Protein perturbations in diseases with emphasis on autophagy process: Insights from the mathematical modelling and network biology perspectives

Thesis submitted for the award
of the degree of

Doctor of Philosophy

by

Dipanka Tanu Sarmah

Under the supervision of

Prof. Nandadulal Bairagi

Center for Mathematical Biology and Ecology
Department of Mathematics
Jadavpur University
Kolkata-700032, India
October 2022

Declaration

I, Dipanka Tanu Sarmah, solemnly declare that this thesis represents my own work which has been done after registration for the degree of PhD at Jadavpur University and has not been previously included in any thesis or dissertation for the purpose of earning a degree, diploma, or any other credential.

Dipanka Tanu Sarmah

Dipanka Tanu Sarmah

19.10.2022

DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
JADAVPUR UNIVERSITY
Kolkata – 700032, India
Telephone: 91 (33) 2414 6717

Dr. Nandadulal Bairagi
Professor and Coordinator
CENTRE FOR MATHEMATICAL BIOLOGY AND ECOLOGY
Mail: nbairagi.math@jadavpuruniversity.in

Date:

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled "**Protein perturbations in diseases with emphasis on autophagy process: Insights from the mathematical modelling and network biology perspectives.**" submitted by **Sri Dipanka Tanu Sarmah**, who got his name registered **(Ref. No.: D-7/SC/114/18 and Index No.: 75/18/Maths./25)** on 21/12/2018 for the award of Ph.D. (Science) degree of Jadavpur University, is absolutely based upon his own work under the supervision of myself, and that neither this thesis nor any part of it has been submitted for either any degree / diploma or any other academic award anywhere before.

19/10/22

Nandadulal Bairagi

Nandadulal Bairagi (Ph. D.)
Professor, Dept. of Mathematics
Jadavpur University
Kolkata-700032

Dedicated to Maa, Deuta,

Majani,

Lovi,

and

Dr. Samrat Chatterjee

"Come my love and do not spurn,

From a little flower to learn.

Let your temper be as sweet,

As the lily at your feet;

Be as gentle, be as mild;

Be as modest, simple child.”

————————(The lilly of the valley)

O my thesis, be my Lily, please be gentle, please be mild

Acknowledgement

Ok.

The hard part.

What if I miss someone?

There are so many people to say thanks.

So many to acknowledge.

So many to say what if you guys were not there?

Let's start with the best.

Thank you, Ganesh Ji, for blessing me, in all aspects of my life and helping me to reach here. Nothing would have been possible had you not blessed me in those tough conditions of my life. I owe you a lot.

Now chronologically.

I would like to thank all my teachers of *"Nagar Aamkatia Prathamik Vidyalay"*. My first school, to whom I owe my elementary education. I would like to thank all my teachers of "Dhakukahana Higher Secondary School" for their support and belief in me. Thank you B.Borroah college Mathematics department, Anjana Ma'am, Ripa Ma'am, Debashis Sir, Himashri Ma'am, Bandita Ma'am and Kashi Sir.

Two people in my life greatly impacted me. I used to memorise mathematics till seventh grade. Only Chetri Sir's influence inspired me to learn mathematics in a completely new way,

and I shifted from memorisation to understanding. Sir, whatever I am today is mostly due to your influence.

Late Palak borta- the guy who taught me English. His fairly unique method of instruction moulded my English. To the public, he was an oddball, but to me, he was a true hero, an alchemist whose extraordinary vocabulary allowed him to shape the universe via his words. Borta, you are no longer with us. I regret not meeting you in your final days, but I am confident that you would have been extremely proud if you had seen my thesis. Perhaps a few tears would have fallen, or you would have exclaimed, "Shilar maak, Moon r thesis saahi!" (Mother of Shila, come and see the thesis of Moon). I am missing you, borta. Be joyful wherever you are...

I would also like to specially mention Shovan Sir, who translated english grammar into formulas. Sir, whatever my grammatical foundation may be, I owe you everything. I still remember and ask everyone to translate, “Sapor manuhjone sapori sapori saporitur majere goi asil”. Sir, you are a gem of a person.

I am really fortunate to have Professor Nandadulal Bairagi Sir as my PhD supervisor. He provided me with essential encouragement, support, and expert advice throughout this roller-coaster trip. His faith in me and the provision of such a supportive environment are truly admirable. His consistent counseling and career and research assistance have been invaluable. Every time we have a conversation, I always learn something new. His charisma and interest in numerous aspects of life teach us how to effectively balance our professional and personal lives. I am extremely appreciative of his tolerance, motivation, belief, excitement, and vast knowledge, which made him such an admirable human being. I could not have imagined having a better doctoral research supervisor.

I would like to thank Dr Samrat Chatterjee, my mentor and advisor. My coming to the lab was nothing more than serendipity. He selected me from obscurity and trained to attain

this stage. His friendliness and his professional maturity were always beneficial in igniting my excitement for work. Thank you, Sir, for noticing the potential in me and embracing me in my quest to be the man I am today. Without your advice and mentoring, I never would have imagined having the chance to reach this stage. Thank you, Sir, for seeing more in me than I ever saw in myself, I am eternally grateful.

Friends are the family we choose, the wine to life, the sunshine to flowers and the petrichor to the pluviophile. I am glad to have experienced friendship throughout my life. I am blessed that I had my best buddies during the toughest period of my life, during my days at THSTI. I will start with my roommates and brothers, Abhijit and Suvankar. We had a great time together, inside and outside the lab. You guys have taught me many things, from cooking to playing badminton to solving Matlab codes. Specially with Abhijit, it was fun during collaborative work. Mitul, my neighbour at my desk, my elder brother and my teammate in "battlegrounds". Academic or personal, you were always available to help me. I've always been comforted knowing I can ask for help from you whenever I need it. You are one of the most brilliant, exceptional, and remarkable person I've ever met. Thank you, Manitosh bhaiya, for your support and advice. Anita, my elder sister, you were kind and always ready to help. Thank you for the support and advice you have given me throughout my days at THSTI. We three were and always be the group of Chintamanies. Swapna, Paulomi and Mimansa, my three little sisters. I never anticipated the amount of affection you guys have shown me as your brother. Your trust towards me helped me to be a better person. Undoubtedly, you guys have built a home in me. Jay and Dev, or should I write JayDev? Thank you both for coming and filling our lab environment with positivity and enthusiasm. Both of you are gems. Kartik and Suraj, you guys were amazing, and your dedication towards work has always inspired me. I would like to thank Abhijit and Chitta from the Jadavpur University Mathematics Department for their constant help and support. As the dawn transforms into a magnificent sunrise, may God shower you with his loving blessings and always guide you to the correct path. I would like to specially mention two names, Phoni Da and Rajat Sir, thank you both for your guidance and support throughout the years.

