-13

regulation of locomotion (GO:0040012)	87	41.77	+	2.08	8.69E-10	1.01E-07
positive regulation of hydrolase activity (GO:0051345)	68	32.76	+	2.08	9.09E-08	7.33E-06
positive regulation of neurogenesis (GO:0050769)	43	20.76	+	2.07	1.93E-05	8.54E-04
negative regulation of protein phosphorylation (GO:0001933)	36	17.44	+	2.06	1.35E-04	4.61E-03
positive regulation of establishment of protein localization (GO:1904951)	42	20.42	+	2.06	2.78E-05	1.20E-03
cellular response to chemical stimulus (GO:0070887)	259	125.96	+	2.06	4.50E-29	7.18E-25
positive regulation of catalytic activity (GO:0043085)	129	62.83	+	2.05	8.15E-14	1.53E-11
regulation of phosphorylation (GO:0042325)	142	69.34	+	2.05	4.45E-15	1.11E-12
regulation of lipid metabolic process (GO:0019216)	36	17.61	+	2.04	1.45E-04	4.89E-03
positive regulation of cell differentiation (GO:0045597)	87	42.58	+	2.04	2.05E-09	2.16E-07
immune effector process (GO:0002252)	94	46.07	+	2.04	4.45E-10	5.34E-08
myeloid cell activation involved in immune response (GO:0002275)	45	22.08	+	2.04	1.96E-05	8.63E-04
interspecies interaction between organisms (GO:0044419)	173	85.12	+	2.03	2.90E-18	1.54E-15
positive regulation of protein metabolic process (GO:0051247)	150	73.85	+	2.03	8.97E-16	2.56E-13
autophagy (GO:0006914)	25	12.34	+	2.03	1.43E-03	3.40E-02
process utilizing autophagic mechanism (GO:0061919)	25	12.34	+	2.03	1.43E-03	3.40E-02
regulation of blood circulation (GO:1903522)	25	12.34	+	2.03	1.43E-03	3.39E-02
blood vessel development (GO:0001568)	43	21.23	+	2.03	3.59E-05	1.48E-03
regulation of body fluid levels (GO:0050878)	43	21.23	+	2.03	3.59E-05	1.48E-03
response to organic cyclic compound (GO:0014070)	80	39.56	+	2.02	1.22E-08	1.16E-06
positive regulation of signal transduction (GO:0009967)	147	72.70	+	2.02	3.87E-15	9.95E-13
granulocyte activation (GO:0036230)	43	21.27	+	2.02	3.68E-05	1.51E-03
regulation of protein modification process (GO:0031399)	161	79.76	+	2.02	1.48E-16	5.13E-14
response to oxygen-containing compound (GO:1901700)	138	68.45	+	2.02	3.93E-14	7.83E-12
response to organic substance (GO:0010033)	260	129.02	+	2.02	5.25E-28	2.79E-24
myeloid leukocyte mediated immunity (GO:0002444)	44	21.87	+	2.01	3.05E-05	1.29E-03
second-messenger-mediated signaling (GO:0019932)	29	14.42	+	2.01	7.46E-04	1.99E-02
regulation of response to stress (GO:0080134)	139	69.13	+	2.01	3.14E-14	6.52E-12
blood vessel morphogenesis (GO:0048514)	35	17.44	+	2.01	2.27E-04	7.30E-03
response to external stimulus (GO:0009605)	211	105.37	+	2.00	8.50E-22	9.69E-19
cellular response to drug (GO:0035690)	35	17.48	+	2.00	2.33E-04	7.45E-03
regulation of cellular response to stress (GO:0080135)	63	31.48	+	2.00	6.72E-07	4.34E-05

positive regulation of cell development (GO:0010720)	48	23.99	+	2.00	1.63E-05	7.43E-04
anatomical structure homeostasis (GO:0060249)	29	14.51	+	2.00	7.89E-04	2.07E-02
regulation of kinase activity (GO:0043549)	77	38.71	+	1.99	5.55E-08	4.64E-06
positive regulation of signaling (GO:0023056)	165	83.00	+	1.99	1.89E-16	6.27E-14
positive regulation of nervous system development (GO:0051962)	47	23.65	+	1.99	2.32E-05	1.01E-03
positive regulation of cell communication (GO:0010647)	164	82.66	+	1.98	2.73E-16	8.89E-14
response to inorganic substance (GO:0010035)	46	23.18	+	1.98	3.05E-05	1.29E-03
negative regulation of developmental process (GO:0051093)	82	41.35	+	1.98	2.28E-08	2.04E-06
vasculature development (GO:0001944)	44	22.21	+	1.98	5.24E-05	2.04E-03
response to extracellular stimulus (GO:0009991)	48	24.29	+	1.98	2.87E-05	1.23E-03
negative regulation of cell communication (GO:0010648)	120	60.75	+	1.98	8.26E-12	1.23E-09
negative regulation of transport (GO:0051051)	42	21.27	+	1.97	9.35E-05	3.40E-03
intracellular signal transduction (GO:0035556)	143	72.53	+	1.97	4.64E-14	9.14E-12
negative regulation of signaling (GO:0023057)	120	60.92	+	1.97	8.91E-12	1.32E-09
cellular response to growth factor stimulus (GO:0071363)	42	21.36	+	1.97	9.62E-05	3.48E-03
immune system development (GO:0002520)	55	28.03	+	1.96	7.20E-06	3.61E-04
response to growth factor (GO:0070848)	44	22.59	+	1.95	6.42E-05	2.43E-03
neutrophil activation (GO:0042119)	41	21.06	+	1.95	1.37E-04	4.65E-03
cardiovascular system development (GO:0072358)	44	22.63	+	1.94	6.59E-05	2.48E-03
regulation of phosphate metabolic process (GO:0019220)	151	77.68	+	1.94	2.52E-14	5.37E-12
regulation of phosphorus metabolic process (GO:0051174)	151	77.72	+	1.94	2.58E-14	5.42E-12
regulation of immune response (GO:0050776)	94	48.62	+	1.93	5.05E-09	5.07E-07
negative regulation of cell cycle (GO:0045786)	47	24.38	+	1.93	4.79E-05	1.91E-03
positive regulation of cellular catabolic process (GO:0031331)	31	16.08	+	1.93	9.19E-04	2.35E-02
hemopoiesis (GO:0030097)	47	24.42	+	1.92	4.90E-05	1.93E-03
positive regulation of molecular function (GO:0044093)	150	77.93	+	1.92	6.96E-14	1.34E-11
regulation of cytosolic calcium ion concentration (GO:0051480)	28	14.55	+	1.92	1.96E-03	4.43E-02
regulation of transferase activity (GO:0051338)	84	43.69	+	1.92	5.25E-08	4.41E-06
negative regulation of signal transduction (GO:0009968)	105	54.62	+	1.92	6.94E-10	8.15E-08
regulation of secretion by cell (GO:1903530)	64	33.31	+	1.92	2.06E-06	1.21E-04
regulation of epithelial cell proliferation (GO:0050678)	28	14.59	+	1.92	2.00E-03	4.48E-02

positive regulation of developmental process (GO:0051094)	114	59.60	+	1.91	1.53E-10	1.97E-08
leukocyte degranulation (GO:0043299)	41	21.44	+	1.91	1.66E-04	5.56E-03
negative regulation of defense response (GO:0031348)	33	17.27	+	1.91	8.46E-04	2.21E-02
regulation of protein transport (GO:0051223)	62	32.46	+	1.91	3.69E-06	2.03E-04
neutrophil mediated immunity (GO:0002446)	40	20.97	+	1.91	2.17E-04	7.00E-03
regulation of cell population proliferation (GO:0042127)	134	70.45	+	1.90	3.89E-12	6.03E-10
negative regulation of protein modification process (GO:0031400)	49	25.78	+	1.90	4.88E-05	1.92E-03
response to stress (GO:0006950)	293	154.63	+	1.89	8.91E-28	3.55E-24
regulation of peptide transport (GO:0090087)	64	33.82	+	1.89	3.64E-06	2.01E-04
hematopoietic or lymphoid organ development (GO:0048534)	50	26.46	+	1.89	4.18E-05	1.69E-03
positive regulation of response to stimulus (GO:0048584)	197	104.56	+	1.88	1.97E-17	7.87E-15
taxis (GO:0042330)	44	23.35	+	1.88	1.32E-04	4.52E-03
cellular response to organic cyclic compound (GO:0071407)	44	23.40	+	1.88	1.35E-04	4.61E-03
negative regulation of phosphorylation (GO:0042326)	36	19.14	+	1.88	6.55E-04	1.78E-02
transcription, DNA-templated (GO:0006351)	53	28.20	+	1.88	2.98E-05	1.27E-03
response to nutrient levels (GO:0031667)	43	22.89	+	1.88	1.71E-04	5.69E-03
immune system process (GO:0002376)	222	118.18	+	1.88	1.12E-19	7.42E-17
nucleic acid-templated transcription (GO:0097659)	53	28.25	+	1.88	3.05E-05	1.29E-03
negative regulation of cell differentiation (GO:0045596)	57	30.42	+	1.87	1.64E-05	7.43E-04
negative regulation of response to external stimulus (GO:0032102)	45	24.04	+	1.87	1.54E-04	5.17E-03
negative regulation of response to stimulus (GO:0048585)	146	78.10	+	1.87	1.19E-12	2.00E-10
response to wounding (GO:0009611)	46	24.63	+	1.87	1.23E-04	4.26E-03
neutrophil degranulation (GO:0043312)	38	20.50	+	1.85	6.53E-04	1.78E-02
regulation of establishment of protein localization (GO:0070201)	62	33.52	+	1.85	1.16E-05	5.53E-04
response to peptide (GO:1901652)	38	20.55	+	1.85	6.61E-04	1.79E-02
chemotaxis (GO:0006935)	43	23.27	+	1.85	2.82E-04	8.79E-03
RNA biosynthetic process (GO:0032774)	53	28.76	+	1.84	5.25E-05	2.04E-03
positive regulation of transport (GO:0051050)	80	43.43	+	1.84	6.32E-07	4.12E-05
neutrophil activation involved in immune response (GO:0002283)	38	20.67	+	1.84	6.90E-04	1.86E-02
positive regulation of catabolic process (GO:0009896)	35	19.06	+	1.84	1.02E-03	2.56E-02
regulation of hydrolase activity (GO:0051336)	101	55.09	+	1.83	1.61E-08	1.48E-06

