Scale-free networks in molecular biology:

Algorithms and Random Walks Analyses

SILVA KONINI

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Graduate Program in Mathematics and Statistics

York University

Toronto, Ontario

June 2017

Abstract

In this research, I focus on I) the mean field analysis of algorithms for scale-free networks in molecular biology and II) the analysis of biological networks using random walks and related algorithms.

I: Many systems in nature and society are described by means of complex networks. Research indicates that these complex networks exhibit scale-free properties. Studying the organizing principles of scale-free networks has significant implications in different fields including developing better drugs, defending the internet from hackers, halting the spread of deadly epidemics, developing marketing strategies, etc.

The sampling of scale-free networks in molecular biology is usually achieved by growing networks from a seed using recursive algorithms with elementary moves which include the addition and deletion of nodes and bonds. These algorithms include the Barabasi-Albert algorithm [1]. Later algorithms, such as the Duplication-Divergence algorithm [27, 29], the Solé algorithm [25, 26] and the iSite algorithm [15, 16], were inspired by biological processes underlying the evolution of protein networks, and the networks they produce differ essentially from networks grown by the Barabasi-Albert algorithm. The mean field analysis of these algorithms is reconsidered, and extended to variant and modified implementations of the algorithms.

II: The second part of this research focuses on improving biological networks using random walks and related algorithms.

I use different algorithms with the goal of finding highly connected hubs and clusters of proteins which are closely related to one another. This is done by building up protein-protein interaction networks and miRNA-gene interaction networks which are then subjected to the action of two algorithms.

The first algorithm used is the random walk with resistance algorithm. As an alternative, I am proposing solving the lattice laplacian on a network as a method to discover clusters of biologically related genes. These approaches seek to find ways of solving complex pathway membership problems in protein interaction databases. The clusters obtained provide more biological insight as opposed to a process of local pairwise comparison between interacting proteins. They may also predict new members in functional pathways or clusters. Underlying these algorithms are simulated biased random walks on the network for determining membership of proteins in given clusters.

Dedicated to the loving memory of my grandmother

Acknowledgments

I would like to express my special appreciation and thanks to my advisor, Professor E J Janse van Rensburg, for supporting my research all the way from when I first started the PhD program in the Department of Mathematics and Statistics, to the completion of this dissertation. It was a pleasure working with him and I am thankful for his excellent guidance provided throughout the past five years.

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Description
PINs	Protein-protein Interaction Networks
PPIs	Protein-protein Interactions
MGINs	MicroRNA-Gene Interaction Networks
MGIs	MicroRNA-Gene Interactions
mRNA	messenger RNA
miRNA	microRNA
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
miRTarBase	Experimentally validated microRNA-target interactions database
RWR	Random Walk with Resistance
LL	Lattice Laplacian
LLR	Lattice Laplacian with Resistance
WLL	Weighted Lattice Laplacian
WLLR	Weighted Lattice Laplacian with Resistance
DWLL	Double Weighted Lattice Laplacian
DWLLR	Double Weighted Lattice Laplacian with Resistance

Chapter 1

Introduction

This dissertation consists of two parts: A) Mean field analysis of algorithms for scale-free networks in molecular biology and B) Analysis of biological networks using random walks and related algorithms.

The contributions in this dissertation consist in modifying algorithms (modified Barabasi-Albert, modified Duplication-Divergence and iSite algorithms) to create denser networks and developing their mean field analyses in ways not done before in the literature. The contributions in the second part of the dissertation include introducing and implementing the lattice laplacian algorithm as an alternative to the random walk algorithm with the goal of improving biological networks, finding clusters of biologically related proteins and suggesting new targets of miRNAs.

A. Mean field analysis of algorithms for scale-free networks in molecular biology

Many systems in nature and society are described by means of complex networks [9]. Some of these systems include the cell [19], chemical reactions [17], the world wide web [6], social

interactions [7], etc. It is generally found that many systems, though different in nature, produce networks which are scale-free and exhibit similar properties [1, 3].

Knowledge about scale-free networks is important for understanding the spread of computer viruses, diseases and providing pharmaceutical researchers with new strategies for selecting drug targets, potentially leading to cures that would kill only harmful cells or bacteria by selectively targeting their hubs, while leaving healthy tissue unaffected [83].

The main property of scale-free networks is that their degree distribution decays as a power law [1, 4] – this shows that there is no characteristic scale for the degrees, which is why the networks are called scale-free. The average degree of a scale-free network offers little insight into the real topology of the network [3] since most nodes have degrees which are far away from the average degree of the network. Nodes of high degree are called *hubs* and though small in number for realistic networks, they are statistically overrepresented compared to the number of hubs in random networks. These hubs play an important role in dynamical processes which occur in scale-free networks.

For example, hubs are important in social networks. The fact that biological viruses spread in social networks, which in many cases appear to be scale-free, suggests that scientists should take a second look at the interplay of network topology and epidemics. In a scale-free network, the traditional public health approach of random immunization could easily fail because it would very likely neglect a number of the hubs [83].

A vaccination for measles, for instance, must reach 90% of the population to be effective. Instead of random immunizations, though, what if doctors targeted the hubs, or the most connected individuals? Research in scale-free networks indicates that this alternative

approach could be effective even if the immunizations reached only a small fraction of the overall population, provided that the fraction contained the hubs. Targeting hubs could be a solution for distribution of vaccines in countries and regions that do not have the resources to treat the entire population [83].

Dezso and Barabasi [84] studied the diffusion and spreading of viruses on a scale-free network, including both biological and computer based viruses. Methods designed to eradicate viruses usually aim at reducing the spreading rate of the virus. Even when the virus has a zero epidemic threshold, there is little guarantee that it will eradicate it [84]. This is because in scale-free networks, the hubs are in contact with a large number of nodes, and are therefore easily infected. Once the hubs are infected, they pass the virus to a significant fraction of nodes in the system. Thus even weakly infectious viruses can spread and prevail on a scale-free network. This negates what diffusion studies used to believe prior to the scale-free network theory that viruses whose spreading rate exceeds a critical threshold will persist, while those under the threshold will die out shortly. Dezso and Barabasi argued that hub-biased curing policies (curing with higher probability the hubs than the less connected nodes) can restore the epidemic threshold, which can stop the virus spreading [84].

Hubs are important in marketing techniques as well. Popular influencer marketing techniques (closely related to word-of-mouth or viral marketing), are based on the premise that a large number of people are connected to everyone else through a small number of hubs. Thus, identifying and focusing marketing activities around these hubs could increase the likelihood of initiating a cascading adoption of products or services- a type of social

epidemic [85].

Determining if a network is scale-free is important in understanding the system's behaviour, but it is not the only parameter which deserves attention [83]. The knowledge of a network's general topology is only one aspect in understanding the overall characteristics and behaviour of such systems. When it comes to social interactions, even though the networks that model these interactions are scale-free, ties between household members are much stronger than connections to casual acquaintances, so diseases are more likely to spread through such linkages. In the case of biological networks, the strength of the interaction between molecules does affect the network's dynamical behaviour as well. Therefore, the nature of the nodes and their interactions plays an important role in the behaviour of the systems.

Unexplained by previous network theories, hubs offer convincing proof that various complex systems have a strict architecture, ruled by fundamental laws, laws that appear to apply equally to cells, computers, languages and society [83]. Furthermore, these organizing principles have significant implications for developing better drugs, defending the internet from hackers, and halting the spread of deadly epidemics [84].

Scale-free networks also exhibit an unexpected degree of robustness – this is the property that such networks maintain their dynamic properties even when many nodes and bonds fail to transmit signals (suffer high failure rates) [9]. In the case of protein-protein interaction networks, it is difficult to disrupt the network: despite a high level of random mutations being introduced, the remaining proteins will continue to work together.

However, these networks remain vulnerable to failure of hub nodes, since these nodes

play a significant role in maintaining the network's connectivity. The presence of scale-free emerging properties in many real-world networks provides initial evidence that these self-organizing phenomena do not only depend on the characteristics of individual systems, but are general laws of evolving networks [83]. The responsible mechanisms for the emergence of scale-free networks are important in understanding why different systems converge to networks with similar architecture [4].

In the first part of this research I describe four evolutionary algorithms able to generate scale-free networks in molecular biology, with a focus on their mean field analysis. These algorithms are the Barabasi-Albert [1], Duplication-Divergence [27, 29], Solé [25, 26] and iSite [15, 16] algorithms.

The Duplication-Divergence, Solé and iSite algorithms were inspired by modelling networks in biological models of protein-protein interaction evolution, and all these algorithms are based in one way or another on two ideas: growth by preferential attachment [12], and growth and changes (*mutations*) in networks induced by the duplication, deletion or replacement of nodes or bonds (these are elementary moves which *mutate* the network by adding, deleting or moving some of its bonds or nodes).

Growth by preferential attachment is implemented by adding bonds preferentially to nodes of high degree. This increases the probability that a node will grow to be a hub in the network, and the resulting network has an increased probability that it will contain hubs [4]. The Barabasi-Albert algorithm uses preferential attachment to grow scale-free networks by attaching bonds to nodes with a probability which is proportional to the degrees of nodes [1]. A mean field analysis of the Barabasi-Albert algorithm was done in

reference [5].

The Duplication-Divergence algorithm [27, 29] generates scale-free networks by implementing elementary moves which mutate and grow the network. These are *duplication* (the duplication of existing nodes and bonds) and *divergence* (local changes made to existing bonds and nodes) elementary moves. These moves model processes which are thought to underlie the evolutionary mechanisms by which protein interaction networks evolve [25, 27, 29]: The *duplication* of genes is a mechanism which generates genes coding for new proteins during evolution and the *divergence* step is a model for the mutation of duplicated genes. After a duplication of a gene, two genes (one the *progenitor* gene, the other the *progeny* gene) coding for the same protein are obtained, and these mutate over time to drift away from one another in gene space, giving rise to modified proteins when translated by cellular machinery [25]. Biologically, the duplication step may result in a new protein interaction between two mutating copies of the same gene (this is called heteromerization), and the divergence step is a model of subfunctionalization (a process whereby interactions between proteins are lost).

Closely related to the Duplication-Divergence algorithm is the Solé algorithm [25, 26]. This algorithm grows networks by duplication of nodes, and mutates the network by rewiring it (this algorithm does not implement the heteromerization of the duplicated genes) [6]. It then implements a process of deleting some bonds on the duplicated nodes (modelling evolutionary changes due to subfunctionalization).

The iSite algorithm [15, 16] is a refinement of the Duplication-Divergence and Solé algorithms. This algorithm introduces more complex nodes which each contain *interaction sites*

as models of protein and protein complexes with localized interaction sites where the interactions with other proteins take place. These localized interaction sites are *iSites*. Such iSites may be involved in many interactions, but each interaction is related to only two iSites, one on each of the proteins involved. That is, iSites are models of the concept of domains on protein surfaces where the actual interactions take place between two proteins. The implementation of the algorithm on nodes containing iSites proceeds by duplication of nodes, and the mutation of iSites through subfunctionalization and heteromerization (namely, the subfunctionalization of iSites leading to loss of protein interactions, and heteromerization where new interactions are introduced between existing iSites). In this model the subfunctionalization is of iSites, leading to the loss of all bonds incident with the iSite (contrary to the situation in other algorithms, for example the Duplication-Divergence algorithm, where subfunctionalization leads to the loss of bonds, rather than nodes).

The first part of this dissertation is organized as follows. I first consider the general properties of scale-free networks, including their scaling and connectivity properties. These ideas are then applied to the analysis of particular algorithms. The Barabasi-Albert model is considered first together with a modified version of the algorithm, and a variant of the algorithm. Mean field theory for the modified and variant algorithms is developed, giving mean field values for the scaling exponent γ. These results are compared to numerical results obtained by generating networks using implementations of the algorithms.

The Duplication-Divergence algorithm and networks generated by it are considered next. The algorithm is also modified, and mean field theory is developed to find mean field values for the scaling exponent. The mean field predictions are then compared to

numerical results generated by implementing the algorithm and sampling networks.

A similar approach is followed for the Solé algorithm. However, in this model the degree distribution may not be integrable, and our results indicate that the networks generated by this algorithm are not scale-free. Modifying the distribution of degrees gives a testable scaling hypothesis for Solé networks, which is tested numerically by generating networks and examining their scaling, as well as by computing the connectivity of Solé networks and comparing it to the mean field predictions. This shows that the size of Solé networks of order n is $O(n^2)$, while the connectivity is $O(n)$ – this implies that Solé networks are dense.

Finally, the iSite algorithm is presented and examined developing a mean field approach to determine its scaling properties. The algorithm is also modified, and the resulting mean field results are tested numerically.

The first part of the research is completed in section 4.5, where our main results are briefly considered and reviewed.

B. Analysis of biological networks using random walks and related algorithms.

The second part of this dissertation focuses on improving the analysis of biological networks using random walks and related algorithms. Algorithms in molecular biology are used to predict new interactions between molecules, assign functions to previously unknown molecules, discover clusters of molecules which are closely related to one another, or predict new targets of a molecule. We study two important biological networks, protein-protein interaction networks (PINs) and microRNA-gene interaction networks (MGINs).

MicroRNAs (miRNAs) are small noncoding RNAs which are involved in post transcriptional regulation of gene expression usually through cleavage of messenger RNA [66].

MiRNAs are involved in diverse cellular functions such as development, differentiation, proliferation, apoptosis and metabolism. The input data used in this research was provided as microarray data by the Peng Lab [37] at York University. In addition we have downloaded miRNA-protein and protein-protein interaction data from the miRTarBase [39] and the STRING [38] databases.

Two network topology-based algorithms are presented with the goal of discovering pathways in protein-protein interaction networks and to suggest new targets of miRNAs in miRNA-gene interaction networks. The underlying idea in network based analysis is the discovery of cluster structures (of complexes and pathways) in PINs and MGINs. These structures give information on biologically related proteins and their functions. The key idea is that two proteins sharing higher "topological" similarities are likely interacting with each other and might belong to the same protein complex and cluster in the network. We test two algorithms. The first is to estimate similarities of the proteins in a network by using a *Random Walk with Resistance* (RWR) algorithm [36]. The second algorithm is to solve the *Lattice Laplacian with Resistance* (LLR) on a network, or its modifications, namely *Weighted Lattice Laplacian with Resistance* (WLLR) and *Double Weighted Lattice Laplacian with Resistance* (DWLLR) algorithms as an alternative to RWR.

Using data on upregulated and downregulated genes by the human miRNA hsa-miR-218-5p provided by the Peng lab [37], PINs were constructed by examining the environment of the genes in the STRING [38] database of protein-protein interactions. The structure of the networks was discovered with the RWR and LL algorithms, and visualized using Cytoscape [11]. Protein clusters are discovered by joining two proteins in a network when

there is a higher value of the Pearson correlation coefficient between their corresponding columns in the RWR probability matrix, or the LL solution matrix.

The algorithms do not produce identical networks, but both show similar networks of biologically related clustered proteins. Clusters can be examined individually by RWR and LL algorithms to predict novel protein functions and reaction pathways. The clusters correspond to complexes of functionally related proteins and in the case of the RWR a reaction pathway involving protein clusters is revealed. All these algorithms simulate biased random walks on the network. This places the algorithms in a class of random walk algorithms examined in the literature.

The focus of this research is on biological networks of the cell, hence in the next chapter we are giving a review of the cell structure and function.

Chapter 2

Cell structure and function

2.1 The structure of a cell

Cells are the basic unit of structure and function in living organisms [51]. Cells contain tiny compartments called organelles. Each organelle carries out a specific function within the cell. Cells contain hereditary material and are able to make copies of themselves [51]. Some of the structures and organelles of the cell are:

Cell membrane: The cell membrane is an outer layer that surrounds the cell and controls which substances enter or leave the cell. It is considered to be the security guard of the cell. It is sometimes called the plasma membrane or the cytoplasmic membrane. It separates the interior of the cell from the outside environment. The cell membrane is selectively permeable to ions and organic molecules. The basic function of the cell membrane is to protect the cell from its surroundings. It consists of a phospholipid bilayer with embedded proteins. The cell membrane is involved in a variety of processes such as cell adhesion, ion

conductivity and cell signalling [51].

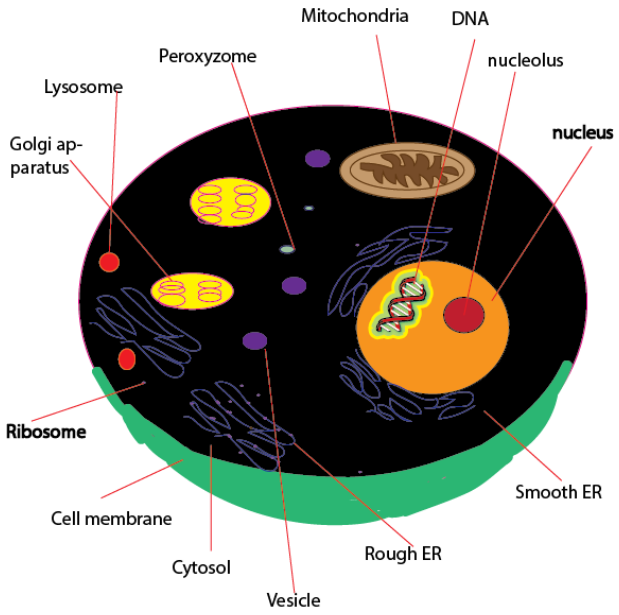


Figure 2.1: *Cell organelles*

The cell consists of the nucleus and the cytoplasm. The cell membrane surrounds the cell. Cytoplasm contains the organelles and the space between them, called the cytosol. Inside the nucleus are the nucleolus and the DNA. Some of the organelles in the cell include mitochondria, rough ER, smooth ER, Golgi apparatus, lysosomes, ribosomes, peroxisomes, etc.

Nucleus: The nucleus contains the hereditary information of a cell. It provides chemical instructions that direct all the cell's activities. It serves as the cell's command centre, sending directions to the cell to grow, mature, divide, or die. It houses DNA (deoxyribonucleic acid), the cells' hereditary material. It is surrounded by a membrane called the nuclear membrane or the nuclear envelope. The membrane protects the DNA and separates the nucleus from the cytoplasm. It contains most of the cell's genetic material, but not all of it. The genetic material is organized as multiple long linear DNA molecules in combination with a large variety of proteins such as histones. DNA in combination with proteins forms the chromosomes. The genes within these chromosomes form the cell's nuclear genome.

The function of the nucleus is to maintain the integrity of the genes and to control the cell activity by regulating gene expression [51].

Cytoplasm: The cytoplasm is the region between the cell membrane and the nucleus. In organisms without a nuclear membrane, it is the region located inside the cell membrane. It is made up of a jelly-liked fluid, called the cytosol which contains other structures outside the nucleus. The cytoplasm is composed of about 80% water and it is usually colourless. Most cellular activities occur within the cytoplasm. The cytoplasm is made up of the inner, granular mass and the outer, clear and glassy layer. The inner mass is called the endoplasm and the outer one is called the ectoplasm [51].

Mitochondria: Mitochondria are rod-shaped cell structures that convert the energy in food molecules to energy that cells can actually use to carry out their functions. Mitochondria are complex organelles and they do have their own genetic material, separate from the DNA in the nucleus. They can copy themselves. Mitochondria are found in most eukaryotic cells and are thought to have been free living organisms which were incorporated into cells over the evolutionary era. The dimensions of mitochondria range from 0.5 to 1.0 micrometer in diameter. They supply the cell with adenosine triphosphate (ATP), which is used as a source of chemical energy. Mitochondria are involved in supplying the cellular energy, signalling, cellular differentiation, cell death, as well as the control of cell cycle and cell growth. They have been implicated in several human diseases, including mitochondrial disorders and cardiac dysfunction. They play a role in the aging process as well. Each mitochondria is composed of compartments that carry out specialized functions. These compartments are the outer membrane, the intermembrane space, the inner membrane,

the cristae and the matrix [51].

Endoplasmic reticulum (ER): There are two different types of the endoplasmic reticulum, the smooth type and the rough type. The smooth endoplasmic reticulum lacks ribosomes. The rough endoplasmic reticulum contains ribosomes on its surface. The smooth endoplasmic reticulum is involved in lipid metabolism, carbohydrate metabolism and detoxification. The rough endoplasmic reticulum is prominent in hepatocytes where active protein synthesis occurs. The endoplasmic reticulum helps process molecules created by the cell. It also transports these molecules to their specific destinations inside or outside the cell. The endoplasmic reticulum forms an interconnected network of membrane-enclosed sacs or tubes known as cisternae. The membranes of the endoplasmic reticulum are a continuation of the outer membrane of the nuclear envelope [51].

Golgi apparatus: Golgi apparatus is composed of stacks of membrane-bound structures known as cisternae. Each cisterna comprises a flat, membrane enclosed disc that includes special Golgi enzymes. The functions of Golgi apparatus involve packaging molecules processed by the endoplasmic reticulum that are meant to be transported out of the cell. It helps in moving material within the cell and out of the cell. It is found in both animal and plant cells. The role of Golgi enzymes is to modify cargo proteins that travel through this organelle depending on where they reside. Golgi apparatus is integral in modifying, sorting and packaging macromolecules. It is involved in the transport of lipids around the cell and the creation of lysosomes [51].

Ribosomes: Ribosomes are small crucial organelles that process the cell's genetic instructions. Ribosomes are made of protein and ribosomal RNA. These organelles can float

freely in the cytoplasm or can be connected to the endoplasmic reticulum. Each cell contains thousands of these organelles. Ribosomes are large and complex molecular machines, found within all living cells. Ribosomes link amino acids together in the order specified by messenger RNA. Ribosomes consist of two major components, the small ribosomal subunit and the large ribosomal subunit. The small ribosomal subunit reads the messenger RNA and the large subunit joins amino acids to form a polypeptide chain. Each subunit is composed of one or more ribosomal RNA molecules and a variety of proteins. Ribosomes are part of the translational apparatus. When a ribosome finishes reading an mRNA molecule, the two subunits of the ribosome split apart. The ribosomal RNA performs the catalytic peptidil transferase activity that links the amino acids together [51].

Lysosomes: Lysosomes are membrane-bound cell organelles found in animal cells. They are spherical vesicles that contain hydrolytic enzymes. These enzymes are capable of breaking down all kinds of biomolecules, including proteins, nucleic acids, carbohydrates and lipids. These organelles are the recycling center of the cell. They digest foreign bacteria that invade the cell. They rid the cell of toxic substances and recycle worn-out cell components. Lysosomes contain around 50 different enzymes which are active at an acidic environment of about pH 5. They act as waste disposal systems of the cell. They digest unwanted material in the cytoplasm. Lysosomes are responsible for cellular homeostasis because of their involvement in secretion, cell signalling, energy metabolism and plasma membrane repair [51].

Peroxisomes: Peroxisomes are also known as the microbodies. They are found in all eukaryotic cells. They are involved in the catabolism of very long chain fatty acids. Per-

oxisomes break down very long chain fatty acids and convert them to medium chain fatty acids, which are shuttled to mitochondria where they eventually break down to carbon dioxide and water [51].

Cytoskeleton: The cytoskeleton is a network of fibres that make up the cell's structural framework. It is a dynamic system, parts of which are constantly destroyed, renewed or newly constructed. The cytoskeleton has three major elements. These elements are microfilaments, microtubules and intermediate filaments. Microfilaments are composed of the protein actin. Microtubules are composed of the protein tubulin and the intermediate filaments consist of more than 60 different building block proteins. The cytoskeleton gives the cell shape and mechanical resistance to deformation. The contraction of the cytoskeleton allows the cell to deform and migrate. It also provides a track-like system that directs the movement of organelles and other substances within the cell. The structure, function and dynamic behaviour of the cytoskeleton depends on the organism and cell type. The structure and function of cytoskeleton changes depending on its association with other proteins [51].

2.2 Inside the nucleus

The nucleus is the largest organelle in animals. It occupies around 10% of the total volume of the cell. The viscous liquid within the nucleus is called the nucleoplasm and is similar in composition with the cytosol. The nucleus appears as a dense spherical organelle.

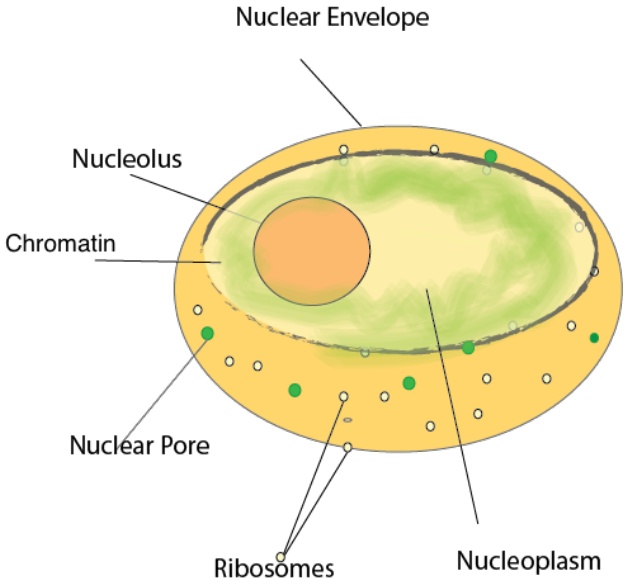


Figure 2.2: *Cell nucleus*

The nucleus is made of the nuclear envelope, the nucleolus, the nucleoplasm and the chromatin. The nuclear envelope separates the nucleus from the cytosol. The nuclear envelope contains pores through which different molecules get in and out of the nucleus.

The nucleus consists of the **nuclear envelope, the nucleolus, the nucleoplasm and the chromatin or chromosomes.**

The nuclear envelope is otherwise known as the nuclear membrane. It consists of two cellular membranes, the inner and the outer membrane. They are parallel to one another with a space of 10 to 50 nanometers in between. The nuclear envelope completely encloses the nucleus and separates the genetic material from the cytoplasm. It serves as a barrier to prevent macromolecules from diffusing freely between the nucleoplasm and the cytoplasm. The outer nucleus membrane is continuous with the membrane of the rough endoplasmic reticulum, and is similarly embedded with ribosomes. The space between the membranes is called the perinuclear space and is continuous with the rough endoplasmic reticulum lu-

men. The nuclear envelope has many nuclear pores in it. They provide aqueous channels through the envelope. The nuclear pores are made of multiple proteins and are referred to as nucleoporins. They consist of approximately several hundred proteins in vertebrates. Their diameter is about 100 nanometer. However, the gap in nuclear pores through which molecules diffuse freely is about 9 nanometer wide due to the presence of regulatory systems within the center of the pores. These pores allow small water-soluble molecules to pass through them. They prevent large molecules, such as nucleic acids and larger proteins from inappropriately entering or exiting the nucleus. These large molecules can only be actively transported into the nucleus [51].

The nucleolus is a densely stained structure. It doesn't have a membrane around it. The main function of the nucleolus is to synthesize ribosomal RNA and assemble ribosomes [51].

The nuclear lamina consists of two networks of intermediate filaments. It provides the nucleus with mechanical support. The nuclear lamina forms an organized meshwork on the internal face of the envelope. If forms a less organized support on the cytosolic face of the envelope. It provides structural support for the nuclear envelope and anchoring sites for nuclear pores and chromosomes. The nuclear lamina is composed mostly of lamin proteins. These proteins are synthesized in the cytoplasm and after transported to the interior of nucleus [51].

Chromosomes are structures made of DNA, protein and RNA. The chromosome is a single piece of coiled DNA containing many genes, regulatory elements and other nucleotide sequences. Chromosomes contain DNA-bound proteins, which package the DNA

and control its functions. The DNA in chromosomes encodes most or all of the genetic information in an organism. Each human cell contains roughly two meters of DNA. Most of the cell cycle, the DNA is in the form of a DNA-protein complex known as chromatin. During cell division the chromatin forms well-defined structures called chromosomes. A small fraction of the genes is located in the mitochondria. There are two types of chromatin, the euchromatin and the heterochromatin. The euchromatin is the less compact DNA form and contains genes that are expressed by the cell. The heterochromatin is the more compact form and contains DNA that is infrequently transcribed. During interphase both types of chromatin can be distinguished. The interphase is the period of the cell cycle where the cell is not dividing [51]. Chromosomes in humans can be divided into two types: autosomes and sex chromosomes. Human cells have 22 pairs of autosomes and one pair of sex chromosomes giving a total of 46 chromosomes per cell. Sequencing of human genome has provided a great deal of information about each chromosome. Each chromosome has one centromere and one or two arms projecting from the centromere. The centromere is the part of the chromosome that links sister chromatids [51].

Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic information. DNA is a nucleic acid. The three major molecules essential for all forms of life are nucleic acids, proteins and carbohydrates. Most DNA molecules consist of two biopolymer strands coiled around each other to form a double helix. The DNA strands are known as polynucleotides. The units from which they are formed are called nucleotides. Each nucleotide is composed of a nitrogen-containing nucleobase, a monosaccaride sugar and a phosphate group [51].

The nucleobase linked to a sugar is called a nucleoside. The nucleobase linked to a

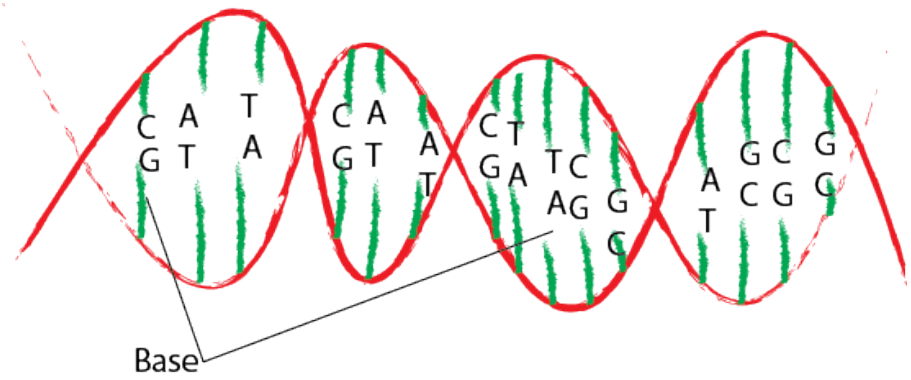


Figure 2.3: *Double helix DNA*

Double helix DNA structure. It contains two strands of DNA which are hybridized together. Hybridization is the process of complementary base pairs binding to form a double helix. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T, and C with G), with hydrogen bonds to make double-stranded DNA.

sugar and one or more phosphate groups is called a nucleotide. There are four nitrogen-containing nucleobases: guanine (G), adenine (A), thymine (T), and cytosine (C). The monosaccharide sugar is called deoxyribose. The sugar is a pentose (five carbon sugar). The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms that are in two adjacent sugar rings. The nucleotides are joined to one another by means of covalent bonds. The covalent bonds are formed between the sugar of one nucleotide and the phosphate group of the other giving so an alternating sugar-phosphate backbone. These asymmetric bonds make the strands of DNA have a direction. The direction of the nucleotides in one strand is opposite to their direction in the other strand. The strands are antiparallel. The asymmetric ends of the DNA are called the 5′ end and the 3′ end. The 5′ end has a terminal phosphate group and the 3′ end has a terminal hydroxyl group [51].

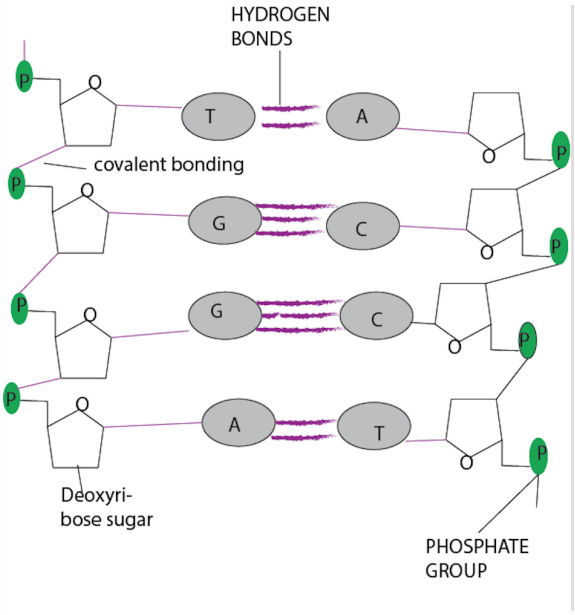


Figure 2.4: *DNA structure*

The two DNA strands (polynucleotides) are composed of monomer units called nucleotides. Each nucleotide is composed of one of four nitrogen-containing nucleobases - cytosine (C), guanine (G), adenine (A), or thymine (T) - a sugar called deoxyribose, and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T, and C with G), with hydrogen bonds to make double-stranded DNA.

The nitrogenous bases of two nucleotides in the two separate polynucleotide strands are bound together by means of hydrogen bonds. This forms the double-stranded DNA. Adenine is bounded to thymine through 2 hydrogen bonds and guanine is bounded to cytosine through 3 hydrogen bonds [51].

The DNA backbone is resistant to cleavage. Both strands of the double-stranded structure store the same biological information. The two strands are separated when the biological information is replicated. 98% of DNA is noncoding, which means these parts of DNA don't encode proteins. The two strands run in opposite directions to each other, therefore antiparallel. One strand is in the 3'-5' direction, while the other strand is in the

5′-3′ direction. The nucleobases are attached to each sugar. It is the sequence of these four nucleobases that encodes biological information [51].

Ribonucleic acid (RNA) is a family of large biological molecules that perform multiple vital roles in the coding, regulation and expression of genes. There are different types of RNAs. There are messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), microRNA (miRNA), small nuclear RNA (snRNA), etc. RNA is a polynucleotide strand. It contains a ribose sugar, a base and a phosphate group. The differences between DNA and RNA are: First, DNA is a double stranded helix, while the RNA molecule consists of one polynucleotide strand only. The sugar in DNA is a deoxyribose sugar, while in RNA is a ribose sugar. RNA has the same nitrogen nucleobase as DNA with the exception of thymine being replaced by uracil (U). Adenine and guanine are purines. Cytosine and uracil are pyrimidines. A phosphate group is attached to the 3′ position of one ribose and the 5′ position of the next. The phosphate groups have a negative charge, making RNA a charged molecule [51].

Messenger RNA conveys genetic information from DNA to the ribosome, where they specify the amino acid sequence of the protein products. The genetic information in the messenger RNA is in the sequence of nucleotides. The nucleotides are arranged in codons consisting of three bases each. Each codon encodes for a specific amino acid, except the stop codons, which terminate the protein synthesis. Messenger RNA is a single stranded molecule that is complementary to one of the DNA strands of a gene. It is an RNA version of the gene that leaves the cell nucleus and moves to the cytoplasm where proteins are made. During protein synthesis, the ribosome moves along the mRNA, reads its base se-

quence, and uses the genetic code to translate each three-base triplet, into its corresponding amino acid [51].

Transfer RNA is an RNA molecule usually 73 to 79 nucleotides in length. It serves as a physical link between the nucleotide sequence in messenger RNA and the amino acid sequence in the protein. It carries an amino acid to the ribosome as directed by the codon in the messenger RNA. Thus transfer RNAs are crucial components in protein translation. The codon in mRNA specifies which amino acid is incorporated into the protein product of the gene from which mRNA is transcribed. The role of transfer RNA is to specify which sequence from the genetic code corresponds to which amino acid. One end of the transfer RNA is called the anticodon. It matches the codon in the messenger RNA. The other end of the tRNA is a covalent attachment to the amino acid that corresponds to the anticodon sequence. Each type of transfer RNA can be attached to only one type of amino acid. Thus there are many types of transfer RNA in each organism [51].

Ribosomal RNA is the RNA component of the ribosomes. Ribosomes are cell organelles made of proteins and ribosomal RNA. Ribosomes contain 60% ribosomal RNA and 40% protein in weight. They contain two major ribosomal RNAs and 50 different types of proteins. The ribosomal RNAs form two subunits, the large subunit and the small subunit. The large subunit acts as a ribozyme. It catalyses the peptide bond formation in proteins. The messenger RNA is sandwiched between the large and the small subunits of the ribosome [51].

2.2.1 Transcription and translation

Transcription is the process in which segments of DNA are copied into RNA by the enzyme RNA polymerase. It is the first step of gene expression. A gene is a stretch of DNA that encodes information.

There is a particular region in DNA that helps initiating the transcription process. This region of DNA is called the promoter. It is located near the transcription start site of a gene, towards the 5′ region of the sense strand. The sense strand is the segment of the double-stranded DNA that runs from the 5′ to 3′ direction. The antisense strand of DNA is the strand that runs in the 3′ to 5′ direction. The enzyme that produces the primary transcript of messenger RNA is called RNA polymerase. In eukaryotes transcription is done by three different RNA polymerases. RNA polymerase II is responsible for transcription of protein coding genes and some other noncoding RNAs, like miRNAs. RNA polymerase creates a transcription bubble. It separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides. RNA polymerase adds matching RNA nucleotides to the complementary nucleotides of the DNA strand. RNA polymerase forms the sugar-phosphate backbone in the newly created RNA strand. The hydrogen bonds of the untwisted RNA-DNA helix break and in this way the newly synthesized RNA strand is released. The primary transcript of RNA undergoes other processes like polyadenylation, capping and splicing. The RNA strand exits the nucleus through the nuclear pore complex. The part of DNA transcribed into an RNA is called a transcription unit. It encodes at least one gene. If the gene transcribed encodes a protein,

the RNA formed is called messenger RNA. The transcribed gene, if not coding for a protein, will encode for a non-coding RNA like miRNA, ribosomal RNA, transfer RNA, or other ribozymes.

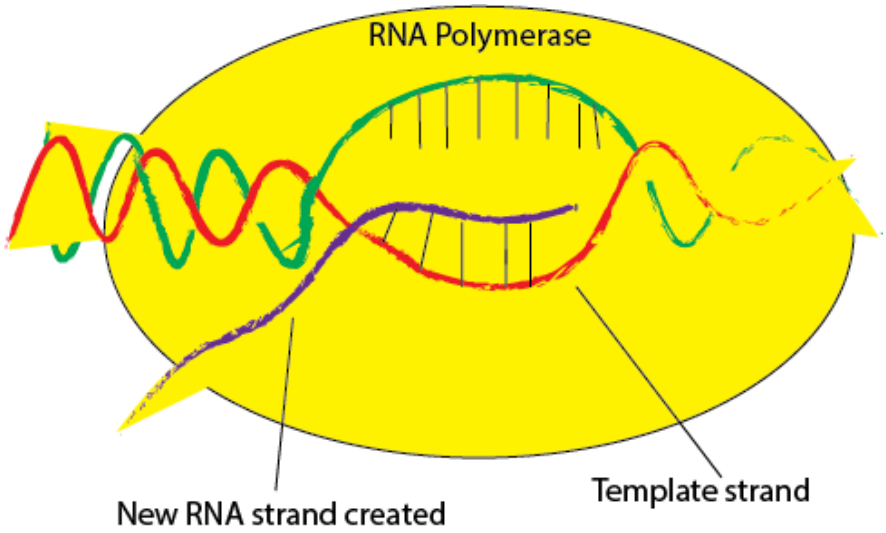


Figure 2.5: *Transcription*

Transcription is the first step of gene expression. A particular segment of DNA is copied into RNA by the enzyme RNA polymerase. Both DNA and RNA are nucleic acids, which use base pairs of nucleotides as a complementary language. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, antiparallel RNA strand. RNA polymerase creates a transcription bubble, which separates the two strands of the DNA helix. This is done by breaking the hydrogen bonds between complementary DNA nucleotides.

The primary transcript produced by RNA polymerase II is called the pre-mRNA, which undergoes several modifications to become mature mRNA. These include 5'capping which is a set of enzymatic reactions that add 7-methylguanosine to the 5' end of the pre-mRNA. It protects the RNA from degradation of exonucleases. This G cap is then bounded to a cap binding complex. Another modification done is the polyadenylation at the 3' end of the pre-mRNA. This occurs if the polyadenylation signal sequence (5'-AAUAAA-3') is present in the pre-mRNA. This signal is usually in between protein-coding sequence and termina-

tor. The pre-mRNA is first cleaved and after around 200 adenines A are added to form the poly(A) tail, which protects RNA from degrading. Another modification of pre-mRNA is RNA splicing. The majority of pre-mRNAs are made up of alternating segments called exons and introns. During splicing, spliceosomes, which are RNA-protein catalytic complexes, catalyze two transesterification reactions. They remove and release an intron and splice neighbouring exons together. Sometimes introns and exons can be either removed or retained in mature mRNA, called alternative splicing creating series of different transcripts originating from a single gene [51].

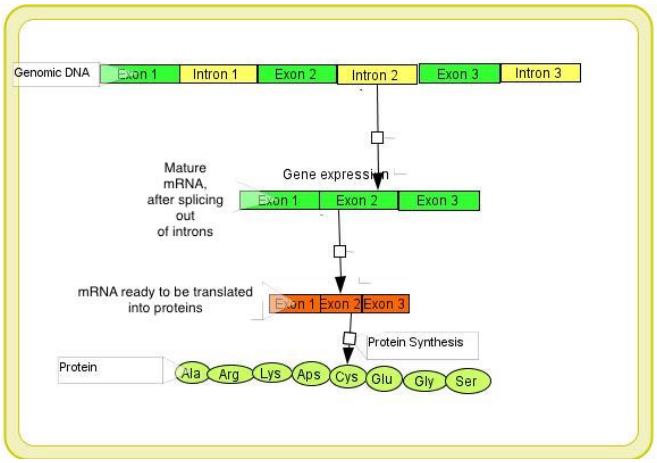


Figure 2.6: *Transcription and translation*

In most eukaryotic genes, coding regions (exons) are interrupted by noncoding regions (introns). During transcription, the entire gene is copied into a pre-mRNA, which includes exons and introns. During the process of RNA splicing, introns are removed and exons joined to form a coding sequence. This "mature" mRNA is ready for translation into proteins.

Translation is the process where messenger RNA is read and translated into a string of amino acids. This process takes place at ribosomes. Ribosomes are organelles in the cytoplasm. Some of them are attached to the endoplasmic reticulum and some stand freely in the cytoplasm. Ribosomes bind and slide along the messenger RNA and serve as a frame-

work for translating the genetic message. As each triplet is read, a transfer RNA molecule brings a specific amino acid to the ribosome. This amino acid is then chemically joined to the previous amino acid by a peptide bond. These tRNA molecules are like waiters. Each is trained to take a specific order from a certain codon. One arm of the tRNA contains an anticodon loop containing the complementary triplet codon. For example, if the codon in the messenger RNA is ACG, the corresponding anticodon in the transfer RNA will be UGC. On another arm is an acceptor stem that attaches to the amino acid corresponding to the triplet codon. After the transfer RNA delivers its amino acid to the translation complex, it floats away to carry another amino acid. The enzyme aminoacyl-tRNA synthetase (ARS), recognizes both the anticodon loop and the acceptor stem of tRNAs. This enzyme attaches the corresponding amino acid to the transfer RNA. These aminoacyl-tRNA synthetase ensure that the transfer RNAs pick up the right amino acids [51].

Translation involves three stages: initiation, elongation, and termination.

Initiation of translation. The start codon that initiates translation is the codon AUG. The region between the start codon and the first codon that is being translated is called the untranslated region or the UTR. The purpose of the untranslated region is important since it contains a ribosome binding site. The translation process begins after the formation of a complex structure. Three initiation factor proteins known as IF1, IF2, and IF3 bind to the small subunit of the ribosome. The pre-initiation complex and a transfer RNA carrying methionine bind to the messenger RNA, near the AUG start codon. This forms the initiation complex. Methionine is always the first amino acid incorporated into any protein but it is not always the first amino acid in mature proteins. It is removed after

translation. Once the initiation complex is formed the large subunit of the ribosome binds to this complex. This causes the release of the initiation factors. The large subunit of the ribosome has three sites where transfer RNA can bind. The A site, or amino acid site is where aminoacyl-tRNA anticodon base pairs up with the messenger RNA codon. This ensures that the correct amino acid is added to the growing polypeptide chain. The polypeptide site, or the P site is where the amino acid is transferred from its transfer RNA to the growing polypeptide chain. The third site is the exit site or the E site. It is the location where the empty tRNA sits before it is released back into the cytoplasm. The only transfer RNA that can bind in the P site of the ribosome, is the methionine -tRNA. The A site is aligned with the second codon of messenger RNA. The ribosome is ready to bind the second aminoacyl-tRNA at the A site. This amino acid will be joined to the initiator methionine by the first peptide bond [51].

Elongation of translation. The next phase is the elongation phase. The ribosome moves along the messenger RNA in the 5′ to 3′ direction. The ribosome shifts or translocates leaving the A site empty for the second amino acid. The transfer RNA that corresponds to the second codon, can after bind to the A site of the ribosome. A peptide bond is formed between the first and the second amino acid. The peptidyl transferase activity that ensures the bonding between the amino acids is a catalytic activity of the ribosome. After the peptide bond is formed, the ribosome shifts, or translocates again. The tRNA occupies the exit site. Then the tRNA is released in the cytoplasm to pick up another amino acid. The A site of the ribosome is empty and ready to receive the next tRNA for the next codon. This process is repeated till all the codons in the messenger RNA are been read by transfer RNA.

The amino acids are linked together in the growing polypeptide chain in the proper order. After all codons are read, translation is terminated and the polypeptide chain is released from the messenger RNA and ribosome [51].

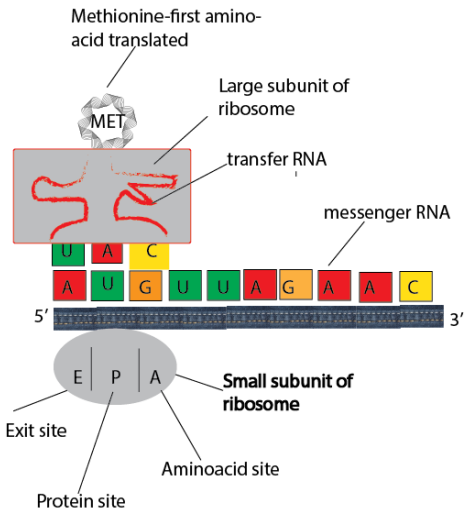


Figure 2.7: *Ttranslation*

Translation is the process in which ribosomes create proteins, following transcription of DNA to RNA in the cell's nucleus. In translation, messenger RNA (mRNA) is decoded by a ribosome, outside the nucleus, to produce a specific amino acid chain, or polypeptide. The ribosome facilitates decoding by inducing the binding of complementary tRNA anticodon sequences to mRNA codons. Ribosomes contain the small subunit, which reads the RNA, and the large subunit, which joins amino acids to form a polypeptide chain. The tRNAs carry specific amino acids that are chained together into a polypeptide as the mRNA passes through and is "read" by the ribosome.

Termination of translation. There are three terminating codons in the messenger RNA. They are UAA, UAG and UGA. No transfer RNAs recognize these three codons. Thus, instead of the tRNA, it's one of the several proteins, called release factors, which binds and facilitate release of the messenger RNA from the ribosome and the dissociation of the ribosome [51].

Chapter 3

Network Classification and Properties

3.1 Network Classification

Network biology provides a description of networks that characterize different biological systems [20]. Complex networks are compared using some basic network measures. These measures include, but are not limited to, average degree, degree distribution, shortest path length and clustering coefficient.

The **degree** of a node is defined as the number of bonds connected to each node. The number of nodes in a network is called the network *order* and the number of bonds is called the network *size*. The **average degree** of a network is defined by $\langle k \rangle = \frac{2L}{N}$, where L is the total number of bonds in the network and N is the number of nodes in the network. In the network shown in figure 3.1, the degree of node i is 7 since it is adjacent to 7 other nodes. The order of the network is 9 since there are 9 nodes and the size of the network is 14 since there are 14 bonds. The average degree of the network can be computed: $\langle k \rangle = \frac{2 \times 14}{9} = 3.1$.

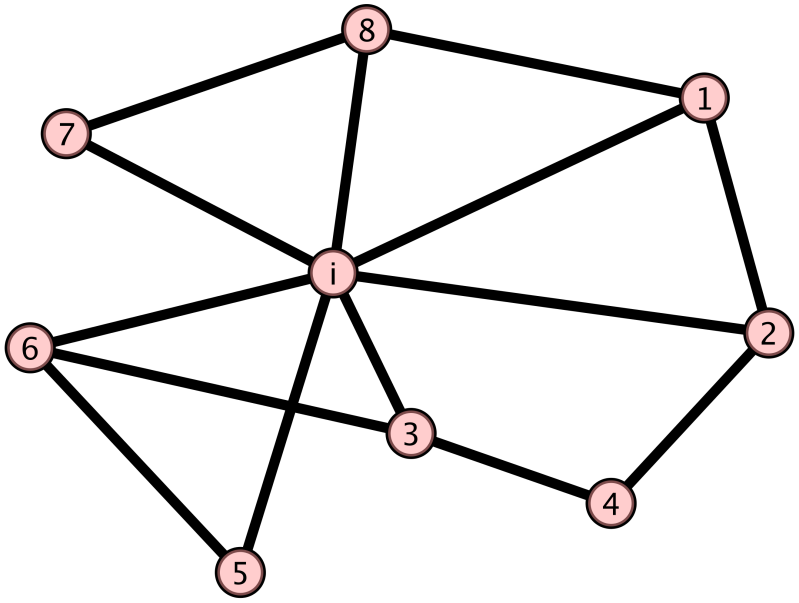


Figure 3.1: *Network*

The **degree distribution** of a network $P(k)$ is the probability distribution of the degrees of the nodes over the whole network. The degree distribution is calculated by counting the number of nodes with $0, 1, 2, 3...$ bonds and dividing it by the total number of nodes. The degree distribution of the network shown in figure 3.1 is $[0, 0, \frac{3}{9}, \frac{5}{9}, 0, 0, 0, \frac{1}{9}]$ since there are no nodes of degree 0, no nodes of degree 1, 3 nodes of degree 2, 5 nodes of degree 3, no nodes of degrees 4, 5 or 6 and one node of degree 7.

Two nodes can be connected to each other through different paths. A shortest path is one that involves the fewest number of bonds between any two given nodes. **The mean path length** is calculated as the average over shortest paths between all pairs of nodes.

The **clustering coefficient** for a particular node i is defined as $\langle c_i \rangle = \frac{2n_i}{k_i(k_i-1)}$, where n_i is the number of neighbours of i that are adjacent to each other and k_i is the degree of node i. Taking the average of these individual clustering coefficients over all nodes i gives the

average clustering coefficient. The clustering coefficient of node i in the network shown in figure 3.1 can be calculated as $\langle c_i \rangle = \frac{2\times5}{7\times6} = 0.24$ since five neighbours of node i are linked to each other as well. In other words, there are five triangles formed with i being one of the vertices of these triangles. The denominator in the clustering coefficient formula gives the maximum possible number of triangles formed if all neighbours of node i were linked to each other. Thus $\frac{k_i\times(k_i-1)}{2}$ gives the possible number of triangles formed around node i. An alternative way of measuring the average clustering coefficient is by using the formula $\langle C \rangle = \frac{3T}{N_3}$, where T is the total number of triangles present in the network and N_3 is the number of connected triplets. Connected triplets are paths of length 2.

Networks can be classified as follows:

Random Networks are networks in which the node degree distribution follows a binomial distribution with probability distribution: $P(k) \sim \binom{n-1}{k} p^k (1-p)^{n-1-k}$. In this case each of n nodes is connected (or not) with independent probability p (or 1-p) (Bernoulli random networks). If n is large the degree distribution of random networks is given by: $P(k) = \frac{n^k}{k!}\frac{e^{-n}}{(1-p)^n} e^{\log(1-p)k}$, where $\lambda = -\log(1-p)$ (Poisson distribution). Nodes with large degrees are very rare. Random networks are also called exponential, because the probability that a node is connected to k other nodes decreases exponentially for large k.

The clustering coefficient of random networks does not depend on the nodes' degrees. Thus, the graph that expresses the relationship between $C(k)$ and k is a horizontal line. Random networks also exhibit the small world property: They are characterized by a small average path length between nodes. The average path length is given by $\langle l \rangle \sim \log N$, where N is the order of the network [20].

The degree distribution $P(k)$ of **scale-free networks** decays as a powerlaw. That is, the probability that a node is connected to k other nodes in the network is given by $P(k) \sim k^{-\gamma}$ where γ is the scale-free exponent. This means nodes of large degree are over-represented in scale-free networks. Nodes of large degrees are called hubs and are adjacent to many other nodes (while the majority of nodes have only a few bonds attached to them). The average degree of the network is not a significant indicator of the network topology since most nodes have degrees which vary considerably from the average degree. On a log - log scale, power law distributions are straight lines. Scale-free networks are also characterized by the small world property: Their average path length is given by $\langle l \rangle \sim \log\log N$, thus much smaller than $\log N$, which characterizes average path lengths in random networks.

3.2 Scale-free networks

Scale-free networks of order n are characterized by degree sequences $\{d_k\}$ which follow a power law distribution (where d_k is the number of nodes of degree k and $\frac{1}{n}d_k$ is the fraction of nodes of degree k).

If $\langle d_k \rangle$ is the average degree distribution over randomly generated scale-free networks, then $\frac{1}{n}\langle d_k \rangle$ is proportional to the probability $P(k)$ that a node has degree k. In scale-free networks, the probability $P(k)$ decays like a powerlaw with exponent γ:

$$P(k) \simeq C_o^{-1}\, k^{-\gamma}. \tag{3.1}$$

Here, γ is the *scale-free network exponent*. The constant C_o is a normalization constant given by

$$C_o = \sum_{k=1}^{n} P(k). \tag{3.2}$$

As $n \to \infty$, it is necessary that $\gamma > 1$ for $P(k)$ to be summable (and $C_o < \infty$). In this case C_o converges to a constant as $n \to \infty$. Thus, if $\gamma > 1$ then the network is said to be integrable with scaling exponent γ (in this event equation (3.1) is the scaling of the limiting degree distribution with $C_o > 0$ finite and $P(k) \to 0$ as $k \to \infty$).

The case that $\gamma = 1$ gives rise to a logarithmic correction. Since $\sum_{k=1}^{n} k^{-1} \sim \log n$, this gives the distribution

$$P(k) \sim \frac{1}{\log n} k^{-1} \tag{3.3}$$

for networks of (large) order n. This network is said to be not integrable, but for asymptotic values and fixed values of n the decay of $P(k)$ will appear to be proportional to k^{-1}.

Since $P(k)$ is the probability that a node in a network has degree k, the average degree sequence $\{\langle d_k \rangle_n\}$ over randomly generated networks of order n is given approximately by $\langle d_k \rangle \sim nP(k)$, for n large. It is not known that the degree sequence is self-averaging (that is, that the degree sequence $\{d_k\}$ has asymptotic distribution $d_k \sim nP(k)$ as $n \to \infty$ for a single randomly generated scale-free network).

The powerlaw decay of degree sequences shows that nodes of large degree (that is, for large k) are more common in scale-free networks (compared to randomly generated networks, where they are exponentially rare). These nodes of large degree are called *hubs*. A precise definition of a hub in a network is somewhat arbitrary, but for the purpose of this

research, a "hub" in a network of order n is defined as a node of degree bigger or equal to $\lfloor \sqrt{n} \rfloor$.

The exponent γ can be estimated from numerical data by computing the average degree sequence $\{\langle d_k \rangle\}$ and then plotting $\log P(k) / \log k$ against $1/\log k$ (for networks of order $n \gg k$). Extrapolating the data to $k = \infty$ using a linear or a quadratic regression gives the value of γ as the y-intercept of the graph. This method works well if $P(k)$ scales with k as in equation (3.1). However, strong corrections to the powerlaw behaviour may make the extrapolation difficult or inaccurate.

A second method to estimate γ is to note that if $\gamma > 1$ and if equation (3.1) holds, then for a fixed value of $\alpha > 0$,

$$\zeta(k) = \log P(\alpha k) - \log P(k) = -\gamma \log \alpha + o(1). \tag{3.4}$$

Experimentation with numerical data shows that by plotting $\zeta(k)$ against $\frac{1}{k} \log k$ good results are obtained, and linear or quadratic regressions of $\zeta(k)$ against $\frac{1}{k} \log k$ can be used to estimate γ.

If it is assumed that $P(k)$ is well approximated by equation (3.1) for all $k \geq 1$, then the average *connectivity* of a network of order n with average degree distribution proportional

to $P(k) = C_o k^{-\gamma}$ is given by

$$\begin{aligned}\langle k \rangle_n &= \frac{\sum_{k=1}^{n} k\,P(k)}{\sum_{k=1}^{n} P(k)} \simeq \frac{\int_1^n k\,P(k)\,dk}{\int_1^n P(k)\,dk} \simeq \left(\frac{\gamma-1}{\gamma-2}\right)\frac{n^\gamma - n^2}{n^\gamma - n} \\ &\simeq \begin{cases} \left(\frac{\gamma-1}{2-\gamma}\right) n^{2-\gamma}, & \text{if } 1 < \gamma < 2; \\ \left(\frac{\gamma-1}{\gamma-2}\right), & \text{if } \gamma > 2. \end{cases}\end{aligned} \tag{3.5}$$

Observe that the asymptotic estimate is very poor if $\gamma \approx 2$, and if n is small.

The cases $\gamma = 1$ and $\gamma = 2$ can also be determined; this gives

$$\langle k \rangle_n \simeq \begin{cases} \frac{n}{\log n}, & \text{if } \gamma = 1; \\ \log n, & \text{if } \gamma = 2. \end{cases} \tag{3.6}$$

The coefficient $\frac{\gamma-1}{\gamma-2}$ may be modifed if $P(k)$ is not well approximated by the powerlaw decay for smaller values of k in equation (3.1). These results, however, do show that the connectivity is a constant independent of n (for large n) if $\gamma > 2$.

The expected number of bonds in the network is given by

$$E_n = \begin{cases} \frac{n^2}{2\log n}, & \text{if } \gamma = 1; \\ \left(\frac{\gamma-1}{2(\gamma-2)}\right) n^{3-\gamma}, & \text{if } 1 < \gamma < 2; \\ \frac{n \log n}{2}, & \text{if } \gamma = 2; \\ \left(\frac{\gamma-1}{2(\gamma-2)}\right) n, & \text{if } \gamma > 2. \end{cases} \tag{3.7}$$

Of course, if $\gamma < 1$, then $E_n = \Theta(n^2)$ and since a complete network has $\frac{1}{2}n(n-1)$ bonds, this implies that these networks are dense in the sense that $\liminf_{n\to\infty} \frac{1}{n^2} E_n > 0$. For all values of $\gamma \geq 1$ the above shows that $\limsup_{n\to\infty} \frac{1}{n^2} E_n = 0$, and the networks are sparse.

These results are useful in examining numerical data for scale-free networks. For example, γ can be estimated by examining degree sequences averaged over randomly sampled networks (from equation (3.1)), or alternatively by using equation (3.4). The connectivity $\langle k \rangle_n$ approaches a constant if $\gamma > 2$ (as in equation (3.5)) or grows as a powerlaw with n if $\gamma < 2$, and with logarithmic corrections if $\gamma = 1$ or $\gamma = 2$ (as in equation (3.6)). Alternatively, the average size E_n (the number of bonds in a network of order n) can be considered, using the results in equation (3.7).

3.3 Protein-protein Interaction Networks (PINs)

Proteins are large, complex molecules that play many critical roles in the body. They do most of the work in cells and are required for the structure, function, and regulation of the body's tissues and organs. Proteins are made up of hundreds or thousands of smaller units called amino acids, which are attached to one another in long chains. There are 20 different types of amino acids that can be combined to make a protein. The sequence of amino acids determines each protein's unique 3-dimensional structure and its specific function. Proteins function as antibodies, enzymes, structural components, messenger molecules, etc.

Physical and chemical processes in the molecular biology of living cells are largely controlled by proteins. Some proteins function independently, but most of them interact with

each other in order to perform biological activities. These interactions are very complex for even the simplest organisms. Knowing the interactions between proteins may help determine the functions of certain proteins which were unknown before. Since the majority of proteins interact with each other, their functions should be studied in the context of these interactions to fully understand their role in the cell. The bonds that keep the proteins together are hydrophobic bonding, van der Waals forces and salt bridges at particular domains on each protein. The binding domains differ in size. They can be a few peptides long or consist of hundreds of amino acids. The size of the binding domains have a direct impact on the strength of interactions.

Protein-protein interactions can be physical interactions or functional associations. Proteins interact physically, meaning they can bind in specific ways and sites, and this binding can produce changes in those same proteins (like conformational changes), which alter their properties.

Two proteins are physically interacting if some of their residues are in physical contact at some point in time. The physical contacts between proteins are specific, occur between defined binding regions in the proteins, and have a particular biological meaning (i.e., they serve a specific function). Often a conformational change in a protein induced by an interaction with another protein activates or inactivates it. For example, an interaction can create a conformational change that enables (activates) or disables (inactivates) the protein to catalyze a given reaction, and so the referred interaction plays an important function in regulating the protein's activity. The results of a physical interaction between two proteins can be diverse. They involve altering the kinetic properties of enzymes (due to

subtle changes in substrate binding), creating new binding sites, inactivating or destroying a protein, changing the specificity of a protein for its substrate through the interaction with different binding partners (demonstrating a new function that neither protein can exhibit alone).

Functional associations do not require physical contact between the associated proteins. Assume that protein A activates protein B at time T1, separates from protein B at time T2 and protein B regulates protein C at time T3. Proteins A and C do not interact, instead, they are associated (functional association). Even for T1 = T2 and the three proteins form a somehow stable complex, proteins A and C are still considered to functionally (not physically) interact. Protein B can undergo conformational changes after its interaction with protein A which are necessary for its interaction with protein C. Therefore, proteins A and C even though not physically interacting are functionally associated with each other.

Protein-protein interactions (PPIs) can be visualized using networks. Protein-protein interaction networks (PINs) are mathematical representations of the physical/functional interactions between proteins in the cell.

The data currently available indicate that protein interaction networks are characterized by degree heterogeneity, the small-world property, and modularity [82]. The first two properties imply the resilience of the network to random disruptions of proteins or interactions due to mutations. They also imply the fragility of these networks to the disruption of hub proteins. The modularity is thought to be both the cause and effect of evolution [82]. The modules represent protein complexes, signaling cascades, and other cell components that evolve partially independently.

The average connectivity in protein interaction networks is small. The networks exhibit the small world property meaning it takes only a handful of links to get from one node to another. The connectivity of different nodes varies considerably. While there are plenty of nodes with very low degree, there is a handful of super connected hubs that have very many connections.

Protein interaction network evolution models based on the concept of gene duplication and divergence provide a good explanation to the observed network properties. Thus, rather than unexpected, the three properties mentioned above are the natural outcome of evolution [82].

3.3.1 Degree distributions of PINs

If protein-protein interaction networks are scale-free, then their degree distribution follows a power law distribution, namely $P(k) \sim Ck^{-\gamma}$. We use data downloaded from the STRING [38] database to show that protein-protein interaction networks are scale-free networks.

STRING stands for search tool for the retrieval of interacting genes/proteins. It contains known and predicted protein-protein interactions. STRING aims to provide a critical assessment and integration of protein-protein interactions, including direct (physical) as well as indirect (functional) associations [38].

The files downloaded on the STRING database were: "protein.links.v9.1.txt.gz" and "protein.links.v10.txt.gz". These files consist of three columns. The first two columns are the identifiers of the two interactors and the third one is the confidence score. Confidence scores show the probability that there is an interaction between the two proteins. Those

probabilities are multiplied by $1,000$. If the confidence score is above 700, it means there's high confidence that the two proteins interact with each other. If the confidence score is between 300 and 700, it shows a medium confidence and a score below 300 is an indicator of a low confidence that there is an interaction between the two proteins.

The degree sequences $\{\langle d_k \rangle\}$ of the proteins in each of the files were computed and then $\log P(k)/\log k$ against $1/\log k$ was plotted. Extrapolating the data to $k = \infty$ using a linear or a quadratic regression gives the value of γ as the y-intercept of the graph. This method works well if $P(k)$ scales with k as in equation (3.1). However, strong corrections to the powerlaw behaviour may make the extrapolation difficult or inaccurate.

$$\log P(k) \sim C - \gamma \log k \tag{3.8}$$

$$\epsilon^2_{(C,\gamma)} = \Sigma_k[(C - \gamma \log k) - \log P(k)]^2 \tag{3.9}$$

$$\frac{\log P(k)}{\log k} = -\gamma + \frac{C}{\log k}. \tag{3.10}$$

First, interactions in the "proteins.links.v9.1.txt.gz" file downloaded from the STRING [38] database were considered (figure 3.2). Extrapolating the data to $k = \infty$ using a quadratic regression gives the value of $\gamma = 2.7$ as the y-intercept of the graph. Next, interactions in the "proteins.links.v10.txt.gz" file downloaded from the STRING [38] database were considered (figure 3.3). Extrapolating the data to $k = \infty$ using a quadratic regression gives the value of $\gamma = 2.5$ as the y-intercept of the graph.

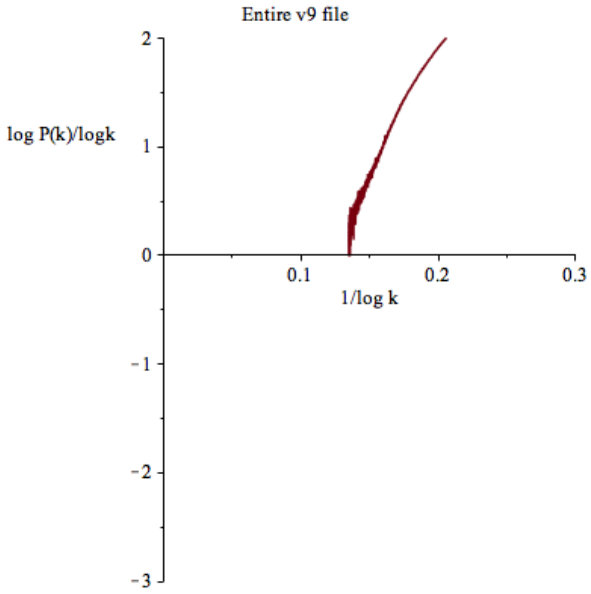


Figure 3.2: $\log P(k)/\log k$ *vs* $1/\log k$ *for protein-protein interactions on "proteins.links.v9.1.txt.gz" file*

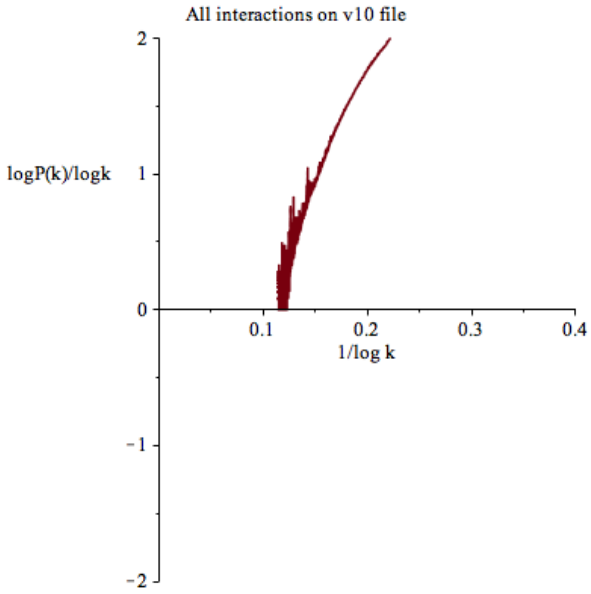


Figure 3.3: $\log P(k)/\log k$ *vs* $1/\log k$ *for protein-protein interactions on "proteins.links.v10.txt.gz" file*

The scaling exponent γ is greater than 2 in both the above networks. Comparing these results with equations (3.5) and (3.7), indicates that the connectivity of the protein-protein

interaction networks is a constant independent of n for large n. The expected number of bonds in these networks should grow linearly in n for large n. This indicates protein-protein interaction networks are sparse networks. They are characterized by a large number of false negatives, meaning missing interactions which are not yet predicted or experimentally verified.

Chapter 4

Mean Field Analysis of Algorithms for Scale-free Networks in Molecular Biology

The scale-free networks model is considered a significant discovery because it has been successfully applied to many complex real-world networks and proved valid. The successful application of this model deemed the other model, the random network model, questionable. The presence of scale-free emerging properties in many real-world networks provides initial evidence that these self-organizing phenomena do not only depend on the characteristics of individual systems, but are general laws of evolving networks [83]. The responsible mechanisms for the emergence of scale-free networks are important in understanding why different systems converge to networks with similar architecture [4].

Four evolutionary algorithms able to generate scale-free networks are described below, with a focus on their mean field analysis.

4.1 Barabasi-Albert networks and the Barabasi-Albert algorithm

The Barabasi-Albert algorithm is a recursive algorithm which grows networks (or clusters of nodes and bonds) from a seed node. This algorithm was introduced in reference [4] and reviewed in 2002 in a seminal paper [1], and its elementary move was inspired by processes underlying the (presumed) evolution of scale-free networks seen in the physical world. The elementary move is a preferential attachment of new nodes (and bonds) to hubs (nodes of high degree) in the network. Thus the two main ideas of the model are *growth* and *preferential attachment*. Growth means that the number of nodes in the network increases in time. Preferential attachment means that new nodes have a tendency to connect to nodes with high degree.

A social network modelled by the Barabasi-Albert model is the co-authorship network of scientists. Barabasi-Albert captured the dynamic and the structural mechanisms that govern the evolution and topology of this complex system by mapping the electronic database containing all relevant journals in mathematics and neuroscience for an 8-year period (1991-98). Each node in the network represents an author and a bond between two nodes means those authors have published together. The network constantly expands by the addition of new authors to the database and the addition of new internal links representing papers co-authored by authors that were part of the database, as well as external links, links representing papers for authors that were not in the database.

The two main ideas that underlie the evolution of the co-authorship network are growth

and preferential attachment. The network continuously grows as new authors publish their papers. For a new author, who appears for the first time on a publication, preferential attachment has a simple meaning: it is more likely that the first paper will be co-authored with somebody that already has a large number of co-authors (links) than with somebody less connected. As a result old authors with more bonds will increase their number of co-authors at a higher rate than those with fewer bonds [1]. A large number of new links appear between old nodes as the network evolves, representing papers written by authors that were part of the network, but did not collaborate before. Such internal links are known to effect both the topology and dynamics of the network [5]. These internal links are also subject to preferential attachment.

The Barabasi-Albert Algorithm is described below. The algorithm is initiated by a single node, and then new nodes and bonds are recursively attached, with new bonds preferentially attached to existing nodes of large degree.

A Barabasi-Albert network of order N nodes is grown as follows:

Barabasi-Albert algorithm:

1. Initiate the network with one node x_0;

2. Suppose that the network consists of nodes $\{x_0, x_1, \ldots, x_{n-1}\}$ of degrees $\{k_0, k_1, \ldots, k_{n-1}\}$;

3. Append a new node x_n by executing step (a) or step (b):

(a) With probability p: Select x_j uniformly and attach x_n to it by inserting the bond $\langle x_j \sim x_n \rangle$;

(b) With default probability $1-p$: Attach x_n by adding bonds $\langle x_j \sim x_n \rangle$ independently with probability $\frac{k_j}{\sum_j k_j}$;

4. Repeat step 3 until a network of order N is grown.

Step 3(a) is a random attachment of a node and bond, and step 3(b) attaches a node with bonds *preferentially* to existing nodes of high degree. The algorithm has a single parameter p. If $p = 1$ then the algorithm grows acyclic (and connected) networks of order N (these are random trees).

On the other hand, if $p = 0$, then step 3(b) is executed on each iteration. New bonds are created with probabilities $q_j = \frac{k_j}{\sum_j k_j}$ for $j = 0, 1, \ldots, n-1$ when the n-th node is added. This shows that the expected number of bonds added in this step is on average $\sum_j q_j = 1$. That is, on average 1 bond is added in each iteration, and the average sum of degrees $\sum_j k_j$ should be equal to $2n$ by handshaking after n iterations. This suggests that the algorithm grows a sparse graph with increasing n. However, since bonds are appended preferentially on growing hubs, the largest clusters in the network should be dominated by growing hubs.

For values of $p \in (0, 1)$ the algorithm adds either (with probability p) a single bond randomly, or it adds a collection of bonds (on average one bond) preferentially. This grows simple networks of order N and size $N-1$, typically not connected unless acyclic.

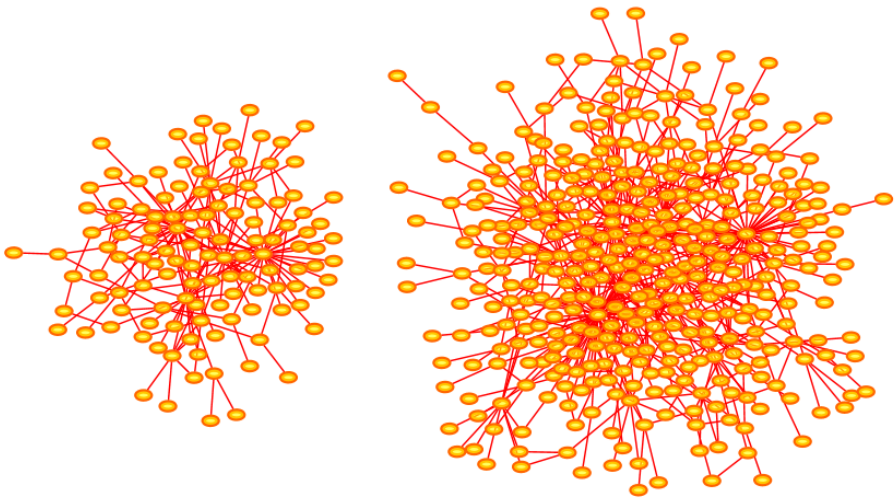


Figure 4.1: *Barabasi-Albert networks with $p = 0$:*

The network on the left was grown to order $n = 122$. It has 5 hubs of degrees $\{12, 17, 18, 19, 31\}$ exceeding $\sqrt{122}$. The network on the right was grown to order $n = 380$. This network has 3 hubs of degrees $\{29, 47, 63\}$ exceeding $\sqrt{380}$. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].

In figure 4.1 an example of a Barbasi-Albert network of order 122 with $p = 0$ is shown (left) and the right is a network of size 380. The appearance of hubs in these networks is clearly seen: In the network on the left there are 5 nodes of degrees exceeding $\sqrt{122}$, the largest of degree 31, and in the network on the right there are 3 hubs of degrees exceeding $\sqrt{380}$, the largest of degree 63.

4.1.1 Modified Barbasi-Albert networks

Barabasi-Albert networks are relatively sparse networks. A modification of the algorithm can be introduced to grow denser networks. For example, one may replace step 3(b) by

3(b). With default probability $1 - p$: Attach x_n by adding bonds $\langle x_j \sim x_n \rangle$ with probability $q_j = \min\{\frac{\lambda k_j + A}{\sum_j k_j}, 1\}$ (where λ and A are non-negative parameters of the

algorithm);

Since $k_j \ll \sum_j k_j$ in Barabasi-Albert networks, one may assume that $\lambda k_j + A \leq \sum_j k_j$ for values of λ and A which are not too large (and so $q_j \leq 1$).

In the modifed Barabasi-Albert Model, the effect of the parameter λ is to increase the density of the network if $\lambda > 1$ and decrease it otherwise. Every node has an initial attractiveness A. This means even the isolated nodes will have a chance to connect to the new nodes entering the system. In the Barabasi-Albert Model, the isolated nodes were not able to receive any new links during the growth process and remained isolated regardless the growth of the network.

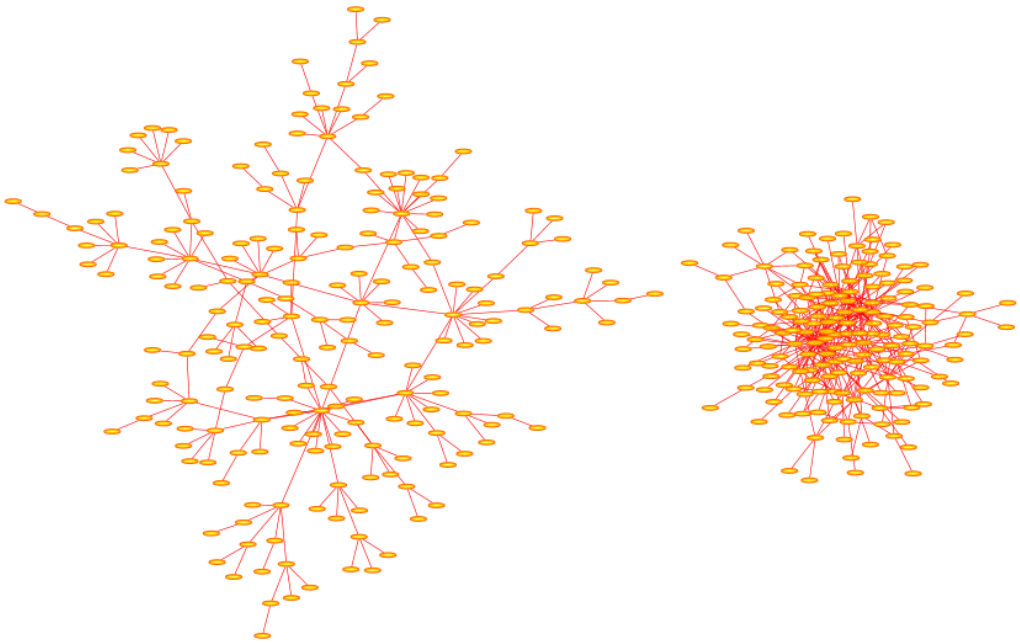


Figure 4.2: *Modified Barabasi-Albert networks:*

The network on the left was grown with $\lambda = 0.5$ to order $n = 201$. It has two hubs of degrees $\{15, 17\}$ which exceed $\sqrt{201}$. The network on the right was grown with $\lambda = 2$ to order $n = 172$. This network contains hubs of degrees $\{15, 15, 16, 17, 19, 27, 33\}$ exceeding $\sqrt{172}$. In both cases the algorithm was implemented with $p = 0$. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].

In figure 4.2 two examples of Modified Barabasi-Albert networks are shown, one a sparse network with $\lambda = 0.5$, $A = 0$ and $p = 0$, and the second a denser network with $\lambda = 2.0$, $A = 0$ and $p = 0$. In both cases the algorithm was iterated 200 times; the sparse network has order 201 and two hubs of degrees $\{15, 17\}$, and the dense network has order 172 with seven hubs of degrees $\{15, 15, 16, 17, 19, 27, 33\}$.

4.1.2 Variant Barbasi-Albert networks

A variant Barbasi-Albert algorithm can be introduced by changing step 3(b) in the Barbasi-Albert algorithm to

3(b). With default probability $1 - p$: Attach x_n by adding bonds $\langle x_j \sim x_n \rangle$ with probability $q_j = \min\{\frac{k_j^\alpha + A}{\sum_j k_j}, 1\}$, (where α and A are non-negative parameters of the algorithm);

The effect of the parameter α is to increase the probability of adding bonds to the hubs of the network if $\alpha > 1$, and to decrease this probability if $\alpha < 1$.

In the variant Barabasi-Albert Model every node has an initial attractiveness A. This means even the isolated nodes will have a chance to connect to the new nodes entering the system. In the Barabasi-Albert Model, the isolated nodes were not able to receive any new links during the growth process and remained isolated regardless the growth of the network.

In the case that $\alpha > 1$ networks dominated by a single very large hub are obtained (see figure 4.3 (right network)), while networks with $\alpha < 1$ are more sparse and not dominated

by a few hubs (see figure 4.3 (left network)). The left network in figure 4.3 was grown by putting $\alpha = 0.15$ and $A = 0$ and has order 327. None of the nodes in this network has degree which exceeds $\sqrt{327}$, and so none qualify as hubs. A denser network is obtained if $\alpha = 1.15$ and $A = 0$, as shown in figure 4.3 on the right. This network is dominated by hubs of degrees $\{22, 24, 26, 42, 43, 116\}$ and has order 351.

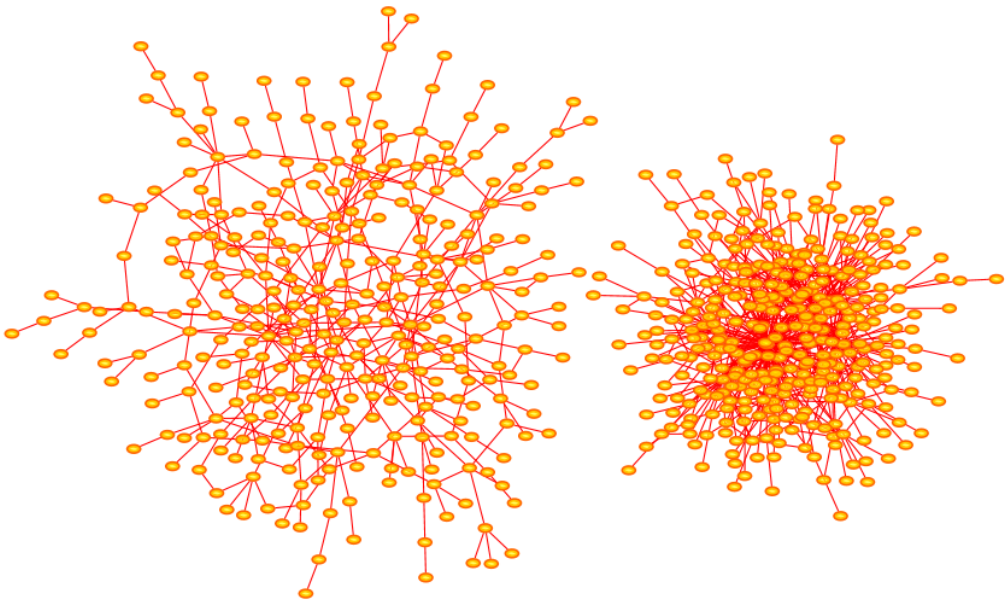


Figure 4.3: *Variant Barabasi-Albert networks:*

The network on the left was grown using $\alpha = 0.15$ and $A = 0$ to a total of $n = 327$ nodes. This graph is very sparse, and none of its nodes qualify as hubs. The network on the right was grown to order $n = 351$ with $\alpha = 1.15$ and $A = 0$. This is a dense network with several nodes qualifying as hubs of degrees $\{22, 24, 26, 42, 43, 116\}$. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].

4.1.3 Mean field theory for Modified Barabasi-Albert networks

Let $k_j(n)$ be the degree of node j after n iterations of the modified Barabasi-Albert algorithm. A mean field calculation of $k_j(n)$ is done by assuming that $k_j(n)$ is equal to its

expected value for each n; that is, $k_j(n) = \langle k_j(n) \rangle$ for each j and n.

The modified Barabasi-Albert algorithm appends bonds to a network of order n as follows: Step 3(a) is executed with probability p, and a bond (and the $(n+1)$-th node) is appended with uniform probability on one of the n existing nodes. The probability that node j gets a bond in this way is $\frac{p}{n}$ and on average one bond is attached with probability p.

If step 3(b) is done instead, then the expected number of bonds added in the mean field is approximately $\sum_j \frac{\lambda k_j(n)+A}{\sum_j k_j(n)} = \lambda + \frac{nA}{\sum_j k_j(n)}$. The total number of bonds in the network is

$$2E_n = \sum_j k_j(n) \tag{4.1}$$

by handshaking. Thus, the increment in the number of bonds when the next node is appended is

$$\Delta E_n = p + (1-p)\lambda + (1-p)\frac{nA}{2E_n}. \tag{4.2}$$

Approximate this by a differential equation

$$2E_n \frac{d}{dn} E_n = 2(p + (1-p)\lambda)E_n + (1-p)nA. \tag{4.3}$$

This can be solved to obtain

$$E_n = \frac{n}{2}((p + (1-p)\lambda) + \sqrt{(p + (1-p)\lambda)^2 + 2(1-p)A}) = Cn, \tag{4.4}$$

where C is a function of (p, λ, A) defined by this expression. Notice that E_n grows approx-

imately linearly in n, so that modified Barabasi-Albert networks will be necessarily sparse as $n \to \infty$ (and by equation (3.7) the scaling exponent is $\gamma > 2$).