Shivam, chota bhai, and my parents' other son. In every event of my life at THSTI, your presence was a comfort. I was always certain someone would manage the situation if anything went awry, which was always the case with me. You are someone I could trust completely. I feel incredibly privileged to have you as my brother. Thank you.

I would like to thank my extended family, Suren Mama, Bupai Mama, Dhan Mahi, late moina mahi, oni and beula mami, koka and bouti, my sisters priyakshi, junuka, thun, dhan mahir cuali, brothers, kalpajyoti, kokoli and moina for your love, belief and constant support.

There aren't any sufficient words to express my gratitude to my parents-in-law for their assistance, blessings, and well wishes. I am indebted to chota bhai Chinu for the unwavering love, compassion, immense moral support, and assistance he has consistently provided.

Thank you, Lovi, my best friend, my soon-to-be wife and the first reader of the thesis. You were there when I was roaming through the alleys of uncertainties, you were there to motivate me when I was deciding to quit, and you were there to cheer me up when the rejections emails of the papers were piling up. You held me high when I felt like a bird with shattered wings; when I was hesitant to speak out of fear of failing, you carried the unspoken words and created beautiful melodies. For me, your presence was "Ariadne's thread" to cross the labyrinth of the struggles.

Thank you, maa and deuta, my constant source of light. Thank you, Majani, for the smile, support, and trust. Standing on the edges of a river, have you ever seen beautiful sunshine and then lost your words to explain the serenity? Trust me, among the people who have helped me, I thought writing about you people would have been easy. Now I don't have words to explain my gratitude towards you guys. Maa, deuta, I have been bestowed with your strength and encouragement. I will never be able to return the faith you have placed in me, no matter how hard I try. Thank you for being my continual support when I thought I could no longer manage and

for giving me hope and boosting my spirits. I am lucky enough and blessed enough to have you as parents. Majani, you bring so much happiness and joy to my life I can not explain in words. When we were kids, I always thought, you are an angel from utopia who gave away her wings and came to earth to play with me. You are the best thing that has ever happened to me. Thank you for being my instinct of hope.

Dipanka Tanu Sarmah

Abstract

The proper functioning of the cellular mechanisms that underlie the makeup of living systems is governed by the intricate interactions between the proteins, which are frequently perturbed in disease conditions. The advancement of high-throughput technologies has led to an unprecedented wealth of quantitative data to trace these perturbations. It is of utmost importance to identify the key set of proteins responsible for modulating these perturbations to obtain the potential targets in a disease. Notably, systematic efforts to detect these core sets of proteins have spurred the expeditious growth of network biology, providing a framework ideal for describing disease characteristics and predicting prospective therapeutic targets. In addition, comprehending how these spatially and temporally dispersed perturbations culminate in imperative biological processes is crucial to understanding cellular homeostasis and, by extension, disease pathogenesis. One such cardinal biological process is autophagy, which remains at the crossroads of numerous other biological processes and pathways. Autophagy plays a crucial role in maintaining cellular homeostasis by degrading unwanted materials like damaged mitochondria and misfolded proteins. However, the contribution of autophagy toward a healthy cell environment is not limited to the cleaning process. It also assists in protein synthesis when the system lacks the amino acids' inflow from the extracellular environment due to diet consumption. Reduction in the autophagy process is associated with diseases like cancer, diabetes, non-alcoholic steatohepatitis, etc., while uncontrolled autophagy may facilitate cell death. In many diseases, therefore, Autophagy is seen to act as Janus.

Nevertheless, despite decades of prominent research focus, it is still a puzzle with various

missing pieces due to its complex mechanism in numerous biological processes and diseases. This necessitates the integration of systems biology into the autophagy scenario, which can investigate a system both in pieces and as a whole. The veracity of these investigations hinges on their capacity to capture effective system dynamics. The development of a purely theoretical algorithm may find crucial nodes in the network by resolving all spatiotemporal scales, often at a cost that ignores the effect on the clinical characteristics of a disease. At the same time, their findings may not allow for generalisation. Conversly, an algorithm that investigates only the primary network properties, or clinical characteristics, is limited by the inability to look at the in-depth association between proteins. Therefore, an unmet need exists for a systematic framework that bridges protein perturbations, large-scale theoretical simulations, autophagy, and clinical characteristics of a disease to learn effective disease pathophysiology. Addressing these piers, in this thesis, we have incorporated mathematical modelling and network biology approaches to develop computational frameworks to investigate the protein perturbations in diseases with an accentuation on the autophagy process. We first developed a framework for identifying autophagy-related targets in diabetic retinopathy by forming algorithmic alloys between disease and autophagic proteins. Investigating the perturbations of proteins at both the gene and metabolic levels and using network controllability, we developed a methodology to identify potential targets in NASH. In the same disease, we next incorporated a guilt-by-association methodology, machine learning, constrained network controllability, and metabolic analysis to identify another set of potential targets. Intriguingly, we noted that a subset of the targets identified by both frameworks was either autophagy-related or was surrounded by autophagy-related proteins, suggesting an autophagy-mediated mode of operation for these targets. Finally, we formulated a mathematical model to investigate the mechanistic understanding of the autophagy process, where we addressed the interplay between DNA damage and Autophagy. Overall, the study discussed in this thesis suggests some potential targets and therapeutic interventions which are either directly or indirectly autophagy-related. On one hand, the frameworks used in this study exploited a quintessential biological process by using an extensive clinical dataset and mathematical modelling, while, on the other hand, the protein perturbations at the gene and metabolic levels were investigated to identify potential therapeu-

tic targets. The proposed methodologies in this thesis are general and can be applied to study any potential disease. We believe that learning the disease dynamics with these frameworks will provide potent novel modalities for accurately targeting diseases and, thereby will assist in the advancement of the drug-discovery process.