regulation of anatomical structure morphogenesis (GO:0022603)	86	47.09	+	1.83	2.44E-07	1.79E-05
regulation of secretion (GO:0051046)	65	35.61	+	1.83	9.39E-06	4.57E-04
leukocyte mediated immunity (GO:0002443)	59	32.42	+	1.82	2.89E-05	1.24E-03
regulation of neurogenesis (GO:0050767)	65	35.73	+	1.82	9.88E-06	4.72E-04
response to abiotic stimulus (GO:0009628)	90	49.56	+	1.82	1.99E-07	1.52E-05
regulation of cell cycle (GO:0051726)	95	52.75	+	1.80	9.90E-08	7.82E-06
regulation of cellular protein metabolic process (GO:0032268)	204	113.50	+	1.80	4.02E-16	1.26E-13
cellular response to stress (GO:0033554)	135	75.25	+	1.79	1.84E-10	2.35E-08
cell migration (GO:0016477)	73	40.71	+	1.79	4.99E-06	2.63E-04
regulation of catalytic activity (GO:0050790)	182	101.80	+	1.79	5.89E-14	1.15E-11
regulation of molecular function (GO:0065009)	233	130.73	+	1.78	3.46E-18	1.73E-15
regulation of proteolysis (GO:0030162)	56	31.48	+	1.78	7.52E-05	2.79E-03
negative regulation of phosphate metabolic process (GO:0045936)	43	24.21	+	1.78	5.58E-04	1.55E-02
negative regulation of phosphorus metabolic process (GO:0010563)	43	24.25	+	1.77	5.68E-04	1.58E-02
negative regulation of transcription, DNA-templated (GO:0045892)	93	52.45	+	1.77	2.87E-07	2.09E-05
cell surface receptor signaling pathway (GO:0007166)	188	106.05	+	1.77	3.28E-14	6.70E-12
regulation of signal transduction (GO:0009966)	237	134.43	+	1.76	5.47E-18	2.57E-15
regulation of protein metabolic process (GO:0051246)	212	121.03	+	1.75	1.50E-15	4.07E-13
regulation of multicellular organismal process (GO:0051239)	239	136.60	+	1.75	8.98E-18	4.10E-15
positive regulation of transcription by RNA polymerase II (GO:0045944)	91	52.07	+	1.75	8.01E-07	5.16E-05
regulation of cell differentiation (GO:0045595)	138	79.00	+	1.75	4.77E-10	5.69E-08
nucleobase-containing compound biosynthetic process (GO:0034654)	83	47.56	+	1.75	2.67E-06	1.53E-04
negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	106	60.75	+	1.74	9.44E-08	7.57E-06
circulatory system development (GO:0072359)	65	37.26	+	1.74	3.24E-05	1.36E-03
metal ion homeostasis (GO:0055065)	47	26.97	+	1.74	5.06E-04	1.42E-02
negative regulation of transcription by RNA polymerase II (GO:0000122)	66	37.95	+	1.74	3.62E-05	1.49E-03
cellular response to nitrogen compound (GO:1901699)	49	28.20	+	1.74	3.40E-04	1.03E-02
negative regulation of nucleic acid-templated transcription (GO:1903507)	95	54.71	+	1.74	4.94E-07	3.33E-05
negative regulation of RNA biosynthetic process (GO:1902679)	95	54.79	+	1.73	5.11E-07	3.39E-05
regulation of cell development (GO:0060284)	71	41.09	+	1.73	1.80E-05	8.05E-04
regulation of multicellular organismal development (GO:2000026)	153	88.57	+	1.73	1.21E-10	1.59E-08
positive regulation of immune response (GO:0050778)	55	31.86	+	1.73	1.78E-04	5.88E-03

negative regulation of cellular biosynthetic process (GO:0031327)	115	66.79	+	1.72	3.95E-08	3.41E-06
regulation of mitotic cell cycle (GO:0007346)	50	29.05	+	1.72	3.98E-04	1.17E-02
negative regulation of biosynthetic process (GO:0009890)	117	68.06	+	1.72	3.52E-08	3.05E-06
regulation of growth (GO:0040008)	50	29.10	+	1.72	4.04E-04	1.19E-02
negative regulation of RNA metabolic process (GO:0051253)	101	58.79	+	1.72	3.87E-07	2.67E-05
heterocycle biosynthetic process (GO:0018130)	87	50.67	+	1.72	2.59E-06	1.49E-04
cellular response to organonitrogen compound (GO:0071417)	44	25.78	+	1.71	1.15E-03	2.83E-02
aromatic compound biosynthetic process (GO:0019438)	87	51.01	+	1.71	3.91E-06	2.12E-04
symbiotic process (GO:0044403)	64	37.56	+	1.70	7.19E-05	2.67E-03
regulation of nervous system development (GO:0051960)	69	40.50	+	1.70	4.40E-05	1.77E-03
positive regulation of cell population proliferation (GO:0008284)	67	39.35	+	1.70	5.08E-05	1.99E-03
negative regulation of molecular function (GO:0044092)	85	50.11	+	1.70	4.89E-06	2.60E-04
negative regulation of macromolecule biosynthetic process (GO:0010558)	109	64.28	+	1.70	2.03E-07	1.53E-05
negative regulation of nucleobase-containing compound metabolic process (GO:0045934)	107	63.21	+	1.69	3.28E-07	2.35E-05
negative regulation of protein metabolic process (GO:0051248)	81	47.90	+	1.69	1.20E-05	5.68E-04
cellular metal ion homeostasis (GO:0006875)	41	24.25	+	1.69	1.76E-03	4.03E-02
regulation of signaling (GO:0023051)	264	156.80	+	1.68	9.62E-18	4.15E-15
regulation of cell communication (GO:0010646)	261	155.10	+	1.68	1.59E-17	6.49E-15
ion homeostasis (GO:0050801)	57	33.95	+	1.68	2.74E-04	8.54E-03
negative regulation of cellular protein metabolic process (GO:0032269)	75	44.71	+	1.68	2.68E-05	1.17E-03
negative regulation of cellular component organization (GO:0051129)	51	30.54	+	1.67	7.57E-04	2.02E-02
regulation of system process (GO:0044057)	42	25.18	+	1.67	2.09E-03	4.67E-02
regulation of response to stimulus (GO:0048583)	312	187.30	+	1.67	4.80E-21	4.26E-18
organic cyclic compound biosynthetic process (GO:1901362)	95	57.09	+	1.66	3.21E-06	1.78E-04
homeostatic process (GO:0042592)	118	70.96	+	1.66	1.78E-07	1.37E-05
positive regulation of nitrogen compound metabolic process (GO:0051173)	229	137.79	+	1.66	1.43E-14	3.16E-12
regulation of developmental process (GO:0050793)	187	112.77	+	1.66	1.28E-11	1.82E-09
positive regulation of macromolecule metabolic process (GO:0010604)	241	145.36	+	1.66	2.72E-15	7.22E-13
localization of cell (GO:0051674)	76	45.86	+	1.66	4.45E-05	1.79E-03

cell motility (GO:0048870)	76	45.86	+	1.66	4.45E-05	1.78E-03
regulation of transport (GO:0051049)	134	81.08	+	1.65	2.61E-08	2.30E-06
inorganic ion homeostasis (GO:0098771)	51	30.88	+	1.65	8.31E-04	2.17E-02
positive regulation of gene expression (GO:0010628)	143	86.70	+	1.65	9.04E-09	8.69E-07
regulation of protein localization (GO:0032880)	75	45.52	+	1.65	5.82E-05	2.23E-03
cation homeostasis (GO:0055080)	50	30.37	+	1.65	1.04E-03	2.60E-02
positive regulation of metabolic process (GO:0009893)	260	158.08	+	1.64	3.78E-16	1.21E-13
positive regulation of nucleic acid-templated transcription (GO:1903508)	116	70.53	+	1.64	3.33E-07	2.37E-05
regulation of neuron differentiation (GO:0045664)	47	28.59	+	1.64	1.47E-03	3.44E-02
positive regulation of RNA biosynthetic process (GO:1902680)	116	70.57	+	1.64	3.38E-07	2.39E-05
positive regulation of transcription, DNA-templated (GO:0045893)	109	66.36	+	1.64	8.52E-07	5.40E-05
chemical homeostasis (GO:0048878)	79	48.11	+	1.64	3.51E-05	1.45E-03
tube morphogenesis (GO:0035239)	46	28.03	+	1.64	1.82E-03	4.16E-02
positive regulation of cellular metabolic process (GO:0031325)	237	144.85	+	1.64	2.23E-14	4.88E-12
negative regulation of cellular process (GO:0048523)	340	207.94	+	1.64	4.94E-22	7.17E-19
viral process (GO:0016032)	55	33.65	+	1.63	6.97E-04	1.88E-02
response to chemical (GO:0042221)	314	192.28	+	1.63	7.59E-20	5.27E-17
negative regulation of gene expression (GO:0010629)	123	75.42	+	1.63	2.06E-07	1.53E-05
response to nitrogen compound (GO:1901698)	76	46.71	+	1.63	7.18E-05	2.68E-03
negative regulation of nitrogen compound metabolic process (GO:0051172)	166	102.65	+	1.62	1.61E-09	1.75E-07
negative regulation of biological process (GO:0048519)	381	235.71	+	1.62	9.06E-25	1.81E-21
positive regulation of RNA metabolic process (GO:0051254)	120	74.36	+	1.61	6.10E-07	3.99E-05
regulation of localization (GO:0032879)	194	120.73	+	1.61	6.61E-11	9.18E-09
negative regulation of cellular metabolic process (GO:0031324)	178	111.03	+	1.60	5.93E-10	7.01E-08
response to drug (GO:0042493)	70	43.82	+	1.60	2.21E-04	7.12E-03
secretion by cell (GO:0032940)	68	42.63	+	1.60	3.32E-04	1.01E-02
positive regulation of cellular biosynthetic process (GO:0031328)	137	86.23	+	1.59	1.60E-07	1.25E-05
exocytosis (GO:0006887)	53	33.52	+	1.58	1.76E-03	4.04E-02
response to organonitrogen compound (GO:0010243)	68	43.14	+	1.58	3.73E-04	1.11E-02
positive regulation of biosynthetic process (GO:0009891)	138	87.59	+	1.58	2.60E-07	1.91E-05
locomotion (GO:0040011)	88	55.90	+	1.57	6.29E-05	2.40E-03
positive regulation of biological process (GO:0048518)	419	267.07	+	1.57	1.71E-25	4.55E-22
export from cell (GO:0140352)	70	44.62	+	1.57	4.45E-04	1.29E-02
regulation of transcription by RNA polymerase II (GO:0006357)	151	96.31	+	1.57	7.53E-08	6.10E-06
positive regulation of cellular process (GO:0048522)	379	243.54	+	1.56	1.87E-21	1.75E-18