To determine the value of γ, a recurrence for $k_j(n)$ can be written. With each iteration the mean field value of $k_j(n)$ (the degree of the j-th node after n iterations) increments by

$$k_j(n+1) = k_j(n) + \frac{p}{n} + \frac{(1-p)(\lambda k_j(n)+A)}{2E_n} \tag{4.5}$$

since $2E_n = \sum_j k_j(n) = 2Cn$, and since the probabilty of adding a bond to node j is $\frac{\lambda k_j(n)+A}{\sum_j k_j(n)}$. This can again be approximated by a differential equation: Take $n \to t$, a continuous time variable, and let $k_j(n) \to k_j(t)$, the continuous mean field degree of node j. Then

$$\frac{d}{dt}k_j(t) = \frac{p}{t} + \frac{(1-p)(\lambda k_j(t)+A)}{2Ct}. \tag{4.6}$$

The initial condition is to assume that node j is added at time t_j. Putting $A = 0$ and $\lambda = 1$ gives $C = 1$ and the equation

$$\frac{d}{dt}k_j(t) = \frac{p}{t} + \frac{(1-p)k_j(t)}{2t} \tag{4.7}$$

which was also derived in reference [5]. In this event the solution is (assuming the initial condition $k_j(t_j) = 1$):

$$k_j(t) = \frac{1+p}{1-p}(t/t_j)^{(1-p)/2} - \frac{2p}{1-p} \tag{4.8}$$

More generally, equation (4.6) can be cast in the general form

$$\frac{d}{dt}k_j(t) = \frac{Q}{t} + \frac{P}{t}k_j(t) \tag{4.9}$$

where $Q = p + \frac{(1-p)A}{2C}$ and $P = \frac{(1-p)\lambda}{2C}$, with solution

$$k_j(t) = \left(1 + \frac{Q}{P}\right)\,(t/t_j)^P - \frac{Q}{P} \tag{4.10}$$

using again the initial condition $k_j(t_j) = 1$.

The mean field degree distribution can be determined from this solution. The probability that node j has degree $k_j(t)$ smaller than κ at time t is denoted by $P[k_j(t) < \kappa]$. Since $k_j(t) < \kappa$ if

$$\left(1 + \frac{Q}{P}\right)(t/t_j)^P - \frac{Q}{P} < \kappa \quad \text{or, equivalently,} \quad t_j > t\left(\frac{Q/P+\kappa}{1+Q/P}\right)^{-1/P},$$

this is also the probability $P\left[(t_j/t) > \left(\frac{Q/P+\kappa}{1+Q/P}\right)^{-1/P}\right]$. If the node t_j is chosen uniformly from the n available, then

$$P[k_j(t) < \kappa] = P\left[(t_j/t) > \left(\frac{Q/P+\kappa}{1+Q/P}\right)^{-1/P}\right] = 1 - \left(\frac{Q/P+\kappa}{1+Q/P}\right)^{-1/P}. \tag{4.11}$$

The mean field degree distribution is the derivative of this to κ:

$$P(\kappa) = P[k_j(t) = \kappa] = \frac{\partial}{\partial\kappa}P[k_j(t) < \kappa] = \frac{(P+Q)^{1/P}}{(P\kappa+Q)^{1+1/P}}. \tag{4.12}$$

For large κ this shows that the modified Barbasi-Albert network is scale-free with exponent

$$\begin{aligned}\gamma &= 1+\tfrac{1}{P} = 1+\tfrac{2C}{(1-p)\lambda} \\ &= 1+\tfrac{((p+(1-p)\lambda)+\sqrt{(p+(1-p)\lambda)^2+2(1-p)A})}{(1-p)\lambda}.\end{aligned} \tag{4.13}$$

This is the mean field exponent of a modified Barabasi-Albert network. Putting $A = 0$ gives the exponent

$$\gamma = 3+\tfrac{2p}{(1-p)\lambda}. \tag{4.14}$$

For small $\lambda < 1$ the exponent is large, indicating a network with few nodes (if any) of high degree. For large $\lambda > 1$, $\gamma \searrow 3^+$. This is a lower bound on γ for modified Barabasi-Albert networks.

If $\lambda = 1$, then the exponent γ is given by

$$\gamma = 1+\tfrac{1}{1-p}+\tfrac{\sqrt{1+2(1-p)A}}{1-p}. \tag{4.15}$$

In this model one similarly finds that $\gamma \geq 3$, and in fact, if $p = 0$, then $\gamma = 2+\sqrt{1+2A}$. The parameter A may be used to tune the exponent γ for any given p.

If both $\lambda = 1$ and $A = 0$, then the known expression for γ for Barabasi-Albert networks is recovered, namely

$$\gamma = \tfrac{3-p}{1-p}. \tag{4.16}$$

Notice that $\gamma \geq 3$ and that $\gamma = 3$ if $p = 0$ [5].

The connectivity of modified Barabasi-Albert networks is given by

$$\langle k \rangle_n \simeq \frac{\int_1^n k\, P(k)\, dk}{\int_1^n P(k)\, dk} \simeq \frac{2C}{2C - (1-p)\lambda}, \tag{4.17}$$

where $2C = ((p + (1-p)\lambda) + \sqrt{(p + (1-p)\lambda)^2 + 2(1-p)A})$. Since $2 - \gamma = 1 - \frac{1}{P}$, equation (3.5) gives $\langle k \rangle_n \simeq \frac{1}{1-P}$. Inserting the value of P gives the result above as well.

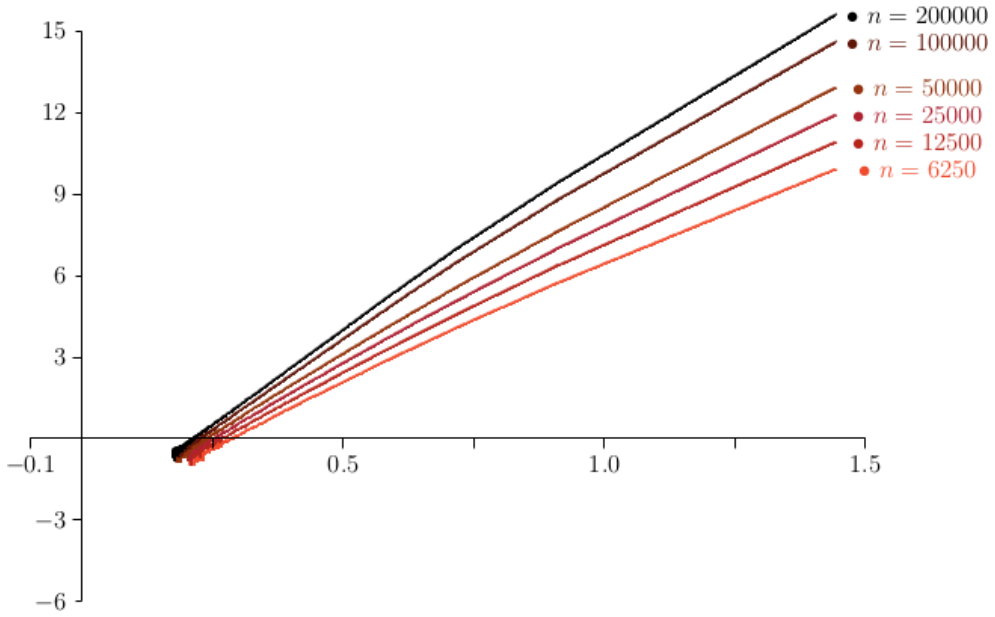


Figure 4.4: *Scaling of Barabasi-Albert networks with $p = 0$:*

Data on networks generated by the Barabasi-Albert algorithm with $p = 0$. In each case 100 networks were grown and the average degree sequence $P_n(k)$ computed. The curves above are plots of $\log P_n(k) / \log(k+1)$ against $1/\log(k+1)$ for $n \in \{6250, 12500, 25000, \cdots, 200000\}$. Least squares fit to the data using a quadratic model gives the y-intercepts which averages to 3.026. This is very close to the value $\gamma = 3$ predicted for the scaling exponent in this model by the mean field approach.

In figure 4.4, the probability $P(k)$, the normalized degree distribution, with the probability that the degree of a node is equal to k, is examined by plotting $\log P(k) / \log(k+1)$ against $1/\log(k+1)$ where $P(k)$ was estimated for values $n \in \{6250, 12500, 25000, 50000, 100000,$

200000} and for $p = 0$. The curves should intersect the vertical axis at $-\gamma$. Least squares fit of the data to quadratic curves gives 6 estimates for γ, which average to $\gamma = 3.026 \pm 0.076$, very close to the theoretical value $\gamma = 3$ (see equation (4.14) for $p = 0$ and $\lambda = 1$).

Data collected for the same values of n and for $p = 0.5$ cannot be successfully analyzed by regressions with quadratic curves, but cubic curves give the average value $\gamma = 5.161 \pm 0.068$, which is not equal to but still fairly well approximated by $\gamma = 5$ prediced by equation (4.14) for $p = 0.5$ and $\gamma = 1$.

When $p = 0.8$ the plots are strongly curved and extrapolation to estimate γ is more difficult. In this case a different approach is needed. Putting $\alpha = \frac{1}{2}$ in equation (3.4) gives

$$\log P(k) - \log P(\tfrac{1}{2}k) = -\gamma \log 2 + O(1) \tag{4.18}$$

so that a plot of $\zeta(k) = (\log P(k) - \log P(\frac{1}{2}k)) / \log 2 \to -\gamma$ as $k \to \infty$. That is, plotting $\zeta(k)$ against $\frac{1}{k}$ gives a curve with y-intercept equal to $-\gamma$. Better results are obtained when plotting against $\frac{1}{k} \log k$. In this case a linear extrapolation gives $\gamma = 11.67 \pm 0.41$ and a quadratic extrapolation gives $\gamma = 11.6 \pm 2.6$. These results are close to the mean field prediction $\gamma = 11$ for $p = 0.8$. Incidentally, if $p = 0.5$ then this kind of analysis shows that $\gamma = 5.47 \pm 0.14$ (linear extrapolation) or $\gamma = 4.4 \pm 1.0$ (quadratic extrapolation), and if $p = 0$, then the results are $\gamma = 3.088 \pm 0.022$ (linear extrapolation) and $\gamma = 2.86 \pm 0.18$ (quadratic extrapolation).

If $\lambda = 2$ and $p = A = 0$ then the algorithm grows modified Barabasi-Albert networks with $\gamma = 3$ (the mean field estimate given by equation (4.13)). Estimating γ by plotting

$\zeta(k)$ against $\frac{1}{k}\log k$ gives the estimate $\gamma = 3.019 \pm 0.098$ (linear extrapolation) and $\gamma = 2.62 \pm 0.33$ (quadratic extrapolation).

The connectivity of Modified Barabasi-Albert networks should converge quickly to a constant with increasing n (by equation (3.5)) since $\gamma > 2$. Computing it for Barabasi-Albert networks (with $\lambda = 1$ and $A = 0$) gives $\langle k \rangle_n \approx 3.16$ for $p = 0$, $\langle k \rangle_n \approx 2.28$ for $p = 0.5$ and $\langle k \rangle_n \approx 2.08$ for $p = 0.8$, and for $n = 12500$. Increasing n does not change these results.

4.1.4 Mean field theory for Variant Barabasi-Albert networks

In this model the increment in the number of bonds when the $(n+1)$-th node is appended is given by

$$\Delta E_n = p + \frac{(1-p)(\sum_j (k_j(n))^\alpha + A)}{2\,E_n}. \tag{4.19}$$

Approximating this with a differential equation gives

$$2E_n \tfrac{d}{dn} E_n = 2pE_n + (1-p)nA + (1-p)\sum_j (k_j(n))^\alpha. \tag{4.20}$$

The right hand side can be approximated as follows: For $\alpha > 1$ the algorithm should grow dense networks with nodes of high degree. Assuming that $k_j(n) \approx k_\ell(n)$ for all ℓ shows that $\sum_j (k_j(n))^\alpha \approx n(k_j(n))^\alpha \approx n\left(\frac{1}{n}\sum_j k_j(n)\right)^\alpha = n^{1-\alpha}(2E_n)^\alpha$. Using this approximation gives

$$2E_n \tfrac{d}{dn} E_n \approx 2pE_n + (1-p)nA + (1-p)n^{1-\alpha}(2E_n)^\alpha. \tag{4.21}$$

If $A = p = 0$, then the differential equation can be solved directly to obtain $E_n \simeq 2^{(\alpha-1)/(2-\alpha)}n$, provided that $\alpha > 1$. This shows that E_n is linear in n, which may be expected if α is not too much larger than 1.

Numerical experimentation shows that E_n grows linearly in n for values of α not too much larger than 1. For example, if $p = 0.5$, $A = 1$ and $\alpha = 1$ then $\frac{1}{n}E_n \to 1.207\ldots$, if $\alpha = 1.5$ then $\frac{1}{n}E_n \to 1.539\ldots$, but if $\alpha = 2$ then $\frac{1}{n}E_n$ increases slowly with n. Similarly, if $p = 0$, and $A = 1$, then, if $\alpha = 1$, $\frac{1}{n}E_n \to 1.366\ldots$, and if $\alpha = 1.5$, $\frac{1}{n}E_n \to 2.399\ldots$, but if $\alpha = 2$ then $\frac{1}{n}E_n$ increases slowly with n and for even larger values of n this growth accelerates.

The recurrence for the degree of the j-th node may be approximated by a differential equation similar to equation (4.6): Assuming that $E_n = Dn^\beta$, replacing $n \to t$ (a continuous time variable), gives the recurrence

$$k_j(t+1) = k_j(t) + \frac{p}{t} + \frac{(1-p)((k_j(t))^\alpha + A)}{2Dt^\beta}. \tag{4.22}$$

This can be approximated by the differential equation

$$\frac{d}{dt}k_j(t) = \frac{p}{t} + \frac{(1-p)((k_j(t))^\alpha + A)}{2Dt^\beta}. \tag{4.23}$$

If $\alpha = 1$ and $\beta = 1$ then the solution of this equation gives the Barabasi-Albert case with $\gamma = 3$. Proceed by considering the case $A = p = 0$ and the initial condition $k_j(t_j) = 1$.

Assume that $\alpha = 1 + \epsilon$. Then the equation becomes

$$\frac{2Dt^\beta}{k_j(t)} \frac{d}{dt} k_j(t) = (k_j(t))^\epsilon. \tag{4.24}$$

A perturbative approach for small ϵ can be done by expanding $(k_j(t))^\epsilon = \exp(\epsilon \log k_j(t)) = 1 + \epsilon \log k_j(t) + \frac{1}{2}\epsilon^2 \log^2 k_j(t) + \cdots$. Truncating this at $O(\epsilon^2)$ and putting $g(t) = \log k_j(t)$ gives the differential equation

$$2Dt^\beta \frac{d}{dt} g(t) = 1 + \epsilon g(t) + \tfrac{1}{2}\epsilon^2 g^2(t). \tag{4.25}$$

Using the initial condition $g(t_j) = \log k_j(t_j) = 0$ the solution of this equation is

$$\epsilon\, g(t) = \begin{cases} -1 + \tan\left(\frac{\pi}{4} + \frac{\epsilon}{4D} \log(\frac{t}{t_j})\right), & \text{if } \beta = 1; \\ -1 + \tan\left(\frac{\pi}{4} + \frac{\epsilon}{4D(\beta-1)}(t_j^{1-\beta} - t^{1-\beta})\right), & \text{if } \beta > 1. \end{cases} \tag{4.26}$$

In the case $\beta > 1$ suppose that $\delta = \beta - 1$ and that δ is small. Then approximate

$$t_j^{1-\beta} - t^{1-\beta} = e^{-\delta \log t_j} - e^{-\delta \log t} \approx \delta \log\left(\frac{t}{t_j}\right) - \tfrac{1}{2}\delta^2 \log\left(\frac{t}{t_j}\right) \log\left(t t_j\right) + O(\delta^3).$$

With this approximation the solution for $g(t)$ above can be expanded in ϵ and δ to give the

first order approximations

$$g(t) \simeq \begin{cases} \frac{1}{2D}\log(\frac{t}{t_j}) + \frac{\epsilon}{8D^2}\log^2\frac{t}{t_j}, & \text{if } \beta = 1; \\ \frac{1}{2D}\log(\frac{t}{t_j}) + \frac{\epsilon}{8D^2}\log^2\frac{t}{t_j} - \frac{\delta}{4D^2}\left(D\log^2(\frac{t}{t_j}) + \log t_j \log(\frac{t}{t_j})\right), & \text{if } \beta > 1. \end{cases}$$

Proceed by solving the above quadratics for $\log(\frac{t}{t_j})$ in terms of $g(t)$. Expand the solution in ϵ and δ and keep only the first few terms. In the case that $\beta = 1$ this gives

$$\log(\tfrac{t}{t_j}) \approx 2D\, g(t) - \epsilon D\, g^2(t). \tag{4.27}$$

Since $g(t) = \log k_j(t)$, the probability that $k_j(t) < \kappa$ is given by

$$P[k_j(t) < \kappa] = P\left[\frac{t_j}{t} > \kappa^{\epsilon D \log\kappa - 2D}\right] \approx 1 - \kappa^{\epsilon D \log\kappa - 2D}. \tag{4.28}$$

Taking the derivative with respect to κ gives the distribution function in the case that $\beta = 1$:

$$P(k) \sim D(2 - D\epsilon \log k)\, k^{-1-2D+D\epsilon\, \log k}. \tag{4.29}$$

These networks are not scale-free. For small values of k the $\log k$ terms are slowly varying, and the networks will appear to be scale-free with $\gamma = 1 + 2D$. However, with increasing k the exponent reduces in value and the connectivity of the network will become dependent on k in the way seen in equation (3.5) for small values of γ.

Notice that if $D = 1$ and $\epsilon = 0$ (or $\alpha = 1$), then the above reduces to $P(k) \sim k^{-3}$, as

expected for Barabasi-Albert networks.

If $\beta > 1$, then a similar approach to the above may be considered. Solving the expression for $g(t)$ above for $\log(\frac{t}{t_j})$ and keeping only terms to $O(\epsilon)$ and $O(\delta)$ gives

$$\log(\tfrac{t}{t_j}) \approx 2D\, g(t) - \epsilon D\, g^2(t) + \delta(2D^2 g^2(t) + g(t)\log t_j). \tag{4.30}$$

This shows that

$$\begin{aligned} P(k_j(t) < \kappa) &= P\left(\tfrac{t_j}{t} > \kappa^{\epsilon D \log\kappa - 2D - 2D^2\delta\log\kappa - \delta\log t_j}\right) \\ &\approx 1 - \kappa^{\epsilon D\log\kappa - 2D - 2D^2\delta\log\kappa - \delta\log t_j}. \end{aligned}$$

This shows that

$$P(k) \sim (2D(1 + 2D\delta\log k - \epsilon\log k))\, k^{-1-2D-\delta\log t_j - D(2D\delta-\epsilon)\log k}. \tag{4.31}$$

This gives an effective exponent $\gamma_k = 1 + 2D + \delta\log t_j + D(2D\delta - \epsilon)\log k$ which decreases in size if $2D\delta - \epsilon < 0$ and increases in size if $2D\delta - \epsilon > 0$. Since $\delta = \beta - 1$ and $\epsilon = \alpha - 1$, and for small α numerical simulations show that $\beta \approx 1$, it is normally the case that $2D\delta - \epsilon < 0$. This means that the networks will first appear scale-free with constant connectivity until k becomes large enough in which case the connectivity will increase with k, as seen above.

4.1.5 Numerical results on Variant Barabasi-Albert networks

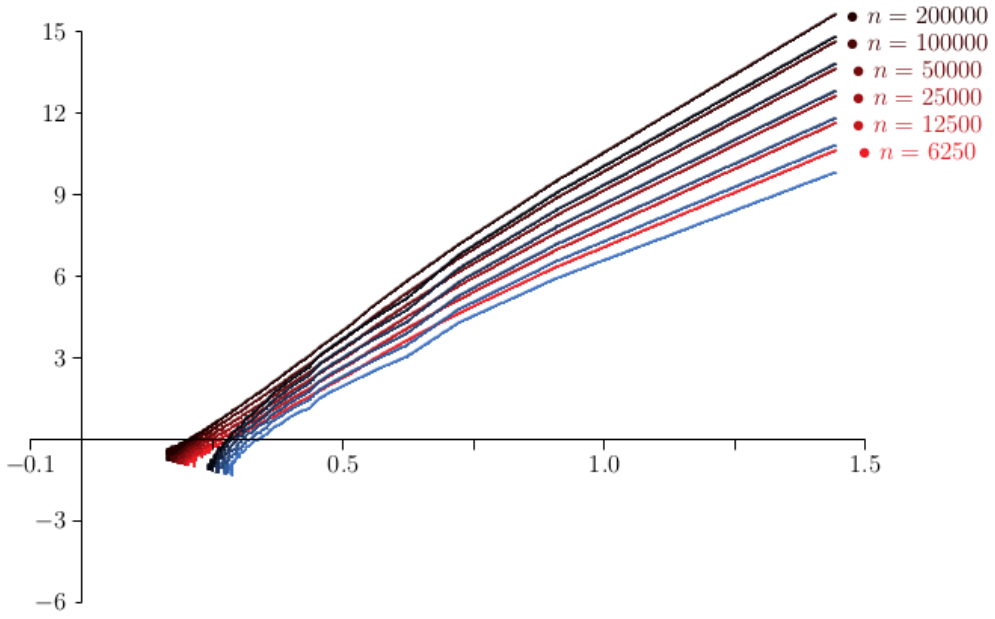


Figure 4.5: *Variant Barabasi-Albert networks with* $p = 0$*:*

Data on networks generated by the variant Barabasi-Albert algorithm with $p = 0$ *and* $\alpha = 1.1$ *(red curves) and* $\alpha = 0.5$ *(blue curves). In each case* 100 *networks were grown and the average degree sequence* $P_n(k)$ *computed. The curves above are plots of* $\log P_n(k)/\log(k+1)$ *against* $1/\log(k+1)$ *for* $n \in \{6250, 12500, 25000, \cdots, 200000\}$.

In figure 4.5 data for networks with $p = 0$ and $\alpha = 1.1$ and $\alpha = 0.5$ is shown. Since $\alpha = 1.1$ is still very close to 1, the results above show that these networks should still appear scale-free, and with connectivity a constant. This is indeed the case. For $n = 6250$ the data gives $\langle k \rangle_n = 3.149$, and increasing n to $n = 200000$ gives $\langle k \rangle_n = 3.176$. That is, the connectivity of the networks is insensitive to n over this range. Least squares fits to the curves with quadratic polynomials in order to determinate the value of γ give the average $\gamma = 2.857 \pm 0.068$. This result is consistent with a constant value of the connectivity of networks of these size ranges. With increasing n, it is expected that γ will decrease in value (that is, the value given here is an effective value), and eventually, the connectivity will

start to increase.

Networks generated with $p = 0$ and $\alpha = 0.5$ turned out to be sparse with low connectivity. For example, for $n = 100000$, the connectivity is $\langle k \rangle_n = 1.036$ and this decreases even further for $n = 200000$, where $\langle k \rangle_n = 1.020$. Attempts to extract an exponent γ from the data for these networks were not successful, the regressions did not settle on a value, but are strongly dependent on n. Notice that the mean field analysis above does not apply to networks with $\alpha < 1$.

Putting $\alpha = 2$ gives networks with average connectivity which increases with n. For example, if $n = 100$, then $\langle k \rangle_n = 43$, for $n = 500$, $\langle k \rangle_n = 260$ and for $n = 1000$, $\langle k \rangle_n = 527$. On the other hand, if $\alpha = \frac{3}{2}$, then $\langle k \rangle_n = 3.08$ if $n = 100$, $\langle k \rangle_n = 3.27$ if $n = 500$, and $\langle k \rangle_n = 3.31$ if $n = 1000$, and it appears that for small values of n the connectivity does not change quickly with increasing n.

4.2 Duplication-Divergence networks

Biological models of protein evolution are usually presented in terms of two processes, namely (1) a *duplication event* involving a gene sequence in DNA, and (2) a *(random) mutation* of duplicated genes which then drift from one another in genetic space [8, 18, 32]. The mutations of duplicated and mutated genes change the proteome and the network of protein interactions: If the protein is self-interacting, then the duplicated proteins interact, and the mutated genes code for proteins with altered interactions (some gained, others weakened or lost) with other proteins.

The Duplication-Divergence algorithm models these processes in order to grow a network, and was used to estimate the rates of duplication and mutation in the protein interaction networks [29]. There is a rich and large literature reporting on modeling protein interaction networks using models which include processes of duplication and divergence [14, 21, 24, 28].

Since proteomic networks appear to be scale-free [13, 22], it seems likely that duplication and divergence processes should grow scale-free networks and that this should also be seen in computer algorithms which grow networks using duplication and divergence elementary moves. Duplication can be implemented by selecting nodes and duplicating them, and their incident bonds, in a network. Divergence is implemented by altering the bonds incident on particular nodes, namely either by deleting, adding or moving bonds.

In the Duplication-Divergence algorithm these moves are implemented by selecting nodes uniformly for duplication to progenitor-progeny pairs, and by deleting bonds incident to either the progenitor node or its progeny. Notice that since nodes of high degree have a larger probability of being adjacent to a node selected for duplication, these nodes have a larger probability of receiving new bonds in the duplication process – in this way there are events of preferential attachment in this algorithm [12, 25].

The basic elementary moves of the Duplication-Divergence algorithm (duplication and divergence) are illustrated in figure 4.6.

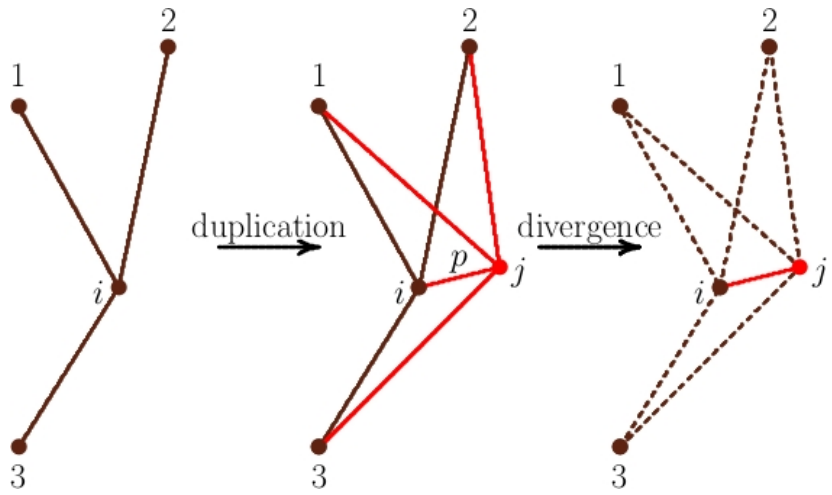


Figure 4.6: *The Duplication-Divergence algorithm: Duplication-Divergence iterations: A node i and its incident bonds are duplicated to create a node j with its incident bonds. The bond $\langle i{\sim}j\rangle$ is added with probability p. In the divergence step one of the pair of bonds $(\langle i{\sim}m\rangle, \langle j{\sim}m\rangle)$ is deleted with probability q, for each value of $m \in \{1,2,3\}$.*

The algorithm is implemented as follows:

Duplication-Divergence algorithm:

1. Initiate the network with one node x_0 and apply the following steps iteratively;

2. Duplication: Choose a node v uniformly and duplicate by creating node v';

3. For all bonds $\langle w{\sim}v\rangle$ incident with v, add the bonds $\langle w{\sim}v'\rangle$;

4. With probability p add the bond $\langle v{\sim}v'\rangle$;

5. Divergence: delete one bond of the pair $\{\langle w{\sim}v\rangle, \langle w{\sim}v'\rangle\}$ incident with v or with its duplicated node v' with probability q (for each w adjacent to both v and v' independently);

6. Stop the algorithm when a network of order N is grown.

The algorithm has two parameters (p, q). The parameter p is the probability that the protein corresponding to the progenitor node v is self-interacting. If it is (with probability p) then the bond $\langle v \sim v' \rangle$ is added to the network and it represents the interaction between v and v'.

The parameter q controls the divergence in this algorithm. As v and v' diverge from one another, one bond in each pair of bonds incident with v and v' is lost independently, with probability q. The result is that the network mutates as bonds (interactions) are lost (while they are created by the duplication process).

A slightly modified algorithm is found by changing step 5 in the algorithm to find a modification of the Duplication-Divergence algorithm which assumes that one of the duplicated pair mutates, while the other remains stable.

5. Divergence: Consider all bonds $\langle w \sim v' \rangle$ incident with the duplicated node v' and delete these independently with probability q.

The Duplication-Divergence algorithm tends to grow disconnected networks, while the Modified Duplication-Divergence algorithm is more likely to grow networks with a single component (that is, connected networks).

4.2.1 Mean field theory for Duplication-Divergence networks

Let $k_j(n)$ be the degree of node j after n iterations. The algorithm appends nodes by duplicating them (the probability that a node v is duplicated in a network of order n is $\frac{1}{n}$), adds bonds by inserting a bond between a node and its duplicate with probability p,

and removes bonds by selecting one bond between node-duplicate pairs and other nodes independently and deleting it with probability q. Let $2E_n = \sum_j k(n)$ be twice the total number of bonds after n iterations. Then, if $k_j(n)$ is the degree of node j at time n, and node j is duplicated, the number of bonds in the network E_n increases in the mean field by

$$E_{n+1} = E_n + p + k_j(n) - q\, k_j(n). \tag{4.32}$$

This follows since $k_j(n)$ bonds are created in the duplication move in the mean field, and another bond is created between the j-th node and its duplicate with probability p. The number of deleted bonds in the mean field is $q\, k_j(n)$.

Notice that $2E_n = \sum_j k_j(n) = n a_n$ where $a_n = \langle k_j(n) \rangle$ is the average degree. In the mean field approximation one substitutes $k_j(n)$ in the recurrence (4.32) by its network average a_n. Then equation (4.32) can be casted as a recurrence for a_n:

$$(n+1)\, a_{n+1} = n\, a_n + 2p + 2(1-q)\, a_n. \tag{4.33}$$

Let $n \to t$, where t is a continuous time variable, and approximate this recurrence by the differential equation

$$t \tfrac{d}{dt} a_t = 2p + (1-2q)\, a_t. \tag{4.34}$$

The initial condition is $a_1 = 1$, and this has solution

$$a_t = \tfrac{1-2(q-p)}{1-2q}\, t^{1-2q} - \tfrac{2p}{1-2q}. \tag{4.35}$$

Since $E_n \simeq \frac{1}{2} n\, a_n$, it follows that

$$E_n = \tfrac{1-2(q-p)}{2(1-2q)}\, n^{2(1-q)} - \tfrac{pn}{1-2q}. \tag{4.36}$$

Comparison to equation (3.7) shows that, if $q < \frac{1}{2}$,

$$\gamma = 1 + 2q. \tag{4.37}$$

In this case $E_n = O(n^{2(1-q)}) + O(n)$ and that while $2(1-q) > 1$, the term $O(n)$ is a strong correction to the growth in E_n for even large values of n. In other words, the degree distribution $P(k)$ of the network will be strongly corrected from the powerlaw distribution in equation (3.1).

If $q = \frac{1}{2}$, then by solving equation (4.34), $a_t = 1 + 2p \log t$ (so that $a_1 = 1$). Since $E_n = \frac{1}{2} n\, a_n$, this shows that

$$E_n = \tfrac{1}{2} n + pn \log n, \quad \text{if } q = \tfrac{1}{2}. \tag{4.38}$$

In this case $\gamma = 2$ by equation (3.7), but notice the subtle domination by the $n \log n$ term. In numerical work this will be very hard to see.

The case $q > \frac{1}{2}$ is considered by noting that $a_t \simeq \frac{2p}{2q-1}$ as $t \to \infty$. This shows that

$$E_n \simeq \tfrac{p\, n}{2q-1}, \quad \text{if } q > \tfrac{1}{2}. \tag{4.39}$$

This shows that $\gamma \geq 2$ by equation (3.7).

Putting the above together gives

$$\gamma \begin{cases} = 1 + 2q, & \text{if } q \leq \frac{1}{2}; \\ \geq 2, & \text{if } q > \frac{1}{2} \end{cases} \tag{4.40}$$

with a logarithmic correction if $q = \frac{1}{2}$.

Comparing the coefficient in equation (3.7) with equation (4.39) gives a refined estimate $\gamma = 1 + \frac{2p}{1+2p-2q} \geq 2$, provided that $2q < 1 + 2p$. For example, if $q = 0.75$ then $p > 0.25$. However, numerical work shows this estimate to be too small, and estimating γ in this regime for this model remains an open question.

The power law decrease in $P(k)$ in equation (3.1) is only asymptotic for this algorithm; and there should be corrections in particular for $q < \frac{1}{2}$. From the results above the average connectivity can be computed: Since $E_n = \frac{1}{2} n \langle k_j(n) \rangle$,

$$\langle k \rangle_n \simeq \begin{cases} \frac{1-2(q-p)}{1-2q}\, n^{1-2q} - \frac{2p}{1-2q}, & \text{if } q < \frac{1}{2}; \\ 2p \log n + 1, & \text{if } q = \frac{1}{2}; \\ \text{Constant}, & \text{if } q > \frac{1}{2}. \end{cases} \tag{4.41}$$

From these results $P(k)$ can be calculated. Since $\langle k \rangle_n \simeq \int_1^n k\, P(k)\, dk$, it follows that $\frac{d}{dn} \langle k \rangle_n =$

$n\,P(n)$. Thus, using this approach gives

$$P(k) \sim \begin{cases} (1-2(q-p))\,k^{-1-2q}, & \text{if } q < \frac{1}{2}; \\ 2p\,k^{-2}, & \text{if } q = \frac{1}{2}; \\ C_0\,k^{-\gamma}, & \text{if } q > \frac{1}{2}, \end{cases} \tag{4.42}$$

where the case $q > \frac{1}{2}$ is unknown since the dependence of the exponent γ on the parameters (p, q) is not known. Notice the change in behaviour at the critical value $q = \frac{1}{2}$; this was already observed numerically in reference [29].

The modified Duplication-Divergence algorithm has the same recurrence (4.35), and so the values for γ and relations for $\langle k \rangle_n$ and $P(k)$ remain unchanged for this algorithm. Notice that this implementation preserves the degree of the selected node, and tends to give a duplicated node with lower degree (while the (unmodified) implementation tends to lower the degrees of both the selected and duplicated nodes). As a result, networks generated with the modified algorithm have, on average, more nodes of degree equal to one (and so appear more tree-like).

4.2.2 Numerical results on Duplication-Divergence networks

In figure 4.7 two networks grown with the Duplication-Divergence algorithm are shown. Both networks were grown with $p = 1$ and have order 300. The network on the left was grown with divergence parameter $q = 0.4$, and that on the right, with the higher mutation rate $q = 0.6$.

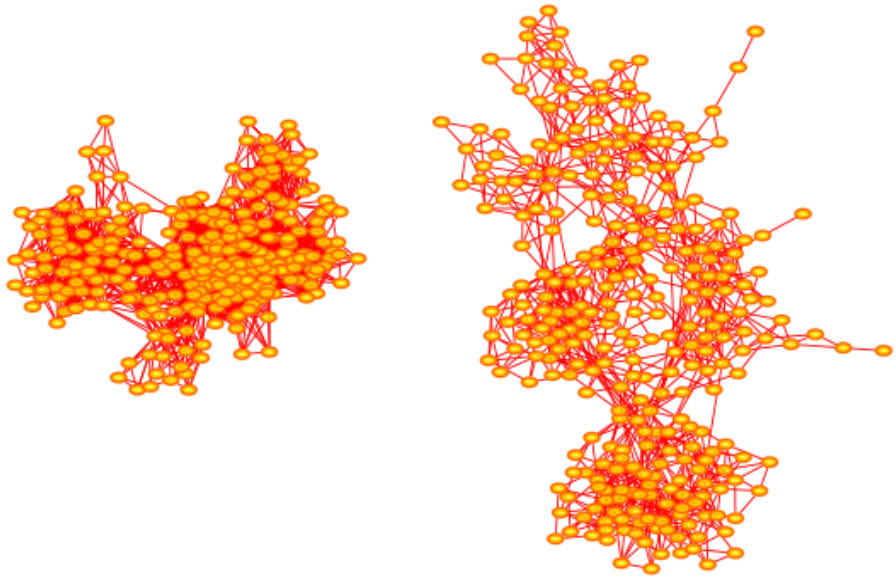


Figure 4.7: *Duplication-Divergence network:*

The network on the left is a network generated with $p = 1$ *and* $q = 0.40$*. It has order* 300 *and it has* 114 *nodes with degrees exceeding* $\sqrt{300}$ *and so qualify as hubs. The largest few of these hubs have degrees* $\{43, 45, 47, 47, 50\}$*. The network on the right is similarly a network generated with* $p = 1$ *and* $q = 0.60$*. It is more extended but has only three nodes of degree equal to one. Its order is* 300*, and it has* 5 *nodes of degrees* $\{18, 18, 19, 20, 23\}$ *which qualify as hubs. Networks generated with the Modified Duplication-Divergence algorithm have a similar appearance, with the exception that more nodes of degree* 1 *are seen. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].*

In figure 4.8 data for networks grown with $p = 0.75$ and $q = 0.4$ are shown. The curves on the right were obtained by plotting $\log P(k) / \log(k+1)$ averaged over 100 networks of sizes $\{3125, 6250, 12500, 25000, 50000, 100000, 200000\}$ against $1/\log(k+1)$. The mean field value of γ is denoted by the bullet on the left-hand axis. These data show that convergence to this value is very slow – this indicates strong corrections to scaling arising in equation (4.36), and noted after equation (4.37).

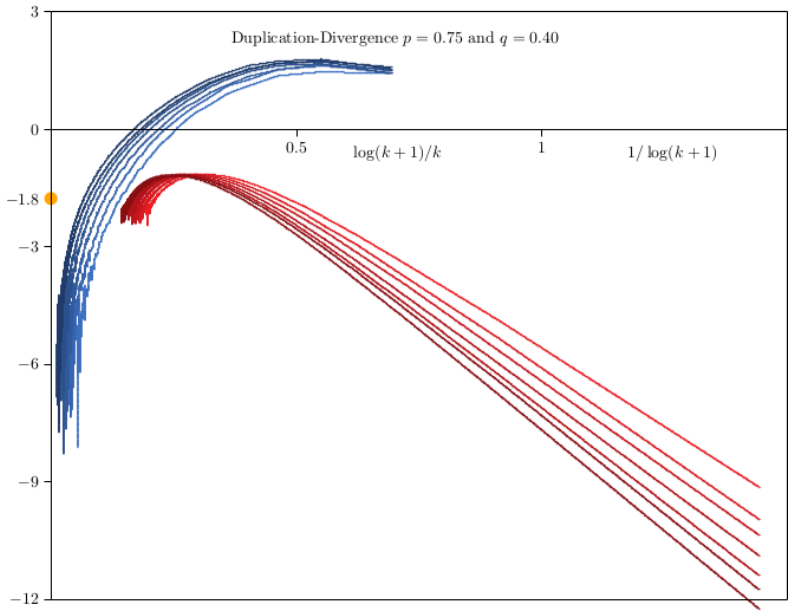


Figure 4.8: *The distribution of degrees in Duplication-Divergence networks with* $p = 0.75$ *and* $q = 0.40$*:*

Data on networks generated by the Duplication-Divergence algorithm. In each case 100 *networks were grown and the average degree sequence* $P_n(k)$ *computed. The curves on the right are plots of* $\log P_n(k)/\log(k+1)$ *against* $1/\log(k+1)$ *for* $n \in \{3125, 6250, 12500, \cdots, 200000\}$*, while those on the left are plots of* $(\log P(2k) - \log P(k))/\log 2$ *as a function of* $\log(k+1)/k$*. The mean field estimate for the exponent* γ *is marked at* $-\gamma = -1.8$ *on the left hand axis. The strong correction to scaling evident in these curves makes it difficult to extrapolate to the mean field value for* γ*.*

An alternative approach is to estimate γ by plotting $\zeta(k) = (\log P(2k) - \log P(k))/\log 2$ as a function of $\log(k+1)/k$ (see equation (3.4) with $\alpha = 2$). The results are also strongly curved data (left in figure 4.8), and while the results are not inconsistent with the mean field value $\gamma \approx 1.8$ in this model, however, it seems difficult to extrapolate these curves to a limiting value of γ. 2). The turnover of the curves around large k happens when k approaches n in these figures.

If $q = 0.60 > \frac{1}{2}$ then the results in figure 4.9 are seen.

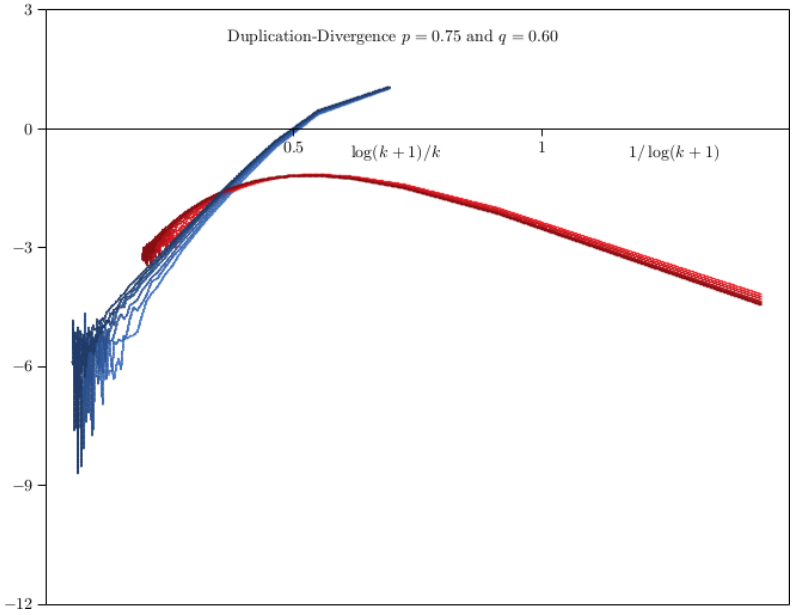


Figure 4.9: *The distribution of degrees in Duplication-Divergence networks with* $p = 0.75$ *and* $q = 0.60$*:*

Data on networks generated by the Duplication-Divergence algorithm. In each case 100 *networks were grown and the average degree sequence* $P_n(k)$ *computed. The curves on the right are plots of* $\log P_n(k)/\log(k+1)$ *against* $1/\log(k+1)$ *for* $n \in \{3125, 6250, 12500, \cdots, 200000\}$*, while those on the left are plots of* $(\log P(2k) - \log P(k))/\log 2$ *as a function of* $\log(k+1)/k$*. Each of these curves can be extrapolated by a quadratic least squares fit to obtain estimates of* γ*. This gives the estimates* γ_n *for* $n = 3125 \times 2^\ell$ *for* $\ell = 0, 1, 2, \ldots, 6$*.*

The curves of $\zeta(k) = (\log P(2k) - \log P(k))/\log 2$ as a function of $\log(k+1)/k$ have straightened considerably, and each can be extrapolated by a quadratic least squares to obtain an estimate γ_n for each value of $n = 3125 \times 2^\ell$ (for $\ell = 0, 1, 2, \ldots, 6$). This gives estimates $\{9.68, 8.52, 7.99, 7.95, 7.82, 7.58, 7.05\}$ which can be extrapolated by a least squares fit of $\gamma_n = \gamma + A/\log n$, giving the estimate $\gamma \approx 2.87$, which is slightly larger than the value predicted by the mean field formula $\gamma = 1 + \frac{2p}{1+2p-2q}$ (see the paragraph following equation (4.40)). This suggests that the approach to limiting behaviour in this model is quite slow, consistent with the remarks after equation (4.40) in the previous section.

The average connectivity $\langle k \rangle_n$ is expected to behave according to equation (4.41). In

table 4.1 $\langle k \rangle_n$ is listed for $p = 0.75$ and $q = 0.40$, $q = 0.50$ and $q = 0.60$.

Table 4.1: *Connectivity data for Duplication-Divergence Networks.*

n	$q = 0.4$	$q = 0.5$	$q = 0.6$
3125	25.9	11.4	5.93
6250	31.3	12.6	6.14
12500	37.3	13.6	6.33
25000	44.8	14.4	6.55
50000	51.9	15.5	6.64
100000	60.1	16.8	6.75
200000	70.3	17.7	6.88

If $q = 0.4$, then equation (4.41) suggests that $\langle k \rangle_n \simeq 8.5\, n^{0.2}$. Computing $\langle k \rangle_n \times n^{-0.2}$ from the data in table 4.1 gives $\{5.18, 5.45, 5.65, 5.91, 5.96, 6.01, 6.12\}$. Plotting these results against $1/\log n$ and then linearly extrapolating as $n \to \infty$ gives 7.98, close to the value of 8.5 predicted in equation (4.41).

If $q = 0.5$, then equation (4.41) suggests that $\langle k \rangle_n \simeq 1.5 \log n$ since $p = 0.75$. Dividing the results in table 4.1 by $\log n$ for each value of n gives the results $\{1.42, 1.44, 1.44, 1.42, 1.43,$ $1.46, 1, 45\}$. The average of this is close to the predicted value of 1.5.

Finally, if $q = 0.6$ then the data appear to approach a constant. Extrapolating these results using the model $A + B/\log n$ gives the estimated limiting value 8.72. By equation (3.5) this indicates that $\gamma = 2.13$, a value which is quite close to 2.15, the value predicted by the formula $\gamma = 1 + \frac{2p}{1+2p-2q}$ in the paragraph following equation (4.40).

4.3 Solé evolutionary networks

The Solé model [25, 26] modifies the Duplication-Divergence model by using duplication and network rewiring as the basic elementary moves. As before, the duplication of nodes is an implementation of gene duplication, and the network rewiring is based on the loss and gain of protein interactions in the bulk of the network [6]. Thus, the algorithm grows networks based on a model of gene duplication and the rewiring of protein interactions; both these processes drive the evolution of the interactome.

The elementary move of the algorithm is as follows: A node in the network is chosen uniformly and randomly, and duplicated to form a progenitor-progeny pair. The progeny will have the same interactions as the progenitor. This network is updated in the rewiring step which has two parts: Bonds incident with the progeny protein are deleted with probability δ, and new bonds, added in the network between nodes (excluding the progenitor protein) are created with probability α. This implementation differs in two ways from the Duplication-Divergence algorithm. In the Solé model there are no self-interacting nodes, and the formation of new bonds in the rewiring steps only occurs in the Solé model.

The basic iterative step of the Solé algorithm is shown in figure 4.10 and a Solé evolutionary network of order N nodes is grown as follows:

Solé evolutionary algorithm:

1. Initiate the network with one node x_0 and apply the following steps iteratively;

2. Choose a node v uniformly and duplicate it to a new node v';

3. For each bond $\langle w{\sim}v \rangle$ incident with the chosen node v, add the bond $\langle w{\sim}v' \rangle$ incident with the duplicated node v';

4. Delete each bond $\langle w{\sim}v' \rangle$ added in step 3 with probability δ independently;

5. For all nodes u not adjacent to the chosen node v, create the bond $\langle u{\sim}v' \rangle$ with probability α;

6. Stop the algorithm when a network of order N is grown.

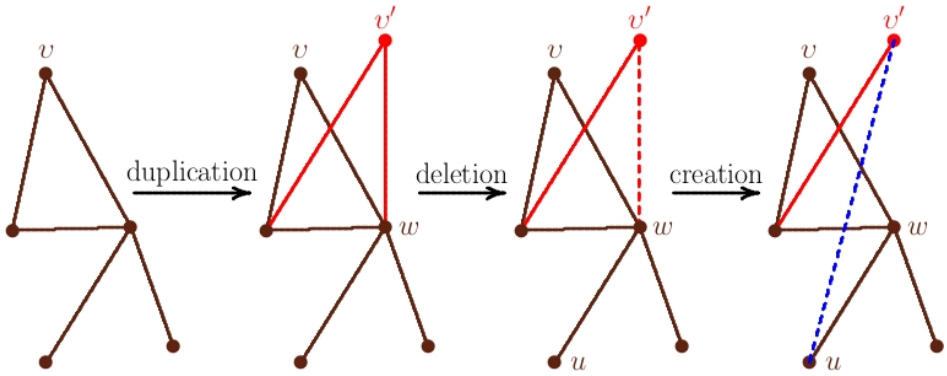


Figure 4.10: *The Solé evolutionary algorithm:*

The duplication-deletion-creation iterations of the Solé algorithm. A node is duplicated, some bonds incident on it are deleted with probability δ and new bonds incident on it are created with probability α.

The algorithm has two parameters (δ, α). If $\delta = 0$ and $\alpha = 1$ then the algorithm grows complete simple networks. More generally, if $\alpha > 0$ then on average roughly αN bonds are added to a network of order N. This shows that the algorithm grows networks of size $O(N^2)$ – that is, Solé networks are dense.

4.3.1 Mean field theory for Solé networks

Let E_n be the total number of bonds in a Solé network after n iterations of the algorithm, and let $\langle k\rangle_n$ be the connectivity of the network (that is, the average degree of nodes) after n iterations (so that $2\,E_n = n\langle k\rangle_n$). In the mean field approximation the node in step 2 of the algorithm has degree $\langle k\rangle_n$ and this number of bonds is added in step 3, while, in a similar way, $\delta\langle k\rangle_n$ bonds are removed in step 4. In step 5 there are $n - \langle k\rangle_n$ choices in the mean field for the node u not adjacent to v' and each bond $\langle u{\sim}v'\rangle$ is added with probability α. This shows that the number of bonds after $n+1$ iterations is given by the recurrence relation

$$E_{n+1} = E_n + (1-\delta)\langle k\rangle_n + \alpha(n - \langle k\rangle_n). \tag{4.43}$$

Since $2\,E_n = n\langle k\rangle_n$ this becomes

$$E_{n+1} - E_n = \alpha n + \frac{2}{n}(1 - \delta - \alpha)\,E_n, \tag{4.44}$$

which is a mean field recurrence relation for E_n.

Taking $n \to t$, a continuous time variable, and approximating E_n by E_t, and approximating the finite difference as a derivative, gives the following differential equation for E_n:

$$\tfrac{d}{dt}E_t = \alpha t + \tfrac{2}{t}(1 - \alpha - \delta)\,E_t. \tag{4.45}$$

Solving this equation and letting $t \to n$ again gives the approximate mean field solution

for E_n:

$$E_n \approx \frac{\alpha n^2}{2(\alpha+\delta)} + \frac{(\alpha+2\delta)n^{2(1-\alpha-\delta)}}{2(\alpha+\delta)}. \tag{4.46}$$

Equation (4.46) shows that the number of bonds is proportional to αn^2, so that networks created by this algorithm are dense, except when $\alpha = 0$ and $\delta \to 0$. Comparison to equation (3.7) suggests that $\gamma \leq 1$ in this model if $\alpha > 0$. Notice that there is no logarithmic factor in the denominator, and that $E_n = \Theta(n^2)$. This is consistent with a mean field value $\gamma < 1$ (and this requires that $P_n(k)$ be modified so that it is a normalizable probability distribution). With these results, it is reasonable to expect that, in the mean field,

$$\gamma \leq 1. \tag{4.47}$$

If $\alpha = 0$ then equation (4.46) gives $E_n \sim n^{2-2\delta}$ and comparison to equation (3.7) gives

$$\gamma = 1 + 2\delta, \quad \text{if } \alpha = 0 \tag{4.48}$$

since the algorithm grows clusters which are not $O(n^2)$ in this case.

4.3.2 Numerical results for Solé networks

Similar to Barabasi-Albert and Duplication-Divergence networks, Solé networks can be grown numerically by implementing the algorithm as given above, using sparse matrix routines to efficiently store the adjacency matrix of the network. The larger size of networks makes these more difficult to grow, and our algorithms sampled efficiently to networks of

size 51,200 bonds.

Solé networks are rich in bonds. This is seen, for example, in equation (4.46), which shows that $E_n \propto n^2$ if $\alpha > 0$. In figure 4.11 two examples of networks generated by the Solé algorithm are shown. If $\delta < 0.5$, then the networks have a dense appearance dominated by a few hubs. If $\delta > 0.5$, then the networks appear more extended, often with no nodes qualifying as hubs under the definition that the degree of a hub in a network of order n is at least $\lfloor \sqrt{n} \rfloor$. The networks in figure 4.11 were generated with $\alpha = 0.005$, and increasing the value of α quickly increases the number of bonds.

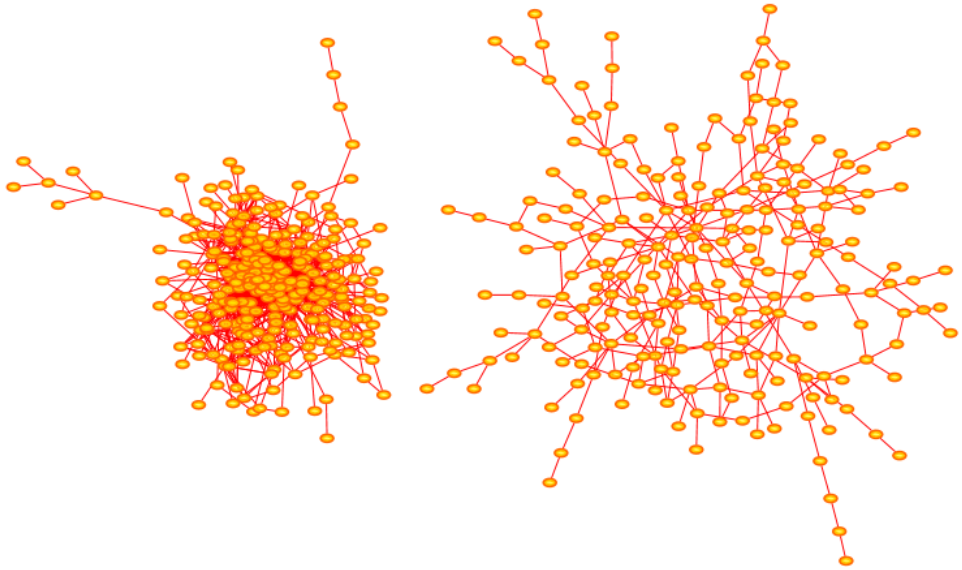


Figure 4.11: *Solé evolutionary networks:*

The network on the left was generated with $\delta = 0.25$ and $\alpha = 0.005$. Its has order 279 and has 47 nodes with degrees exceeding $\sqrt{279}$ and so qualify as hubs. The largest few of these hubs have degrees $\{40, 41, 62, 80\}$. This algorithm creates dense networks as seen here, even for small values of α. Increasing the value of δ gives more extended networks. The network on the right was generated with $\delta = 0.75$ and $\alpha = 0.005$ and grown to order 230. None of its nodes qualify as hubs. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].

The mean field result that $\gamma \leq 1$ has implications for the scaling of Solé networks. In

particular, $P_n(k)$ in equation (3.1) is not normalizable for infinite networks if $\gamma \leq 1$ and so is not a valid candidate degree distribution in this model. The degree distribution can be modified to

$$P(k) \simeq C_o\, k^{-\gamma} D(n^{-\phi} k) \tag{4.49}$$

where $D(x)$ is a function of the combined (or scaled) variable $x = n^{-\phi}k$. That is, as $n \to \infty$, k is rescaled by $n^{-\phi}$ and $k^{\gamma}P(k)$ approaches a limiting distribution proportional to $D(x)$.

This can be tested numerically by plotting $k^{\gamma}P(k)$ as a function of $x = n^{-\phi}k$. For the proper choices of γ and ϕ it is expected that $k^{\gamma}P(k) \simeq C_o D(x)$ for a wide range of values of n (that is, the data should approach a limiting curve as $n \to \infty$). The result is shown in figure 4.12 for $(\delta = 0.25, \alpha = 0.005)$ and $(\delta = 0.75, \alpha = 0.005)$. These are plots on the same graph for $n = 100 \times 2^n$ for $n \in \{6, 7, 8, 9\}$ (other curves at smaller values of N are left away to give a clearer picture).

The data for $\delta = 0.75$ are the cluster of peaks to the left (blue curves), rescaled by choosing $\phi = 1$ and $\gamma = \frac{1}{2}$, while the cluster of peaks to the right is for $\delta = 0.25$ with $\phi = 1$ and $\gamma = \frac{2}{3}$. With increasing n the data appear to approach a single underlying curve if $\gamma = \frac{1}{2}$ in the one instance, and $\gamma = \frac{2}{3}$ in the other instance. Both these values are consistent with the mean field expectation that $\gamma \leq 1$ in this model. Further refinements in this scaling assumption may be necessary, since the curves are still becoming narrower with increasing n. It is not clear that these approach a limiting curve as $n \to \infty$, although the data for $\delta = 0.75$ suggest this to be the case. In these cases the curves are sharply peaked with a mean of about 0.02 if $\delta = 0.25$ and about 0.007 if $\delta = 0.75$.

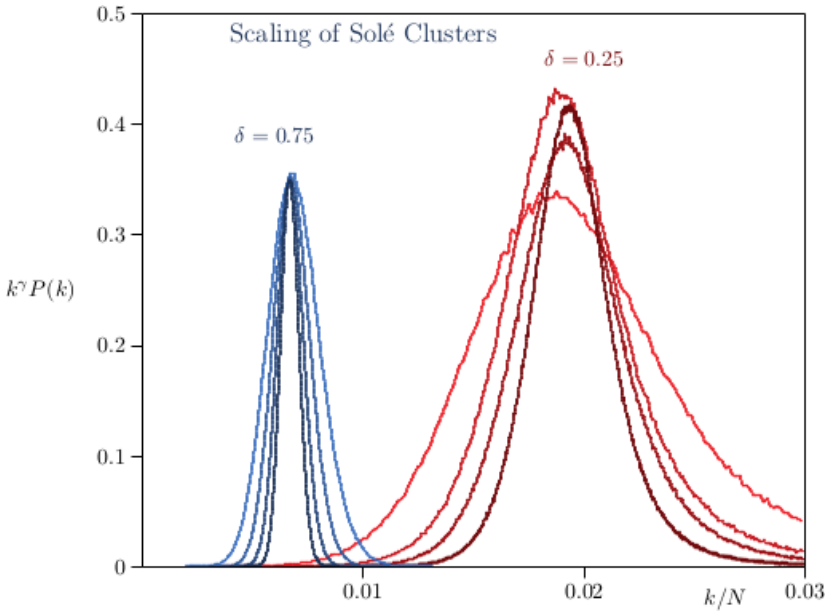


Figure 4.12: *Scaling of Solé evolutionary networks:*

Plotting $k^\gamma P_N(k)$ against $N^{-\phi}k$ for networks generated by the Solé Evolutionary algorithm gives the distributions above. On the left the results are shown for networks grown with $\delta = 0.75$ and $\alpha = 0.005$. The choices $\gamma = 1/2$ and $\phi = 1$ uncovers a distribution as shown where the order of the networks are $N = 100 \times 2^n$ for $n = 6, 7, 8, 9$. A similar distribution, but with $\gamma = 2/3$ and $\phi = 1$, is seen when networks are grown with $\delta = 0.25$ and $\alpha = 0.005$. It is not known that the value of γ changes discontinuously as δ increases from 0.25 to 0.75.

Since the curve $D(x)$ is sharply peaked at a constant value c_o of the rescaled variable x, the connectivity of Solé networks is estimated by treating $D(x)$ as concentrated at c_o and then (assuming that $\phi = 1$ and approximating the connectivity)

$$\langle k \rangle_n \sim \frac{\int_0^\infty k^{1-\gamma} D(k/n^\phi)\, dk}{\int_0^\infty k^{-\gamma} D(k/n^\phi)\, dk} \sim \frac{(n^\phi)^{2-\gamma}}{(n^\phi)^{1-\gamma}} \sim n^\phi. \tag{4.50}$$

In other words, the connectivity of Solé networks should increase linearly with n^ϕ (and since $\phi = 1$, linearly with n). In table 4.2 the connectivities of Solé networks for $\delta = 0.25$ and $\delta = 0.75$ (with $\alpha = 0.005$) are listed.

Table 4.2: *Connectivity data for Solé Networks.*

n	$\delta = 0.25$	$\delta = 0.75$
100	2.95	1.50
200	4.46	1.94
400	7.59	2.94
800	14.75	5.36
1600	30.46	10.64
3200	59.94	21.26
6400	122.78	45.57
12800	245.35	85.18
25600	496.87	170.35
51200	994.54	340.76

Non-linear least squares fits to the data show that $\phi = 1.01$ when $\delta = 0.25$ and $\phi = 0.99$ when $\delta = 0.75$. That is, these results are consistent with the value $\phi = 1$ seen above.

4.4 The iSite model of network evolution

Protein interaction networks evolve by mutations in proteins which change the interactions of the proteins in the network. In the Duplication-Divergence algorithm, a mutated protein loses its interactions randomly. This random deletion of interactions is a good first order approximation to the evolution of networks. The iSite model refines this by giving structure to nodes in the network by introducing *iSites* on nodes as localities of the interaction sites on a protein [15, 16]. Subfunctionalization of interaction sites in the iSite model is implemented by silencing iSites, and adding interactions with reduced probability if the iSite is not silenced.

The implementation of the iSite algorithm relies in the first place on duplication of

nodes, and then subfunctionalization of iSites on the nodes. The subfunctionalization of iSites is implemented by randomly deleting of bonds incident to duplicated iSites, *and* by the silencing of iSites by turning them off. These processes are models of random mutations which cause the loss of information in the genome (and leave behind non-coding remnants of genes). A process of spontaneously creating new iSites is not in the iSites algorithm, although this is a possible refinement which may be introduced. The elementary move of the iSite algorithm is illustrated schematically in figure 4.13.

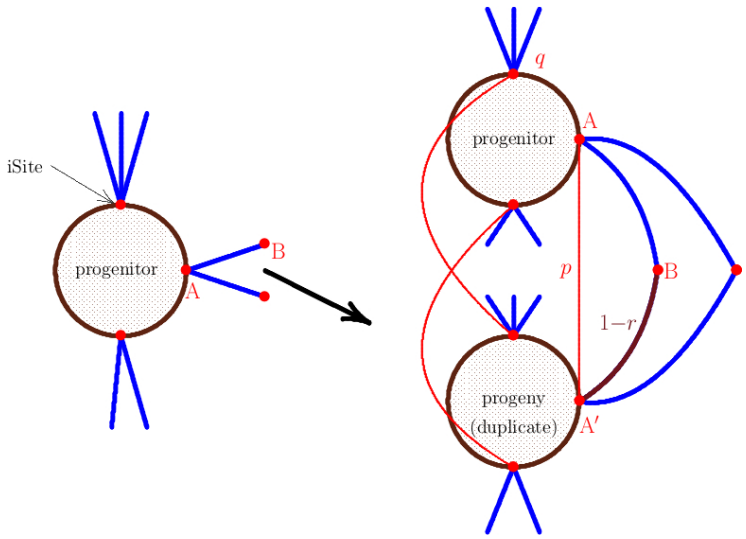


Figure 4.13: *The iSite evolutionary algorithm:*

The duplication-deletion iterations of the iSite algorithm. A node together with its iSites is duplicated, and some bonds incident with the duplicated iSites are deleted with probability r. New bonds between a self-interacting iSite and its duplicate are inserted with probability p, and iSites are silenced with probability q.

A uniformly chosen node is duplicated into a progenitor-progeny pair (and so also duplicating the iSites of the progenitor onto the progeny). If the duplicated iSite is self-interacting, then bonds are added between the iSite on the progenitor and the duplicated iSite on the progeny with probability p – this allows for subfunctionalization of the dupli-

cated iSites. Bonds incident with the iSites on the progenitor are duplicated with reduced probability $1 - r$, and iSites on the progenitor or progeny nodes are silenced with probability q. If an iSite is silenced, then all bonds incident with it are deleted. Notice that subfunctionalization enters in several ways, both in the duplication of self-interacting iSites, in the duplication of bonds, and in the silencing of iSites.

iSite evolutionary algorithm:

1. Initiate the network with one node x_0 with I active iSites (each of which is self-interacting with probability p) and iterate the following steps;

2. Choose a progenitor protein v uniformly in the network and duplicate it, and its associated iSites A, to a successor protein v' with duplicated iSites A';

(a) A duplicated iSite $A' \in v'$ is active with probability $1 - q$ if it is duplicated from an active iSite on $A \in v$, and silenced otherwise;

(b) An active duplicated iSite $A' \in v'$ is self-interacting with probability p if it is duplicated from a self-interacting iSite on $A \in v$, and not self-interacting otherwise;

(c) If a silenced iSite A is duplicated to iSite A', then A' is also silenced;

3. Add bonds as follows:

(a) If iSite $A \in v$ is self-interacting and A is duplicated to iSite $A' \in v'$, then add the bond $\langle A{\sim}A' \rangle$ if A' is not silenced;

(b) If $\langle A{\sim}B \rangle$ is a bond incident with iSite A on the progenitor v, and A is duplicated to iSite A' on the duplicate v', then $\langle A{\sim}B \rangle$ is duplicated to $\langle A'{\sim}B \rangle$ with

probability $1 - r$ provided that A' is not silenced;

4. Iterate the algorithm from step (2) and stop the iterations when a network of order N is grown.

4.4.1 Mean field theory for the iSite model

Let nodes in the network correspond to proteins, and let $i_j(n)$ be the number of active iSites on node j after n iterations of the algorithm. Denote the degree of node j by $k_j(n)$ (that is the total number of bonds with one end-point in node j), and let E_n be the number of bonds of the network. Then $2E_n = \sum_j k_j(n)$.

The average number of active iSites per node is $i(n) = \frac{1}{n}\sum_j i_j(n)$. With each iteration $i(n)$ iSites are created, of which $qi(n)$ are silenced, in the mean field. This gives the following recurrence relation for $i(n)$:

$$(n+1)\, i(n+1) = n\, i(n) + (1-q)\, i(n). \tag{4.51}$$

The exact solution of this recurrence is

$$i(n) = \frac{i(0)\,\Gamma(1-q+n)}{n!\,\Gamma(1-q)} \tag{4.52}$$

where Γ is the gamma function with the property that $\Gamma(x+1) = x\,\Gamma(x)$ and $\Gamma(1) = 1$. Notice that $i(0) = I$, where I is the number of iSites on the source node x_0.

For large n the Γ-function and the factorial have well known asymptotics (namely the Stirling approximation [31]), so that

$$i(n) \simeq \frac{I\, n^{-q}}{\Gamma(1-q)}. \tag{4.53}$$

This shows that with increasing n the total number of iSites grows proportionally to n^{1-q}. If $q = 0$, then this is linear in n since no iSites become silenced, and if $q = 1$, then the number approaches a constant.

The total number of bonds in the network increases after n iterations by the recurrence

$$E_{n+1} = E_n + \frac{2(1-r)}{n}\, E_n + p\, i(n), \tag{4.54}$$

since there are on average $\frac{2}{n}E_n$ bonds incident to each node, and the probability that each one of them is duplicated is $1 - r$, and there are on average $i(n)$ iSites per node, and the probability that each of these is self-interacting is p.

Using the asymptotic solution for $i(n)$ and approximating this recurrence by a differential equation gives

$$\tfrac{d}{dt}\, E_t = \frac{2(1-r)}{t}\, E_t + \frac{pI}{\Gamma(1-q)}\, t^{-q}. \tag{4.55}$$

This equation can be solved, and using the initial condition $E_1 = 0$, the result is

$$E_t = \frac{pI}{(1+q-2r)\,\Gamma(1-q)} \left(t^{2-2r} - t^{1-q} \right). \tag{4.56}$$

Thus, the average degree of a node is equal to $\frac{2}{n}E_n$, so that the connectivity of iSite evolutionary networks is given by

$$\langle k\rangle_n \simeq \frac{2pI}{(1+q-2r)\,\Gamma(1-q)}\left(n^{1-2r}-n^{-q}\right) \tag{4.57}$$

in the mean field. This shows that the large n value of $\langle k\rangle_n$ is dominated by the larger of $-q$ and $1-2r$. In particular, if $r<\frac{1}{2}(1+q)$, then $\langle k\rangle_n \sim n^{1-2r}$. If $r>\frac{1}{2}(1+q)$, then $\langle k\rangle_n \sim n^{-q}$.

By equation (3.7) one may determine the mean field value of γ for this model:

$$\gamma = \begin{cases} 1+2r, & \text{if } r<\frac{1}{2}(1+q); \\ \\ 2+q, & \text{if } r>\frac{1}{2}(1+q). \end{cases} \tag{4.58}$$

If $2r=1+q$, then a different solution is obtained, namely

$$E_t = \frac{pI}{\Gamma(1-q)}\,t^{1-q}\log t. \tag{4.59}$$

This shows that $\gamma = 2+q$ in this case as well, but there is also a logarithmic correction to the growth of $E(t)$, and so there is a logarithmic factor in the expression for $\langle k\rangle_n$.

4.4.2 Modified iSite evolutionary algorithm

The subfunctionalization of proteins can be refined by introducing in the iSite algorithm the probability of creating new iSites on the progeny node with a probability s. This changes the algorithm as follows.

Modified iSite evolutionary algorithm: Implement the algorithm as above but introduce the parameter s and create new active iSites by replacing step 2 in the iSite evolutionary algorithm by

2. Choose a progenitor node v uniformly in the network and duplicate it, and its associated iSites A, to a progeny node v' with duplicated iSites A';

(a) A duplicated iSite $A' \in v'$ is active with probability $1-q$ if it is duplicated from an active iSite on $A \in v$, and silenced otherwise;

(b) An active duplicated iSite $A' \in v'$ is self-interacting with probability p if it is duplicated from a self-interacting iSite on $A \in v$, and not self-interacting otherwise;

(c) If a silenced iSite A is duplicated to iSite A', then A' is also silenced;

(d) With probability s create an active iSite C on the progeny node v', where C is self-interacting with probability p.

The recurrence for the average number of active iSites per node $i(n)$ (see equation (4.52)) is modified to

$$(n+1)\, i(n+1) = n\, i(n) + (1+s-q)\, i(n) \tag{4.60}$$

in the Modified iSite evolutionary algorithm. The exact solution is obtained by replacing q by $q-s$ in equation (4.52), and the asymptotic approximation of the solution is given by

$$i(n) \simeq \frac{I\, n^{s-q}}{\Gamma(1+s-q)}, \tag{4.61}$$

as seen in equation (4.53).

The total number of bonds in the network, E_n, still satisfies equation (4.54), and so it follows from equations (4.56), (4.57) and (4.58), that for the modified iSite evolutionary algorithm (notice the condition that $q < r + s$):

$$E_n = \frac{pI}{(1+q-s-2r)\,\Gamma(1+s-q)} \left(n^{2-2r} - n^{1+s-q} \right). \tag{4.62}$$

This shows that the connectivity of Modified iSite networks is given by

$$\langle k \rangle_n \simeq \frac{2pI}{(1+q-s-2r)\,\Gamma(1+s-q)} \left(n^{1-2r} - n^{s-q} \right). \tag{4.63}$$

The value of the scaling exponent is seen from above to be given by

$$\gamma = \begin{cases} 1 + 2r, & \text{if } r < \frac{1}{2}(1 + q - s); \\ \\ 2 + q - s, & \text{if } r > \frac{1}{2}(1 + q - s) \end{cases} \tag{4.64}$$

with a correction factor in the expression for $\langle k \rangle_n$ if $2r = (1 + q - s)$.

4.4.3 Numerical results for iSite networks

The iSite algorithm was coded and networks were grown to compute averaged statistics. Examples of iSite networks generated by the algorithm are shown in figure 4.14.

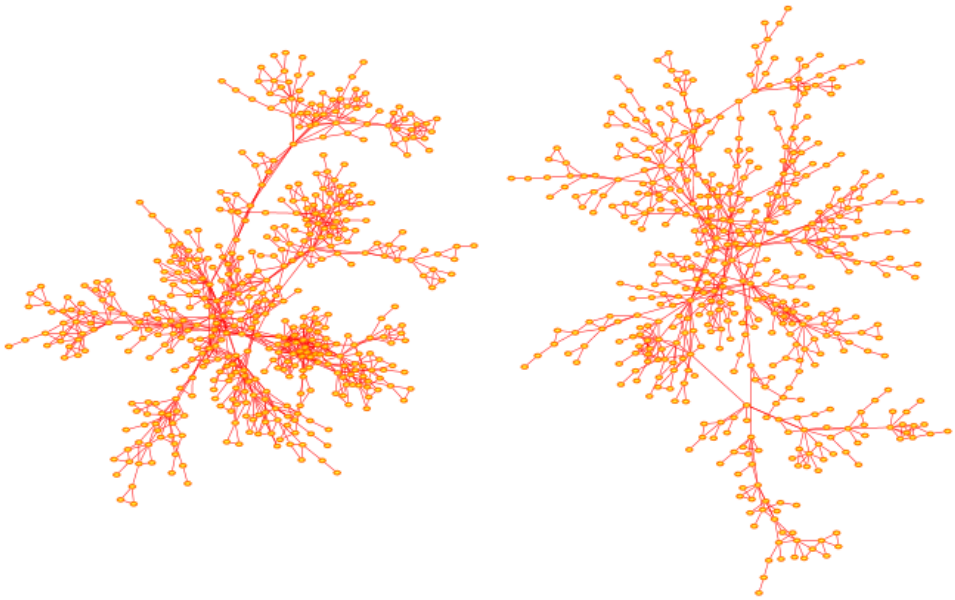


Figure 4.14: *iSite evolutionary networks:*

The network on the left was generated with 4 *iSites per node,* $p = 0.5$, $q = 0.1$ *and* $r = 0.8$, *and the network on the right was generated with* 2 *iSites per node, and with* $p = 0.5$, $q = 0.1$ *and* $r = 0.8$. *The order of the network on the left is* 501 *and on the right,* 491. *The network on the left has two nodes qualifying as hubs, of degrees* $\{23, 25\}$, *while the network on the right has none. The arrangement of nodes and bonds in these networks was created using the prefuse force directed layout in Cytoscape 3.4.0 [11].*

The algorithm was then used to sample networks of order up to $200,000$. The connectivity $\langle k \rangle_n$ of iSite networks for $I = 3$ iSites per node, and with $p = 0.5$, $q = 0.4$ and $r = 0.3$, is shown in table 4.3. By equation (3.5), $\log\langle k \rangle \simeq \log \frac{\gamma-1}{2-\gamma} + (2-\gamma)\log n$. Least squares fit to the data in Column 2 gives $\log \frac{\gamma-1}{2-\gamma} \approx 1.0211$, and $(2-\gamma) = 0.258$. Solving for γ gives in the first instance $\gamma = 1.735$ and in the second $\gamma = 1.742$. Since $2r < 1 + q$ in this case, the mean field value of γ is $\gamma = 1 + 2r = 1.6$, close to these estimated values.

Table 4.3: *Connectivity data for iSite Networks.*

n	Column 2	Column 3	Column 4	Column 5
3125	22.385	20.701	4.756	6.648
6250	26.524	25.752	4.770	6.556
12500	31.395	29.137	4.677	6.579
25000	37.808	35.308	4.733	6.358
50000	45.931	42.244	4.579	6.299
100000	54.830	50.035	4.584	6.204
200000	64.668	59.284	4.649	6.071
Column 2:	$I = 3$, $p = 0.5$	$q = 0.4$, $r = 0.3$		
Column 3:	$I = 5$, $p = 0.5$	$q = 0.4$, $r = 0.3$		
Column 4:	$I = 3$, $p = 0.5$	$q = 0.05$, $r = 0.8$		
Column 5:	$I = 5$, $p = 0.5$	$q = 0.05$, $r = 0.8$		

Data for $I = 5$ and with the same values of $(p, q, r) = (0.5, 0.4, 0.3)$ are shown in table 4.3 as well. Changing the value of I (the number of iSites per node) should not change the value of γ, and this appears to be the case here. A least squares fit to the data in Column 3 and determining γ as above gives $\gamma = 1.737$ and $\gamma = 1.7498$, very close to the values above.

If $p = 0.5$, $q = 0.05$ and $r = 0.8$, then $2r > 1 + q$, and in this case $\gamma = 2 + q$. If the number of iSites per node is $I = 3$, then the data in table 4.3 gives a constant value for $\langle k \rangle$, and for $I = 5$ a slightly decreasing numerical estimate. The mean field value of γ in these cases is 2.05, and a least squares fit gives $\gamma \approx 2.009$ if $I = 3$ and $\gamma \approx 2.022$ if $I = 5$ (where the coefficient of $\log n$ in the least squares fit is $2 - \gamma$). These results are consistent with the mean field results obtained above, since it shows that the value of γ is close to $2 + q$.

Data on networks generated by the iSite evolutionary algorithm with parameters $I = 3,\ p = 0.5,\ q = 0.4,\ r = 0.3$ were collected. In each case 500 networks were grown and the average degree sequence $P_n(k)$ computed.

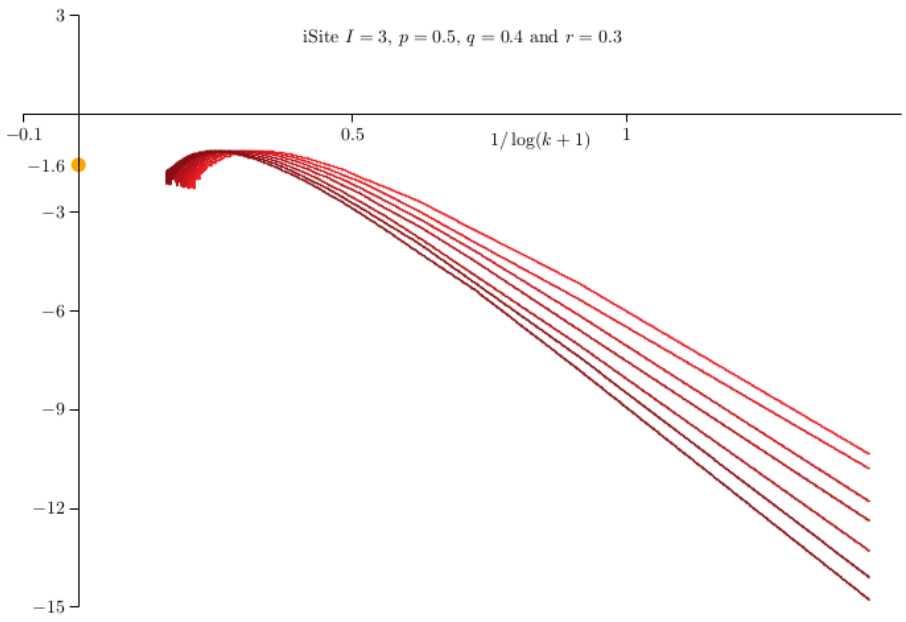


Figure 4.15: *iSite evolutionary networks with* $I = 3$, $p = 0.5$, $q = 0.4$ *and* $r = 0.3$*:*

Data on networks generated by the iSite evolutionary algorithm. In each case 500 *networks were grown and the average degree sequence* $P_n(k)$ *computed. The curves are plots of* $\log P_n(k)/\log(k+1)$ *against* $1/\log(k+1)$ *for* $n \in \{3125, 6250, 12500, \cdots, 200000\}$*. As* $k \to \infty$*, then the curves are expected to pass through* $-\gamma$ *on the y-axis, and its mean field value is* $\gamma = 1 + 2r = 1.6$ *– this value is marked on the y-axis.*

The curves in figure 4.15, are plots of $\log P_n(k)/\log(k+1)$ against $1/\log(k+1)$ for $n \in \{3125, 6250, 12500, \cdots, 200000\}$. As $k \to \infty$, then the curves are expected to pass through $-\gamma$ on the y-axis, and its mean field value is $\gamma = 1 + 2r = 1.6$ – this value is marked on the y-axis. As seen from the graph, the curves are approaching well to the mean field γ degree exponent. This indicates good agreement between the mean field analysis and numerical results for the iSite evolutionary algorithm.

4.5 Conclusions

In this research a number of algorithms used for generating networks in molecular biology were examined. Mean field theory for the algorithms was in some cases reviewed, and in other cases (Modified Barabasi-Albert, Modified Duplication-Difergence, iSite) newly presented, and also refined. The algorithms include the Barabasi-Albert [1], Duplication-Divergence [27], Solé [26] and iSite algorithms [15, 16], and these were in some cases modified by the introduction of more general elementary moves.

The efficient implementation of these algorithms was also examined, and sparse matrix routines (or, more general, hash-coding; see for example reference [23]) were used to optimize the implementation. This gives computer algorithms which can generate very large networks efficiently, and networks of order $200,000$ nodes were routinely sampled. We also explored even larger networks, up to order 3 million, but did not use those in our data analysis.

The adjacency matrix of a network of size E bonds can be stored (using sparse matrix routines) in an array of size $O(E)$. This means that the implementation of these network growth algorithms has average case space complexity $O(E)$.

Hash coding allows for the efficient implementation of routines which search, insert or delete entries in arrays storing the networks. These routines have average time complexity $O(1)$ [10], (and worst case time complexity $O(E)$ for searches, inserting and deleting bonds, due to collisions if a hash table is densely populated).

Generally, the time complexity of algorithms should grow as $O(E^\tau)$ if networks of size

E are grown (where τ is an exponent dependent on the particular algorithm). For example, networks of size E bonds can be generated using $O(E)$ computer memory, and the Duplication-Divergence and iSite algorithms can be implemented with $O(n^\tau)$ time complexity to grow networks of order n nodes (and where $n \leq E$). An examination of these algorithms (the Duplication-Divergence and iSite algorithms) suggests that an optimal implementation will have $\tau \approx 1$ (if the size of the hash tables is much larger than n).

The Barabasi-Albert and Solé algorithms (with their modified and variant implementations) should have average time complexity of $O(n^2)$ for growing networks of order n nodes. This follows because each iteration of the algorithms has to explore all nodes in the current network for the possible insertion of new bonds. Data on the time complexity of the algorithms are shown in table 4.4.

Table 4.4: *Computational Time Complexity of Implemented Algorithms.*

Algorithm	$n = 6250$	$n = 12500$	$n = 25000$	$n = 50000$	τ
Bar-Alb ($p = 0$)	0.602	2.51	9.03	38.0	1.97
Mod Bar-Alb ($\lambda = 2$, $p = A = 0$)	0.618	2.55	10.1	36.3	1.96
Var Bar-Alb ($\alpha = 2$, $a = 0$)	1.35	4.46	16.4	$--$	$--$
Dupl-Div ($p = 1$, $q = 0.4$)	0.349	0.862	2.04	5.01	1.28
Dupl-Div ($p = 1$, $q = 0.6$)	0.155	0.319	0.635	1.31	1.02
Mod Dupl-Div ($p = 1$, $q = 0.4$)	0.340	0.891	2.45	7.09	1.46
Mod Dupl-Div ($p = 1$, $q = 0.6$)	0.165	0.338	0.699	1.44	1.04
Solé ($\delta = 0.25$, $\alpha = 0.005$)	4.84	20.5	91.0	436.0	2.16
Solé ($\delta = 0.75$, $\alpha = 0.005$)	6.10	20.0	79.5	323.2	1.92
iSite ($p = 0.5$, $q = 0.01$, $r = 0.8$, $I = 1$)	0.114	0.234	0.454	0.925	1.00
iSite ($p = 0.5$, $q = 0.01$, $r = 0.8$, $I = 2$)	0.110	0.216	0.458	0.878	1.01
iSite ($p = 0.5$, $q = 0.01$, $r = 0.8$, $I = 3$)	0.106	0.217	0.432	0.857	1.00
iSite ($p = 0.5$, $q = 0.01$, $r = 0.8$, $I = 4$)	0.107	0.231	0.422	0.848	0.98
iSite ($p = 0.25$, $q = 0.01$, $r = 0.8$, $I = 4$)	0.104	0.249	0.415	0.844	0.98
iSite ($p = 0.75$, $q = 0.01$, $r = 0.8$, $I = 4$)	0.108	0.216	0.437	0.867	1.00
Mod iSite ($p = 0.5$, $q = 0.1$, $r = 0.8$, $s = 0.1$, $I = 4$)	0.288	0.560	1.102	2.53	1.04

The data displayed are the average time T to grow one network of order n. Assuming that $T = C_0 n^\tau$ and fitting $\log T$ to $\log n$, least squares estimates of τ can be obtained. For example, it is expected that $\tau = 2$ for the Barabasi-Albert algorithm, while the estimate

obtained in the table is $\tau \approx 1.97$. This is consistent with the expectation that the time complexity of the algorithm is $O(n^2)$ in an optimal implementation. This is similarly seen for the modified and variant implementation of the Barabasi-Albert algorithm, and for the Solé algorithm. The time complexity of the remaining algorithms is $O(n)$, and this is found consistently, except for the Duplication-Divergence algorithm for $p = 1$ and $q = 0.4$ (and also for the modified implementation of this algorithm). In these cases the algorithm samples denser networks (see figure 4.7) which takes up larger amounts of memory, making the implementation less efficient.