Abbreviations

ARP-Autophagy related proteins

BCL2 - B-cell lymphoma 2

BH3 - BCL2 homology domain 3

CAP - Candidate Autophagic Proteins

CMap - Connectivity Map database

CP - Candidate proteins

DEG - Differentially expressed genes

DDE - Delay differential equation

DGIdb - The Drug Gene Interaction Database

DM - Diabetes Mellitus

DPN - Directional PPI network

DR - Diabetic Retinopathy

DRAM - DNA damage regulated autophagy modulator 1

DRP - Diabetic Retinopathy related proteins

ER - Endoplasmic Raticulam

FDR - False discovery rate

FVA - Flux variability analysis

gpr - Gene-protein-reaction association

GSA - global sensitivity analysis

GSEA - Gene set enrichment analysis.

GSMM - Genome scale metabolic model

GWAS - Genome-wide association studies

ICP - Indispensable candidate protein

KNN - K-Nearest Neighbor

LHS - Latin Hypercube Sampling

LSA - Local sensitivity analysis

MCG - Metabolic candidate genes

ML - Machine learning

MOMA - Minimization of metabolic adjustment

MTA - Metabolic transformation algorithm

Mtor - The mammalian target of rapamycin

NAFL - Non-alcoholic fatty liver

NAFLD - Non-alcoholic fatty liver disease

NASH - Non-alcoholic steatohepatitis

NES - Normalized enrichment scores

PPI - Protein-protein interaction

PRCC - Partial Ranked Correlation Coefficient

RFE - The recursive feature elimination

RWR - Random walk with restart

RWRMLA - Random walk restart multilayer approach

RWRPCC - Random walk restart proteins with classification capability

RWRPs - Random walk restart proteins

SA - Sensitivity analysis

SVM - Support Vector Machines

TP - Transformation score

TPM - Transcripts Per Kilobase Million

WGCNA - Weighted gene co-expression network analysis

Contents

List of Figures

List of Tables

1

Introduction

1.1 Autophagy

"Frenemy"- a friend who may turn into a potential enemy. This is how a cell might introduce autophagy to its compartments and cellular constituents. This crosstalk is necessary because the process of autophagy, which is otherwise helpful, may turn into Thanatos if the situations inside the cell or its immediate neighbourhood go berserk [1].

Autophagy refers to the process of self-eating. The term derives from the Greek word "Auto", which means self and "phagy", which means death. However, just as we cannot say Mozart was only a musician or Da Vinci was only a painter, enclosing autophagy to this simple definition does not justify its role. But how and why does autophagy need to play a role in the cells? Let us first go through what if it doesn't.

More than 200 years of biochemical research have gathered a pile of data to address every minute detail of the formation of proteins and their functions. Interestingly, like autophagy, the name protein also comes from the Greek word "Proteios", which means "first or foremost". The name unequivocally justifies their importance. They are the building blocks of the cell and carries out the necessary task to maintain the structure, and functions of the tissues and organs of the body. Proteins need to assemble themselves with a sequence of amino acids to perform their function. This process, known as the central dogma of molecular biology, starts with the transcription and ends with the translation of proteins. As soon as the protein attains its structure, it starts performing its assigned function whenever the need occurs. But what happens if a protein does not acquire its necessary amino acid sequence? Does such a situation even exist? The answer is yes. Every 1 in 7 proteins does not get to perform its function due to the incorrect occurrence of amino acid sequence, occurring due to factors like mutations, ribosomal error, temperature and pH etc. Such proteins are termed misfolded proteins, and they remain in the cytoplasm, idle. This leads to two types of problems inside the cell. Firstly, shortage of proteins to do a specific task. For instance, if the proteins responsible for metabolising sugar continuously misfold, the cell will grow slowly due to the lack of energy. Another problem is the contortion of the misfolded proteins into shapes unfavourable for the cellular environment. Both these problems lead to multiple diseases. Cystic Fibrosis, Marfan syndrome, and Tay-Sachs disease are among the diseases caused by the first scenario, while Huntington's, Parkinson and Alzheimer's disease are to be named among the diseases caused by the second.

However, it is not only the proteins, which, despite being essential, may act harmful when not folded properly. Another type of such essential to lethal transition is shown by mitochondria, the powerhouse of the cell. Initially identified and declared as an elementary organism by Altman in 1890 (however, he named them bioblasts), mitochondria finally got their name from Carl Benda in 1898, which is acquired from two Greek words, "mitos" (thread) and "chondros" (granule). It plays a multifaceted role in cell biology due to its involvement with energy production, apoptosis activation, calcium homeostasis, phospholipid synthesis, and multiple metabolic pathways such as fatty acid activation, gluconeogenesis etc. However, due to the

mutation in nuclear or mitochondrial DNA, mitochondria lose their function and pave the way to various clinically heterogeneous diseases, which primarily affect oxidative phosphorylation resulting in a decrease in ATP production. Damaged mitochondria can release an excessive amount of reactive oxygen species (ROS) to the cytosol, increasing inflammation. Some such diseases include Leigh's syndrome, deafness, diabetes etc. Can a cell overcome these hiccups? It, of course, can not survive without these important molecular/ organelles. Therefore eliminating them at the early stage is beyond a fact to consider. So what is the mechanism by which cells keep these hiccups at bay and keep on performing their functions? The answer is simple, if anything goes berserk, eat it, and that is where the cellular housekeeping process autophagy comes in.