negative regulation of macromolecule metabolic process (GO:0010605)	179	115.37	+	1.55	5.41E-09	5.40E-07
negative regulation of metabolic process (GO:0009892)	196	126.60	+	1.55	1.02E-09	1.16E-07
positive regulation of nucleobase-containing compound metabolic process (GO:0045935)	126	81.42	+	1.55	2.26E-06	1.32E-04
anatomical structure formation involved in morphogenesis (GO:0048646)	59	38.20	+	1.54	1.87E-03	4.25E-02
secretion (GO:0046903)	73	47.60	+	1.53	5.07E-04	1.43E-02
positive regulation of macromolecule biosynthetic process (GO:0010557)	126	82.27	+	1.53	3.42E-06	1.89E-04
cell cycle (GO:0007049)	88	58.32	+	1.51	2.07E-04	6.72E-03
organonitrogen compound catabolic process (GO:1901565)	68	45.56	+	1.49	1.56E-03	3.65E-02
signal transduction (GO:0007165)	325	218.91	+	1.48	1.22E-14	2.74E-12
positive regulation of cellular component organization (GO:0051130)	77	51.90	+	1.48	1.07E-03	2.65E-02
regulation of biosynthetic process (GO:0009889)	271	182.79	+	1.48	9.63E-12	1.41E-09
regulation of nucleic acid-templated transcription (GO:1903506)	223	150.42	+	1.48	1.82E-09	1.95E-07
regulation of RNA biosynthetic process (GO:2001141)	223	150.63	+	1.48	1.89E-09	2.01E-07
neurogenesis (GO:0022008)	105	71.00	+	1.48	9.89E-05	3.56E-03
regulation of transcription, DNA-templated (GO:0006355)	217	147.57	+	1.47	7.28E-09	7.04E-07
regulation of cellular biosynthetic process (GO:0031326)	263	179.39	+	1.47	7.24E-11	9.88E-09
cell communication (GO:0007154)	348	238.65	+	1.46	7.61E-15	1.81E-12
regulation of biological quality (GO:0065008)	256	175.86	+	1.46	3.71E-10	4.55E-08
signaling (GO:0023052)	340	234.14	+	1.45	3.37E-14	6.81E-12
cellular response to stimulus (GO:0051716)	420	289.31	+	1.45	7.02E-19	4.31E-16
regulation of RNA metabolic process (GO:0051252)	235	161.95	+	1.45	3.91E-09	4.00E-07
regulation of primary metabolic process (GO:0080090)	375	258.52	+	1.45	4.65E-16	1.40E-13
response to endogenous stimulus (GO:0009719)	93	64.32	+	1.45	6.22E-04	1.70E-02
generation of neurons (GO:0048699)	96	66.57	+	1.44	4.80E-04	1.37E-02
regulation of nitrogen compound metabolic process (GO:0051171)	361	250.60	+	1.44	7.81E-15	1.83E-12
regulation of cellular metabolic process (GO:0031323)	384	267.32	+	1.44	6.49E-16	1.88E-13
regulation of nucleobase-containing compound metabolic process (GO:0019219)	248	172.92	+	1.43	3.30E-09	3.42E-07
cellular nitrogen compound biosynthetic process (GO:0044271)	102	71.17	+	1.43	4.34E-04	1.26E-02
regulation of gene expression (GO:0010468)	273	191.17	+	1.43	4.28E-10	5.17E-08
regulation of macromolecule biosynthetic process (GO:0010556)	247	173.09	+	1.43	5.64E-09	5.56E-07
regulation of metabolic process (GO:0019222)	410	288.08	+	1.42	1.12E-16	4.07E-14

regulation of cellular macromolecule biosynthetic process (GO:2000112)	237	167.14	+	1.42	2.17E-08	1.95E-06
regulation of cellular component organization (GO:0051128)	146	103.29	+	1.41	3.29E-05	1.38E-03
movement of cell or subcellular component (GO:0006928)	94	66.79	+	1.41	1.23E-03	2.99E-02
vesicle-mediated transport (GO:0016192)	116	82.70	+	1.40	3.43E-04	1.03E-02
regulation of macromolecule metabolic process (GO:0060255)	369	264.39	+	1.40	3.47E-13	6.08E-11
cell differentiation (GO:0030154)	221	159.27	+	1.39	4.81E-07	3.25E-05
response to stimulus (GO:0050896)	504	364.65	+	1.38	3.36E-20	2.44E-17
macromolecule biosynthetic process (GO:0009059)	105	75.98	+	1.38	1.22E-03	2.98E-02
anatomical structure morphogenesis (GO:0009653)	128	92.78	+	1.38	2.91E-04	9.04E-03
cellular developmental process (GO:0048869)	222	161.57	+	1.37	8.77E-07	5.51E-05
phosphate-containing compound metabolic process (GO:0006796)	123	91.46	+	1.34	1.13E-03	2.76E-02
nervous system development (GO:0007399)	136	101.88	+	1.33	7.41E-04	1.99E-02
phosphorus metabolic process (GO:0006793)	123	92.61	+	1.33	1.76E-03	4.04E-02
system development (GO:0048731)	251	189.73	+	1.32	2.00E-06	1.19E-04
organic substance biosynthetic process (GO:1901576)	161	123.03	+	1.31	5.38E-04	1.51E-02
multicellular organism development (GO:0007275)	282	216.10	+	1.30	9.51E-07	5.95E-05
biosynthetic process (GO:0009058)	163	126.34	+	1.29	8.85E-04	2.29E-02
regulation of cellular process (GO:0050794)	618	480.19	+	1.29	2.72E-20	2.17E-17
developmental process (GO:0032502)	322	251.75	+	1.28	6.48E-07	4.20E-05
regulation of biological process (GO:0050789)	639	503.03	+	1.27	3.03E-20	2.30E-17
anatomical structure development (GO:0048856)	294	232.65	+	1.26	7.69E-06	3.83E-04
biological regulation (GO:0065007)	671	532.81	+	1.26	9.25E-22	9.84E-19
localization (GO:0051179)	310	246.82	+	1.26	6.52E-06	3.32E-04
animal organ development (GO:0048513)	172	137.19	+	1.25	2.12E-03	4.73E-02
establishment of localization (GO:0051234)	244	199.09	+	1.23	5.72E-04	1.59E-02
organonitrogen compound metabolic process (GO:1901564)	279	228.44	+	1.22	2.01E-04	6.55E-03
multicellular organismal process (GO:0032501)	365	299.35	+	1.22	7.05E-06	3.56E-04
transport (GO:0006810)	236	193.56	+	1.22	9.28E-04	2.37E-02
cellular macromolecule metabolic process (GO:0044260)	263	218.91	+	1.20	9.88E-04	2.49E-02
nitrogen compound metabolic process (GO:0006807)	353	300.84	+	1.17	3.43E-04	1.03E-02
macromolecule metabolic process (GO:0043170)	311	265.79	+	1.17	1.49E-03	3.48E-02
cellular metabolic process (GO:0044237)	385	329.34	+	1.17	1.96E-04	6.40E-03
cellular process (GO:0009987)	765	657.58	+	1.16	9.54E-18	4.23E-15
primary metabolic process (GO:0044238)	372	320.62	+	1.16	5.41E-04	1.51E-02
organic substance metabolic process (GO:0071704)	387	336.79	+	1.15	7.95E-04	2.08E-02
metabolic process (GO:0008152)	415	361.25	+	1.15	3.73E-04	1.11E-02
biological_process (GO:0008150)	818	757.64	+	1.08	1.40E-09	1.55E-07

Unclassified (UNCLASSIFIED)	69	129.36	-	.53	1.40E-09	1.54E-07
detection of stimulus (GO:0051606)	12	29.95	-	.40	3.44E-04	1.03E-02
detection of chemical stimulus (GO:0009593)	7	22.08	-	.32	4.63E-04	1.33E-02
detection of stimulus involved in sensory perception (GO:0050906)	5	23.23	-	.22	1.51E-05	6.93E-04
sensory perception of chemical stimulus (GO:0007606)	4	22.97	-	.17	2.84E-06	1.61E-04
detection of chemical stimulus involved in sensory perception (GO:0050907)	3	20.55	-	.15	4.11E-06	2.22E-04
sensory perception of smell (GO:0007608)	1	19.91	-	.05	9.61E-08	7.67E-06
detection of chemical stimulus involved in sensory perception of smell (GO:0050911)	0	18.68	-	<0.01	1.46E-08	1.36E-06

Table A7: All GO-terms from lightyellow module. The number of relevant genes from the network (#genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process complete	#genes	exp	FE	raw P-value	FDR
post-embryonic camera-type eye development (GO:0031077)	4	.19	20.94	1.40E-04	8.70E-03
surfactant homeostasis (GO:0043129)	4	.25	16.29	2.91E-04	1.54E-02
chemical homeostasis within a tissue (GO:0048875)	4	.30	13.33	5.33E-04	2.50E-02
response to peptidoglycan (GO:0032494)	5	.38	13.09	1.11E-04	7.20E-03
regulation of plasma membrane organization (GO:1903729)	5	.41	12.21	1.45E-04	8.92E-03
response to muramyl dipeptide (GO:0032495)	6	.55	10.99	5.01E-05	3.81E-03
cellular copper ion homeostasis (GO:0006878)	4	.38	10.47	1.12E-03	4.55E-02
positive regulation of T-helper 1 type immune response (GO:0002827)	5	.57	8.72	5.38E-04	2.52E-02
positive regulation of interleukin-1 beta secretion (GO:0050718)	8	.96	8.38	1.51E-05	1.44E-03
positive regulation of toll-like receptor signaling pathway (GO:0034123)	5	.65	7.63	9.10E-04	3.86E-02
regulation of interleukin-6 biosynthetic process (GO:0045408)	5	.65	7.63	9.10E-04	3.85E-02
regulation of interleukin-1 beta secretion (GO:0050706)	10	1.31	7.63	2.68E-06	3.14E-04
cytokine secretion (GO:0050663)	10	1.39	7.19	4.31E-06	4.68E-04
positive regulation of interleukin-1 secretion (GO:0050716)	8	1.12	7.15	4.08E-05	3.27E-03
T cell differentiation involved in immune response (GO:0002292)	6	.87	6.87	4.58E-04	2.23E-02

positive regulation of transcription from RNA polymerase II promoter in response to stress (GO:0036003)	6	.90	6.66	5.30E-04	2.51E-02
negative regulation of cytokine biosynthetic process (GO:0042036)	6	.90	6.66	5.30E-04	2.50E-02
MyD88-dependent toll-like receptor signaling pathway (GO:0002755)	6	.90	6.66	5.30E-04	2.49E-02
regulation of interleukin-1 secretion (GO:0050704)	10	1.53	6.54	8.96E-06	9.05E-04
T cell activation involved in immune response (GO:0002286)	11	1.72	6.40	3.88E-06	4.33E-04
purinergic nucleotide receptor signaling pathway (GO:0035590)	8	1.26	6.37	8.37E-05	5.78E-03
lipopolysaccharide-mediated signaling pathway (GO:0031663)	6	.96	6.28	6.98E-04	3.09E-02
acute-phase response (GO:0006953)	7	1.12	6.26	2.57E-04	1.40E-02
interferon-gamma-mediated signaling pathway (GO:0060333)	12	1.96	6.11	2.18E-06	2.59E-04
macrophage activation (GO:0042116)	8	1.42	5.64	1.79E-04	1.07E-02
regulation of cytokine biosynthetic process (GO:0042035)	17	3.03	5.61	5.17E-08	9.28E-06
alpha-beta T cell activation (GO:0046631)	10	1.80	5.55	3.20E-05	2.71E-03
positive regulation of interleukin-6 production (GO:0032755)	14	2.62	5.34	1.31E-06	1.68E-04
positive regulation of interleukin-1 beta production (GO:0032731)	8	1.50	5.33	2.53E-04	1.39E-02
positive regulation of cytokine biosynthetic process (GO:0042108)	10	1.96	5.09	6.21E-05	4.57E-03
regulation of interferon-beta production (GO:0032648)	7	1.39	5.03	8.37E-04	3.61E-02
regulation of receptor internalization (GO:0002090)	8	1.61	4.97	3.88E-04	1.95E-02
granulocyte chemotaxis (GO:0071621)	11	2.24	4.92	3.59E-05	2.94E-03
positive regulation of tumor necrosis factor production (GO:0032760)	12	2.46	4.89	1.69E-05	1.60E-03
regulation of osteoclast differentiation (GO:0045670)	9	1.86	4.85	2.00E-04	1.17E-02
negative regulation of cytokine secretion (GO:0050710)	9	1.86	4.85	2.00E-04	1.16E-02
alpha-beta T cell differentiation (GO:0046632)	7	1.45	4.84	1.03E-03	4.24E-02
regulation of interleukin-8 production (GO:0032677)	10	2.07	4.82	9.35E-05	6.27E-03
toll-like receptor signaling pathway (GO:0002224)	13	2.70	4.81	8.79E-06	8.94E-04
regulation of toll-like receptor signaling pathway (GO:0034121)	9	1.88	4.78	2.21E-04	1.26E-02