The results in this research raise some questions about the sampling of scale-free networks by random iterative growth algorithms:

- In some cases, see for example reference [29], the parameters of the algorithms were set to grow networks with properties similar to that of real protein interaction network. The values of the parameters are then used to estimate the rate of subfunctionalization (or mutation) in the genome. The results are dependent on the algorithm, and so further refinement of algorithms may be needed before useful estimates can be made.

- The mean field approaches are useful in some models (for example the Barabasi-Albert algorithm, and the iSite algorithm), but are poorer approximations in other models (the variant Barabasi-Albert algorithm, the Duplication-Divergence algorithm and its modification, and the Solé algorithm). Can the mean field approach be improved to give a better approximation to these algorithms?

- Investigation of some numerical properties of the networks (for example the connectivity) suggests that the algorithms may be self-averaging. That is, networks are generated with properties which converge to the statistical averages of these properties over a sample of networks generated by the algorithm. This is, for example, illustrated in figure 4.16 for the connectivity of Barabasi-Albert networks. As the network is grown, its connectivity appears to approach the average connectivity over a large sample of networks.

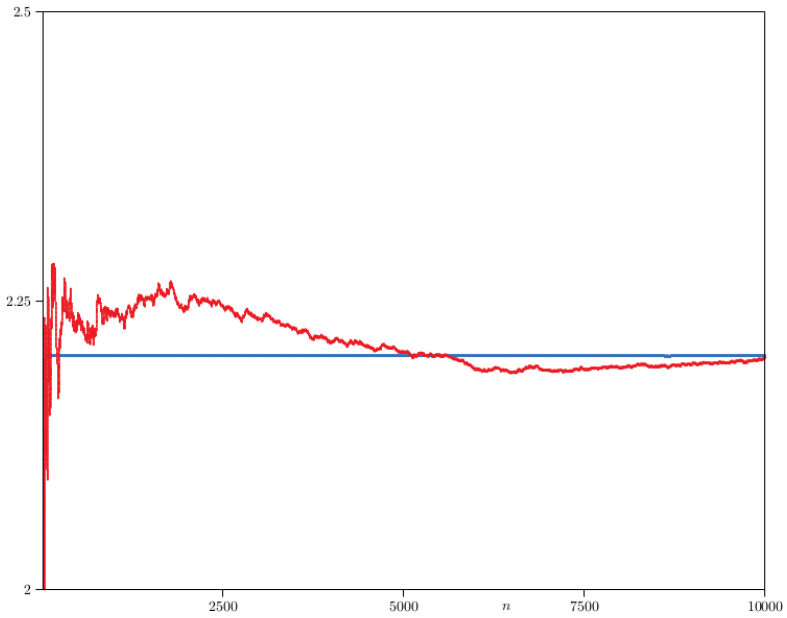


Figure 4.16: *Self-averaging of the connectivity of Barabasi-Albert networks:*
The connectivity of a single network grown with the Barabasi-Albert algorithm with $p = 0.6$ *as a function of the size of the network is given by the noisy red curve as the network is grown to order* $n = 10000$*. The blue curve is the average connectivity of Barabasi-Albert networks, plotted as a function of* n*. Notice that the red data appear to converge, with increasing* n *to the average, so that the connectivity of a randomly grown Barabasi-Albert network appears to converge to its average.*

- In this research some algorithms were modified in ways not done before in the literature (this includes the modified Barabasi-Albert, the Duplication-Divergence, the Solé

and iSite models). Exploring the properties of these modified algorithms, including their usefulness as models of networks in molecular biology, will be the subject of future investigation.

Lastly, these algorithms grow networks using a probabilistic set of rules to implement an elementary move. Each realized network N_n of order n is obtained with some probability $p(N_n)$, so that the function $p(N_n)$ is a probability distribution over networks of order n. Determining $p(N_n)$ for any of the algorithms presented here seems difficult, and general properties of $p(N_n)$ remain unknown (other than averages of network properties over $p(N_n)$ are scale-free if the algorithm grows scale-free networks).

Chapter 5

MicroRNAs and microRNA-gene interaction networks

The study of the presence of molecules and their interactions at the cellular and sub-cellular level is the focus of molecular biology. Different algorithms and tools are developed to model these interactions. Their goal is to predict yet unobserved interactions, assign functions to unknown molecules using their relations with known molecules or simply build up biological knowledge in a structured way. These algorithms can be applied to solve a particular biological problem, such as predicting protein interaction/complex formation, but also to derive systems behaviour by breaking down networks into modules or motifs with certain characteristics.

I use different algorithms in this research with the goal of finding highly connected hubs and clusters of genes which are closely related to one another. I start by building up protein-protein interaction networks and miRNA-gene interaction networks which are

then subjected to the action of two algorithms. The first algorithm is the random walk with resistance algorithm on a network. As an alternative, I am proposing the lattice laplacian on a network as a method to discover clusters of biologically related proteins. These approaches seek to find ways of solving complex pathway membership problems in protein interaction databases. The clusters obtained provide more biological insight as opposed to a process of local pairwise comparison between interacting proteins. They may also predict new members in functional pathways or clusters. These algorithms simulate biased random walks on the network for determining membership of proteins in given clusters.

The biological network I am considering is the protein-protein interaction environment of miRNA hsa-miR-218-5p.

5.1 Hsa-miR-218-5p

MicroRNAs (miRNAs) are small RNA molecules involved in various important biological processes inside the cell. They control the expression of many genes both directly and indirectly. There are over $1,000$ miRNAs coded by the human genome [39]. MiRNAs are implicated in numerous disease states and various miRNAs based therapies are being investigated [55].

Hsa-miR-218-5p is a small non-coding RNA that regulates gene expression by antisense binding. Hsa-miR-218-5p appears to be a vertebrate specific miRNA and has now been predicted and experimentally confirmed in a wide range of vertebrate species. Hsa-miR-

218-5p plays key roles in tumor metastasis. It inhibits invasion and metastasis of gastric cancer [33]. Hsa-miR-218-5p inhibits cancer cell proliferation in lung cancer [34]. Hsa-miR-218-5p, along with hsa-miR-585, has been found to be silenced by DNA methylation in oral squamous cell carcinoma [46]. It is downregulated in nasopharyngeal carcinoma [48], with artificially-induced expression serving to slow tumour growth. Hsa-miR-218-5p has been found to be implicated in epilepsy [49].

5.2 MicroRNAs structure and function

A miRNA is a small, non-coding RNA molecule containing around 22 nucleotides [52]. It is found in plants, animals and some viruses. The human genome encodes over 1,000 miRNAs [39] which target about 60% of the genes. A given miRNA may have hundred different messenger RNA targets and a given target may be regulated by multiple miRNAs [52].

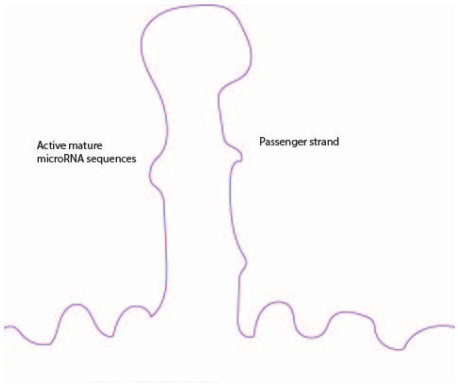


Figure 5.1: *MiRNA hairpin structure*

After transcription into a single stranded miRNA molecule in the nucleus, the miRNA will fold around itself and form a hairpin loop structure.

MiRNAs are first transcribed to pri-miRNAs [53]. The transcripts then undergo several processing steps like endonucleolytic cleavage [66], nuclear export and a strand selection procedure, to yield the single stranded mature miRNA product [54]. Transcription and processing of miRNAs determine the abundance and sequence of mature miRNAs and have important implications for their functions. MiRNAs' genes encode for long hairpin structure RNAs [54]. When processed by a series of RNase III enzymes (Drosha and Dicer), they form a miRNA duplex of 22 nucleotides with 2 nucleotides overhangs on the 3' end [58]. Only one strand of this duplex is incorporated in a protein complex that includes a member of Argonaute family of proteins [62]. MiRNA functions as a guide for this complex to the target messenger RNA. It accelerates messenger RNA deadenylation which causes messenger RNA degradation and translation repression.

Transcription of miRNAs takes place in the nucleus [53]. The primary transcripts of miRNAs are generally long (more than 1kb) and contain a 5' 7methyl guanosine cap and a 3' poly A tail [60]. The enzyme that transcribes the information from the genes to the miRNA is RNA polymerase II [53]. It is the same polymerase that transcribes messenger RNA [53]. The promoters that direct the miRNA transcription also bear the hallmark of Polymerase II promoters. Most miRNAs are products of Polymerase II, although RNA Polymerase III may also be involved. MiRNAs diverge in their expression levels since Polymerase II promoters are highly regulated and can vary greatly in strength. Since 50% of mammalian miRNAs are located within the intronic regions of protein coding or nonprotein-coding genes, these miRNAs could use their host gene transcripts as carriers [53]. Another possibility is that some will be transcribed separately using internal promot-

ers.

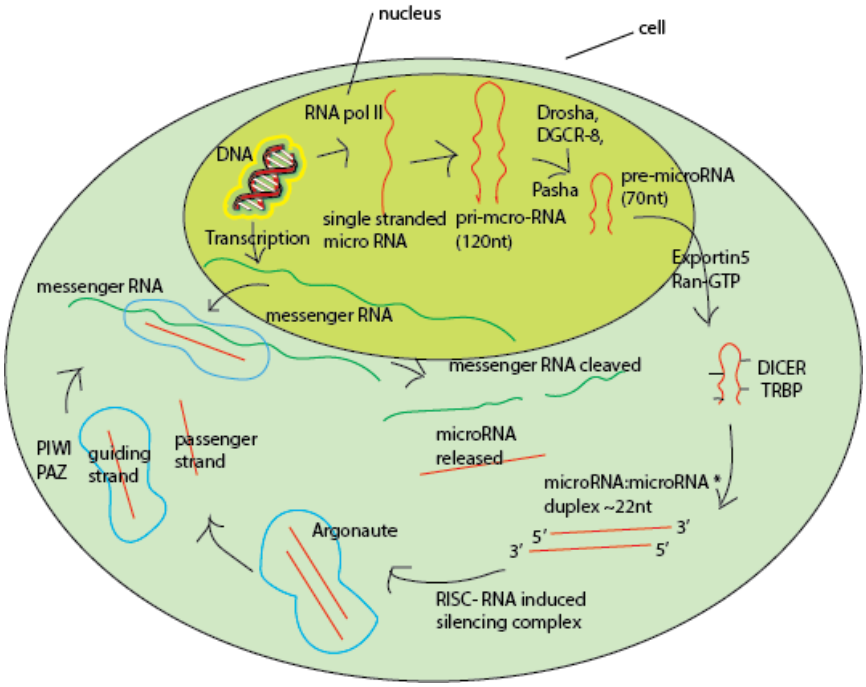


Figure 5.2: *MiRNA processing*

MiRNA is first transcribed in a single stranded miRNA by means of RNA polymerase II. The single stranded miRNA folds around itself to form a pri-miRNA around 120 nucleotides long. The partial complementarity between the base pairs in the pri-miRNA cause the formation of miRNA hairpin structure. The pri-miRNA undergoes the cleavage of the endounuclease Drosha in the nucleus and a pre-miRNA around 70 nucleotides long is formed. Pre-miRNA is exported outside the nucleus in the cytosol by means of Exportin 5 Ran-GTP. GTP is hydrolyzed to GDP and Exp 5/Ran-GTP releases its cargo in the cytosol. In the cytosol the pre-miRNA undergoes a second cleavage by the Dicer enzyme. A miRNA-miRNA duplex with around 22 nucleotides is formed. Dicer, PACT and TRBP form the RISC complex (RNA induced silencing complex). Mature miRNAs are transferred to Argonaute proteins. One strand is released in the cytosol as a passenger strand and the other one serves as a guiding strand. The guiding strand is the one which will interact with messenger RNA and control its translation into proteins.*

From a primary miRNA transcript with more than 1kb, to a mature miRNA with approximately 22 nucleotides, the miRNA must go through a series of processing steps. The first step of processing of the miRNA in animal cells is the production of a miRNA approximately 70 nucleotides long inside the nucleus called the precursor miRNA or pre-miRNA [60]. The precursor is excised from the primary transcript by Drosha enzyme [56]. Drosha

is a RNase III type endonuclease which produces duplex RNA containing a 5′- phosphate and a 3′ -OH, with a 2 nucleotide overhang at the 3′ end. Drosha by itself doesn't have any enzymatic activity [56]. It requires a subunit called DGCR8 in humans or Pasha in flies. DGCR8 contains two double-stranded RNA-binding domains that help the Drosha subunit to recognize the correct substrate [56]. An extension of several base-paired residues is required beyond the final pre-miRNA product. A flanking, single stranded RNA is required for efficient processing and at the other end of the hairpin a large terminal loop is preferred by Drosha. Drosha recognizes the local structure of a relevant hairpin, but doesn't utilize the property of a 5′ 7-methyl guanosine cap or a 3′ polyA tail [56]. Thus the processing of miRNA might happen before the primary transcript is completely synthesized.

Pre-miRNAs are produced in the nucleus, and afterwards Exportin 5 and its Ran-guanosine triphosphate (GTP) cofactor transport them from the nucleus to the cytoplasm. Exportin 5 binds to a minihelix containing RNAs with a 3′ overhang. The Exportin 5/Ran-GTP complex has a very high affinity for pre-miRNAs [57]. In the cytoplasm, GTP is hydrolyzed to GDP and Exp 5/Ran-GDP releases its cargo [57].

After being exported to the cytoplasm, pre-miRNAs are processed to mature miRNAs by means of the endonuclease Dicer [58]. Dicer also initiates the formation of the RISC (RNA induced silencing complex) composed of the Argonaute proteins [63]. Dicer is another RNase III type enzyme and it cleaves pre-miRNAs in the cytoplasm. Dicer proteins have a PAZ domain approximately 130 amino acids long [58]. It binds to single-stranded 3′ ends of double stranded RNAs [59]. As a pre-miRNA generated by Drosha [56], it already contains a 2 nucleotide 3′ overhang. Dicer recognizes the 3′ overhang via its PAZ domain

and cleaves the double-stranded region approximately 20 nucleotides away [59]. This produces a miRNA duplex containing approximately 2 nucleotide 3′ overhangs at both ends. Dicer acts as a ruler to cleave double stranded RNA substrates at a set distance from one end [59]. Proteins like TRBP and PACT in humans bind to Dicer and contribute to its function. TRBP is the transactivation response RNA binding protein [64]. PACT is the protein kinase RNA activator [65]. They enhance the affinity of Dicer for RNAs and participate in the selection of mature miRNA strands and/or the transfer of miRNAs to their final stop, the Argonaute proteins [63]. PACT, TRBP and Dicer form the RISC (RNA induced silencing complex) complex [61]. Dicer produces a miRNA duplex intermediate [59]. One of the two strands can be detected in cells. The strand with the less stable hydrogen bonding at its 5′ end within the original duplex is stabilized and becomes the mature miRNA. The other complementary strand is lost. Many miRNAs have a U residue at their 5′ends. These miRNAs have the highest chance of being selected, since a U:G base pair is less stable than a U:A pair, which in turn is less stable than a G:C pair.

Dicer interacts with a family of conserved proteins called Argonautes. Mature miRNAs are eventually transferred to Argonaute proteins and serve as guides in RNA silencing [63]. The transcription of miRNAs is the same as that of pre-messenger RNA [53], and the differential expression pattern of miRNAs mirrors that of mRNAs. Different cell types produce some but not the entire miRNA repertoire encoded by the same genome. The processing of a small number of miRNAs may be under the control of specific RNA-binding proteins. These proteins can block or allow processing until an appropriate time, or modulate cleavage sites to influence strand selection. Drosha is a key determinant of which

part of a primary transcript will become the mature miRNA [56]. The cleavage sites chosen by Drosha dictate where Dicer will cleave and hence, after strand selection, which RNA strand remains as the final product [58]. Understanding where Drosha and Dicer cleave is important. First, the cleavage sites determine the sequence of mature miRNAs. Second, the cleavage sites directly impact the function of miRNAs.

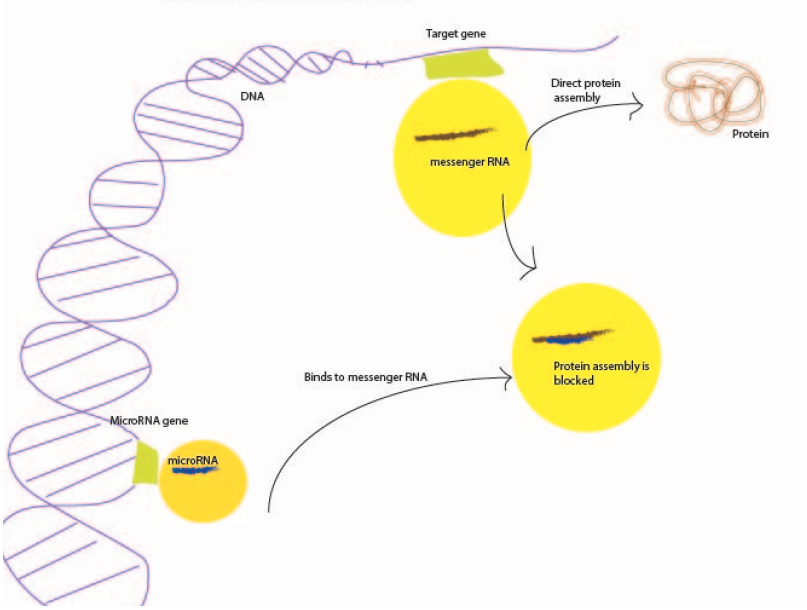


Figure 5.3: *Protein synthesis inhibition by miRNA*

After the transcription process, messenger RNA is translated into proteins in the ribosomes. When miRNA binds to messenger RNA through complementary base pairing, the translation of messenger RNA into proteins is blocked.

The **functions** of miRNA include: RNA silencing and post-transcriptional regulation of gene expression. MiRNAs are encoded by nuclear DNA and they function via base-pairing with complementary sequences within messenger RNA molecules. MiRNAs silence the expression of mRNAs through one or more of the following processes [54]: 1) cleavage of the mRNAs strand into two pieces [66] 2) destabilization of the mRNAs through shortening its poly-A tail 3) decreasing the translation efficiency of mRNAs into proteins by ribosomes.

MiRNAs control the expression of thousands of target mRNAs [67]. Each messenger RNA is believed to be targeted by multiple miRNAs. There is a sub cellular structure called the PB (processing bodies) which is linked to the miRNA pathway in down-regulating the target mRNA [77]. MiRNAs production is altered in cancer cells [67]. This suggests an impact that miRNAs might have in causing cancer. MiRNAs might form another strand of the regulatory system that exists in the cell. MiRNAs, when perfectly base-paired to their target messenger RNA, direct cleavage of a single phosphodiester bond in the target messenger RNA. This cleavage is the result of the Slicer activity of the RISC (RNA-induced silencing complex) [63].

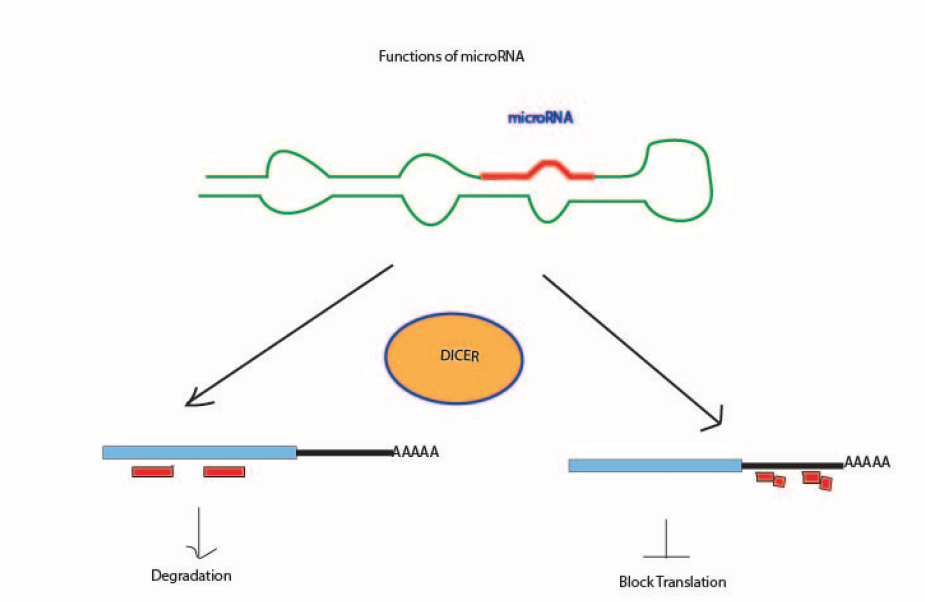


Figure 5.4: *MiRNA and translation*

In animal cells the match ups between mRNA and miRNA are not perfect. Depending on the level of complementarity between the nucleotides on the positions 2-7 on the 5' end of the miRNA and the 3' end of the UTR region on the mRNA, either mRNA degradation (left) or destabilization and blocking of translation (right) occurs.

MiRNAs control gene expression. MiRNAs are complementary to part of one or more

messenger RNAs. If the base pairing between miRNA and messenger RNA is perfect or almost perfect then cleavage of messenger RNA is promoted. This is the main way plant miRNAs function [69]. In animal cells the match ups are not perfect. For these miRNAs, in order to recognize the messenger RNA it's important that nucleotides 2-7 still be perfectly complementary [68].

Animal miRNAs action by inhibiting the translation of messenger RNAs into proteins. MiRNAs that are partially complementary to a target can speed up deadenylation causing messenger RNA to degrade sooner [60]. Depending on the level of complementarity between the nucleotides on the positions 2-7 of the 5′ end of the miRNA and the 3′ end of the UTR region on the messenger RNA either messenger RNA destabilization or degradation of the messenger RNA can occur [68]. Animal miRNAs have a diverse set of target genes. But genes involved in common functions such as gene expression have fewer miRNA target sites and seem to be under selection to avoid targeting by miRNAs [70].

Micro RNAs and diseases

MiRNAs bind to messenger RNAs before they are translated to proteins. They might turn the translation machinery off, blocking the production of proteins [67]. Several miRNAs have been found to have influence in some types of cancer [71].

MiRNAs are related to heart disease. The expression of miRNAs in diseased human hearts is altered. Several studies have indicated that miRNAs play a role not only in heart disease but in its development as well [72]. MiRNAs regulate important factors of cardiogenesis and cardiac conductance [73]. MiRNAs are involved in kidney diseases [74]. They appear to regulate the development and function of nervous systems [75]. MiRNAs

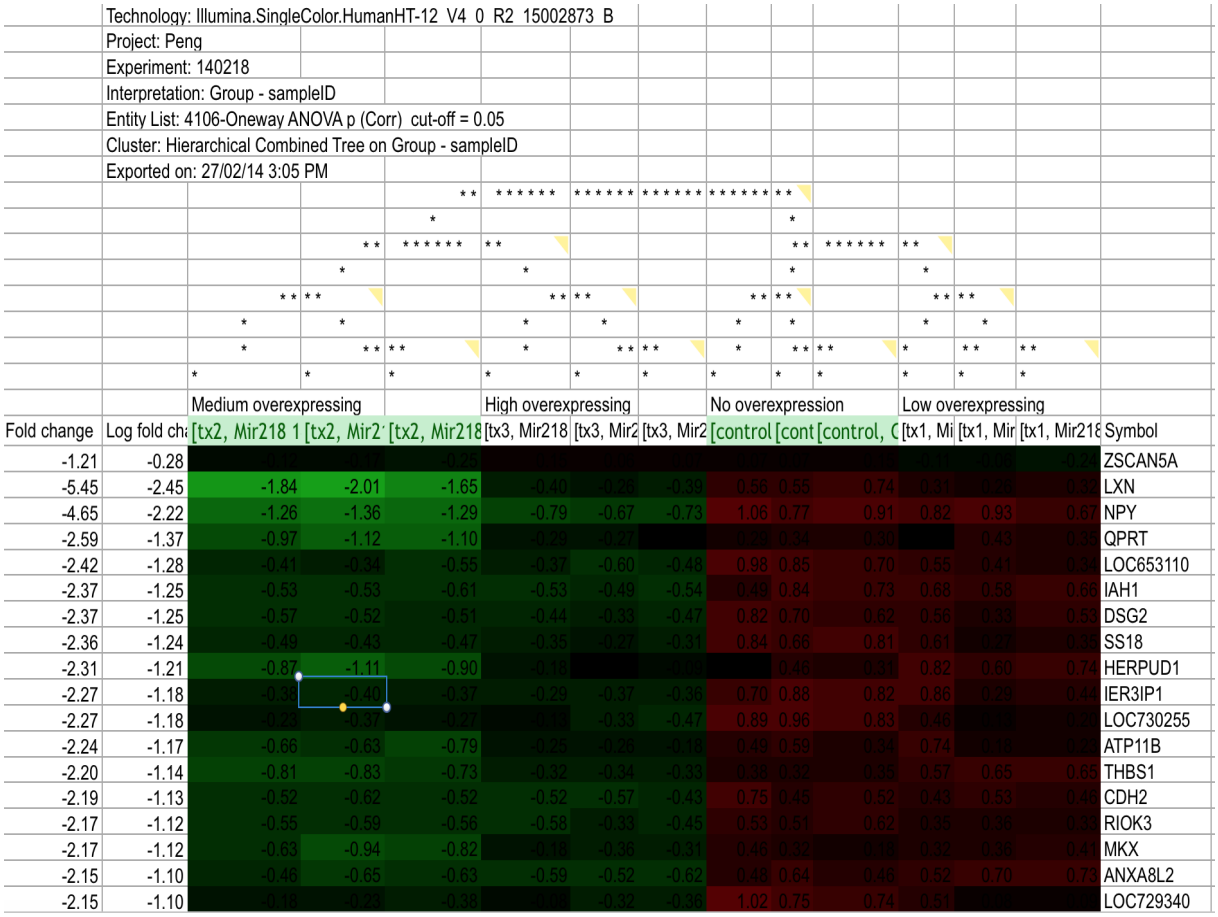
play crucial roles in the regulation of stem cell progenitors differentiating into adipocytes, having an impact on obesity [76].

5.3 Graphical Representation of MicroRNA-Gene Interactions

The graphs in this section are created using Cytoscape. Cytoscape is an open source software which can be used to visualize the interactions between different proteins [11].

The input data used in this research was provided by the Peng Lab [37] at York University. Figures 5.5 and 5.6 show pictures of part of the microarray data provided by the lab, including data on upregulated and downregulated proteins. The entire microarray data is included at the end of this dissertation as an appendix. The microarray contains hsa-miRNA-218-5p along with the upregulated and downregulated messenger RNAs by hsa-miR-218-5p. Changes in messenger RNA levels reflect changes in the protein levels. If a protein has increased expression in the presence of hsa-miR-218-5p, then we say it is "upregulated" (i.e. the mRNA transcript of the gene coding for that protein is upregulated). If a protein's expression decreases in the presence of hsa-miR-218-5p, then we say it is "downregulated" (i.e. the mRNA transcript of the gene coding for that protein is downregulated). Despite the actual interactions seen in the cell are between the miRNA and messenger RNAs, since messenger RNAs are transcripts (copies) of the genes and the changes in messenger RNA levels are reflected in protein level changes, we will refer to these interactions as miRNA-protein interactions. Fold change is a measure describing how much a quantity

changes going from an initial to a final value. Fold changes are defined directly in terms of ratios [86]. If the initial value of a substance is A and the final value is B, the fold change is defined as B/A. In genomics, log ratios are often used for analysis and visualization of fold changes. The log 2 (log with base 2) is most commonly used [86].



Technology: Illumina.SingleColor.HumanHT-12_V4_0_R2_15002873_B
Project: Peng
Experiment: 140218
Interpretation: Group - sampleID
Entity List: 4106-Oneway ANOVA p (Corr) cut-off = 0.05
Cluster: Hierarchical Combined Tree on Group - sampleID
Exported on: 27/02/14 3:05 PM

		Medium overexpressing			High overexpressing			No overexpression			Low overexpressing			
Fold change	Log fold ch	[tx2, Mir218 1	[tx2, Mir2	[tx2, Mir218	[tx3, Mir218	[tx3, Mir2	[tx3, Mir2	[control	[cont	[control, C	[tx1, Mi	[tx1, Mir	[tx1, Mir218	Symbol
-1.21	-0.28													ZSCAN5A
-5.45	-2.45	-1.84	-2.01	-1.65	-0.40	-0.26	-0.39	0.56	0.55	0.74	0.31			LXN
-4.65	-2.22	-1.26	-1.36	-1.29	-0.79	-0.67	-0.73	1.06	0.77	0.91	0.82	0.93	0.67	NPY
-2.59	-1.37	-0.97	-1.12	-1.10								0.43		QPRT
-2.42	-1.28	-0.41	-0.34	-0.55	-0.37	-0.60	-0.48	0.98	0.85	0.70	0.55	0.41		LOC653110
-2.37	-1.25	-0.53	-0.53	-0.61	-0.53	-0.49	-0.54	0.49	0.84	0.73	0.68	0.58	0.66	IAH1
-2.37	-1.25	-0.57	-0.52	-0.51	-0.44	-0.33	-0.47	0.82	0.70	0.62	0.56	0.33	0.59	DSG2
-2.36	-1.24	-0.49	-0.43	-0.47	-0.35	-0.27	-0.31	0.84	0.66	0.81	0.61			SS18
-2.31	-1.21	-0.87	-1.11	-0.90					0.46		0.82	0.60	0.74	HERPUD1
-2.27	-1.18				-0.29	-0.37	-0.36	0.70	0.88	0.82	0.86			IER3IP1
-2.27	-1.18							0.89	0.96	0.83				LOC730255
-2.24	-1.17	-0.66	-0.63	-0.79				0.49	0.59	0.34	0.74			ATP11B
-2.20	-1.14	-0.81	-0.83	-0.73							0.57	0.65	0.65	THBS1
-2.19	-1.13	-0.52	-0.62	-0.52	-0.52	-0.57	-0.43	0.75	0.45	0.52	0.43	0.53	0.49	CDH2
-2.17	-1.12	-0.55	-0.59	-0.56	-0.58	-0.33	-0.45	0.53	0.51	0.62	0.35	0.36	0.39	RIOK3
-2.17	-1.12	-0.63	-0.94	-0.82										MKX
-2.15	-1.10	-0.46	-0.65	-0.63	-0.59	-0.52	-0.62	0.48	0.64	0.46	0.52	0.70	0.73	ANXA8L2
-2.15	-1.10							1.02	0.75	0.74				LOC729340

Figure 5.5: *Microarray data negative fold change*

A picture of the microarray data showing the most downregulated gene products (proteins) by hsa-miR-218-5p (the expression of the genes is decreased in the presence of hsa-miR-218-5p).

2.19	1.13	0.32	0.48	0.47	0.10	0.27	[illegible]	-0.61	-0.78	-0.74	-0.22	[illegible]	-0.26	LOC642956
2.20	1.14	0.67	0.71	0.57	0.13	0.48	0.12	-0.61	-0.37	-0.48	-0.12	-0.21	-0.25	H2AFJ
2.26	1.18	0.46	0.50	0.57	0.09	0.17	0.33	-0.61	-0.77	-0.64	-0.19	-0.09	-0.10	VPS16
2.27	1.18	0.84	0.71	0.91	0.25	0.37	0.29	-0.31	-0.38	-0.41	-0.25	-0.34	-0.32	CKLF
2.27	1.19	0.82	0.73	0.77	0.34	0.46	0.47	-0.34	-0.44	-0.46	-0.66	-0.72	-0.84	MIR1978
2.28	1.19	0.90	0.73	0.88	0.31	0.34	0.31	-0.35	-0.39	-0.31	-0.73	-0.98	-0.85	ACTA2
2.28	1.19	1.08	0.86	0.84	0.25	0.20	0.11	-0.14	-0.39	-0.26	-0.30	-0.10	-0.13	ITGB2
2.30	1.20	0.57	0.74	0.76	0.10	0.18	0.17	-0.51	-0.52	-0.51	-0.16	-0.19	-0.27	KRT86
2.30	1.20	0.83	0.71	0.90	0.10	0.28	0.07	-0.35	-0.52	-0.29	-0.08	-0.07	-0.16	GNG11
2.33	1.22	[illegible]	-0.08	[illegible]	[illegible]	0.38	0.57	-1.34	-1.05	-1.35	-0.46	0.45	0.27	LOC650909
2.40	1.26	1.06	0.84	0.90	[illegible]	0.14	[illegible]	-0.34	-0.32	-0.32	[illegible]	[illegible]	[illegible]	SIPA1L2
2.43	1.28	0.92	0.85	0.84	0.64	0.43	0.39	-0.45	-0.39	-0.39	-0.56	-0.47	-0.64	CDK6
2.44	1.29	0.76	0.84	0.95	0.42	0.34	0.35	-0.34	-0.57	-0.40	-1.11	-1.08	-1.13	None
2.52	1.34	0.39	0.63	0.57	0.45	0.47	0.37	-0.90	-0.69	-0.83	-0.47	-0.45	-0.37	TFPI
2.59	1.37	0.67	0.86	0.71	0.51	0.60	0.48	-0.73	-0.48	-0.66	-1.85	-1.87	-1.83	RGS4
2.59	1.38	0.68	0.76	0.66	0.74	0.65	0.64	-0.64	-0.72	-0.67	-1.27	-1.23	-1.22	ARHGDIB
2.73	1.45	1.24	1.37	1.38	[illegible]	0.35	0.13	-0.12	-0.07	-0.16	-0.26	-0.11	[illegible]	SPOCK1
2.77	1.47	0.92	0.99	0.87	0.53	0.45	0.33	-0.51	-0.49	-0.64	-0.33	-0.57	-0.43	GREM1
2.80	1.49	0.71	0.65	0.93	1.19	1.04	0.96	-0.71	-0.72	-0.75	-0.67	-0.65	-0.72	CD163L1
2.90	1.53	1.15	1.24	1.02	0.60	0.51	0.36	-0.43	-0.40	-0.37	-0.37	-0.58	-0.36	KIAA1199
2.95	1.56	0.79	0.94	0.85	0.62	0.55	0.62	-0.65	-0.74	-0.70	-0.80	-0.55	-0.73	GALR2
2.97	1.57	0.62	0.64	0.74	0.19	0.13	0.37	-0.94	-0.94	-0.82	-0.20	-0.21	-0.16	FABP5
2.98	1.57	0.69	0.72	0.78	0.79	0.68	0.86	-0.82	-0.91	-0.81	-0.79	-0.77	-0.68	ECSCR
3.01	1.59	0.92	0.90	0.97	0.83	0.63	0.67	-0.68	-0.63	-0.68	-0.74	-0.81	-0.73	TMEM200A
3.07	1.62	0.43	0.62	0.60	0.50	0.59	0.45	-1.10	-1.03	-1.08	-0.65	-0.43	-0.57	TFPI
3.17	1.66	1.18	1.36	1.34	0.51	0.25	0.31	-0.39	-0.39	-0.34	-0.31	-0.25	-0.26	ABI3BP
3.33	1.74	0.71	0.76	0.70	0.51	0.89	0.80	-0.66	-1.10	-1.29	-0.69	-0.51	-0.54	CXCL1
3.42	1.77	0.83	0.79	0.75	0.56	0.68	0.41	-0.94	-1.00	-1.01	-0.66	-0.41	-0.70	LAMB3
3.47	1.79	0.87	1.04	0.96	1.01	0.76	1.08	-0.76	-0.82	-0.94	-1.52	-1.41	-1.46	KISS1
3.54	1.82	1.18	1.15	1.08	0.60	0.63	0.44	-0.44	-0.80	-0.82	-0.76	-0.61	-0.72	HIST1H2BK
3.61	1.85	0.94	0.85	0.91	1.34	1.34	1.29	-0.85	-1.04	-0.96	-0.87	-0.91	-0.93	M160
3.95	1.98	0.70	0.83	0.87	0.35	0.21	0.11	-1.20	-1.22	-1.12	-0.18	-0.14	-0.31	FABP5L2
3.98	1.99	-0.18	[illegible]	0.10	[illegible]	0.32	0.94	-2.23	-1.75	-2.08	-0.81	0.54	0.47	LOC10013276
4.05	2.02	1.33	1.48	1.34	0.39	0.60	0.53	-0.64	-0.59	-0.66	-0.46	-0.39	-0.39	IL1B
5.44	2.44	1.32	1.77	1.62	1.38	1.11	0.85	-0.87	-0.91	-0.85	-0.86	-1.06	-1.02	MMP1

Figure 5.6: *Microarray data positive fold change*

A picture of the microarray data showing the most upregulated gene products (proteins) by hsa-miR-218-5p (the expression of the genes is increased in the presence of hsa-miR-218-5p).

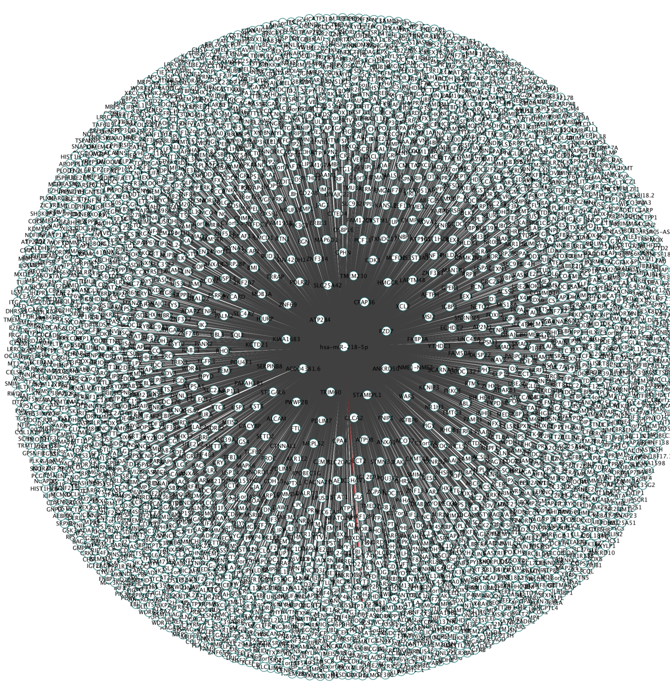


Figure 5.7: *All gene products (proteins) controlled by hsa-miR-218-5p*

All gene products (proteins) controlled by hsa-miR-218 on the microarray data provided by the Peng's lab at York University are shown.

A graphical representation of all proteins controlled by hsa-miR-218-5p is shown in figure 5.7. Graphs depicting proteins that are downregulated and upregulated by hsa-miR-218-5p are shown in figures 5.8 and 5.9 respectively.

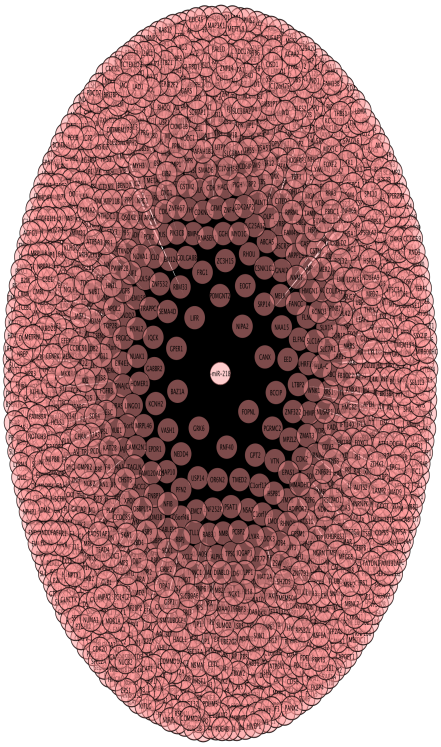


Figure 5.8: *Downregulated gene products (proteins) by hsa-miR-218-5p*

All downregulated gene products (proteins) by hsa-miR-218 on the microarray data provided by the Peng's lab at York University are shown.

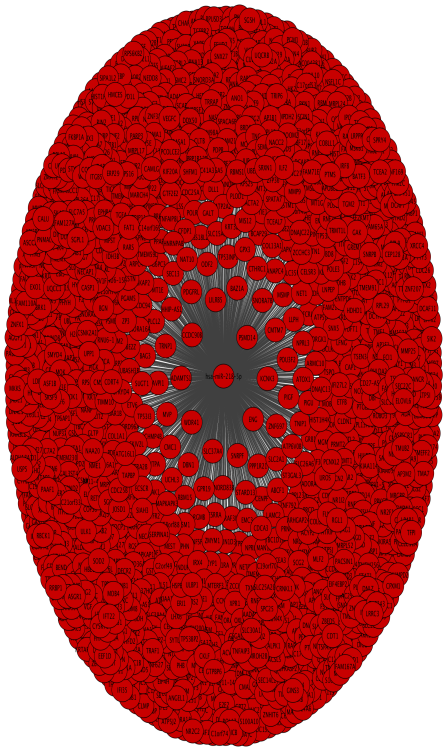


Figure 5.9: *Upregulated gene products (protesin) by hsa-miR-218-5p*

All upregulated gene products (proteins) by hsa-miR-218 on the microarray data provided by the Peng's lab at York University are shown.

Next, we are considering separately only those proteins which are the most downregulated or upregulated by hsa-miR-218-5p. We are using a cutoff fold change of 2, meaning those proteins whose expression is increased twice or more in the presence of hsa-miR-218-5p or those proteins whose expression is decreased twice or more in the presence of hsa-miR-218-5p. Using a cutoff value of 2 is large and this may ignore some useful data but

the purpose of choosing this cutoff is to make the networks tractable. These interactions are shown in figures 5.10 and 5.11. There are 24 proteins which are downregulated and 38 proteins which are upregulated by fold change of 2 or more by hsa-miR-218-5p.

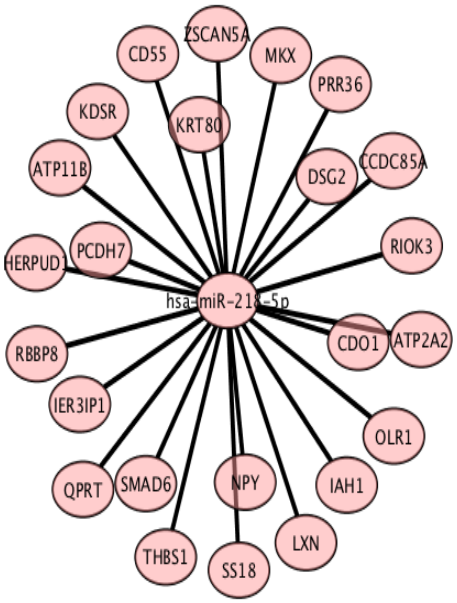


Figure 5.10: *The most downregulated gene products (proteins) by hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. 24 most downregulated gene products (proteins) by hsa-miR-218 on the microarray data provided by the Peng's lab at York University (fold change of 2 or more) are shown.

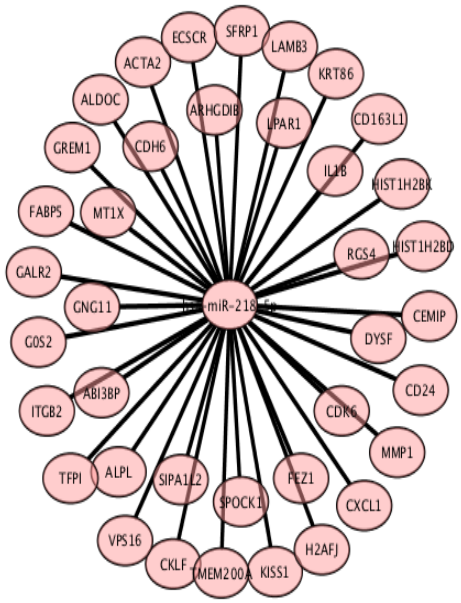


Figure 5.11: *The most upregulated gene products (proteins) by hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. 38 most upregulated gene products (proteins) by hsa-miR-218 on the microarray data provided by the Peng's lab at York University (fold change of 2 or more) are shown.

Figures 5.12 and 5.13 include data downloaded from the miRTarBase. MiRTarBase [39] is a database that contains experimentally validated miRNA-target interactions. It contains more that 50,000 miRNA-target interactions. These interactions are collected by manually surveying literature. The collected miRNA-target interactions (MTIs) are experimentally validated by reporter assay, western blot, microarray and next-generation sequencing ex-

periments. For each of the most upregulated (figure 5.13) or downregulated (figure 5.12) proteins by hsa-miR-218-5p, the set of their interacting miRNAs is downloaded from miR-TarBase [39] and the networks in figures 5.12 and 5.13 are created using Cytoscape [11].

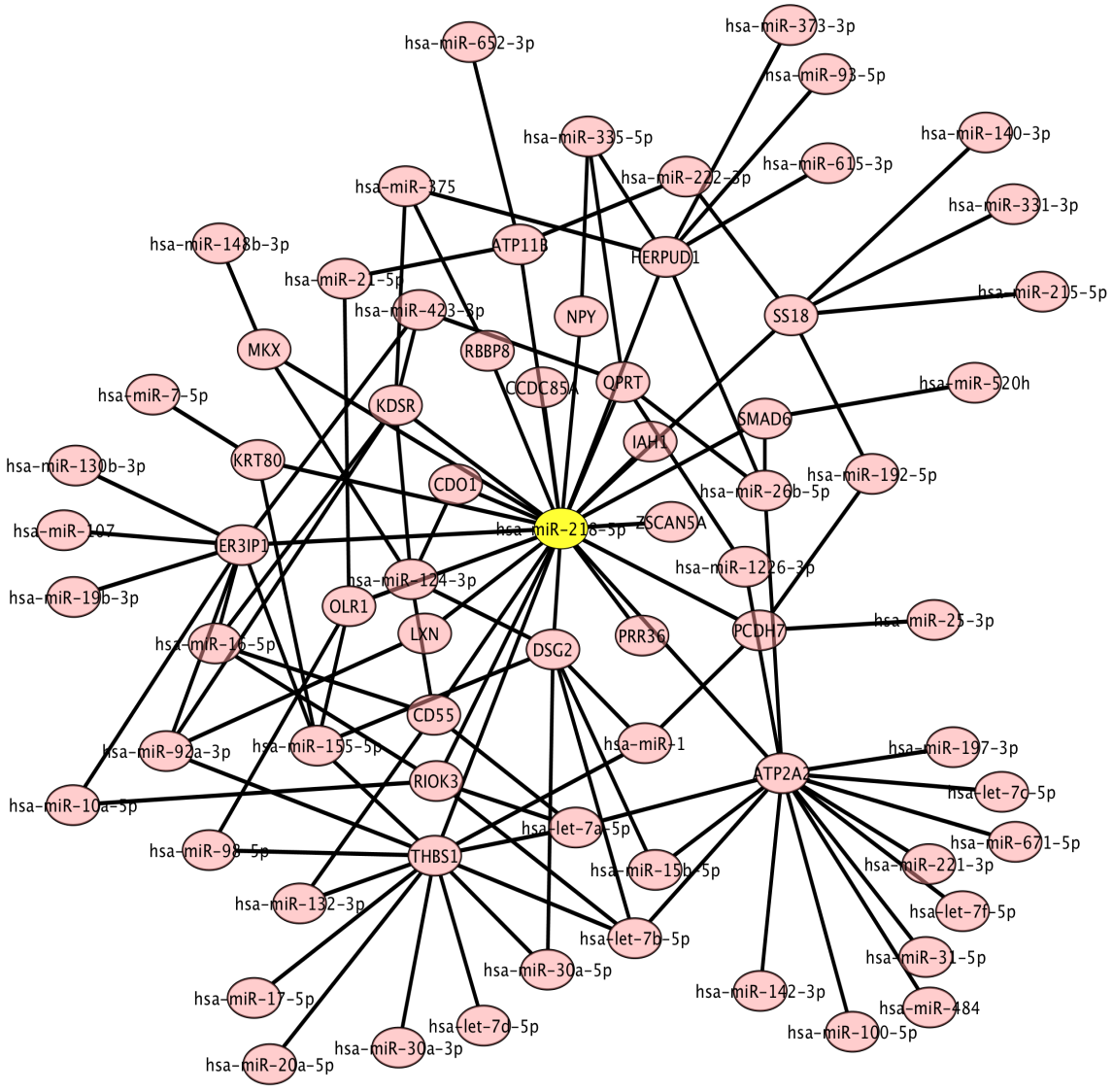


Figure 5.12: *MiRNAs controlling the most donwnregulated gene products (proteins)*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. For each of the most downregulated gene products (proteins) by hsa-miR-218, the set of the miRNAs controlling these proteins is downloaded from the miRTarBase database [39] and Cytoscape [11] is used to create the network. Hsa-miR-218-5p is shown in yellow. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

The networks show that one miRNA controls several proteins and one protein is controlled by many miRNAs. In figure 5.12, the proteins which have the most interactions with miRNAs are: ATP2A2 (15 miRNAs), THBS1 (13 miRNAs) and IER3IP1 (9 miRNAs). The miRNAs which have the largest number of interacting proteins in figure 5.12 are hsa-miR-155-5p (4 proteins) and hsa-miR-124-3p (4 proteins) .

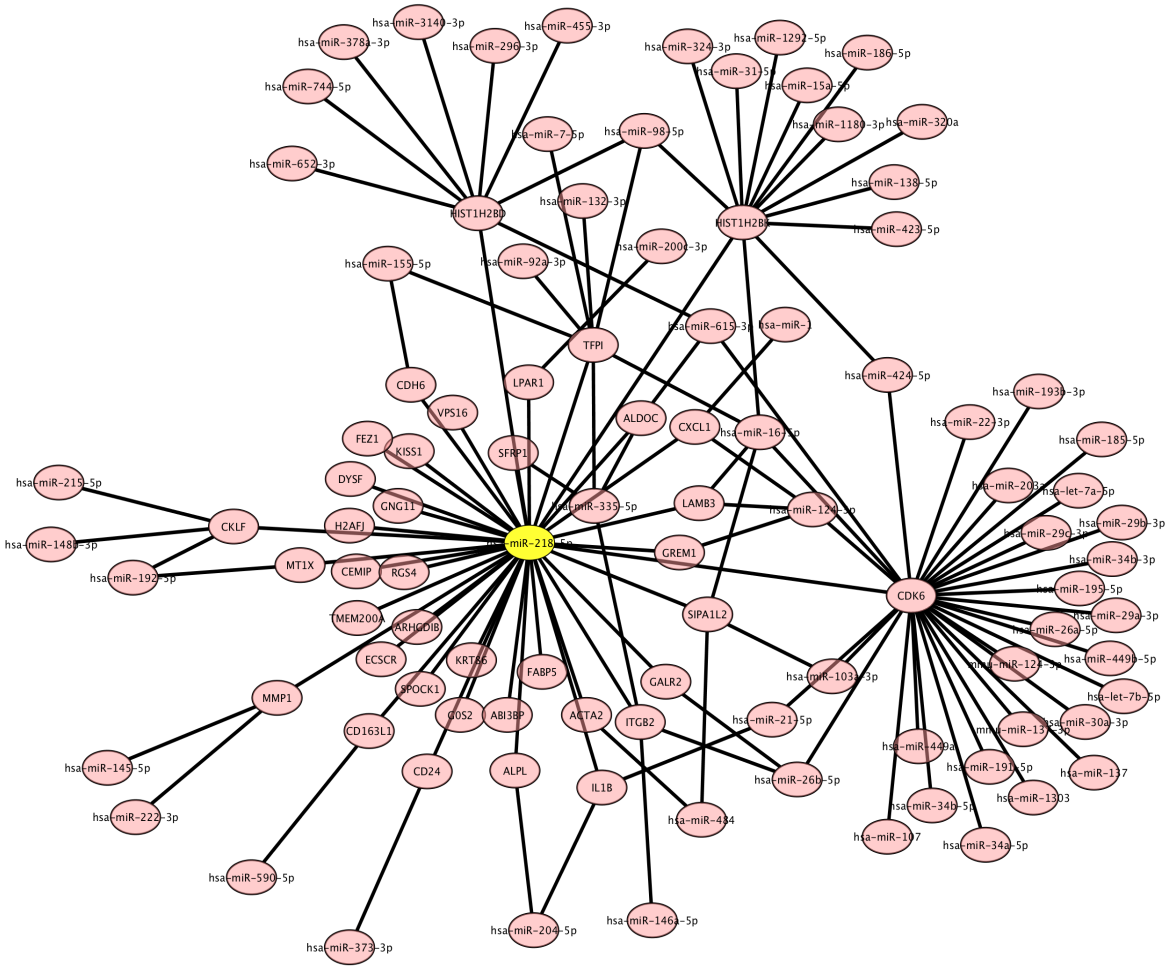


Figure 5.13: *MiRNAs controlling the most upregulated gene products (proteins)*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. For each of the most upregulated gene products (proteins) by hsa-miR-218, the set of the miRNAs controlling these proteins is downloaded from the miRTarBase database [39] and Cytoscape [11] is used to create the network. Hsa-miR-218-5p is shown in yellow. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

In figure 5.13, the proteins which have the most interactions with miRNAs are: CDK6 (31 miRNAs), HIST1H2BK (13 miRNAs) and HIST1H2BK (9 miRNAs). The miRNAs which have the largest number of interacting proteins are hsa-miR-16-5p (5 proteins), hsa-miR-335-5p (4 proteins) and hs-miR-1243p (4 proteins).

For each of the proteins controlled by hsa-miR-218-5p with a fold change of 2 or more, the set of their interacting proteins was downloaded from the STRING [38] database and networks in figures 5.14, 5.15 and 5.16 were created using Cytoscape [11].

STRING stands for search tool for the retrieval of interacting genes/proteins. It is a database of known and predicted protein-protein interactions. The interactions described in STRING [38] include direct (physical) and indirect (functional) interactions. The interactions between proteins are derived from different sources. These sources include genomic context, high-throughput experiments, conserved gene coexpression and previous knowledge [38]. STRING database currently covers $5,214,234$ proteins from $1,133$ organisms. The database aims to simplify access to information about protein associations. The associations are derived from experimental data, mining of databases and literature and genomic context analysis. The STRING database is searched up to 2 steps from hsa-miR-218-5p and the downloaded data was used to create the networks in figures 5.14, 5.15 and 5.16. The networks were visualized using Cytoscape [11].

Figure 5.14 shows protein interactions of the most downregulated proteins by hsa-miR-218-5p and figure 5.15 the protein interaction environment of the most upregulated proteins by hsa-miR-218-5p. In each of the networks, hsa-miR-218-5p is shown in yellow. The most downregulated proteins are shown in blue in figure 5.14, while the most upregulated

proteins are shown in green in figure 5.15. A graphical representation of most up and down regulated proteins combined, controlled by hsa-miR-218-5p, along with their interacting proteins downloaded from STRING database is presented in figure 5.16.

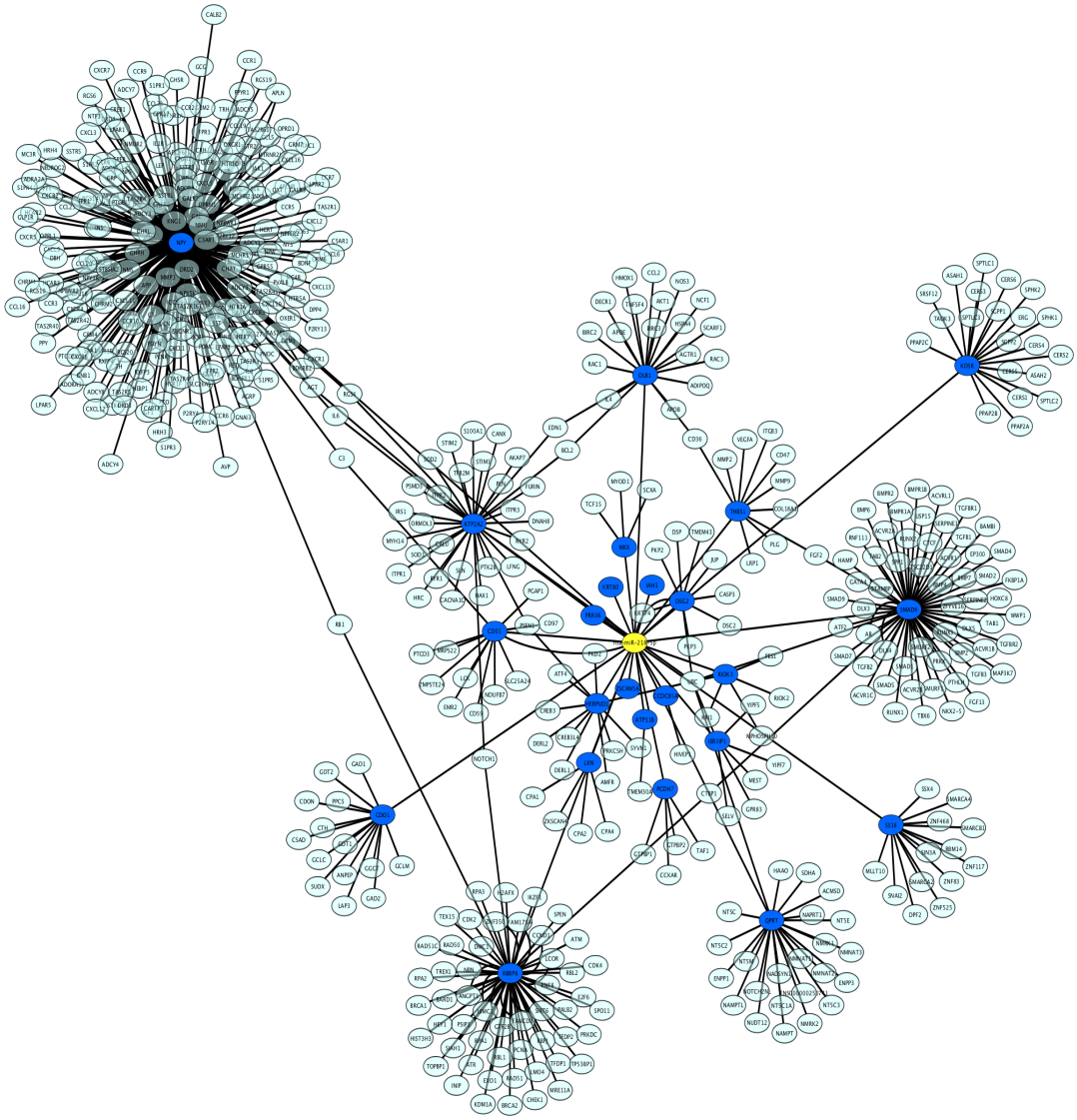


Figure 5.14: *PPI environment of the downregulated proteins by hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The protein-protein interaction environment of proteins downregulated by hsa-miR-218-5p was downloaded from the STRING database [38] and Cytoscape [11] was used to create the network. Hsa-miR-218 is shown in yellow and the most downregulated proteins are shown in blue. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

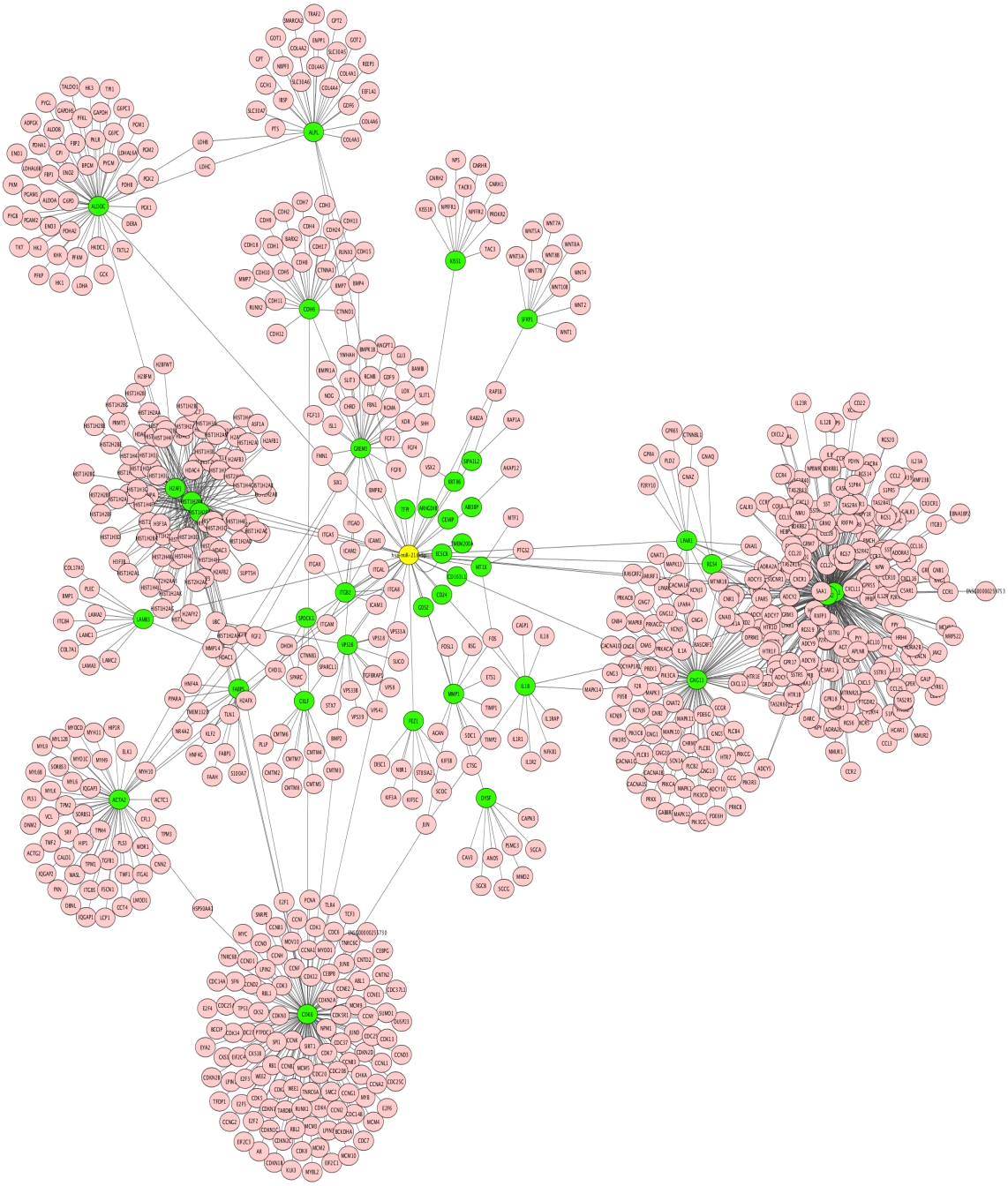


Figure 5.15: *PPI environment of the upregulated proteins by hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The protein-protein interaction environment of proteins upregulated by hsa-miR-218-5p was downloaded from the STRING database [38] and Cytoscape [11] was used to create the network. Hsa-miR-218 is shown in yellow and the most upregulated proteins are shown in green. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

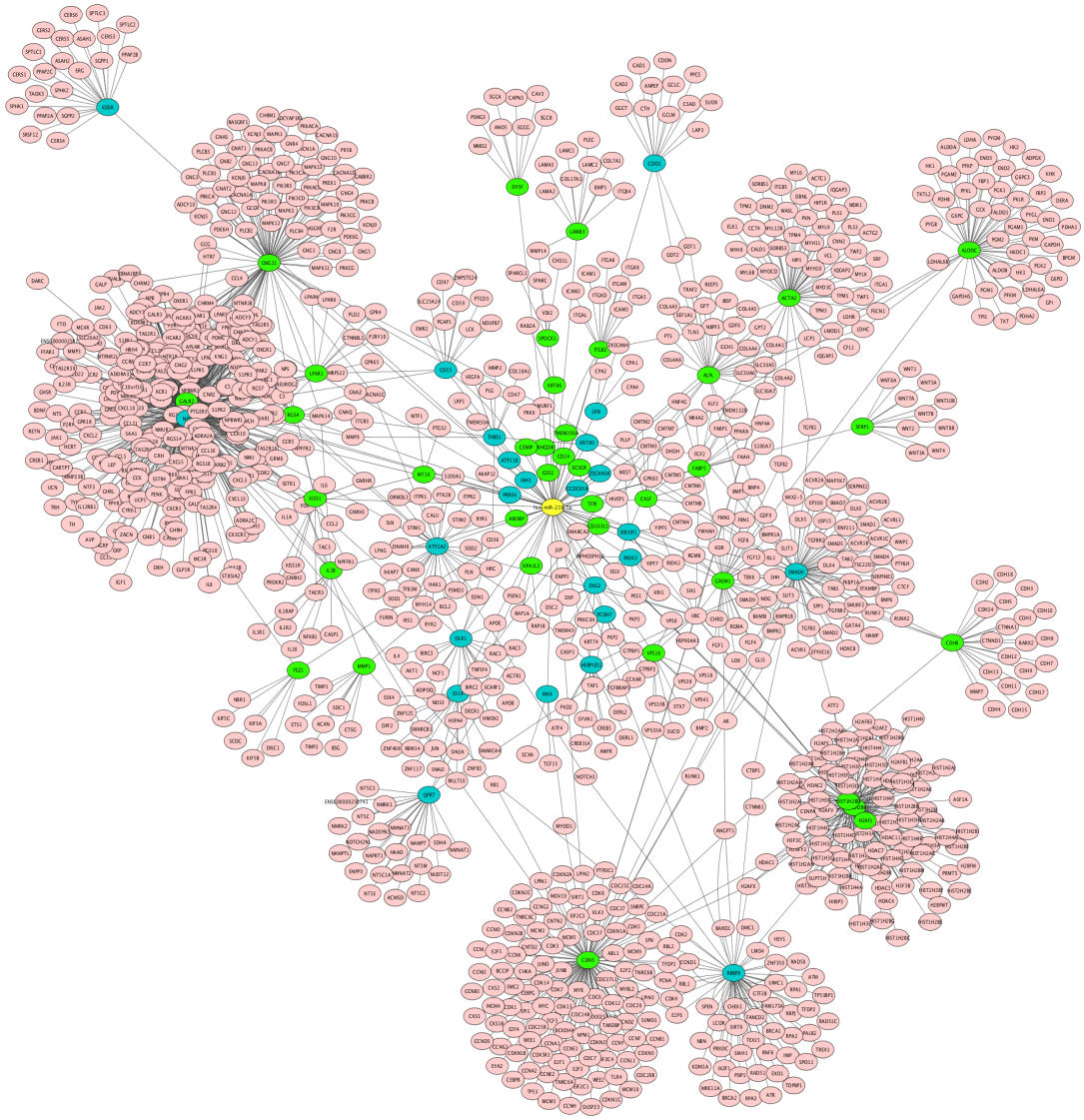


Figure 5.16: *PPI environment of proteins controlled by hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The protein-protein interaction environment of proteins controlled by hsa-miR-218-5p was downloaded from the STRING database [38]. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0 [11]. Hsa-miR-218-5p is shown in yellow. The most upregulated proteins are shown in green and the most downregulated proteins in blue.

The final network in figure 5.16 was subjected to the action of the various algorithms described in the next chapter. The network in figure 5.16 has 1,112 nodes and 1,741 bonds. The most upregulated proteins are shown in green. The most downregulated proteins are

shown in blue and hsa-miR-218-5p is shown in yellow. When applying different algorithms on the protein-protein interaction environment of hsa-miR-218-5p, the cutoff value is tuned so that the reconstructed networks contain roughly the same number of bonds as this original one (figure 5.16).

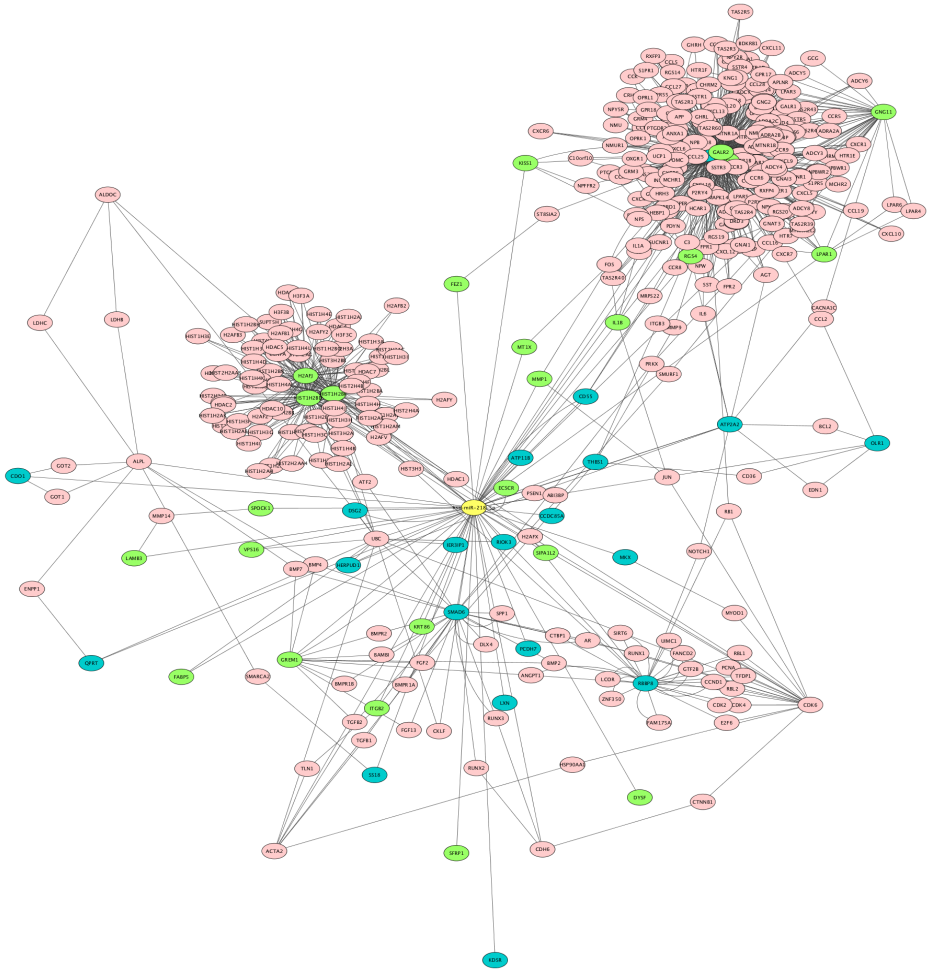


Figure 5.17: *PPI environment of proteins controlled by hsa-miR-218-5p (nodes of degree 1 removed)*

The nodes of degree 1 are removed from the network in figure 5.16. Hsa-miR-218-5p is shown in yellow. The upregulated proteins are shown in green and the downregulated ones in blue. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

From the network in figure 5.16, nodes of degree 1 are removed to clear up the clutter in the figure and the network in figure 5.17 is created.

Chapter 6

Analysis of Biological Networks Using Random Walks and Related Algorithms

Two network topology-based algorithms are presented with the goal of discovering clusters of closely related proteins in protein-protein interaction networks, and suggesting new targets of hsa-miR-218-5p. The underlying idea in network based analysis is the discovery of cluster structures (of complexes and pathways) in protein-protein interaction networks. These structures give information of biologically related proteins and their functions. The key idea is that two proteins sharing higher topological similarities are likely interacting with each other and might belong to the same protein complex and cluster in the network.

I test two algorithms. The first is to estimate similarities of the proteins in a network using a random walk with resistance [1] (RWR) algorithm. The second algorithm is to solve the lattice laplacian (LL) on a network, or its modifications WLLR (weighted lattice laplacian with resistance) and DWLLR (double weighted lattice laplacian with resistance).

Data on up and downregulated proteins by miRNA hsa-miR-218-5p were provided by the Peng lab [37]. Protein-protein interaction networks were constructed on these by examining the environment of these proteins in the STRING [38] database of PINs.

The structure of the networks was discovered with the RWR and LL algorithms, and visualized using Cytoscape [11]. Protein clusters are discovered by joining two proteins in a network when there is a higher value of the Pearson correlation coefficient between their corresponding columns in the RWR probability matrix, or in the LL solution matrix. The algorithms do not produce identical networks, but both show similar networks of biologically related clustered proteins. Clusters can be examined individually by RWR and LL algorithms to predict novel protein functions and reaction pathways.

6.1 Random Walk with Resistance (RWR)

The random walk with resistance algorithm was introduced in reference [36] as a method of improving protein-protein interaction networks. A random walk on nodes of the network is a particle stepping on nodes by choosing the next node uniformly at random from the set of neighbours of that node. Let i be the current node where the random walk is at time k and denote the probability of the random walk to be at node i at time k by $q_i^{(k)}$. The probability of the random walk to take the path from node i to node j is denoted by P_{ij}.

If $\langle i, j \rangle$ is a bond, then clearly $P_{ij} = \frac{1}{d(i)}$, where $d(i)$ is the degree of node i (the number of nodes that node i is adjacent to). The probability of a random walk to go from a node i at time k to a node j at time $k+1$ is given by:

$$f_{ij}^{(k+1)} = q_i^{(k)} P_{ij}. \tag{6.1}$$

The probability of the random walk to arrive at a node j is obtained by adding up the probabilities of random walks to enter node j through all different paths starting at i. This is:

$$q_j^{(k+1)} = \sum_i q_i^{(k)} P_{ij}. \tag{6.2}$$

When applying the random walk approach to a network, probability vectors are generated for every node in the network. The probability vector for a given node j gives the probabilities that random walks arrive at node j.

Random Walk with Resistance Algorithm: RWR Algorithm:

Parameters $\epsilon = \frac{|V|}{|E|^2}$, $\beta = \frac{1}{|E|}$ and threshold t:

1 : G a network with $V = 1, 2, ..., N$ and $M = |E|$ bonds.

2 : The transition probability matrix is P_{ij} and it is uniform, meaning if $\langle i, j \rangle$ is a bond, then $P_{ij} = \frac{1}{d(i)}$ where $d(i)$ is the degree of node i (the number of nodes that node i is adjacent to) and 0 otherwise. The probability of starting at node v and being at node i at time k is $q_{v,i}^{(k)}$.

3 : Initiate a random walk at node v at time $k = 0$. Update $q_{v,j}^{(k)}$ as follows: Compute the

stepping probabilities $f_{ij}^{(k+1)}$ depending on ϵ and β:

$$f_{ij}^{(k+1)} = \begin{cases} max(0, q_{v,i}^{(k)} P_{ij} - \epsilon), & \text{if } q_{v,j}^{(k)} > 0, \\ max(0, q_{v,i}^{(k)} P_{ij} - \epsilon), & \text{if } q_{v,j}^{(k)} = 0 \\ & \& \; max_t(q_{v,t}^{(k)} P_{tj}) \geq \beta, \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

Update node probabilities $q_{v,j}^{(k)}$ by: $q_{v,j}^{(k+1)} = \frac{\sum_i f_{ij}^{(k+1)}}{\sum_{ij} f_{ij}^{(k+1)}}$. Increment k until the $q_{v,j}^{(k)}$ converges to $q_{v,j}$.

4 : Repeat step 3 for nodes v for $n = 1$ to $n = |V|$.

5 : Construct probability matrix Ψ_{ij} with rows $q_i : \Psi_{ij} = [q_{i,j}]$.

6 : Compute H_j=median j-th column of Ψ.

7 : Compute $\Theta = \Psi - H$.

8 : Compute Pearson correlation coefficients C_{ij} between columns i and j of Θ.

9 : Join nodes i and j if $C_{ij} > t$.

Two parameters are introduced, ϵ and β. The probability vectors for each node i are calculated as in equation (6.3).

First, the algorithm checks if the next node in the random walk is already a node of the walk. If this is the case, then the probability to get from node i to node j will be equal

to $q_{v,i}^{(k)} P_{ij} - \epsilon$. If this value is negative, the probability is set to 0. If the next node on the random walk is not a node of the walk already, then the parameter β controls this step: If there exists a path from the initial node v to a node j such that the probability of the random walk to take that path is greater than or equal to β, then node j will be the next node of the walk with probability $q_{v,i}^{(k)} P_{ij} - \epsilon$, or 0 if the later is negative. If node j is a new node in the random walk and there exists no path from the initial node v to node j with probability greater than or equal to β, then the probability of the random walk to get from node i to node j will be 0.

The first parameter ϵ is a resistance term which ensures that the random walk stays close to the initial node. This will ensure that probability vectors will be different for every different starting node.

The second parameter β of the algorithm makes it difficult for new nodes, which were never visited by the random walk to be part of it. Those nodes that are new to the walk and have some sort of connection to the initial node through some paths have a higher chance to be visited by the random walk than nodes which don't. This is particularly important for the hubs in the network. A hub node is a node that is adjacent to many other nodes in the network. Not all the nodes linked to the hub node can be biologically relevant. By introducing the β parameter, it is ensured that the random walk travels only from the hub node to those nodes that are linked to the initial node of the random walk not only through the hub but also through some other paths.

The values of the parameters used in the RWR algorithm are: $\epsilon = \frac{|V|}{|E|^2}$ and $\beta = \frac{1}{|E|}$. The probability to start at a node v and arrive at a node j is calculated by adding up the

probabilities to arrive at that particular node through different paths. The probabilities are normalized so that they sum to 1. Therefore $q_j^{(k+1)} = \frac{\sum_i f_{ij}^{(k+1)}}{\sum_{ij} f_{ij}^{(k+1)}}$.

The above procedure is applied to every node of the network. A probability matrix Ψ_{ij} of dimensions $\mid V \mid$ x $\mid V \mid$ is generated. The matrix describes the probability of a random walk to start at node i and be at node j in the network. Each column vector represents the probabilities of a random walk to start at different nodes i of the network and end at a node j. The row vector of the matrix represents the probabilities of a random walk to start at a node i of the network and end at all nodes j of the network. Thus, the column vector represents the information passed from all the nodes to the current node. The row vector represents the information passed from a given node to all the other nodes in the network. The column vector better represents the topological profile of a given node in the network.

To further magnify the difference between the probability vectors, the median vector H_j of the j-th column is generated. H_j =median $(\psi_{i=1\sim|V|,j})$. The difference between the ψ_{ij} matrix and the median vector H_j is taken to obtain the offset matrix Θ. $\Theta_{ij} = \psi_{ij} - H_j$. The first term of the vector H_j is subtracted from the first column of matrix Θ, the second term of H_j is subtracted from the second column of matrix Θ and so on.

In the offset matrix Θ, the Pearson correlation coefficient is calculated between every pair of columns. These calculations give the correlation matrix. In the correlation matrix, nodes with similar topological profiles exceeding a given threshold are deemed to interact with each other and joined by a bond in the new reconstructed network. This is implemented in steps $5 \rightarrow 9$ in the RWR algorithm.

In 2012, Lei and Ruan in reference [36] introduced and applied the above procedure to

a yeast protein interaction network. In this study, the original yeast PIN network had 2,708 nodes and 7,123 bonds. After generating the offset matrix Θ_{ij}, and the correlation matrix C_{ij}, the new network is created. The pairs of nodes which have the highest topological similarity are linked together and a new network with the same number of bonds as the original network is created by choosing the threshold suitably.

It is found in the study that 40% of the bonds in the new reconstructed network are new bonds which did not exist in the original network. The new interactions generated in the reconstructed networks were verified to be more relevant biologically and functionally compared to the old (removed) bonds in the old network [36]. It was seen that the proteins which are linked by new bonds/interactions in the new reconstructed network have similar biological functions. Moreover, if two proteins are interacting with each other and one of them is (not) an essential protein, then the other one is also expected to be (not) an essential protein as well. The new interactions in the new reconstructed network share a high degree of essentiality compared to the proteins in the removed interactions in the old network.

I consider networks generated based on the data provided by the Peng lab [37] at York University. The data contains hsa-miR-218-5p and the set of proteins which are upregulated and donwregulated by hsa-miR-218-5p. For each of the upregulated and downregulated proteins by hsa-miR-218-5p with a fold change of 2 or more, the protein interactions were downloaded from the STRING [38] database. The network created is shown in figure 5.16. The network contains 1,112 nodes and 1,741 bonds. It was created using Cytoscape [11]

and was subjected to the action of different algorithms.

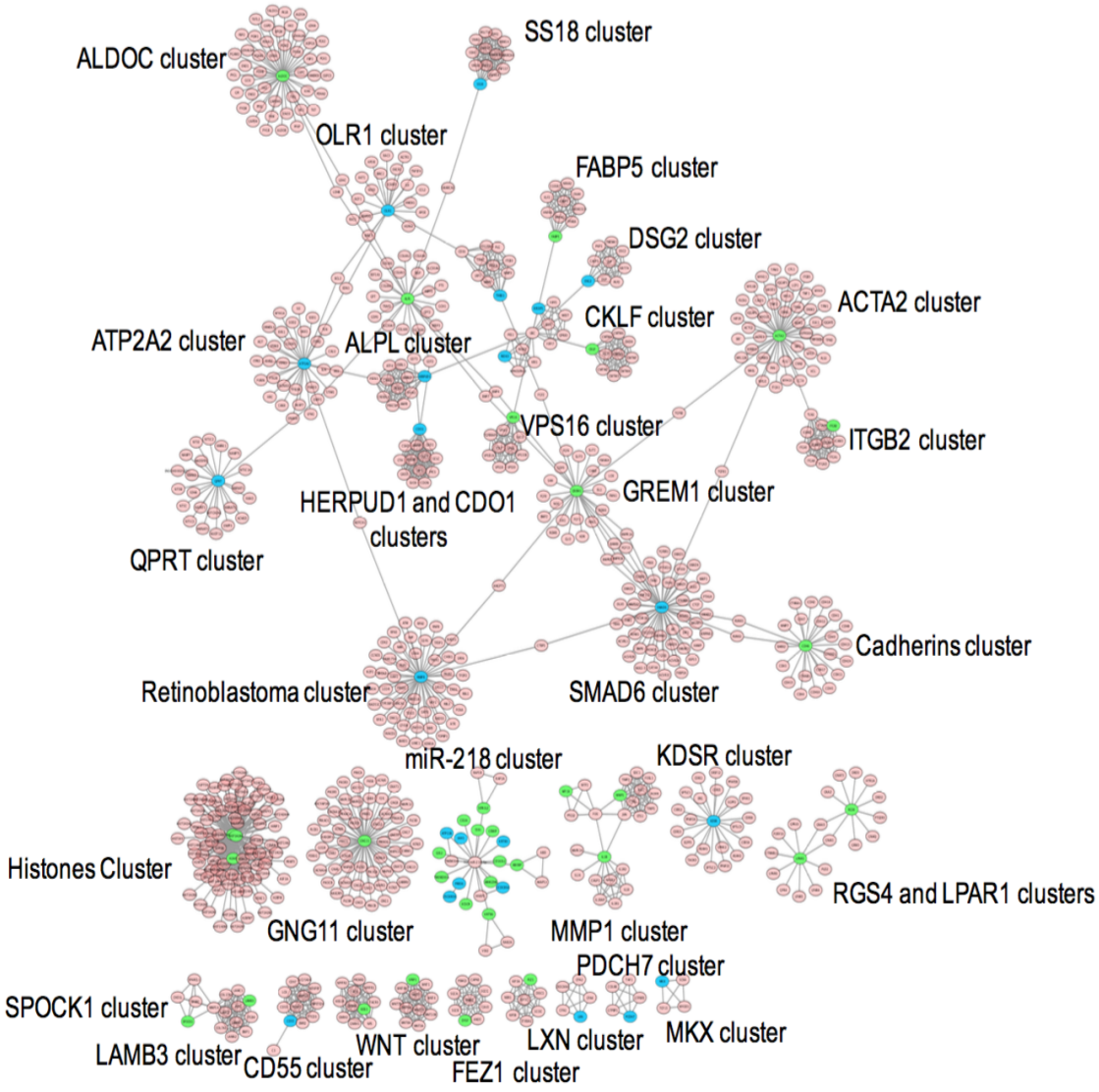


Figure 6.1: *RWR network*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the random walk with resistance (RWR) algorithm is shown. The network in figure 5.16 was subjected to the action of the RWR algorithm. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The network obtained has a clustered structure. The main clusters are labelled using either the protein of the highest degree in the cluster or the proteins which appear the most in the cluster. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

The random walk with resistance algorithm was applied in the network shown in figure 5.16. A new network is reconstructed retaining roughly the same number of bonds as the old network. The new reconstructed network is shown in figure 6.1. The new created network has a clustered structure. The clusters are labelled in the network in figure 6.1. For each cluster, the labelling was done using either the protein of the highest degree in the cluster (ALDOC cluster, QPRT cluster, etc), or the proteins which appear the most in the cluster (histone cluster, cadherins cluster, etc).