As already mentioned, cells experience a frequent appearance of unwanted constituents in the cytoplasm. Necessitating the need for their removal, cells initiate a series of signalling events which trigger the process of autophagy. Autophagy then takes this debris to the lysosome, where they are degraded, and different forms of amino acids are released, which are again used by the body in the construction of proteins and other cellular functions. In other words, when the toxicity has veiled the pool of intracellular tranquillity, autophagy activates, and, after a cascade of events, the cells gladly discover the debris outstretched at the lysosome, being ready to be turned into nutrients.

However, although easy it sounds, autophagy is a complex biological process which needs the synergetic association of numerous proteins for successful completion. It is divided into three parts, i) macroautophagy, ii) microautophagy, and iii) chaperone-mediated autophagy. While the three of them morphologically differ, the motive of the three types are the same, they differ in how they degrade the intracellular debris. For instance, in macrophagy, a double membrane structure called autophagosome is formed, which engulfs and carries the cargo to the lysosome. This structure is not required in microautophagy, and the cytoplasm directly fuses with the lysosome. Chaperone-mediated autophagy, on the other hand, can act only on a certain type of protein and directly takes them to the lysosome for degradation. By the term autophagy, we will be referring to this form only.

Since we now know its importance and about whom we should be talking, let's go back a few

decades and take a brief snap at the history of autophagy.

1.1.1 History of autophagy research

Autophagy was a constant buzz in cell biology even before it was discovered. In the 1860s, there was a concept of having a self-nourishment system in the human body, which allows the individual to survive by eating itself when a nutrient deficiency occurs [2]. In 1859, in a French journal titled “Des Seances de I academic de Science” (Session of the academy of sciences), an impactful article containing the term “autophagie” was published by French physiologist M. Anselmier. Addressing this, on the 14th of March, 1860, an article was published in the New York Times under the section “Scientific Gossip in Paris” where they used the phrase “cannibalism reduced to a civilized and humanitary institution” to summarise the work published by M. Anselmier. In paris, the impact of the article of Anselemier led to the inclusion of the word autophagie in the dictionary “NOUVEAU DICTIONNAIRE DE MÉDECINE ET DE CHIRURGIE PRATIQUES” [2].

However, autophagy, as we know it today, was first defined by de Duve in 1963, which involved serendipity, and some observations gleaned from experiments of many scientists. In 1955, a renowned Belgian cytochemist, Christian de Duve, while studying the effect of insulin on rat liver cells, discovered the presence of a novel organelle, which he named “lysosome”, referring to its “lytic” nature [3]. Soon this became a topic of discussion, and electron microscopic studies were employed to assess their role further. These studies revealed the fusion of phagosomes and lysosomes to degrade foreign materials. Following this, in 1957, the experiment by S. Clark et al. [4] and Novikoff et al.[5] observed some irregular-shaped membrane structures with mitochondria, ribosomes and ER. These structures were found to increase with increasing stress and chemical treatments. Not many days later, the existence of autophagosome was brought to light by the experiment of A U Arstila and B F Trump [6]. Riding on these facts, de Duve presented a report at the Ciba Foundation Lysosome Symposium held in London in 1963 [7]. Entitled ’The concept of Lysosomes’, the report delineated endocytosis and exocytosis and determined the functions of lysosomes on heterophagy and autophagy. The concept of autophagy was proposed for the first time, which was described as the degradation

of cytoplasm and organelles via autophagosomes. Digging it deeper, de Duve and Wattiaux published a review in 1966, where they predicted the function of autophagy and stated that through fragmentation and self-digestion, cells use autophagy when there is nutrient deficiency or self-cleaning of dead cells is required. In 1976 Christian de Duve, due to his discovery of lysosome and peroxidases, was awarded the Nobel prize for Physiology or Medicine.

Since its discovery, the process of autophagy has passed through numerous glorious years where it even found a noble prize given to two of its investigators. Below, the timeline of autophagy is provided, followed by a detailed description of the events that complete autophagic flux.

Autophagic research can be divided into two parts. From its discovery to 1992, the first part can be termed the "pre-ATG" era. The years 1992-till now form the "post-ATG" era. Before the advent of ATG, autophagy was typically explored using biochemical, cellular physiological, and ultrastructural methods. In this era, in the 1960-the 80s, researchers investigated the association of protein turnover in lysosomes. For instance, U. Pfeifer et al. investigated the autophagic turnover of long-lived proteins and revealed the association between circadian rhythm and autophagy [8, 9]. They observed that autophagy is inhibited by feeding while induced by fasting between meals, underpinning the fact that this process is strongly modulated by nutrient conditions. Their study also revealed that the liver loses 30-40% of proteins due to degradation during a starvation period of 48 hours. Two crucial findings in this era were the inhibition of autophagy by amino acids [10] and insulin [11]. The latter established the hormone as a legitimate physiological inhibitor of autophagic degradation and also served as a tool for calculating the average lifetime of the autophagic vesicles based on their rate of demise. Another notable feat was achieved by Seglen et al. [12] when they reported that 3-methyladenine could specifically inhibit autophagy without affecting protein synthesis or intracellular ATP levels. This has become a classic autophagy inhibitor since then. Again, certain types of proteins can be directly sequestered from the cytosol to the lysosomal membrane for destruction without the aid of autophagosomes. This type of autophagy, as already mentioned, is called Chaperone-mediated autophagy and was discovered by J. Fred Dice [13]. In this era, however, the genes and proteins specifically associated with autophagic processes remained unidentified.