positive regulation of tumor necrosis factor superfamily cytokine production (GO:1903557)	12	2.51	4.78	2.06E-05	1.85E-03
granulocyte migration (GO:0097530)	12	2.54	4.73	2.27E-05	2.01E-03
regulation of interleukin-1 beta production (GO:0032651)	11	2.35	4.69	5.32E-05	4.03E-03
neutrophil mediated immunity (GO:0002446)	63	13.45	4.68	1.50E-22	3.98E-19
neutrophil activation involved in immune response (GO:0002283)	62	13.26	4.67	3.60E-22	5.75E-19
positive regulation of interleukin-8 production (GO:0032757)	7	1.50	4.66	1.25E-03	5.01E-02
astrocyte differentiation (GO:0048708)	7	1.50	4.66	1.25E-03	5.00E-02
neutrophil activation (GO:0042119)	63	13.51	4.66	1.81E-22	4.14E-19
neutrophil degranulation (GO:0043312)	61	13.15	4.64	1.15E-21	1.42E-18
granulocyte activation (GO:0036230)	63	13.64	4.62	2.92E-22	5.82E-19
myeloid leukocyte activation (GO:0002274)	73	15.91	4.59	1.45E-25	1.16E-21
acute inflammatory response (GO:0002526)	9	1.96	4.58	2.95E-04	1.56E-02
positive regulation of interleukin-1 production (GO:0032732)	8	1.75	4.58	6.34E-04	2.87E-02
pattern recognition receptor signaling pathway (GO:0002221)	15	3.30	4.54	3.46E-06	3.94E-04
myeloid cell activation involved in immune response (GO:0002275)	64	14.16	4.52	3.72E-22	5.40E-19
tissue regeneration (GO:0042246)	8	1.77	4.51	6.96E-04	3.10E-02
myeloid leukocyte mediated immunity (GO:0002444)	63	14.03	4.49	1.07E-21	1.43E-18
leukocyte degranulation (GO:0043299)	61	13.75	4.44	8.81E-21	8.27E-18
neutrophil migration (GO:1990266)	10	2.26	4.42	1.81E-04	1.07E-02
lymphocyte activation involved in immune response (GO:0002285)	13	2.95	4.41	2.04E-05	1.84E-03
neutrophil chemotaxis (GO:0030593)	9	2.05	4.40	3.88E-04	1.96E-02
cellular response to molecule of bacterial origin (GO:0071219)	24	5.49	4.38	8.35E-09	1.64E-06
regulation of interleukin-1 production (GO:0032652)	12	2.76	4.35	4.74E-05	3.68E-03
regulation of interleukin-6 production (GO:0032675)	17	3.93	4.33	1.44E-06	1.82E-04
cytokine production (GO:0001816)	19	4.39	4.32	3.52E-07	5.15E-05
leukocyte activation involved in immune response (GO:0002366)	73	16.89	4.32	3.82E-24	1.52E-20
cell activation involved in immune response (GO:0002263)	73	17.00	4.29	5.41E-24	1.73E-20
cellular response to lipopolysaccharide (GO:0071222)	22	5.21	4.22	6.24E-08	1.09E-05
exocytic process (GO:0140029)	8	1.94	4.13	1.18E-03	4.75E-02

regulation of tumor necrosis factor production (GO:0032680)	17	4.18	4.07	3.06E-06	3.54E-04
cellular response to mechanical stimulus (GO:0071260)	9	2.21	4.07	6.46E-04	2.91E-02
cellular response to biotic stimulus (GO:0071216)	25	6.14	4.07	1.52E-08	2.88E-06
cellular response to interferon-gamma (GO:0071346)	18	4.48	4.02	1.85E-06	2.25E-04
regulation of tumor necrosis factor superfamily cytokine production (GO:1903555)	17	4.26	3.99	3.88E-06	4.30E-04
response to interferon-gamma (GO:0034341)	20	5.08	3.94	6.70E-07	9.30E-05
negative regulation of myeloid cell differentiation (GO:0045638)	9	2.29	3.93	8.21E-04	3.55E-02
response to molecule of bacterial origin (GO:0002237)	36	9.28	3.88	3.47E-11	1.15E-08
response to lipopolysaccharide (GO:0032496)	34	8.81	3.86	1.42E-10	4.20E-08
regulation of pattern recognition receptor signaling pathway (GO:0062207)	10	2.59	3.86	4.86E-04	2.35E-02
myeloid leukocyte differentiation (GO:0002573)	13	3.38	3.84	7.57E-05	5.37E-03
regulated exocytosis (GO:0045055)	72	18.99	3.79	8.03E-21	8.01E-18
leukocyte activation (GO:0045321)	95	25.13	3.78	2.36E-27	3.77E-23
positive regulation of NF-kappaB transcription factor activity (GO:0051092)	16	4.26	3.76	1.49E-05	1.43E-03
regulation of organ growth (GO:0046620)	10	2.67	3.74	6.08E-04	2.78E-02
positive regulation of cytokine secretion (GO:0050715)	15	4.12	3.64	3.85E-05	3.10E-03
myeloid leukocyte migration (GO:0097529)	12	3.33	3.60	2.45E-04	1.36E-02
exocytosis (GO:0006887)	76	21.50	3.53	2.65E-20	2.35E-17
positive regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043123)	18	5.10	3.53	9.90E-06	9.94E-04
T cell activation (GO:0042110)	23	6.52	3.53	5.98E-07	8.38E-05
T cell differentiation (GO:0030217)	13	3.77	3.45	2.02E-04	1.17E-02
cell activation (GO:0001775)	99	29.14	3.40	3.13E-25	1.66E-21
stress-activated MAPK cascade (GO:0051403)	10	2.95	3.39	1.21E-03	4.88E-02
leukocyte chemotaxis (GO:0030595)	13	3.85	3.38	2.46E-04	1.36E-02
protein secretion (GO:0009306)	15	4.45	3.37	8.58E-05	5.90E-03
regulation of cytokine secretion (GO:0050707)	20	5.95	3.36	6.25E-06	6.65E-04
signal release (GO:0023061)	21	6.25	3.36	3.71E-06	4.20E-04
establishment of protein localization to extracellular region (GO:0035592)	15	4.48	3.35	9.14E-05	6.18E-03
negative regulation of protein secretion (GO:0050709)	12	3.60	3.33	4.74E-04	2.29E-02
negative regulation of hemopoiesis (GO:1903707)	13	3.90	3.33	2.79E-04	1.49E-02
regulation of myeloid leukocyte differentiation (GO:0002761)	11	3.33	3.30	8.58E-04	3.67E-02

inflammatory response (GO:0006954)	46	13.97	3.29	1.17E-11	4.14E-09
secretion by cell (GO:0032940)	89	27.34	3.25	1.66E-21	1.89E-18
cell chemotaxis (GO:0060326)	18	5.57	3.23	2.91E-05	2.51E-03
protein localization to extracellular region (GO:0071692)	15	4.67	3.21	1.40E-04	8.68E-03
negative regulation of peptide secretion (GO:0002792)	12	3.77	3.19	6.83E-04	3.05E-02
leukocyte differentiation (GO:0002521)	30	9.41	3.19	9.37E-08	1.57E-05
calcium-mediated signaling (GO:0019722)	13	4.09	3.18	4.27E-04	2.13E-02
positive regulation of peptidase activity (GO:0010952)	17	5.38	3.16	6.22E-05	4.55E-03
platelet degranulation (GO:0002576)	11	3.49	3.15	1.23E-03	4.96E-02
secretion (GO:0046903)	96	30.54	3.14	3.30E-22	5.86E-19
leukocyte mediated immunity (GO:0002443)	65	20.79	3.13	4.64E-15	2.65E-12
positive regulation of inflammatory response (GO:0050729)	12	3.85	3.12	8.14E-04	3.55E-02
export from cell (GO:0140352)	89	28.63	3.11	2.73E-20	2.30E-17
positive regulation of endopeptidase activity (GO:0010950)	15	4.86	3.09	2.11E-04	1.21E-02
lymphocyte differentiation (GO:0030098)	20	6.55	3.05	2.29E-05	2.02E-03
negative regulation of cytokine production (GO:0001818)	23	7.59	3.03	6.35E-06	6.67E-04
regulation of leukocyte migration (GO:0002685)	17	5.65	3.01	1.09E-04	7.14E-03
peptide secretion (GO:0002790)	15	4.99	3.00	2.78E-04	1.49E-02
cellular response to external stimulus (GO:0071496)	30	10.07	2.98	3.58E-07	5.19E-05
cellular response to starvation (GO:0009267)	15	5.08	2.96	3.27E-04	1.69E-02
lymphocyte activation (GO:0046649)	31	10.51	2.95	2.78E-07	4.14E-05
regulation of myeloid cell differentiation (GO:0045637)	19	6.44	2.95	5.62E-05	4.21E-03
immune effector process (GO:0002252)	87	29.55	2.94	1.91E-18	1.33E-15
positive regulation of protein-containing complex assembly (GO:0031334)	20	6.85	2.92	4.13E-05	3.30E-03
cellular response to nutrient levels (GO:0031669)	21	7.23	2.90	2.88E-05	2.50E-03
regulation of actin polymerization or depolymerization (GO:0008064)	15	5.21	2.88	4.24E-04	2.12E-02
regulation of synaptic plasticity (GO:0048167)	15	5.24	2.86	4.46E-04	2.20E-02
regulation of actin filament length (GO:0030832)	15	5.24	2.86	4.46E-04	2.19E-02
neurotransmitter transport (GO:0006836)	14	4.97	2.82	7.88E-04	3.45E-02
antigen processing and presentation of peptide antigen (GO:0048002)	15	5.32	2.82	5.19E-04	2.47E-02
regulation of endocytosis (GO:0030100)	16	5.70	2.81	3.60E-04	1.84E-02
regulation of ossification (GO:0030278)	15	5.38	2.79	5.73E-04	2.65E-02
positive regulation of cytokine production (GO:0001819)	35	12.61	2.78	1.86E-07	2.91E-05