6.2 Lattice Laplacian with Resistance (LLR) Algorithm

As an alternative to the random walk with resistance (RWR) algorithm, I am proposing solving the lattice laplacian with resistance (LLR) algorithms as a method to analyze biological networks. Laplace's equation is a second-order partial differential equation.

$$\nabla^2 H = 0. \tag{6.4}$$

Let $G = (V, E)$ be a graph with nodes V and bonds E. Let $H : V \to R$ be a function of the vertices taking values in $\mathcal{R}$. Then the discrete laplacian Δ acting on H is defined by:

$$(\Delta H)(i) = \sum_{j:\langle i,j\rangle \text{is a bond}} (H(i) - H(j)) = 0. \tag{6.5}$$

Thus, this sum is over the adjacent nodes of the node i. If the graph has weighted bonds (that is, a weighting function $\alpha : E \to R$), then the discrete laplacian can be defined by:

$$(\Delta_\alpha H)(i) = \sum_{j:\langle i,j\rangle \text{is a bond}} \alpha_{ij}(H(i) - H(j)) = 0, \tag{6.6}$$

where α_{ij} is the weight on the bond $\langle i, j\rangle$. Similar to the lattice laplacian is the average operator:

$$H(i) = \frac{1}{d_i} \sum_{j:\langle i,j\rangle \text{is a bond}} H(j). \tag{6.7}$$

Steps of solving the lattice laplacian in a network:

1. The boundary conditions are specified.

2. Interior points are assigned arbitrary starting potentials H(i) (values of the solution on the nodes). The potentials H(i) are first guessed at 0 except on one node and fixed at 1 on another. The final solution doesn't depend on these initial values, but the solution may converge faster if a good guess is made.

3. The network is swept through, updating the values of the potentials H(i) for every node in the network iteratively.

4. Repeat step 3 until converged within a specified accuracy. The information received from the solution of the lattice laplacian is used to reconstruct a new network.

Solving the Lattice Laplacian with Resistance Algorithm (LLR) is described below:

Solving Lattice Laplacian with Resistance Algorithm: Parameter ϵ:

1 : Network G with $N = |V|$ nodes and $M = |E|$ bonds.

2 : Fix the potential of the node corresponding to hsa-miR-218-5p to be 0.

3 : Iterate for all nodes i in the network:

3a : Fix $y_i^0 = 1$.

3b : $y_j^{(t+1)} = \frac{\sum_{j\sim k} y_k^{(t)} + \epsilon y_j^{(t)}}{deg\ j + \epsilon}$, for $j \neq i$, for t until converged.

3c : Repeat $3b$ until all y_j are converged.

3d : Put $q_i(j) = y_j$ for $i\ \epsilon\ V$.

4 : Collect the potentials $q_i(j)$ into rows of matrix Ψ.

5 : Update network as in steps 6-9 of the RWR algorithm.

If $\epsilon = 0$, then the LL (lattice laplacian) algorithm is recovered. The lattice laplacian algorithm was applied in the protein interactions of proteins controlled by hsa-miR-218-5p described in the previous section. A new network is reconstructed. The threshold value is chosen so that the number of bonds is kept at roughly the same value as the number of bonds in the original network (figure 5.16). The results of implementing the above algorithm for different values of the parameter ϵ are shown in figures 6.2, 6.3 and 6.4. In the networks in figures 6.2, 6.3 and 6.4, the most upregulated proteins are shown in green and the most downregulated proteins are shown in blue.

The reconstructed network of protein-protein interactions of proteins controlled by hsa-miR-218-5p using the lattice laplacian with resistance (LLR) algorithm with $\epsilon = 0$ is shown in figure 6.2. The network was created using Cytoscape [11]. The network obtained has a clustered structure. The main clusters are labelled.

In figure 6.3, the value of the parameter ϵ used is -0.1.

In figure 6.4, the reconstructed network of protein-protein interactions of proteins controlled by hsa-miR-218-5p using the lattice laplacian with resistance (LLR) algorithm with $\epsilon = -0.05$ is shown.

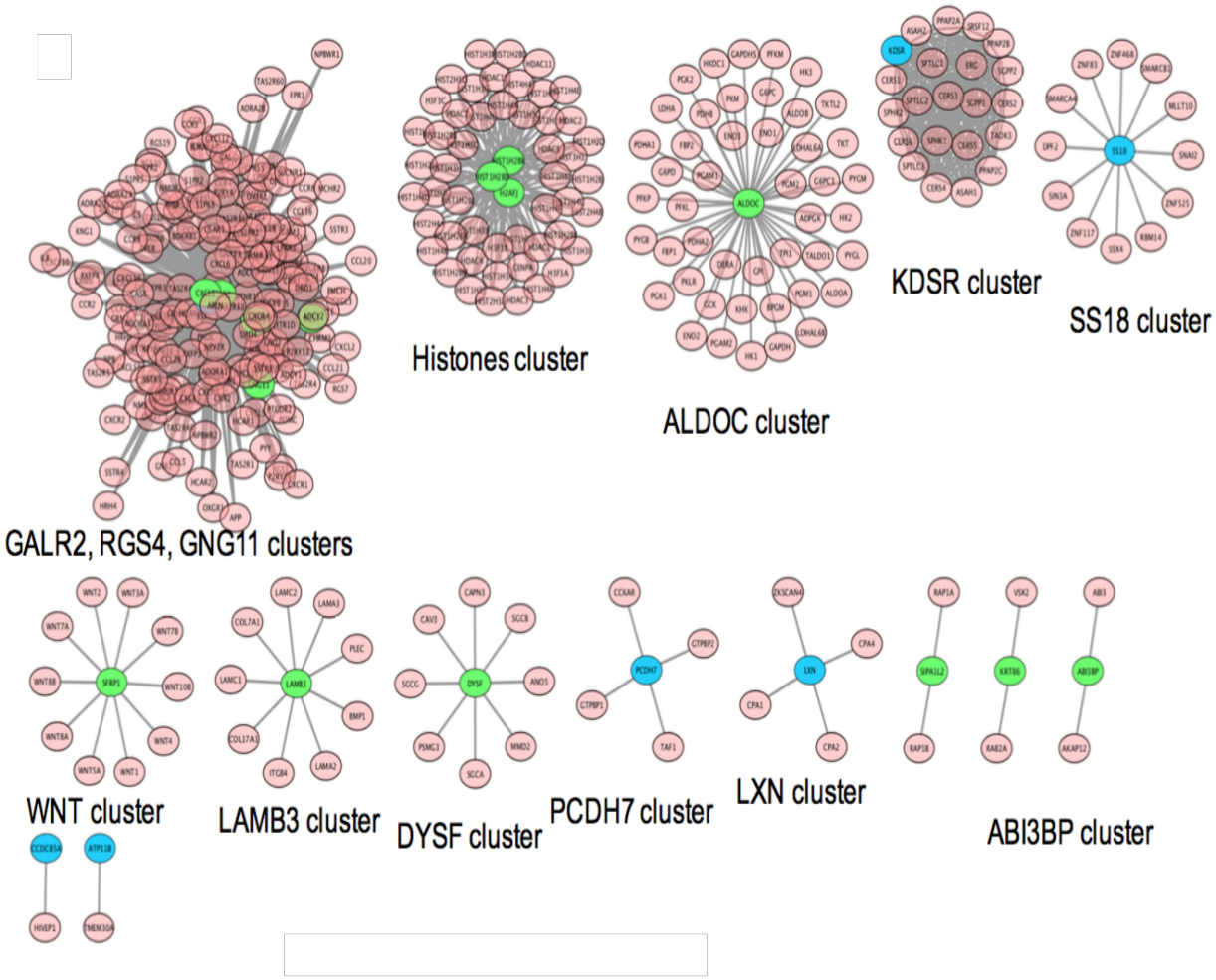


Figure 6.2: *LLR network with $\epsilon = 0$*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of protein-protein interactions of proteins controlled by hsa-miR-218-5p using the lattice laplacian with resistance (LLR) algorithm with $\epsilon = 0$ is shown. The network in figure 5.16 was subjected to the action of the LLR algorithm with $\epsilon = 0$. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure. The main clusters are labelled.

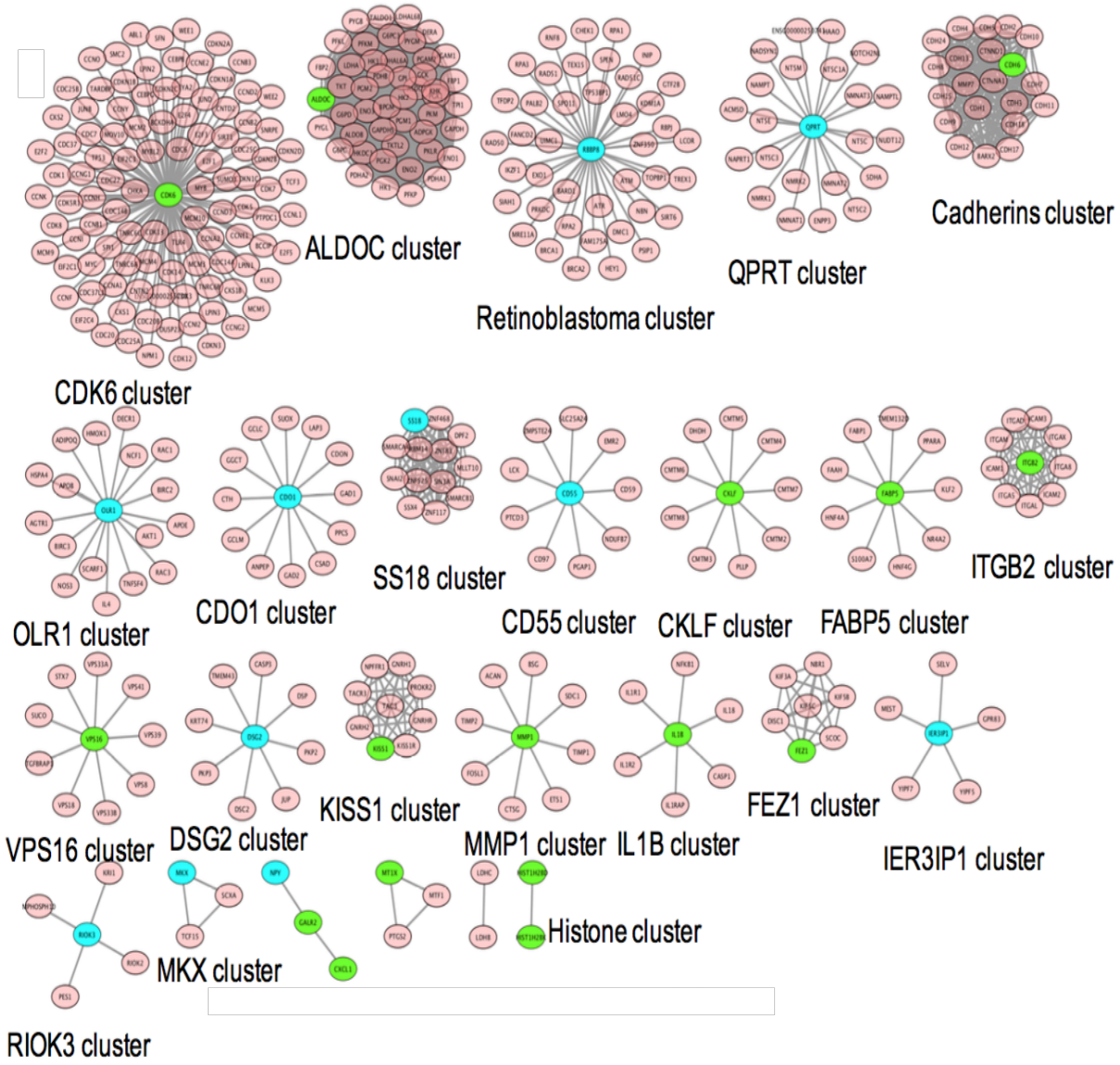


Figure 6.3: *LLR network with* $\epsilon = -0.1$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of protein-protein interactions of proteins controlled by hsa-miR-218-5p using the lattice laplacian with resistance (LLR) algorithm with $\epsilon = -0.1$ *is shown. The network in figure 5.16 was subjected to the action of the LLR algorithm with* $\epsilon = -0.1$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure. The main clusters are labelled.*

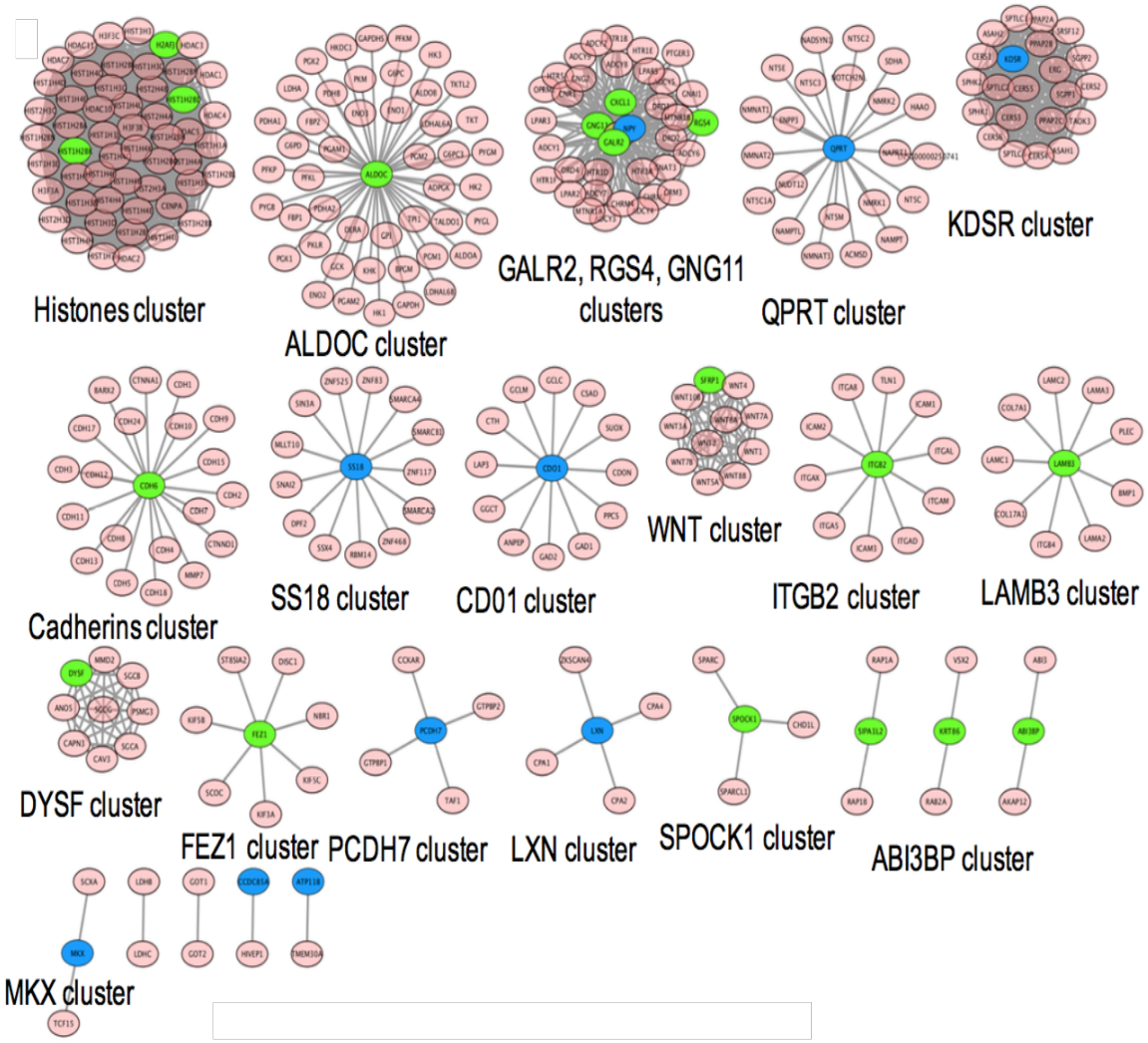


Figure 6.4: *LLR network with* $\epsilon = -0.05$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of protein-protein interactions of proteins controlled by hsa-miR-218-5p using the lattice laplacian with resistance (LLR) algorithm with $\epsilon = -0.05$ *is shown. The network in figure 5.16 was subjected to the action of the LLR algorithm with* $\epsilon = -0.05$*. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure. The main clusters are labelled using either the protein of the highest degree in the cluster or the proteins which appear the most in the cluster.*

RWR algorithm produces more clusters compared to the LLR algorithm. In general, clusters seen in the networks produced by these two algorithms are the same.

6.3 Weighted Lattice Laplacian with Resistance (WLLR) Algorithm

In this section, I modify the LLR algorithm by assigning weights to the interactions between hsa-miR-218-5p and the proteins depending on the nature of the interaction between them. If the protein is upregulated by hsa-miR-218-5p, a weight of $+1$ is assigned to the interaction between hsa-miR-218-5p and the protein. If the protein is downregulated by hsa-miR-218-5p, then a weight of -1 is assigned to the interaction. The purpose of including the weights is to refine and improve the LLR algorithm so that better predictions of protein-protein interactions and miRNA-gene interactions can be obtained.

The weighted lattice laplacian with resistance algorithm (WLLR) works as follows:

Weighted Lattice Laplacian with Resistance Algorithm: Parameter ϵ:

1 : Network G with $N = |V|$ nodes and $M = |E|$ bonds.

2 : Fix the potential of the node corresponding to hsa-miR-218-5p to be 0.

3 : Iterate for all nodes i in the network:

3a : Fix $y_i^0 = 1$.

3b : $y_j^{(t+1)} = \frac{\sum\limits_{j\sim k} w_{kj} y_k^{(t)} + \epsilon y_j^{(t)}}{\sum\limits_{j\sim k} w_{kj} + \epsilon}$, for $j \neq i$, for t until converged.

3c : Repeat b until all y_j are converged.

3d : Put $q_i(j) = y_j$ for $i \, \epsilon \, V$.

4 : Collect the potentials $q_i(j)$ into rows of matrix Ψ.

5 : Update network as in steps 6-9 of RWR algorithm.

The case where $\epsilon = 0$ gives the WLL (weighted lattice laplacian algorithm). The results of implementing the WLLR algorithm for values of the parameter $\epsilon = 0$ and ± 0.1 are presented in figures 6.5, 6.6 and 6.7.

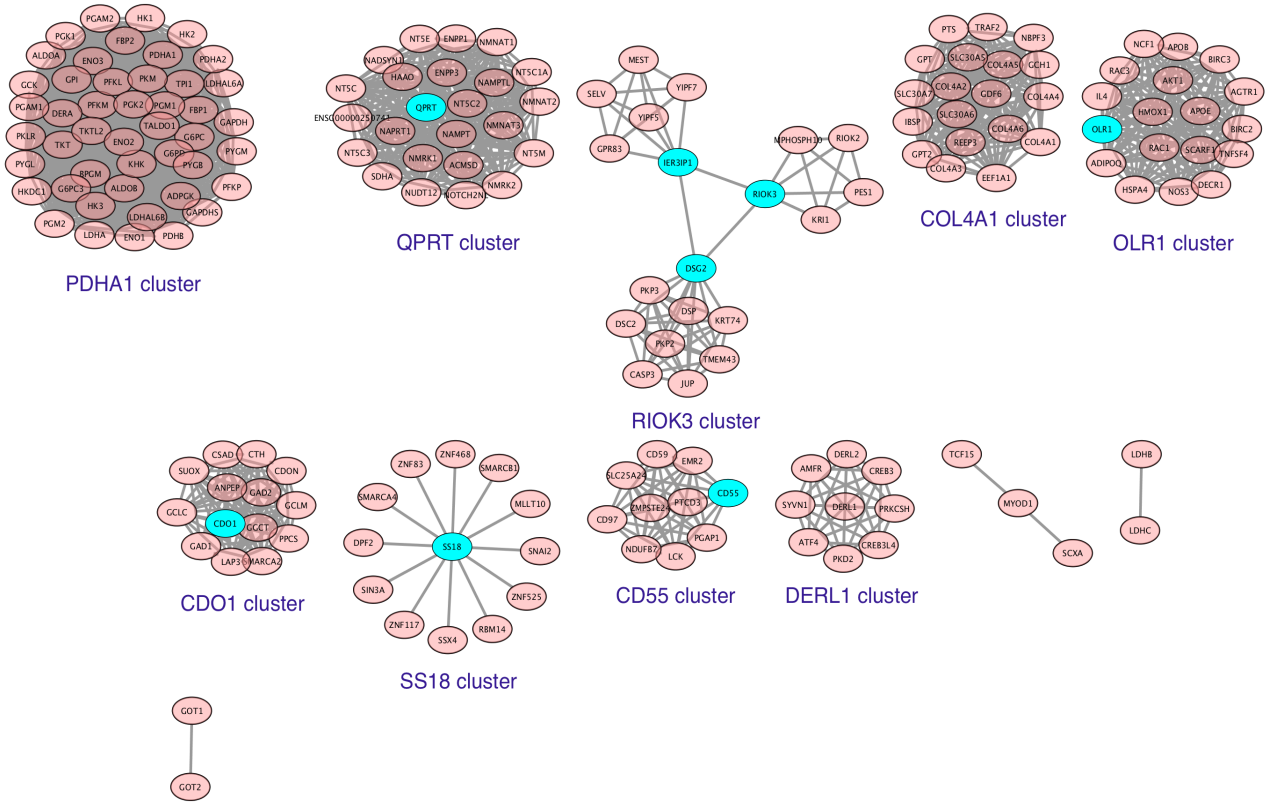


Figure 6.5: *WLLR network with* $\epsilon = 0$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the weighted lattice laplacian (WLLR) algorithm with $\epsilon = 0$ *is shown. The network in figure 5.16 was subjected to the action of the WLLR algorithm with* $\epsilon = 0$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure.*

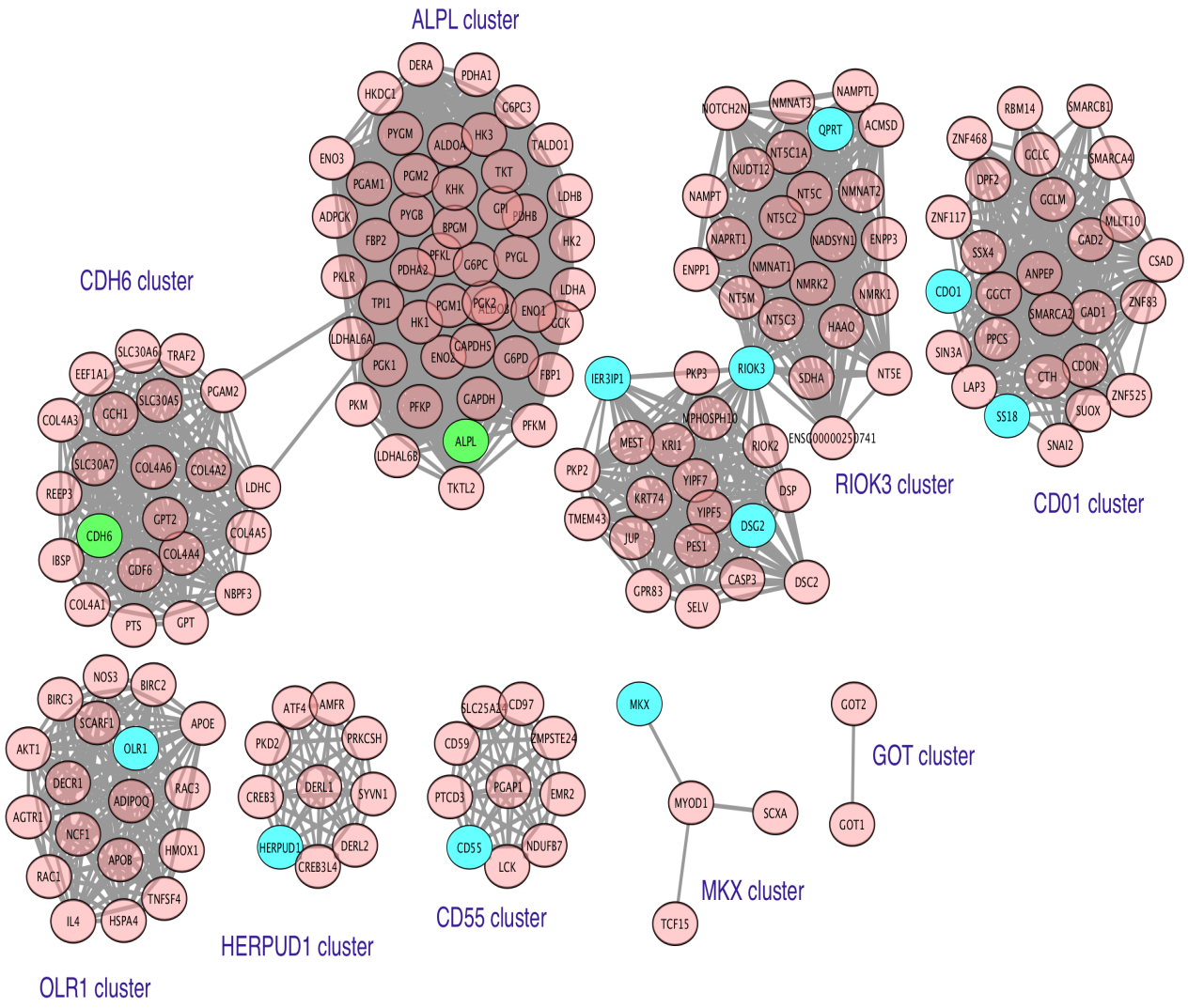


Figure 6.6: *WLLR network with* $\epsilon = 0.1$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the weighted lattice laplacian with resistance (WLLR) algorithm with $\epsilon = 0.1$ *is shown. The network in figure 5.16 was subjected to the action of the WLLR algorithm with* $\epsilon = 0.1$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The network obtained has a clustered structure. The main clusters are labelled. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.*

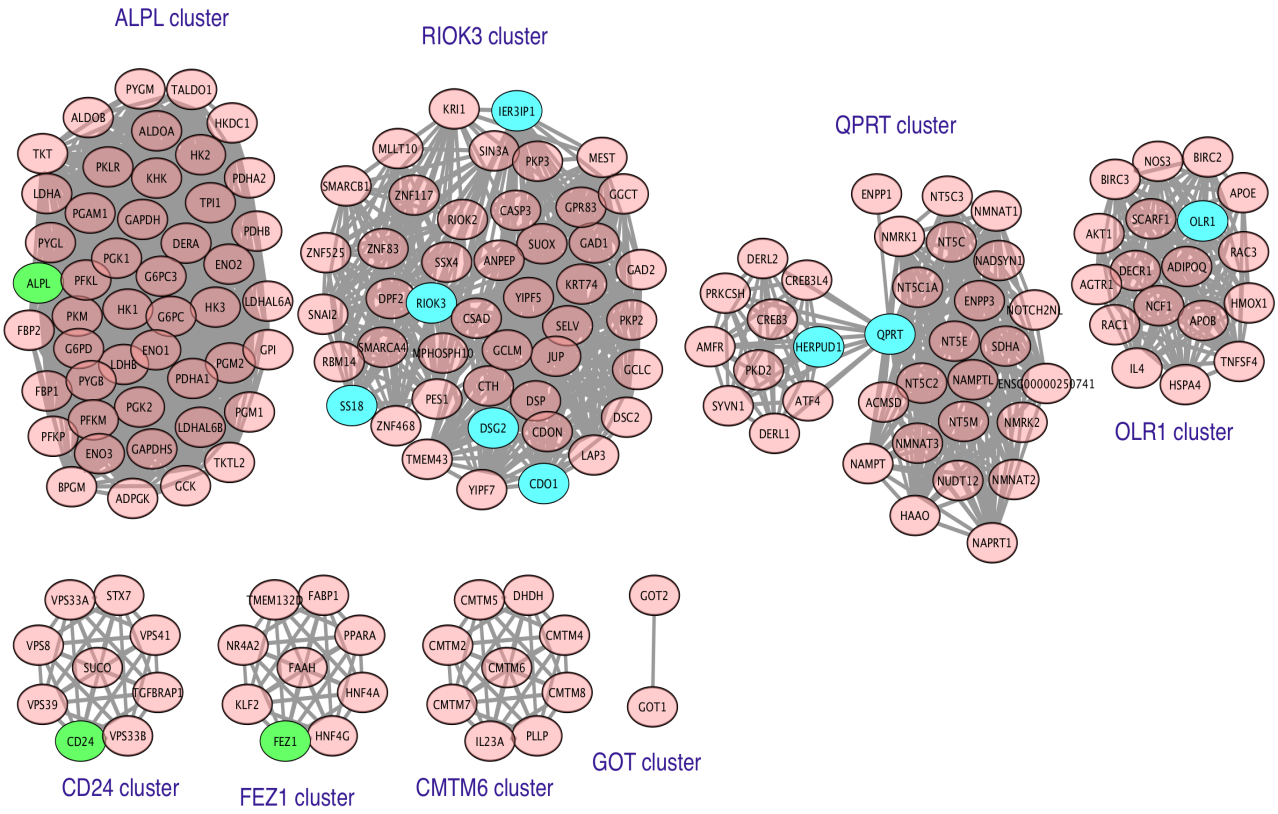


Figure 6.7: *WLLR network with* $\epsilon = -0.1$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the weighted lattice laplacian with resistance (WLLR) algorithm with $\epsilon = -0.1$ *is shown. The network in figure 5.16 was subjected to the action of the WLLR algorithm with* $\epsilon = -0.1$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure.*

Clusters generated are similar in the case of WLLR algorithm as in the network generated by the RWR algorithm. There are fewer clusters generated by WLLR algorithm than

by LLR algorithm, which in turn generates fewer clusters than the RWR algorithm.

6.4 Double Weighted Lattice Laplacian with Resistance (DWLLR) Algorithm

In this section, I modify the WLLR algorithm by assigning weights to both nodes and bonds in the network and the weighted lattice laplacian with resistance algorithm (WLLR) is modified to the double weighted lattice laplacian with resistance (DWLLR) algorithm.

Let y_i be the potential on node i, w_{ij} be the weight on bond $i \sim j$ and s_i be the weight on node i. The weight s_i for each node i is set to be equal to $s_i = \frac{1}{deg(i)}$, where $deg(i)$ is the degree of node i. If the protein is upregulated by hsa-miR-218-5p, a weight of $+1$ is assigned to the interaction between hsa-miR-218-5p and the protein. If the protein is downregulated by hsa-miR-218-5p, then a weight of -1 is assigned to the interaction. The purpose of including weights to both the nodes and bonds in the network is to refine and improve the WLLR algorithm so that better predictions of protein interactions and miRNA-gene interactions can be obtained.

The double weighted lattice laplacian with resistance algorithm (DWLLR) works as follows:

Double Weighted Lattice Laplacian with Resistance Algorithm: Parameter ϵ:

1 : Network G with $N = |V|$ nodes and $M = |E|$ bonds.

2 : Fix the potential of the node corresponding to hsa-miR-218-5p to be 0.

3 : Iterate for all nodes i in the network:

3a : Fix $y_i^0 = 1$.

3b : $y_j^{(t+1)} = \frac{\sum\limits_{j\sim k} y_k^{(t)} w_{kj} s_k + \epsilon y_j^{(t)}}{\sum\limits_{j\sim k} w_{kj} s_k + \epsilon}$, for $j \neq i$, for t until converged.

3c : Repeat b until all y_j are converged.

3d : Put $q_i(j) = y_j$ for $i \, \epsilon \, V$.

4 : Collect the potentials $q_i(j)$ into rows of matrix Ψ.

5 : Update network as in steps 6-9 of RWR algorithm.

If $\epsilon = 0$, then the DWLL algorithm is recovered. The results of implementing the above algorithm are presented in figures 6.8, 6.9 and 6.10 for cases: $\epsilon = 0$, $\epsilon = 0.1$, and $\epsilon = -0.1$.

The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm with $\epsilon = 0$ is presented in figure 6.8. The network in figure 5.16 was subjected to the action of the DWLLR algorithm with $\epsilon = 0$. The threshold was chosen to keep the number of bonds in the newly reconstructed network at roughly the same value as the number of bonds in the original network, shown in figure 5.16. The network was created using Cytoscape [11]. The network obtained has a clustered structure. The main clusters are labelled.

In figure 6.9, the newly reconstructed network is obtained using parameter $\epsilon = 0.1$.

The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm with $\epsilon = -0.1$ is presented in figure 6.10.

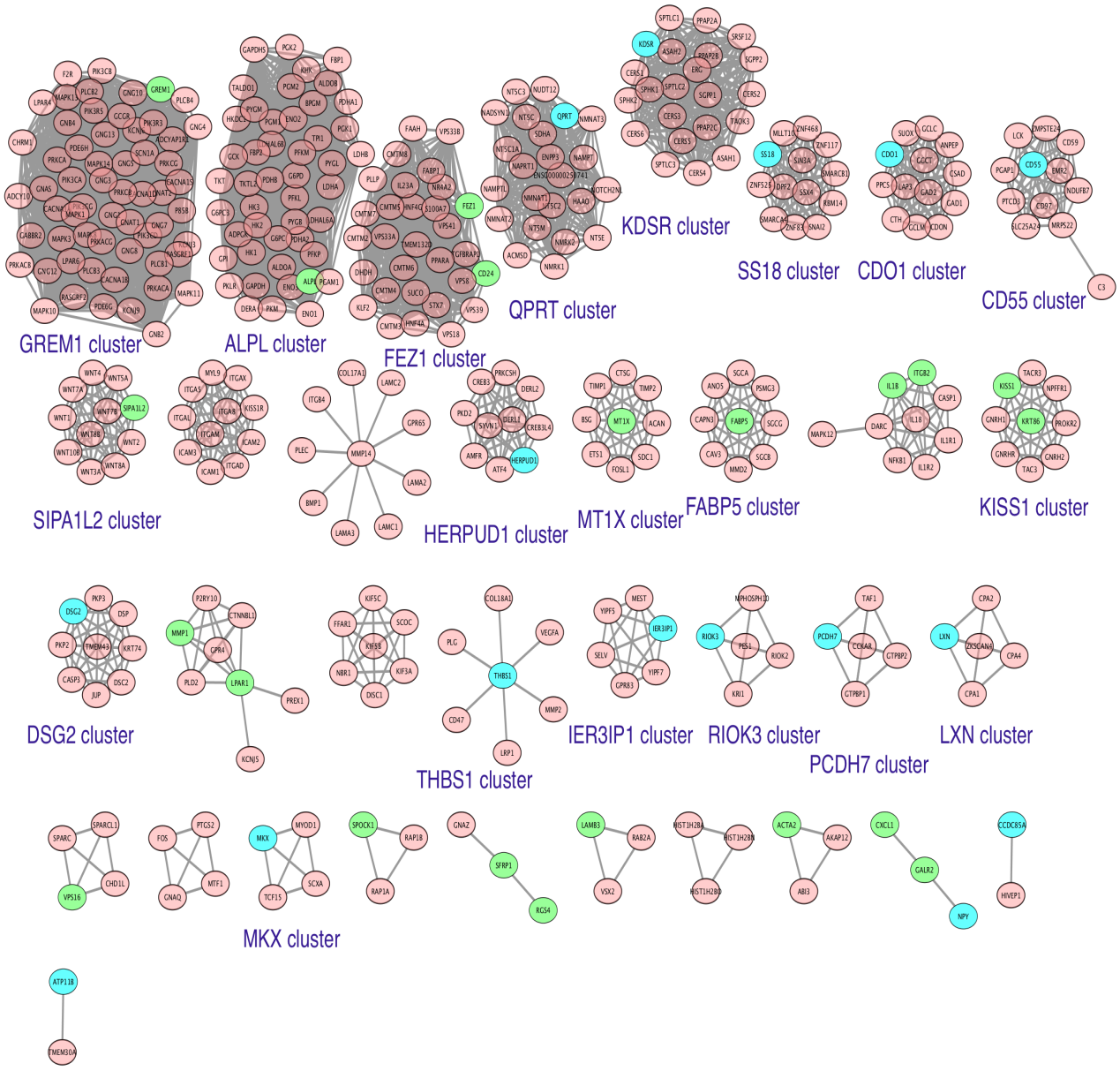


Figure 6.8: *DWLLR network with* $\epsilon = 0$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm with $\epsilon = 0$ *is shown. The network in figure 5.16 was subjected to the action of the DWLLR algorithm with* $\epsilon = 0$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure. The main clusters are labelled.*

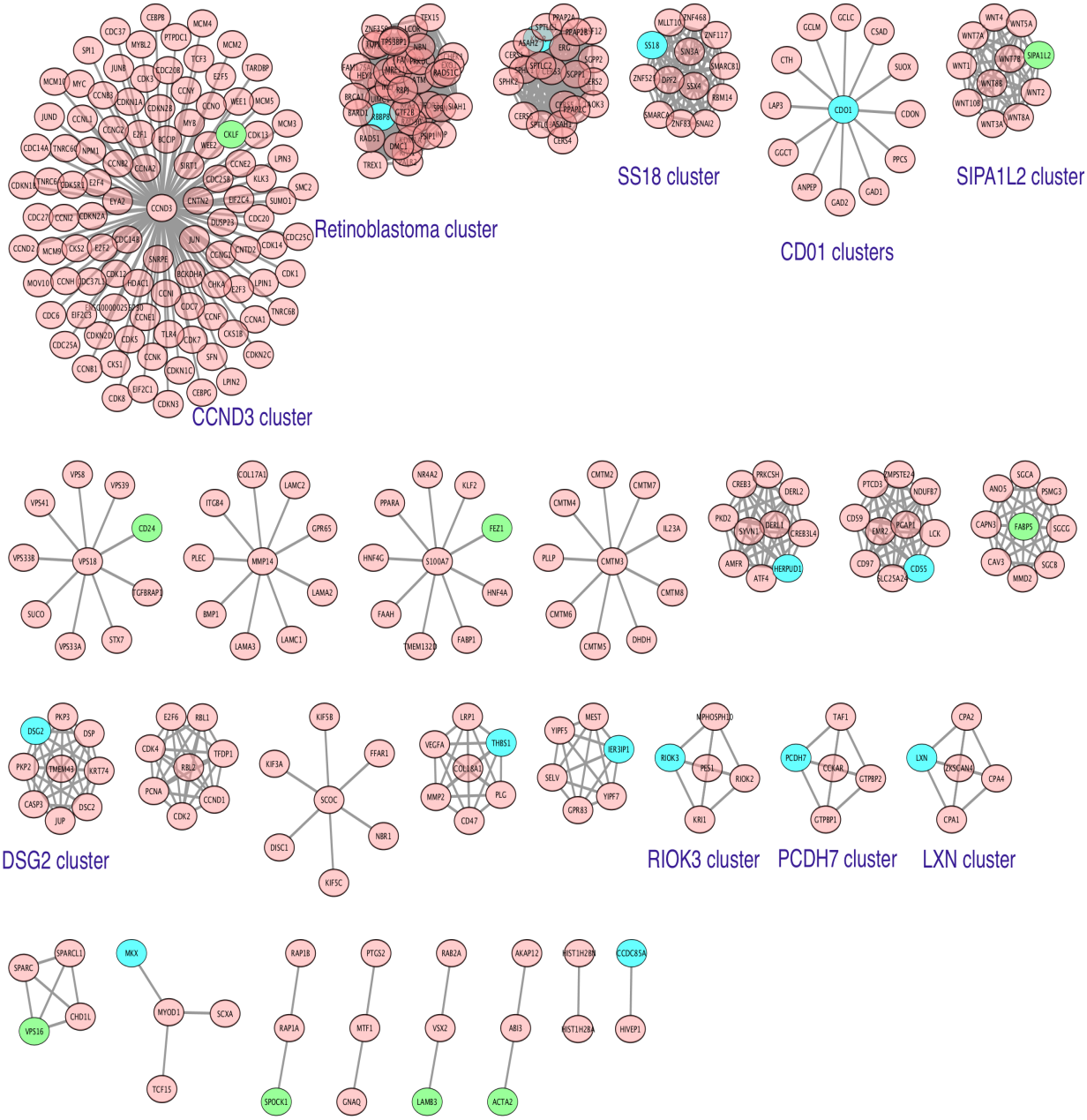


Figure 6.9: *DWLLR network with* $\epsilon = 0.1$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm with $\epsilon = 0.1$ *is shown. The network in figure 5.16 was subjected to the action of the DWLLR algorithm with* $\epsilon = 0.1$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The main clusters are labelled.*

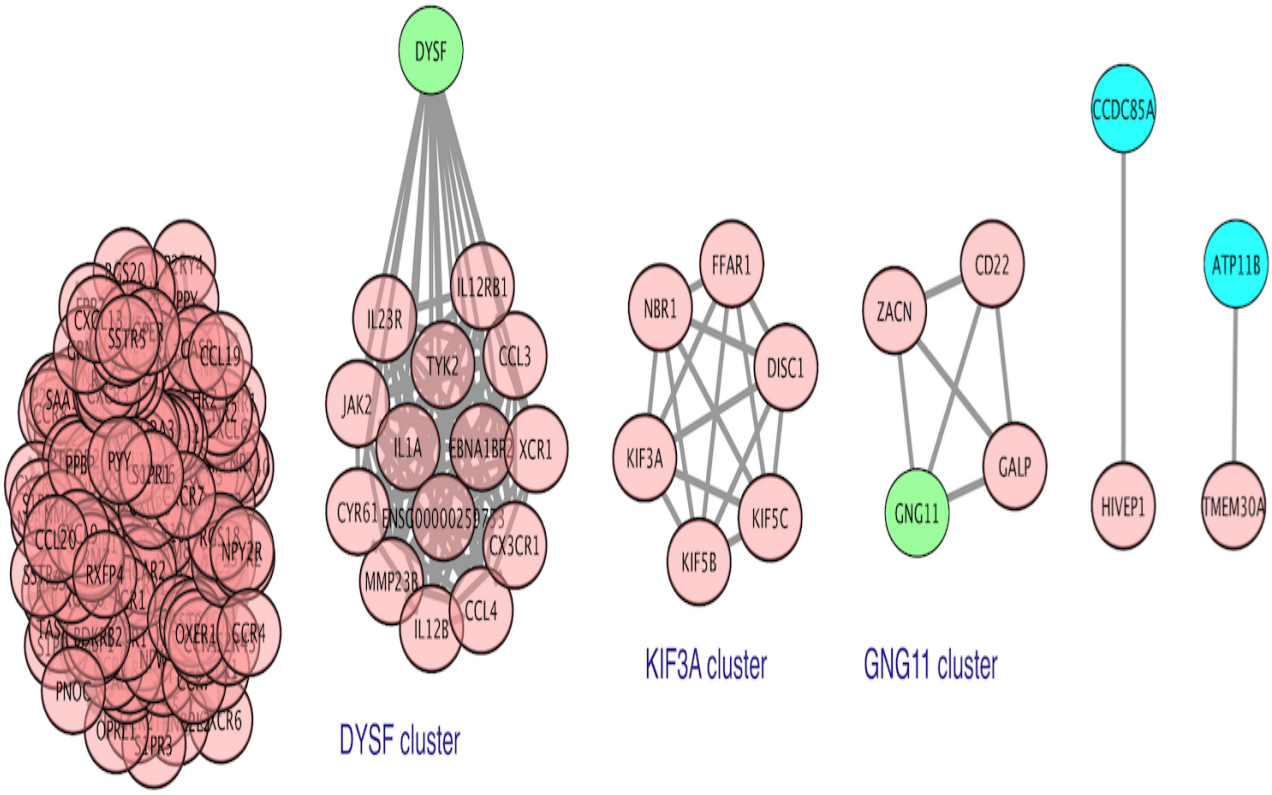


Figure 6.10: *DWLLR network with* $\epsilon = -0.1$

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm with $\epsilon = -0.1$ *is shown. The network in figure 5.16 was subjected to the action of the DWLLR algorithm with* $\epsilon = -0.1$*. The threshold was chosen to keep the number of bonds at roughly the same value as in the network in figure 5.16. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. The network obtained has a clustered structure. The main clusters are labelled.*

DWLLR algorithm produces more clusters in general compared to WLLR algorithm. Only the case where $\epsilon = -0.1$ produces the fewest number of clusters amongst all algorithms described above. There are only 5 clusters in that case.

6.5 Common and sparse clusters

A summary of the most common and sparse clusters seen in the networks generated by different algorithms is presented in table 6.1. Clusters which appear in 3 or more networks described above are listed under the "common clusters" and the others under the "sparse clusters".

Table 6.1: *Network clusters*

Common Clusters (present in 3 or more networks)	**Sparse Clusters**
ALDOC	*OLR*1
Histone	*FABPS*
*SS*18	*SPOCK*1
*GNG*11	*IER*3*P*1
QPRT	*VPS*16
KDSR	*ATP*2*A*2
Cadherines	*KISS*1
*CDO*1	*DSG*2
*ITGB*2	*RIOK*3
DYSF	*CD*55
*FEZ*1	*MMP*1
*PCDH*7	*Retinoblastoma*
LXN	*CKLF*
WNT	*SMAD*6
*ABI*3*BP*	*GREM*1
MKX	*ACTA*2
	*THBS*1
	*HERPUD*1
	LPAR

The most common clusters seen in the networks described in this chapter are clusters of histones, cadherines, WNT, ITGB2, MKX, PCDH7, SS18, GNG 11, QPRT, etc. Some of them share common biological functions. Clusters of WNT, GNG11, IL1B and cadherines have in common their involvement in signalling pathways which impact cell proliferation and development. Histones play a central role in transcription regulation, DNA repair, DNA

replication and chromosomal stability. Clusters like PCDH7 and cadherines are involved in cell-cell recognition and adhesion processes.

Some of the sparse clusters, share similar biological functions as well. For example, SMAD6 and LPAR clusters are involved in processes that mediate signaling pathways. Other sparse clusters like DSG2 cluster, SPOCK1 cluster and THBS1 cluster include proteins that mediate cell-cell interactions.

6.6 Results

6.6.1 BiNGO analysis

The network in figure 5.16 was subjected to the action of the RWR and LL algorithms described in the previous sections. The cutoff used for all these algorithms was 0.9. Networks in figures 6.11, 6.12 and 6.13 were created using Cytoscape [11]. The number of iterations used in each algorithm was $2,000$. All networks created have a clustered structure.

BiNGO [80] analysis in Cytoscape [11] was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in. BiNGO [80] is a tool to determine which Gene Ontology (GO) [81] categories are statistically over-represented in a set of genes or a subgraph of a biological network. Gene Ontology (GO) [81] defines concepts/classes used to describe gene functions. It classifies functions of genes in three aspects: *molecular function* (the elemental activities of a gene product at the molecular level, such as binding or catalysis), *cellular component* (where gene products are active) and *biological processes* (pathways and larger processes made up of the activities of

multiple gene products).

The percentages given beside each cluster in figures 6.11, 6.12 and 6.13 give the portion of the proteins in the cluster that have the molecular function or are involved in the biological process indicated. The top biological processes or molecular functions are given for every cluster in the networks generated by RWR, LLR and DWLLR algorithms.

A) Random Walk Network

The network in figure 5.16 was subjected to the action of the random walk with resistance algorithm and the network in figure 6.11 was obtained. The cutoff used was 0.9. The new network was created using Cytoscape [11] and it has a clustered structure. BiNGO [80] analysis in Cytoscape [11] was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in.

Top pathways, that clusters in RWR network are involved in, include: cell adhesion, signaling, G-protein coupled receptor protein signaling pathway, WNT-receptor signaling pathway, etc. The top enriched biological process terms associated with the proteins were transcription regulation and nucleic acid metabolism, chromatin organization, regulation of transcription, regulation of gene expression, response to stimulus, muscle contraction, protein transport, etc.

B) Lattice laplacian with Resistance, $\epsilon = 0.1$:

The network in figure 5.16 was subjected to the action of the LLR algorithm and the network in figure 6.12 was generated. The cutoff used was 0.9 and $\epsilon = 0.1$. The number of iterations performed was $2,000$. The new network was created using Cytoscape [11] and it has a clustered structure. BiNGO [80] analysis in Cytoscape [11] was used to identify the main

biological pathways and biological processes that genes in each of the clusters are involved in.

Top pathways, that clusters in LLR network are involved in, include cell adhesion, signaling pathway, cell surface receptor linked signaling pathway, G-protein coupled receptor protein signaling pathway, WNT-receptor signaling pathway, etc. The top enriched biological process terms associated with the proteins were regulation of transcription, regulation of gene expression, regulation of nucleobase, nucleoside and nucleotide, and nucleic acid metabolism, chromatin organization, regulation of transcription, regulation of gene expression, response to stimulus, nervous system development, muscle organ development, etc.

C) Double weighted lattice laplacian with resistance, $\epsilon = 0.1$

The network in figure 5.16 was subjected to the action of the DWLLR algorithm and the network in figure 6.13 was generated. The cutoff used was 0.9 and $\epsilon = 0.1$. The number of iterations performed was $2,000$. The new network was created using Cytoscape [11] and it has a clustered structure. BiNGO [80] analysis in Cytoscape [11] was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in.

Top pathways that clusters in DWLLR network are involved in, include: cell adhesion, WNT-receptor signaling pathway, signaling pathways etc. The top enriched biological process terms associated with the genes were: nucleoside phosphate metabolic processes, nucleobase, nucleoside and nucleotide metabolic processes, chromosome organization, chromatin organization, response to chemical stimulus, muscle organ development, microtubule based process, protein transport, etc.

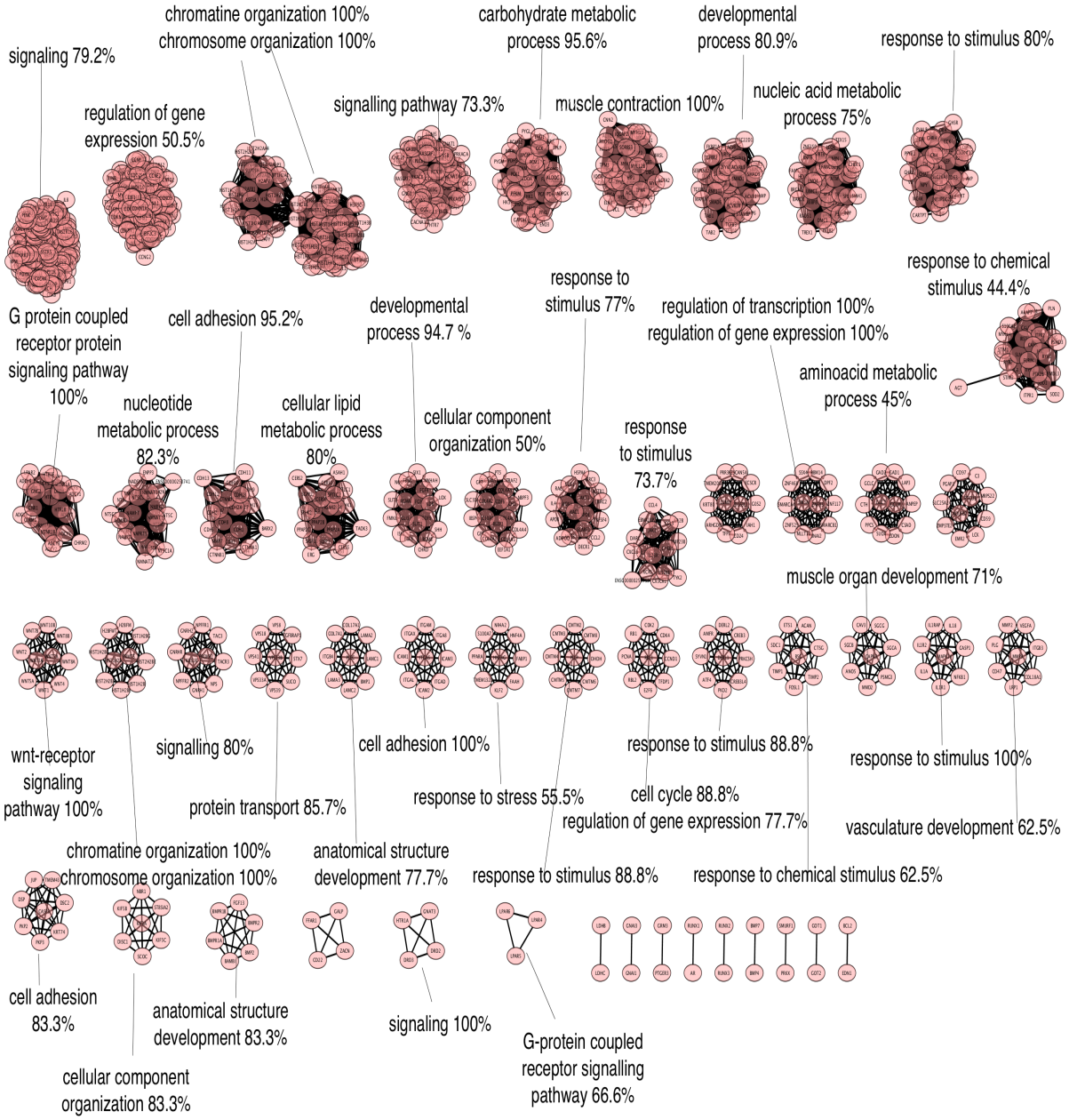


Figure 6.11: *RWR network (BiNGO analysis)*

The network in figure 5.16 was subjected to the action of the random walk with resistance algorithm. The cutoff used was 0.9. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. BiNGO [80] analysis in Cytoscape was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in. These analyses revealed that the top pathways in which hsa-miR-218-5p target proteins were involved include: cell adhesion, G-protein coupled receptor protein signaling pathway, WNT-receptor signaling pathway, etc. Most of these pathways are cancer related pathways. The top enriched biological process terms associated with the genes were transcription regulation and nucleic acid metabolism, chromatin organization, regulation of transcription, regulation of gene expression, response to stimulus, muscle contraction etc. Most of these biological processes are related to cancer development.

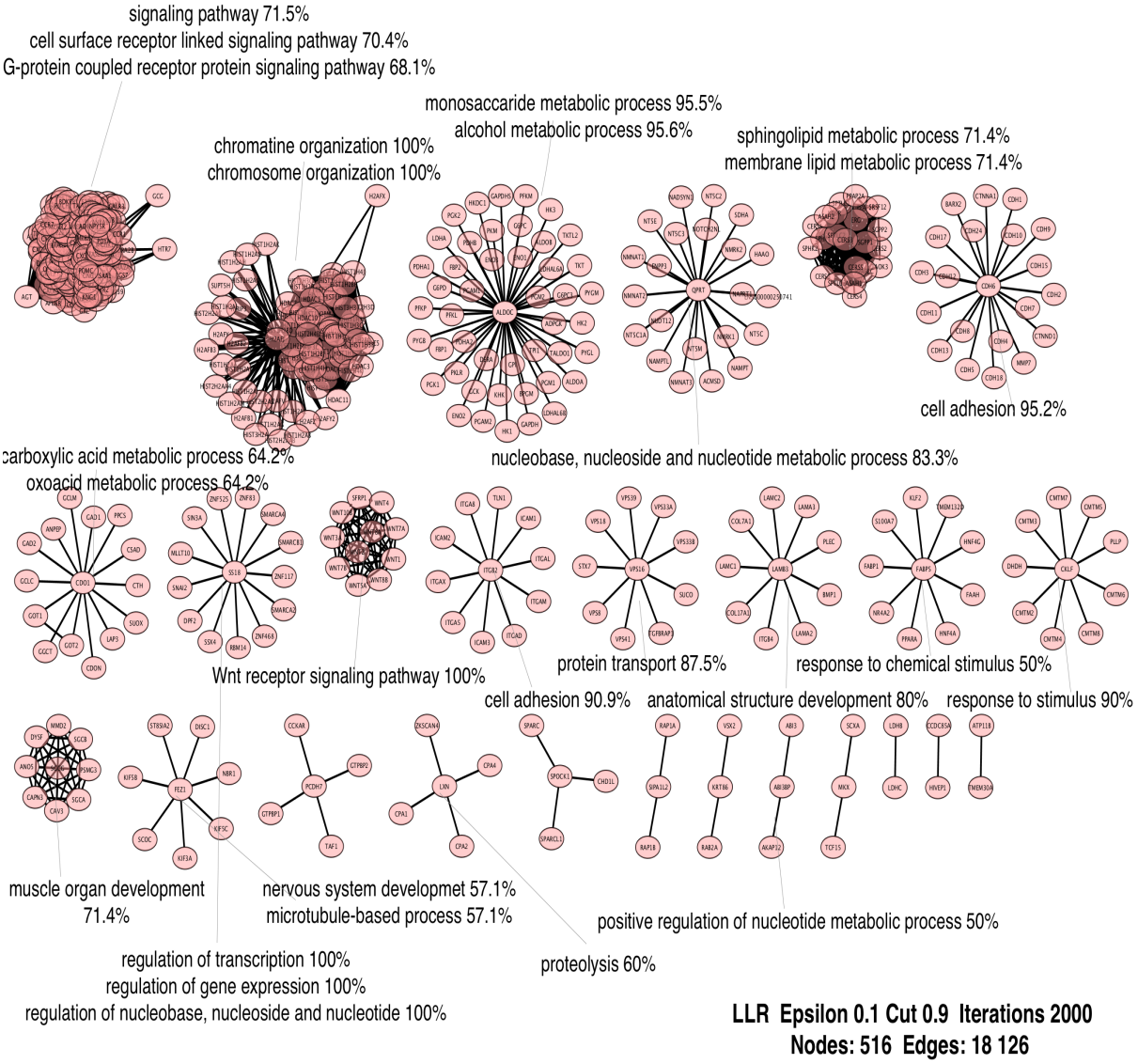


Figure 6.12: *LLR network with* $\epsilon = 0.1$ *(BiNGO analysis)*

The network in figure 5.16 was subjected to the action of the LLR algorithm. The cutoff used was 0.9 and $\epsilon = 0.1$*. The number of iterations performed was* $2,000$*. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. BiNGO [80] analysis in Cytoscape [11] was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in. These analyses revealed that the top pathways in which hsa-miR-218-5p target proteins were involved include: cell adhesion, signaling pathway, cell surgace receptor linked signaling pathway, G-protein coupled receptor protein signaling pathway, WNT-receptor signaling pathway, etc. Most of these pathways are cancer related pathways. The top enriched biological process terms associated with the proteins were regulation of transcription, regulation of gene expression, regulation of nucleobase, nucleoside and nucleotide, and nucleic acid metabolism, chromatin organization, regulation of transcription, regulation of gene expression, response to stimulus, nervous system development, etc. Most of these biological processes are related to cancer development.*

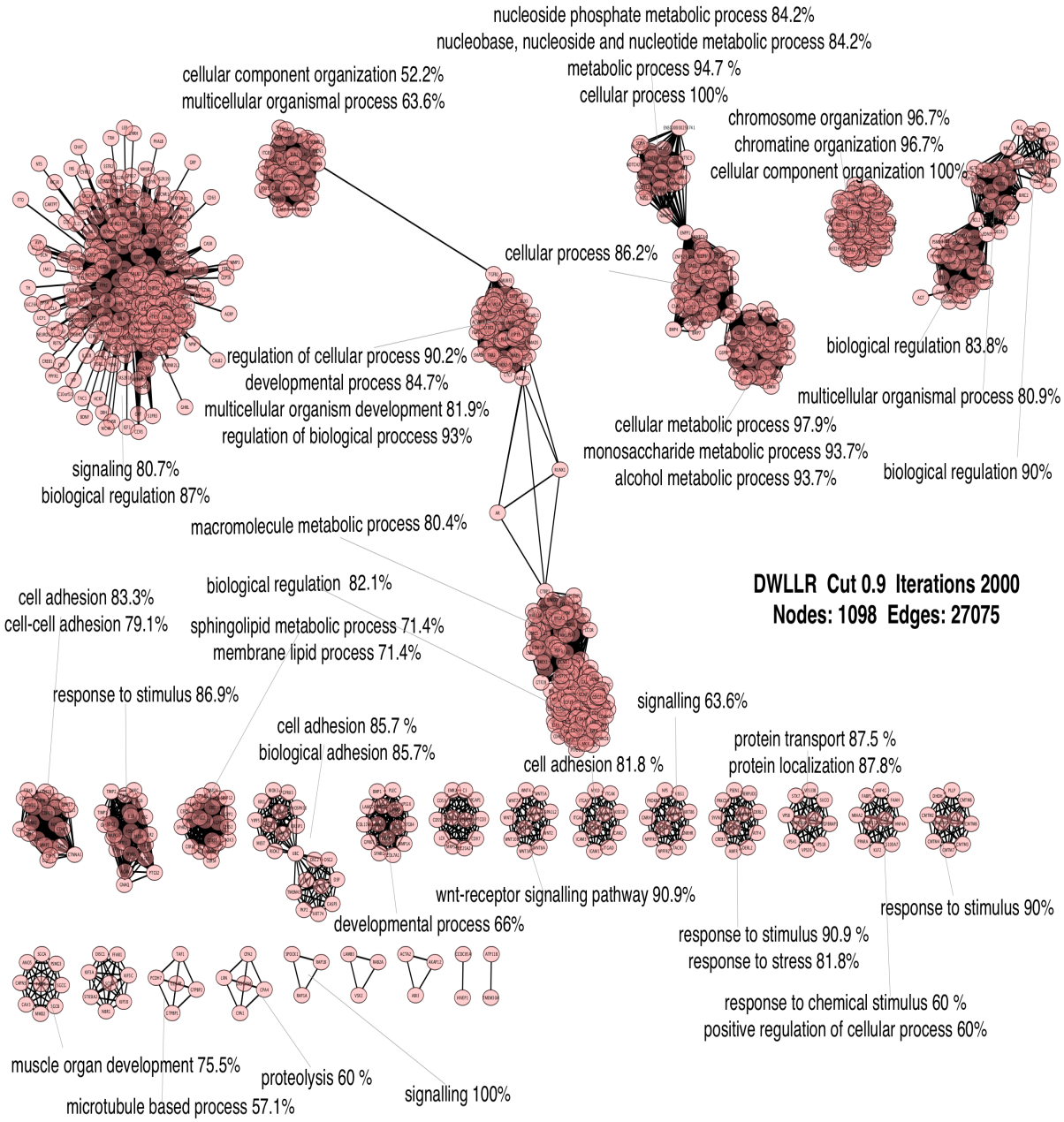


Figure 6.13: *DWLLR network with* $\epsilon = 0.1$ *(BiNGO analysis)*

The network in figure 5.16 was subjected to the action of the DWLLR algorithm. The cutoff used was 0.9 and $\epsilon = 0.1$*. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. BiNGO [80] analysis in Cytoscape [11] was used to identify the main biological pathways and biological processes that proteins in each of the clusters are involved in. These analyses revealed that the top pathways in which hsa-miR-218-5p target proteins were involved include: cell adhesion, WNT-receptor signaling pathway, signaling pathways etc. Most of these pathways are cancer related pathways. The top enriched biological process terms associated with the genes were: nucleoside phosphate metabolic processes, nucleobase, nucleoside and nucleotide metabolic processes, chromosome organization, chromatin organization, response to stimulus, muscle organ development, etc. Most of these biological processes are related to cancer development.*

Although different algorithms do not produce identical networks, they all show similar networks of biologically related clustered proteins. Most biological processes and pathways that clusters of proteins are involved in are repeatedly seen in all algorithms analyzed (RWR, LLR, DWLLR). These analyses revealed that the top pathways in which hsa-miR-218-5p target proteins were involved included cell adhesion and cancer related pathways. The top enriched biological process terms associated with the genes were transcription regulation and nucleic acid metabolism, which are related to cancer development.

6.6.2 The protein-protein interaction environment of hsa-miR-218-5p

The network in figure 5.16 was subjected to the action of the DWLLR algorithm with $\epsilon = 0$. Some of the most upregulated and downregulated proteins (explained in section 5.3) by hsa-miR-218-5p were not picked up by the algorithm in the newly reconstructed network. I reinserted those proteins again in the network presented in figure 6.14. The 24 most downregulated proteins are shown in red and the 38 most upregulated proteins are shown in green. Hsa-miR-218-5p is shown in yellow.

Using BiNGO [80] analysis in Cytoscape [11], the main molecular functions and biological processes of the proteins in each cluster are identified. If beside a cluster the biological process indicated is *"cell-cell adhesion 40%"*, it means, in that particular cluster 40% of the proteins are involved in cell-cell adhesion processes.

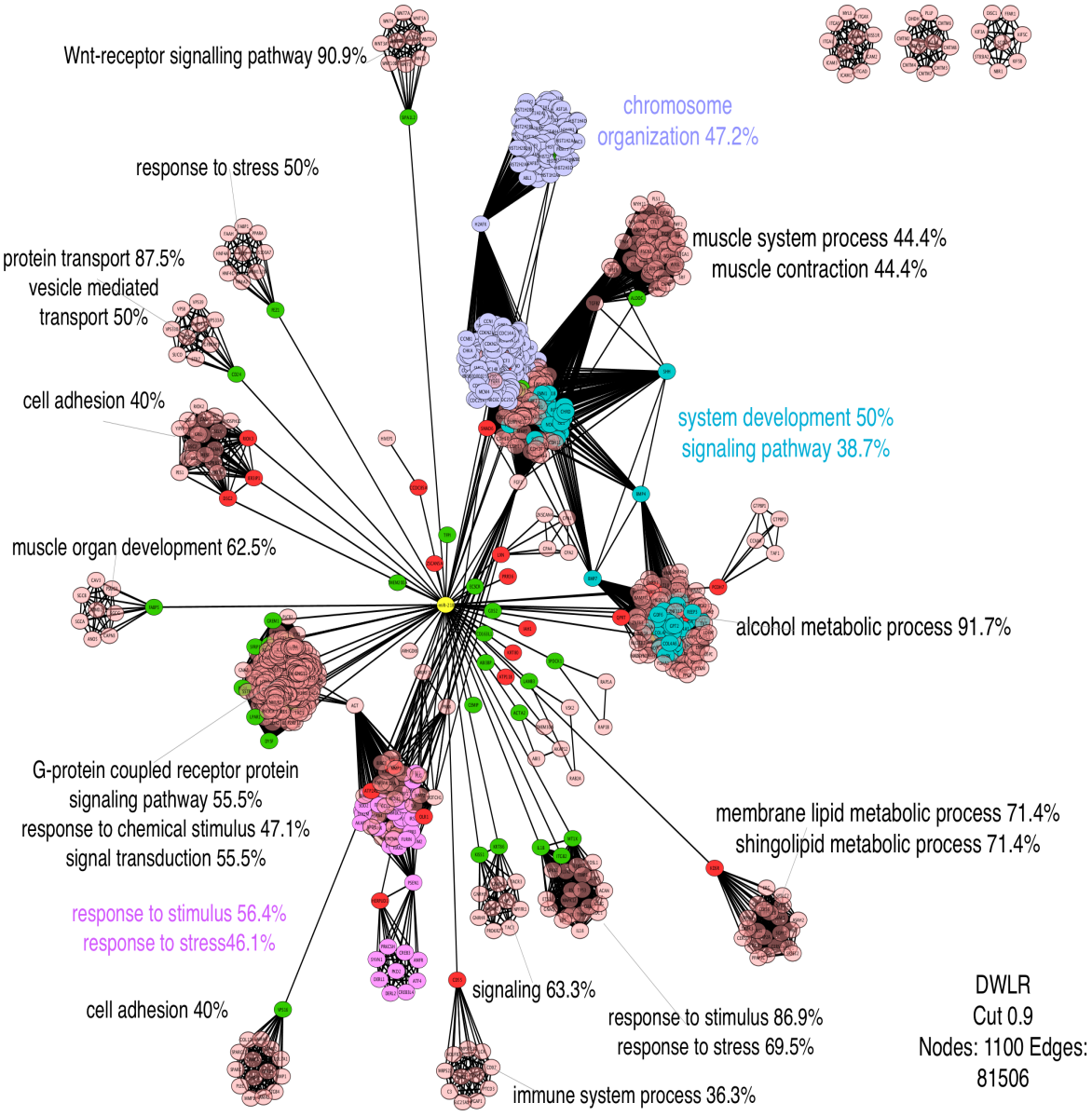


Figure 6.14: *PPI environment of hsa-miR-218-5p*

Gene regulation by hsa-miRNA-218-5p results in up or downregulation of proteins in protein interaction networks. The reconstructed network of PPIs of proteins controlled by hsa-miR-218-5p using the double weighted lattice laplacian with resistance (DWLLR) algorithm is shown. The network in figure 5.16 was subjected to the action of the DWLLR algorithm with $\epsilon = 0$*. The 24 most downregulated (red) proteins and 38 most upregulated (green) proteins by hsa-miR-218-5p were reinserted again in the network. The network has* $1,128$ *nodes and* $81,524$ *bonds. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. Hsa-miR-218-5p is shown in yellow. The network obtained has a clustered structure. Using BiNGO [80] analysis in Cytoscape [11], the main biological functions of the proteins in each cluster were identified. These analyses revealed that the top pathways in which hsa-miR-218-5p target proteins were involved included cell adhesion and cancer related pathways. The top enriched biological process terms associated with the proteins were transcription regulation and nucleic acid metabolism, which are related to cancer development.*

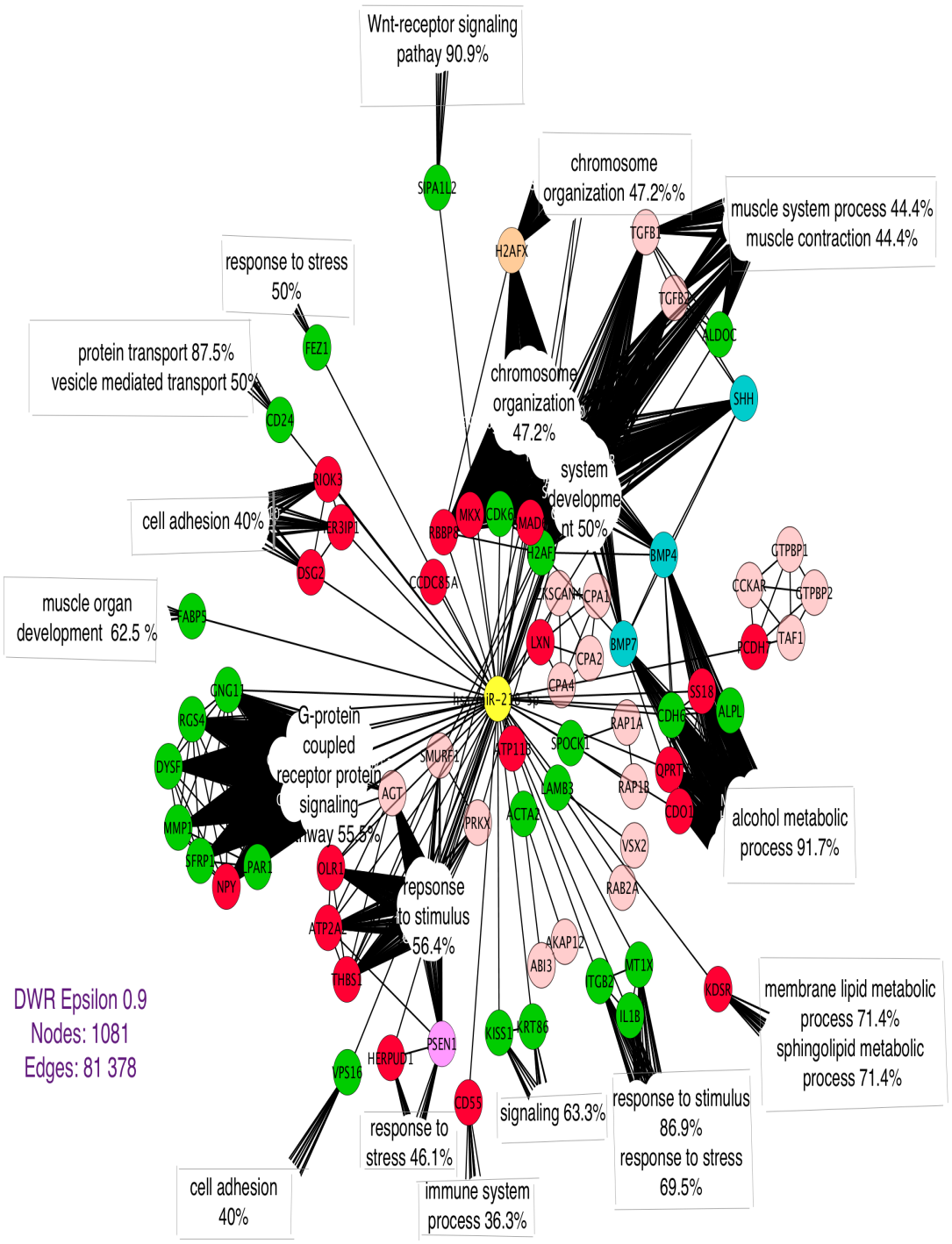


Figure 6.15: *PPI environment of hsa-miR-218-5p (clusters removed)*

The clusters of proteins in the network in figure 6.14 are removed for better clarity. I am replacing the clusters of proteins with the corresponding pathways and biological processes that most proteins in the cluster are involved in. The percentage beside each biological process or pathway shows the percentage of proteins in the cluster that are involved in that particular function.

For simplicity, I am replacing the clusters of proteins with the corresponding pathways and biological processes that most proteins in the cluster are involved in. The percentages given beside each cluster in figures 6.14 and 6.15 give the portion of the proteins in the cluster that have the molecular function or are involved in the biological process indicated.

As seen from the networks in figures 6.14 and 6.15, some of the biological pathways and processes that clusters of proteins in the PPIs of hsa-miR-218-5p are involved in include cell adhesion, signaling pathways, G-protein coupled receptor signaling pathway, WNT-receptor signaling pathway, chromosome organization, response to stress, response to stimulus, muscle organ development, immune system processes, alcohol metabolic processes, membrane lipid metabolic processes, etc.

Finally, we performed the BiNGO analysis in Cytoscape on the entire network in figure 6.14, instead of on individual clusters. The top *biological processes* (pathways and larger processes made up of the activities of multiple gene products), *molecular functions* (the elemental activities of a gene product at the molecular level) and *cellular components* (where gene products are active) of proteins in the PPIs of hsa-miR-218-5p are given in tables 6.2, 6.3 and 6.4.

The results in tables 6.2, 6.3 and 6.4 indicate that signaling pathway was one of the top pathways identified as a main biological process of proteins in the PPIs of hsa-miR-218-5p. Protein binding, DNA binding and nucleotide binding were amongst the top molecular functions of proteins in the PPIs of hsa-miR-218-5p. About 36.8% of proteins in PPIs of hsa-miR-218-5p perform their biological functions in the organelles of the cell and 38.2% in the nucleus.

Table 6.2: *Biological processes of proteins in hsa-miR-218-5p PPIs environment*

Biological process	**Percentage of genes in the network**
cellular process	82.5%
biological regulation	66.7%
regulation of cellular process	60.9%
multicellular organismal process	47.9%
response to stimulus	47.3%
signaling	45.3%
regulation of metabolic process	40.9%
signaling pathway	38.7%
developmental process	35.9%

Table 6.3: *Molecular functions of proteins in hsa-miR-218-5p PPIs environment*

Molecular Function	**Percentage of genes in the network**
binding	90.9%
protein binding	70.8%
signal transducer activity	27.4%
molecular transducer activity	27.4%
DNA binding	21.6%
receptor activity	19.1%
nucleotide binding	17.9%

Table 6.4: *Cellular components of proteins in hsa-miR-218-5p PPIs environment*

Cellular Components	**Percentage of genes in the network**
intracellular part	70.6%
membrane-bounded organelle	55.1%
intracellular membrane-bounded organelle	54.8%
organelle part	38.9%
nucleus	38.2%
cytoplasmic part	36.8%
plasma membrane	35.4%

In conclusion, these analyses revealed that the top processes that hsa-miR-218-5p target proteins were involved included immune system processes, muscle organ development, response to stress and cancer development.

6.6.3 Identifying hubs

The networks described in the previous section were analyzed with the goal of identifying hubs (proteins of high degrees) in each cluster. These hubs are shown in figures 6.16, 6.17 and 6.18.

The network in figure 5.16 was subjected to the action of the RWR, LLR and DWLLR algorithms. The cutoff used was 0.9. The number of iterations was 2,000. The new networks were created using Cytoscape [11] and they all show clustered structures. Hubs (nodes of high degrees) were identified for each cluster. The results are summarized in table 6.5.

Table 6.5: *Hubs in PPIs of hsa-miR-218-5p (figures 6.16, 6.17 and 6.18)*

Hubs	**Degrees**
NPY	308
GALR2	267
RGS4	267
CXCL1	267
PRGER3	252
GRM3	252
H2AFJ	78
HIST2H2B2	78
HIST1H2BK	78
HIST1H2BDE	76
IL6	27

Some of the identified hubs in the RWR network include: HIST1H2BD (degree 76), HIST1H2BK (degree 75) and IL6 (degree 27). In the LLR network, some of the hubs identified are: H2AFJ (degree 78), HIST2H2BD (degree 78), HIST1H2BK (degree 78) and ALDOC (degree 46). Identified hubs in the DWLLR network include: NPY (degree 308), GALR2 (degree 267), RGS4 (degree 267), CXCL1 (degree 267), PRGER3 (degree 252) and GRM3 (degree 252). Out of the above genes, the only one which is already a target of hsa-miR-218-5p in

miRTarBase [39] is HIST1H2BK. The other genes might be potential targets of hsa-miR-218-5p.

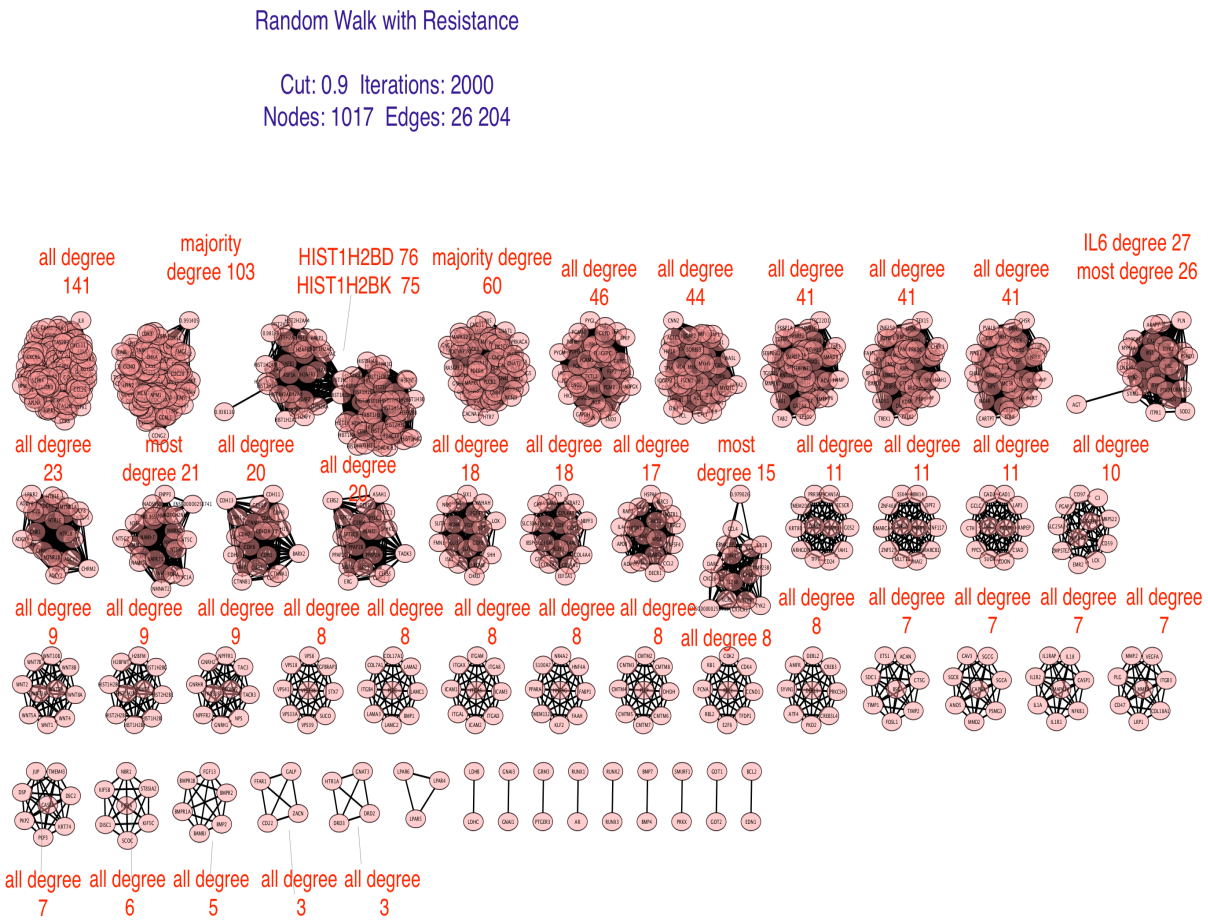


Figure 6.16: *RWR network (Hubs)*

The network in figure 5.16 was subjected to the action of the random walk with resistance algorithm. The cutoff used was 0.9. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. Hubs (nodes of high degrees) are identified for each cluster. Some of the hubs are HIST1H2BD (degree76), HIST1H2BK (degree 75), IL6 (degree 27), etc.