A quantum leap in autophagy research came from a yeast genetic screen performed by Yoshinori Ohsumi, marking the start of the post-ATG era. In 1992, Ohsumi et al. discovered that nutritional shortage triggered an autophagic breakdown in *Saccharomyces Cerevisiae* [14]. This discovery substantially benefited autophagy research. However, the phenotypes as well as the physiological roles of the autophagy-deficient mutants were yet to be understood at that time. In 1993, Ohsumi and Tsukada discovered Apg1, the first autophagy-deficient mutant, by using light microscopic selection to acquire mutants, which, under nitrogen starvation conditions, fail to accumulate autophagic bodies [15]. Although this mutant thrived properly in a nutrient-rich environment, it perished after prolonged nitrogen deprivation. Using this phenotype as a preliminary screen, they discovered approximately one hundred autophagy-deficient mutants, and their genetic analysis showed fifteen complementation groups, leading to the discovery of several autophagy-related genes, coined as APG. Within a span of a few years, various autophagy-related genes were discovered by employing independent yeast genetic screens, which elongated the list of Ohsumi's APGs. For instance, Michael Thumm et al. isolated six autophagy-deficient aut mutants in 1994 by employing antibody staining to identify colonies with impaired cytoplasmic enzyme degradation [16]. Due to its comparable membrane properties, the Cytoplasm to Vacuole Targeting (Cvt) pathway has been explored extensively as a model for selective autophagy [17]. In 1995, D. Klionsky and his group isolated several Cvt mutants defective in this pathway, which were shown to be predominantly allelic to APG mutants [18]. Various other labs also worked on this domain in different species of Yeast [19–22]. For example, in *Pichia pastoris*, William A. Dunn et al. [21] showed that glucose-induced microautophagy needs the alpha-subunit of phosphofructokinase, while Suresh Subramani et al.[20] investigated the Peroxisome Degradation by Microautophagy. There was a high overlapping between the autophagy-related genes identified in these studies. However, ascribing mutants their own names created a confusing circumstance. For example, APG1, PAZ1, AUT3, GSA10, CVT10, and PDD7 all refer to the *Saccharomyces cerevisiae* gene now known as ATG1. To mitiigate this, later in 2003, all these genes were brought under the term "autophagy-related genes" (ATGs) [23]. Amidst this buzz of APGs, in 1994, Rapamycin silently entered autophagic research when Meijer et al. discovered that Rapamycin attenuated the inhibitory impact of amino

acids on autophagic proteolysis [24]. This was indeed a crucial discovery as the mammalian target of Rapamycin (mTOR) was later found to play a quintessential role in autophagy.
The ATG proteins work in several functional units. Numerous studies looking into these relationships have discovered multiple conjugate systems. In 1998, the first such system, the Atg12 conjugation, was discovered by Ohsumi et al. [25] which was followed by the Atg8 conjugation system by the same group in 2000 [26].
In the interim, autophagy translational research was also taking shape. Marking as the first instance, in 1999, Beth Levine et al. described the involvement of beclin1 [27] in cancer. They stated that it has a reduced expression in breast cancer and can inhibit tumorigenesis. In the following years, numerous other studies have reported the translational landscape of autophagy, and its role in diseases like Crohn's disease [28], Aging [29], neurological diseases [30] etc. were identified. These studies were assisted by the discoveries like Atg14-PI3K complex [31], the role of TFEB in autophagy [32] etc. The timeline of crucial events of autophagy is shown in **Figure 1.1**.

1.1.2 Bird's eye view to the process of autophagy

Due to the importance it carries, the process of autophagy remains at the crossroads of multiple biological pathways and processes. As a result, autophagy is orchestrated by the synergistic association of a large number of proteins. Each of these steps is modulated by some specific set of proteins. The core autophagic genes, ATGs, are involved throughout this process. Based on their functions, they can be placed in five multifunctional modules. In yeast, these modules are shown in **Table 1.1**. Except for red algae, these core molecules are conserved across all the species. Autophagy is a multistep process that includes five sequential stages. These are i) initiation, ii) double-membrane nucleation, iii) phagophore elongation, iv) cargo sequestration, and v) degradation. The initiation of autophagy is dependent on the protein mTOR. It is comprised of two functionally and structurally distinct conserved protein complexes, mTORC1 and mTORC2. Nevertheless, only the former is susceptible to Rapamycin [33] and serves an unquestionably vital function in autophagy. Although mTORC2 inhibition was also reported to induce autophagy under fasting conditions, this is predominantly medi-