negative regulation of protein transport (GO:0051224)	14	5.05	2.77	9.13E-04	3.86E-02
negative regulation of establishment of protein localization (GO:1904950)	14	5.16	2.71	1.11E-03	4.53E-02
regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043122)	18	6.69	2.69	2.53E-04	1.38E-02
cytokine-mediated signaling pathway (GO:0019221)	50	18.94	2.64	1.95E-09	4.32E-07
positive regulation of peptide secretion (GO:0002793)	22	8.40	2.62	8.00E-05	5.60E-03
antigen processing and presentation (GO:0019882)	16	6.11	2.62	7.27E-04	3.21E-02
cellular response to extracellular stimulus (GO:0031668)	21	8.08	2.60	1.26E-04	7.92E-03
innate immune response (GO:0045087)	55	21.20	2.59	6.45E-10	1.66E-07
regulation of leukocyte differentiation (GO:1902105)	20	7.75	2.58	1.99E-04	1.16E-02
positive regulation of DNA-binding transcription factor activity (GO:0051091)	19	7.37	2.58	2.87E-04	1.53E-02
myeloid cell differentiation (GO:0030099)	16	6.22	2.57	8.68E-04	3.69E-02
positive regulation of protein secretion (GO:0050714)	20	7.78	2.57	2.07E-04	1.19E-02
response to starvation (GO:0042594)	16	6.28	2.55	9.46E-04	3.96E-02
regulation of cytokine production (GO:0001817)	51	20.03	2.55	4.61E-09	9.44E-07
regulation of leukocyte cell-cell adhesion (GO:1903037)	22	8.65	2.54	1.18E-04	7.60E-03
regulation of cysteine-type endopeptidase activity (GO:2000116)	17	6.69	2.54	6.82E-04	3.05E-02
positive regulation of leukocyte cell-cell adhesion (GO:1903039)	16	6.30	2.54	9.87E-04	4.11E-02
positive regulation of cell-cell adhesion (GO:0022409)	19	7.50	2.53	3.56E-04	1.82E-02
response to wounding (GO:0009611)	40	15.80	2.53	2.86E-07	4.23E-05
positive regulation of response to external stimulus (GO:0032103)	35	13.89	2.52	1.89E-06	2.28E-04
regulation of protein polymerization (GO:0032271)	16	6.39	2.51	1.12E-03	4.54E-02
response to biotic stimulus (GO:0009607)	94	37.63	2.50	1.42E-15	9.42E-13
regulation of cell-cell adhesion (GO:0022407)	29	11.63	2.49	1.86E-05	1.71E-03
cellular response to lipid (GO:0071396)	36	14.44	2.49	1.56E-06	1.94E-04
response to other organism (GO:0051707)	91	36.70	2.48	1.01E-14	5.40E-12
response to external biotic stimulus (GO:0043207)	91	36.76	2.48	1.05E-14	5.40E-12
defense response to other organism (GO:0098542)	65	26.47	2.46	1.20E-10	3.61E-08
cellular response to cytokine stimulus (GO:0071345)	69	28.13	2.45	3.05E-11	1.04E-08

response to cytokine (GO:0034097)	75	30.59	2.45	3.70E-12	1.41E-09
regulation of cellular component size (GO:0032535)	26	10.62	2.45	6.53E-05	4.76E-03
positive regulation of protein transport (GO:0051222)	31	12.66	2.45	1.18E-05	1.17E-03
immune response (GO:0006955)	125	51.08	2.45	5.01E-20	3.81E-17
response to bacterium (GO:0009617)	47	19.35	2.43	8.85E-08	1.50E-05
regulation of protein-containing complex assembly (GO:0043254)	30	12.39	2.42	1.94E-05	1.78E-03
defense response (GO:0006952)	90	37.36	2.41	7.39E-14	3.37E-11
positive regulation of cell adhesion (GO:0045785)	28	11.68	2.40	4.33E-05	3.42E-03
vesicle organization (GO:0016050)	20	8.40	2.38	6.66E-04	2.99E-02
positive regulation of defense response (GO:0031349)	24	10.10	2.38	1.78E-04	1.07E-02
positive regulation of establishment of protein localization (GO:1904951)	31	13.10	2.37	2.02E-05	1.83E-03
response to lipid (GO:0033993)	56	23.71	2.36	1.19E-08	2.32E-06
vesicle-mediated transport (GO:0016192)	125	53.05	2.36	1.02E-18	7.39E-16
negative regulation of immune system process (GO:0002683)	40	17.00	2.35	1.64E-06	2.03E-04
positive regulation of secretion (GO:0051047)	29	12.33	2.35	4.19E-05	3.33E-03
regulation of hemopoiesis (GO:1903706)	30	12.77	2.35	4.90E-05	3.76E-03
phagocytosis (GO:0006909)	21	8.95	2.35	5.39E-04	2.51E-02
wound healing (GO:0042060)	31	13.24	2.34	3.53E-05	2.90E-03
regulation of T cell activation (GO:0050863)	21	9.03	2.32	5.87E-04	2.70E-02
regulation of inflammatory response (GO:0050727)	34	14.65	2.32	1.43E-05	1.39E-03
hemopoiesis (GO:0030097)	36	15.66	2.30	8.68E-06	8.89E-04
chemotaxis (GO:0006935)	34	14.93	2.28	1.85E-05	1.72E-03
positive regulation of cellular protein localization (GO:1903829)	21	9.22	2.28	7.25E-04	3.20E-02
taxis (GO:0042330)	34	14.98	2.27	1.96E-05	1.79E-03
regulation of MAP kinase activity (GO:0043405)	21	9.33	2.25	8.20E-04	3.56E-02
regulation of cell adhesion (GO:0030155)	44	19.59	2.25	1.53E-06	1.92E-04
regulation of response to biotic stimulus (GO:0002831)	25	11.13	2.25	3.60E-04	1.83E-02
positive regulation of secretion by cell (GO:1903532)	26	11.60	2.24	2.65E-04	1.44E-02
regulation of cell migration (GO:0030334)	53	23.99	2.21	1.93E-07	2.98E-05
regulation of response to external stimulus (GO:0032101)	72	32.64	2.21	1.20E-09	2.78E-07
response to external stimulus (GO:0009605)	149	67.59	2.20	4.61E-20	3.68E-17
negative regulation of transport (GO:0051051)	30	13.64	2.20	1.01E-04	6.70E-03
regulation of peptide secretion (GO:0002791)	31	14.16	2.19	1.18E-04	7.62E-03
regulation of DNA-binding transcription factor activity (GO:0051090)	26	11.90	2.19	3.41E-04	1.76E-02
regulation of defense response (GO:0031347)	51	23.36	2.18	4.55E-07	6.54E-05
regulation of anatomical structure size (GO:0090066)	31	14.22	2.18	1.21E-04	7.74E-03
regulation of cell motility (GO:2000145)	56	25.71	2.18	1.49E-07	2.38E-05
positive regulation of cell migration (GO:0030335)	31	14.24	2.18	1.23E-04	7.83E-03
regulation of peptidase activity (GO:0052547)	27	12.42	2.17	2.71E-04	1.46E-02

regulation of protein secretion (GO:0050708)	29	13.34	2.17	2.35E-04	1.31E-02
regulation of peptide transport (GO:0090087)	47	21.69	2.17	1.68E-06	2.07E-04
regulation of protein transport (GO:0051223)	45	20.82	2.16	3.13E-06	3.59E-04
cellular cation homeostasis (GO:0030003)	38	17.63	2.16	2.71E-05	2.37E-03
regulation of endopeptidase activity (GO:0052548)	25	11.60	2.16	5.25E-04	2.49E-02
immune system process (GO:0002376)	163	75.81	2.15	4.22E-21	4.49E-18
negative regulation of response to external stimulus (GO:0032102)	33	15.42	2.14	8.11E-05	5.65E-03
regulation of locomotion (GO:0040012)	57	26.80	2.13	2.65E-07	3.99E-05
negative regulation of multicellular organismal process (GO:0051241)	71	33.40	2.13	6.13E-09	1.24E-06
cellular metal ion homeostasis (GO:0006875)	33	15.55	2.12	9.13E-05	6.20E-03
regulation of plasma membrane bounded cell projection organization (GO:0120035)	40	18.86	2.12	1.76E-05	1.65E-03
hematopoietic or lymphoid organ development (GO:0048534)	36	16.97	2.12	5.74E-05	4.28E-03
positive regulation of transport (GO:0051050)	59	27.86	2.12	1.58E-07	2.50E-05
response to nutrient levels (GO:0031667)	31	14.68	2.11	1.68E-04	1.01E-02
cellular ion homeostasis (GO:0006873)	38	18.01	2.11	3.35E-05	2.79E-03
regulation of vesicle-mediated transport (GO:0060627)	32	15.17	2.11	1.30E-04	8.09E-03
regulation of secretion (GO:0051046)	48	22.84	2.10	3.01E-06	3.50E-04
regulation of cell projection organization (GO:0031344)	40	19.10	2.09	2.16E-05	1.92E-03
regulation of establishment of protein localization (GO:0070201)	45	21.50	2.09	7.50E-06	7.77E-04
positive regulation of locomotion (GO:0040017)	32	15.34	2.09	2.20E-04	1.25E-02
positive regulation of cellular component movement (GO:0051272)	32	15.36	2.08	2.22E-04	1.26E-02
positive regulation of cell motility (GO:2000147)	31	14.93	2.08	3.07E-04	1.61E-02
modulation of chemical synaptic transmission (GO:0050804)	26	12.53	2.08	8.03E-04	3.51E-02
regulation of cellular protein localization (GO:1903827)	32	15.42	2.08	2.27E-04	1.28E-02
regulation of trans-synaptic signaling (GO:0099177)	26	12.55	2.07	8.17E-04	3.55E-02
cellular divalent inorganic cation homeostasis (GO:0072503)	27	13.04	2.07	6.22E-04	2.83E-02
regulation of cellular component movement (GO:0051270)	58	28.03	2.07	4.83E-07	6.82E-05
positive regulation of cellular component biogenesis (GO:0044089)	30	14.57	2.06	4.36E-04	2.16E-02
divalent inorganic cation homeostasis (GO:0072507)	28	13.62	2.06	5.10E-04	2.44E-02
regulation of immune system process (GO:0002682)	96	46.88	2.05	6.58E-11	2.02E-08
regulation of protein localization (GO:0032880)	59	29.20	2.02	8.13E-07	1.08E-04
regulation of secretion by cell (GO:1903530)	43	21.37	2.01	3.28E-05	2.74E-03

positive regulation of protein metabolic process (GO:0051247)	95	47.37	2.01	2.33E-10	6.51E-08
immune system development (GO:0002520)	36	17.98	2.00	1.57E-04	9.49E-03
cation homeostasis (GO:0055080)	39	19.48	2.00	7.19E-05	5.12E-03
positive regulation of hydrolase activity (GO:0051345)	42	21.01	2.00	4.65E-05	3.64E-03
negative regulation of phosphate metabolic process (GO:0045936)	31	15.53	2.00	4.29E-04	2.13E-02
negative regulation of phosphorus metabolic process (GO:0010563)	31	15.55	1.99	4.37E-04	2.16E-02
response to extracellular stimulus (GO:0009991)	31	15.58	1.99	4.46E-04	2.19E-02
regulation of immune response (GO:0050776)	62	31.19	1.99	6.85E-07	9.35E-05
cellular chemical homeostasis (GO:0055082)	41	20.79	1.97	7.10E-05	5.08E-03
inorganic ion homeostasis (GO:0098771)	39	19.81	1.97	1.23E-04	7.81E-03
metal ion homeostasis (GO:0055065)	34	17.30	1.97	3.22E-04	1.67E-02
positive regulation of multicellular organismal process (GO:0051240)	95	48.66	1.95	1.03E-09	2.43E-07
positive regulation of intracellular signal transduction (GO:1902533)	55	28.22	1.95	6.30E-06	6.66E-04
negative regulation of cell differentiation (GO:0045596)	38	19.51	1.95	1.78E-04	1.07E-02
apoptotic process (GO:0006915)	48	24.81	1.94	2.98E-05	2.56E-03
establishment of localization in cell (GO:0051649)	126	65.14	1.93	1.15E-12	4.95E-10
positive regulation of immune system process (GO:0002684)	57	29.77	1.91	6.90E-06	7.20E-04
positive regulation of immune response (GO:0050778)	39	20.44	1.91	2.48E-04	1.36E-02
positive regulation of apoptotic process (GO:0043065)	34	17.82	1.91	6.10E-04	2.78E-02
interspecies interaction between organisms (GO:0044419)	104	54.61	1.90	4.73E-10	1.24E-07
cellular homeostasis (GO:0019725)	48	25.21	1.90	3.65E-05	2.98E-03
regulation of cell activation (GO:0050865)	34	17.90	1.90	6.31E-04	2.86E-02
positive regulation of cellular protein metabolic process (GO:0032270)	84	44.34	1.89	4.13E-08	7.49E-06
positive regulation of programmed cell death (GO:0043068)	34	17.98	1.89	6.55E-04	2.95E-02
negative regulation of cellular component organization (GO:0051129)	37	19.59	1.89	4.56E-04	2.23E-02
regulation of transport (GO:0051049)	98	52.01	1.88	3.28E-09	6.88E-07
cell migration (GO:0016477)	49	26.12	1.88	4.75E-05	3.66E-03
regulation of growth (GO:0040008)	35	18.67	1.88	5.60E-04	2.60E-02
regulation of response to stress (GO:0080134)	83	44.34	1.87	6.93E-08	1.19E-05
locomotion (GO:0040011)	67	35.86	1.87	2.23E-06	2.64E-04
positive regulation of cellular component organization (GO:0051130)	62	33.29	1.86	5.74E-06	6.19E-04
regulation of MAPK cascade (GO:0043408)	37	19.89	1.86	5.10E-04	2.44E-02