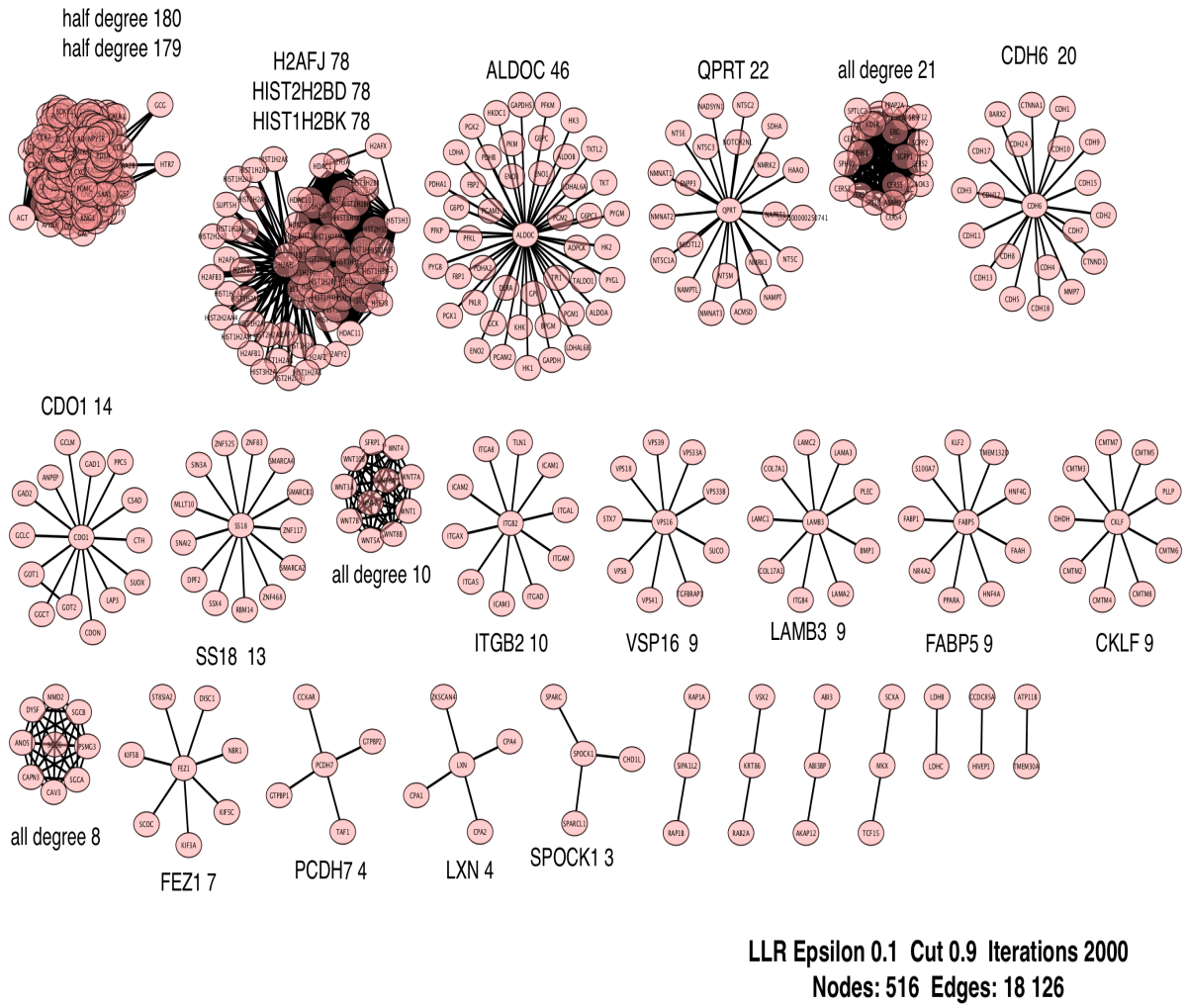


Figure 6.17: *LLR network with* $\epsilon = 0.1$ *(Hubs)*

The network in figure 5.16 was subjected to the action of the LLR algorithm. The cutoff used was 0.9 and $\epsilon = 0.1$*. The number of iterations performed was* $2,000$*. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. Hubs (nodes of high degrees) are identified for each cluster. Some of the hubs are H2AFJ (degree 78), HIST2H2BD (degree 78), HIST1H2BK (degree 78), ALDOC (degree 46), etc.*

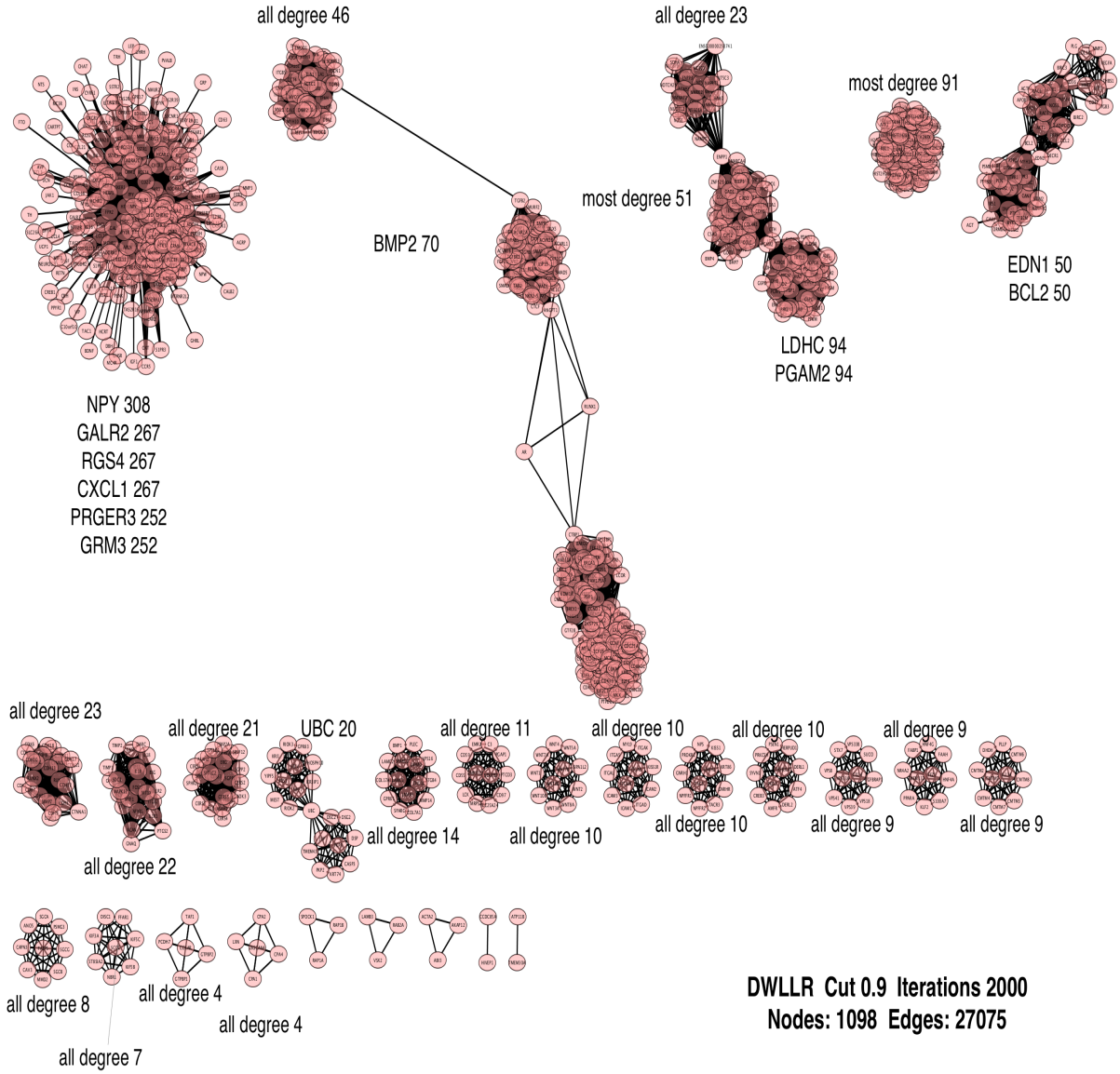


Figure 6.18: *DWLLR network with* $\epsilon = 0.1$ *(Hubs)*

The network in figure 5.16 was subjected to the action of the DWLLR algorithm. The cutoff used was 0.9 and $\epsilon = 0.1$*. The number of iterations performed was* $2,000$*. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0. Hubs (nodes of high degrees) are identified for each cluster. Some of the hubs are: NPY (degree 308), GALR2 (degree 267), RGS4 (degree 267), CXCL1 (degree 267), PRGER3 (degree 252), GRM3 (degree 252), etc.*

Network analysis in Cytoscape [11] was used to identify important proteins (hubs) in the PPI environment of hsa-miR-218-5p network in figure 6.14 and the results are shown in table 6.6.

Table 6.6: *Hubs in PPIs of hsa-miR-218-5p (figure 6.14)*

Hubs	**Degrees**
RGS4	309
SFRP1	309
MMP1	309
GREM1	309
GNG11	309
DYSF	309
LPAR1	309
CXCL1	309
GALR2	309
NPY	309
GNAZ	309
GNG1	309
PLD2	308
CTNNBL1	308

The above genes might be potential targets of hsa-miR-218-5p. None of them is a verified or predicted target of hsa-miR-218-5p in the miRTarBase [39] database.

6.6.4 Hsa-miR-218-5p target genes implicated in diseases

In this case study, the upregulated and downregulated proteins by hsa-miR-218-5p were taken as source proteins for a search in the STRING [38] database to generate a network which was then subjected to the action of the algorithms. In general the algorithms above reveal a clustered structure in the networks. The clusters correspond to complexes of functionally related proteins and in the case of the RWR reveals a reaction pathway involving

protein clusters.

These approaches seek to find ways of solving complex pathway membership problems in protein interaction databases. The clusters obtained provide more biological insight as opposed to a process of local pairwise comparison between interacting proteins. They may also predict new members in functional pathways or clusters.

Underlying these algorithms are simulated biased random walks on the network for determining membership of proteins in given clusters. This places the algorithms in a class of random walk algorithms examined in the literature.

The output consists in many cases of identifiable clusters of functionally related proteins or pathways. Several of the genes seen in the final network (figure 6.14) were identified as targets of hsa-miR-218-5p in several studies in the literature (discussed below). These proteins were implicated in different states of several diseases.

Hsa-miR-218-5p and metastatic bone disease of breast cancer

In 2016, Yu Xin et al [41], found that levels of hsa-miR-218-5p are positively correlated with the activation of WNT signaling pathway in breast cancer cells. WNT signaling is implicated in bone formation and activated in breast cancer cells promoting primary and metastatic tumor growth [41]. The WNT cluster is seen in the networks in figures 6.1 and 6.2.

Hsa-miR-218-5p is highly expressed in bone metastases from breast cancer patients. Inhibition of this miRNA impaired the growth of bone metastases from breast cancer cells in the bone microenvironment in vivo [41]. Hsa-miR-218-5p targets the WNT inhibitors

Sclerostin (SOST) and sFRP-2, which highly enhance WNT signaling. Contrary, delivery of antimiR-218-5p decreases WNT activity and the expression of metastasis-related proteins, including bone sialoprotein (BSP/IBSP), osteopontin (OPN/SPP1) and CXCR-4 [41].

Most of the proteins discussed in the above study, involved in hsa-miR-218-5p action mechanisms in metastatic bone disease of breast cancer cell like WNT, FRP-2, SSP1, IBSP, CXCR-4 are also seen in our final networks generated by solving the DWLLR on the PPI environment of hsa-miR-218-5p (figure 6.14).

A subset of the final network containing the above mentioned proteins is shown in figure 6.19. The full network with the marked proteins is shown in figure 6.20. The proteins shown in green are the ones involved in metastatic bone disease of breast cancer cells.

The WNT cluster of genes consists of structurally related genes which encode secreted signaling proteins [79]. These proteins have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis. They are very conserved in evolution [79].

The CXCR-4 gene encodes a CXC chemokine receptor specific for stromal cell-derived factor-1. The protein has 7 transmembrane regions and is located on the cell surface. It acts with the CD4 protein to support HIV entry into cells and is also highly expressed in breast cancer cells [79].

SPP1 codes for a protein which is involved in the attachment of osteoclasts to the mineralized bone matrix. The encoded protein is secreted and binds hydroxyapatite with high affinity. This protein is also a cytokine that upregulates expression of interferon-gamma and interleukin-12 [79].

IBSP encodes a major structural protein of the bone matrix. This protein binds to calcium and hydroxyapatite via its acidic amino acid clusters, and mediates cell attachment [79].

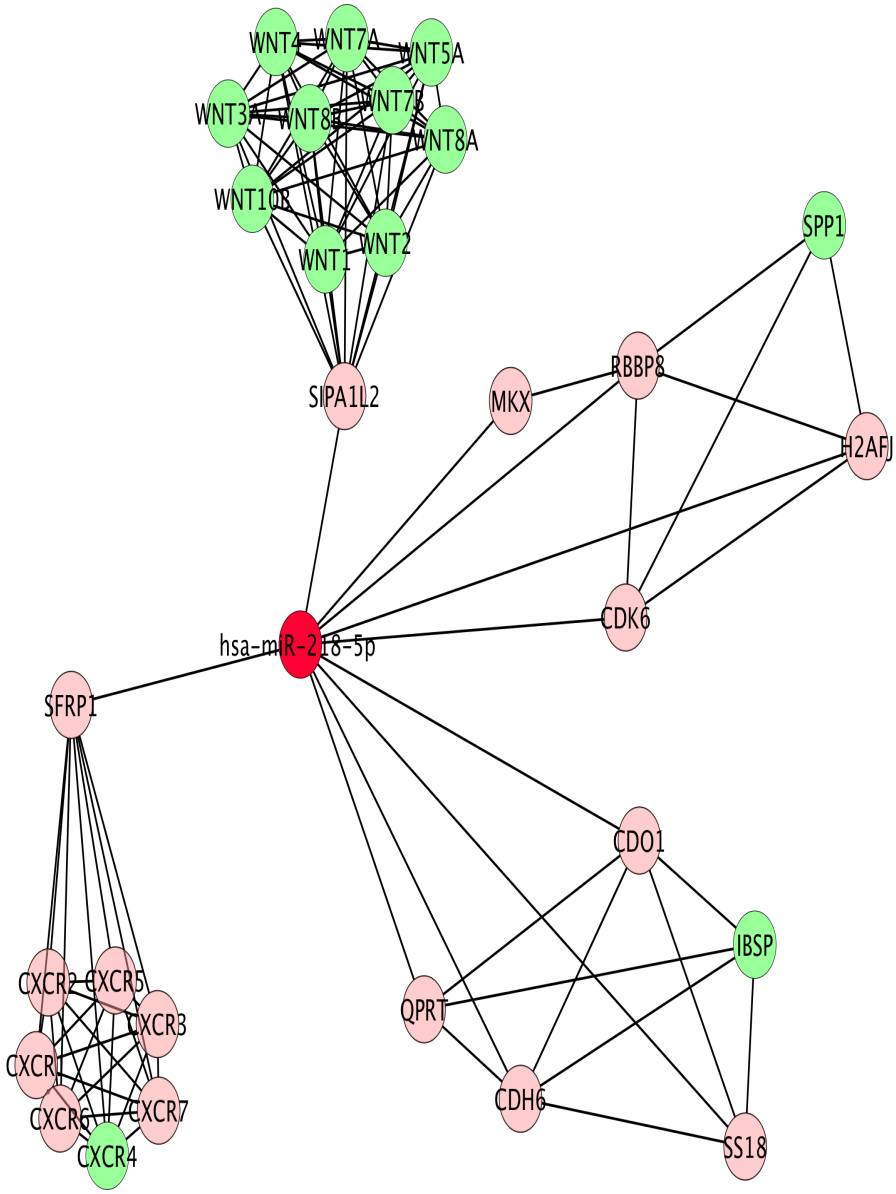


Figure 6.19: *Hsa-miR-218-5p implications in metastatic bone disease of breast cancer cells*

The network is a subset of the network in figure 6.14. The proteins involved in hsa-miR-218-5p action mechanisms in metastatic bone disease of breast cancer are shown in green. These proteins are WNT, SSP1, IBSP, CXCR-4.

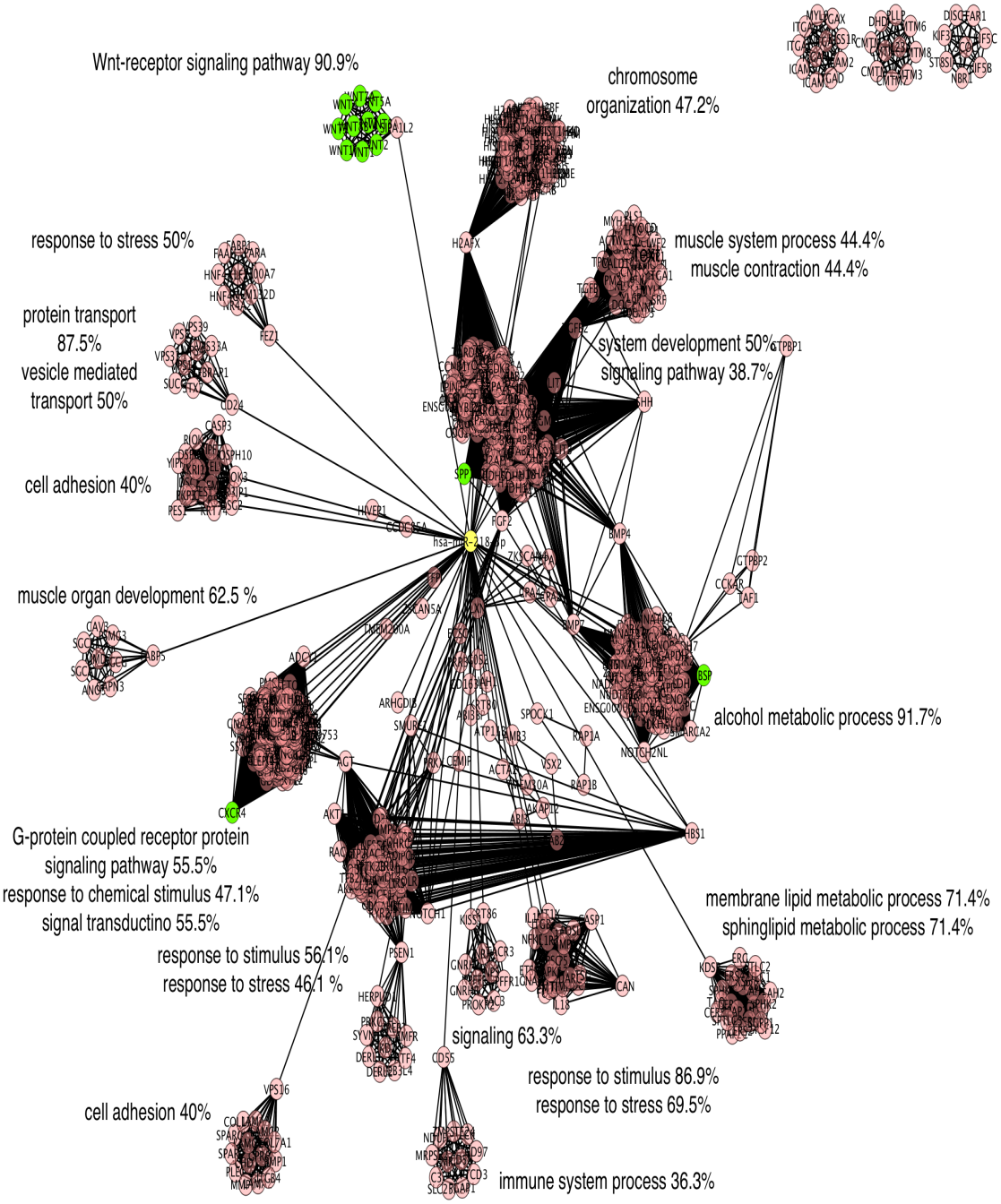


Figure 6.20: *Hsa-miR-218-5p implications in metastatic bone disease of breast cancer cells (full network)*

The network in figure 6.14 with proteins in figure 6.19 marked. The proteins involved in hsa-miR-218-5p action mechanisms in metastatic bone disease of breast cancer are shown in green. These proteins are WNT, SSP1, IBSP, CXCR-4. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

Hsa-miR-218-5p and prostate cancer

In 2015, Sanchez et al showed that a group of 19 miRNAs were overexpressed in prostate cancer [42]. Hsa-miR-218-5p was one of them. The different prostate cancer cell populations (bulk and cancer stem cells, CSCs) release exosomes that contain miRNAs that could modify the local or premetastatic niche [42]. The analysis of the differential expression of miRNAs in exosomes (cell-derived vesicles) allows evaluating the differential biological effect of both populations in the niche, and the identification of potential biomarkers and therapeutic targets [42]. 1839 miRNAs were identified in the exosomes. Of these, 990 were known miRNAs, from which only 19 were significantly differentially expressed: 6 were overexpressed in CSCs and 13 in bulk cells exosomes [42]. One of the 13 main miRNAs that is overexposed in bulk cells is hsa-miR-218-5p. One of the effects of the overexpression of these miRNAs was increased levels of metalloproteinases MMP2, MMP9 and MMP13. Metalloproteinases are zinc-dependent enzymes capable of cleaving components of the extracellular matrix and molecules involved in signal transduction [79].

In prostate cancer, secreted frizzled-related protein 1 (SFRP1), TGFβ1 and stromal cell-derived factor 1 (SDF-1/CXCL12) are all candidate molecules for inducing tumourigenicity [43]. Members of SFRP family of proteins act as soluble modulators of WNT signaling; silencing of SFRP genes leads to deregulated activation of the WNT-pathway which is associated with cancer [79]. SDF-1 is a cell adhesion molecule and a member of the immunoglobulin superfamily. The receptor for SDF-1 (SDFR-1, or neuroplastin) is expressed by prostate stem cells 48, indicating that in cancer SDF-1 signaling pathways are likely to be important. SDF-1 - CXCR4 signaling can induce cancer-like behaviour, such as activa-

tion of anti-apoptotic pathways 53, motility, homing and adhesion during embryogenesis, organogenesis and metastasis [43].

The above are two separate studies, one describing the different miRNAs involved in prostate cancer and the second different proteins which produce candidate molecules for inducing tumourigenicity in prostate cancer. Some of these genes like SFRP1, CXCR4, CXCL12, MMP2, MMP9 were also seen in our final PPI environment of hsa-miR-218-5p network (figure 6.14) generated by applying the DWLLR algorithm on the network in figure 5.16. The subset of this network containing the above proteins is extracted from the full network and is given in figure 6.21. The proteins involved in prostate cancer are shown in green. The full network is shown in figure 6.22.

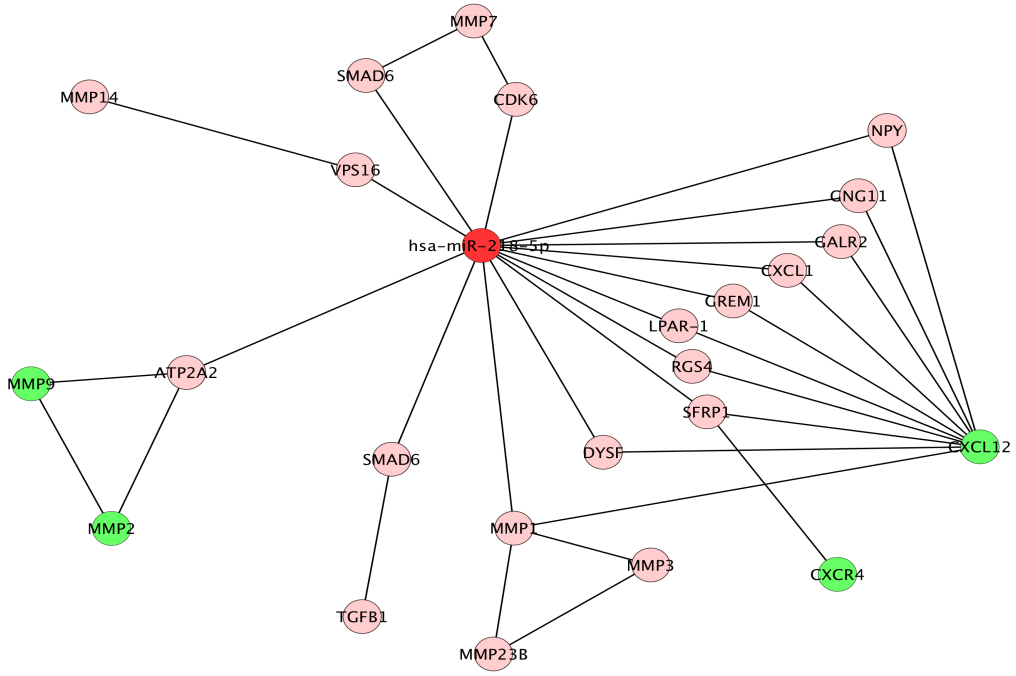


Figure 6.21: *Hsa-miR-218-5p implications in prostate cancer*

The network is a subset of the network in figure 6.14. Hsa-miR-218-5p target proteins implicated in metastatic prostate cancer are shown in green. These proteins are SFRP1, CXCR4, CXCL12, MMP2, MMP9.

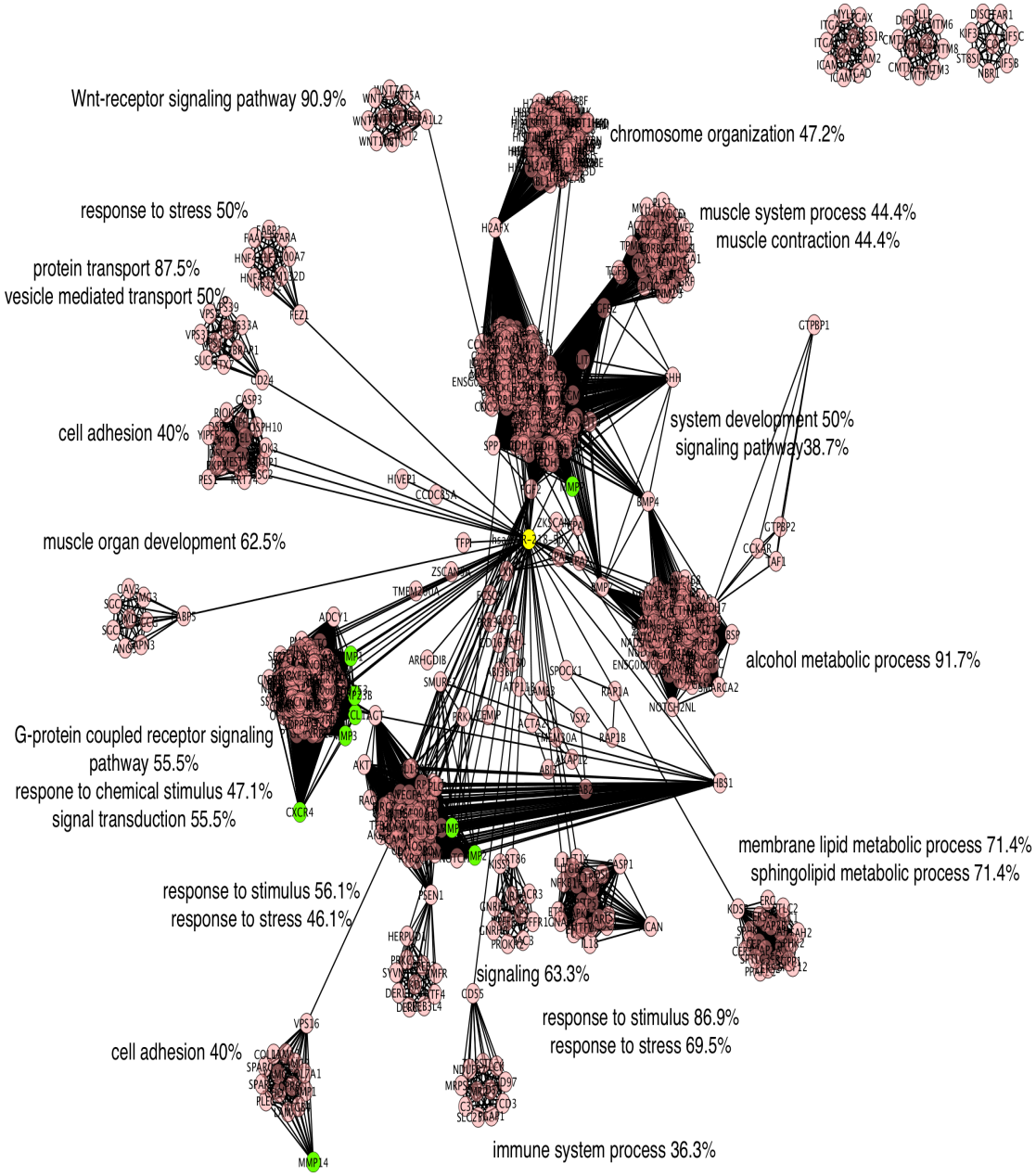


Figure 6.22: *Hsa-miR-218-5p implications in prostate cancer (full network)*

The network in figure 6.14 with proteins in figure 6.21 marked. The target proteins of hsa-miR-218-5p implicated in prostate cancer are shown in green. These proteins are SFRP1, CXCR4, CXCL12, MMP2, MMP9. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

Hsa-miR218-5p and gallbladder cancer

Gallbladder cancer is the fifth most frequent gastrointestinal malignancy [41]. Colon cancer-associated transcript-1 (CCAT1) is a long non-coding RNA [41]. Its expression is higher in gallblader cancer tissues compared with adjacent normal tissues. CCAT1 over-expression increased the expression of Bmi1, a target gene of hsa-miR-218-5p [41]. The knockdown (experimental technique by which the expression of one or more of an organism's genes are reduced) of CCAT1 inhibited the proliferation of gallbladder cancer cells, partially through regulation of Bmi1 and involvement of hsa-miR-218-5p [41].

The Bmi1 expression interacts with several signaling containing WNT, AKT, NOTCH, Hedgehog and receptor tyrosine kinase (RTK) pathway [44]. In Ewing sarcoma family of tumors (ESFT), the knockdown of the Bmi1 gene would greatly influence the Notch and WNT signaling pathway which are important for ESFT formation and development [44].

Protein kinase B (PKB), also known as AKT, is a serine/threonine-specific protein kinase. The NOTCH signaling pathway is a highly conserved cell signaling system present in most multicellular organisms. Mammals possess four different NOTCH receptors, referred to as NOTCH1, NOTCH2, NOTCH3, and NOTCH4 [45].

Even though Bmi1, the gene overexpressed by hsa-miR-218-5p in gallbladder cancer is not found in our final graphs, the genes involved in the signaling pathways controlled by Bmi1 like AKT, NOTCH and WNT genes are seen on it. The subset of the network containing these genes is given in figure 6.23. The proteins in green are the ones involved in the gallbladder cancer. The full network is shown in figure 6.24.

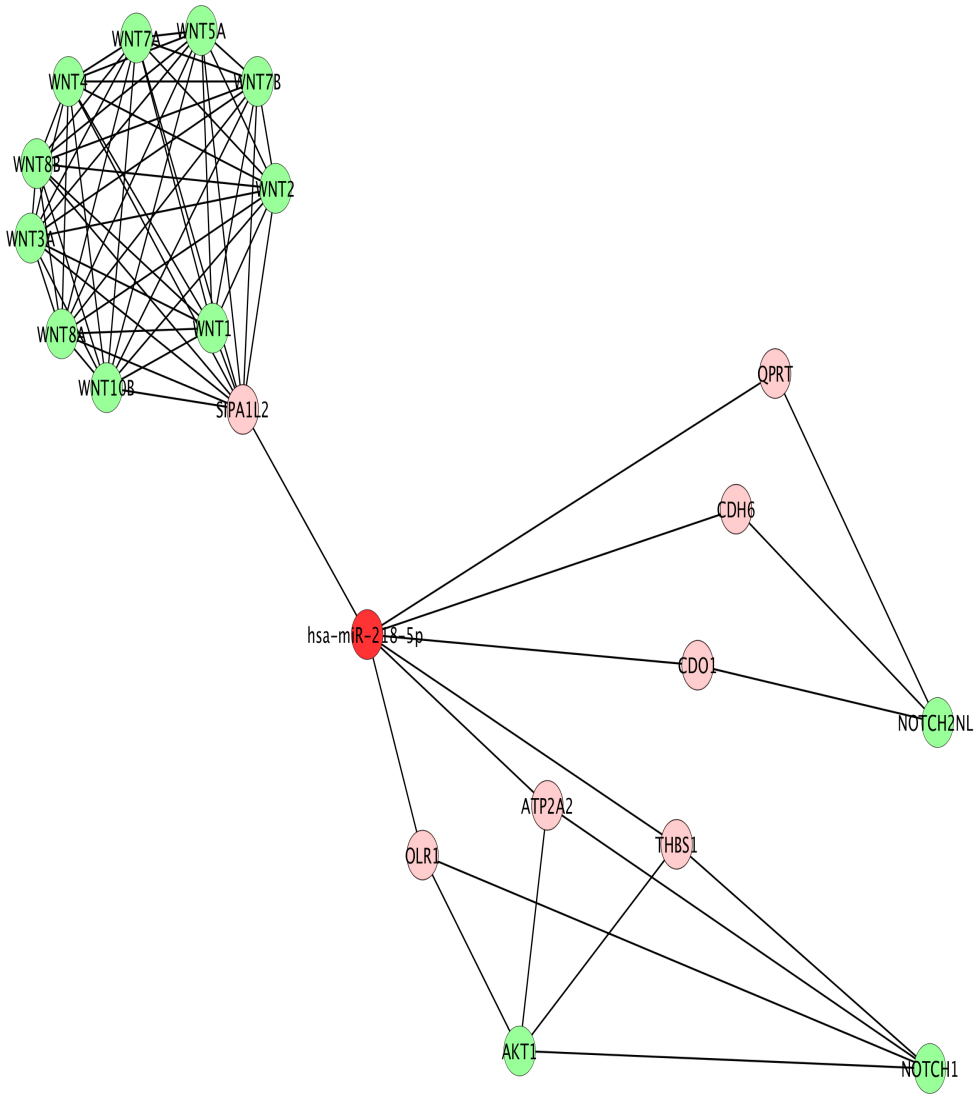


Figure 6.23: *Hsa-miR-218-5p implications in gallbladder cancer*

The graph is a subset of the network in figure 6.14. The proteins involved in the signalling pathways controlled by Bmi1 in gallbladder cancer are AKT, NOTCH and WNT genes, shown in green.

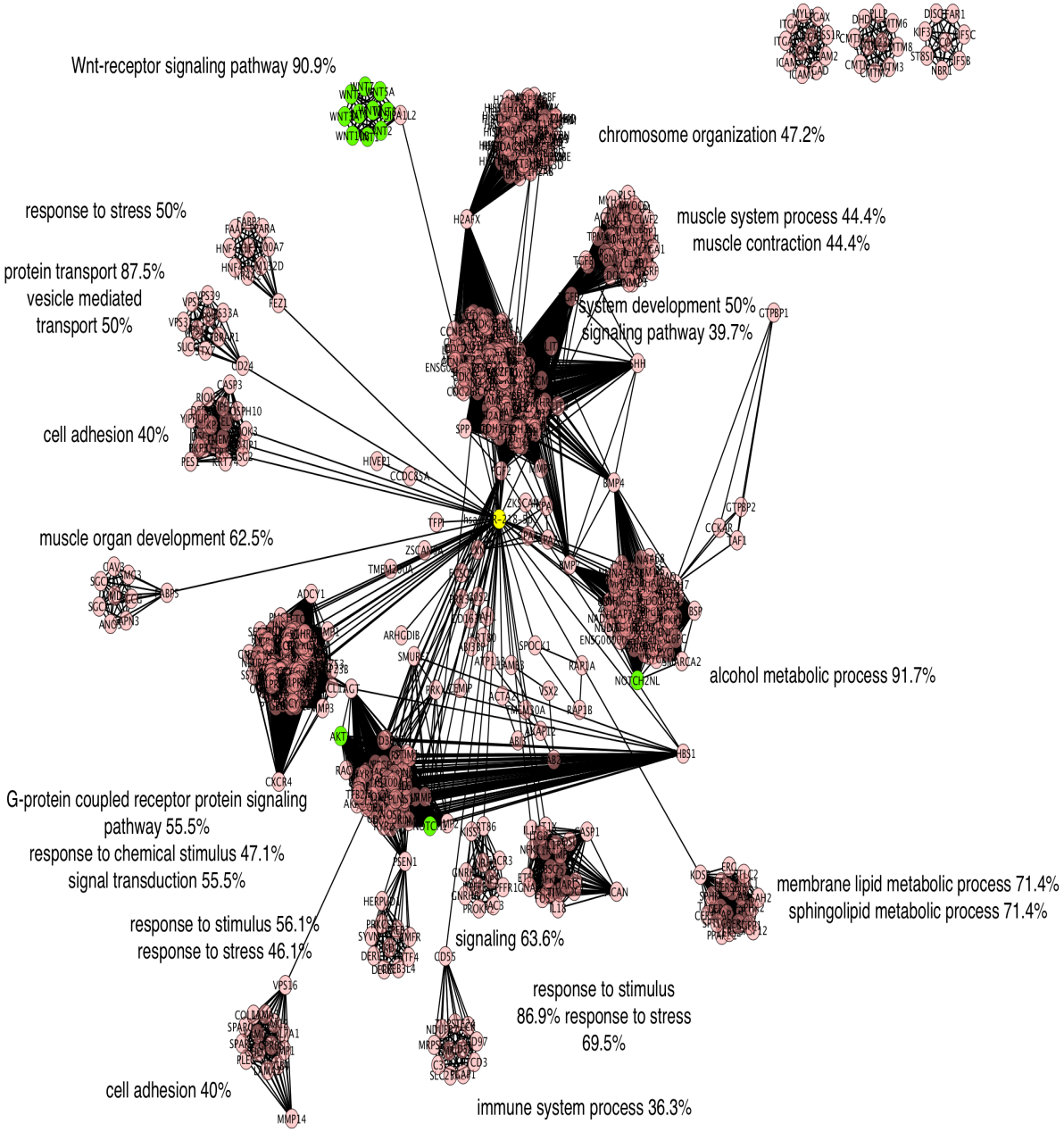


Figure 6.24: *Hsa-miR-218-5p implications in gallbladder cancer (full network)*

The network in figure 6.14 with proteins in figure 6.23 marked. The target proteins of hsa-miR-218-5p implicated in gallbladder cancer are shown in green. These proteins are AKT, NOTCH and WNT. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

Hsa-miR218-5p and epilepsy

There is increasing evidence that expression of miRNAs is dysregulated in neuronal disorders, including epilepsy, a chronic neurological disorder characterized by spontaneous recurrent seizures [49]. A miRNA screen was performed in hippocampal focal and non-focal brain tissue samples obtained from the temporal neocortex of Mesial Temporal Lobe Epilepsy patients. Recent reports from screening studies in animal epilepsy models and from human specimens of MTLE (mesial temporal lobe epilepsy) patients show that miRNA expression is also altered in epilepsy [49]. In 2015, Haenisch et al suggested that differential miRNA expression in neurons could contribute to an altered function of the transcription factor SOX11 and other proteins in the setting of epilepsy, resulting not only in impaired neural differentiation, but also in imbalanced neuronal excitability and accelerated drug export.

After screening genome-wide miRNA expression in hippocampal focal and non-focal brain tissue from the temporal neocortex of MTLE patients who underwent temporal lobectomy, the interaction between dysregulated miRNAs and predicted target genes potentially involved in hippocampal cellular remodeling during epileptogenesis was investigated [49].

Hsa-miR-218-5p is amongst the top 40 expressed miRNAs in healthy hippocampus. The intention underlying this filtering approach was to use the most highly expressed miRNAs to uncover genes that are more likely to be critical for normal brain function and to result in crucial functional impairments as a result of altered miRNA expression. The targets of hsa-miR-218-5p are ADCY1, BSN, MECP2 and SOX11. Only one of these proteins (shown in figure 6.25), the ADCY1, is observed in our final network in figure 6.14 generated with

the DWLLR algorithm. The full network is shown in figure 6.26.

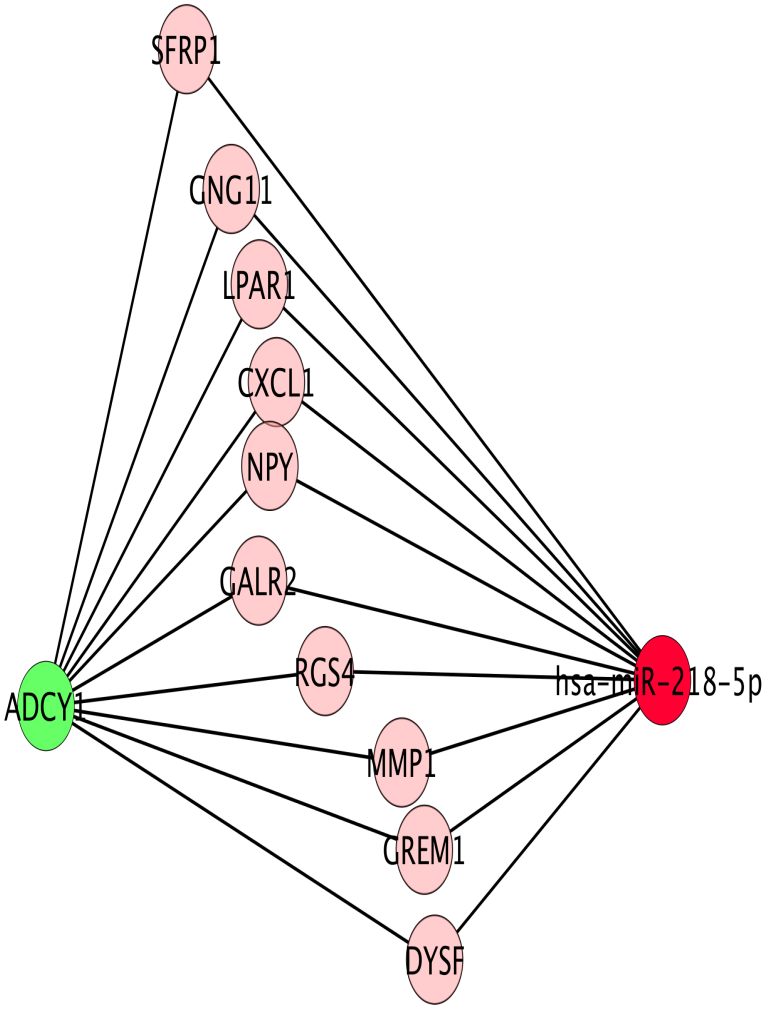


Figure 6.25: *Hsa-miR-218-5p implications in epilepsy*

The graph is a subset of the network in figure 6.14. One of the target proteins of hsa-miR-218-5p, involved in epilepsy and also seen in our final graph is ADCY1 (shown in green).

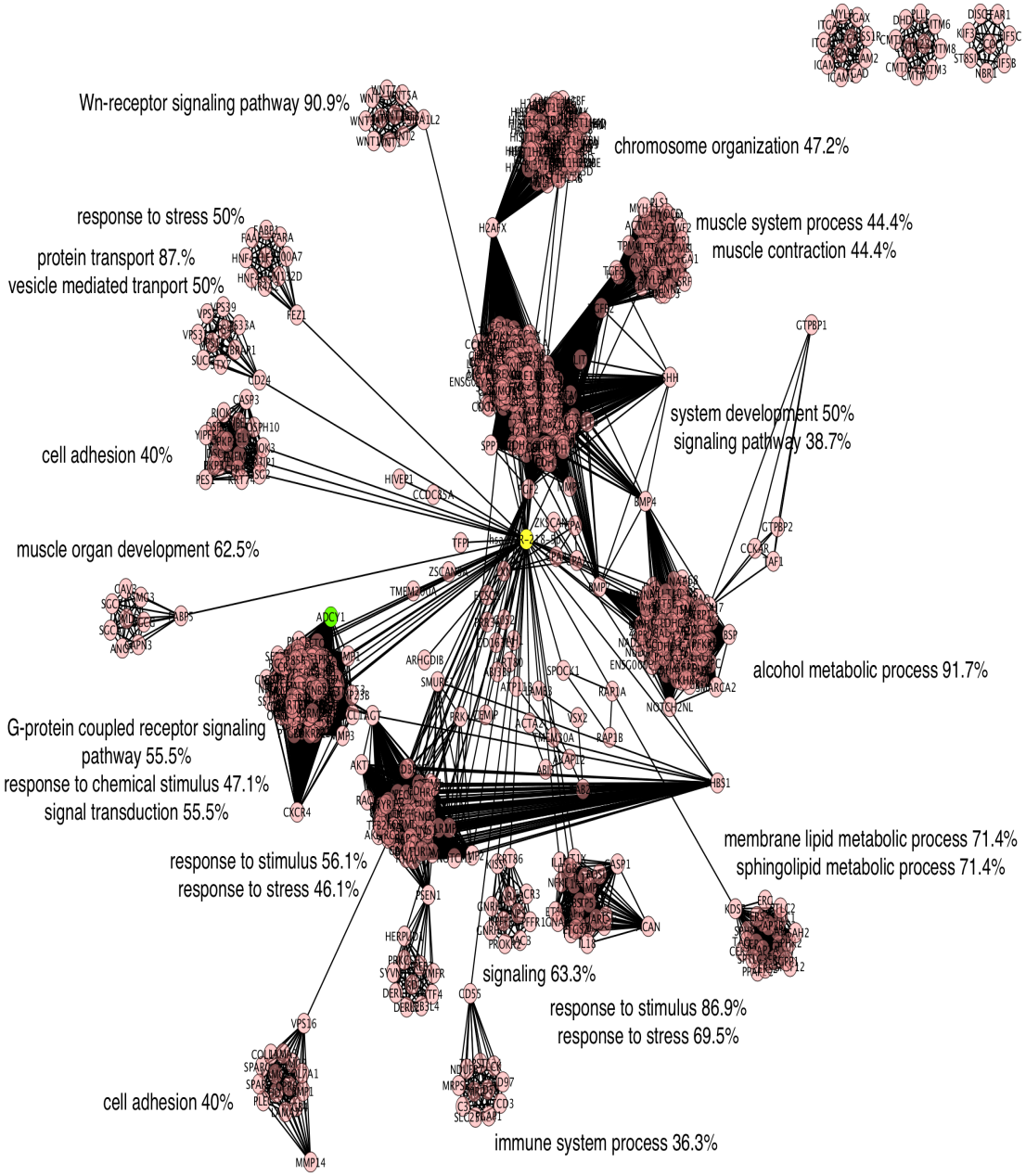


Figure 6.26: *Hsa-miR-218-5p implications in epilepsy (full network)*

The network in figure 6.14 with proteins in figure 6.25 marked. The target protein of hsa-miR-218-5p implicated in epilepsy (ADCY1) is shown in green. The arrangement of nodes and bonds in this network was created using the prefuse force directed layout in Cytoscape 3.3.0.

Chapter 7

Discussion and Future Work

7.1 Research Outcomes

7.1.1 Mean field analysis of algorithms for scale-free networks in molecular biology

In this research a number of algorithms used for generating networks in molecular biology were examined. Mean field theory for the algorithms was in some cases reviewed, and in other cases newly presented, and also refined. The algorithms include the Barabasi-Albert [1], Duplication-Divergence [27], Solé [26] and iSite algorithms [15, 16], and these were in some cases modified by the introduction of more general elementary moves. The mean field theory was newly presented in the cases of modified Barabasi-Albert, modified Duplication-Divergence and iSite algorithms. The above work sets a pattern for analyzing algorithms which generate scale-free networks in molecular biology. The majority of algo-

rithms described in this research model evolutionary processes which are thought to be the underlying mechanisms by which protein-protein interaction networks evolve. The same framework can be used to analyze other biological algorithms that model these evolutionary processes.

7.1.2 Random walk and laplacian analysis of microRNA-protein interaction networks

Analysis of biological networks (hsa-miR-218-5p protein-protein interaction environment) was considered and two different algorithms (random walk and lattice laplacian) were used with the goal of finding clusters of biologically related proteins and important genes (hubs) in the network.

Although different algorithms do not produce identical networks, they all show similar networks of biologically related clustered proteins. They reveal a cluster structure in the networks. Most biological processes and pathways that identified clusters of genes are involved in are repeatedly seen in the networks generated by all algorithms analyzed (RWR, LLR, WLLR, DWLLR).

Both the random walk and lattice laplacian algorithms are diffusive processes on a lattice. The RWR algorithm is a method aiming to improve the quality of the protein-protein interaction networks purely based on the topology of the network, with no additional biological information involved.

In the lattice laplacian with resistance algorithms, in addition to the basic idea of com-

paring the topological similarities of the nodes in the network, biological information obtained from the microarray data is involved as well. By including the weights on the bonds, according to the biological information provided by the microarray data, the aim is to improve the algorithms and provide networks which are biologically more meaningful.

The protein-protein interaction networks are usually sparse. The reconstruction of the networks relies on the information of the original networks. Since the networks of PPIs contain many false positive and false negatives, this will affect the completeness and the accuracy of the networks reconstructed by the algorithms.

The cutoff value we are using in the original data provided as microarray data by the Peng's lab is a fold change of 2. This means we are leaving out a large number of genes which might have important biological relevance and be possible targets of hsa-miR-218-5p. The size of the networks obtained using this cutoff value is around $2,000$ edges. Changing the cutoff value of the considered proteins beyond 2 will significantly increase the sizes of the networks obtained.

These analyses revealed that the top pathways in which hsa-miR-218-5p target genes were involved included cell adhesion and cancer related pathways. The top enriched biological process terms associated with the genes were transcription regulation and nucleic acid metabolism, which are related to cancer development.

Some of the genes which seem important in the network (hubs of high degrees) include HIST1H2BD, HIST1H2BK, IL6, H2AFJ, ALDOC, NPY, GALR2, RGS4, CXCL1, PRGER3 and GRM3. Out of the above genes, the only one which is already a target of hsa-miR-218-5p in miRtaRBase [39] is HIST1H2BK. The other genes might be potential targets of hsa-miR-

218-5p.

7.2 Future Work

7.2.1 Mean field analysis of algorithms for scale-free networks in molecular biology

The results in this research raise some questions about the sampling of scale-free networks by random iterative growth algorithms:

- In some cases, see for example reference [29], the parameters of the algorithms were set to grow networks with properties similar to that of real protein interaction networks. The values of the parameters are then used to estimate the rate of subfunctionalization (or mutation) in the genome. The results are dependent on the algorithm, and so further refinement of algorithms may be needed before useful estimates can be made.

- The mean field approaches are useful in some models (for example the Barabasi-Albert algorithm, and the iSite algorithm), but are poorer approximations in other models (the variant Barabasi-Albert algorithm, the Duplication-Divergence algorithm and its modification, and the Solé algorithm). Can the mean field approach be improved to give a better approximation to these algorithms?

- Investigation of some numerical properties of the networks (for example the connectivity) suggests that the algorithms may be self-averaging. That is, networks are

generated with properties which converge to the statistical averages of these properties over a sample of networks generated by the algorithm. This is, for example, illustrated in figure 4.16 for the connectivity of Barabasi-Albert networks. As the network is grown, its connectivity appears to approach the average connectivity over a large sample of networks.

- In this research some algorithms were modified in ways not done before in the literature (this includes the modified Barabasi-Albert, the Duplication-Divergence, the Solé and iSite models). Exploring the properties of these modified algorithms, including their usefulness as models of networks in molecular biology, will be the subject of future investigation.

Lastly, these algorithms grow networks using a probabilistic set of rules to implement an elementary move. Each realized network N_n of order n is obtained with some probability $p(N_n)$, so that the function $p(N_n)$ is a probability distribution over networks of order n. Determining $p(N_n)$ for any of the algorithms presented here seems difficult, and general properties of $p(N_n)$ remain unknown (other than averages of network properties over $p(N_n)$ are scale-free if the algorithm grows scale-free networks).

7.2.2 Random walk and laplacian analysis of microRNA-protein interaction networks

Future work regarding MGINs includes being able to predict functions of miRNAs using recursive algorithms. Both the random walk and solving the lattice laplacian algorithms are diffusive processes on a lattice. The random walk algorithm diffuses and looks at probability of ending somewhere, which is basically the same idea involved in solving the lattice laplacian on a network. *Are there different ways of doing it?*

Some of the genes which seem important in the networks generated using the above algorithms include HIST1H2BD, HIST1H2BK, IL6, H2AFJ, ALDOC, NPY, GALR2, RGS4, CXCL1, PRGER3, GRM3. Out of the above genes, the only one which is already a target of hsa-miR-218-5p in miRtaRBase [39] is HIST1H2BK. *Are the other genes indeed targets of hsa-miR-218-5p?*

Bibliography

[1] Albert, R and Barabási, A-L. *Statistical mechanics of complex networks*. Reviews of Modern Physics. 74(1):47–97, 2002.

[2] Albert, Réka and Jeong, Hawoong and Barabási, Albert-László. *Error and attack tolerance of complex networks*. Nature Publishing Group. 406(6794):378–382, 2000.

[3] Barabási, A-L, *Scale-free networks: a decade and beyond*. American Association for the Advancement of Science. Science. 325(5939):412–413, 2009.

[4] Barabási, A-L and Albert, R. *Emergence of scaling in random networks*. Science. American Association for the Advancement of Science 286(5439): 509–512, 1999.

[5] Barabási, A-L and Albert, R and Jeong, H. *Mean-field theory for scale-free random networks*. Physica A: Statistical Mechanics and its Applications. 272(1): 173–187, 1999.

[6] *Scale-free characteristics of random networks: the topology of the world-wide web*, Barabási, A-L and Albert, R and Jeong, H. *Scale-free characteristics of random networks: the topology of the world-wide web*. Physica A: Statistical Mechanics and its Applications. 281(1):69–77, 2000.

[7] Barrat, A and Barthelemy, M and Pastor-Satorras, R and Vespignani, A. *The architecture of complex weighted networks*. Proceedings of the National Academy of Sciences of the United States of America. 101(11): 3747–3752, 2004. year=2004.

[8] Berg, J and Lässig, M and Wagner, A. *Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications*. BMC evolutionary biology. 4(1): 1–12, 2004.

[9] Boccaletti, S and Latora, V and Moreno, Y and Chavez, M and Hwang, D-U. *Complex networks: Structure and dynamics*. Physics reports. 424(4):175–308, 2006.

[10] Cormen, TH. *Introduction to algorithms*. 2009

[11] Cytoscape 3.4.0. Cytoscape Developers, NRNB, 2016.

[12] Eisenberg, Eli and Levanon, Erez Y. *Preferential attachment in the protein network evolution*. Physical review letters. 91(13):138701, 2003.

[13] Estrada, E. *Virtual identification of essential proteins within the protein interaction network of yeast*. Proteomics. 6(1):35–40, 2006.

[14] Evlampiev, K and Isambert, H. *Conservation and topology of protein interaction networks under duplication-divergence evolution*. Proceedings of the National Academy of Sciences. 105(29):9863–9868, 2008.

[15] Gibson, T A and Goldberg, D S. *Improving evolutionary models of protein interaction networks*. Bioinformatics. 27(3):376–382, 2011.

[16] Gibson, T A and Goldberg, D S. *The Topological Profile of a Model of Protein Network Evolution Can Direct Model Improvement*. International Workshop on Algorithms in Bioinformatics: 40–52, 2015.

[17] Guimera, R and Amaral, L A N. *Functional cartography of complex metabolic networks*. Nature. 433(7028): 895–900, 2005.

[18] He, Xionglei and Zhang, Jianzhi. *Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution*. Genetics. 169(2):1157–1164, 2005.

[19] Jeong, H and Tombor, B and Albert, R and Oltvai, Z N and Barabási, A-L. *The large-scale organization of metabolic networks*. Nature. 407(6804): 651–654, 2000.

[20] Barabasi, Albert-Laszlo and Oltvai, Zoltan N. *Network biology: understanding the cell's functional organization*. Nature reviews genetics. 5(2):101–113, 2004.

[21] Kellis, M and Birren, B W and Lander, E S. *Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae*. Nature. 428(6983): 617–624, 2004.

[22] Khanin, Raya and Wit, Ernst. *How scale-free are biological networks*. Journal of computational biology. 13(3):810–818, 2006.

[23] Knuth, DE. *The art of computer programming: Sorting and searching*. 3, 1973.

[24] Näsvall, J and Sun, L and Roth, J R and Andersson, D I. *Real-time evolution of new genes by innovation, amplification, and divergence*. Science. 338(6105):384–387, 2012.

[25] Pastor-Satorras, R and Smith, E and Solé, R V. *Evolving protein interaction networks through gene duplication*. Journal of Theoretical biology. 222(2):199–210, 2003.

[26] Solé, R V and Pastor-Satorras, R and Smith, E and Kepler, T B. *A model of large-scale proteome evolution*. Advances in Complex Systems. 5(01):43–54, 2002.

[27] Taylor, J S and Raes, J. *Duplication and divergence: the evolution of new genes and old ideas*. Annu. Rev. Genet., 38: 615–643, 2004.

[28] Tautz, D and Domazet-Lošo, T. *The evolutionary origin of orphan genes*. Nature Reviews Genetics. 12(10): 692–702, 2011.

[29] Vázquez, A and Flammini, A and Maritan, A and Vespignani, A. :Modeling of protein interaction networks, Complexus. 1(1):38–44, 2002.

[30] Wagner, Andreas. *How the global structure of protein interaction networks evolves*. Proceedings of the Royal Society of London B: Biological Sciences. 270(1514):457–466, 2003.

[31] Wright, HB. *Tables of integrals and other mathematical data*. 1961

[32] Wagner, Andreas. *How the global structure of protein interaction networks evolves*. Proceedings of the Royal Society of London B: Biological Sciences. 270(1514):457–466, 2003.

[33] Tie, Jun and Pan, Yanglin and Zhao, Lina and Wu, Kaichun and Liu, Jie and Sun, Shiren and Guo, Xuegang and Wang, Biaoluo and Gang, Yi and Zhang, Yongguo and others. *MiR-218 inhibits invasion and metastasis of gastric cancer by targeting the Robo1 receptor*. PLoS Genet 6(3):e1000879, 2010.

[34] Zhu, Kegan and Ding, Hanying and Wang, Wengong and Liao, Zhicong and Fu, Zheng and Hong, Yeting and Zhou, Yong and Zhang, Chen-Yu and Chen, Xi. *Tumor-suppressive miR-218-5p inhibits cancer cell proliferation and migration via EGFR in non-small cell lung cancer*, author=Zhu, Kegan and Ding, Hanying and Wang, Wengong and Liao, Zhicong and Fu, Zheng and Hong, Yeting and Zhou, Yong and Zhang, Chen-Yu and Chen, Xi. Oncotarget, 7(19):28075, 2016.

[35] Uesugi, Atsushi and Kozaki, Ken-ichi and Tsuruta, Tomohiko and Furuta, Mayuko and Morita, Kei-ichi and Imoto, Issei and Omura, Ken and Inazawa, Johji. *The tumor suppressive microRNA miR-218 targets the mTOR component Rictor and inhibits AKT phosphorylation in oral cancer*. Cancer research, canres–0368, 2001.

[36] Lei, Chengwei and Ruan, Jianhua. *A random walk based approach for improving protein-protein interaction network and protein complex prediction*. Bioinformatics and Biomedicine (BIBM). 2012 IEEE International Conference on. 1–6, 2012.

[37] Peng C. *The Peng Lab*. Department of Biology. Faculty of Science. York University. 2017

[38] Szklarczyk, Damian and Franceschini, Andrea and Kuhn, Michael and Simonovic, Milan and Roth, Alexander and Minguez, Pablo and Doerks, Tobias and Stark, Manuel and Muller, Jean and Bork, Peer and others. *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored*. Nucleic acids research. 39(1):D561–D568, 2011.

[39] Hsu, Sheng-Da and Lin, Feng-Mao and Wu, Wei-Yun and Liang, Chao and Huang,

Wei-Chih and Chan, Wen-Ling and Tsai, Wen-Ting and Chen, Goun-Zhou and Lee, Chia-Jung and Chiu, Chih-Min and others. *miRTarBase: a database curates experimentally validated microRNA–target interactions*. Nucleic acids research. gkq1107, 2010.

[40] Taipaleenmäki, Hanna and Farina, Nicholas H and van Wijnen, Andre J and Stein, Janet L and Hesse, Eric and Stein, Gary S and Lian, Jane B. *Antagonizing miR-218-5p attenuates Wnt signaling and reduces metastatic bone disease of triple negative breast cancer cells*. Oncotarget 7(48):79032–79046, 2016.

[41] Xin, Yu and Li, Zheng and Shen, Jianxiong and Chan, Matthew TV and Wu, William Ka Kei, *CCAT1: a pivotal oncogenic long non-coding RNA in human cancers*. Cell proliferation. 49(3): 255–260, 2016.

[42] Sánchez, Catherine A and Andahur, Eliana I and Valenzuela, Rodrigo and Castellon Vera, Enrique and Fulla Ortiz, Juan and Ramos, Christian G and Trivino, Juan C. *Exosomes from bulk and stem cells from human prostate cancer have a differential microRNA content that contributes cooperatively over local and pre-metastatic niche*, 2015.

[43] Collins, Anne T and Maitland, Norman J. *Prostate cancer stem cells*. European Journal of Cancer, 42(9):1213–1218, 2006.

[44] Douglas, Dorothea and Hsu, Jessie Hao-Ru and Hung, Long and Cooper, Aaron and Abdueva, Diana and van Doorninck, John and Peng, Grace and Shimada, Hiro and Triche, Timothy J and Lawlor, Elizabeth R. *BMI-1 promotes ewing sarcoma tumorigenicity independent of CDKN2A repression*. Cancer research 68(16):6507–6515, 2008.

[45] Artavanis-Tsakonas, Spyros and Rand, Matthew D and Lake, Robert J. . Science, 284(5415): 770–776, 1999.

[46] Uesugi, Atsushi and Kozaki, Ken-ichi and Tsuruta, Tomohiko and Furuta, Mayuko and Morita, Kei-ichi and Imoto, Issei and Omura, Ken and Inazawa, Johji. *The tumor suppressive microRNA miR-218 targets the mTOR component*.Cancer research. canres–0368, 2011.

[47] Hohenester, Erhard.*Structural insight into Slit–Robo signaling*, 2008.

[48] Liu, Na and Chen, Nian-Yong and Cui, Rui-Xue and Li, Wen-Fei and Li, Yan and Wei, Rong-Rong and Zhang, Mei-Yin and Sun, Ying and Huang, Bi-Jun and Chen, Mo and others. *Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis*. The lancet oncology. 13(6): 633–641, 2012.

[49] . Haenisch, Sierk and Zhao, Yi and Chhibber, Aparna and Kaiboriboon, Kitti and Do, Lynn V and Vogelgesang, Silke and Barbaro, Nicholas M and Alldredge, Brian K and Lowenstein, Daniel H and Cascorbi, Ingolf and others. *SOX11 identified by target gene evaluation of miRNAs differentially expressed in focal and non-focal brain tissue of therapy-resistant epilepsy patients*. Neurobiology of disease. 77:127–140, 2015.

[50] Bartel, David P. *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 116(2):281–297, 2004.

[51] Cooper, Geoffrey M and Hausman, Robert E. The cell, 2000

[52] Ambros, Victor. *The functions of animal microRNAs*. Nature. 431(7006), 350–355, 2004.

[53] Lee, Yoontae and Kim, Minju and Han, Jinju and Yeom, Kyu-Hyun and Lee, Sanghyuk and Baek, Sung Hee and Kim, V Narry. *MicroRNA genes are transcribed by RNA polymerase II*. The EMBO journal 23(20): 4051–4060, 2004.

[54] Bartel, David P, *MicroRNAs: target recognition and regulatory functions*. Cell. 136(2): 215–233, 2009.

[55] Li, Chunsheng and Feng, Yi and Coukos, George and Zhang, Lin. *Therapeutic microRNA strategies in human cancer*. The AAPS journal. 11(4):747, 2009.

[56] Lee, Yoontae and Ahn, Chiyoung and Han, Jinju and Choi, Hyounjeong and Kim, Jaekwang and Yim, Jeongbin and Lee, Junho and Provost, Patrick and Rådmark, Olof and Kim, Sunyoung and others, *The nuclear RNase III Drosha initiates microRNA processing*. Nature. 425(6956): 415–419, 2003.

[57] Murchison, Elizabeth P and Hannon, Gregory J. *miRNAs on the move: miRNA biogenesis and the RNAi machinery*.Current opinion in cell biology. 16(3):223–229, 2004.

[58] Lund, E and Dahlberg, JE *Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs*, Cold Spring Harbor symposia on quantitative biology. 71: 59–66, 2006.

[59] Ji, Xinhua. *The mechanism of RNase III action: how dicer dices*, RNA interference, 99–116, 2008.

[60] Cai, Xuezhong and Hagedorn, Curt H and Cullen, Bryan R. *Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs*. Rna. volume=10(12): 1957–1966, 2004.

[61] Rana, Tariq M. *Illuminating the silence: understanding the structure and function of small RNAs*. Nature reviews Molecular cell biology. 8(1): 23–36, 2007.

[62] Krol, Jacek and Sobczak, Krzysztof and Wilczynska, Urszula and Drath, Maria and Jasinska, Anna and Kaczynska, Danuta and Krzyzosiak, Wlodzimierz J. *Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design*. Journal of Biological Chemistry. 279(40):42230–42239,2004.

[63] Pratt, Ashley J and MacRae, Ian J. *The RNA-induced silencing complex: a versatile gene-silencing machine*. Journal of Biological Chemistry. 284(27):17897–17901, 2009.

[64] MacRae, Ian J and Ma, Enbo and Zhou, Min and Robinson, Carol V and Doudna, Jennifer A. *In vitro reconstitution of the human RISC-loading complex*. Proceedings of the National Academy of Sciences. volume=105(2):512–517, 2008.

[65] Meister, Gunter and Landthaler, Markus and Peters, Lasse and Chen, Po Yu and Urlaub, Henning and Lührmann, Reinhard and Tuschl, Thomas. *Identification of novel argonaute-associated proteins*. Current biology, 15(23): 2149–2155, 2005.

[66] Kawasaki, Hiroaki and Taira, Kazunari. *MicroRNA-196 inhibits HOXB8 expression in myeloid differentiation of HL60 cells*. Nucleic acids symposium series. volume=48(1): 211–212, 2004.

[67] Fabian, Marc Robert and Sonenberg, Nahum and Filipowicz, Witold. *Regulation of*

mRNA translation and stability by microRNAs. Annual review of biochemistry, 79:351–379, 2010.

[68] Lewis, Benjamin P and Burge, Christopher B and Bartel, David P. *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets*. cell, 120(1), 2005.

[69] Jones-Rhoades, Matthew W and Bartel, David P and Bartel, Bonnie. *MicroRNAs and their regulatory roles in plants*. Annu. Rev. Plant Biol.. 57:19–53, 2006.

[70] Stark, Alexander and Brennecke, Julius and Bushati, Natascha and Russell, Robert B and Cohen, Stephen M.*Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3? UTR evolution*. Cell. 123(6): 1133–1146, 2005.

[71] Musilova, Katerina and Mraz, Marek. *MicroRNAs in B-cell lymphomas: how a complex biology gets more complex*.Leukemia. 29(5): 1004–1017, 2015.

[72] Chen, Jian-Fu and Murchison, Elizabeth P and Tang, Ruhang and Callis, Thomas E and Tatsuguchi, Mariko and Deng, Zhongliang and Rojas, Mauricio and Hammond, Scott M and Schneider, Michael D and Selzman, Craig H and others. *Targeted deletion of Dicer in the heart leads to dilated cardiomyopathy and heart failure*. Proceedings of the National Academy of Sciences. 105(6): 2111–2116, 2008.

[73] Zhao, Yong and Ransom, Joshua F and Li, Ankang and Vedantham, Vasanth and von Drehle, Morgan and Muth, Alecia N and Tsuchihashi, Takatoshi and McManus, Michael T and Schwartz, Robert J and Srivastava, Deepak. *Dysregulation of cardiogenesis,*

cardiac conduction, and cell cycle in mice lacking miRNA-1-2. Cell. volume=129(2): 303–317, 2007.

[74] Phua, Yu Leng and Chu, Jessica YS and Marrone, April K and Bodnar, Andrew J and Sims-Lucas, Sunder and Ho, Jacqueline. *Renal stromal miRNAs are required for normal nephrogenesis and glomerular mesangial survival*. Physiological reports.3(10): pages=e12537, 2015.

[75] Maes, Olivier C and Chertkow, Howard M and Wang, Eugenia and Schipper, Hyman M. *MicroRNA: implications for Alzheimer disease and other human CNS disorders*. Current genomics, volume=10(3):154–168, 2009.

[76] Romao, Josue M and Jin, Weiwu and Dodson, Michael V and Hausman, Gary J and Moore, Stephen S and others. *MicroRNA regulation in mammalian adipogenesis*. Experimental Biology and Medicine. 236(9): 997–1004, 2011.

[77] Pillai, Ramesh S. *MicroRNA function: multiple mechanisms for a tiny RNA?*. Rna, volume=11(12): 1753–1761, 2005.

[78] Osaki, M and Oshimura, M and and Ito, H. *PI3K-Akt pathway: its functions and alterations in human cancer*. Apoptosis. 9(6): 667–676, 2004.

[79] *http://www.genecards.org* Crown Human Genome Center. Department of Molecular Genetics. the Weizmann Institute of Science.

[80] Maere, Steven and Heymans, Karel and Kuiper, Martin. *BiNGO: a Cytoscape plugin to*

assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 21(16): 3448–3449, 2005.

[81] Gene Ontology Consortium and others. *The gene ontology project in 2008*. Gene Ontology Consortium and others. Nucleic acids research. volume=36(suppl 1): D440–D444, 2008.

[82] Vazquez, Alexei and Alzate, O. *Protein interaction networks*. Neuroproteomics. Boca Raton, FL, 2010.

[83] Tolba, Ahmed. *Scale free networks; a literature review*, Proceedings of the International Conference on Complex Systems, 2007.

[84] Dezső, Zoltán and Barabási, Albert-László. Halting viruses in scale-free networks. Physical Review E volume=65(5):055103, 2002.

[85] Fricker, M and Boddy, L and Nakagaki, T and Bebber, DP and Gross, T and Sayama, H. *Adaptive networks: Theory, models and applications*, 2009.

[86] Love, Michael I and Huber, Wolfgang and Anders, Simon. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. Genome biology. 15(12):550, 2014.

Appendix

Appendix

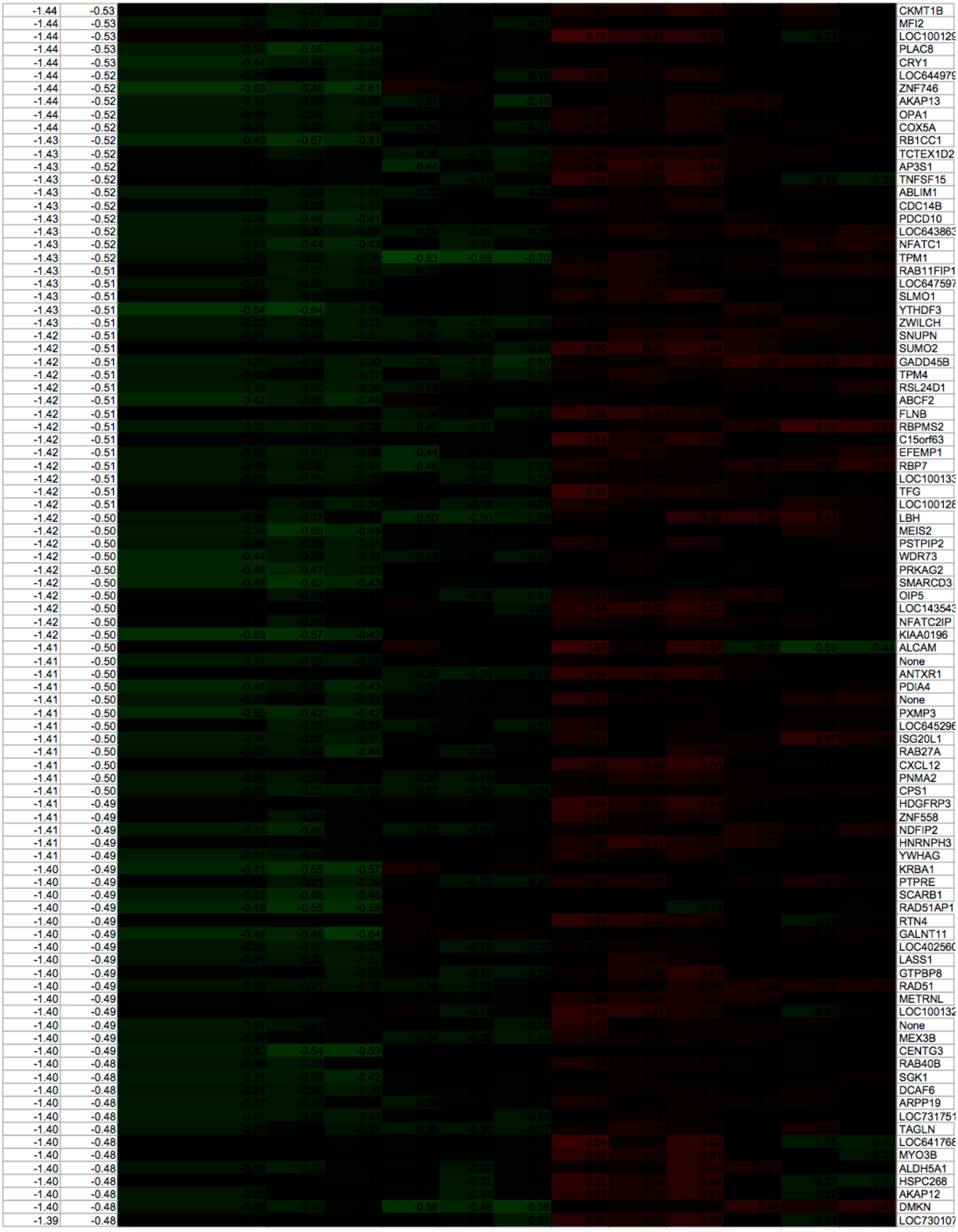
Technology: Illumina.SingleColor.HumanHT-12_V4_0_R2_15002873_B
Project: Peng
Experiment: 140218
Interpretation: Group - sampleID
Entity List: 4106-Oneway ANOVA p (Corr) cut-off = 0.05
Cluster: Hierarchical Combined Tree on Group - sampleID
Exported on: 27/02/14 3:05 PM

		Medium overexpressing			High overexpressing			No overexpression			Low overexpressing			
Fold chang	Log fold cha	[tx2, Mir218 14-3]	[tx2, Mir21	[tx2, Mir21	[tx3, Mir218	[tx3, Mir21	[tx3, Mir218	[control, GF	[control, GF	[control, GF	[tx1, Mir218	[tx1, Mir218	[tx1, Mir218	Symbol
-1.21	-0.28													ZSCAN5A
-5.45	-2.45	-1.84	-2.01	-1.65	-0.40	-0.26	-0.39	0.56	0.55	0.74	0.31			LXN
-4.65	-2.22	-1.26	-1.36	-1.29	-0.79	-0.67	-0.73	1.06	0.77	0.91	0.82	0.93	0.67	NPY
-2.59	-1.37	-0.97	-1.12	-1.10	-0.28	-0.27			0.34	0.30		0.43		QPRT
-2.42	-1.28	-0.41	-0.34	-0.55	-0.37	-0.60	-0.48	0.98	0.85	0.70	0.55	0.41		LOC65311C
-2.37	-1.25	-0.53	-0.53	-0.61	-0.53	-0.49	-0.54	0.48	0.84	0.73	0.68	0.58	0.65	IAH1
-2.37	-1.25	-0.57	-0.52	-0.51	-0.44	-0.33	-0.47	0.82	0.70	0.62	0.56	0.33	0.53	DSG2
-2.36	-1.24	-0.49	-0.43	-0.47	-0.35	-0.27	-0.31	0.84	0.66	0.81	0.61			SS18
-2.31	-1.21	-0.87	-1.11	-0.90	-0.18				0.46	0.31	0.82	0.60	0.74	HERPUD1
-2.27	-1.18	-0.38	-0.40	-0.37	-0.28	-0.37	-0.36	0.70	0.88	0.82	0.88	0.59	0.44	IER3IP1
-2.27	-1.18	-0.29	-0.37	-0.27		-0.33	-0.47	0.89	0.96	0.83	0.48			LOC730255
-2.24	-1.17	-0.66	-0.63	-0.79	-0.28	-0.26	-0.18	0.48	0.89	0.94	0.74			ATP11B
-2.20	-1.14	-0.81	-0.83	-0.73	-0.32	-0.34	-0.33	0.38	0.32	0.35	0.57	0.65	0.65	THBS1
-2.19	-1.13	-0.52	-0.62	-0.52	-0.52	-0.57	-0.43	0.75	0.45	0.52	0.43	0.53	0.48	CDH2
-2.17	-1.12	-0.55	-0.59	-0.56	-0.58	-0.33	-0.45	0.53	0.51	0.62	0.35	0.35	0.33	RIOK3
-2.17	-1.12	-0.63	-0.94	-0.82	-0.18	-0.36	-0.31	0.46	0.32	0.39	0.32	0.36	0.41	MKX
-2.15	-1.10	-0.46	-0.65	-0.63	-0.59	-0.52	-0.62	0.48	0.64	0.48	0.52	0.70	0.73	ANXA8L2
-2.15	-1.10	-0.18	-0.28	-0.38		-0.32	-0.36	1.02	0.75	0.74	0.61			LOC729340
-2.12	-1.09	-0.36	-0.47	-0.45				0.64	0.72	0.72				CDO1
-2.11	-1.08	-0.79	-0.73	-0.66	-0.22	-0.38	-0.43	0.42	0.50	0.41	0.62	0.41	0.42	CCDC85A
-2.10	-1.07		-0.23		-0.25	-0.20	-0.35	0.90	0.80	1.03	0.72	0.31		LOC727984
-2.06	-1.04	-0.76	-0.67	-0.69	-0.61				0.67	0.38	0.68	0.59	0.72	HERPUD1
-2.06	-1.04	-0.61	-0.66	-0.48	-0.20	-0.39	-0.19	0.54	0.29	0.54	0.33	0.32		FLJ22184
-2.05	-1.04	-0.40	-0.55	-0.63	-0.51	-0.61	-0.60	0.59	0.48	0.47	0.43	0.43	0.48	KRT80
-2.04	-1.03	-0.52	-0.72	-0.39	-0.49	-0.66	-0.58	0.80	0.57	0.36	0.82	0.85	0.89	SMAD6
-2.03	-1.02	-0.48	-0.60	-0.52	-0.27	-0.46	-0.48	0.58	0.39	0.50				RBBP8
-2.02	-1.02	-0.69	-0.83	-0.78		-0.23	-0.26	0.31	0.36					ATP2A2
-2.02	-1.02	-0.67	-0.87	-0.58	-0.41	-0.48	-0.25	0.29	0.33		0.85	0.83	0.90	OLR1
-2.02	-1.02	-0.35	-0.49	-0.64	-0.25	-0.27	-0.32	0.62	0.36	0.69	0.38	0.44	0.44	FVT1
-2.01	-1.01	-0.63	-0.28	-0.55				0.85	0.46	0.36				CD55
-2.00	-1.00	-0.50	-0.50	-0.51	-0.53	-0.43	-0.45	0.56	0.43	0.59	0.85	0.99	1.01	PCDH7
-1.99	-0.99	-0.25	-0.40	-0.28		-0.24		0.81	0.64	0.59	0.37			PHLDB2
-1.98	-0.98	-0.31	-0.34	-0.21	-0.20		-0.23	0.73	0.60	0.77	0.31	0.44		LYL1
-1.96	-0.97	-0.44	-0.28	-0.45	-0.65	-0.45	-0.64	0.66	0.50	0.68	0.41		0.43	DMKN
-1.96	-0.97	-0.60	-0.76	-0.62	-0.19	-0.30	-0.18	0.36	0.29					MATN2
-1.95	-0.96	-0.39	-0.49	-0.36	-0.31	-0.34	-0.29	0.51	0.63	0.50	0.53			LIMA1
-1.94	-0.96	-0.29	-0.42	-0.58	-0.23	-0.46	-0.50	0.55	0.55	0.56	0.42			LOC729301
-1.92	-0.94	-0.70	-0.79	-0.87						0.36				CUL1
-1.91	-0.94	-0.50	-0.59	-0.57	-0.53	-0.35	-0.30	0.40	0.45	0.31	0.51	0.36		ME2
-1.91	-0.93	-0.38	-0.46	-0.64	-0.80	-0.65	-0.79	0.48	0.49	0.38	0.68	0.88	0.84	EDN1
-1.89	-0.92	-0.42	-0.85	-0.72	-0.33	-0.35	-0.20	0.30			0.33	0.48		HSPA5
-1.89	-0.92	-0.80	-0.98	-0.89	-0.26									RHEB
-1.87	-0.91	-0.51	-0.44	-0.37	-0.21	-0.27	-0.29	0.38	0.55	0.46	0.33			C18orf21
-1.87	-0.90	-0.32	-0.42	-0.34	-0.60	-0.50	-0.46	0.66	0.32	0.64	0.49	0.32		HTRA1
-1.86	-0.90				-0.21		-0.33	0.70	0.78	0.91	0.73			ANXA2P1
-1.86	-0.90	-0.31	-0.28		-0.18		-0.43	0.63	0.67	0.72	0.43			ANXA2P1
-1.84	-0.88	-0.55	-0.49	-0.47	-0.33	-0.38	-0.40	0.39	0.38	0.41	0.33	0.32		CXADR
-1.84	-0.88	-0.51	-0.61	-0.66						0.29				MCM4
-1.83	-0.88	-0.28	-0.36	-0.27	-0.53	-0.43	-0.25	0.51	0.73	0.52				KIAA0101
-1.83	-0.87	-0.53	-0.53	-0.36	-0.50	-0.33	-0.27	0.35	0.37	0.61	0.45			C4orf49
-1.82	-0.87	-0.45	-0.68	-0.50			-0.44	0.47	0.38	0.31				AIF1L
-1.82	-0.86	-0.38	-0.67	-0.43	-0.52	-0.43	-0.31	0.38	0.35	0.33	0.31	0.35		ATP5A1
-1.82	-0.86	-0.32	-0.53	-0.34	-0.35	-0.34	-0.25	0.44	0.41	0.55	0.34	0.39		OSAP
-1.80	-0.85	-0.32	-0.43	-0.42	-0.39	-0.31	-0.43	0.52	0.42	0.43	0.32	0.31		None
-1.79	-0.84	-0.54	-0.41	-0.59		-0.27	-0.22	0.43		0.41				MAP2K1
-1.79	-0.84			-0.38			-0.31	0.69	0.53	0.77	0.31			PTPLA
-1.79	-0.84	-0.41	-0.50	-0.44	-0.31	-0.41	-0.57	0.36	0.39	0.48	0.31	0.48	0.64	ACAA2
-1.78	-0.83	-0.30	-0.30	-0.50	-0.24	-0.26	-0.36	0.51	0.36	0.58	0.33			KIAA1012
-1.78	-0.83	-0.44	-0.55	-0.57	-0.18			0.58		0.53				PFN2
-1.78	-0.83	-0.79	-0.96	-0.83										FAM62B
-1.78	-0.83	-0.46	-0.59	-0.75										LYN
-1.78	-0.83	-0.28		-0.28		-0.22	-0.52	0.66	0.45	0.66	0.50			LOC646688
-1.77	-0.83	-0.44	-0.47	-0.44	-0.18	-0.31	-0.43	0.39	0.49					DIS3L
-1.77	-0.83	-0.58	-0.61	-0.55	-0.24	-0.21	-0.32	0.42			0.31	0.52		NARS
-1.77	-0.83	-0.33	-0.28	-0.28	-0.40	-0.29	-0.27	0.50	0.51	0.57	0.35	0.32		EPDR1
-1.77	-0.82	-0.92	-0.91	-0.83						-0.19		0.31		REPIN1
-1.77	-0.82	-0.52	-0.53	-0.60		-0.19	-0.23	0.29						MYBL1
-1.77	-0.82	-0.36	-0.34	-0.36			-0.18	0.31	0.46	0.65	0.48			MYC
-1.77	-0.82	-0.33	-0.50	-0.50			-0.30	0.58		0.39				PTPLB
-1.77	-0.82	-0.18	-0.21	-0.32	-0.22	-0.26	-0.37	0.59	0.61	0.55	0.37			RHOBTB3
-1.76	-0.82	-0.29	-0.44	-0.43	-0.40	-0.55	-0.26	0.34	0.51	0.43	0.46	0.50		DMKN
-1.76	-0.82	-0.37	-0.50	-0.26	-0.21	-0.32	-0.20	0.42	0.44	0.61	0.50			LOC100129
-1.76	-0.82			-0.26				0.56	0.72	0.72				EEF1A2
-1.76	-0.81	-0.23	-0.22	-0.26	-0.22			0.58	0.50	0.63				TXNL1
-1.75	-0.81	-0.34	-0.37	-0.21		-0.35	-0.30	0.48	0.49	0.62				PRKCI
-1.75	-0.81	-0.36	-0.19				-0.22	1.11	0.48	0.83		-0.42	-0.30	ZNF91
-1.75	-0.80	-0.34	-0.36	-0.49	-0.36	-0.19	-0.36	0.43	0.32	0.48		0.49		C18orf55