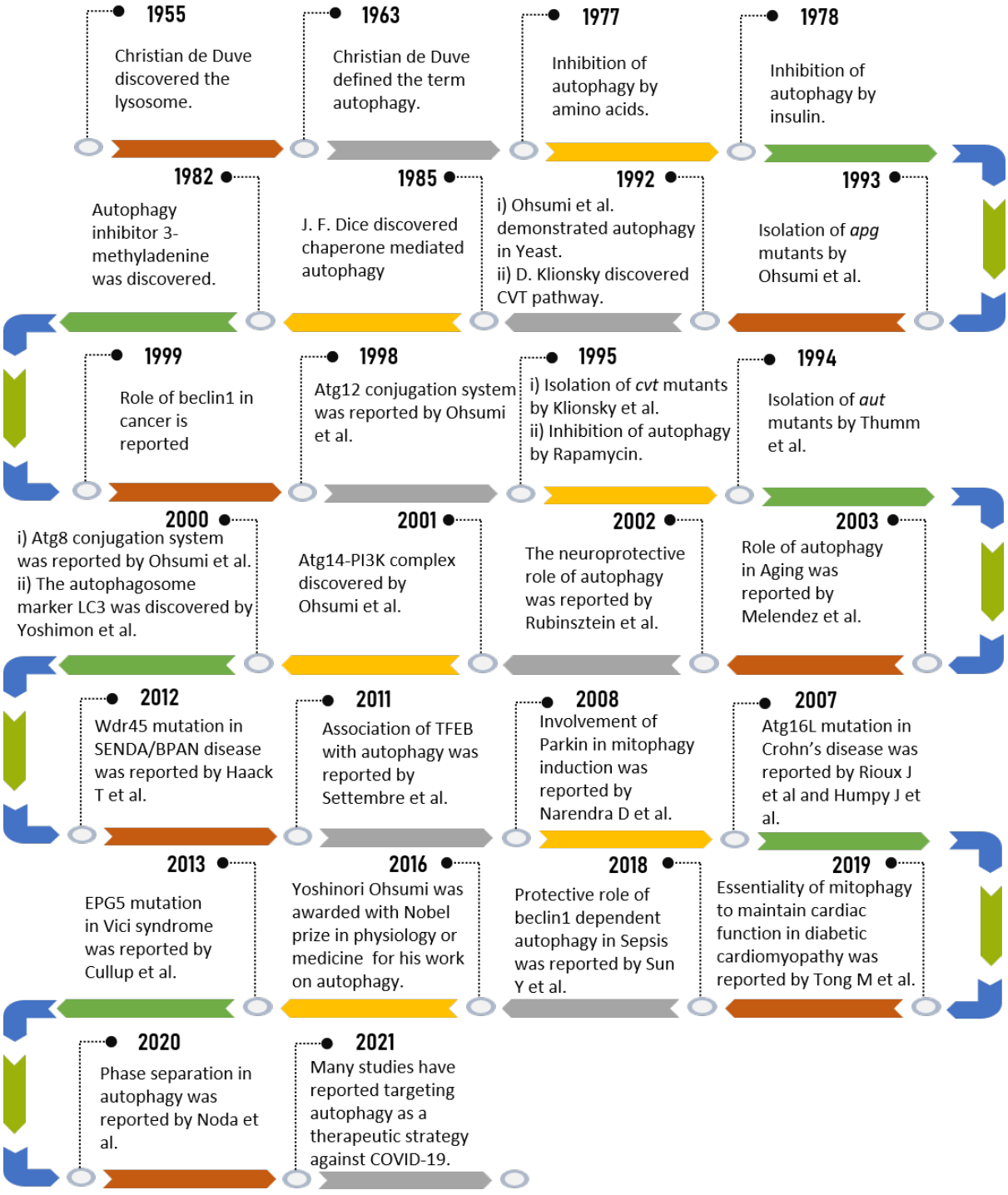


Figure 1.1: **The timeline of autophagy.**

Table 1.1: **Distribution of the autophagy related genes (ATGs) accorss five disticnt modules in yeast.**

Module	ATGs
Atg8-phosphatidylethanoamine and Atg5-Atg12 conjugation system	Atg3-5, Atg7-8, Atg10, Atg12, Atg16
Atg1 kinase complex	Atg1, Atg13, Atg17, Atg29, Atg31
PI3K complex	Atg6, Atg14, Atg38, Vps15, Vps34
Atg2-Atg18 complex	Atg2, Atg18
Atg9 vesicles	Atg9

ated via FoxO3 [34, 35]. Nutrient abundance or growth factor signalling prompts the lysosome translocation of mTORC1, where it activates and initiates growth-promoting processes while suppressing autophagy. It accomplishes this by inhibiting the autophagy initiation complex [36] and the nuclear translocation of the TFEB, which regulates the transcription levels of a plethora of lysosomal and autophagy genes [32]. In contrast, starvation disassociates mTORC1 from lysosomes, resulting in the induction of autophagy [36]. However, after prolonged starvation, mTORC1 becomes reactivated, and forms proto-lysosomal tubules and vesicles that eventually mature into functioning lysosomes [37]. In other words, autophagy walks the tightrope between mTORC1 and lysosomes for a steady autophagic flux.

Initiation of autophagy necessitates the ULK1 complex composed of FIP200, ATG13 and ATG101, which is positively regulated by AMPK and negatively regulated by mTOR [38]. Following autophagy induction, the ULK1 complex translocates to autophagy initiation sites and regulates the recruitment of the vacuolar protein sorting 34 (VPS34) complex composed of the VPS34, beclin1, VPS15 and ATG14-like (ATG14L). This complex is responsible for producing the phospholipid phosphatidylinositol 3-phosphate (PI3P) at the autophagosome forming site called the phagophore [39]. Multiple autophagy proteins have been found to bind beclin1 to activate or inhibit the beclin1/PI3KIII complex. For instance, AMBRA1 binds directly to beclin1 and enhances its interaction with VPS34, resulting in higher activation of VPS34 and the production of autophagosomes [40]. The autophagy process encounters its final obstacle at the fusing of autophagosomes and lysosomes, where a significant energy barrier must be over-

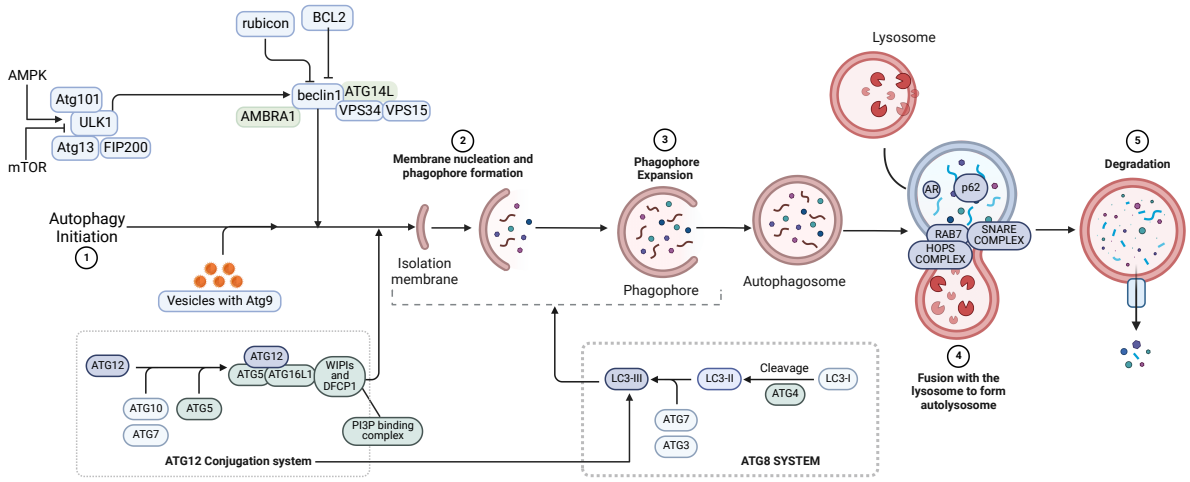


Figure 1.2: **The schematic diagram of autophagy.** The primary regulators of autophagy are AMPK and mTOR, with the former serving as an inducer and the latter as an inhibitor. The process is completed in five steps and the the core autophagic genes, ATGs play vital role in different stages by forming into several complexes.