negative regulation of developmental process (GO:0051093)	49	26.52	1.85	8.23E-05	5.71E-03
response to oxygen-containing compound (GO:1901700)	81	43.91	1.84	2.15E-07	3.30E-05
ion homeostasis (GO:0050801)	40	21.78	1.84	3.87E-04	1.96E-02
positive regulation of catalytic activity (GO:0043085)	74	40.31	1.84	9.80E-07	1.27E-04
regulation of localization (GO:0032879)	142	77.45	1.83	1.29E-12	5.43E-10
negative regulation of catalytic activity (GO:0043086)	41	22.38	1.83	3.13E-04	1.64E-02
regulation of cellular component biogenesis (GO:0044087)	48	26.33	1.82	1.19E-04	7.61E-03
regulation of intracellular signal transduction (GO:1902531)	89	49.09	1.81	1.11E-07	1.81E-05
regulation of multicellular organismal development (GO:2000026)	103	56.82	1.81	6.15E-09	1.23E-06
membrane organization (GO:0061024)	44	24.29	1.81	2.38E-04	1.33E-02
cell death (GO:0008219)	53	29.36	1.81	5.62E-05	4.23E-03
positive regulation of cell death (GO:0010942)	35	19.40	1.80	1.09E-03	4.48E-02
localization of cell (GO:0051674)	53	29.42	1.80	5.77E-05	4.29E-03
cell motility (GO:0048870)	53	29.42	1.80	5.77E-05	4.27E-03
positive regulation of molecular function (GO:0044093)	90	49.99	1.80	9.85E-08	1.62E-05
programmed cell death (GO:0012501)	51	28.38	1.80	9.41E-05	6.29E-03
negative regulation of cell communication (GO:0010648)	70	38.97	1.80	3.72E-06	4.18E-04
positive regulation of phosphorus metabolic process (GO:0010562)	57	31.74	1.80	3.21E-05	2.71E-03
positive regulation of phosphate metabolic process (GO:0045937)	57	31.74	1.80	3.21E-05	2.70E-03
positive regulation of protein modification process (GO:0031401)	61	34.03	1.79	1.75E-05	1.64E-03
positive regulation of cell communication (GO:0010647)	95	53.02	1.79	5.98E-08	1.06E-05
negative regulation of signaling (GO:0023057)	70	39.08	1.79	3.92E-06	4.32E-04
regulation of neurogenesis (GO:0050767)	41	22.92	1.79	5.41E-04	2.52E-02
regulation of cellular localization (GO:0060341)	50	28.00	1.79	1.28E-04	7.99E-03
positive regulation of signaling (GO:0023056)	95	53.24	1.78	6.46E-08	1.12E-05
intracellular signal transduction (GO:0035556)	83	46.53	1.78	4.81E-07	6.85E-05
cell surface receptor signaling pathway (GO:0007166)	121	68.03	1.78	6.78E-10	1.66E-07
negative regulation of response to stimulus (GO:0048585)	89	50.10	1.78	2.24E-07	3.41E-05
cellular response to oxygen-containing compound (GO:1901701)	52	29.47	1.76	1.26E-04	7.92E-03
response to stress (GO:0006950)	175	99.20	1.76	2.07E-14	1.00E-11
regulation of catalytic activity (GO:0050790)	115	65.30	1.76	4.11E-09	8.52E-07
positive regulation of response to stimulus (GO:0048584)	118	67.08	1.76	1.97E-09	4.32E-07
regulation of multicellular organismal process (GO:0051239)	154	87.62	1.76	2.20E-12	8.55E-10
regulation of anatomical structure morphogenesis (GO:0022603)	53	30.21	1.75	1.49E-04	9.08E-03
regulation of hydrolase activity (GO:0051336)	62	35.34	1.75	3.45E-05	2.86E-03

negative regulation of programmed cell death (GO:0043069)	44	25.13	1.75	4.62E-04	2.24E-02
cellular localization (GO:0051641)	143	81.84	1.75	3.56E-11	1.14E-08
regulation of phosphate metabolic process (GO:0019220)	87	49.83	1.75	6.85E-07	9.42E-05
regulation of phosphorus metabolic process (GO:0051174)	87	49.86	1.74	6.91E-07	9.34E-05
positive regulation of phosphorylation (GO:0042327)	52	29.83	1.74	1.99E-04	1.16E-02
response to drug (GO:0042493)	49	28.11	1.74	2.83E-04	1.51E-02
negative regulation of apoptotic process (GO:0043066)	43	24.67	1.74	5.99E-04	2.75E-02
response to organic substance (GO:0010033)	144	82.77	1.74	4.40E-11	1.38E-08
regulation of cell differentiation (GO:0045595)	88	50.68	1.74	8.28E-07	1.09E-04
regulation of kinase activity (GO:0043549)	43	24.83	1.73	8.65E-04	3.69E-02
positive regulation of protein phosphorylation (GO:0001934)	48	27.73	1.73	3.79E-04	1.93E-02
cellular response to organic substance (GO:0071310)	113	65.49	1.73	1.48E-08	2.84E-06
positive regulation of cell differentiation (GO:0045597)	47	27.32	1.72	5.02E-04	2.41E-02
cellular response to chemical stimulus (GO:0070887)	139	80.80	1.72	2.19E-10	6.23E-08
regulation of cell development (GO:0060284)	45	26.36	1.71	8.44E-04	3.63E-02
regulation of protein metabolic process (GO:0051246)	132	77.64	1.70	1.29E-09	2.95E-07
positive regulation of developmental process (GO:0051094)	65	38.23	1.70	4.70E-05	3.66E-03
response to organic cyclic compound (GO:0014070)	43	25.38	1.69	1.04E-03	4.27E-02
regulation of nervous system development (GO:0051960)	44	25.98	1.69	1.12E-03	4.55E-02
negative regulation of signal transduction (GO:0009968)	59	35.04	1.68	1.46E-04	8.97E-03
transport (GO:0006810)	209	124.16	1.68	1.67E-15	1.07E-12
regulation of protein modification process (GO:0031399)	86	51.17	1.68	4.08E-06	4.46E-04
regulation of transferase activity (GO:0051338)	47	28.03	1.68	8.46E-04	3.63E-02
positive regulation of transcription by RNA polymerase II (GO:0045944)	56	33.40	1.68	3.00E-04	1.58E-02
positive regulation of signal transduction (GO:0009967)	78	46.64	1.67	1.43E-05	1.39E-03
negative regulation of cell death (GO:0060548)	46	27.64	1.66	1.12E-03	4.56E-02
regulation of cellular protein metabolic process (GO:0032268)	121	72.81	1.66	3.10E-08	5.75E-06
establishment of localization (GO:0051234)	212	127.71	1.66	4.71E-15	2.59E-12
regulation of developmental process (GO:0050793)	120	72.34	1.66	4.07E-08	7.46E-06
nitrogen compound transport (GO:0071705)	85	51.28	1.66	6.24E-06	6.68E-04
chemical homeostasis (GO:0048878)	51	30.86	1.65	7.32E-04	3.22E-02
positive regulation of nitrogen compound metabolic process (GO:0051173)	146	88.39	1.65	9.81E-10	2.34E-07
establishment of protein localization (GO:0045184)	73	44.21	1.65	3.76E-05	3.05E-03
protein transport (GO:0015031)	69	41.89	1.65	6.88E-05	4.94E-03
regulation of cell population proliferation (GO:0042127)	74	45.19	1.64	4.46E-05	3.51E-03
regulation of protein phosphorylation (GO:0001932)	65	39.87	1.63	1.80E-04	1.07E-02
regulation of cell communication (GO:0010646)	162	99.50	1.63	2.35E-10	6.46E-08

negative regulation of protein metabolic process (GO:0051248)	50	30.73	1.63	1.03E-03	4.24E-02
amide transport (GO:0042886)	71	43.66	1.63	9.19E-05	6.19E-03
regulation of programmed cell death (GO:0043067)	70	43.09	1.62	8.73E-05	5.96E-03
regulation of apoptotic process (GO:0042981)	69	42.49	1.62	1.07E-04	7.05E-03
regulation of molecular function (GO:0065009)	136	83.86	1.62	1.68E-08	3.15E-06
cell-cell signaling (GO:0007267)	52	32.06	1.62	9.19E-04	3.87E-02
regulation of signaling (GO:0023051)	163	100.59	1.62	2.80E-10	7.59E-08
positive regulation of macromolecule metabolic process (GO:0010604)	151	93.25	1.62	1.60E-09	3.59E-07
regulation of phosphorylation (GO:0042325)	72	44.48	1.62	7.84E-05	5.52E-03
regulation of cellular component organization (GO:0051128)	107	66.26	1.61	9.70E-07	1.27E-04
peptide transport (GO:0015833)	69	42.73	1.61	1.52E-04	9.25E-03
positive regulation of cellular metabolic process (GO:0031325)	150	92.92	1.61	2.20E-09	4.76E-07
regulation of response to stimulus (GO:0048583)	193	120.15	1.61	3.73E-12	1.38E-09
organic substance transport (GO:0071702)	97	60.96	1.59	7.99E-06	8.23E-04
regulation of cell death (GO:0010941)	74	46.55	1.59	1.10E-04	7.14E-03
positive regulation of metabolic process (GO:0009893)	160	101.41	1.58	2.70E-09	5.74E-07
positive regulation of cellular process (GO:0048522)	246	156.23	1.57	1.95E-15	1.20E-12
movement of cell or subcellular component (GO:0006928)	67	42.84	1.56	4.03E-04	2.02E-02
neurogenesis (GO:0022008)	71	45.55	1.56	3.19E-04	1.66E-02
positive regulation of biosynthetic process (GO:0009891)	87	56.19	1.55	6.72E-05	4.85E-03
cell communication (GO:0007154)	237	153.09	1.55	8.54E-14	3.79E-11
localization (GO:0051179)	244	158.33	1.54	3.68E-14	1.73E-11
response to chemical (GO:0042221)	190	123.35	1.54	2.13E-10	6.18E-08
positive regulation of cellular biosynthetic process (GO:0031328)	85	55.31	1.54	1.09E-04	7.14E-03
positive regulation of macromolecule biosynthetic process (GO:0010557)	81	52.78	1.53	1.89E-04	1.11E-02
nervous system development (GO:0007399)	100	65.36	1.53	2.71E-05	2.36E-03
positive regulation of transcription, DNA-templated (GO:0045893)	65	42.57	1.53	9.45E-04	3.97E-02
signaling (GO:0023052)	229	150.20	1.52	1.62E-12	6.63E-10
generation of neurons (GO:0048699)	65	42.71	1.52	9.72E-04	4.06E-02
positive regulation of biological process (GO:0048518)	260	171.32	1.52	1.54E-14	7.70E-12
signal transduction (GO:0007165)	213	140.43	1.52	3.49E-11	1.14E-08
positive regulation of gene expression (GO:0010628)	84	55.61	1.51	2.03E-04	1.17E-02
regulation of signal transduction (GO:0009966)	130	86.23	1.51	2.13E-06	2.55E-04
positive regulation of nucleic acid-templated transcription (GO:1903508)	68	45.24	1.50	1.00E-03	4.17E-02
positive regulation of RNA biosynthetic process (GO:1902680)	68	45.27	1.50	1.01E-03	4.18E-02
protein localization (GO:0008104)	90	60.25	1.49	1.52E-04	9.23E-03