IGFBP3
RAC2
LIFR
OSBPL1A
ATP5A1
TATDN1
MGC18216
SLC38A1
COL3A1
APCDD1L
FCAR
ANKRD28
NR2F2
LRRN2
FVT1
CEBPG
CEBPD
LOC643308
FAM82A2
None
HNRNPU
CDO1
C18orf25
IQGAP1
TJP1
TNS3
TRA1P2
OTUD1
PDIA6
GRAMD3
LOC645233
MBOAT2
SEMA3E
ROD1
ME2
TMUB1
RFX7
C18orf55
DNAJB6
C18orf8
NFIB
LOC100132
LOC728843
FECH
LOC648526
TERF1
SEC11C
CNBP
KIDINS220
ATP1B3
ATP1B3
TUFT1
LMO4
C1orf198
CHSY1
CHP
ERRFI1
PHLDB2
EHD4
TDG
RNF138
FAM96A
STXBP6
DYM
C18orf10
LOC441484
PDIA3P
TSPAN3
None
MTFR1
RG9MTD1
NPTN
MTHFS
STC2
TRMT12
TSPYL5
HMGB1
ATP9A
C4orf49
NPC1
TRAM2
LYSMD2
UBE3C
FAM63A
ZNF532
DDX17
ADNP2
KRT80
CNDP2
DEM1
DDIT4
SHROOM4
PQLC1
SNAPC1

-1.57	-0.65								0.53	0.41	0.51				LMO4
-1.57	-0.65	-0.43	-0.50	-0.41	-0.28	-0.21						0.64	0.86	0.96	MBP
-1.57	-0.65	-0.35	-0.45	-0.24	-0.22	-0.26				0.42	0.39	0.56	0.62	0.5	SELS
-1.57	-0.65	-0.23	-0.34	-0.31	-0.24		-0.29		0.45	0.29	0.33	0.29			LOC100132
-1.57	-0.65	-0.51	-0.40	-0.38		-0.40	-0.19				0.39		0.40		NMB
-1.57	-0.65	-0.21		-0.22					0.41	0.45	0.60				SAP30
-1.57	-0.65	-0.37		-0.37	-0.18	-0.28	-0.51		0.45		0.63		0.30		ATF5
-1.57	-0.65	-0.25	-0.33		-0.26	-0.23	-0.28		0.42	0.37	0.41		0.29		RPL17
-1.57	-0.65	-0.61	-0.43	-0.54											WNK1
-1.57	-0.65						-0.36		0.74	0.48	0.63		-0.23	-0.2	LOC345041
-1.56	-0.65	-0.28	-0.33	-0.39	-0.21	-0.25	-0.22		0.43	0.29			0.52	0.3	LINGO1
-1.56	-0.64			-0.19		-0.25			0.73	0.48	0.49		-0.26	-0.3	CHRNA5
-1.56	-0.64		-0.31	-0.47		-0.17	-0.24		0.44		0.39		0.35		LOC653344
-1.56	-0.64	-0.29	-0.41	-0.41					0.42						USP13
-1.56	-0.64	-0.39	-0.49	-0.38	-0.23					0.33		0.38	0.44	0.4	RIOK3
-1.56	-0.64					-0.19	-0.18		0.46	0.47	0.69	0.45			OIP5
-1.56	-0.64	-0.34	-0.48	-0.48	-0.18	-0.42	-0.42					0.43	0.37	0.3	SMAD4
-1.56	-0.64	-0.70	-0.63	-0.59											SQLE
-1.56	-0.64	-0.34	-0.44			-0.27	-0.23		0.52						ROD1
-1.56	-0.64	-0.37	-0.33	-0.29							0.41			0.4	OSBPL1A
-1.56	-0.64						-0.26		0.52	0.56	0.74			-0.2	GGCT
-1.56	-0.64								0.50	0.56	0.68				CXCL12
-1.55	-0.64	-0.28	-0.22	-0.28	-0.24	-0.26			0.47	0.32	0.35				SS18
-1.55	-0.64	-0.38	-0.36	-0.27	-0.28	-0.23	-0.30		0.36	0.31			0.55	0.3	PIAS2
-1.55	-0.63	-0.26	-0.33	-0.21	-0.38	-0.44	-0.36		0.45		0.45		0.48	0.3	TXNRD2
-1.55	-0.63	-0.27	-0.48	-0.49			-0.36								XBP1
-1.55	-0.63								0.75	0.61	0.67		-0.20	-0.2	C3orf34
-1.55	-0.63					-0.22			0.77	0.64	0.56		-0.29	-0.2	LOC100128
-1.54	-0.63	-0.56	-0.58	-0.55										-0.1	ATP6V0E2
-1.54	-0.62	-0.43	-0.41	-0.34		-0.34	-0.26		0.42				0.55	0.3	HDHD2
-1.54	-0.62	-0.35	-0.42	-0.26	-0.18	-0.20	-0.35		0.38		0.30	0.29	0.30	0.3	LRRC1
-1.54	-0.62	-0.33		-0.17	-0.53	-0.41	-0.43			0.65	0.47	0.58	0.53	0.6	NNMT
-1.53	-0.62				-0.19				0.37	0.51	0.55				RAB3IP
-1.53	-0.62	-0.20	-0.37	-0.48		-0.36				0.30					PDIA3P
-1.53	-0.62	-0.20				-0.19	-0.24		0.59	0.42	0.51	0.32			TMED7
-1.53	-0.62								0.68	0.61	0.61		-0.25		HSPD1
-1.53	-0.62	-0.31	-0.27	-0.39	-0.41	-0.19	-0.40			0.39			0.40	0.3	PTPLAD1
-1.53	-0.62		-0.19		-0.23	-0.30			0.38	0.52	0.47				CKMT1A
-1.53	-0.61	-0.26	-0.35	-0.25	-0.42		-0.35			0.40	0.42	0.45	0.44	0.3	CCDC5
-1.53	-0.61	-0.46	-0.53	-0.54						0.30		0.72	0.49	0.6	DDIT3
-1.53	-0.61	-0.33	-0.58	-0.41	-0.40	-0.21									EIF3J
-1.53	-0.61					-0.23	-0.22		0.70	0.50	0.56			-0.3	LMOD3
-1.53	-0.61	-0.27	-0.27	-0.35		-0.21	-0.32		0.47				0.30		NPTN
-1.53	-0.61					-0.57	-0.40		0.81	0.52	0.59		-0.31	-0.3	LOC100130
-1.53	-0.61	-0.43	-0.45	-0.39											EEF1A1
-1.53	-0.61	-0.61	-0.53	-0.51											MRPS35
-1.53	-0.61	-0.23	-0.33		-0.30	-0.20			0.38	0.36	0.35				LOC649970
-1.53	-0.61	-0.56	-0.51	-0.39								0.42		0.3	FGF2
-1.52	-0.61	-0.42	-0.47	-0.26	-0.21	-0.31	-0.19							0.3	TMEM44
-1.52	-0.61	-0.49	-0.52	-0.38									-0.29		TMEM106B
-1.52	-0.61		-0.18	-0.21		-0.24	-0.29		0.46	0.43	0.43				SPATA5L1
-1.52	-0.61	-0.29	-0.31	-0.38			-0.20		0.57					0.3	NIPA1
-1.52	-0.60	-0.39	-0.39	-0.42	-0.41							0.34	0.35	0.3	ANXA2
-1.52	-0.60	-0.35	-0.28	-0.36					0.42						PLOD2
-1.52	-0.60	-0.28	-0.42	-0.44					0.31					-0.2	ITGB1BP1
-1.52	-0.60	-0.84	-0.54	-0.69											LOC387820
-1.52	-0.60	-0.30		-0.28	-0.62	-0.38	-0.66			0.49	0.52	0.58		0.3	TAGLN
-1.52	-0.60					-0.36	-0.42		0.71	0.55	0.62		-0.29	-0.3	MYL12A
-1.52	-0.60	-0.55	-0.73	-0.63											FASTK
-1.51	-0.60								0.75	0.52	0.47		-0.23		ZNF549
-1.51	-0.60	-0.52	-0.44	-0.31		-0.21	-0.28		0.38						APPBP2
-1.51	-0.60		-0.29						0.52	0.46	0.49			-0.2	LZTFL1
-1.51	-0.60	-0.46	-0.44	-0.40						0.31			0.33		TCEA1
-1.51	-0.60	-0.34	-0.39	-0.31	-0.32	-0.20	-0.35		0.36	0.28	0.29	0.34	0.32	0.3	HEG1
-1.51	-0.60	-0.58	-0.51	-0.57											ZNF212
-1.51	-0.60	-0.47	-0.33	-0.31		-0.20	-0.25			0.39					BNIP2
-1.51	-0.59	-0.24	-0.26						0.41	0.42	0.47		-0.25		KDSR
-1.51	-0.59	0.32				-0.26	-0.29		0.72	0.52	0.73		-0.32		GRIPAP1
-1.51	-0.59	-0.43	-0.60	-0.50		-0.35									NDUFB9
-1.51	-0.59	-0.60	-0.39	-0.30			-0.18						0.46	0.3	None
-1.51	-0.59	-0.32	-0.33	-0.26	-0.77	-0.80	-0.68		0.34	0.29	0.29	0.33			TPM1
-1.51	-0.59		-0.38	-0.36	-0.39	-0.44	-0.44		0.54			0.42	0.36		LOC730101
-1.51	-0.59	-0.19	-0.30	-0.24		-0.29	-0.36		0.33	0.36	0.35	0.34		0.3	FANCI
-1.51	-0.59	-0.54	-0.45	-0.36					0.28				0.40		ATP7B
-1.51	-0.59	-0.45	-0.63	-0.58		-0.18						0.38	0.54	0.5	SDF2L1
-1.50	-0.59	-0.29	-0.38	-0.38		-0.25	-0.18			0.31	0.34	0.33	0.49	0.4	GYG2
-1.50	-0.59		-0.27	-0.29					0.47		0.48				LMCD1
-1.50	-0.59					-0.33			0.90	0.39	0.63		-0.34	-0.2	XRCC2
-1.50	-0.59	-0.36	-0.32	-0.31					0.31						NCBP2
-1.50	-0.59	-0.28					-0.17		0.45	0.41	0.41				MEX3C
-1.50	-0.59	-0.24	-0.21				-0.21		0.50	0.29	0.39				SRP14
-1.50	-0.59	-0.27	-0.25	-0.32		-0.36	-0.20		0.41		0.39	0.34	0.37		PSMA4
-1.50	-0.59	-0.40	-0.40	-0.39					0.50						LOC729148
-1.50	-0.59	-0.18	-0.29	-0.26		-0.37			0.44	0.29	0.31				PDCD7
-1.50	-0.59	-0.42	-0.39	-0.45					0.39			0.45	0.46	0.4	MAPK6
-1.50	-0.58								0.68	0.55	0.49				DCUN1D5
-1.50	-0.58	-0.28	-0.28	-0.24	-0.50	-0.32	-0.31		0.40		0.33	0.50	0.34	0.3	ARID5B
-1.50	-0.58	-0.48	-0.43	-0.33	-0.37	-0.20						0.29	0.31		LOC729082
-1.50	-0.58	-0.18	-0.30	-0.49	-0.31	-0.41	-0.48		0.32			0.55	0.63	0.6	FHL1
-1.50	-0.58	-0.35	-0.46	-0.39			-0.21					0.35	0.57	0.6	AKAP13
-1.50	-0.58	-0.24	-0.37	-0.36											GGH
-1.50	-0.58								0.73		0.56				IGF2BP2
-1.50	-0.58	-0.36	-0.40	-0.54			-0.20								UBE2V2
-1.50	-0.58	-0.21	-0.35	-0.55											STK40

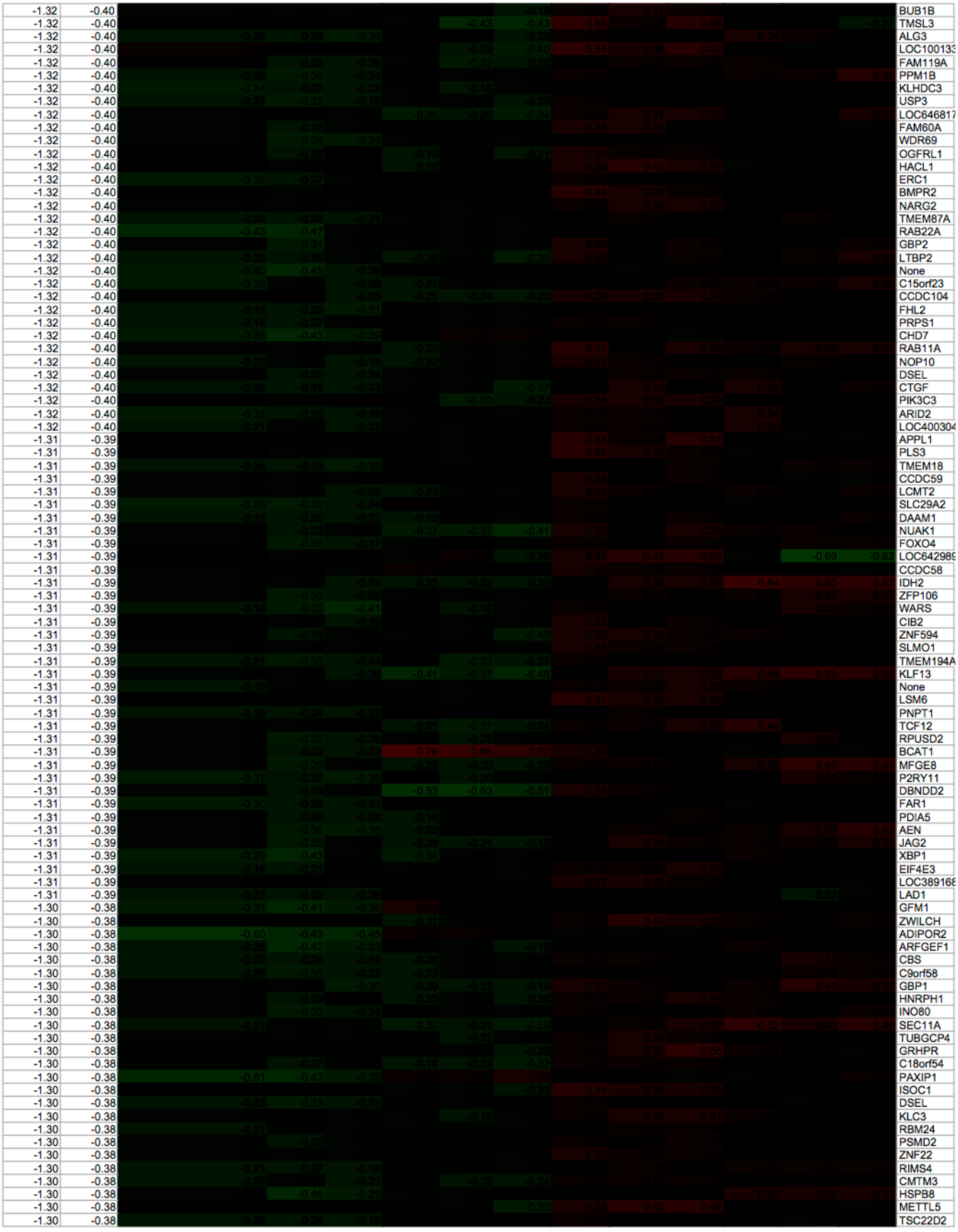
-1.50	-0.58	-0.39	-0.49	-0.48	-0.24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	MCCC1
-1.50	-0.58	-0.29	-0.23	-0.31	[illegible]	[illegible]	-0.25	0.29	0.31	0.31	[illegible]	[illegible]	[illegible]	IFFO2
-1.49	-0.58	-0.28	-0.30	-0.32	-0.24	[illegible]	-0.25	[illegible]	[illegible]	0.32	[illegible]	[illegible]	[illegible]	GYG2
-1.49	-0.58	[illegible]	-0.21	[illegible]	[illegible]	-0.26	[illegible]	0.71	0.34	0.48	[illegible]	-0.19	[illegible]	MAGT1
-1.49	-0.58	-0.34	-0.20	-0.33	[illegible]	[illegible]	[illegible]	0.33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ALDH1B1
-1.49	-0.58	-0.28	-0.38	-0.41	[illegible]	-0.27	[illegible]	0.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	NAV2
-1.49	-0.58	-0.29	-0.35	-0.29	-0.27	-0.29	-0.23	[illegible]	[illegible]	0.33	[illegible]	[illegible]	[illegible]	LOC64225
-1.49	-0.58	-0.27	-0.41	-0.48	-0.24	[illegible]	-0.24	[illegible]	[illegible]	0.31	[illegible]	[illegible]	[illegible]	RCN2
-1.49	-0.58	-0.21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.54	0.60	[illegible]	[illegible]	[illegible]	PPIB
-1.49	-0.58	-0.46	-0.31	-0.21	-0.34	-0.23	-0.19	[illegible]	[illegible]	[illegible]	0.51	0.52	0.54	RTTN
-1.49	-0.58	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.78	0.42	0.64	[illegible]	-0.18	[illegible]	C14orf153
-1.49	-0.58	-0.35	-0.56	-0.39	-0.22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.38	[illegible]	RPL17
-1.49	-0.57	-0.18	-0.40	-0.32	-0.23	[illegible]	-0.21	[illegible]	[illegible]	0.29	[illegible]	[illegible]	[illegible]	C18orf32
-1.49	-0.57	-0.25	-0.23	-0.24	-0.25	-0.25	-0.44	0.45	[illegible]	0.33	0.46	0.52	0.58	PDGFRB
-1.49	-0.57	[illegible]	-0.21	-0.18	[illegible]	[illegible]	-0.23	0.34	0.41	0.45	0.36	[illegible]	[illegible]	LOC647276
-1.49	-0.57	-0.70	-0.72	-0.68	[illegible]	[illegible]	[illegible]	-0.27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	DNAJB6
-1.49	-0.57	-0.58	-0.57	-0.66	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	MAP6D1
-1.49	-0.57	-0.44	-0.48	-0.36	[illegible]	-0.18	[illegible]	[illegible]	[illegible]	[illegible]	0.30	[illegible]	[illegible]	C6orf141
-1.49	-0.57	-0.41	-0.51	-0.49	[illegible]	[illegible]	-0.26	[illegible]	[illegible]	[illegible]	[illegible]	0.46	0.42	BDNF
-1.49	-0.57	-0.31	-0.48	-0.32	-0.31	-0.23	[illegible]	0.33	[illegible]	[illegible]	0.30	0.30	0.29	KBTBD11
-1.49	-0.57	-0.27	-0.26	-0.26	[illegible]	[illegible]	-0.25	[illegible]	0.32	0.38	[illegible]	[illegible]	[illegible]	COMMD2
-1.49	-0.57	[illegible]	-0.29	-0.37	[illegible]	[illegible]	[illegible]	0.42	[illegible]	[illegible]	0.39	[illegible]	[illegible]	None
-1.49	-0.57	-0.29	-0.27	-0.28	-0.29	-0.24	-0.20	0.39	[illegible]	[illegible]	[illegible]	0.33	[illegible]	ZNF271
-1.48	-0.57	-0.52	-0.42	-0.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.36	[illegible]	0.33	SQLE
-1.48	-0.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.29	0.49	0.40	0.57	[illegible]	[illegible]	-0.22	RAD23B
-1.48	-0.57	[illegible]	[illegible]	[illegible]	-0.29	[illegible]	-0.24	0.58	0.51	0.46	0.30	-0.18	[illegible]	NAE1
-1.48	-0.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.52	0.52	0.69	0.38	[illegible]	[illegible]	HMGB1
-1.48	-0.57	-0.35	-0.32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.31	0.33	0.38	[illegible]	[illegible]	SLC39A6
-1.48	-0.57	-0.47	-0.42	-0.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	XRN1
-1.48	-0.57	[illegible]	[illegible]	[illegible]	[illegible]	-0.19	-0.36	0.70	0.56	0.54	-0.20	-0.57	-0.43	HIGD1A
-1.48	-0.57	[illegible]	[illegible]	[illegible]	-0.20	[illegible]	-0.23	0.56	0.35	0.47	[illegible]	[illegible]	[illegible]	TUB
-1.48	-0.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.69	0.30	0.56	[illegible]	-0.37	-0.21	SSTR2
-1.48	-0.57	-0.50	-0.48	-0.54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TDG
-1.48	-0.56	-0.26	-0.31	-0.28	[illegible]	[illegible]	[illegible]	0.44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CTXN1
-1.48	-0.56	-0.21	-0.40	-0.25	[illegible]	-0.24	-0.30	0.34	[illegible]	[illegible]	0.29	[illegible]	[illegible]	NFE2L2
-1.48	-0.56	[illegible]	[illegible]	[illegible]	[illegible]	-0.20	[illegible]	0.66	0.41	0.48	[illegible]	-0.21	[illegible]	BMS1P5
-1.48	-0.56	-0.30	-0.26	-0.33	[illegible]	-0.18	-0.23	0.40	[illegible]	[illegible]	[illegible]	0.43	0.40	CRTC3
-1.48	-0.56	[illegible]	[illegible]	[illegible]	[illegible]	-0.18	[illegible]	0.60	0.43	0.40	0.31	[illegible]	[illegible]	HNRPH3
-1.48	-0.56	-0.30	-0.23	-0.35	-0.42	-0.40	-0.34	0.30	[illegible]	[illegible]	0.38	0.66	0.42	SEL1L3
-1.48	-0.56	-0.43	-0.55	-0.66	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ODC1
-1.47	-0.56	-0.27	-0.21	-0.27	[illegible]	[illegible]	[illegible]	0.34	[illegible]	0.33	[illegible]	[illegible]	[illegible]	SLC35F2
-1.47	-0.56	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.77	0.41	0.46	[illegible]	-0.29	-0.40	FKBP14
-1.47	-0.56	[illegible]	[illegible]	[illegible]	[illegible]	-0.17	-0.17	0.81	0.43	0.51	[illegible]	-0.26	-0.27	ZNF394
-1.47	-0.55	-0.28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.41	0.44	0.40	[illegible]	-0.22	[illegible]	ZNF286C
-1.47	-0.55	[illegible]	-0.24	-0.19	[illegible]	[illegible]	-0.31	0.35	0.34	0.38	[illegible]	[illegible]	[illegible]	KCTD20
-1.47	-0.55	-0.60	-0.60	-0.44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	C8orf76
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.90	0.35	0.59	[illegible]	-0.22	-0.30	LOC728903
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	-0.18	[illegible]	0.57	0.60	0.51	[illegible]	-0.35	-0.22	SACM1L
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.21	0.48	0.68	0.53	[illegible]	[illegible]	[illegible]	FAM49B
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	-0.25	-0.29	0.57	0.54	0.50	[illegible]	[illegible]	-0.23	ZMAT3
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	-0.24	-0.29	0.64	0.45	0.64	[illegible]	-0.37	[illegible]	BLZF1
-1.46	-0.55	-0.26	-0.22	[illegible]	-0.48	-0.21	-0.35	[illegible]	0.35	0.42	[illegible]	[illegible]	[illegible]	C18orf10
-1.46	-0.55	[illegible]	-0.34	-0.18	-0.22	[illegible]	[illegible]	[illegible]	[illegible]	0.40	[illegible]	0.33	[illegible]	CSNK1G1
-1.46	-0.55	-0.42	-0.28	-0.57	-0.32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	LOC40075
-1.46	-0.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.64	0.39	0.60	[illegible]	-0.21	-0.23	LOC100128
-1.46	-0.54	-0.51	-0.63	-0.42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TRAM1
-1.46	-0.54	[illegible]	[illegible]	[illegible]	-0.27	-0.34	-0.39	0.60	0.68	0.42	0.34	[illegible]	[illegible]	CALD1
-1.46	-0.54	-0.30	-0.29	-0.27	[illegible]	-0.26	[illegible]	0.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	YEATS2
-1.46	-0.54	[illegible]	-0.20	-0.21	[illegible]	-0.19	[illegible]	0.54	[illegible]	0.29	0.33	[illegible]	[illegible]	TM2D3
-1.46	-0.54	-0.33	-0.40	-0.36	[illegible]	[illegible]	-0.20	[illegible]	[illegible]	[illegible]	0.33	0.37	0.36	VPS4B
-1.46	-0.54	-0.42	-0.47	-0.33	[illegible]	-0.27	[illegible]	0.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	EFEMP1
-1.46	-0.54	-0.49	-0.55	-0.28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CD59
-1.46	-0.54	[illegible]	[illegible]	[illegible]	[illegible]	-0.28	-0.52	0.72	0.57	0.67	-0.20	-0.52	-0.74	LSM3
-1.46	-0.54	-0.19	-0.20	[illegible]	-0.22	[illegible]	[illegible]	0.64	[illegible]	0.37	[illegible]	[illegible]	[illegible]	ULK2
-1.46	-0.54	-0.29	-0.31	-0.24	-0.38	-0.47	-0.45	0.30	[illegible]	[illegible]	0.53	0.52	0.41	GSTM2
-1.46	-0.54	-0.20	-0.50	[illegible]	-0.52	-0.46	-0.56	0.45	[illegible]	[illegible]	0.98	0.96	0.91	LIMCH1
-1.45	-0.54	-0.49	-0.30	-0.48	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	ACTL6A
-1.45	-0.54	-0.29	-0.30	-0.36	[illegible]	[illegible]	[illegible]	[illegible]	0.30	[illegible]	[illegible]	[illegible]	[illegible]	PRC1
-1.45	-0.54	-0.36	-0.39	-0.29	-0.18	-0.29	-0.25	[illegible]	[illegible]	[illegible]	0.41	0.47	0.33	SELS
-1.45	-0.54	-0.32	-0.47	-0.44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	HSPB1
-1.45	-0.54	[illegible]	[illegible]	[illegible]	[illegible]	-0.21	[illegible]	0.57	0.50	0.48	[illegible]	-0.28	-0.18	DUSP19
-1.45	-0.54	-0.39	-0.37	-0.29	-0.23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.30	GTF2IP1
-1.45	-0.54	[illegible]	-0.29	[illegible]	[illegible]	[illegible]	[illegible]	0.44	0.31	0.38	[illegible]	[illegible]	[illegible]	NUCB2
-1.45	-0.54	-0.23	-0.29	-0.46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.39	0.32	[illegible]	ATP1B1
-1.45	-0.54	-0.24	-0.22	-0.24	[illegible]	[illegible]	[illegible]	0.52	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CLPX
-1.45	-0.54	-0.33	-0.30	-0.47	[illegible]	[illegible]	[illegible]	0.36	[illegible]	[illegible]	[illegible]	[illegible]	0.30	CSK
-1.45	-0.53	-0.27	-0.22	-0.28	[illegible]	-0.20	[illegible]	[illegible]	0.37	[illegible]	[illegible]	[illegible]	[illegible]	GTF2A2
-1.45	-0.53	[illegible]	[illegible]	-0.22	-0.43	-0.28	-0.33	0.40	0.39	0.34	[illegible]	0.40	[illegible]	HOMER2
-1.45	-0.53	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.47	0.42	0.32	[illegible]	[illegible]	[illegible]	SORD
-1.45	-0.53	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.19	0.54	0.39	0.54	0.29	[illegible]	[illegible]	YAP1
-1.45	-0.53	-0.36	-0.24	-0.41	[illegible]	-0.27	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	SNUPN
-1.45	-0.53	-0.39	-0.52	-0.48	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CADM4
-1.45	-0.53	-0.39	-0.35	-0.43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.21	TCEB1
-1.45	-0.53	-0.37	-0.36	[illegible]	-0.26	-0.21	-0.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.35	PTPLAD1
-1.45	-0.53	-0.55	-0.51	-0.32	[illegible]	-0.26	-0.23	[illegible]	[illegible]	[illegible]	0.46	[illegible]	[illegible]	SLC38A2
-1.45	-0.53	-0.18	-0.52	-0.25	-0.39	-0.43	-0.37	[illegible]	[illegible]	[illegible]	0.54	0.55	0.68	KIAA0363
-1.45	-0.53	-0.32	-0.27	-0.22	-0.24	[illegible]	-0.32	0.30	[illegible]	0.29	[illegible]	[illegible]	[illegible]	PPP2R2B
-1.44	-0.53	[illegible]	-0.22	[illegible]	[illegible]	[illegible]	[illegible]	0.55	[illegible]	0.46	0.30	[illegible]	[illegible]	LOC284988
-1.44	-0.53	[illegible]	[illegible]	[illegible]	[illegible]	-0.24	[illegible]	0.78	0.39	0.50	[illegible]	-0.24	-0.18	ZNF69
-1.44	-0.53	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.21	0.48	0.42	0.58	[illegible]	-0.24	-0.26	PPP2CB
-1.44	-0.53	-0.30	-0.28	-0.40	[illegible]	[illegible]	-0.19	[illegible]	0.29	[illegible]	[illegible]	0.39	0.29	MYC
-1.44	-0.53	-0.31	-0.18	-0.26	-0.24	[illegible]	-0.35	[illegible]	[illegible]	0.38	[illegible]	[illegible]	[illegible]	GNB5
-1.44	-0.53	-0.31	-0.35	[illegible]	-0.21	-0.20	[illegible]	0.34	[illegible]	[illegible]	0.41	0.43	0.43	ELP2
-1.44	-0.53	-0.21	[illegible]	[illegible]	[illegible]	[illegible]	-0.29	0.32	0.41	0.48	[illegible]	-0.18	[illegible]	GBAS

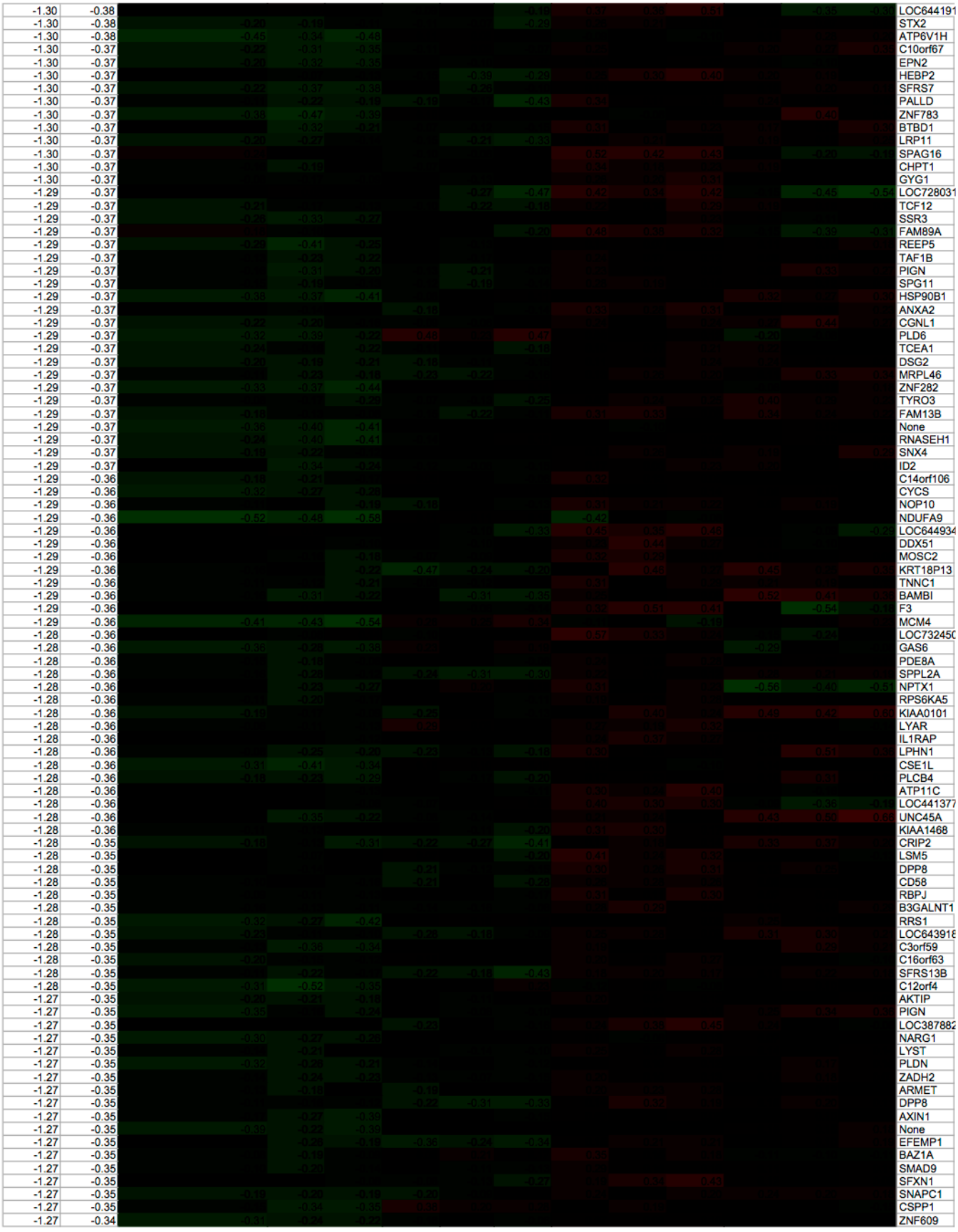


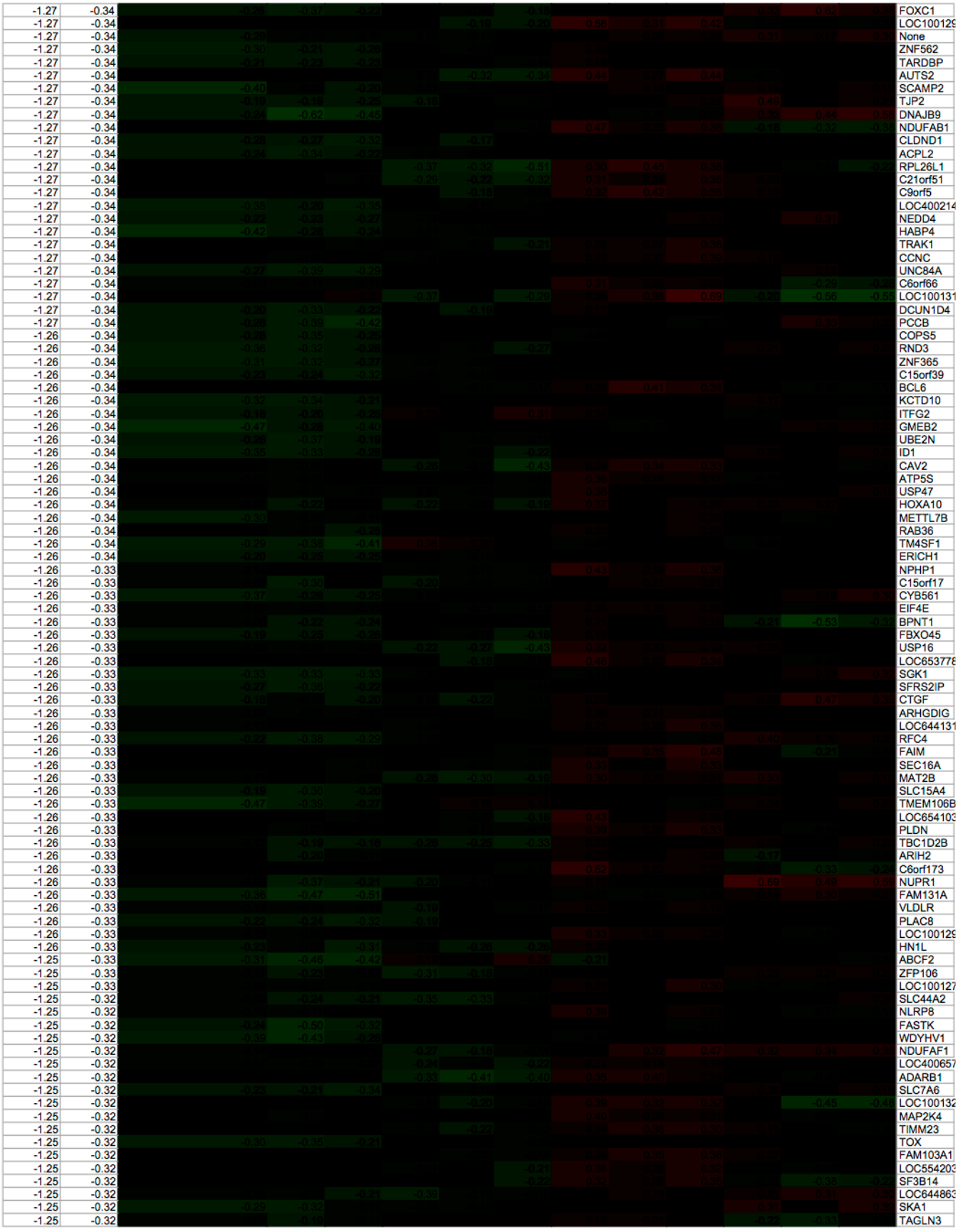
-1.44	-0.53			-0.27			-0.18			0.38	0.30	0.41			0.38		CKMT1B
-1.44	-0.53		-0.26	-0.37	-0.39				-0.21								MFI2
-1.44	-0.53									0.76	0.47	0.59			-0.22		LOC100129
-1.44	-0.53		-0.32	-0.58	-0.44												PLAC8
-1.44	-0.53		-0.44	-0.34	-0.39												CRY1
-1.44	-0.52		-0.25		-0.22				-0.19	0.50		0.29					LOC644979
-1.44	-0.52		-0.58	-0.48	-0.61		-0.34										ZNF746
-1.44	-0.52		-0.34	-0.28	-0.25		-0.21		-0.49			0.31	0.37				AKAP13
-1.44	-0.52		-0.19	-0.24	-0.33					0.30		0.32					OPA1
-1.44	-0.52		-0.21	-0.29	-0.26		-0.26		-0.21	0.34		0.31					COX5A
-1.43	-0.52		-0.40	-0.57	-0.51												RB1CC1
-1.43	-0.52			-0.24			-0.26	-0.22	-0.36	0.40	0.31	0.40					TCTEX1D2
-1.43	-0.52						-0.44		-0.26	0.36	0.45	0.42					AP3S1
-1.43	-0.52							-0.18		0.61	0.36	0.55			-0.19	-0.30	TNFSF15
-1.43	-0.52		-0.32	-0.25	-0.32		-0.25		-0.28								ABLIM1
-1.43	-0.52			-0.23	-0.33					0.33		0.31					CDC14B
-1.43	-0.52		-0.34	-0.46	-0.41												PDCD10
-1.43	-0.52		-0.30	-0.22	-0.27		-0.19	-0.22	-0.20	0.18	0.29	0.21		0.34	0.26	0.40	LOC643863
-1.43	-0.52		-0.34	-0.44	-0.43			-0.32							0.31	0.39	NFATC1
-1.43	-0.52		-0.26	-0.25	-0.36		-0.83	-0.68	-0.70								TPM1
-1.43	-0.51			-0.33	-0.25		-0.25			0.33	0.34				0.32	0.39	RAB11FIP1
-1.43	-0.51		-0.21	-0.24	-0.23												LOC647597
-1.43	-0.51			-0.19						0.35	0.38	0.42					SLMO1
-1.43	-0.51		-0.54	-0.64	-0.29									0.34			YTHDF3
-1.43	-0.51		-0.22	-0.25	-0.32		-0.18	-0.30	-0.38						0.35		ZWILCH
-1.42	-0.51		-0.21	-0.24	-0.30		-0.20	-0.24			0.32	0.39		0.34			SNUPN
-1.42	-0.51								-0.34	0.60	0.45	0.59		0.35			SUMO2
-1.42	-0.51		-0.35	-0.18	-0.30		-0.30	-0.33	-0.50			0.33		0.60	0.45	0.41	GADD45B
-1.42	-0.51		-0.18		-0.37			-0.24	-0.30		0.31						TPM4
-1.42	-0.51		-0.33	-0.40	-0.36		-0.23										RSL24D1
-1.42	-0.51		-0.42	-0.38	-0.46		-0.18										ABCF2
-1.42	-0.51						-0.24		-0.21	0.58	0.44	0.39					FLNB
-1.42	-0.51		-0.32	-0.38	-0.28		-0.40	-0.37						0.32	0.68	0.61	RBPMS2
-1.42	-0.51									0.64	0.32	0.48					C15orf63
-1.42	-0.51		-0.26	-0.31	-0.28		-0.44				0.33						EFEMP1
-1.42	-0.51		-0.18	-0.28	-0.34		-0.46	-0.22	-0.33	0.31				0.33	0.35	0.42	RBP7
-1.42	-0.51		-0.29	-0.29	-0.29			-0.22	-0.29								LOC100133
-1.42	-0.51									0.65	0.36	0.49					TFG
-1.42	-0.51			-0.25	-0.34			-0.18	-0.21	0.40					0.30		LOC100128
-1.42	-0.50		-0.20	-0.37			-0.50	-0.50	-0.38			0.48		0.45	0.43		LBH
-1.42	-0.50		-0.34	-0.60	-0.45										0.39		MEIS2
-1.42	-0.50		-0.20	-0.28	-0.31					0.31		0.32					PSTPIP2
-1.42	-0.50		-0.44	-0.29	-0.33		-0.24		-0.25								WDR73
-1.42	-0.50		-0.46	-0.47	-0.33												PRKAG2
-1.42	-0.50		-0.48	-0.62	-0.43												SMARCD3
-1.42	-0.50			-0.28				-0.25	-0.31	0.37		0.37		0.32			OIP5
-1.42	-0.50								-0.28	0.44	0.42	0.52					LOC143543
-1.42	-0.50			-0.28						0.36	0.36	0.31					NFATC2IP
-1.42	-0.50		-0.53	-0.57	-0.42												KIAA0196
-1.41	-0.50						-0.29			0.65		0.54		-0.31	-0.59	-0.44	ALCAM
-1.41	-0.50		-0.31	-0.38	-0.26												None
-1.41	-0.50						-0.29	-0.29	-0.19	0.54	0.44	0.47					ANTXR1
-1.41	-0.50		-0.45	-0.28	-0.43		-0.20										PDIA4
-1.41	-0.50		-0.26		-0.21					0.41					0.37		None
-1.41	-0.50		-0.50	-0.42	-0.42												PXMP3
-1.41	-0.50			-0.33	-0.23		-0.20		-0.21	0.36							LOC645296
-1.41	-0.50		-0.34	-0.27	-0.31					0.31					0.57		ISG20L1
-1.41	-0.50		-0.22	-0.24	-0.44			-0.24				0.34					RAB27A
-1.41	-0.50									0.42	0.48	0.56					CXCL12
-1.41	-0.50		-0.26	-0.26			-0.20	-0.18			0.35	0.31					PNMA2
-1.41	-0.50		-0.19	-0.33	-0.19		-0.27	-0.26	-0.24								CPS1
-1.41	-0.49									0.56	0.31	0.44					HDGFRP3
-1.41	-0.49			-0.23						0.49		0.34					ZNF558
-1.41	-0.49		-0.31	-0.46			-0.22	-0.20						0.32			NDFIP2
-1.41	-0.49									0.39	0.41	0.32		0.39			HNRNPH3
-1.41	-0.49		-0.31	-0.19						0.35		0.39					YWHAG
-1.40	-0.49		-0.41	-0.55	-0.57		0.30										KRBA1
-1.40	-0.49			-0.21	-0.29			-0.23	-0.20	0.56	0.32				0.30		PTPRE
-1.40	-0.49		-0.31	-0.46	-0.46												SCARB1
-1.40	-0.49		-0.48	-0.55	-0.59							-0.18					RAD51AP1
-1.40	-0.49									0.62	0.44	0.37			-0.19		RTN4
-1.40	-0.49		-0.46	-0.48	-0.64												GALNT11
-1.40	-0.49		-0.20	-0.30	-0.23			-0.18	-0.20								LOC402560
-1.40	-0.49		-0.21	-0.22	-0.32						0.32						LASS1
-1.40	-0.49				-0.28			-0.17	-0.29		0.36	0.41					GTPBP8
-1.40	-0.49		-0.25	-0.21	-0.19		-0.28	-0.22						0.33	0.38		RAD51
-1.40	-0.49									0.39	0.39	0.37		0.29			METRNL
-1.40	-0.49							-0.18		0.53	0.51	0.39			-0.23		LOC100132
-1.40	-0.49		-0.31	-0.21					-0.18	0.44							None
-1.40	-0.49		-0.24				-0.24	-0.28	-0.36	0.38							MEX3B
-1.40	-0.49		-0.40	-0.54	-0.53												CENTG3
-1.40	-0.48		-0.19	-0.33						0.33							RAB40B
-1.40	-0.48		-0.31	-0.39	-0.42												SGK1
-1.40	-0.48		-0.21	-0.28	-0.36					0.36	0.43	0.41					DCAF6
-1.40	-0.48		-0.37	-0.24			-0.28			0.35							ARPP19
-1.40	-0.48		-0.39	-0.23	-0.35				-0.22		0.30						LOC731751
-1.40	-0.48		-0.26		-0.27		-0.28	-0.25	-0.27						0.30		TAGLN
-1.40	-0.48									0.64		0.52			-0.38		LOC641768
-1.40	-0.48									0.62	0.36	0.51					MYO3B
-1.40	-0.48		-0.19	-0.32				-0.27				0.45		0.31			ALDH5A1
-1.40	-0.48							-0.32		0.62	0.36	0.48			-0.21		HSPC268
-1.40	-0.48		-0.25					-0.34		0.41	0.31	0.51			-0.22		AKAP12
-1.40	-0.48		-0.26	-0.40			-0.58	-0.48	-0.56					0.45	0.33		DMKN
-1.39	-0.48								-0.31	0.50	0.38	0.46			-0.24		LOC730107

-1.39	-0.48				-0.20	-0.27	0.38	0.47	0.43			-0.35	-0.43	UGP2
-1.39	-0.48	-0.28	-0.19	-0.36										PRKAG2
-1.39	-0.48	-0.40	-0.38	-0.22							0.36			MATN2
-1.39	-0.48		-0.21					0.41	0.40	0.32				CTNNAL1
-1.39	-0.48		-0.30		-0.20			0.37						C4BPB
-1.39	-0.48	-0.53	-0.55	-0.46										DERL1
-1.39	-0.48							0.52		0.34				NDRG3
-1.39	-0.48							0.56	0.36	0.44			-0.39	FAM119A
-1.39	-0.47	-0.19	-0.35		-0.29	-0.25	-0.35	0.29			0.41	0.38		CYB5A
-1.39	-0.47	-0.18		-0.28	-0.44	-0.53	-0.57	0.35						TGFBI
-1.39	-0.47	-0.51	-0.59	-0.38										USP13
-1.39	-0.47				0.44			0.58	0.30			-0.18	-0.25	LYAR
-1.39	-0.47	-0.22						0.38	0.35	0.39				GNB4
-1.39	-0.47	-0.31	-0.28	-0.42							0.32			HMGB2
-1.39	-0.47	-0.28			-0.20		-0.19		0.43	0.32	0.44		0.33	SNAP23
-1.39	-0.47		-0.36	-0.36		-0.22								TMX3
-1.39	-0.47	-0.28		-0.21				0.37		0.40	0.54	0.52		LOC728014
-1.39	-0.47	-0.23	-0.44	-0.35										LARP1B
-1.39	-0.47							0.60		0.53				CBL
-1.38	-0.47			-0.23				0.54		0.39		-0.30	-0.25	LOC92659
-1.38	-0.47		-0.26	-0.20	-0.17	-0.34	-0.20	0.30		0.32	0.40	0.61	0.69	DSP
-1.38	-0.47	-0.18	-0.40	-0.38		-0.27								CLIP1
-1.38	-0.47	-0.30	-0.18	-0.31						0.31				FJX1
-1.38	-0.47							0.53	0.43	0.45				CRELD1
-1.38	-0.47		-0.34	-0.22	-0.27	-0.28	-0.18				0.52	0.61	0.59	VTN
-1.38	-0.47		-0.18		-0.22	-0.18	-0.23	0.38		0.31	0.30			ZWILCH
-1.38	-0.47		-0.21					0.41	0.35	0.31				USP14
-1.38	-0.47	-0.24	-0.25	-0.20	-0.54	-0.39	-0.48				0.45	0.71	0.44	GLIPR2
-1.38	-0.47		-0.23		-0.33	-0.28		0.50					0.31	GOLGA8B
-1.38	-0.47	-0.35	-0.34	-0.23	-0.37	-0.25						0.35	0.37	TSPAN3
-1.38	-0.47	-0.47	-0.57	-0.58					-0.20					NSMAF
-1.38	-0.46	-0.21						0.38		0.38	0.31			SFRS3
-1.38	-0.46	-0.24		-0.18	-0.28					0.40		0.35		LEO1
-1.38	-0.46	-0.22					-0.24	0.54		0.35				ADNP
-1.38	-0.46	-0.35	-0.25	-0.18	-0.25		-0.25				0.32	0.35	0.40	MORF4L1
-1.38	-0.46	-0.45	-0.30			-0.18		0.30			0.48	0.29	0.35	MBNL2
-1.38	-0.46		-0.32		-0.33	-0.38	-0.29	0.25		0.35				SMAD7
-1.38	-0.46	-0.20		-0.27	-0.24	-0.21	-0.37		0.31			0.35	0.33	ETS2
-1.38	-0.46		-0.23					0.25		0.30				ALDH1B1
-1.38	-0.46	-0.27	-0.27	-0.38										SLTM
-1.38	-0.46							0.45	0.39	0.40				LOC402057
-1.38	-0.46	-0.25	-0.54	-0.29							0.53		0.35	PMAIP1
-1.38	-0.46	-0.23	-0.25	-0.30			-0.23							HMGB2
-1.38	-0.46		-0.40	-0.29	-0.24			0.37						HLTF
-1.38	-0.46						-0.35	0.56	0.51	0.56		-0.19		LOC653147
-1.38	-0.46	-0.21	-0.33	-0.27				0.38						PTPLB
-1.38	-0.46								0.47	0.41				NCBP2
-1.38	-0.46	-0.27	-0.40					0.38						ACTR6
-1.38	-0.46							0.41		0.37				AGL
-1.38	-0.46	-0.22	-0.25	-0.19			-0.30							FAM96A
-1.37	-0.46	-0.30	-0.30	-0.40										RARA
-1.37	-0.46				-0.18	-0.22		0.44						PTGFRN
-1.37	-0.46	-0.34	-0.39	-0.32							0.41	0.37	0.45	MTMR10
-1.37	-0.46			-0.20			-0.18	0.44	0.36	0.34				LDOC1L
-1.37	-0.46	-0.35	-0.25	-0.23		-0.28	-0.42					0.40		KIAA0427
-1.37	-0.46	-0.23			-0.28	-0.31	-0.27	0.31						ANXA2
-1.37	-0.46	-0.35	-0.45	-0.27										MFSD1
-1.37	-0.46			-0.33				0.38						LOC644979
-1.37	-0.45	-0.20	-0.30		-0.20		-0.17		0.34					LOC647349
-1.37	-0.45							0.57		0.52				HMGB1L1
-1.37	-0.45	-0.30	-0.29			-0.26	-0.22							TJP1
-1.37	-0.45	-0.18	-0.23			-0.18								DAAM1
-1.37	-0.45	-0.23	-0.48	-0.22	-0.28							0.36		ATP1B1
-1.37	-0.45			-0.30				0.33						MGAT2
-1.37	-0.45	-0.48	-0.29	-0.33	0.29									LYRM5
-1.37	-0.45	-0.32	-0.23	-0.25	-0.21		-0.33					0.48		MYL7
-1.37	-0.45	-0.45	-0.53	-0.28							0.30	0.38		ZNF467
-1.37	-0.45	-0.31	-0.31	-0.34				0.30						CITED2
-1.37	-0.45	-0.32	-0.31	-0.20	0.30									SDCBP
-1.37	-0.45			-0.28	-0.17		-0.19		0.37					FAM103A1
-1.37	-0.45	-0.24	-0.26	-0.28										LOC653381
-1.36	-0.45							0.53		0.32				MGA
-1.36	-0.45	-0.18	-0.24	-0.31	-0.47	-0.45	-0.29				0.71	0.60	0.51	IGFBP3
-1.36	-0.45	-0.34	-0.38	-0.54										LMBR1
-1.36	-0.45					-0.27		0.48	0.30	0.45		-0.34	-0.18	KIAA1751
-1.36	-0.45						-0.20	0.50		0.50		-0.21	-0.20	AUP1
-1.36	-0.45		-0.21	-0.27	-0.20	-0.23		0.33						BEND7
-1.36	-0.45	-0.24		-0.28	-0.24		-0.18		0.34		0.40	0.38	0.45	TMEM85
-1.36	-0.45	-0.29	-0.30	-0.20				0.29						AGFG1
-1.36	-0.44	-0.42	-0.30	-0.34										ACP1
-1.36	-0.44		-0.23	-0.25			-0.25	0.30			0.34			GALNT1
-1.36	-0.44	-0.27	-0.20				-0.25	0.38			0.30	0.33		CCBE1
-1.36	-0.44								0.34	0.45				THYN1
-1.36	-0.44	-0.36	-0.37	-0.41										ZNF462
-1.36	-0.44	-0.28	-0.42	-0.36										RFC4
-1.36	-0.44	-0.39	-0.31	-0.21	-0.28						0.31	0.32		ZNF770
-1.36	-0.44	-0.36	-0.40	-0.35										TMEM22
-1.36	-0.44		-0.17		-0.21	-0.22	-0.25		0.37	0.31			0.30	ROR2
-1.36	-0.44	-0.42	-0.57	-0.54										KRAS
-1.36	-0.44	-0.31		-0.29		-0.21	-0.30	0.28			0.30	0.33		CD276
-1.36	-0.44	-0.19						0.41	0.30					TIPIN
-1.36	-0.44							0.47	0.30					WRNIP1
-1.36	-0.44	-0.18	-0.21		-0.36	-0.22			0.44		0.29			NUSAP1
-1.36	-0.44		-0.18	-0.43		-0.20								TAF1B

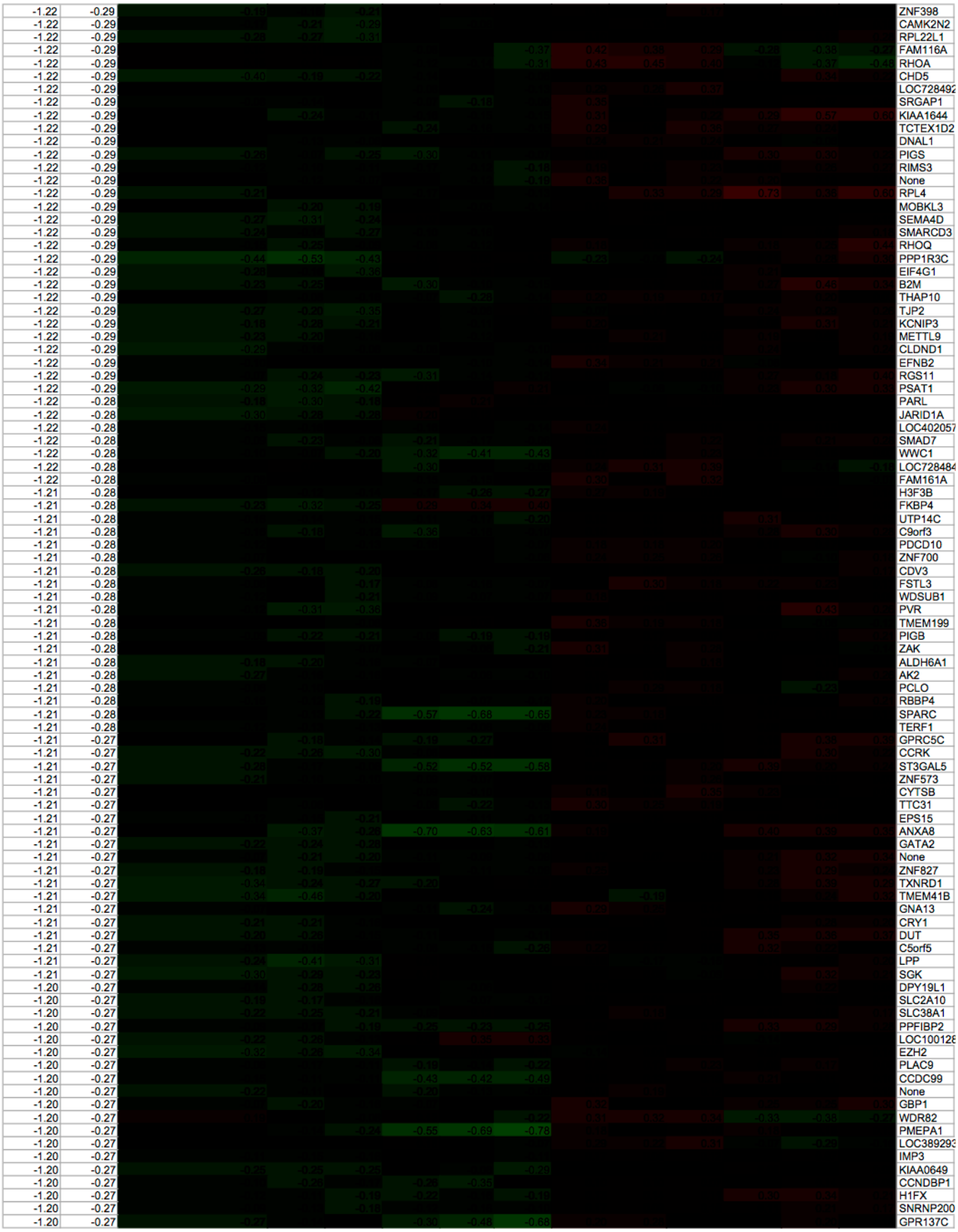
-1.36	-0.44		-0.20					0.44						SENP6
-1.36	-0.44	-0.24				-0.18	-0.18	0.35		0.34	0.35			UBE2Q2
-1.36	-0.44		-0.27					0.35	0.29					None
-1.36	-0.44	-0.34	-0.53	-0.37									0.32	IL1RAP
-1.36	-0.44					-0.19		0.59	0.34	0.38			-0.27	LOC100130
-1.36	-0.44				-0.20		-0.28	0.31	0.39	0.33	0.29			LOC100133
-1.35	-0.44					-0.19	-0.39	0.48	0.33	0.59				IGF2BP3
-1.35	-0.44		-0.21			-0.19		0.30	0.40					GFRA1
-1.35	-0.44	-0.36	-0.38	-0.29										PPP1CB
-1.35	-0.44	-0.21	-0.23	-0.37		-0.31	-0.22							PLEKHG4
-1.35	-0.44							0.45	0.31	0.30				NGFRAP1
-1.35	-0.44	-0.24	-0.22					0.34			-0.21			None
-1.35	-0.44		-0.24	-0.25	-0.24	-0.29	-0.28	0.29			0.32			LOC647946
-1.35	-0.44		-0.28											KLHDC5
-1.35	-0.44							0.45	0.31	0.46				ALG5
-1.35	-0.43	-0.18	-0.38	-0.27			-0.21							KIAA1598
-1.35	-0.43		-0.32	-0.33	-0.40	-0.29	-0.26				0.30		0.30	ZFAND6
-1.35	-0.43	-0.30	-0.39	-0.40		-0.19	-0.20							PSPH
-1.35	-0.43	-0.43	-0.41	-0.35										USP36
-1.35	-0.43							0.48		0.36				SKP2
-1.35	-0.43						-0.19	0.52	0.34	0.42				C9orf30
-1.35	-0.43	-0.48	-0.54	-0.40										DNAJB11
-1.35	-0.43	-0.42	-0.39	-0.31										FNBP1
-1.35	-0.43							0.42		0.37				NIPSNAP1
-1.35	-0.43	-0.21	-0.18	-0.22				0.38						NAP1L1
-1.35	-0.43	-0.23	-0.23	-0.26				0.30			0.30			HERC2
-1.35	-0.43							0.53	0.38	0.39				CXCL12
-1.35	-0.43	-0.27	-0.21	-0.25				0.33						LOC646909
-1.35	-0.43		-0.31	-0.32			-0.24							BNC1
-1.35	-0.43	-0.32	-0.46	-0.47										DYRK4
-1.34	-0.43	-0.36	-0.21	-0.35										RAB2A
-1.34	-0.43							0.53	0.49	0.53				LOC644131
-1.34	-0.43	-0.19						0.48	0.29	0.34				XPO1
-1.34	-0.43	-0.38	-0.41	-0.30		-0.25	-0.25							CHPT1
-1.34	-0.42	-0.48	-0.46	-0.42										AGAP3
-1.34	-0.42	-0.42	-0.38	-0.29	-0.29									NCAPG2
-1.34	-0.42	-0.23	-0.23			-0.34	-0.28	0.30						SYNM
-1.34	-0.42			-0.27					0.30	0.34				SCHIP1
-1.34	-0.42	-0.25	-0.50	-0.38										LOC646316
-1.34	-0.42		-0.22	-0.19				0.35						GALC
-1.34	-0.42	-0.29	-0.32											FAM69B
-1.34	-0.42				-0.24	-0.28		0.51	0.33	0.38				KAT2B
-1.34	-0.42		-0.25	-0.28										IVD
-1.34	-0.42	-0.29	-0.31	-0.36			-0.17							CYB561
-1.34	-0.42			-0.18	-0.20	-0.18	-0.35		0.33	0.38				SLTM
-1.34	-0.42	-0.43	-0.44	-0.39										IFIT2
-1.34	-0.42	-0.25	-0.32		-0.25	-0.35	-0.36				0.36	0.39	0.48	PALM
-1.34	-0.42	-0.27	-0.26		-0.19	-0.38	-0.42				0.48	0.36		FSTL1
-1.34	-0.42							0.34	0.34	0.33				PLXNA1
-1.34	-0.42						-0.21	0.56		0.34				TLK1
-1.34	-0.42		-0.37	-0.28			-0.24				0.48		0.41	ATP11B
-1.34	-0.42	-0.51	-0.35	-0.47				-0.19						SERP1
-1.34	-0.42	-0.66	-0.45	-0.50				-0.23		-0.21	1.03	0.74	0.91	CCL2
-1.34	-0.42				-0.26	-0.48	-0.43	0.46						MNS1
-1.34	-0.42		-0.22					0.40	0.31					PPP4R1
-1.34	-0.42	-0.23	-0.32		-0.19									C4BPB
-1.34	-0.42								0.52	0.52	-0.32	-0.35	-0.36	FCGRT
-1.34	-0.42	-0.20				-0.19			0.35		0.36	0.33		PSMA4
-1.34	-0.42							0.45	0.31	0.35				CBFB
-1.34	-0.42	-0.27	-0.17	-0.33										CDK4
-1.33	-0.42							0.30						SUV420H1
-1.33	-0.42	-0.25		-0.26	-0.22						0.44	0.43	0.37	RPL17
-1.33	-0.42		-0.26	-0.25		-0.29	-0.26	0.38			0.52	0.56	0.51	ADM2
-1.33	-0.42						-0.18	0.49		0.40		-0.18		AVEN
-1.33	-0.42	-0.22	-0.23				-0.24				0.51	0.46	0.61	BCKDHA
-1.33	-0.42	-0.27	-0.32	-0.28										STK38L
-1.33	-0.42	-0.32	-0.39	-0.40							0.29			WDYHV1
-1.33	-0.41			-0.30	-0.19	-0.21	-0.33			0.34	0.75	0.81	0.74	GSTM1
-1.33	-0.41	-0.30	-0.33	-0.32										KLHL29
-1.33	-0.41		-0.25	-0.19	-0.24	-0.23	-0.22	0.37						BASP1
-1.33	-0.41	-0.31	-0.37	-0.29										LOC100128
-1.33	-0.41	-0.26	-0.31	-0.24	-0.23	-0.20						0.40	0.43	CCNDBP1
-1.33	-0.41		-0.22					0.30		0.35	-0.82	-0.57	-0.85	TMEM158
-1.33	-0.41	-0.24	-0.34	-0.34			-0.21						0.31	SLC39A6
-1.33	-0.41	-0.29	-0.38	-0.34		-0.28								SLC43A3
-1.33	-0.41	-0.48	-0.42	-0.53	0.44	0.29						0.29		ZNF777
-1.33	-0.41							0.59		0.38				BFAR
-1.33	-0.41					-0.26	-0.30	0.47		0.34				KIAA1279
-1.33	-0.41	-0.42	-0.43	-0.32										WDR67
-1.33	-0.41		-0.27					0.35	0.34	0.30				DFNA5
-1.33	-0.41		-0.29	-0.25			-0.19							STAU2
-1.33	-0.41						-0.33	0.42	0.48	0.41		-0.25		PNN
-1.33	-0.41	-0.27		-0.28	-0.22					0.35	0.61	0.48	0.60	ID1
-1.33	-0.41							0.50				-0.21	-0.24	ZNF14
-1.33	-0.41	-0.26		-0.19		-0.37	-0.30	0.35			0.61	0.67	0.69	ABCA8
-1.33	-0.41						-0.24	0.65	0.33	0.40		-0.39	-0.47	LOC727821
-1.33	-0.41	-0.24			-0.22					0.35				LOC643438
-1.33	-0.41	-0.32			-0.21	-0.27	-0.31						0.30	JPH3
-1.33	-0.41							0.44		0.34				LOC100130
-1.33	-0.41	-0.44	-0.30	-0.27										PAK2
-1.32	-0.41						-0.19	0.34	0.38					NCL
-1.32	-0.41	-0.27	-0.27				-0.19							TRIP4
-1.32	-0.40	-0.25	-0.23											LOC646100
-1.32	-0.40	-0.29	-0.36	-0.23				0.30						GOLPH4

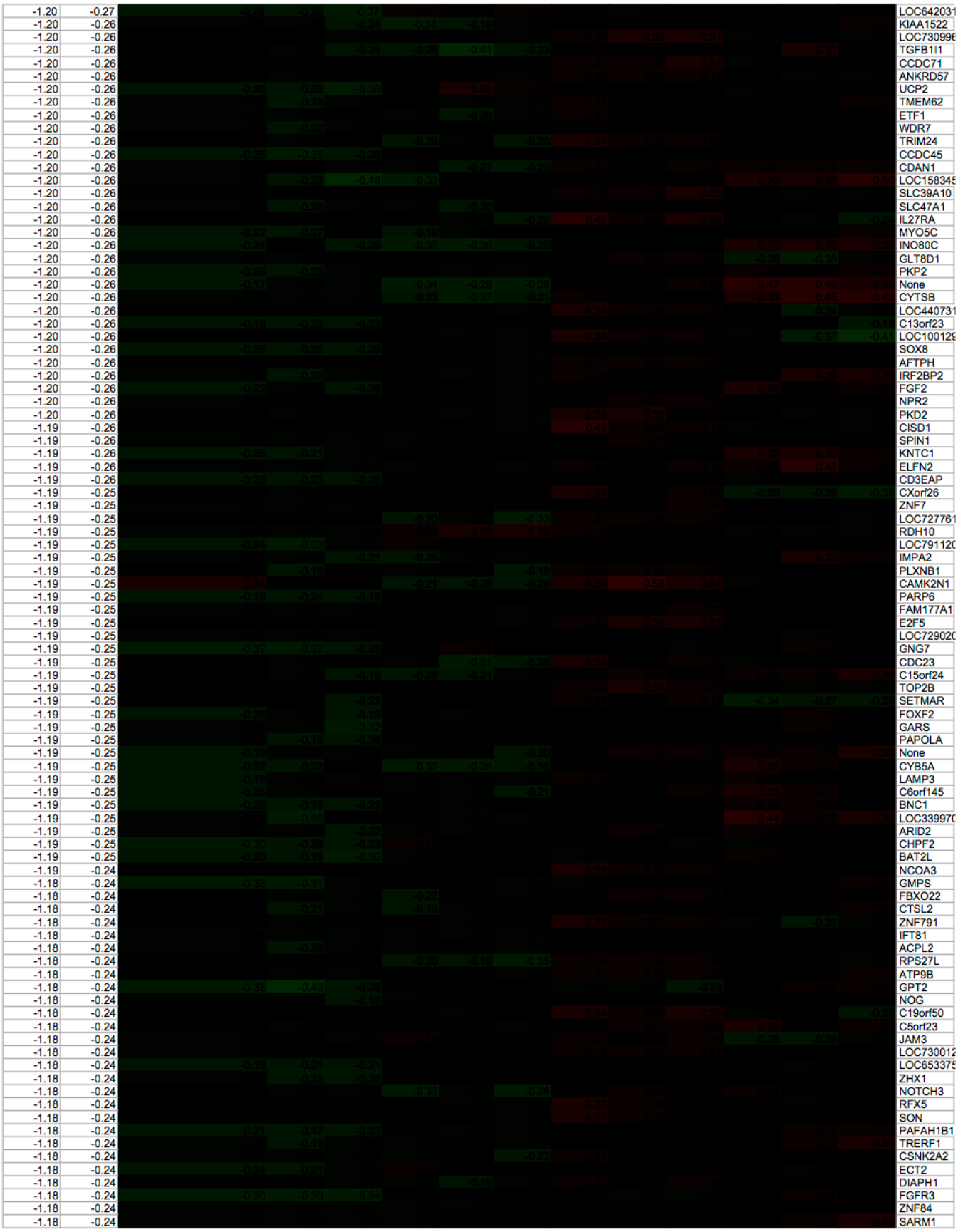




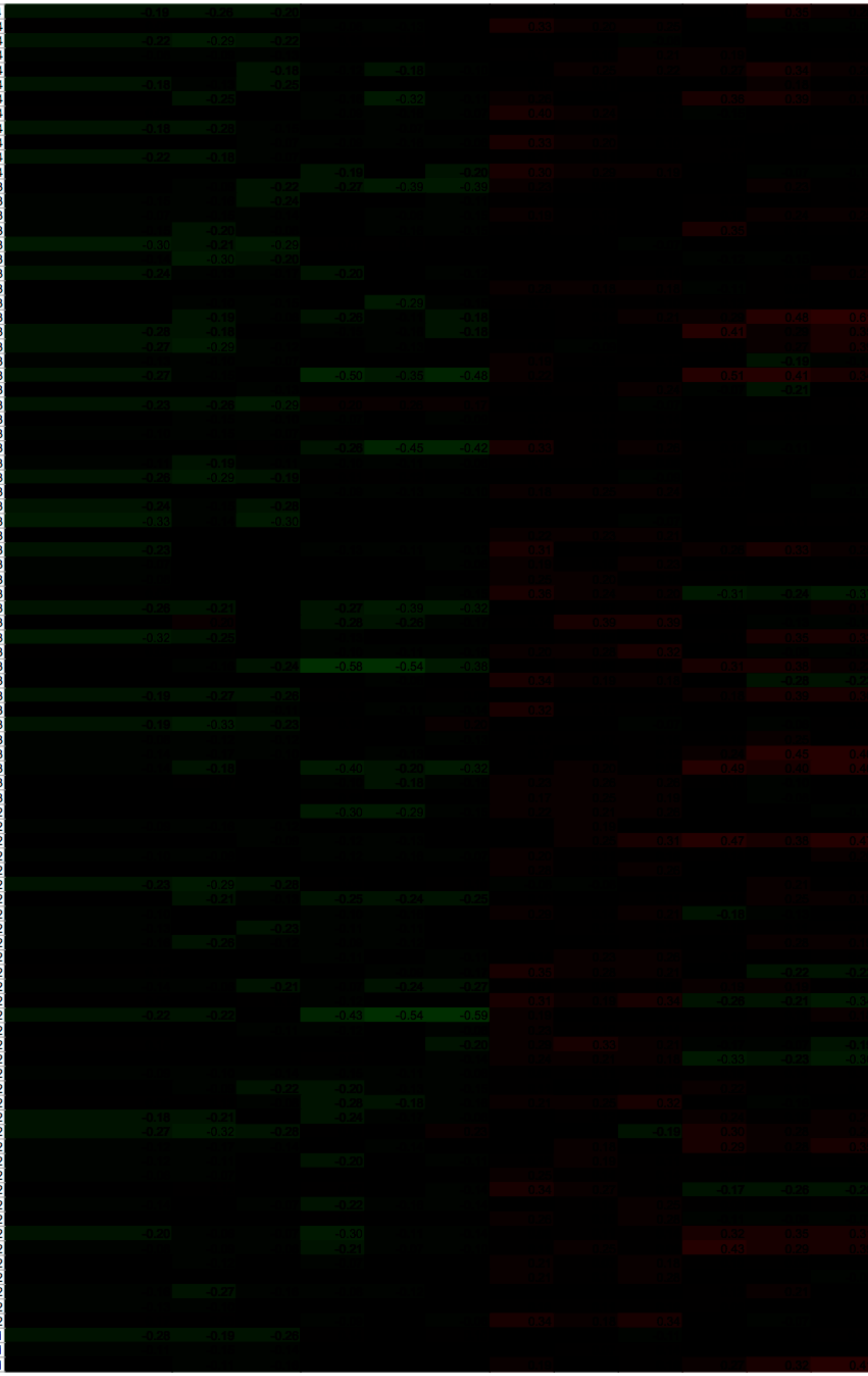


-1.25	-0.32	-0.27	-0.36		-0.20									ALDH5A1
-1.25	-0.32	-0.29	-0.23		-0.21	-0.17					0.41			LOC646688
-1.25	-0.32							0.32						CSTF2
-1.25	-0.32	-0.35	-0.35	-0.22										MYNN
-1.25	-0.32		-0.41	-0.33										ALDH6A1
-1.25	-0.32	-0.35	-0.31	-0.33										C12orf32
-1.25	-0.32	-0.30	-0.38											PRKCZ
-1.25	-0.32			-0.34							0.45	0.43	0.41	CENPV
-1.24	-0.32							0.32		0.30				EPM2AIP1
-1.24	-0.32								0.50	0.54			-0.20	LOC730029
-1.24	-0.32	-0.35	-0.30	-0.34										SLC4A2
-1.24	-0.32		-0.28	-0.27										NT5C3
-1.24	-0.32		-0.19			-0.18	-0.26							AAGAB
-1.24	-0.32	-0.28	-0.19											CAMK2B
-1.24	-0.31		-0.20											INO80
-1.24	-0.31					-0.19	-0.35	0.29		0.38		-0.22		SUMO1
-1.24	-0.31		-0.23								-0.32	-0.23		ADAP2
-1.24	-0.31	-0.19	-0.27	-0.33										HSPBL2
-1.24	-0.31							0.37		0.42		-0.32	-0.33	MRPL1
-1.24	-0.31	-0.51	-0.40	-0.48				-0.23		-0.17		0.29		NSMCE2
-1.24	-0.31	-0.19	-0.20	-0.29	-0.20						0.29	0.44	0.42	WDR61
-1.24	-0.31		-0.28	-0.34							0.42	0.42	0.47	LOC728014
-1.24	-0.31	-0.21	-0.27	-0.27										MRPS22
-1.24	-0.31		-0.28											CDC42EP1
-1.24	-0.31		-0.21							0.30				LEO1
-1.24	-0.31							0.44		0.32		-0.31	-0.35	LOC100131
-1.24	-0.31													LOC285550
-1.24	-0.31	-0.29	-0.35	-0.28										BMI1
-1.24	-0.31		-0.21											HNRPK
-1.24	-0.31		-0.37	-0.22								0.45	0.48	CENTA1
-1.24	-0.31	-0.26			-0.19	-0.30	-0.25							RICH2
-1.24	-0.31	-0.24	-0.22	-0.18	-0.23									BMP2K
-1.24	-0.31	-0.37	-0.31	-0.46										NUB1
-1.24	-0.31													ZNF469
-1.24	-0.31					-0.27	-0.40		0.33	0.38				SERINC1
-1.24	-0.31	-0.24	-0.18	-0.28										DNAJC3
-1.24	-0.31	-0.23	-0.32	-0.37								0.34	0.31	NKD2
-1.24	-0.31	-0.21	-0.31	-0.37								0.42	0.33	TJP2
-1.24	-0.31													VGLL3
-1.24	-0.31		-0.22	-0.20										METAP1
-1.24	-0.31	-0.28		-0.19	-0.35	-0.30								AKAP12
-1.24	-0.31			-0.21										UACA
-1.24	-0.30							0.38	0.33	0.37			-0.20	LOC100130
-1.23	-0.30				-0.41	-0.42	-0.48					0.34		CXXC5
-1.23	-0.30	-0.34	-0.39	-0.28							0.45	0.49	0.63	TXNRD1
-1.23	-0.30		-0.26											LOC100129
-1.23	-0.30		-0.20		-0.35									ID2
-1.23	-0.30	-0.30	-0.27	-0.28							-0.25			CHKA
-1.23	-0.30		-0.20		-0.24	-0.21								PALM
-1.23	-0.30	-0.25	-0.19				-0.19							PCF11
-1.23	-0.30	-0.25	-0.18	-0.22		-0.20								HSPA12A
-1.23	-0.30		-0.24											STRBP
-1.23	-0.30		-0.28	-0.21										TXNDC16
-1.23	-0.30	-0.17				-0.26	-0.18		0.32		0.75	0.70	0.83	BDKRB1
-1.23	-0.30					-0.21	-0.24	0.30						CELSR2
-1.23	-0.30					-0.18	-0.21	0.39						C16orf58
-1.23	-0.30		-0.18				-0.28							FAM119A
-1.23	-0.30	-0.29	-0.26	-0.25										NUSAP1
-1.23	-0.30		-0.33	-0.19										ATG12
-1.23	-0.30	-0.30	-0.20											LOC647954
-1.23	-0.30							0.38						LRRC20
-1.23	-0.30		-0.23	-0.19										VPS37B
-1.23	-0.30	-0.20				-0.18	-0.19							CANX
-1.23	-0.30				-0.33	-0.22	-0.35	0.44	0.33	0.33		-0.27	-0.44	LOC100132
-1.23	-0.30													RRAS2
-1.23	-0.30		-0.23										0.31	SERF2
-1.23	-0.30	-0.22	-0.26	-0.42										SDHALP1
-1.23	-0.30	-0.17	-0.24											MPZL1
-1.23	-0.30	-0.27	-0.27	-0.18							0.31			CCNB2
-1.23	-0.30						-0.23	0.29						C7orf28A
-1.23	-0.30							0.39		0.30			-0.31	LARP7
-1.23	-0.30							0.33						C1orf131
-1.23	-0.30													PEX11A
-1.23	-0.30									0.33				INPP5A
-1.23	-0.30	-0.23	-0.24	-0.22										SERPINB8
-1.23	-0.30	-0.37	-0.28	-0.29										UFM1
-1.23	-0.30		-0.23											TTC39C
-1.23	-0.30	-0.43	-0.29	-0.26										EIF4G1
-1.23	-0.30	-0.18				-0.20								FAM119A
-1.23	-0.30		-0.19	-0.19	-0.31	-0.34	-0.24					0.42		COMMD4
-1.23	-0.30	-0.23	-0.20											SPAG1
-1.23	-0.30													TCF4
-1.23	-0.30	-0.20			-0.36	-0.35	-0.23		0.36		0.39	0.46	0.49	CXXC1
-1.23	-0.30						-0.23	0.32	0.41	0.54		-0.25		NISCH
-1.23	-0.29							0.34		0.37		-0.19		MSH2
-1.23	-0.29	-0.22	-0.34	-0.22	0.30									TRIB1
-1.23	-0.29		-0.21		0.89	0.62	0.50							KITLG
-1.23	-0.29						-0.31	0.32						NCK2
-1.23	-0.29	-0.27	-0.26	-0.24	0.40									None
-1.23	-0.29			-0.27	-0.21		-0.19			0.30				SERF2
-1.23	-0.29					-0.20	-0.19						0.30	ZFYVE19
-1.23	-0.29													HSPA1B
-1.22	-0.29	-0.42	-0.55	-0.45				-0.22		-0.21	0.65	0.81	0.85	TRIB3
-1.22	-0.29	-0.18	-0.17	-0.18								0.32	0.32	NGRN