come. This fusion mechanism is tightly regulated by either STX17-SNAP29-VAMP7/VAMP8 [41] or STX7-SNAP29-YKT6 SNARE complex [36]. Again during fusion, the two vesicles must be kept close for which HOPS complex, PLEKHM115, and EPG5 simultaneously interact with proteins present on the autophagosomal and autolysosomal membrane. PLEKHM115 binds with Arl8b^{ase} and RAB7^{ase} on the lysosome and LC3 on the autophagosome, EPG5 binds to RAB7^{ase} and LC3[42], and the HOPS complex interacts with lysosomal Arl8b^{ase} [43] and autophagosomal Qa-SNARE STX17 [44]. Once the autolysosome forms, the inner autophagosomal membrane degrades, and more than 60 lysosomal hydrolases work simultaneously to digest the confined material [45]. The molecular mechanism of autophagy is shown in **Figure 1.2**.

1.1.3 Autophagy research in numbers: how much have we dug?

We obtained articles from PubMed that matched the search word "autophag*" within the Title/Abstract field. A few publications discussed the term "autophagia" rather than autophagy/autophagosome and were therefore excluded. We collected 69405 items in the end [1]. In accordance with Mizhushima et al.[46], we classified these publications into three time periods: 1965 to 1975,

[1] At the time of compiling the thesis.

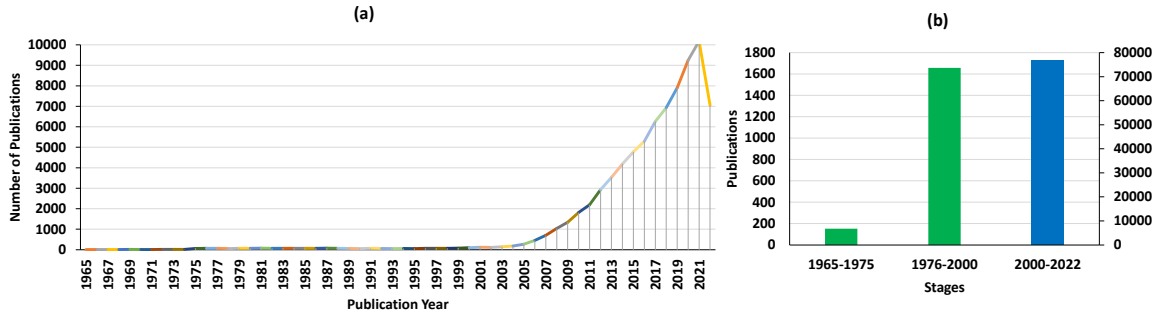


Figure 1.3: **The fast emerging study field of autophagy.** a) The number of publications in autophagy as per the PubMed records. b) The number of publications on autophagy in three different periods. It can be seen there has been a humongous rise in autophagy research in the last few years.

1975 to 2000, and 2001 to the present. In the first period, an average of 15.20 articles were published, followed by an average of 165.7 articles in the second period, and a whopping 7665.5 articles since 2001, establishing the fact that in recent years, there has been a meteoric rise in the number of articles published about autophagy. The number of publications in autophagy per year is shown in **Figure 1.3**.

1.1.4 Autophagy and diseases: despondency or hope?

The enormous number of publications on autophagy meant that it has been plotted, pieced, and ploughed into the landscape of translational science to find its impact on multiple diseases. Despite that, whether autophagy is despondency or hope is still unclear for many diseases. The crucial role of autophagy in diseases can be attributed to multiple facts. For instance, this is a quintessential process to maintain cellular homeostasis, which it accomplishes by aiding the breakdown of long-lived or misfolded proteins, damaged organelles, and protein aggregates. Again, autophagy's ability to orchestrate various stress responses is another crowning achievement that keeps it at the crossroads of multiple diseases and disorders **Figure 1.4**. This complex and paradoxical role of autophagy in regulating the course of the disease has been the subject of extensive research. Autophagy acts as a Janus in cancer by playing a role both in tumour suppressor and activator. The assessment of the ultimate fate of tumour cells by autophagy relies on the type, stage, and genetic context of cancer. Autophagy provides a cytoprotective impact by eliminating unwanted materials, thereby minimising the genomic damage leading to

abnormal mutations and cancer. However, when cancer advances, the stress-suppressing features of autophagy are co-opted by tumour cells to satisfy the increased metabolic demands required for tumour survival and rapid growth. Autophagy-related proteins are associated with preventing cancer cell growth in various cancers, including the colon, gastric, breast, lung and prostate cancer [47–50]. However, autophagy also helps in tumorigenesis by promoting the proliferation of cancer-cell and tumour growth [51, 52]. Abnormal lipid metabolism and the excessive accumulation of triglycerides stored in lipid droplets trigger the non-alcoholic fatty liver disease, which may eventually lead to non-alcoholic steatohepatitis (NASH) [53]. *In vitro* and *in vivo* studies have revealed that autophagy plays a protective role in NASH by selective degradation of these lipid droplets [54]. Hence, the autophagy pathway can be a potential target in treating NASH. In various neurodegenerative diseases, including Parkinson's disease, Alzheimer's disease, and Amyotrophic lateral sclerosis [55], misfolded protein accumulation is considered a pathological hallmark. Since the accumulation of misfolded proteins is directly affected by a decrease in the neuronal autophagy level, autophagy is considered a target pathway in neurodegenerative diseases. The importance of autophagy can be mapped to insulin resistance and type 2 diabetes, as it plays an indispensable role in the physiology of beta cells. Autophagy takes part in the regulation of insulin homeostasis and is necessary for normal beta cell homeostasis [56, 57]. The disrupted autophagic activity has been reported in the beta cells of type 2 diabetes mellitus (T2DM) patients [58]. Metformin has been widely used in type 2 diabetes clinical therapy and protects pancreas beta cells from injury through autophagy activation by the AMP-activated protein kinase (AMPK) pathway [59]. Due to its crucial role in cellular housekeeping, autophagy also plays a role in anti-ageing mechanisms [60]. It also plays an essential role in cell remodeling during development [61] and in cellular defense against pathogens [1].