macromolecule localization (GO:0033036)	103	69.56	1.48	7.68E-05	5.42E-03
regulation of biological quality (GO:0065008)	167	112.81	1.48	9.62E-08	1.60E-05
cellular response to stimulus (GO:0051716)	264	185.59	1.42	1.72E-11	5.96E-09
cell differentiation (GO:0030154)	144	102.17	1.41	1.60E-05	1.52E-03
cellular developmental process (GO:0048869)	146	103.64	1.41	1.39E-05	1.36E-03
response to stimulus (GO:0050896)	329	233.92	1.41	3.49E-15	2.06E-12
developmental process (GO:0032502)	221	161.50	1.37	1.26E-07	2.03E-05
negative regulation of biological process (GO:0048519)	206	151.21	1.36	7.12E-07	9.55E-05
anatomical structure development (GO:0048856)	203	149.24	1.36	1.05E-06	1.35E-04
negative regulation of cellular process (GO:0048523)	180	133.39	1.35	1.27E-05	1.26E-03
system development (GO:0048731)	163	121.71	1.34	6.64E-05	4.81E-03
multicellular organism development (GO:0007275)	182	138.63	1.31	4.99E-05	3.81E-03
regulation of cellular metabolic process (GO:0031323)	219	171.48	1.28	3.05E-05	2.61E-03
organonitrogen compound metabolic process (GO:1901564)	186	146.54	1.27	2.71E-04	1.47E-02
regulation of primary metabolic process (GO:0080090)	210	165.83	1.27	8.61E-05	5.89E-03
multicellular organismal process (GO:0032501)	243	192.03	1.27	1.29E-05	1.27E-03
regulation of nitrogen compound metabolic process (GO:0051171)	202	160.76	1.26	2.32E-04	1.31E-02
regulation of macromolecule metabolic process (GO:0060255)	210	169.60	1.24	3.49E-04	1.80E-02
biological regulation (GO:0065007)	423	341.79	1.24	1.91E-12	7.61E-10
regulation of cellular process (GO:0050794)	381	308.04	1.24	8.96E-10	2.17E-07
regulation of biological process (GO:0050789)	397	322.69	1.23	3.01E-10	8.00E-08
regulation of metabolic process (GO:0019222)	227	184.80	1.23	2.40E-04	1.33E-02
cellular process (GO:0009987)	491	421.83	1.16	4.31E-12	1.56E-09
biological_process (GO:0008150)	534	486.01	1.10	6.48E-10	1.64E-07
Unclassified (UNCLASSIFIED)	35	82.99	.42	6.48E-10	1.62E-07
sensory perception of chemical stimulus (GO:0007606)	3	14.74	.20	5.84E-04	2.69E-02
ribonucleoprotein complex biogenesis (GO:0022613)	1	12.42	.08	9.51E-05	6.32E-03

Table A8: All GO-terms from yellow4 module. The number of related genes in the reference list (ref), number of relevant genes from the network (genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process	#ref	#genes	exp	+/-	FE	raw P-value	FDR
oxygen transport (GO:0015671)	14	3	.02	+	>100	2.19E-06	9.71E-04
membrane disruption in other organism (GO:0051673)	10	2	.02	+	>100	1.49E-04	3.00E-02
gas transport (GO:0015669)	19	3	.03	+	>100	4.93E-06	2.02E-03
iron coordination entity transport (GO:1901678)	13	2	.02	+	>100	2.36E-04	4.28E-02

response to herbicide (GO:0009635)	13	2	.02	+	>100	2.36E-04	4.23E-02
hydrogen peroxide catabolic process (GO:0042744)	28	4	.04	+	93.08	1.58E-07	9.03E-05
hydrogen peroxide metabolic process (GO:0042743)	39	4	.06	+	66.83	5.37E-07	2.96E-04
bicarbonate transport (GO:0015701)	42	4	.06	+	62.06	7.08E-07	3.77E-04
actin filament capping (GO:0051693)	33	3	.05	+	59.24	2.25E-05	6.79E-03
negative regulation of actin filament depolymerization (GO:0030835)	36	3	.06	+	54.30	2.88E-05	8.06E-03
antibiotic catabolic process (GO:0017001)	54	4	.08	+	48.27	1.82E-06	8.53E-04
antibacterial humoral response (GO:0019731)	45	3	.07	+	43.44	5.39E-05	1.39E-02
defense response to fungus (GO:0050832)	46	3	.07	+	42.50	5.74E-05	1.43E-02
regulation of actin filament depolymerization (GO:0030834)	51	3	.08	+	38.33	7.69E-05	1.80E-02
negative regulation of protein polymerization (GO:0032272)	72	4	.11	+	36.20	5.39E-06	2.10E-03
cofactor catabolic process (GO:0051187)	72	4	.11	+	36.20	5.39E-06	2.05E-03
negative regulation of actin filament polymerization (GO:0030837)	58	3	.09	+	33.70	1.11E-04	2.42E-02
response to fungus (GO:0009620)	60	3	.09	+	32.58	1.22E-04	2.56E-02
cellular iron ion homeostasis (GO:0006879)	66	3	.10	+	29.62	1.60E-04	3.15E-02
negative regulation of protein depolymerization (GO:1901880)	66	3	.10	+	29.62	1.60E-04	3.11E-02
cellular oxidant detoxification (GO:0098869)	94	4	.14	+	27.73	1.48E-05	5.03E-03
tetrapyrrole metabolic process (GO:0033013)	71	3	.11	+	27.53	1.97E-04	3.65E-02
negative regulation of protein-containing complex disassembly (GO:0043242)	76	3	.12	+	25.72	2.39E-04	4.23E-02
antibiotic metabolic process (GO:0016999)	102	4	.16	+	25.55	2.02E-05	6.58E-03
cellular detoxification (GO:1990748)	105	4	.16	+	24.82	2.26E-05	6.67E-03
erythrocyte differentiation (GO:0030218)	81	3	.12	+	24.13	2.86E-04	4.97E-02
reactive oxygen species metabolic process (GO:0072593)	113	4	.17	+	23.07	2.98E-05	8.06E-03
response to hydrogen peroxide (GO:0042542)	119	4	.18	+	21.90	3.63E-05	9.65E-03
drug transport (GO:0015893)	151	5	.23	+	21.58	3.73E-06	1.61E-03
detoxification (GO:0098754)	133	4	.20	+	19.60	5.53E-05	1.40E-02
negative regulation of protein-containing complex assembly (GO:0031333)	137	4	.21	+	19.02	6.18E-05	1.52E-02
drug catabolic process (GO:0042737)	141	4	.22	+	18.48	6.89E-05	1.67E-02
antimicrobial humoral response (GO:0019730)	142	4	.22	+	18.35	7.08E-05	1.69E-02
negative regulation of supramolecular fiber organization (GO:1902904)	146	4	.22	+	17.85	7.86E-05	1.82E-02
cellular response to toxic substance (GO:0097237)	220	6	.34	+	17.77	1.08E-06	5.37E-04
negative regulation of cytoskeleton organization (GO:0051494)	153	4	.23	+	17.04	9.38E-05	2.11E-02

response to reactive oxygen species (GO:0000302)	196	5	.30	+	16.62	1.28E-05	4.45E-03
neutrophil degranulation (GO:0043312)	482	12	.74	+	16.22	3.95E-12	3.15E-08
neutrophil activation involved in immune response (GO:0002283)	486	12	.75	+	16.09	4.34E-12	2.31E-08
neutrophil mediated immunity (GO:0002446)	493	12	.76	+	15.86	5.11E-12	2.04E-08
neutrophil activation (GO:0042119)	495	12	.76	+	15.80	5.35E-12	1.71E-08
granulocyte activation (GO:0036230)	500	12	.77	+	15.64	6.00E-12	1.60E-08
leukocyte degranulation (GO:0043299)	504	12	.77	+	15.51	6.57E-12	1.50E-08
myeloid leukocyte mediated immunity (GO:0002444)	514	12	.79	+	15.21	8.22E-12	1.64E-08
myeloid cell activation involved in immune response (GO:0002275)	519	12	.80	+	15.07	9.18E-12	1.63E-08
myeloid leukocyte activation (GO:0002274)	583	13	.89	+	14.53	1.55E-12	2.47E-08
cellular response to lipopolysaccharide (GO:0071222)	191	4	.29	+	13.65	2.16E-04	3.96E-02
response to molecule of bacterial origin (GO:0002237)	340	7	.52	+	13.42	7.75E-07	3.99E-04
cellular response to molecule of bacterial origin (GO:0071219)	201	4	.31	+	12.97	2.61E-04	4.59E-02
leukocyte activation involved in immune response (GO:0002366)	619	12	.95	+	12.63	6.78E-11	7.73E-08
cell activation involved in immune response (GO:0002263)	623	12	.96	+	12.55	7.30E-11	7.76E-08
regulated exocytosis (GO:0045055)	696	13	1.07	+	12.17	1.37E-11	2.18E-08
response to lipopolysaccharide (GO:0032496)	323	6	.50	+	12.10	9.37E-06	3.48E-03
exocytosis (GO:0006887)	788	13	1.21	+	10.75	6.25E-11	7.67E-08
cofactor metabolic process (GO:0051186)	438	7	.67	+	10.41	4.04E-06	1.70E-03
response to antibiotic (GO:0046677)	315	5	.48	+	10.34	1.17E-04	2.53E-02
leukocyte mediated immunity (GO:0002443)	762	12	1.17	+	10.26	7.05E-10	5.62E-07
response to oxidative stress (GO:0006979)	393	6	.60	+	9.95	2.79E-05	7.95E-03
defense response to bacterium (GO:0042742)	339	5	.52	+	9.61	1.65E-04	3.16E-02
leukocyte activation (GO:0045321)	921	13	1.41	+	9.20	4.15E-10	3.90E-07
response to toxic substance (GO:0009636)	520	7	.80	+	8.77	1.22E-05	4.32E-03
cell activation (GO:0001775)	1068	14	1.64	+	8.54	1.81E-10	1.81E-07
secretion by cell (GO:0032940)	1002	13	1.54	+	8.45	1.15E-09	8.33E-07
export from cell (GO:0140352)	1049	13	1.61	+	8.08	1.99E-09	1.38E-06
drug metabolic process (GO:0017144)	493	6	.76	+	7.93	9.65E-05	2.14E-02
secretion (GO:0046903)	1119	13	1.72	+	7.57	4.31E-09	2.87E-06
response to bacterium (GO:0009617)	709	8	1.09	+	7.35	9.40E-06	3.41E-03
immune effector process (GO:0002252)	1083	12	1.66	+	7.22	3.46E-08	2.05E-05
response to inorganic substance (GO:0010035)	545	6	.84	+	7.17	1.66E-04	3.15E-02
response to drug (GO:0042493)	1030	10	1.58	+	6.33	2.12E-06	9.69E-04
innate immune response (GO:0045087)	777	7	1.19	+	5.87	1.51E-04	3.02E-02
vesicle-mediated transport (GO:0016192)	1944	17	2.98	+	5.70	4.59E-10	4.07E-07