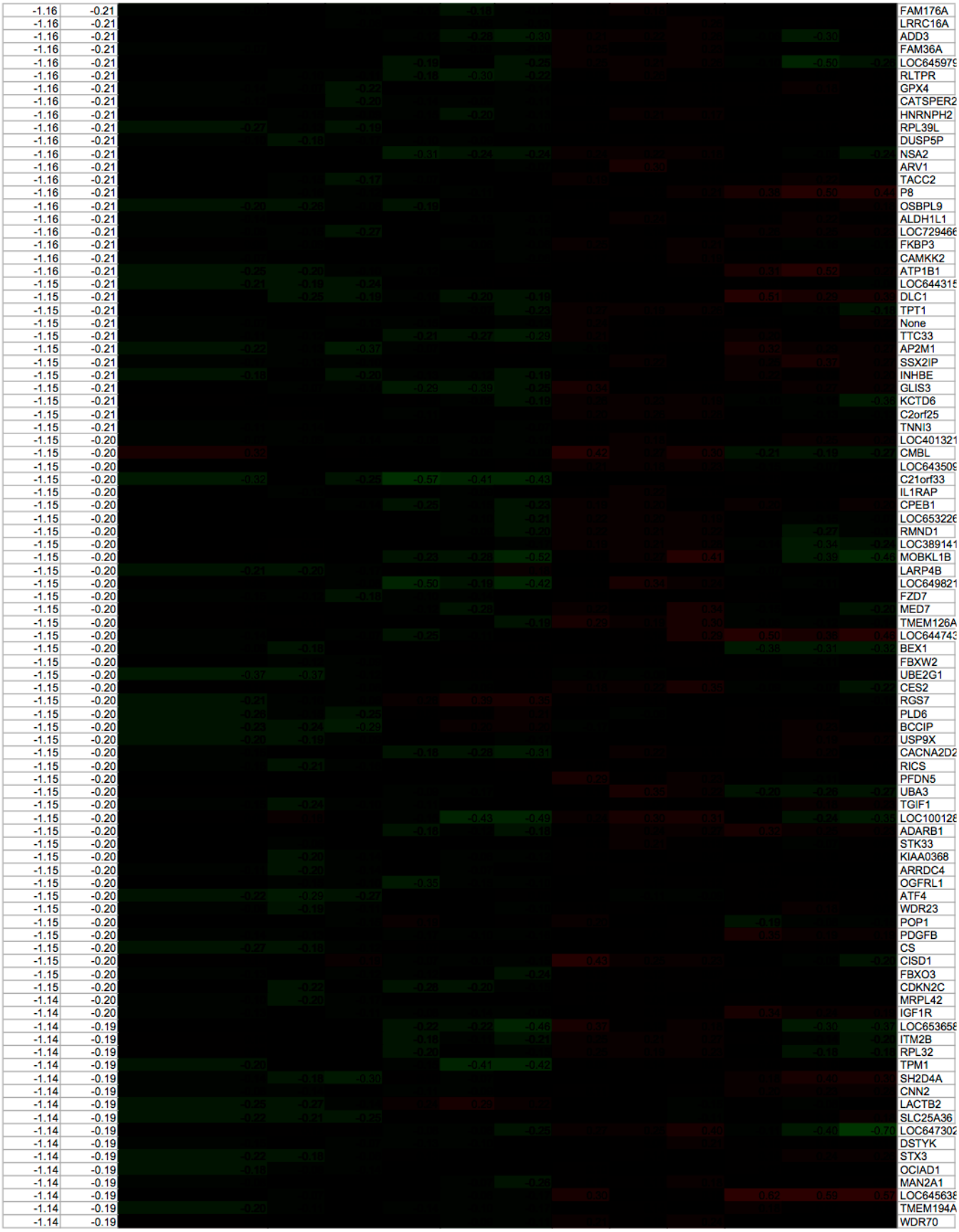


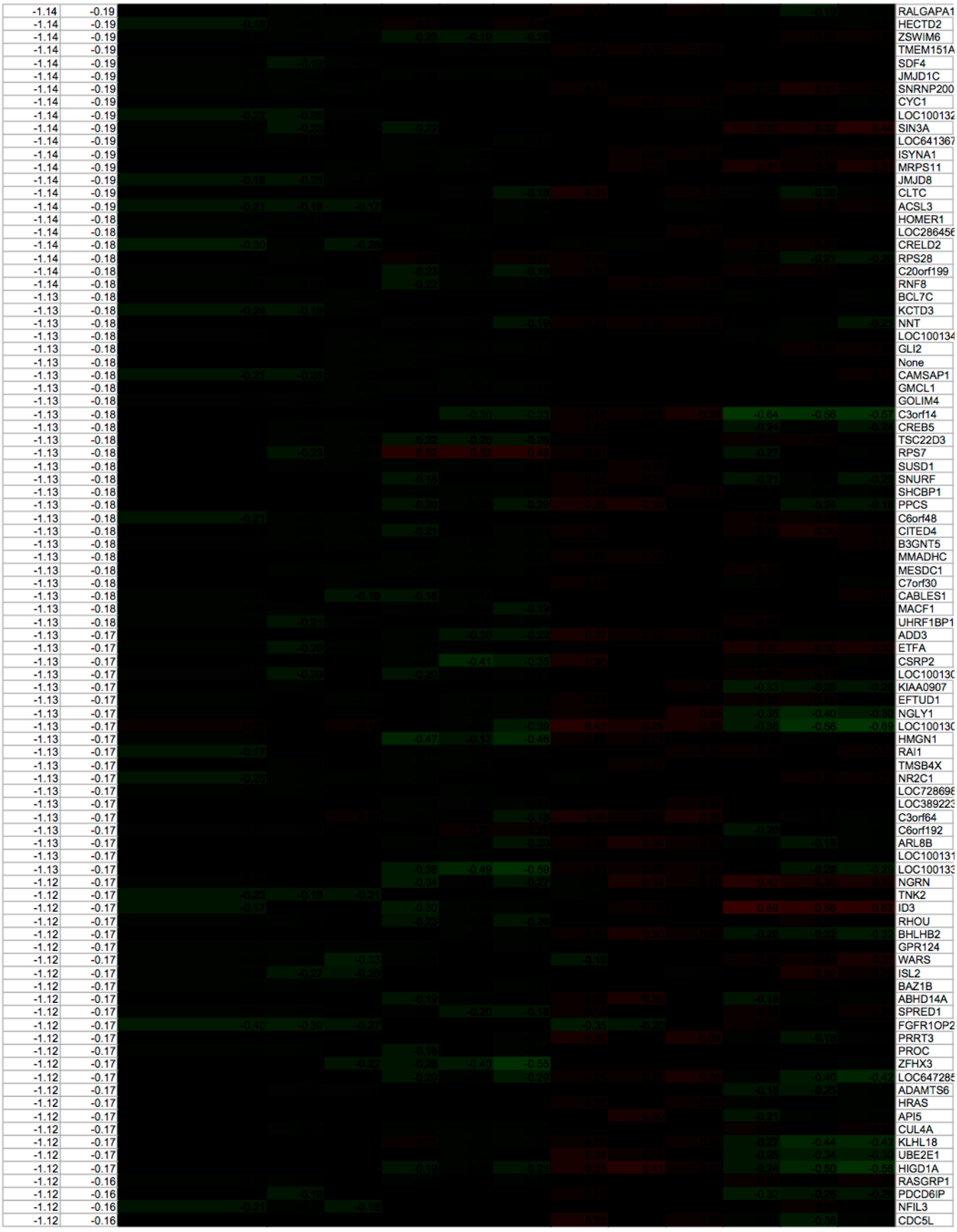


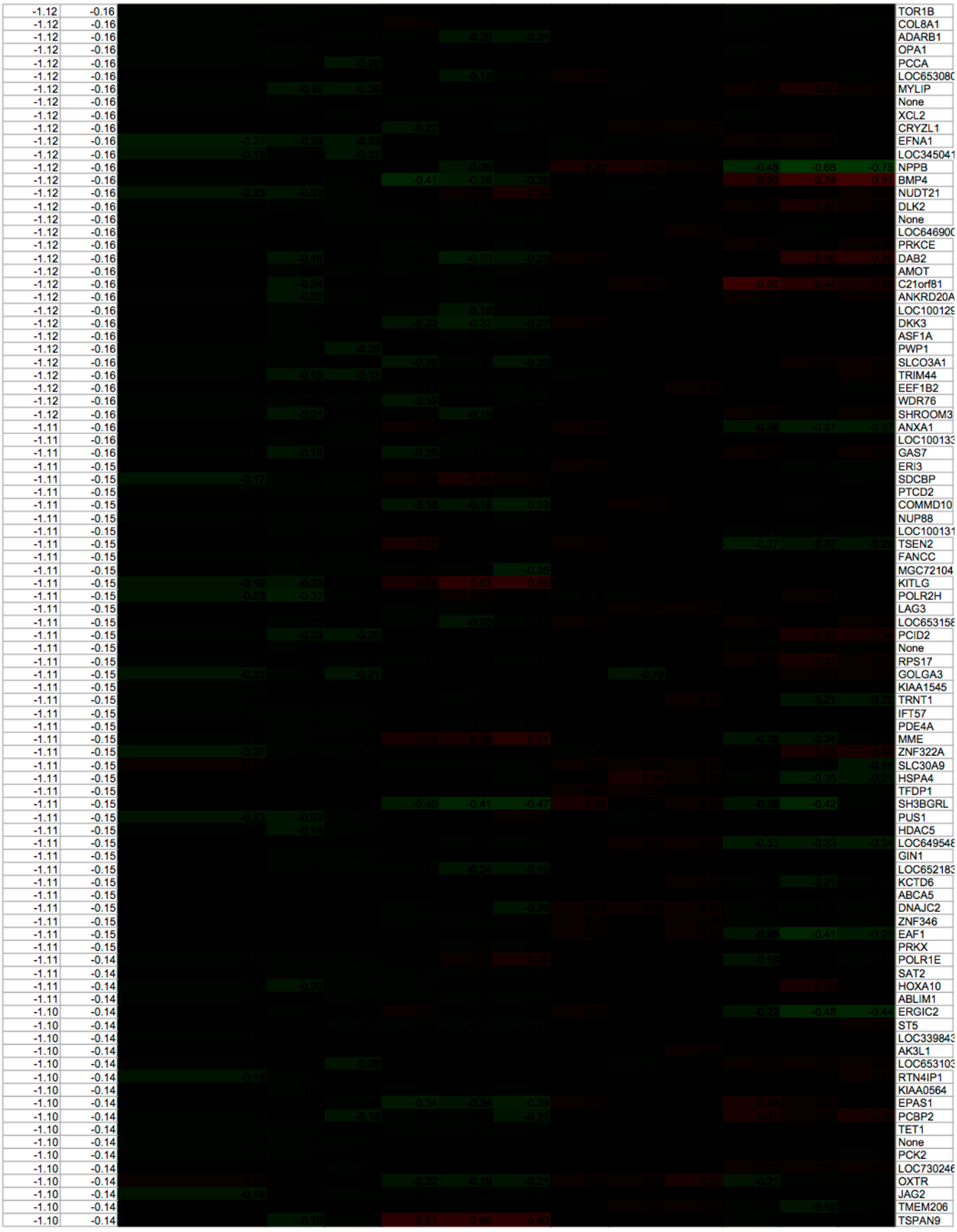
-1.20	-0.27	LOC642031
-1.20	-0.26	KIAA1522
-1.20	-0.26	LOC730996
-1.20	-0.26	TGFB1I1
-1.20	-0.26	CCDC71
-1.20	-0.26	ANKRD57
-1.20	-0.26	UCP2
-1.20	-0.26	TMEM62
-1.20	-0.26	ETF1
-1.20	-0.26	WDR7
-1.20	-0.26	TRIM24
-1.20	-0.26	CCDC45
-1.20	-0.26	CDAN1
-1.20	-0.26	LOC158345
-1.20	-0.26	SLC39A10
-1.20	-0.26	SLC47A1
-1.20	-0.26	IL27RA
-1.20	-0.26	MYO5C
-1.20	-0.26	INO80C
-1.20	-0.26	GLT8D1
-1.20	-0.26	PKP2
-1.20	-0.26	None
-1.20	-0.26	CYTSB
-1.20	-0.26	LOC440731
-1.20	-0.26	C13orf23
-1.20	-0.26	LOC100129
-1.20	-0.26	SOX8
-1.20	-0.26	AFTPH
-1.20	-0.26	IRF2BP2
-1.20	-0.26	FGF2
-1.20	-0.26	NPR2
-1.20	-0.26	PKD2
-1.19	-0.26	CISD1
-1.19	-0.26	SPIN1
-1.19	-0.26	KNTC1
-1.19	-0.26	ELFN2
-1.19	-0.26	CD3EAP
-1.19	-0.25	CXorf26
-1.19	-0.25	ZNF7
-1.19	-0.25	LOC727761
-1.19	-0.25	RDH10
-1.19	-0.25	LOC791120
-1.19	-0.25	IMPA2
-1.19	-0.25	PLXNB1
-1.19	-0.25	CAMK2N1
-1.19	-0.25	PARP6
-1.19	-0.25	FAM177A1
-1.19	-0.25	E2F5
-1.19	-0.25	LOC729020
-1.19	-0.25	GNG7
-1.19	-0.25	CDC23
-1.19	-0.25	C15orf24
-1.19	-0.25	TOP2B
-1.19	-0.25	SETMAR
-1.19	-0.25	FOXF2
-1.19	-0.25	GARS
-1.19	-0.25	PAPOLA
-1.19	-0.25	None
-1.19	-0.25	CYB5A
-1.19	-0.25	LAMP3
-1.19	-0.25	C6orf145
-1.19	-0.25	BNC1
-1.19	-0.25	LOC339970
-1.19	-0.25	ARID2
-1.19	-0.25	CHPF2
-1.19	-0.25	BAT2L
-1.19	-0.24	NCOA3
-1.18	-0.24	GMPS
-1.18	-0.24	FBXO22
-1.18	-0.24	CTSL2
-1.18	-0.24	ZNF791
-1.18	-0.24	IFT81
-1.18	-0.24	ACPL2
-1.18	-0.24	RPS27L
-1.18	-0.24	ATP9B
-1.18	-0.24	GPT2
-1.18	-0.24	NOG
-1.18	-0.24	C19orf50
-1.18	-0.24	C5orf23
-1.18	-0.24	JAM3
-1.18	-0.24	LOC730012
-1.18	-0.24	LOC653375
-1.18	-0.24	ZHX1
-1.18	-0.24	NOTCH3
-1.18	-0.24	RFX5
-1.18	-0.24	SON
-1.18	-0.24	PAFAH1B1
-1.18	-0.24	TRERF1
-1.18	-0.24	CSNK2A2
-1.18	-0.24	ECT2
-1.18	-0.24	DIAPH1
-1.18	-0.24	FGFR3
-1.18	-0.24	ZNF84
-1.18	-0.24	SARM1

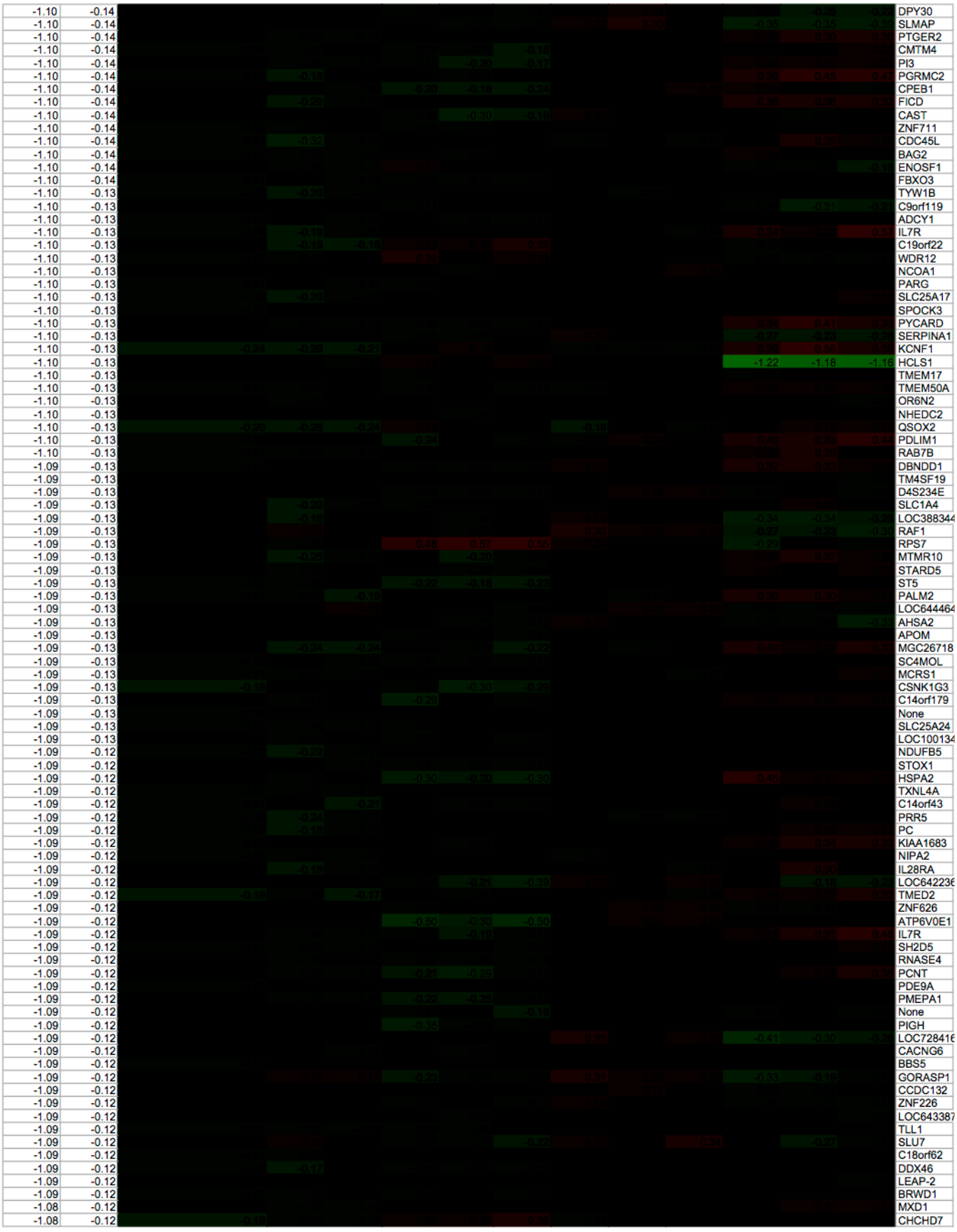


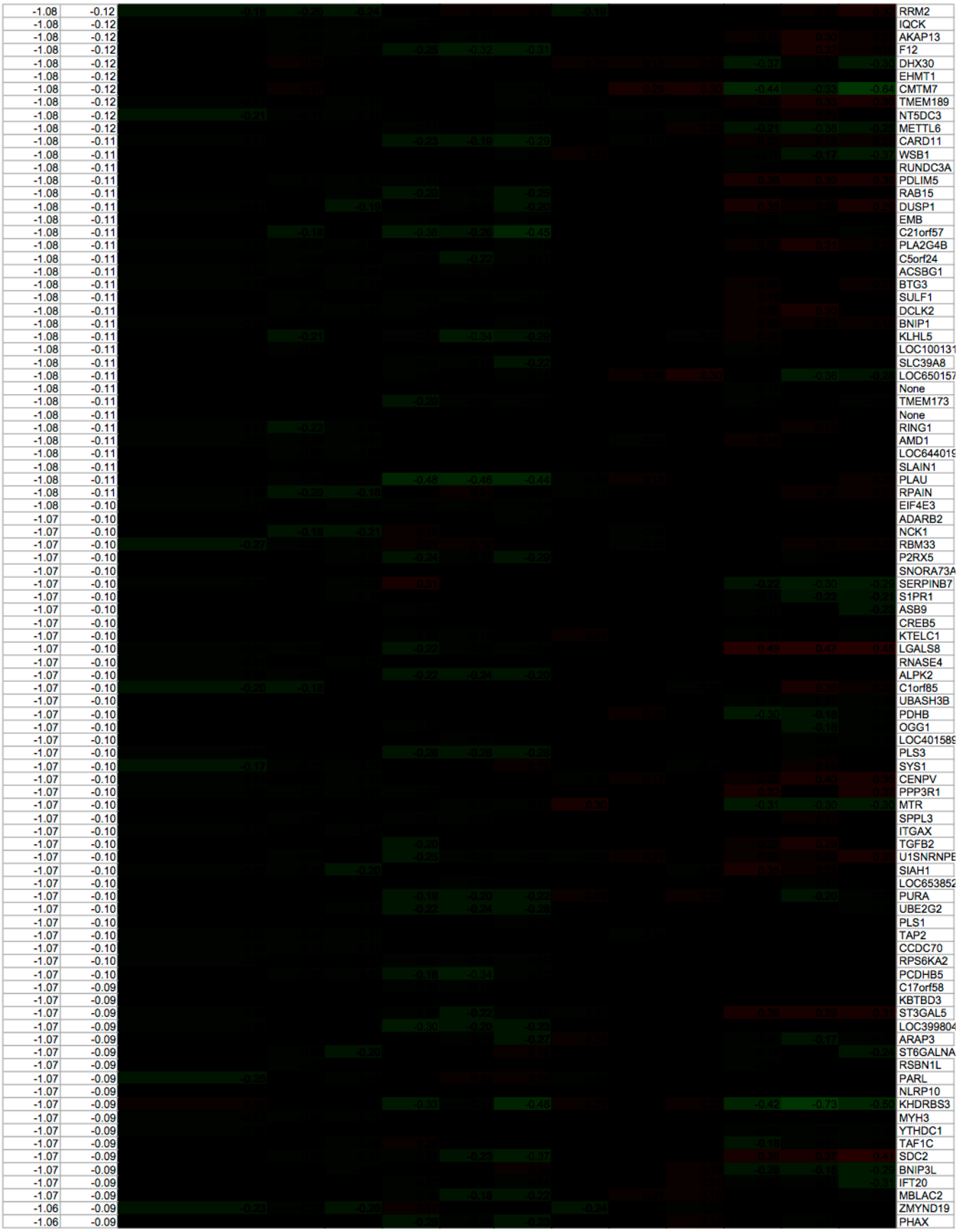
-1.18	-0.24	SLC6A9
-1.18	-0.24	MCART1
-1.18	-0.24	COPB2
-1.18	-0.24	PIK3CB
-1.18	-0.24	HLA-C
-1.18	-0.24	WDR21A
-1.18	-0.24	CAP2
-1.18	-0.24	TMEM151A
-1.18	-0.24	C9orf45
-1.18	-0.24	PCDHB3
-1.18	-0.24	MIR630
-1.18	-0.24	SNAI2
-1.18	-0.23	TPM1
-1.18	-0.23	FUK
-1.18	-0.23	FHL2
-1.18	-0.23	COL5A2
-1.18	-0.23	NOVA1
-1.18	-0.23	ATAD1
-1.18	-0.23	CHST10
-1.18	-0.23	DYX1C1
-1.18	-0.23	HCG18
-1.18	-0.23	OLFML2A
-1.18	-0.23	SMAD3
-1.17	-0.23	SLTM
-1.17	-0.23	LOC100133
-1.17	-0.23	None
-1.17	-0.23	LOC654189
-1.17	-0.23	FOXM1
-1.17	-0.23	RPP25
-1.17	-0.23	LOC642987
-1.17	-0.23	FAT1
-1.17	-0.23	SPEF2
-1.17	-0.23	COLEC12
-1.17	-0.23	ACP2
-1.17	-0.23	CKB
-1.17	-0.23	HPS1
-1.17	-0.23	LARGE
-1.17	-0.23	PROS1
-1.17	-0.23	ANKRD10
-1.17	-0.23	ENAH
-1.17	-0.23	LOC644464
-1.17	-0.23	TMED7
-1.17	-0.23	UQCRC1
-1.17	-0.23	DUT
-1.17	-0.23	LOC731878
-1.17	-0.23	LOC730994
-1.17	-0.23	LCLAT1
-1.17	-0.23	MAT2A
-1.17	-0.23	PPP2R2C
-1.17	-0.23	TSSC1
-1.17	-0.23	ARHGEF5
-1.17	-0.23	PDXK
-1.17	-0.23	PPM1H
-1.17	-0.23	CENPK
-1.17	-0.23	C14orf45
-1.17	-0.22	WRB
-1.17	-0.22	DCAF6
-1.17	-0.22	MRPS11
-1.17	-0.22	SHC3
-1.17	-0.22	PPP3CA
-1.17	-0.22	LOC650803
-1.17	-0.22	PTGFRN
-1.17	-0.22	LOC642946
-1.17	-0.22	LOC440926
-1.17	-0.22	ALDH4A1
-1.17	-0.22	BCL10
-1.17	-0.22	LDHA
-1.17	-0.22	CASC4
-1.17	-0.22	LOC388654
-1.17	-0.22	MAP3K1
-1.16	-0.22	ATP8B1
-1.16	-0.22	CAMK1
-1.16	-0.22	GMPR2
-1.16	-0.22	HIVEP1
-1.16	-0.22	NUMA1
-1.16	-0.22	TRIM7
-1.16	-0.22	SECISBP2L
-1.16	-0.22	KIAA1285
-1.16	-0.22	FERMT2
-1.16	-0.22	ERBB3
-1.16	-0.22	DHX15
-1.16	-0.22	ZNF621
-1.16	-0.22	CHRNA5
-1.16	-0.22	YTHDC1
-1.16	-0.22	B2M
-1.16	-0.22	AP4E1
-1.16	-0.22	LOC100129
-1.16	-0.22	PIK3R4
-1.16	-0.22	CAPSL
-1.16	-0.22	CENPJ
-1.16	-0.22	STAU1
-1.16	-0.21	LOC650369
-1.16	-0.21	CCRK
-1.16	-0.21	CSPG4

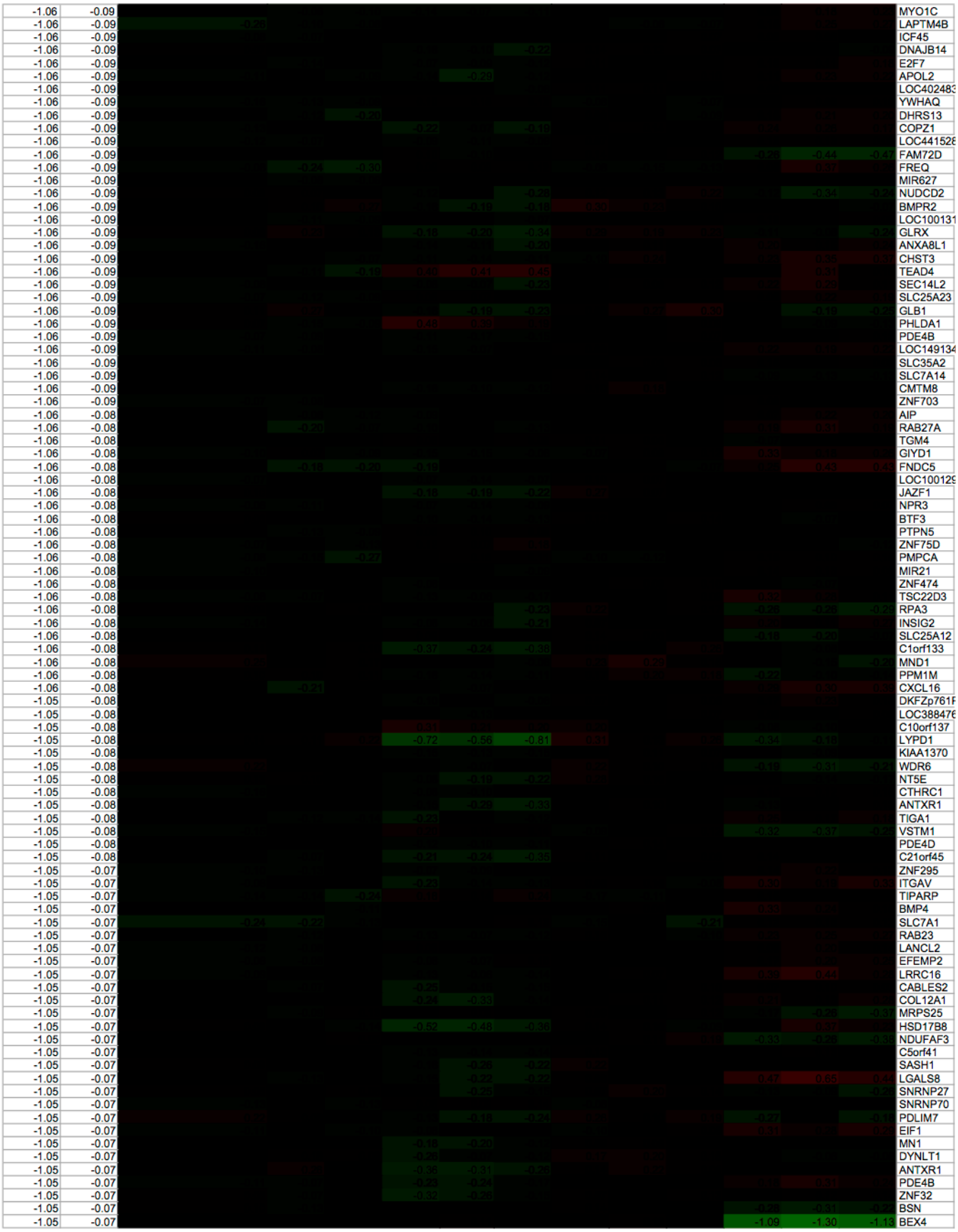




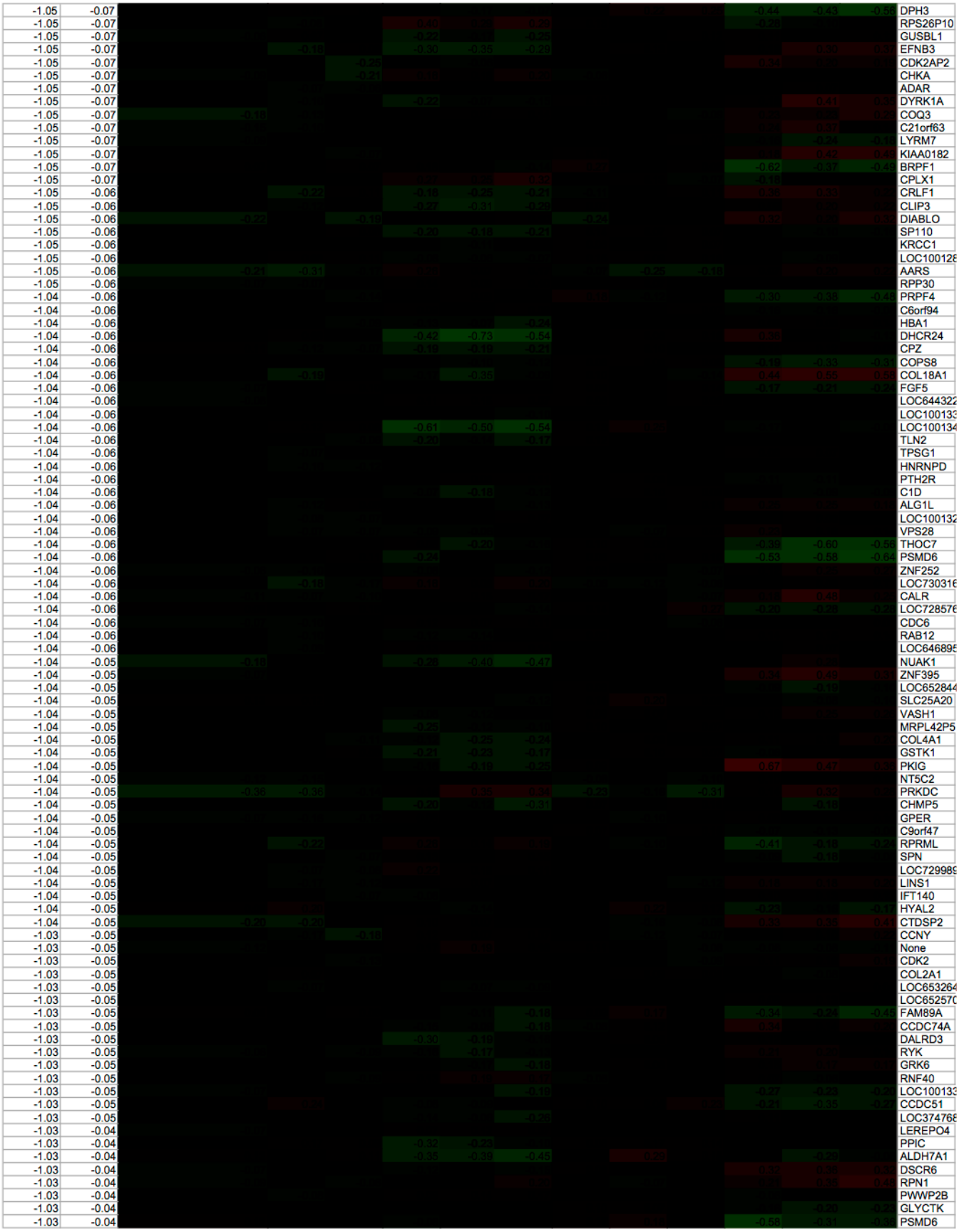

-1.12 -0.16 TOR1B
-1.12 -0.16 COL8A1
-1.12 -0.16 ADARB1
-1.12 -0.16 OPA1
-1.12 -0.16 PCCA
-1.12 -0.16 LOC65308(
-1.12 -0.16 MYLIP
-1.12 -0.16 None
-1.12 -0.16 XCL2
-1.12 -0.16 CRYZL1
-1.12 -0.16 EFNA1
-1.12 -0.16 LOC345041
-1.12 -0.16 NPPB
-1.12 -0.16 BMP4
-1.12 -0.16 NUDT21
-1.12 -0.16 DLK2
-1.12 -0.16 None
-1.12 -0.16 LOC646900
-1.12 -0.16 PRKCE
-1.12 -0.16 DAB2
-1.12 -0.16 AMOT
-1.12 -0.16 C21orf81
-1.12 -0.16 ANKRD20A
-1.12 -0.16 LOC100129
-1.12 -0.16 DKK3
-1.12 -0.16 ASF1A
-1.12 -0.16 PWP1
-1.12 -0.16 SLCO3A1
-1.12 -0.16 TRIM44
-1.12 -0.16 EEF1B2
-1.12 -0.16 WDR76
-1.12 -0.16 SHROOM3
-1.11 -0.16 ANXA1
-1.11 -0.16 LOC100133
-1.11 -0.16 GAS7
-1.11 -0.15 ERI3
-1.11 -0.15 SDCBP
-1.11 -0.15 PTCD2
-1.11 -0.15 COMMD10
-1.11 -0.15 NUP88
-1.11 -0.15 LOC100131
-1.11 -0.15 TSEN2
-1.11 -0.15 FANCC
-1.11 -0.15 MGC72104
-1.11 -0.15 KITLG
-1.11 -0.15 POLR2H
-1.11 -0.15 LAG3
-1.11 -0.15 LOC653158
-1.11 -0.15 PCID2
-1.11 -0.15 None
-1.11 -0.15 RPS17
-1.11 -0.15 GOLGA3
-1.11 -0.15 KIAA1545
-1.11 -0.15 TRNT1
-1.11 -0.15 IFT57
-1.11 -0.15 PDE4A
-1.11 -0.15 MME
-1.11 -0.15 ZNF322A
-1.11 -0.15 SLC30A9
-1.11 -0.15 HSPA4
-1.11 -0.15 TFDP1
-1.11 -0.15 SH3BGRL
-1.11 -0.15 PUS1
-1.11 -0.15 HDAC5
-1.11 -0.15 LOC649548
-1.11 -0.15 GIN1
-1.11 -0.15 LOC652183
-1.11 -0.15 KCTD6
-1.11 -0.15 ABCA5
-1.11 -0.15 DNAJC2
-1.11 -0.15 ZNF346
-1.11 -0.15 EAF1
-1.11 -0.15 PRKX
-1.11 -0.14 POLR1E
-1.11 -0.14 SAT2
-1.11 -0.14 HOXA10
-1.11 -0.14 ABLIM1
-1.10 -0.14 ERGIC2
-1.10 -0.14 ST5
-1.10 -0.14 LOC339843
-1.10 -0.14 AK3L1
-1.10 -0.14 LOC653103
-1.10 -0.14 RTN4IP1
-1.10 -0.14 KIAA0564
-1.10 -0.14 EPAS1
-1.10 -0.14 PCBP2
-1.10 -0.14 TET1
-1.10 -0.14 None
-1.10 -0.14 PCK2
-1.10 -0.14 LOC730246
-1.10 -0.14 OXTR
-1.10 -0.14 JAG2
-1.10 -0.14 TMEM206
-1.10 -0.14 TSPAN9

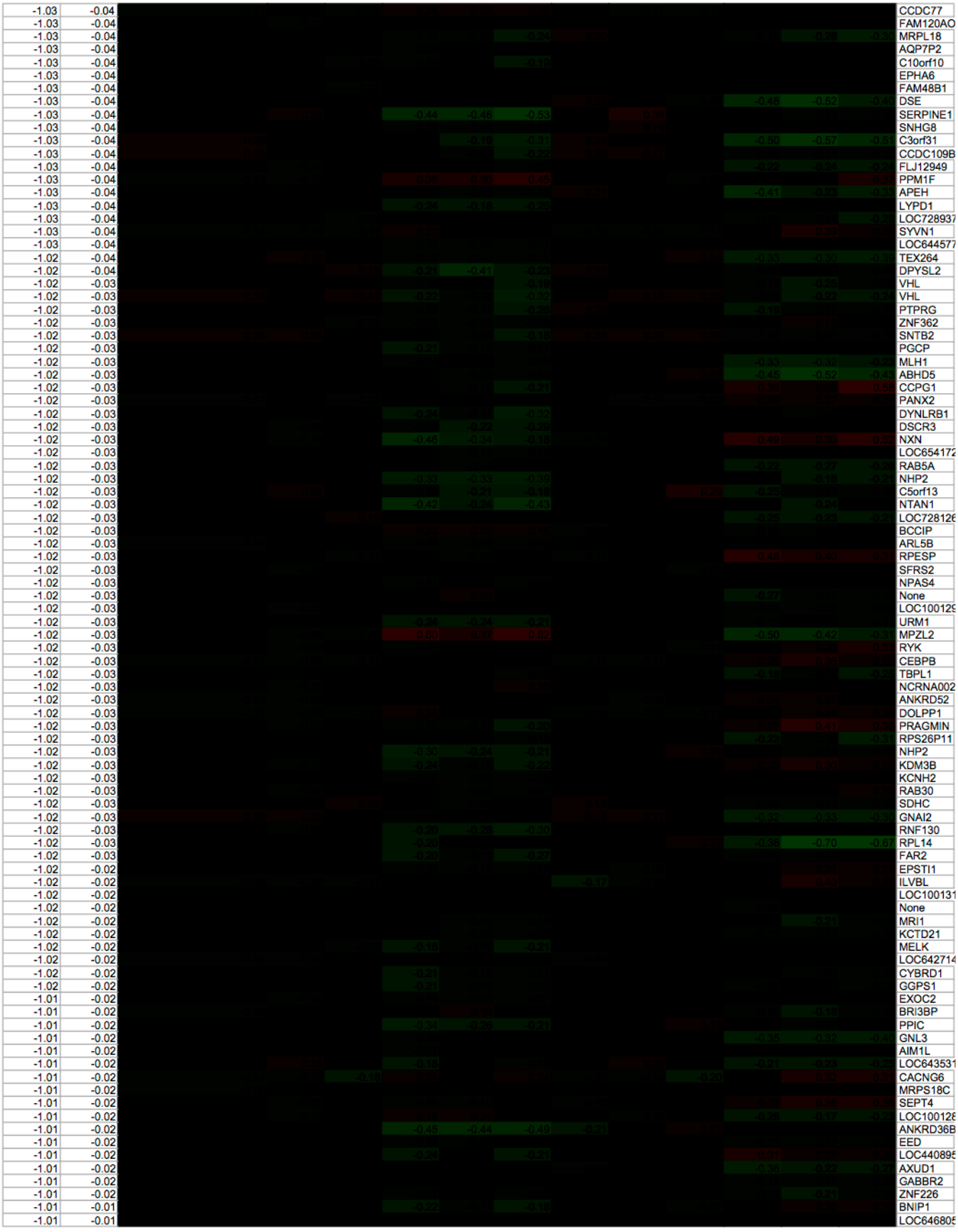


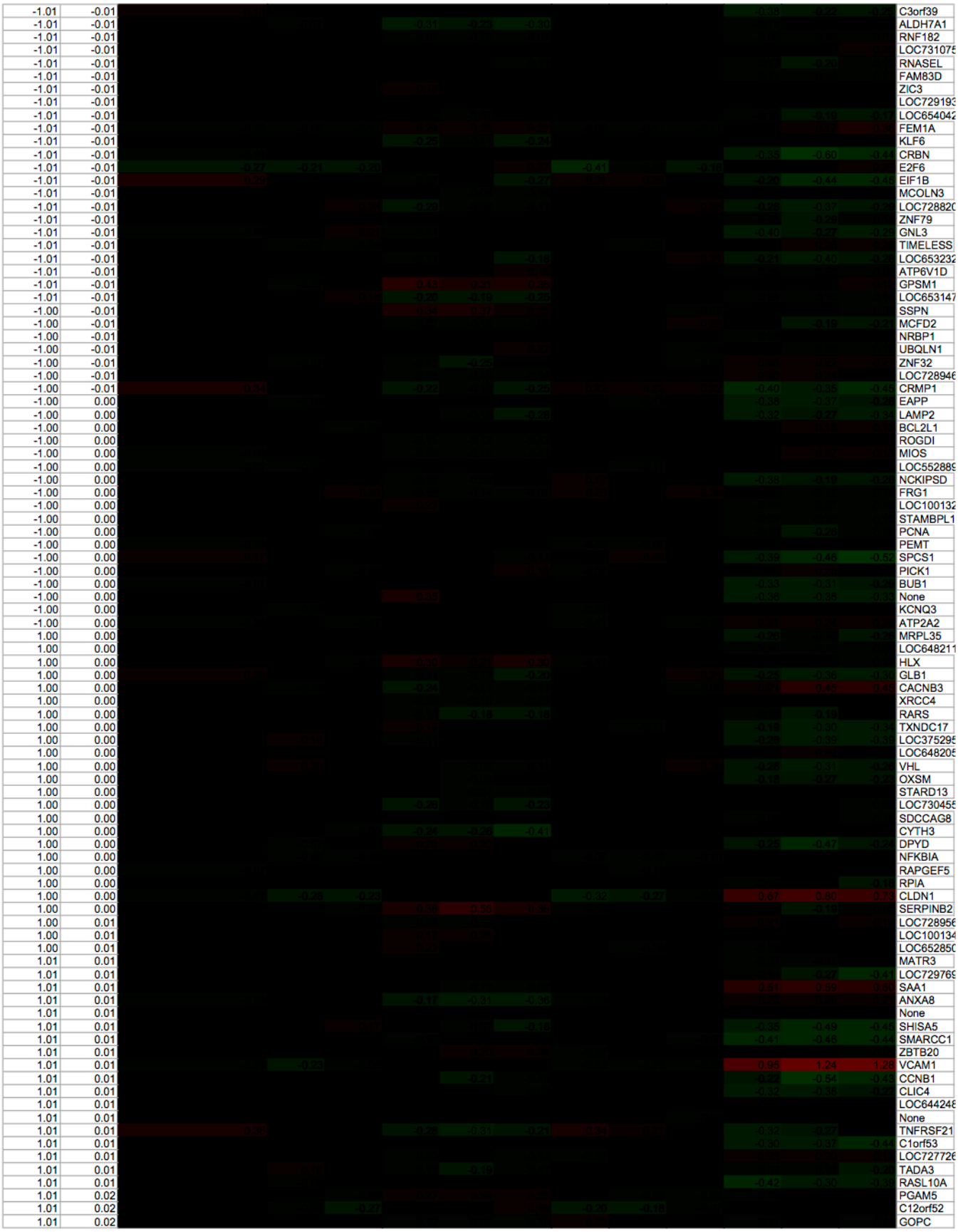




-1.06	-0.09	MYO1C
-1.06	-0.09	LAPTM4B
-1.06	-0.09	ICF45
-1.06	-0.09	DNAJB14
-1.06	-0.09	E2F7
-1.06	-0.09	APOL2
-1.06	-0.09	LOC402483
-1.06	-0.09	YWHAQ
-1.06	-0.09	DHRS13
-1.06	-0.09	COPZ1
-1.06	-0.09	LOC441528
-1.06	-0.09	FAM72D
-1.06	-0.09	FREQ
-1.06	-0.09	MIR627
-1.06	-0.09	NUDCD2
-1.06	-0.09	BMPR2
-1.06	-0.09	LOC100131
-1.06	-0.09	GLRX
-1.06	-0.09	ANXA8L1
-1.06	-0.09	CHST3
-1.06	-0.09	TEAD4
-1.06	-0.09	SEC14L2
-1.06	-0.09	SLC25A23
-1.06	-0.09	GLB1
-1.06	-0.09	PHLDA1
-1.06	-0.09	PDE4B
-1.06	-0.09	LOC149134
-1.06	-0.09	SLC35A2
-1.06	-0.09	SLC7A14
-1.06	-0.09	CMTM8
-1.06	-0.09	ZNF703
-1.06	-0.08	AIP
-1.06	-0.08	RAB27A
-1.06	-0.08	TGM4
-1.06	-0.08	GIYD1
-1.06	-0.08	FNDC5
-1.06	-0.08	LOC100129
-1.06	-0.08	JAZF1
-1.06	-0.08	NPR3
-1.06	-0.08	BTF3
-1.06	-0.08	PTPN5
-1.06	-0.08	ZNF75D
-1.06	-0.08	PMPCA
-1.06	-0.08	MIR21
-1.06	-0.08	ZNF474
-1.06	-0.08	TSC22D3
-1.06	-0.08	RPA3
-1.06	-0.08	INSIG2
-1.06	-0.08	SLC25A12
-1.06	-0.08	C1orf133
-1.06	-0.08	MND1
-1.06	-0.08	PPM1M
-1.06	-0.08	CXCL16
-1.05	-0.08	DKFZp761F
-1.05	-0.08	LOC388476
-1.05	-0.08	C10orf137
-1.05	-0.08	LYPD1
-1.05	-0.08	KIAA1370
-1.05	-0.08	WDR6
-1.05	-0.08	NT5E
-1.05	-0.08	CTHRC1
-1.05	-0.08	ANTXR1
-1.05	-0.08	TIGA1
-1.05	-0.08	VSTM1
-1.05	-0.08	PDE4D
-1.05	-0.08	C21orf45
-1.05	-0.07	ZNF295
-1.05	-0.07	ITGAV
-1.05	-0.07	TIPARP
-1.05	-0.07	BMP4
-1.05	-0.07	SLC7A1
-1.05	-0.07	RAB23
-1.05	-0.07	LANCL2
-1.05	-0.07	EFEMP2
-1.05	-0.07	LRRC16
-1.05	-0.07	CABLES2
-1.05	-0.07	COL12A1
-1.05	-0.07	MRPS25
-1.05	-0.07	HSD17B8
-1.05	-0.07	NDUFAF3
-1.05	-0.07	C5orf41
-1.05	-0.07	SASH1
-1.05	-0.07	LGALS8
-1.05	-0.07	SNRNP27
-1.05	-0.07	SNRNP70
-1.05	-0.07	PDLIM7
-1.05	-0.07	EIF1
-1.05	-0.07	MN1
-1.05	-0.07	DYNLT1
-1.05	-0.07	ANTXR1
-1.05	-0.07	PDE4B
-1.05	-0.07	ZNF32
-1.05	-0.07	BSN
-1.05	-0.07	BEX4

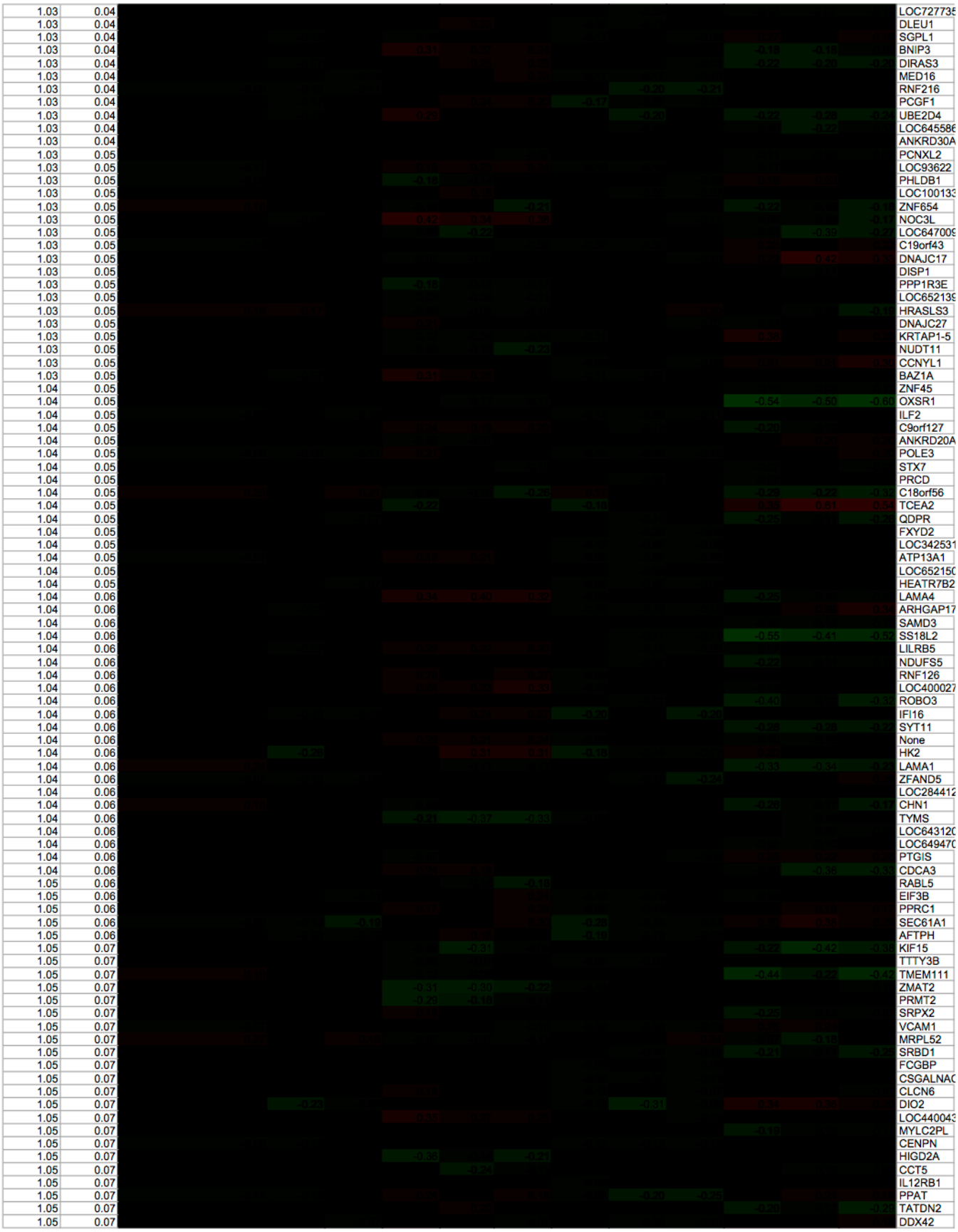






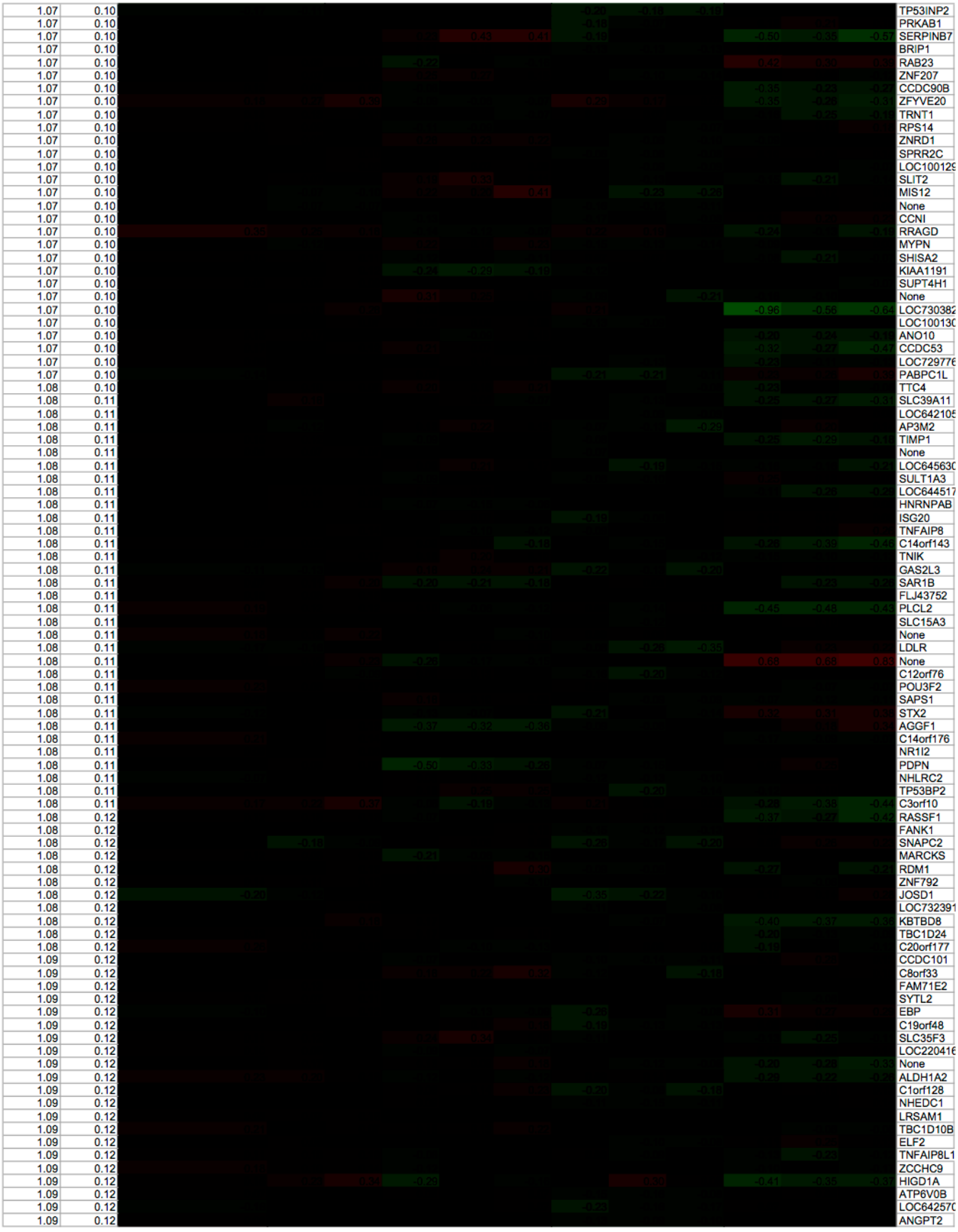
-1.01	-0.01	C3orf39
-1.01	-0.01	ALDH7A1
-1.01	-0.01	RNF182
-1.01	-0.01	LOC731075
-1.01	-0.01	RNASEL
-1.01	-0.01	FAM83D
-1.01	-0.01	ZIC3
-1.01	-0.01	LOC729193
-1.01	-0.01	LOC654042
-1.01	-0.01	FEM1A
-1.01	-0.01	KLF6
-1.01	-0.01	CRBN
-1.01	-0.01	E2F6
-1.01	-0.01	EIF1B
-1.01	-0.01	MCOLN3
-1.01	-0.01	LOC728820
-1.01	-0.01	ZNF79
-1.01	-0.01	GNL3
-1.01	-0.01	TIMELESS
-1.01	-0.01	LOC653232
-1.01	-0.01	ATP6V1D
-1.01	-0.01	GPSM1
-1.01	-0.01	LOC653147
-1.00	-0.01	SSPN
-1.00	-0.01	MCFD2
-1.00	-0.01	NRBP1
-1.00	-0.01	UBQLN1
-1.00	-0.01	ZNF32
-1.00	-0.01	LOC728946
-1.00	-0.01	CRMP1
-1.00	0.00	EAPP
-1.00	0.00	LAMP2
-1.00	0.00	BCL2L1
-1.00	0.00	ROGDI
-1.00	0.00	MIOS
-1.00	0.00	LOC552889
-1.00	0.00	NCKIPSD
-1.00	0.00	FRG1
-1.00	0.00	LOC100132
-1.00	0.00	STAMBPL1
-1.00	0.00	PCNA
-1.00	0.00	PEMT
-1.00	0.00	SPCS1
-1.00	0.00	PICK1
-1.00	0.00	BUB1
-1.00	0.00	None
-1.00	0.00	KCNQ3
-1.00	0.00	ATP2A2
1.00	0.00	MRPL35
1.00	0.00	LOC648211
1.00	0.00	HLX
1.00	0.00	GLB1
1.00	0.00	CACNB3
1.00	0.00	XRCC4
1.00	0.00	RARS
1.00	0.00	TXNDC17
1.00	0.00	LOC375295
1.00	0.00	LOC648205
1.00	0.00	VHL
1.00	0.00	OXSM
1.00	0.00	STARD13
1.00	0.00	LOC730455
1.00	0.00	SDCCAG8
1.00	0.00	CYTH3
1.00	0.00	DPYD
1.00	0.00	NFKBIA
1.00	0.00	RAPGEF5
1.00	0.00	RPIA
1.00	0.00	CLDN1
1.00	0.00	SERPINB2
1.00	0.01	LOC728956
1.00	0.01	LOC100134
1.00	0.01	LOC652850
1.01	0.01	MATR3
1.01	0.01	LOC729769
1.01	0.01	SAA1
1.01	0.01	ANXA8
1.01	0.01	None
1.01	0.01	SHISA5
1.01	0.01	SMARCC1
1.01	0.01	ZBTB20
1.01	0.01	VCAM1
1.01	0.01	CCNB1
1.01	0.01	CLIC4
1.01	0.01	LOC644248
1.01	0.01	None
1.01	0.01	TNFRSF21
1.01	0.01	C1orf53
1.01	0.01	LOC727726
1.01	0.01	TADA3
1.01	0.01	RASL10A
1.01	0.02	PGAM5
1.01	0.02	C12orf52
1.01	0.02	GOPC

1.01	0.02						-0.18							TINP1
1.01	0.02						-0.22					-0.23		LOC647856
1.01	0.02													LOC440829
1.01	0.02					0.37	0.30							ARG2
1.01	0.02										-0.31	-0.21	-0.24	UQCC
1.01	0.02													ZNHIT6
1.01	0.02				-0.21									DCLK2
1.01	0.02											0.41		HAUS4
1.01	0.02													LOC100130
1.01	0.02													LOC642504
1.01	0.02											-0.37	-0.28	PSMD14
1.01	0.02						-0.18				-0.41	-0.52	-0.55	DPH3
1.01	0.02										-0.32	-0.23	-0.33	PIGY
1.01	0.02										-0.54	-0.23	-0.39	GPD1L
1.01	0.02													SRF
1.01	0.02										-0.42	-0.41	-0.35	IP6K1
1.01	0.02													CYP2S1
1.01	0.02						-0.19				-0.36	-0.60	-0.60	BUB3
1.01	0.02					-0.24	-0.26				-0.19	-0.27	-0.24	PLEK2
1.01	0.02				-0.18	-0.25	-0.21							FBXW11
1.01	0.02				-0.19		-0.22				-0.39	-0.45	-0.36	SEC22C
1.01	0.02													ERP29
1.01	0.02													LOC650832
1.01	0.02				-0.53	-0.22	-0.18						0.47	LOC652846
1.01	0.02													GATAD2A
1.02	0.02	-0.18	-0.20			0.34		-0.19	-0.20					PFKM
1.02	0.02													LOC391859
1.02	0.02										-0.40	-0.28	-0.30	SLC4A7
1.02	0.02				-0.24	-0.32	-0.34				-0.25			ADAM19
1.02	0.02				-0.20						-0.21	-0.30	-0.42	C14orf142
1.02	0.02													WDR23
1.02	0.02				-0.32	-0.34	-0.26							C21orf33
1.02	0.02												0.44	WDR41
1.02	0.02						-0.23							ALG8
1.02	0.02													ITPR2
1.02	0.02													WDR12
1.02	0.03													CENPM
1.02	0.03			-0.23					-0.22	-0.21				PHF17
1.02	0.03				-0.39	-0.35	-0.55							FAT1
1.02	0.03													NLE1
1.02	0.03			-0.19				-0.30						MTHFD2
1.02	0.03													PLXND1
1.02	0.03				-0.19									C11orf75
1.02	0.03													COG8
1.02	0.03				-0.41	-0.19	-0.30					-0.24	-0.37	LOC653874
1.02	0.03				0.52	0.56	0.65							NCOA6IP
1.02	0.03								-0.34		-0.58	-0.42	-0.43	LOC729646
1.02	0.03													PNKD
1.02	0.03													PCOLCE2
1.02	0.03										-0.18			GGT3
1.02	0.03													SNORA16A
1.02	0.03										-0.30	-0.65	-0.43	LOC728285
1.02	0.03										-0.33	-0.28	-0.35	NICN1
1.02	0.03													GGT5
1.02	0.03										-0.31			SPHK2
1.02	0.03						0.31							SEH1L
1.02	0.03										-0.18	-0.25	-0.35	LOC100129
1.02	0.03													PLRG1
1.02	0.03				-0.20	-0.19	-0.21							MPST
1.02	0.03													ADAM23
1.02	0.03													HIST1H2BC
1.02	0.03													RNF44
1.02	0.03							-0.23		-0.23				EIF4H
1.02	0.03										0.50	0.58	0.54	ECHDC2
1.02	0.03													CFB
1.02	0.03				-0.23									HES1
1.02	0.03										-0.27		-0.29	CPNE8
1.02	0.03				0.31	0.33	0.32		-0.16					SLC25A19
1.02	0.03													ATP6V0C
1.02	0.03													FAM90A6P
1.02	0.03					0.50								SERPINB2
1.02	0.03										-0.40	-0.59	-0.48	FAM162A
1.02	0.04										-0.18			PODXL
1.03	0.04													FTH1
1.03	0.04													DBN1
1.03	0.04						-0.24				-0.35		-0.39	KLHDC8B
1.03	0.04												-0.21	CCDC34
1.03	0.04					0.39	0.34							SPRYD4
1.03	0.04													NR0B1
1.03	0.04													DAZAP2
1.03	0.04													NR2E3
1.03	0.04										-0.54	-0.53	-0.59	CPNE8
1.03	0.04													LOC729639
1.03	0.04										-0.26		-0.21	SERPINA1
1.03	0.04													LOC100130
1.03	0.04													AASDH
1.03	0.04				-0.20									LOC100131
1.03	0.04					-0.24	-0.18							ATXN1
1.03	0.04							-0.28			0.49			GABARAPL
1.03	0.04				-0.30	-0.21	-0.25							PLCG1
1.03	0.04													SEPT12
1.03	0.04											-0.23	-0.17	ZNF654
1.03	0.04										0.41			CCNYL1
1.03	0.04										-0.75	-0.51	-0.65	LOC440957

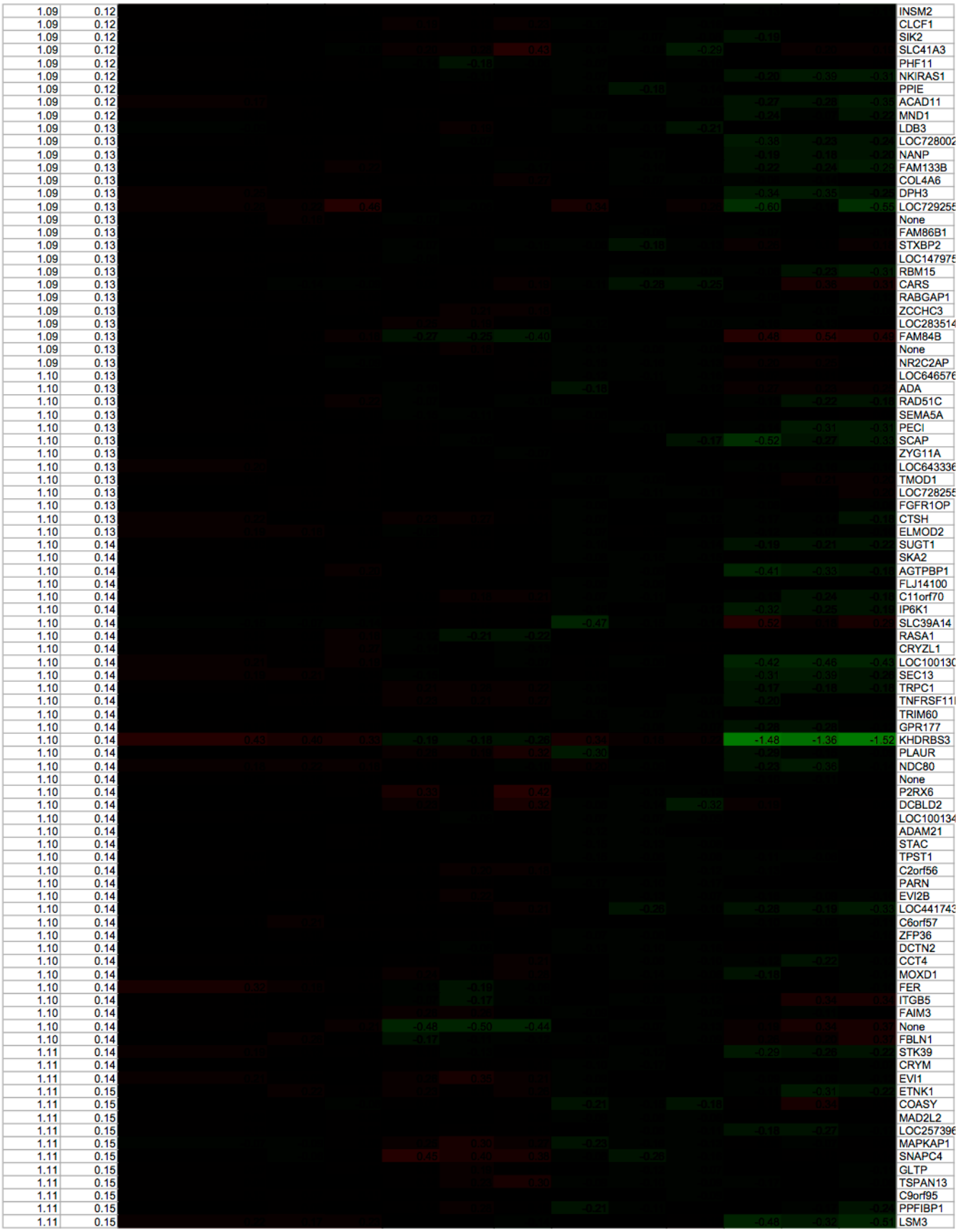


1.03	0.04	LOC727735
1.03	0.04	DLEU1
1.03	0.04	SGPL1
1.03	0.04	BNIP3
1.03	0.04	DIRAS3
1.03	0.04	MED16
1.03	0.04	RNF216
1.03	0.04	PCGF1
1.03	0.04	UBE2D4
1.03	0.04	LOC645586
1.03	0.04	ANKRD30A
1.03	0.05	PCNXL2
1.03	0.05	LOC93622
1.03	0.05	PHLDB1
1.03	0.05	LOC100133
1.03	0.05	ZNF654
1.03	0.05	NOC3L
1.03	0.05	LOC647009
1.03	0.05	C19orf43
1.03	0.05	DNAJC17
1.03	0.05	DISP1
1.03	0.05	PPP1R3E
1.03	0.05	LOC652139
1.03	0.05	HRASLS3
1.03	0.05	DNAJC27
1.03	0.05	KRTAP1-5
1.03	0.05	NUDT11
1.03	0.05	CCNYL1
1.03	0.05	BAZ1A
1.04	0.05	ZNF45
1.04	0.05	OXSR1
1.04	0.05	ILF2
1.04	0.05	C9orf127
1.04	0.05	ANKRD20A
1.04	0.05	POLE3
1.04	0.05	STX7
1.04	0.05	PRCD
1.04	0.05	C18orf56
1.04	0.05	TCEA2
1.04	0.05	QDPR
1.04	0.05	FXYD2
1.04	0.05	LOC342531
1.04	0.05	ATP13A1
1.04	0.05	LOC652150
1.04	0.05	HEATR7B2
1.04	0.06	LAMA4
1.04	0.06	ARHGAP17
1.04	0.06	SAMD3
1.04	0.06	SS18L2
1.04	0.06	LILRB5
1.04	0.06	NDUFS5
1.04	0.06	RNF126
1.04	0.06	LOC400027
1.04	0.06	ROBO3
1.04	0.06	IFI16
1.04	0.06	SYT11
1.04	0.06	None
1.04	0.06	HK2
1.04	0.06	LAMA1
1.04	0.06	ZFAND5
1.04	0.06	LOC284412
1.04	0.06	CHN1
1.04	0.06	TYMS
1.04	0.06	LOC643120
1.04	0.06	LOC649470
1.04	0.06	PTGIS
1.04	0.06	CDCA3
1.05	0.06	RABL5
1.05	0.06	EIF3B
1.05	0.06	PPRC1
1.05	0.06	SEC61A1
1.05	0.06	AFTPH
1.05	0.07	KIF15
1.05	0.07	TTTY3B
1.05	0.07	TMEM111
1.05	0.07	ZMAT2
1.05	0.07	PRMT2
1.05	0.07	SRPX2
1.05	0.07	VCAM1
1.05	0.07	MRPL52
1.05	0.07	SRBD1
1.05	0.07	FCGBP
1.05	0.07	CSGALNAC
1.05	0.07	CLCN6
1.05	0.07	DIO2
1.05	0.07	LOC440043
1.05	0.07	MYLC2PL
1.05	0.07	CENPN
1.05	0.07	HIGD2A
1.05	0.07	CCT5
1.05	0.07	IL12RB1
1.05	0.07	PPAT
1.05	0.07	TATDN2
1.05	0.07	DDX42

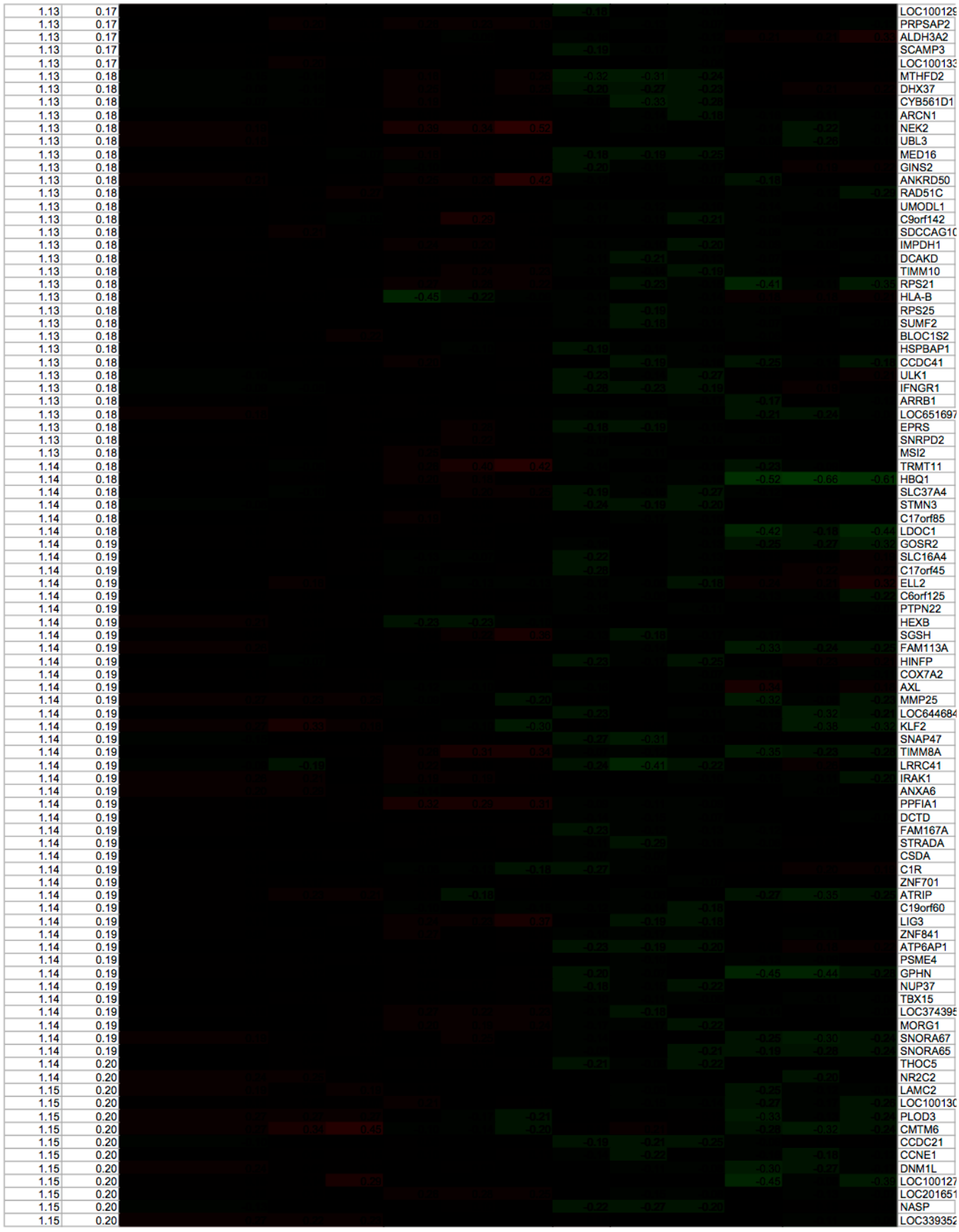
1.05	0.07										-0.47	-0.49	-0.50	CMTM7
1.05	0.07													GPHB5
1.05	0.07										-0.34	-0.23	-0.24	SCARNA13
1.05	0.07													TMEM71
1.05	0.07													MSH5
1.05	0.07										-0.40		-0.32	SETD2
1.05	0.07					0.31	0.33							ULBP1
1.05	0.07													FLJ36070
1.05	0.07										-0.29	-0.39	-0.44	LOC644907
1.05	0.07								-0.29					WDR90
1.05	0.08				-0.51	-0.29	-0.35							PRR16
1.05	0.08													QSER1
1.05	0.08													None
1.05	0.08						0.39							RRP7A
1.05	0.08												-0.20	TMEM136
1.05	0.08										-0.31	-0.26	-0.28	ATRIP
1.05	0.08										-0.31	-0.19		KRTAP2-1
1.05	0.08				0.35	0.38								TFRC
1.05	0.08													LOC651635
1.06	0.08				-0.40						-0.37	-0.41	-0.29	CRTAP
1.06	0.08										-0.26	-0.19	-0.24	ACAA1
1.06	0.08								-0.19		-0.20	-0.20	-0.24	MLF1
1.06	0.08													PAN3
1.06	0.08										0.42	0.30	0.33	CTHRC1
1.06	0.08				-0.21	-0.28	-0.31							FXYD6
1.06	0.08											-0.19		ZNF697
1.06	0.08													ANKRD45
1.06	0.08										-0.34	-0.23	-0.32	RPUSD3
1.06	0.08													LOC731314
1.06	0.08											-0.30	-0.27	TMF1
1.06	0.08													C1orf150
1.06	0.08				-0.22	-0.33	-0.27							SESTD1
1.06	0.08							-0.20						SIX5
1.06	0.08					-0.19					-1.24	-1.25	-1.23	C3orf14
1.06	0.08				0.30	0.29	0.42		-0.25		-0.18			MMP3
1.06	0.08					-0.22	-0.21							IFNGR2
1.06	0.08										-0.28	-0.21		HYAL3
1.06	0.08													LOC654069
1.06	0.08										-0.18			INPP5D
1.06	0.08						-0.17					0.45	0.42	TRNP1
1.06	0.08													LOC642516
1.06	0.08		0.32								-0.28	-0.20		RASSF1
1.06	0.08										-0.19			TSGA14
1.06	0.08													LOC100133
1.06	0.08													LOC541469
1.06	0.08													LOXL3
1.06	0.08													LOC644413
1.06	0.08		-0.45	-0.34	0.44	0.51	0.33	-0.34	-0.31	-0.33				CCL20
1.06	0.08							-0.17						LSM4
1.06	0.08					0.34	0.46	-0.18			-0.41		-0.30	TMEM98
1.06	0.08										-0.38	-0.34	-0.38	SNRK
1.06	0.08											-0.19	-0.20	SOX7
1.06	0.08													CDC26
1.06	0.08													LOC201175
1.06	0.09											-0.27	-0.24	SMYD4
1.06	0.09										-0.21			ALG8
1.06	0.09													KHDC1L
1.06	0.09										-0.35	-0.29	-0.27	LOC646672
1.06	0.09													HOXC8
1.06	0.09													SNORD56
1.06	0.09										-0.33	-0.25	-0.31	PSRC1
1.06	0.09											-0.33	-0.23	NUDCD2
1.06	0.09											0.29		GSDMD
1.06	0.09													ZCCHC17
1.06	0.09													SLC13A5
1.06	0.09				-0.25						-0.38	-0.24	-0.23	ARF4
1.06	0.09													AKR1E2
1.06	0.09				0.33		0.32	-0.18		-0.24				PPP1CC
1.06	0.09				-0.51	-0.48	-0.47				0.32			GJA1
1.07	0.09											0.35		GALT
1.07	0.09											-0.20		ZC3HAV1
1.07	0.09									-0.22				PLK4
1.07	0.09													ANXA6
1.07	0.09													TSKU
1.07	0.09													KLF16
1.07	0.09											-0.31	-0.21	MYOF
1.07	0.09					-0.32	-0.28					-0.24	-0.30	SAR1B
1.07	0.09						0.35							POLR1C
1.07	0.09							-0.18						SFRS9
1.07	0.09											-0.19		KRTAP4-12
1.07	0.09													NLGN4X
1.07	0.09													REXO4
1.07	0.09							-0.27			0.32	0.34	0.41	ATP2A2
1.07	0.09													KLHL21
1.07	0.09			0.36		-0.26	-0.42				-0.51	-0.27	-0.44	MGC39900
1.07	0.09													S100A10
1.07	0.09						0.32							MTERFD1
1.07	0.09										-0.57	-0.50	-0.54	LOC645166
1.07	0.09													SYTL3
1.07	0.09													ZNF383
1.07	0.10					-0.18	-0.20							FN1
1.07	0.10													FLJ10374
1.07	0.10		-0.18					-0.36				0.56	0.61	DCI
1.07	0.10							-0.25		-0.23				C12orf41

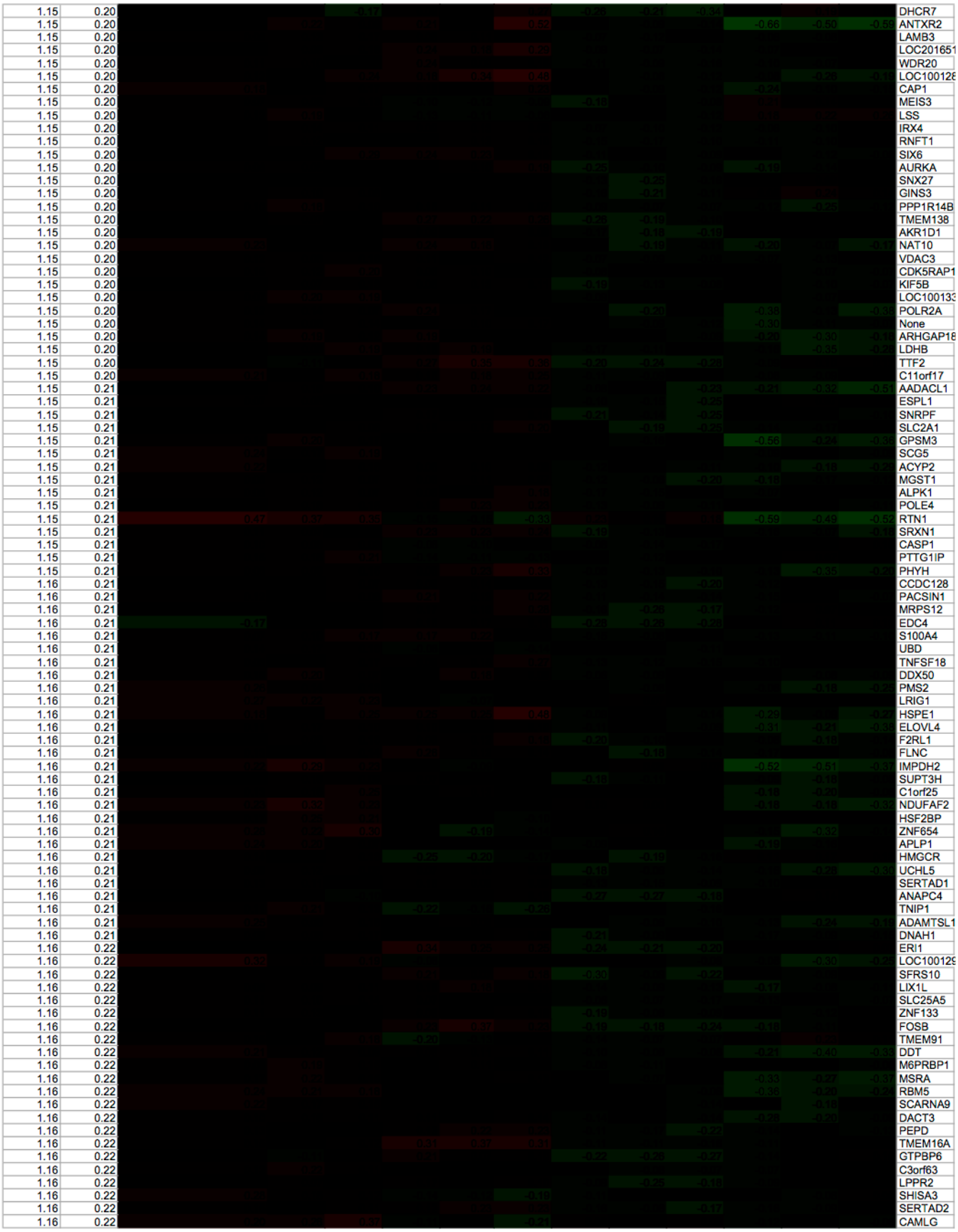


1.07	0.10	TP53INP2
1.07	0.10	PRKAB1
1.07	0.10	SERPINB7
1.07	0.10	BRIP1
1.07	0.10	RAB23
1.07	0.10	ZNF207
1.07	0.10	CCDC90B
1.07	0.10	ZFYVE20
1.07	0.10	TRNT1
1.07	0.10	RPS14
1.07	0.10	ZNRD1
1.07	0.10	SPRR2C
1.07	0.10	LOC100129
1.07	0.10	SLIT2
1.07	0.10	MIS12
1.07	0.10	None
1.07	0.10	CCNI
1.07	0.10	RRAGD
1.07	0.10	MYPN
1.07	0.10	SHISA2
1.07	0.10	KIAA1191
1.07	0.10	SUPT4H1
1.07	0.10	None
1.07	0.10	LOC730382
1.07	0.10	LOC100130
1.07	0.10	ANO10
1.07	0.10	CCDC53
1.07	0.10	LOC729776
1.07	0.10	PABPC1L
1.08	0.10	TTC4
1.08	0.11	SLC39A11
1.08	0.11	LOC642105
1.08	0.11	AP3M2
1.08	0.11	TIMP1
1.08	0.11	None
1.08	0.11	LOC645630
1.08	0.11	SULT1A3
1.08	0.11	LOC644517
1.08	0.11	HNRNPAB
1.08	0.11	ISG20
1.08	0.11	TNFAIP8
1.08	0.11	C14orf143
1.08	0.11	TNIK
1.08	0.11	GAS2L3
1.08	0.11	SAR1B
1.08	0.11	FLJ43752
1.08	0.11	PLCL2
1.08	0.11	SLC15A3
1.08	0.11	None
1.08	0.11	LDLR
1.08	0.11	None
1.08	0.11	C12orf76
1.08	0.11	POU3F2
1.08	0.11	SAPS1
1.08	0.11	STX2
1.08	0.11	AGGF1
1.08	0.11	C14orf176
1.08	0.11	NR1I2
1.08	0.11	PDPN
1.08	0.11	NHLRC2
1.08	0.11	TP53BP2
1.08	0.11	C3orf10
1.08	0.12	RASSF1
1.08	0.12	FANK1
1.08	0.12	SNAPC2
1.08	0.12	MARCKS
1.08	0.12	RDM1
1.08	0.12	ZNF792
1.08	0.12	JOSD1
1.08	0.12	LOC732391
1.08	0.12	KBTBD8
1.08	0.12	TBC1D24
1.08	0.12	C20orf177
1.09	0.12	CCDC101
1.09	0.12	C8orf33
1.09	0.12	FAM71E2
1.09	0.12	SYTL2
1.09	0.12	EBP
1.09	0.12	C19orf48
1.09	0.12	SLC35F3
1.09	0.12	LOC220416
1.09	0.12	None
1.09	0.12	ALDH1A2
1.09	0.12	C1orf128
1.09	0.12	NHEDC1
1.09	0.12	LRSAM1
1.09	0.12	TBC1D10B
1.09	0.12	ELF2
1.09	0.12	TNFAIP8L1
1.09	0.12	ZCCHC9
1.09	0.12	HIGD1A
1.09	0.12	ATP6V0B
1.09	0.12	LOC642570
1.09	0.12	ANGPT2

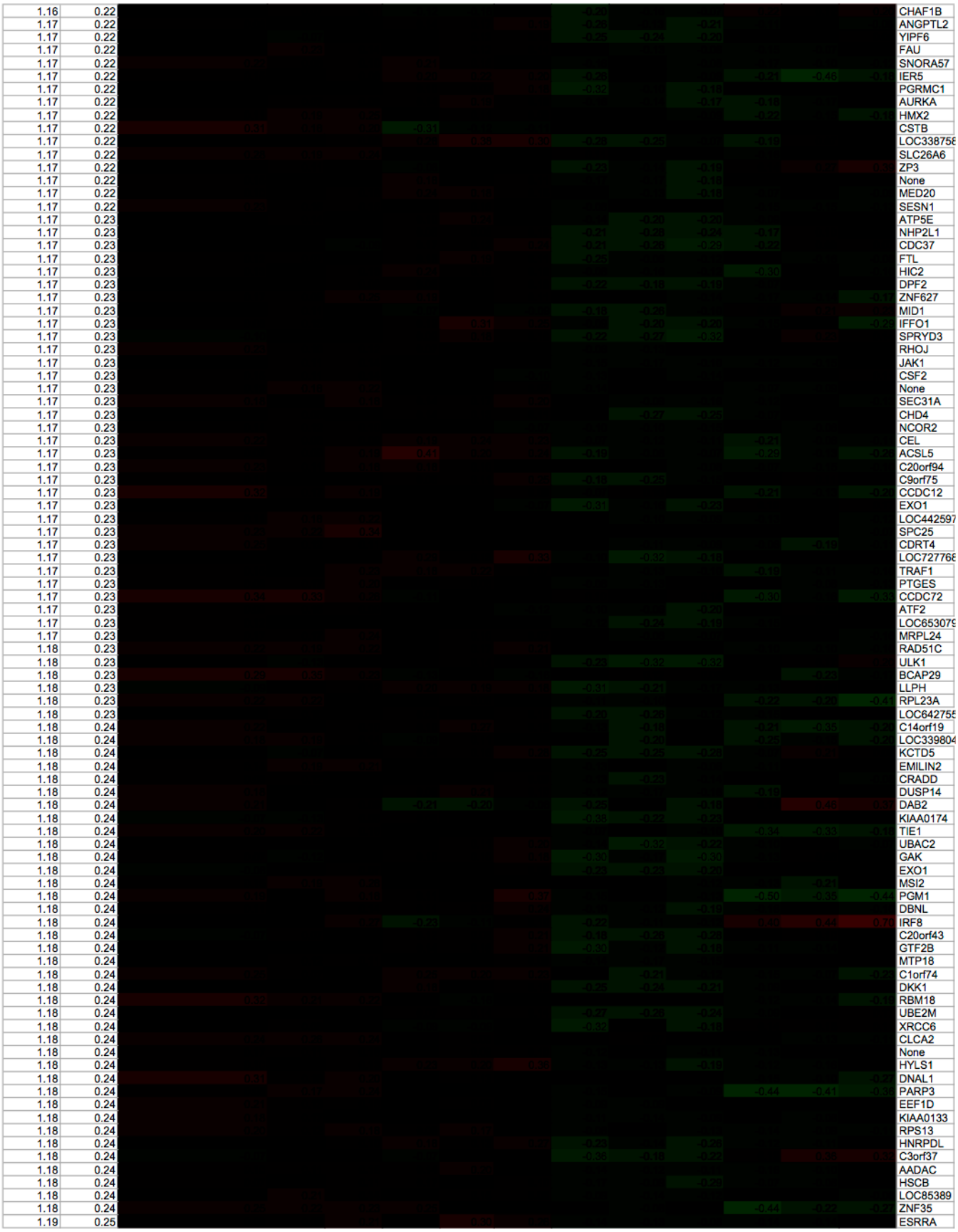


1.11	0.15	-0.07			0.43	0.43	0.47	-0.19	-0.20	-0.11		0.18	0.15	SLC25A22
1.11	0.15		-0.11	-0.11	0.08	0.17	0.18	-0.21	-0.29	-0.16		0.07		SLC30A1
1.11	0.15	0.15		0.09	0.18	0.21	0.10		-0.17	-0.07	-0.14	-0.10		ERLIN2
1.11	0.15		-0.11				0.11	-0.12	-0.25	-0.22		0.08		ASPSCR1
1.11	0.15			0.13	0.26		0.34	-0.13		-0.11	-0.33	-0.21	-0.38	FERMT1
1.11	0.15	0.09	0.10	0.13						-0.07				ST3GAL3
1.11	0.15	0.17		0.16	0.19			0.02		-0.17	-0.32	-0.34	-0.33	KDELC2
1.11	0.15	0.13	0.19	0.11			-0.15						-0.08	KRT32
1.11	0.15				0.17	0.13	0.10		-0.24	-0.11	-0.06	-0.12	-0.12	SENP5
1.11	0.15				0.27	0.10	0.18	-0.13	-0.27	-0.06	-0.06	0.08		C17orf68
1.11	0.15			-0.10		0.17	-0.09	-0.10	-0.27	-0.13	0.37	0.43	0.32	PSG4
1.11	0.15	0.10		0.14	0.23	0.14	0.10		-0.10	-0.06	-0.12		-0.13	TGFBR3
1.11	0.15			0.09	0.19	0.19	0.18		-0.19	-0.15		-0.12	-0.16	LRRC3
1.11	0.15	0.18	0.35	0.19	-0.20		-0.27	0.14		0.12			-0.12	ETV6
1.11	0.15	0.22	0.12	0.15	-0.03		-0.11				-0.09		-0.07	TBCA
1.11	0.15	0.07	0.11					-0.08	-0.11				-0.10	LOC149950
1.11	0.15	0.08	0.10		0.28	0.18	0.17	-0.11	-0.06		-0.29	-0.08	-0.24	ANGEL1
1.11	0.15	0.18	0.16	0.15		-0.08	-0.18				-0.09	-0.22	-0.10	HBEGF
1.11	0.15		0.15	0.08		0.13	0.10	-0.10	-0.06		-0.34	-0.21	-0.27	LOC678655
1.11	0.15	-0.06	-0.14		0.13	0.16	0.25	-0.29	-0.14	-0.28		0.21	0.19	C19orf12
1.11	0.15	0.28	0.16	0.34	-0.22	-0.07			0.20	0.11	-0.14	-0.50	-0.31	LOC100129
1.11	0.15							-0.28	-0.09	-0.12	0.13	0.09		RPL3
1.11	0.15	0.16	0.18	0.30		-0.14	-0.14		0.11		-0.18	-0.15	-0.16	RYBP
1.11	0.15		0.14		0.27	0.12	0.14	-0.09	-0.11					FAM100B
1.11	0.15		0.08		0.13	0.07			-0.16	-0.17	-0.38	-0.19	-0.19	MGC45800
1.11	0.15	0.15		0.23			0.11				-0.28	-0.13	-0.27	ABL2
1.11	0.15							-0.21	-0.10	-0.12	0.08		0.10	ANXA2P2
1.11	0.15	0.15	0.23	0.08			-0.09						-0.06	EPHB4
1.11	0.15	0.15	0.16	0.13			-0.21							LOC653438
1.11	0.15	0.08	0.11		-0.12		-0.12	-0.09		-0.11	0.15	0.08		TFR2
1.11	0.15	0.22	0.25	0.16	-0.08		-0.07	-0.08	0.11	0.14	-0.06		-0.08	C3orf75
1.11	0.15	0.07	0.12	0.09	0.22		0.12	-0.08	-0.06		-0.14	-0.17	-0.19	C14orf145
1.11	0.15	0.11	0.12	0.11		0.08					-0.16	-0.16	-0.16	GLDC
1.11	0.16		0.07			0.12			-0.12	-0.14	-0.16	-0.14	-0.13	TNRC6B
1.11	0.16		0.17	0.11						-0.19	-0.26	-0.40	-0.19	GPRASP2
1.11	0.16	0.10	0.10	0.10					-0.06	-0.07				SLC39A4
1.11	0.16			0.08	0.12	0.09	0.16		-0.11	-0.16	-0.15	-0.31	-0.13	NAV3
1.11	0.16	0.19	0.19	0.14	0.19						-0.26	-0.19	-0.18	SAAL1
1.11	0.16	0.12	0.08	0.09	0.12		0.07		-0.13			-0.09	-0.09	SLC9A6
1.11	0.16	0.13	0.12	0.10			0.11	-0.11			-0.39	-0.13	-0.31	TUSC4
1.11	0.16	0.23		0.13				0.14	-0.11	-0.05	-0.58	-0.53	-0.44	ARHGEF3
1.11	0.16	0.14		0.06	0.24				-0.18	-0.09	-0.43	-0.40	-0.50	NETO2
1.11	0.16	0.08	0.25	0.13			0.10			0.10	-0.35	-0.15	-0.27	C3orf75
1.12	0.16	0.05	0.24	0.24			-0.06		0.10		-0.09	-0.23	-0.21	MED6
1.12	0.16	0.08		0.10	0.28	0.18			-0.13	-0.12	-0.38	-0.45	-0.48	PRDM8
1.12	0.16	0.16	0.11	0.08		-0.06		-0.05	-0.06					None
1.12	0.16					0.12	0.02	-0.14	-0.17	-0.19				C20orf20
1.12	0.16		0.12		-0.15	-0.31	-0.28	-0.16		-0.05	0.23		0.23	C5orf32
1.12	0.16	0.07		0.12	0.12	0.15	0.10		-0.16		-0.12	-0.09	-0.07	LOC124512
1.12	0.16	0.10	0.15	0.06	0.10	0.07			-0.07	-0.08	-0.29	-0.30	-0.22	NCAPG
1.12	0.16			0.09	0.15	0.10	0.07	-0.14	-0.19	-0.05	-0.10	-0.08		TAOK2
1.12	0.16	0.20	0.30	0.19				0.13	0.07		-0.36	-0.18	-0.20	XPC
1.12	0.16		-0.08			0.10	0.15	-0.15	-0.24	-0.13			0.07	ZDHHC6
1.12	0.16	0.09	0.08		0.13	0.24	0.26			-0.15	-0.31	-0.18	-0.15	None
1.12	0.16			-0.09	0.30	0.35	0.21	-0.29	-0.28		-0.13		-0.05	None
1.12	0.16			-0.08	0.11	0.12	0.10	-0.26	-0.11	-0.17	-0.06	0.09		SEPX1
1.12	0.16	0.29			0.29	0.28	0.21	0.08	-0.25		-0.74	-0.80	-0.62	OSBPL6
1.12	0.16	-0.12			0.21	0.13	0.38	-0.14	-0.26	-0.17	-0.12	0.10		ZNF784
1.12	0.16	0.11		0.07	0.08	0.20	0.19	-0.09	-0.14		-0.13	-0.10		ZNF788
1.12	0.16	0.12			-0.18	-0.15	-0.11	-0.21		-0.12	0.27	0.42	0.53	FBLN1
1.12	0.16		0.20	0.17	0.20	0.21	0.27				-0.08		-0.10	KIAA0114
1.12	0.16	0.12	0.17	0.23	-0.08				0.08		-0.14	-0.08	-0.08	LHX3
1.12	0.16						0.11	-0.15	-0.15	-0.24	-0.06	0.08	0.10	ADAMTS4
1.12	0.16				0.08			-0.17	-0.14	-0.17		0.12		ASB11
1.12	0.16	0.26	0.10	0.14	0.09			0.08			-0.19	-0.29	-0.34	C20orf7
1.12	0.16	0.12		0.16	0.09	0.17	0.13	-0.06		-0.10	-0.25		-0.23	LOC100134
1.12	0.16	0.21	0.17	0.09					-0.16		-0.22	-0.15	-0.26	RWDD1
1.12	0.16					0.07		-0.13	-0.20	-0.14		0.10		NCSTN
1.12	0.16					0.07	0.10	-0.18	-0.14	-0.18			-0.09	LOC441662
1.12	0.17	0.19			0.12				-0.20	-0.07	-0.17	-0.34	-0.33	ZIC2
1.12	0.17	0.10	0.09		0.16	0.21	0.14		-0.19	-0.09	-0.25	-0.14		EHD1
1.12	0.17	0.17	0.11		0.11	0.15	0.19	-0.10			-0.08		-0.09	FAM195B
1.12	0.17	0.14	0.16	0.20							-0.06	0.07		DEPDC1B
1.12	0.17			0.13	-0.17	-0.19	0.08	-0.14	-0.12	-0.08	0.22	0.22	0.40	UGCG
1.12	0.17			-0.07	0.23	0.21	0.25	-0.19	-0.19	-0.18		0.09		SKIV2L
1.12	0.17		0.07	0.09	-0.18	-0.17	-0.12		-0.07	-0.24	0.13	0.20	0.32	CTNNA1
1.12	0.17	-0.12	-0.15			0.14	0.13	-0.23	-0.38	-0.21	0.60	0.74	0.63	LTB
1.12	0.17			0.07	0.09				-0.19	-0.23	-0.31	-0.32	-0.35	STS-1
1.12	0.17	0.09	0.12	0.12	-0.08	-0.16	-0.26	-0.08	-0.10				0.10	UBE2Z
1.13	0.17	0.09	0.07	0.11	-0.29	-0.21	-0.22	-0.09	-0.07	-0.06	0.20	0.10	0.25	CTNNA1
1.13	0.17	0.15	0.17	0.26						0.13	-0.16		-0.11	LOC728576
1.13	0.17		0.16		0.10	0.09	0.09		-0.06	-0.18	-0.13	-0.14	-0.12	BNC2
1.13	0.17	0.22	0.20	0.22	-0.07	-0.08		0.08					-0.06	CLTB
1.13	0.17	0.10	0.10	0.13	0.23	0.12		-0.09				-0.10	-0.13	STARD13
1.13	0.17	0.08	0.07					-0.14		-0.19	-0.26	-0.26	-0.25	TCEA3
1.13	0.17	0.13	0.12	0.08		0.11	0.23	-0.07		-0.14	-0.41	-0.24	-0.31	CDCP1
1.13	0.17							-0.14	-0.21	-0.15		0.20	0.10	FAM90A3
1.13	0.17		0.07	0.08				-0.10	-0.08	-0.15	-0.14		-0.07	NME1
1.13	0.17		0.10		0.15	0.20	0.09		-0.06	-0.23	-0.13	-0.15	-0.10	CRNKL1
1.13	0.17	0.07				-0.07		-0.12	-0.14	-0.13				LOC642588
1.13	0.17		-0.08				0.11	-0.23	-0.19	-0.24	0.06	0.10		RRP8
1.13	0.17				0.14	0.08	0.15	-0.23	-0.16	-0.20		0.07	0.19	ASCC2
1.13	0.17	0.11	0.19	0.15	0.10		0.17		-0.08		-0.14	-0.07	-0.18	RASSF2
1.13	0.17				0.11			-0.12	-0.20	-0.21				F2R





1.15	0.20	DHCR7
1.15	0.20	ANTXR2
1.15	0.20	LAMB3
1.15	0.20	LOC201651
1.15	0.20	WDR20
1.15	0.20	LOC100128
1.15	0.20	CAP1
1.15	0.20	MEIS3
1.15	0.20	LSS
1.15	0.20	IRX4
1.15	0.20	RNFT1
1.15	0.20	SIX6
1.15	0.20	AURKA
1.15	0.20	SNX27
1.15	0.20	GINS3
1.15	0.20	PPP1R14B
1.15	0.20	TMEM138
1.15	0.20	AKR1D1
1.15	0.20	NAT10
1.15	0.20	VDAC3
1.15	0.20	CDK5RAP1
1.15	0.20	KIF5B
1.15	0.20	LOC100133
1.15	0.20	POLR2A
1.15	0.20	None
1.15	0.20	ARHGAP18
1.15	0.20	LDHB
1.15	0.20	TTF2
1.15	0.20	C11orf17
1.15	0.21	AADACL1
1.15	0.21	ESPL1
1.15	0.21	SNRPF
1.15	0.21	SLC2A1
1.15	0.21	GPSM3
1.15	0.21	SCG5
1.15	0.21	ACYP2
1.15	0.21	MGST1
1.15	0.21	ALPK1
1.15	0.21	POLE4
1.15	0.21	RTN1
1.15	0.21	SRXN1
1.15	0.21	CASP1
1.15	0.21	PTTG1IP
1.15	0.21	PHYH
1.16	0.21	CCDC128
1.16	0.21	PACSIN1
1.16	0.21	MRPS12
1.16	0.21	EDC4
1.16	0.21	S100A4
1.16	0.21	UBD
1.16	0.21	TNFSF18
1.16	0.21	DDX50
1.16	0.21	PMS2
1.16	0.21	LRIG1
1.16	0.21	HSPE1
1.16	0.21	ELOVL4
1.16	0.21	F2RL1
1.16	0.21	FLNC
1.16	0.21	IMPDH2
1.16	0.21	SUPT3H
1.16	0.21	C1orf25
1.16	0.21	NDUFAF2
1.16	0.21	HSF2BP
1.16	0.21	ZNF654
1.16	0.21	APLP1
1.16	0.21	HMGCR
1.16	0.21	UCHL5
1.16	0.21	SERTAD1
1.16	0.21	ANAPC4
1.16	0.21	TNIP1
1.16	0.21	ADAMTSL1
1.16	0.21	DNAH1
1.16	0.22	ERI1
1.16	0.22	LOC100129
1.16	0.22	SFRS10
1.16	0.22	LIX1L
1.16	0.22	SLC25A5
1.16	0.22	ZNF133
1.16	0.22	FOSB
1.16	0.22	TMEM91
1.16	0.22	DDT
1.16	0.22	M6PRBP1
1.16	0.22	MSRA
1.16	0.22	RBM5
1.16	0.22	SCARNA9
1.16	0.22	DACT3
1.16	0.22	PEPD
1.16	0.22	TMEM16A
1.16	0.22	GTPBP6
1.16	0.22	C3orf63
1.16	0.22	LPPR2
1.16	0.22	SHISA3
1.16	0.22	SERTAD2
1.16	0.22	CAMLG

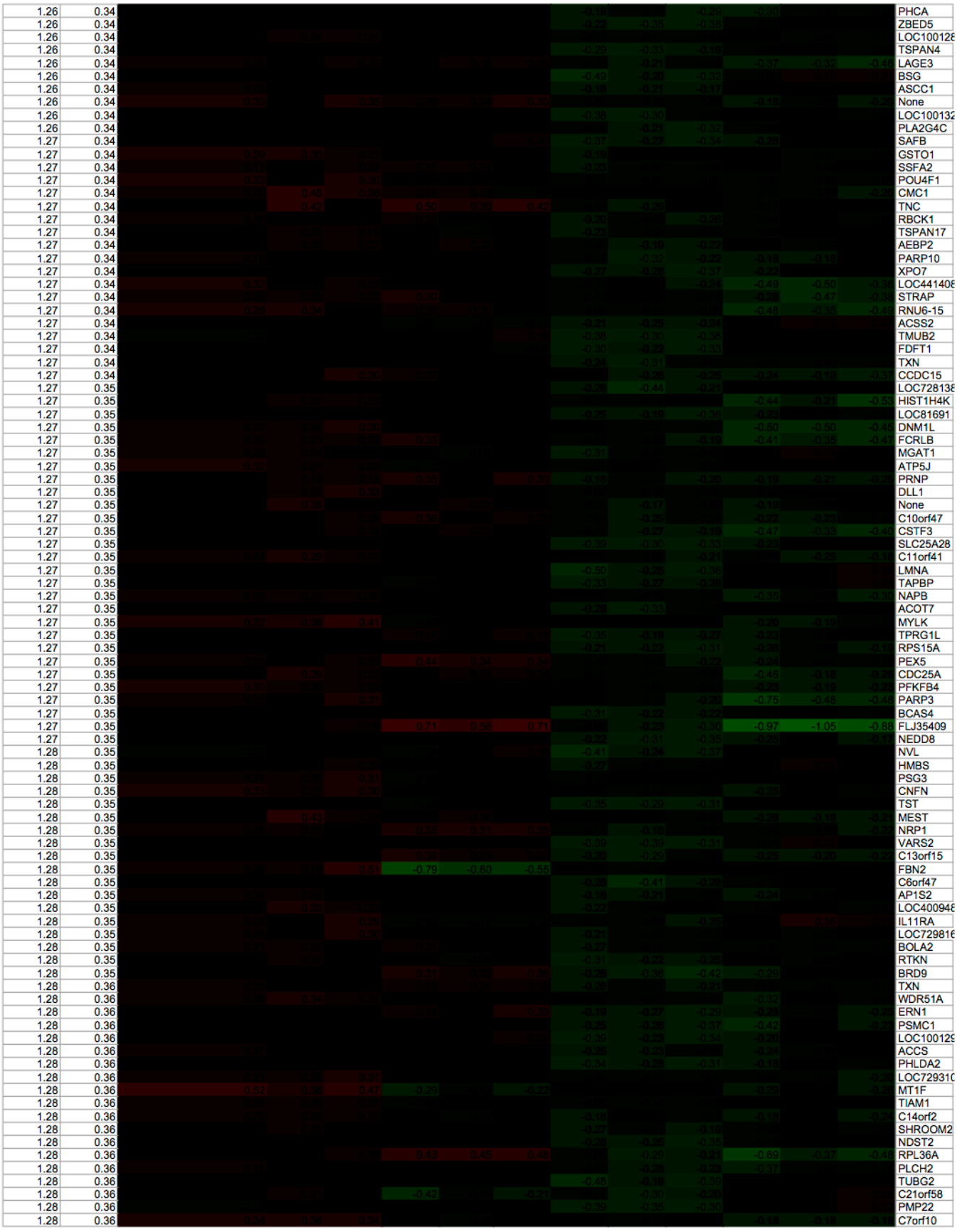


1.16	0.22	CHAF1B
1.17	0.22	ANGPTL2
1.17	0.22	YIPF6
1.17	0.22	FAU
1.17	0.22	SNORA57
1.17	0.22	IER5
1.17	0.22	PGRMC1
1.17	0.22	AURKA
1.17	0.22	HMX2
1.17	0.22	CSTB
1.17	0.22	LOC338758
1.17	0.22	SLC26A6
1.17	0.22	ZP3
1.17	0.22	None
1.17	0.22	MED20
1.17	0.22	SESN1
1.17	0.23	ATP5E
1.17	0.23	NHP2L1
1.17	0.23	CDC37
1.17	0.23	FTL
1.17	0.23	HIC2
1.17	0.23	DPF2
1.17	0.23	ZNF627
1.17	0.23	MID1
1.17	0.23	IFFO1
1.17	0.23	SPRYD3
1.17	0.23	RHOJ
1.17	0.23	JAK1
1.17	0.23	CSF2
1.17	0.23	None
1.17	0.23	SEC31A
1.17	0.23	CHD4
1.17	0.23	NCOR2
1.17	0.23	CEL
1.17	0.23	ACSL5
1.17	0.23	C20orf94
1.17	0.23	C9orf75
1.17	0.23	CCDC12
1.17	0.23	EXO1
1.17	0.23	LOC442597
1.17	0.23	SPC25
1.17	0.23	CDRT4
1.17	0.23	LOC727768
1.17	0.23	TRAF1
1.17	0.23	PTGES
1.17	0.23	CCDC72
1.17	0.23	ATF2
1.17	0.23	LOC653079
1.17	0.23	MRPL24
1.18	0.23	RAD51C
1.18	0.23	ULK1
1.18	0.23	BCAP29
1.18	0.23	LLPH
1.18	0.23	RPL23A
1.18	0.23	LOC642755
1.18	0.24	C14orf19
1.18	0.24	LOC339804
1.18	0.24	KCTD5
1.18	0.24	EMILIN2
1.18	0.24	CRADD
1.18	0.24	DUSP14
1.18	0.24	DAB2
1.18	0.24	KIAA0174
1.18	0.24	TIE1
1.18	0.24	UBAC2
1.18	0.24	GAK
1.18	0.24	EXO1
1.18	0.24	MSI2
1.18	0.24	PGM1
1.18	0.24	DBNL
1.18	0.24	IRF8
1.18	0.24	C20orf43
1.18	0.24	GTF2B
1.18	0.24	MTP18
1.18	0.24	C1orf74
1.18	0.24	DKK1
1.18	0.24	RBM18
1.18	0.24	UBE2M
1.18	0.24	XRCC6
1.18	0.24	CLCA2
1.18	0.24	None
1.18	0.24	HYLS1
1.18	0.24	DNAL1
1.18	0.24	PARP3
1.18	0.24	EEF1D
1.18	0.24	KIAA0133
1.18	0.24	RPS13
1.18	0.24	HNRPDL
1.18	0.24	C3orf37
1.18	0.24	AADAC
1.18	0.24	HSCB
1.18	0.24	LOC85389
1.18	0.24	ZNF35
1.19	0.25	ESRRA

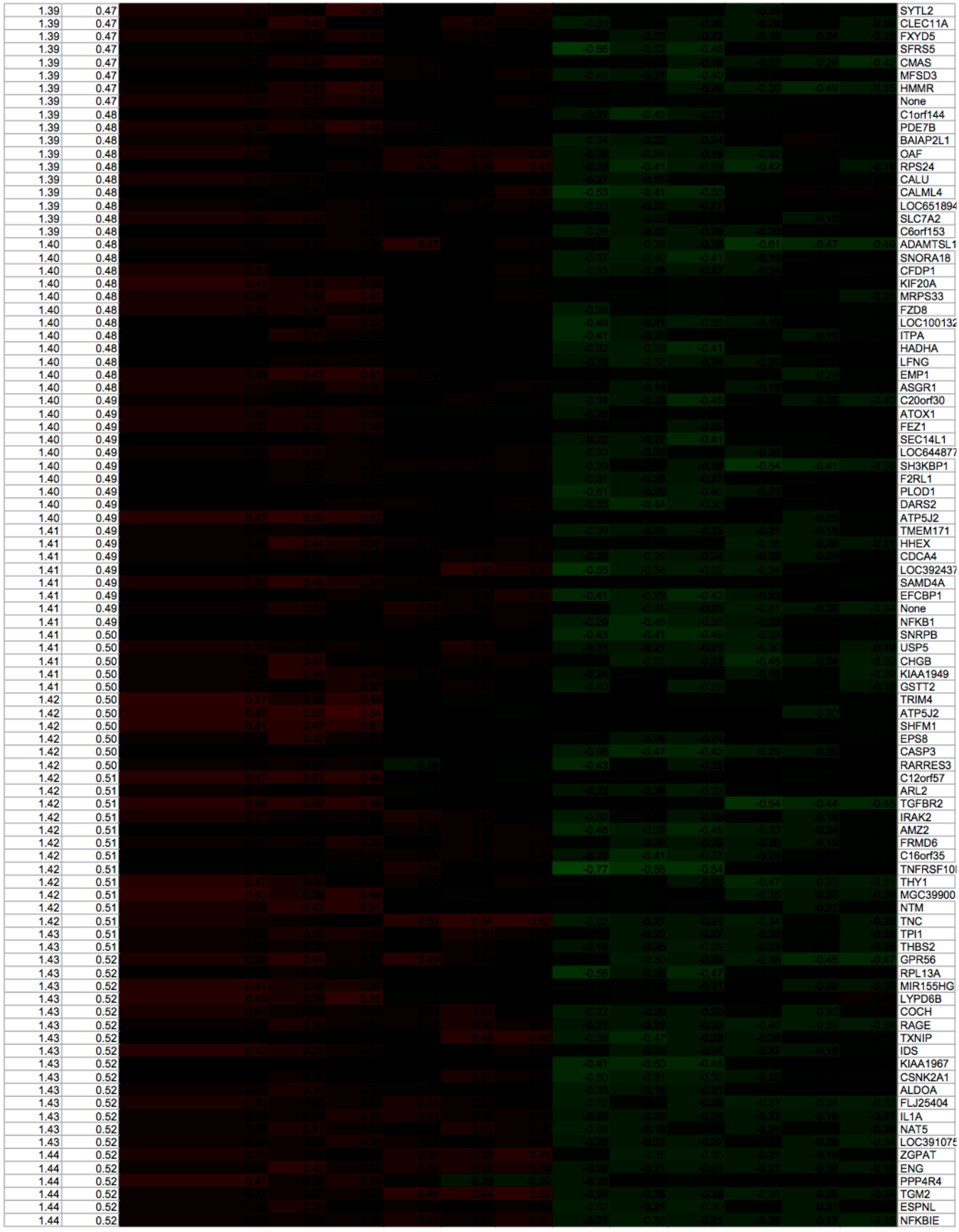
1.19	0.25	C14orf169
1.19	0.25	BAPX1
1.19	0.25	LHX6
1.19	0.25	ZIK1
1.19	0.25	DCTN5
1.19	0.25	CDKN3
1.19	0.25	KIF5C
1.19	0.25	OKL38
1.19	0.25	PREI3
1.19	0.25	PFKFB3
1.19	0.25	TSEN34
1.19	0.25	PPP4R4
1.19	0.25	SEPT2
1.19	0.25	GNPDA1
1.19	0.25	NTSR1
1.19	0.25	THUMPD2
1.19	0.25	DDX10
1.19	0.25	C12orf47
1.19	0.25	CDC16
1.19	0.25	UBE2I
1.19	0.25	UBE2D3
1.19	0.25	KIAA1618
1.19	0.25	IL1RAPL1
1.19	0.25	DNAJC22
1.19	0.25	ATP2B4
1.19	0.25	FCRLA
1.19	0.25	TWF2
1.19	0.25	BEXL1
1.19	0.25	RAB38
1.19	0.25	ARPC4
1.19	0.25	CYP4V2
1.19	0.26	TIGD7
1.19	0.26	ZNFX1
1.19	0.26	DCAF15
1.19	0.26	LOC730323
1.19	0.26	SNRPB2
1.19	0.26	MAN1B1
1.19	0.26	MRRF
1.20	0.26	None
1.20	0.26	ALDH18A1
1.20	0.26	TUBA1C
1.20	0.26	BOP1
1.20	0.26	TMEM97
1.20	0.26	TP53INP1
1.20	0.26	COG5
1.20	0.26	PIGF
1.20	0.26	RPL29
1.20	0.26	CRKL
1.20	0.26	GLRX5
1.20	0.26	KIAA0586
1.20	0.26	KLF9
1.20	0.26	WRN
1.20	0.26	EBF3
1.20	0.26	SNORD96A
1.20	0.26	ODF2
1.20	0.26	TRAF3IP2
1.20	0.26	SSH2
1.20	0.26	C14orf156
1.20	0.26	MEG3
1.20	0.26	ZMYND8
1.20	0.26	None
1.20	0.26	C19orf56
1.20	0.26	MRPS18A
1.20	0.26	C3orf38
1.20	0.26	EXOSC1
1.20	0.26	SEC23B
1.20	0.26	LOC730324
1.20	0.26	ZNF343
1.20	0.26	AKT1
1.20	0.26	TUBB3
1.20	0.26	APH1B
1.20	0.26	SNORD36A
1.20	0.26	XPR1
1.20	0.26	C12orf10
1.20	0.26	GULP1
1.20	0.26	CPVL
1.20	0.26	SNRPF
1.20	0.26	UBIAD1
1.20	0.26	HCST
1.20	0.26	ATG16L1
1.20	0.27	ARHGEF6
1.20	0.27	LBX2
1.20	0.27	GMPPB
1.20	0.27	PAFAH1B3
1.20	0.27	NET1
1.20	0.27	DARS
1.20	0.27	INTS1
1.20	0.27	PDXDC1
1.20	0.27	CBR3
1.20	0.27	NPC2
1.20	0.27	COX8A
1.20	0.27	NINJ1
1.20	0.27	TGM2
1.20	0.27	SPATA7

1.22	0.29							-0.25						MRPS34
1.23	0.29				0.47						-0.22	-0.27	-0.33	IL24
1.23	0.29								-0.24	-0.20	-0.34	-0.33	-0.20	AK3
1.23	0.29							-0.27	-0.22	-0.33				CDC16
1.23	0.29													SLC20A2
1.23	0.29				-0.50	-0.56	-0.66		-0.19	-0.21				COL5A1
1.23	0.29	0.34				0.36			-0.20					PLEKHA4
1.23	0.29		0.42		-0.22									COL6A2
1.23	0.29							-0.18	-0.24	-0.24				C4orf41
1.23	0.29													BARD1
1.23	0.30								-0.20					GPR19
1.23	0.30							-0.26		-0.22				RETSAT
1.23	0.30	0.30	0.42								-0.49	-0.47	-0.32	ING2
1.23	0.30							-0.22	-0.26	-0.24	-0.25			TPCN2
1.23	0.30			0.37										OTOF
1.23	0.30				0.66	0.47	0.36				-0.25	-0.19	-0.20	DCLK1
1.23	0.30	-0.18						-0.39	-0.33	-0.35				YY1AP1
1.23	0.30							-0.26		-0.24				RNF169
1.23	0.30								-0.23		-0.21			GTF3C6
1.23	0.30							-0.36	-0.22	-0.28				COMT
1.23	0.30									-0.22				PDPR
1.23	0.30								-0.21	-0.19				MRE11A
1.23	0.30								-0.30	-0.20	-0.33		-0.22	DLL3
1.23	0.30		0.29											TMEM59L
1.23	0.30													MCEE
1.23	0.30										-0.43	-0.17	-0.19	LOC100133
1.23	0.30							-0.20	-0.29	-0.23				LOC730235
1.23	0.30									-0.19	-0.23			TBC1D16
1.23	0.30							-0.19	-0.26	-0.21				NDUFA12
1.23	0.30	0.30							-0.30					PDE2A
1.23	0.30					0.31			-0.18	-0.30	-0.24		-0.26	RPS26
1.23	0.30				0.56	0.41		-0.19		-0.24	-0.44	-0.38	-0.40	SLC2A3
1.23	0.30	0.29							-0.23		-0.68	-0.45	-0.70	RPL29
1.23	0.30									-0.34	-0.36	-0.30	-0.29	PAM
1.23	0.30	0.32	0.30										-0.22	PILRB
1.23	0.30						0.31	-0.31	-0.28		-0.31			KLRAQ1
1.23	0.30								-0.20	-0.22	-0.17			TFB1M
1.23	0.30			0.31							-0.42	-0.26	-0.54	LOC388789
1.23	0.30							-0.30						DEF6
1.23	0.30		0.31											LOC389137
1.23	0.30							-0.21	-0.24		-0.19	-0.22	-0.28	ADK
1.23	0.30													TMEM93
1.23	0.30												-0.32	NSF
1.23	0.30			0.29					-0.24		-0.21	-0.21		PRIC285
1.23	0.30													FAM127A
1.23	0.30									-0.17	-0.16	-0.17	-0.22	APOBEC3F
1.23	0.30							-0.32	-0.26					PTPRA
1.23	0.30			0.30										LOC200030
1.23	0.30			0.29		-0.28	-0.22				0.32		0.59	RNF150
1.23	0.30							-0.29	-0.36	-0.18			-0.19	SNHG8
1.23	0.30							-0.26	-0.27	-0.31				CDT1
1.23	0.30							-0.34	-0.25	-0.21				EXOSC1
1.23	0.30	0.27										-0.26		PBRM1
1.23	0.30			0.30					-0.19		-0.29	-0.21		CETN3
1.23	0.30								-0.21	-0.23	-0.34	-0.27	-0.28	None
1.23	0.30							-0.27	-0.34	-0.34				BAG3
1.23	0.30							-0.21	-0.37	-0.21				SIPA1
1.23	0.30		0.29	0.32							-0.35	-0.31	-0.45	BOLA2
1.24	0.30													LOC731486
1.24	0.30						0.32	-0.28	-0.40	-0.30				ZFAND2A
1.24	0.31				-0.30			-0.28		-0.20				UBE2L6
1.24	0.31										-0.31	-0.25		RGMB
1.24	0.31									-0.18				RHOC
1.24	0.31				0.33				-0.20	-0.21	-0.27			CD248
1.24	0.31							-0.33	-0.21	-0.33				BSCL2
1.24	0.31	0.32						-0.18						PMS2L4
1.24	0.31							-0.25						MSRB2
1.24	0.31		0.37	0.33			0.34				-0.20	-0.34	-0.18	GBE1
1.24	0.31							-0.27	-0.36	-0.29	-0.26			PIGU
1.24	0.31		0.34	0.33								-0.19		ACYP1
1.24	0.31									-0.18		-0.20		ST3GAL6
1.24	0.31							-0.25		-0.24		-0.20		RNF24
1.24	0.31							-0.21	-0.23	-0.35				GPS1
1.24	0.31							-0.32	-0.21	-0.42				FAM39E
1.24	0.31													ACVRL1
1.24	0.31	0.29	0.29		-0.40	-0.25	-0.18							P4HA2
1.24	0.31							-0.24		-0.29	-0.21			NECAP1
1.24	0.31			0.34										LOC88523
1.24	0.31		0.32		-0.39	-0.40	-0.26	-0.22						P4HA2
1.24	0.31								-0.29	-0.30		-0.20		MYB
1.24	0.31							-0.28	-0.23	-0.42				LOC148413
1.24	0.31	0.33												SAMD8
1.24	0.31							-0.26	-0.41	-0.29			-0.23	SCYL1
1.24	0.31		0.41	0.33							-0.26	-0.45	-0.37	LOC391019
1.24	0.31							-0.18	-0.26	-0.34				SFTA1P
1.24	0.31													RNMTL1
1.24	0.31													ANKRA2
1.24	0.31									-0.18	-0.48	-0.29	-0.32	PRKCD
1.24	0.31	0.31						-0.18	-0.19			-0.30		VAMP8
1.24	0.31							-0.25	-0.39	-0.27				TSPO
1.24	0.31								-0.18					DECR2
1.24	0.31							-0.24	-0.20	-0.35				CEP192
1.24	0.31		0.33						-0.26		-0.47	-0.24	-0.33	FXYD5
1.24	0.31	0.30		0.32										BAD

1.24	0.31								-0.22		-0.18			LOC391370
1.24	0.31		0.36	0.37								-0.30	-0.41	CDH13
1.24	0.31							-0.30	-0.21	-0.28				CFL1
1.24	0.31	0.47	0.38									-0.26	-0.39	BUD31
1.24	0.31							-0.21	-0.23	-0.25				RPRC1
1.24	0.31	0.40	0.47	0.40								-0.24		POLR2J2
1.24	0.31		0.37								-0.23		-0.46	LOC100129
1.24	0.31							-0.21	-0.23	-0.24	-0.24			LOC728153
1.24	0.32								-0.29	-0.35				FXR2
1.24	0.32							-0.17						TFIP11
1.25	0.32								-0.17					SFRS13A
1.25	0.32							-0.18						EIF4EBP2
1.25	0.32							-0.39		-0.29				FBXW2
1.25	0.32		0.42								-0.51	-0.43	-0.40	ACY1
1.25	0.32				0.33	0.30		-0.19	-0.26	-0.37	-0.33			ATN1
1.25	0.32							-0.19	-0.17					TTLL3
1.25	0.32						0.38	-0.34	-0.38	-0.30				TOMM34
1.25	0.32	0.40		0.40							-0.27	-0.21	-0.31	GCA
1.25	0.32										-0.28		-0.50	NOP56
1.25	0.32							-0.20		-0.23	-0.21			ICAM3
1.25	0.32				0.37		0.35		-0.23				-0.29	DNLZ
1.25	0.32	0.32							-0.19	-0.18			-0.19	RIN2
1.25	0.32							-0.22	-0.27	-0.25				CD82
1.25	0.32							-0.23	-0.24	-0.22	-0.17			PRDX6
1.25	0.32							-0.27		-0.17	-0.18			IMP4
1.25	0.32				0.36	0.36			-0.26	-0.22				E4F1
1.25	0.32		0.29		0.30	0.40	0.45					-0.24	-0.26	NRP1
1.25	0.32							-0.34						CSNK2A1P
1.25	0.32	0.33					0.39				-0.30	-0.23	-0.24	PHYH
1.25	0.32								-0.29		-0.20	-0.19	-0.37	TGFA
1.25	0.32				0.40		0.39	-0.25				-0.28	-0.28	SNORD3C
1.25	0.32							-0.25	-0.42	-0.21				NR1H3
1.25	0.32							-0.23	-0.42	-0.26				TJAP1
1.25	0.32							-0.32		-0.21				TMEM55B
1.25	0.32								-0.31	-0.18	-0.28	-0.24	-0.21	AP1S2
1.25	0.32							-0.20						CHMP4B
1.25	0.32	0.41							-0.19		-0.24	-0.26		FIS1
1.25	0.32													WDFY2
1.25	0.32							-0.19		-0.22		-0.23		LOC100128
1.25	0.32							-0.46	-0.28	-0.23				RPL12
1.25	0.32								-0.19	-0.22		-0.24		DLGAP5
1.25	0.32							-0.31	-0.26	-0.35	-0.32			ALDH9A1
1.25	0.32								-0.19		-0.28		-0.17	RPL34
1.25	0.32								-0.23	-0.24	-0.46	-0.55	-0.52	FAM98A
1.25	0.32							-0.38		-0.20				ITM2C
1.25	0.33		0.38				0.30							C21orf69
1.25	0.33							-0.19	-0.31	-0.22	-0.20			MTMR11
1.25	0.33								-0.18		-0.25	-0.32	-0.24	SNORD57
1.25	0.33	0.39		0.34							-0.38	-0.29	-0.30	HIST1H2BH
1.25	0.33				0.34		0.37						-0.20	None
1.25	0.33										-0.35			TMEM115
1.25	0.33	0.38												HIST1H2BJ
1.25	0.33							-0.42	-0.25	-0.30				NSDHL
1.25	0.33							-0.26	-0.20	-0.24	-0.20			CECR5
1.25	0.33		0.32											LAMP2
1.25	0.33	0.30		0.30							-0.22	-0.27	-0.24	GTF2H5
1.26	0.33							-0.27		-0.25				TCEB3
1.26	0.33								-0.17		-0.26	-0.26	-0.20	HSZFP36
1.26	0.33							-0.25						C8orf13
1.26	0.33								-0.19		-0.21			WDR68
1.26	0.33							-0.18	-0.40	-0.39				LMNB2
1.26	0.33							-0.53	-0.19	-0.36		0.31		C17orf49
1.26	0.33							-0.30	-0.31	-0.35				PRIM2A
1.26	0.33								-0.22	-0.18				CD34
1.26	0.33							-0.18		-0.27				GPR137
1.26	0.33	0.32									-0.46	-0.22	-0.25	MAPKAPK3
1.26	0.33								-0.22		-0.22			MPRIP
1.26	0.33	0.31	0.32								-0.23			SULF2
1.26	0.33								-0.24	-0.22	-0.20			CDS2
1.26	0.33				0.41	0.37	0.40	-0.24	-0.30	-0.32	-0.32			NOP2
1.26	0.33						0.37		-0.35	-0.28			-0.22	RIOK1
1.26	0.33										-0.36	-0.59	-0.34	PYGL
1.26	0.33	0.32	0.31								-0.50	-0.29	-0.40	LRRFIP2
1.26	0.33							-0.43	-0.21	-0.19		0.33		GPX3
1.26	0.33	0.32	0.34											GRB14
1.26	0.33							-0.46		-0.48				AVPI1
1.26	0.33							-0.48	-0.22	-0.33				COX4I1
1.26	0.33							-0.26		-0.19				TMED10P
1.26	0.33							-0.25						ASAM
1.26	0.33						0.30	-0.25		-0.20		-0.28	-0.21	EMG1
1.26	0.33							-0.39	-0.22					BCS1L
1.26	0.33	0.41	0.31											ARMC10
1.26	0.33						-0.19	-0.24		-0.35	0.59		0.50	NCOA7
1.26	0.33		0.30									-0.24	-0.18	FMNL2
1.26	0.33							-0.27	-0.33	-0.29				ACVR1B
1.26	0.33									-0.19	-0.49	-0.58	-0.52	RNU6-1
1.26	0.33							-0.38	-0.24	-0.31				MRPL17
1.26	0.33							-0.29	-0.25	-0.27				CCNE2
1.26	0.34							-0.22		-0.18				SLCO4A1
1.26	0.34						0.33	-0.34	-0.39	-0.36				LPCAT3
1.26	0.34					0.35			-0.26		-0.43	-0.43	-0.24	C20orf46
1.26	0.34									-0.17		-0.29	-0.32	DNER
1.26	0.34		0.30								-0.17			BEND4
1.26	0.34						0.42	-0.24		-0.28	-0.29			TTC27

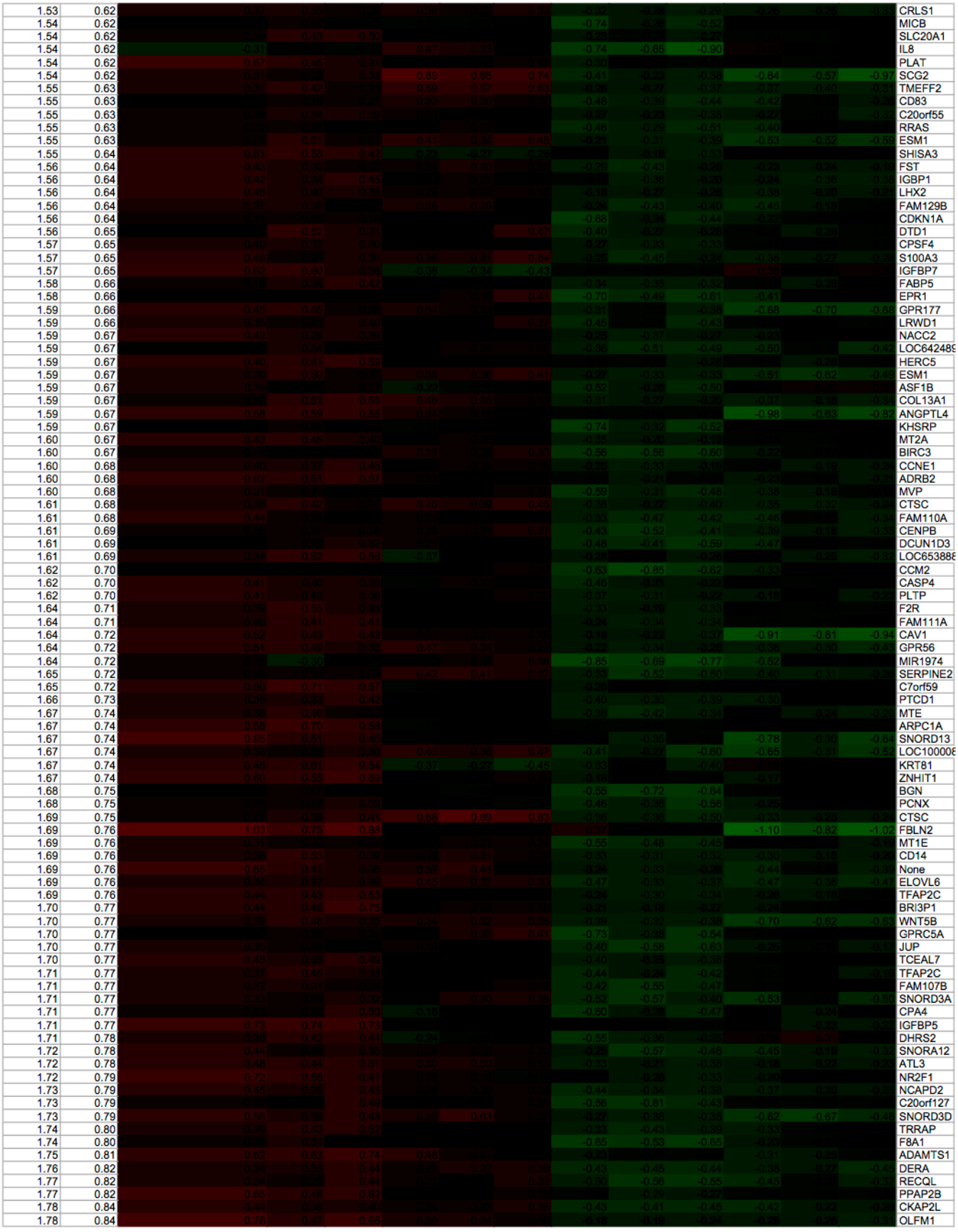


1.28	0.36	0.30	0.25		0.42	0.41	0.44		-0.33					PPPDE2
1.28	0.36							-0.43	-0.29	-0.32				SBDSP
1.29	0.36	0.21		0.47		0.28					-0.64	-0.86	-0.61	PEG10
1.29	0.36							-0.27	-0.36	-0.37	-0.29			GMDS
1.29	0.36	0.31	0.36		-0.18			-0.19						PTPRA
1.29	0.36							-0.38	-0.37	-0.37	-0.32			ZDHHC16
1.29	0.36			0.21	-0.25			-0.30	-0.29	-0.27				None
1.29	0.36													PSME2
1.29	0.37							-0.36	-0.21	-0.31				C17orf53
1.29	0.37		0.36					-0.20		-0.17			-0.22	LOC653557
1.29	0.37							-0.34	-0.30	-0.30	-0.31			CDC20
1.29	0.37	0.35	0.33	0.32							-0.42	-0.24	-0.18	DNM1L
1.29	0.37							-0.44	-0.39	-0.34				CDK5RAP2
1.29	0.37							-0.19	-0.29	-0.21				C1orf86
1.29	0.37	0.33								-0.20			-0.17	DHX29
1.29	0.37	0.33	0.34	0.41			-0.18				-0.19			DUSP6
1.29	0.37								-0.24	-0.25	-0.32		-0.38	ATP5I
1.29	0.37							-0.29		-0.32				TNFRSF1A
1.29	0.37	0.31	0.40											NCRNA000
1.29	0.37	0.30		0.31					-0.25		-0.23			UQCRB
1.29	0.37	0.30	0.38										-0.29	AP1S1
1.29	0.37	0.45								-0.19	-0.37	-0.17	-0.23	MARCH3
1.29	0.37		0.35					-0.26				-0.38	-0.27	ATP5G1
1.29	0.37								-0.18		-0.27			UBR2
1.29	0.37					0.44	0.44	-0.25		-0.21	-0.29	-0.36	-0.33	SEMA3A
1.29	0.37	0.41												LOC550643
1.29	0.37						0.30	-0.24	-0.29	-0.18	-0.24			CCDC56
1.29	0.37									-0.19	-0.32			KLHL35
1.29	0.37		0.36										-0.18	PAMR1
1.29	0.37							-0.24	-0.31	-0.34	-0.30		-0.18	IQCC
1.29	0.37	0.35		0.35	0.48	0.32	0.42		-0.22	-0.18				FAM116B
1.29	0.37							-0.29		-0.18	-0.22	-0.26		LOC387841
1.29	0.37													FAM171B
1.29	0.37							-0.45		-0.30	-0.30	-0.36		CDCA8
1.29	0.37				-0.32	-0.30	-0.48	-0.43		-0.22				HLA-H
1.29	0.37							-0.23				-0.20	-0.32	LOC100134
1.29	0.37							-0.40		-0.39	-0.28			SDHAF2
1.30	0.37								-0.27	-0.19				LRCH2
1.30	0.37							-0.26	-0.38	-0.27	-0.31	-0.26		AP1B1
1.30	0.37								-0.25	-0.18	-0.22		-0.20	HDHD1A
1.30	0.37							-0.32	-0.35	-0.40				GPN2
1.30	0.37								-0.27		-0.22			CSNK2A1P
1.30	0.37							-0.39	-0.25	-0.47	-0.18			C1orf112
1.30	0.37		0.29					-0.25						MSMP
1.30	0.37			0.31			0.34	-0.18	-0.19		-0.20	-0.18		None
1.30	0.37							-0.22	-0.22		-0.21			UROS
1.30	0.37							-0.20	-0.45	-0.34				PPP1R15A
1.30	0.37							-0.55	-0.36	-0.40		0.41	0.38	ASNS
1.30	0.37							-0.43	-0.24	-0.41				ALDH1A3
1.30	0.37	0.31				0.30			-0.18	-0.19	-0.35	-0.23	-0.25	SNORA61
1.30	0.37			0.29				-0.20		-0.20	-0.24		-0.21	TSPYL2
1.30	0.38								-0.32	-0.37				EFR3A
1.30	0.38							-0.38	-0.26	-0.28				HJURP
1.30	0.38	0.41		0.36			-0.23							RHOB
1.30	0.38	0.34			0.30				-0.22	-0.20	-0.27	-0.23		PRMT3
1.30	0.38				0.32		0.38		-0.34	-0.24	-0.44	-0.26	-0.22	MED28
1.30	0.38							-0.46	-0.36	-0.36				COQ10B
1.30	0.38	0.32		0.35						-0.18				CSTF3
1.30	0.38		0.41	0.29										RN5S9
1.30	0.38						0.30	-0.18	-0.40	-0.32	-0.20			CENPF
1.30	0.38								-0.18			-0.22		SYCE2
1.30	0.38							-0.27	-0.30		-0.20			RPL13
1.30	0.38	0.33		0.44										CD24
1.30	0.38							-0.22		-0.19	-0.22	-0.21	-0.34	MRPL51
1.30	0.38	0.46	0.49											TRIP6
1.30	0.38	0.45		0.39							-1.20	-1.05	-1.07	UBE2E2
1.30	0.38		0.29					-0.29				-0.19		MLF2
1.30	0.38							-0.19						DENND3
1.30	0.38				-0.18			-0.26	-0.28	-0.30				PSMB10
1.30	0.38	0.39									-0.50	-0.49	-0.58	SH3KBP1
1.30	0.38								-0.20	-0.30				STAT1
1.30	0.38							-0.30	-0.27	-0.31				RBMS1
1.31	0.38							-0.23	-0.19					MRPS26
1.31	0.38	0.31								-0.25				PI4K2B
1.31	0.38								-0.24	-0.21	-0.38		-0.24	NRIP3
1.31	0.38		0.37					-0.35			-0.18		-0.25	NP
1.31	0.39		0.38	0.40										FAM46A
1.31	0.39							-0.28	-0.34	-0.30				KRT34
1.31	0.39							-0.45	-0.28	-0.35	-0.19			INPP1
1.31	0.39									-0.29				ZNF451
1.31	0.39	0.32						-0.18		-0.30	-0.41		-0.22	KCNK3
1.31	0.39							-0.36	-0.19	-0.28	-0.30			PHB
1.31	0.39							-0.32	-0.29	-0.35				BMS1
1.31	0.39							-0.24		-0.20	-0.18			C7orf44
1.31	0.39	0.36		0.33									-0.18	CLDND2
1.31	0.39								-0.23	-0.20	-0.37		-0.20	CD70
1.31	0.39			0.36				-0.23	-0.24	-0.18	-0.28			C11orf17
1.31	0.39							-0.27	-0.21	-0.30	-0.38		-0.23	POLR3F
1.31	0.39							-0.40	-0.39	-0.41				SLC7A5
1.31	0.39	0.33	0.38										-0.21	MOCS1
1.31	0.39							-0.28	-0.31	-0.30			-0.28	SIGIRR
1.31	0.39	0.32	0.40											CCND1
1.31	0.39							-0.46	-0.33	-0.34				UBE2L6
1.31	0.39	0.35		0.37				-0.25						SRGN



1.39	0.47			0.31			0.56			0.31					-0.29			SYTL2
1.39	0.47			0.29		0.41				0.34			-0.30		-0.26	-0.29		CLEC11A
1.39	0.47			0.31			0.33							-0.23	-0.22	-0.19	-0.24	FXYD5
1.39	0.47												-0.66	-0.32	-0.46			SFRS5
1.39	0.47			0.29		0.37	0.45			0.31					-0.19	-0.27	-0.29	CMAS
1.39	0.47									0.34			-0.40	-0.27	-0.40			MFSD3
1.39	0.47			0.28		0.31	0.51								-0.26	-0.30	-0.49	HMMR
1.39	0.47			0.37		0.32	0.40											None
1.39	0.48												-0.27	-0.43	-0.27			C1orf144
1.39	0.48			0.39		0.36	0.49			0.39								PDE7B
1.39	0.48												-0.34	-0.23	-0.24			BAIAP2L1
1.39	0.48			0.37							0.29	0.31	0.30	-0.26	-0.34	-0.19	-0.32	OAF
1.39	0.48										0.35	0.39	0.43	-0.23	-0.41	-0.35	-0.42	RPS24
1.39	0.48			0.40										-0.27	-0.26			CALU
1.39	0.48									0.38			-0.53	-0.41	-0.50			CALML4
1.39	0.48			0.29						0.34			-0.33	-0.22	-0.27			LOC651894
1.39	0.48			0.35		0.37								-0.26			-0.19	SLC7A2
1.39	0.48												-0.39	-0.22	-0.28	-0.20		C6orf153
1.40	0.48			0.27		0.29			0.47					-0.29	-0.26	-0.61	-0.47	ADAMTSL1
1.40	0.48												-0.37	-0.40	-0.41	-0.19		SNORA18
1.40	0.48			0.31						0.38			-0.33	-0.28	-0.27	-0.24		CFDP1
1.40	0.48			0.43		0.36	0.36											KIF20A
1.40	0.48			0.38		0.34	0.41			0.36								MRPS33
1.40	0.48			0.36		0.31							-0.36					FZD8
1.40	0.48						0.32						-0.49	-0.21	-0.35	-0.18		LOC100132
1.40	0.48					0.31							-0.41	-0.26			-0.19	ITPA
1.40	0.48												-0.32	-0.39	-0.41			HADHA
1.40	0.48												-0.18	-0.32	-0.24	-0.20		LFNG
1.40	0.48			0.38		0.47	0.41										-0.29	EMP1
1.40	0.48			0.35			0.38							-0.18		-0.18		ASGR1
1.40	0.49												-0.34	-0.23	-0.45		-0.25	C20orf30
1.40	0.49			0.29		0.35	0.39						-0.26					ATOX1
1.40	0.49			0.32		0.33	0.36								-0.25			FEZ1
1.40	0.49												-0.22	-0.27	-0.41			SEC14L1
1.40	0.49			0.29		0.30							-0.33	-0.23		-0.26		LOC644877
1.40	0.49												-0.39		-0.28	-0.54	-0.41	SH3KBP1
1.40	0.49												-0.31	-0.28	-0.27			F2RL1
1.40	0.49												-0.51	-0.28	-0.40	-0.25		PLOD1
1.40	0.49												-0.33	-0.44	-0.30			DARS2
1.40	0.49			0.47		0.55	0.52										-0.25	ATP5J2
1.41	0.49			0.29		0.29							-0.39	-0.25	-0.23	-0.21	-0.18	TMEM171
1.41	0.49			0.27		0.44	0.39									-0.35	-0.26	HHEX
1.41	0.49												-0.35	-0.29	-0.24	-0.28	-0.26	CDCA4
1.41	0.49									0.40	0.31		-0.55	-0.34	-0.32	-0.34		LOC392437
1.41	0.49			0.32		0.46	0.35											SAMD4A
1.41	0.49										0.31		-0.41	-0.29	-0.42	-0.23		EFCBP1
1.41	0.49			0.27		0.39				0.32				-0.33	-0.21	-0.31	-0.29	None
1.41	0.49												-0.29	-0.46	-0.35	-0.23		NFKB1
1.41	0.50												-0.43	-0.41	-0.49	-0.33		SNRPB
1.41	0.50			0.37						0.30			-0.21	-0.21	-0.21	-0.30		USP5
1.41	0.50			0.29		0.41								-0.32	-0.23	-0.45	-0.24	CHGB
1.41	0.50			0.29		0.46	0.32						-0.24			-0.19		KIAA1949
1.41	0.50						0.37						-0.40		-0.29			GSTT2
1.42	0.50			0.57		0.38	0.49											TRIM4
1.42	0.50			0.47		0.55	0.64										-0.30	ATP5J2
1.42	0.50			0.51		0.47	0.51											SHFM1
1.42	0.50					0.42								-0.24	-0.24			EPS8
1.42	0.50												-0.36	-0.47	-0.42	-0.29	-0.26	CASP3
1.42	0.50									0.28			-0.43		-0.28			RARRES3
1.42	0.51			0.47		0.51	0.48											C12orf57
1.42	0.51												-0.23	-0.28	-0.33			ARL2
1.42	0.51			0.48		0.49	0.49									-0.54	-0.44	TGFBR2
1.42	0.51			0.31						0.32			-0.32		-0.28		-0.18	IRAK2
1.42	0.51												-0.48	-0.29	-0.45	-0.33	-0.24	AMZ2
1.42	0.51			0.27			0.39							-0.20	-0.29	-0.20	-0.19	FRMD6
1.42	0.51												-0.37	-0.41	-0.37	-0.28		C16orf35
1.42	0.51									0.29			-0.77	-0.55	-0.54			TNFRSF10I
1.42	0.51			0.47		0.40									-0.29	-0.47	-0.33	THY1
1.42	0.51			0.63		0.38	0.49									-0.35	-0.22	MGC39900
1.42	0.51			0.39		0.42	0.51										-0.22	NTM
1.42	0.51			0.38						0.52	0.54	0.62	-0.32	-0.27	-0.21	-0.34		TNC
1.43	0.51			0.28		0.35	0.32				0.30			-0.20	-0.27	-0.20		TPI1
1.43	0.51												-0.19	-0.28	-0.29	-0.22		THBS2
1.43	0.52			0.32		0.41				0.49				-0.30	-0.20	-0.36	-0.45	GPR56
1.43	0.52												-0.58	-0.28	-0.47			RPL13A
1.43	0.52			0.41		0.36	0.37								-0.21		-0.29	MIR155HG
1.43	0.52			0.43		0.28	0.55											LYPD6B
1.43	0.52			0.34		0.40				0.31			-0.32	-0.20	-0.25		-0.32	COCH
1.43	0.52					0.35				0.30			-0.21	-0.22	-0.22	-0.40	-0.39	RAGE
1.43	0.52									0.49	0.45		-0.38	-0.47	-0.28	-0.26		TXNIP
1.43	0.52			0.42		0.29								-0.23	-0.24	-0.27	-0.18	IDS
1.43	0.52												-0.41	-0.50	-0.45			KIAA1967
1.43	0.52										0.37		-0.50	-0.31	-0.30	-0.20		CSNK2A1
1.43	0.52			0.28		0.26							-0.35	-0.18	-0.21			ALDOA
1.43	0.52			0.30		0.36				0.31			-0.32		-0.28	-0.37	-0.34	FLJ25404
1.43	0.52									0.32			-0.35	-0.23	-0.28	-0.22	-0.18	IL1A
1.43	0.52			0.28		0.31					0.31		-0.39	-0.18		-0.24		NAT5
1.43	0.52			0.29			0.39						-0.26	-0.23	-0.27		-0.26	LOC391075
1.44	0.52			0.39						0.30	0.33	0.45		-0.35	-0.30	-0.21	-0.19	ZGPAT
1.44	0.52					0.42				0.36	0.29	0.29	-0.26	-0.21	-0.21	-0.21	-0.26	ENG
1.44	0.52			0.47			0.39				-0.29	-0.20	-0.26					PPP4R4
1.44	0.52									0.46	0.44	0.52	-0.24	-0.28	-0.29	-0.35	-0.25	TGM2
1.44	0.52			0.28		0.30							-0.32	-0.21	-0.30		-0.21	ESPNL
1.44	0.52									0.40		0.31	-0.27	-0.33	-0.21	-0.26	-0.17	NFKBIE

1.44	0.52	0.49	0.30	0.39	0.20						-0.20			OLFM1
1.44	0.53		0.41					-0.39		-0.36				UBR4
1.44	0.53	0.31	0.35	0.31				-0.28	-0.31					TMEM205
1.44	0.53	0.51	0.44	0.46							-0.22			MMP9
1.44	0.53			0.30			0.38		-0.45	-0.37	-0.34	-0.22	-0.38	C3orf26
1.44	0.53	0.45	0.30	0.34				-0.18	-0.20		-0.41		-0.28	ATG7
1.44	0.53		0.31		0.43	0.38	0.36	-0.22	-0.40	-0.24	-0.41	-0.21	-0.30	C20orf27
1.45	0.53		0.36					-0.40		-0.33	-0.21	-0.19		G6PC3
1.45	0.53						0.44	-0.27	-0.39	-0.43	-0.45		-0.25	HMGA1
1.45	0.53							-0.34	-0.33					DHRS4
1.45	0.53	0.59	0.32	0.53							-0.24			CYFIP2
1.45	0.53	0.38	0.30	0.38		0.21		-0.19	-0.17	-0.20	-0.42	-0.33	-0.19	CYP27B1
1.45	0.53							-0.45	-0.28	-0.33				TUBG1
1.45	0.53	0.40	0.51	0.44										COL6A2
1.45	0.53	0.37	0.43	0.46							-0.21		-0.19	PPARG
1.45	0.53	0.31		0.37				-0.18	-0.30					UPLP
1.45	0.53	0.31	0.29					-0.29	-0.36		-0.23	-0.24	-0.25	RGS10
1.45	0.53	0.32	0.33	0.32	0.36				-0.23	-0.26	-0.21	-0.24	-0.26	WBP11
1.45	0.53					0.35		-0.23	-0.21	-0.33	-0.35	-0.41	-0.35	FKBP1A
1.45	0.54	0.29						-0.33	-0.43	-0.33	0.40	0.58	0.48	C15orf48
1.45	0.54	0.30		0.32				-0.22	-0.33	-0.29				RRBP1
1.45	0.54	0.32		0.31				-0.31	-0.33					CEACAM1
1.45	0.54							-0.46	-0.33	-0.27	-0.19	-0.21		SNRPB
1.46	0.54			0.35					-0.41	-0.26	-0.20			HEBP1
1.46	0.54							-0.32	-0.59	-0.49	-0.43		-0.21	SNHG5
1.46	0.54	0.35	0.43					-0.25				-0.36	-0.29	ING2
1.46	0.55				0.51	0.60	0.53	-0.41	-0.29	-0.24	-0.18	-0.30	-0.31	CTSC
1.46	0.55			0.29				-0.23	-0.42	-0.30			-0.18	TRIP11
1.47	0.55	0.59	0.55	0.61							-0.31	-0.60	-0.54	RGL1
1.47	0.56	0.30		0.40	0.40		0.37		-0.38	-0.35	-0.43		-0.20	GALNT9
1.47	0.56	0.31		0.29				-0.28	-0.25	-0.27	-0.25		-0.42	SPSB2
1.47	0.56							-0.55	-0.41	-0.42				ETFB
1.47	0.56		0.50	0.35				-0.32		-0.19		-0.21		RIPK2
1.47	0.56	0.43	0.37	0.35					-0.22	-0.18		-0.22		APOBEC3C
1.47	0.56	0.57	0.35				0.39		-0.19	-0.17	-0.20	-0.20		PYGB
1.47	0.56	0.45	0.41	0.30					-0.24	-0.24	0.39			C21orf7
1.48	0.56			0.37					-0.31	-0.39	-0.29			BIRC5
1.48	0.56		0.38	0.29				-0.31	-0.23	-0.25	-0.20	-0.26		FER1L3
1.48	0.56	0.31	0.43	0.32	0.40			-0.18	-0.26	-0.19	-0.31	-0.33	-0.30	ARHGAP24
1.48	0.56							-0.20	-0.35	-0.33				ECM1
1.48	0.57	0.46			0.33			-0.22	-0.27	-0.25	-0.68	-0.79	-0.62	C11orf1
1.48	0.57				0.33			-0.22	-0.41	-0.37	-0.33		-0.20	VASN
1.48	0.57	0.32	0.40	0.30				-0.41						TAX1BP3
1.49	0.57	0.44		0.32				-0.38	-0.22	-0.24		-0.21	-0.22	LOC641700
1.49	0.58	0.49	0.36	0.54					-0.18					PDLIM3
1.49	0.58	0.42	0.40	0.36	-0.18		-0.30			-0.24				JAG1
1.49	0.58							-0.44	-0.36	-0.33	-0.28	-0.40		PRNP
1.49	0.58	0.33	0.53	0.31				-0.24	-0.17		-0.18	-0.45	-0.38	ITGA2
1.49	0.58							-0.50	-0.27	-0.51	-0.19			VPS24
1.50	0.58	0.49	0.50	0.58		0.33					-0.58	-0.50	-0.39	ACN9
1.50	0.58	0.42		0.34				-0.19	-0.39	-0.32	-0.28			CEACAM1
1.50	0.58	0.34	0.59	0.42				-0.27						SRGN
1.50	0.58	0.38	0.36	0.81							-0.39	-0.24		VGLL4
1.50	0.58	0.33		0.33				-0.28	-0.22	-0.38	-0.30			C17orf90
1.50	0.58							-0.39	-0.55	-0.47	-0.49		-0.30	PNPO
1.50	0.58	0.41		0.30	0.33	0.31	0.32	-0.27	-0.27	-0.28	-0.23	-0.25	-0.25	BATF3
1.50	0.59	0.32	0.45	0.33				-0.24		-0.29	-0.36			SLC12A9
1.50	0.59	0.43	0.61				0.38		-0.23	-0.23	-0.17			CPXM1
1.50	0.59				0.40		0.34	-0.23	-0.47	-0.54	-0.30			LAMC3
1.50	0.59							-0.53	-0.43	-0.53				VPS24
1.51	0.59		0.40	0.46		0.31		-0.37	-0.29		-0.18	-0.26	-0.30	RGS10
1.51	0.59	0.30							-0.43	-0.45	-0.36	-0.23	-0.23	COPS7A
1.51	0.59					0.31		-0.31	-0.40	-0.37	-0.28	-0.36		SLC16A5
1.51	0.59	0.49		0.29				-0.21	-0.33	-0.20	-0.27			None
1.51	0.59		0.49	0.29				-0.34	-0.18	-0.22	-0.22		-0.20	SMTN
1.51	0.60	0.32	0.40	0.41	0.29			-0.24	-0.22	-0.20		-0.22	-0.22	GRP
1.51	0.60	0.30		0.50					-0.23	-0.36		-0.20	-0.47	HIST2H2AA
1.51	0.60	0.44	0.44	0.50				-0.22						QPCT
1.51	0.60	0.36	0.32	0.37	0.42			-0.25	-0.30	-0.19	-0.23			FAM129B
1.51	0.60							-0.39	-0.43	-0.54	-0.25		-0.21	IFI35
1.52	0.60		0.34	0.36				-0.47		-0.23	-0.33	-0.27	-0.34	CMAS
1.52	0.60	0.36	0.32		0.45	0.48	0.38	-0.34	-0.28	-0.29	-0.21	-0.34	-0.24	MGST1
1.52	0.61	0.51	0.45	0.61								-0.20	-0.36	HIST1H2BD
1.52	0.61							-0.56	-0.44	-0.40	-0.28			KLHL36
1.52	0.61							-0.49	-0.27	-0.34				FKBP1A
1.52	0.61	0.32					0.52	-0.32	-0.36	-0.36	-0.49	-0.18	-0.34	SNRPB2
1.52	0.61			0.29				-0.51	-0.25	-0.34	-0.29		-0.21	KIF2C
1.53	0.61						0.35	-0.57	-0.40	-0.57	-0.37			NAMPT
1.53	0.61	0.42	0.50	0.35				-0.22		-0.20	-0.32	-0.34	-0.23	TNFRSF10
1.53	0.61		0.46				0.32	-0.25	-0.31	-0.28	-0.34	-0.24	-0.29	YARS2
1.53	0.61							-0.48	-0.52	-0.49	-0.31			EFHD2
1.53	0.61	0.50	0.49	0.45					-0.23		-0.77	-0.94	-0.76	TGFBR2
1.53	0.61	0.46	0.54	0.36	-0.20			-0.29						POP7
1.53	0.61					0.34	0.33	-0.49	-0.45	-0.52	-0.31			DDX47
1.53	0.61	0.50	0.38	0.47			0.37			-0.18	-0.22	-0.20		VAMP5
1.53	0.61	0.40	0.36	0.30				-0.18	-0.28	-0.32	-0.33		-0.24	CST3
1.53	0.61	0.47	0.42	0.31				-0.35			-0.25	-0.32	-0.18	PDGFRL
1.53	0.61	0.38	0.44	0.33				-0.30	-0.21	-0.18	-0.18			MT1A
1.53	0.61	0.38						-0.35	-0.22	-0.51	-0.23	-0.24		ZNF185
1.53	0.61		0.39	0.32				-0.44	-0.31					TRIOBP
1.53	0.61	0.38	0.42			0.37	0.41	-0.19	-0.50	-0.26	-0.39		-0.33	FKBP1A
1.53	0.62		0.54	0.57				-0.28			-0.43	-0.35	-0.25	CAV1
1.53	0.62	0.62	0.54	0.47					-0.28		-0.80	-0.68	-0.64	VGF
1.53	0.62	0.38	0.31	0.58				-0.17		-0.29	-0.24	-0.33		HMMR



1.78	0.84	0.81	0.82	0.83	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.23	-0.28	[illegible]	IGFBP5
1.79	0.84	0.62	0.55	0.64	[illegible]	[illegible]	[illegible]	-0.22	-0.21	-0.28	-0.19	[illegible]	-0.18	GAL3ST4
1.80	0.85	0.38	0.47	0.60	[illegible]	[illegible]	[illegible]	-0.44	-0.23	-0.42	0.32	[illegible]	[illegible]	DHRS2
1.80	0.85	0.43	0.43	0.40	[illegible]	0.29	[illegible]	-0.45	-0.48	-0.35	[illegible]	[illegible]	[illegible]	SMOX
1.80	0.85	0.53	0.78	0.73	[illegible]	[illegible]	-0.23	-0.40	[illegible]	[illegible]	[illegible]	[illegible]	-0.26	ARPC1B
1.81	0.85	1.04	0.86	0.81	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.19	-1.25	-1.17	FBLN2
1.81	0.85	-0.54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.13	-0.79	-1.07	[illegible]	-0.23	[illegible]	IL8
1.82	0.86	0.67	0.64	0.62	[illegible]	[illegible]	[illegible]	-0.27	-0.24	[illegible]	-0.38	-0.44	-0.38	LPHN2
1.82	0.87	0.40	0.43	0.48	[illegible]	[illegible]	[illegible]	-0.54	-0.42	-0.33	-0.25	-0.21	-0.18	C20orf72
1.83	0.87	0.77	0.70	0.62	[illegible]	[illegible]	[illegible]	-0.23	-0.26	[illegible]	-0.24	-0.23	-0.34	TFPI2
1.83	0.87	0.77	0.45	0.49	[illegible]	[illegible]	[illegible]	[illegible]	-0.40	-0.38	-0.18	-0.17	[illegible]	ZCWPW1
1.85	0.89	0.40	0.56	0.45	0.48	0.44	[illegible]	-0.39	-0.49	-0.38	-0.44	-0.27	-0.33	VEGFC
1.85	0.89	0.48	0.37	0.53	[illegible]	[illegible]	[illegible]	-0.43	-0.44	-0.42	-0.29	[illegible]	[illegible]	EBI3
1.86	0.89	0.34	[illegible]	[illegible]	0.31	0.39	0.42	-0.55	-0.63	-0.66	-0.49	-0.21	-0.28	INA
1.86	0.89	0.70	0.74	0.65	0.37	[illegible]	[illegible]	[illegible]	[illegible]	-0.36	-0.59	-0.75	-0.90	DUSP6
1.86	0.89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.50	-0.92	-0.70	-0.81	-0.35	[illegible]	[illegible]	LOC100130
1.86	0.90	0.58	0.55	0.44	0.39	0.32	[illegible]	-0.22	-0.44	-0.47	-0.70	-0.36	-0.57	KCNMA1
1.87	0.90	0.71	0.67	0.73	[illegible]	[illegible]	[illegible]	-0.29	[illegible]	-0.21	[illegible]	[illegible]	[illegible]	COL6A1
1.87	0.90	0.65	0.69	0.52	[illegible]	[illegible]	[illegible]	-0.20	-0.25	-0.40	-0.29	[illegible]	[illegible]	ITGB2
1.88	0.91	0.49	0.42	0.46	[illegible]	0.53	0.48	-0.45	-0.44	-0.49	-0.39	[illegible]	[illegible]	FKBP1A
1.88	0.91	0.64	0.74	0.70	0.32	[illegible]	[illegible]	-0.22	-0.21	-0.23	[illegible]	-0.70	-0.55	GREM1
1.89	0.92	0.29	0.32	[illegible]	[illegible]	[illegible]	[illegible]	-0.78	-0.38	-0.72	-0.28	[illegible]	[illegible]	PRKCDBP
1.90	0.93	0.62	0.72	0.72	[illegible]	[illegible]	[illegible]	-0.30	[illegible]	-0.29	-0.28	[illegible]	[illegible]	COPS6
1.90	0.93	0.41	0.49	0.40	[illegible]	[illegible]	[illegible]	-0.54	-0.45	-0.50	[illegible]	[illegible]	[illegible]	FABP5L2
1.91	0.93	[illegible]	0.36	[illegible]	0.59	0.55	0.60	-0.49	-0.78	-0.66	-0.70	-0.23	-0.65	CD68
1.92	0.94	0.62	0.65	0.58	0.41	0.39	[illegible]	-0.35	-0.29	-0.35	-0.45	-0.42	-0.35	MLPH
1.94	0.95	0.62	0.32	0.62	[illegible]	0.33	[illegible]	-0.56	-0.33	-0.41	-0.29	[illegible]	-0.31	TNFAIP3
1.95	0.97	[illegible]	[illegible]	0.40	[illegible]	[illegible]	0.46	-0.79	-0.57	-0.63	-0.38	-0.33	-0.23	LOC286016
1.96	0.97	0.46	0.62	0.54	0.42	0.43	0.36	-0.43	-0.36	-0.50	-0.58	-0.52	-0.41	FARS2
1.97	0.98	1.00	0.81	0.94	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	TMSB15A
1.98	0.98	0.56	0.52	0.69	0.38	0.38	0.29	-0.29	-0.47	-0.42	-1.02	-0.84	-0.84	HS3ST3A1
2.00	1.00	0.83	0.69	0.69	[illegible]	0.43	0.41	-0.33	-0.24	-0.22	-0.32	-0.23	-0.27	MT1X
2.00	1.00	0.66	0.62	0.78	[illegible]	[illegible]	[illegible]	-0.30	-0.31	-0.32	[illegible]	[illegible]	[illegible]	FEZ1
2.00	1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.46	-1.14	-0.76	-0.90	-0.37	[illegible]	[illegible]	LOC389286
2.00	1.00	0.69	0.70	0.86	[illegible]	[illegible]	[illegible]	-0.20	-0.23	-0.31	-0.31	-0.35	-0.48	HIST1H2BD
2.01	1.01	0.59	0.66	0.59	0.55	0.38	0.48	-0.41	-0.38	-0.39	-0.89	-0.68	-0.69	CDH6
2.02	1.02	0.68	0.68	0.54	0.31	0.32	0.41	-0.43	-0.36	-0.37	-0.53	-0.31	-0.41	DYSF
2.10	1.07	1.08	0.89	0.92	-0.25	[illegible]	-0.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	CD24
2.10	1.07	-0.18	-0.21	[illegible]	[illegible]	[illegible]	[illegible]	-1.10	-1.31	-1.23	[illegible]	[illegible]	[illegible]	ALDOC
2.13	1.09	0.44	0.68	0.69	1.05	1.02	0.78	-0.54	-0.48	-0.44	-1.06	-1.04	-1.04	SFRP1
2.14	1.10	0.48	0.42	0.51	0.63	0.74	0.86	-0.66	-0.54	-0.67	-0.71	-0.42	-0.49	LOC100132
2.15	1.10	0.52	0.63	0.47	0.52	0.58	0.71	-0.47	-0.63	-0.59	-0.59	-0.62	-0.57	ALPL
2.17	1.12	0.74	0.69	0.92	[illegible]	0.40	0.41	-0.34	-0.29	-0.38	-0.30	-0.26	-0.29	CKLF
2.17	1.12	0.57	0.40	0.59	0.67	0.83	0.75	-0.63	-0.57	-0.59	-0.40	-0.45	-0.40	G0S2
2.18	1.12	0.55	0.61	0.64	0.58	0.47	0.50	-0.47	-0.49	-0.60	-0.62	-0.53	-0.52	LPAR1
2.19	1.13	0.70	1.05	0.95	[illegible]	[illegible]	[illegible]	-0.29	[illegible]	-0.28	[illegible]	-0.38	-0.28	HIST1H2BK
2.19	1.13	0.32	0.48	0.47	[illegible]	[illegible]	[illegible]	-0.61	-0.78	-0.74	-0.22	[illegible]	-0.28	LOC642956
2.20	1.14	0.67	0.71	0.57	[illegible]	0.48	[illegible]	-0.61	-0.37	-0.48	[illegible]	-0.21	-0.25	H2AFJ
2.26	1.18	0.46	0.50	0.57	[illegible]	[illegible]	0.33	-0.61	-0.77	-0.64	-0.19	[illegible]	[illegible]	VPS16
2.27	1.18	0.84	0.71	0.91	[illegible]	0.37	0.29	-0.31	-0.38	-0.41	-0.25	-0.34	-0.32	CKLF
2.27	1.19	0.82	0.73	0.77	0.34	0.48	0.47	-0.34	-0.44	-0.46	-0.66	-0.72	-0.84	MIR1978
2.28	1.19	0.90	0.73	0.88	0.31	0.34	0.31	-0.35	-0.39	-0.31	-0.73	-0.98	-0.85	ACTA2
2.28	1.19	1.08	0.86	0.84	[illegible]	[illegible]	[illegible]	[illegible]	-0.39	-0.28	-0.30	[illegible]	[illegible]	ITGB2
2.30	1.20	0.57	0.74	0.76	[illegible]	[illegible]	[illegible]	-0.51	-0.52	-0.51	[illegible]	-0.19	-0.27	KRT86
2.30	1.20	0.83	0.71	0.90	[illegible]	[illegible]	[illegible]	-0.35	-0.52	-0.29	[illegible]	[illegible]	[illegible]	GNG11
2.33	1.22	[illegible]	[illegible]	[illegible]	[illegible]	0.38	0.57	-1.34	-1.05	-1.35	-0.46	0.45	[illegible]	LOC650909
2.40	1.26	1.06	0.84	0.90	[illegible]	[illegible]	[illegible]	-0.34	-0.32	-0.32	[illegible]	[illegible]	[illegible]	SIPA1L2
2.43	1.28	0.92	0.85	0.84	0.64	0.43	0.39	-0.45	-0.39	-0.39	-0.56	-0.47	-0.64	CDK6
2.44	1.29	0.76	0.84	0.95	0.42	0.34	0.35	-0.34	-0.57	-0.40	-1.11	-1.08	-1.13	None
2.52	1.34	0.39	0.63	0.57	0.45	0.47	0.37	-0.90	-0.69	-0.83	-0.47	-0.45	-0.37	TFPI
2.59	1.37	0.67	0.86	0.71	0.51	0.60	0.48	-0.73	-0.48	-0.66	-1.85	-1.87	-1.83	RGS4
2.59	1.38	0.68	0.76	0.66	0.74	0.65	0.64	-0.64	-0.72	-0.67	-1.27	-1.23	-1.22	ARHGDIB
2.73	1.45	1.24	1.37	1.38	[illegible]	0.35	[illegible]	[illegible]	[illegible]	[illegible]	-0.26	[illegible]	[illegible]	SPOCK1
2.77	1.47	0.92	0.99	0.87	0.53	0.45	0.33	-0.51	-0.49	-0.64	-0.33	-0.57	-0.43	GREM1
2.80	1.49	0.71	0.65	0.93	1.19	1.04	0.96	-0.71	-0.72	-0.75	-0.67	-0.65	-0.72	CD163L1
2.90	1.53	1.15	1.24	1.02	0.60	0.51	0.36	-0.43	-0.40	-0.37	-0.37	-0.58	-0.36	KIAA1199
2.95	1.56	0.79	0.94	0.85	0.62	0.55	0.62	-0.65	-0.74	-0.70	-0.80	-0.55	-0.73	GALR2
2.97	1.57	0.62	0.64	0.74	[illegible]	[illegible]	0.37	-0.94	-0.94	-0.82	-0.20	-0.21	[illegible]	FABP5
2.98	1.57	0.69	0.72	0.78	0.79	0.68	0.86	-0.82	-0.91	-0.81	-0.79	-0.77	-0.68	ECSCR
3.01	1.59	0.92	0.90	0.97	0.83	0.63	0.67	-0.68	-0.63	-0.68	-0.74	-0.81	-0.73	TMEM200A
3.07	1.62	0.43	0.62	0.60	0.50	0.59	0.45	-1.10	-1.03	-1.08	-0.65	-0.43	-0.57	TFPI
3.17	1.66	1.18	1.36	1.34	0.51	[illegible]	0.31	-0.39	-0.39	-0.34	-0.31	-0.25	-0.28	ABI3BP
3.33	1.74	0.71	0.76	0.70	0.51	0.89	0.80	-0.66	-1.10	-1.29	-0.69	-0.51	-0.54	CXCL1
3.42	1.77	0.83	0.79	0.75	0.56	0.68	0.41	-0.94	-1.00	-1.01	-0.66	-0.41	-0.70	LAMB3
3.47	1.79	0.87	1.04	0.96	1.01	0.76	1.08	-0.76	-0.82	-0.94	-1.52	-1.41	-1.46	KISS1
3.54	1.82	1.18	1.15	1.08	0.60	0.63	0.44	-0.44	-0.80	-0.82	-0.76	-0.61	-0.72	HIST1H2BK
3.61	1.85	0.94	0.85	0.91	1.34	1.34	1.29	-0.85	-1.04	-0.96	-0.87	-0.91	-0.93	M160
3.95	1.98	0.70	0.83	0.87	0.35	[illegible]	[illegible]	-1.20	-1.22	-1.12	-0.18	[illegible]	-0.31	FABP5L2
3.98	1.99	-0.18	[illegible]	[illegible]	[illegible]	0.32	0.94	-2.23	-1.75	-2.08	-0.81	0.54	0.47	LOC100132
4.05	2.02	1.33	1.48	1.34	0.39	0.60	0.53	-0.64	-0.59	-0.66	-0.46	-0.39	-0.39	IL1B
5.44	2.44	1.32	1.77	1.62	1.38	1.11	0.85	-0.87	-0.91	-0.85	-0.86	-1.06	-1.02	MMP1