defense response to other organism (GO:0098542)	970	8	1.49	+	5.37	8.66E-05	1.97E-02
establishment of localization in cell (GO:0051649)	2387	18	3.66	+	4.91	1.14E-09	8.66E-07
response to other organism (GO:0051707)	1345	10	2.06	+	4.84	2.21E-05	6.91E-03
response to external biotic stimulus (GO:0043207)	1347	10	2.07	+	4.84	2.24E-05	6.86E-03
cellular response to oxygen-containing compound (GO:1901701)	1080	8	1.66	+	4.83	1.81E-04	3.41E-02
response to biotic stimulus (GO:0009607)	1379	10	2.12	+	4.73	2.74E-05	7.94E-03
immune response (GO:0006955)	1872	13	2.87	+	4.52	1.70E-06	8.24E-04
response to oxygen-containing compound (GO:1901700)	1609	11	2.47	+	4.45	1.66E-05	5.51E-03
ion transport (GO:0006811)	1342	9	2.06	+	4.37	1.39E-04	2.85E-02
cellular localization (GO:0051641)	2999	20	4.60	+	4.35	5.83E-10	4.90E-07
immune system process (GO:0002376)	2778	18	4.26	+	4.22	1.31E-08	8.03E-06
transport (GO:0006810)	4550	25	6.98	+	3.58	2.02E-11	2.93E-08
establishment of localization (GO:0051234)	4680	25	7.18	+	3.48	3.86E-11	5.14E-08
response to external stimulus (GO:0009605)	2477	13	3.80	+	3.42	3.67E-05	9.60E-03
localization (GO:0051179)	5802	25	8.90	+	2.81	5.10E-09	3.26E-06
response to stress (GO:0006950)	3635	15	5.58	+	2.69	1.18E-04	2.52E-02
regulation of biological quality (GO:0065008)	4134	17	6.34	+	2.68	2.96E-05	8.15E-03
response to chemical (GO:0042221)	4520	18	6.94	+	2.59	2.19E-05	6.99E-03
response to stimulus (GO:0050896)	8572	26	13.16	+	1.98	5.13E-06	2.05E-03
cellular process (GO:0009987)	15458	32	23.72	+	1.35	1.31E-04	2.71E-02

Table A9: All GO-terms from darkolivegreen module. The number of related genes in the reference list (ref), number of relevant genes from the network (genes), the expected number genes found, over/under-representation (+/-), fold enrichment (FE), the raw p-value and the false discovery rate (FDR) are provided for each term.

GO biological process complete	#ref	#genes	exp	+/-	FE	raw P-value	FDR
regulation of ribonuclease activity (GO:0060700)	9	3	.05	+	55.16	4.47E-05	1.30E-02
dicarboxylic acid biosynthetic process (GO:0043650)	13	3	.08	+	38.19	1.12E-04	3.03E-02
positive regulation of defense response to virus by host (GO:0002230)	30	6	.18	+	33.10	7.00E-08	3.02E-05
type I interferon signaling pathway (GO:0060337)	67	13	.40	+	32.11	1.58E-15	2.80E-12
cellular response to type I interferon (GO:0071357)	67	13	.40	+	32.11	1.58E-15	2.52E-12
antigen processing and presentation of endogenous peptide antigen (GO:0002483)	21	4	.13	+	31.52	1.42E-05	4.64E-03

antigen processing and presentation of endogenous peptide antigen via MHC class I (GO:0019885)	21	4	.13	+	31.52	1.42E-05	4.54E-03
response to type I interferon (GO:0034340)	72	13	.44	+	29.88	3.61E-15	5.23E-12
regulation of defense response to virus by host (GO:0050691)	41	7	.25	+	28.25	1.43E-08	6.92E-06
response to interferon-beta (GO:0035456)	32	5	.19	+	25.86	2.70E-06	1.00E-03
antigen processing and presentation of endogenous antigen (GO:0019883)	28	4	.17	+	23.64	3.92E-05	1.16E-02
interferon-gamma-mediated signaling pathway (GO:0060333)	72	10	.44	+	22.98	6.32E-11	5.04E-08
negative regulation of viral genome replication (GO:0045071)	62	8	.37	+	21.35	9.36E-09	4.67E-06
defense response to virus (GO:0051607)	203	24	1.23	+	19.56	2.93E-23	4.67E-19
negative regulation of viral life cycle (GO:1903901)	85	10	.51	+	19.47	2.80E-10	1.86E-07
negative regulation of innate immune response (GO:0045824)	60	7	.36	+	19.31	1.54E-07	6.46E-05
regulation of defense response to virus (GO:0050688)	79	9	.48	+	18.85	2.92E-09	1.55E-06
negative regulation of viral process (GO:0048525)	104	11	.63	+	17.50	9.89E-11	7.52E-08
regulation of viral genome replication (GO:0045069)	99	10	.60	+	16.72	1.10E-09	6.50E-07
response to virus (GO:0009615)	288	25	1.74	+	14.36	3.54E-21	2.83E-17
regulation of viral life cycle (GO:1903900)	147	12	.89	+	13.51	2.19E-10	1.59E-07
negative regulation of response to biotic stimulus (GO:0002832)	95	7	.57	+	12.19	2.75E-06	9.96E-04
antigen processing and presentation of peptide antigen via MHC class I (GO:0002474)	103	7	.62	+	11.25	4.55E-06	1.58E-03
cellular response to interferon-gamma (GO:0071346)	164	11	.99	+	11.10	8.80E-09	4.53E-06
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent (GO:0002479)	76	5	.46	+	10.89	1.28E-04	3.42E-02
response to interferon-gamma (GO:0034341)	186	12	1.12	+	10.68	2.72E-09	1.50E-06
antigen processing and presentation of exogenous peptide antigen via MHC class I (GO:0042590)	80	5	.48	+	10.34	1.61E-04	4.15E-02
regulation of viral process (GO:0050792)	209	13	1.26	+	10.29	8.75E-10	5.37E-07
regulation of symbiosis, encompassing mutualism through parasitism (GO:0043903)	225	13	1.36	+	9.56	2.04E-09	1.17E-06
regulation of innate immune response (GO:0045088)	304	15	1.84	+	8.17	8.72E-10	5.57E-07
regulation of response to biotic stimulus (GO:0002831)	408	20	2.47	+	8.11	1.18E-12	1.17E-09

regulation of type I interferon production (GO:0032479)	130	6	.79	+	7.64	1.73E-04	4.32E-02
innate immune response (GO:0045087)	777	31	4.70	+	6.60	7.04E-17	3.75E-13
positive regulation of innate immune response (GO:0045089)	212	8	1.28	+	6.24	5.60E-05	1.60E-02
cytokine-mediated signaling pathway (GO:0019221)	694	25	4.19	+	5.96	9.74E-13	1.04E-09
positive regulation of response to biotic stimulus (GO:0002833)	251	9	1.52	+	5.93	2.80E-05	8.60E-03
defense response to other organism (GO:0098542)	970	34	5.86	+	5.80	7.18E-17	2.87E-13
regulation of immune effector process (GO:0002697)	477	15	2.88	+	5.20	2.74E-07	1.09E-04
response to cytokine (GO:0034097)	1121	32	6.77	+	4.72	1.77E-13	2.02E-10
response to other organism (GO:0051707)	1345	38	8.13	+	4.68	6.84E-16	1.82E-12
response to external biotic stimulus (GO:0043207)	1347	38	8.14	+	4.67	7.17E-16	1.64E-12
cellular response to cytokine stimulus (GO:0071345)	1031	29	6.23	+	4.65	4.34E-12	3.85E-09
immune effector process (GO:0002252)	1083	30	6.54	+	4.58	2.49E-12	2.34E-09
response to biotic stimulus (GO:0009607)	1379	38	8.33	+	4.56	1.50E-15	3.00E-12
defense response (GO:0006952)	1369	37	8.27	+	4.47	7.15E-15	9.51E-12
positive regulation of defense response (GO:0031349)	370	10	2.24	+	4.47	1.01E-04	2.78E-02
positive regulation of cytokine production (GO:0001819)	462	11	2.79	+	3.94	1.35E-04	3.53E-02
regulation of defense response (GO:0031347)	856	20	5.17	+	3.87	2.71E-07	1.11E-04
interspecies interaction between organisms (GO:0044419)	2001	46	12.09	+	3.80	4.78E-16	1.52E-12
viral process (GO:0016032)	791	18	4.78	+	3.77	1.65E-06	6.28E-04
symbiotic process (GO:0044403)	883	20	5.34	+	3.75	4.38E-07	1.71E-04
immune response (GO:0006955)	1872	41	11.31	+	3.62	1.66E-13	2.04E-10
regulation of cytokine production (GO:0001817)	734	14	4.44	+	3.16	1.64E-04	4.15E-02
positive regulation of immune response (GO:0050778)	749	14	4.53	+	3.09	2.01E-04	4.93E-02
regulation of immune response (GO:0050776)	1143	20	6.91	+	2.90	1.98E-05	6.20E-03
response to external stimulus (GO:0009605)	2477	42	14.97	+	2.81	2.76E-10	1.92E-07
regulation of response to external stimulus (GO:0032101)	1196	20	7.23	+	2.77	3.73E-05	1.12E-02
immune system process (GO:0002376)	2778	46	16.79	+	2.74	5.67E-11	4.76E-08
cellular response to organic substance (GO:0071310)	2400	38	14.50	+	2.62	1.75E-08	8.22E-06
regulation of immune system process (GO:0002682)	1718	26	10.38	+	2.50	1.22E-05	4.05E-03
response to organic substance (GO:0010033)	3033	43	18.33	+	2.35	3.73E-08	1.70E-05

cell surface receptor signaling pathway (GO:0007166)	2493	33	15.06	+	2.19	1.22E-05	4.13E-03
response to stress (GO:0006950)	3635	48	21.97	+	2.19	4.53E-08	2.01E-05
cellular response to chemical stimulus (GO:0070887)	2961	38	17.89	+	2.12	4.44E-06	1.57E-03
response to chemical (GO:0042221)	4520	47	27.31	+	1.72	7.69E-05	2.15E-02

Camilla Mauroy

NTNU
Norwegian University of
Science and Technology