Nevertheless, despite playing a protective role in various diseases, uncontrolled autophagy may lead to excessive degradation of the cellular constituents and may cause cell death [62–64]. Hence, although important, the autophagy process needs to be strictly monitored for the smooth functioning of the cellular homeostasis [65, 66].

1.1.5 Autophagy as a therapeutic target: cancer as an example

In light of the aforementioned facts, pharmaceutical methods to upregulate or inhibit this system are currently garnering substantial interest. As a physiological process, autophagy maintains the health of cells and inhibits carcinogenesis, but it can also contribute to the treatment of numerous diseases by delivering nutrients or initiating cell death. Therefore, the strategy to target autophagy with pharmaceutical intervention must be cautious and thoroughly supported by preclinical facts pertaining to autophagy's function and state. The majority of cells treated with chemotherapy induce autophagy. This autophagy activation is a last-ditch effort by tumour cells to survive. Various studies have reported that inhibiting autophagy sensitises cancer cells to anti-cancer drugs. In ER-positive breast cancer cells, for example, the genetic deletion of Atg5, Atg7, or beclin1 reverses tamoxifen resistance [67]. The efficacy of chemotherapy in HER2-positive breast cancer cells can also be enhanced by combining inhibitor 3-MA, and trastuzumab [68]. However, currently, CQ/HCQ, which inhibits the autophagosome-lysosome fusion by disrupting lysosomal acidification, is the only FDA-approved autophagy inhibitor for clinical trials [69, 70]. A list of autophagy modulators which are currently in clinical stages for targeting cancer is shown in **Table 1.2**.

Table 1.2: **Cancer-targeting autophagy modulators currently in clinical trials.**

Clinical trials	Mechanism/target of chemotherapy drugs	Current state	Type of cancer	Identifier
Sirolimus or vorinostat with HCQ (Phase I)	mTOR and HDAC inhibitor	Completed	Advanced solid tumors	NCT01266057
Vorinostat with HCQ (Phase I)	HDAC inhibitor	Ongoing	Advanced solid tumors	NCT01023737

Table 1.2 continued from previous page

Clinical trials	Mechanism/target of chemotherapy drugs	Current state	Type of cancer	Identifier
Gemcitabine/ abraxane with HCQ (Phase I/II)	Nucleoside analog/ antimicrotubule agent	Completed	Pancreatic adenocarcinoma	NCT01506973
Paclitaxel, carboplatin and bevacizumab with HCQ (Phase II)	Microtubule disrupting agents/ inhibitor of DNA synthesis/ VEGF-A inhibitor	Completed	Non-small lung cancer	NCT01649947
FOLFOX and bevacizumab with HCQ (Phase I/II)	Folinic acid/ thymidylate synthase inhibitor/ inhibition of DNA synthesis/VEGFA inhibitor	Completed	Colorectal cancer	NCT01206530
Vorinostat with HCQ versus regorafenib (Phase II)	HDAC inhibitor/ Multi kinase inhibitor	Completed	Refractory metastatic colorectal cancer	NCT02316340
MLN9708 and vorinostat (Phase I)	Proteasome inhibitor/ HDAC inhibitor	Completed	Advanced p53 mutant malignancies	NCT02042989
Temsirolimus with HCQ (Phase I)	Cell cycle arrest	Completed	Refractory solid tumors	NCT00909831

Table 1.2 continued from previous page

Clinical trials	Mechanism/target of chemotherapy drugs	Current state	Type of cancer	Identifier
CQ (Phase I/II)	Lysosomal inhibitor	Completed	Breast ductal carcinoma	NCT01023477
RAD001 with HCQ (Phase I/II)	mTOR inhibitor	Completed	Renal cell carcinoma	NCT01510119
Navitoclax and abiraterone acetate with or without HCQ (Phase II)	Bcl2 inhibitor/ Androgen synthesis inhibitor	Completed	Refractory prostate cancer	NCT01828476
Sunitinib malate with HCQ (Phase I)	Receptor tyrosine kinase inhibitor	Completed	Advanced solid tumors	NCT00813423
CQ (Phase II)	Lysosomal inhibitor	Completed	Breast cancer	NCT02333890
Sorafenib with HCQ (Phase I)	Multi kinase inhibitor	Completed	Refractory or relapsed solid tumors	NCT01634893
Dabrafenib and trametinib with HCQ (Phase I/II)	BRAF inhibitor/ MEK inhibitor	Completed	Advanced BRAF mutant melanoma	NCT02257424
Velcade and cyclophosphamide with CQ (Phase II)	Proteasome inhibitor/ DNA replication inhibitor	Completed	Refractory or relapsed multiple myeloma	NCT01438177

Table 1.2 continued from previous page

Clinical trials	Mechanism/target of chemotherapy drugs	Current state	Type of cancer	Identifier
Cisplatin and etoposide with CQ (Phase I)	DNA replication inhibitor/ topoisomerase inhibitor	Completed	Stage 4 Small cell lung cancer	NCT00969306
MK2206 with HCQ (Phase I)	Akt inhibitor	Completed	Advanced solid tumors (Prostate, melanoma or kidney cancer)	NCT01480154
Enzalutamide with metformin hydrochloride	Anti-androgen	Completed	Prostate cancer	NCT02339168
Gemcitabine and abraxane with or without HCQ (Phase II)	Nucleoside analog/ Antimicrotubule agent	Completed	Pancreatic cancer	NCT01978184

1.1.6 The need for a helping hand

The question arises whether the rising techniques and technological advancements for monitoring autophagy are sufficient to reveal its potential. Can it address all the ifs and buts in drug discovery? Also, what does the practicality indicate about investigating the effect of each protein involved in this process?

The answer is no. Due to the kaleidoscopic nature of autophagy, this process needs to be investigated through various methods. Many of these may not be feasible in the first place from the experimental perspective. Addressing the crossover between intracellular mechanisms and tissue-level phenotypes is one example, as it necessitates the computation of the probable

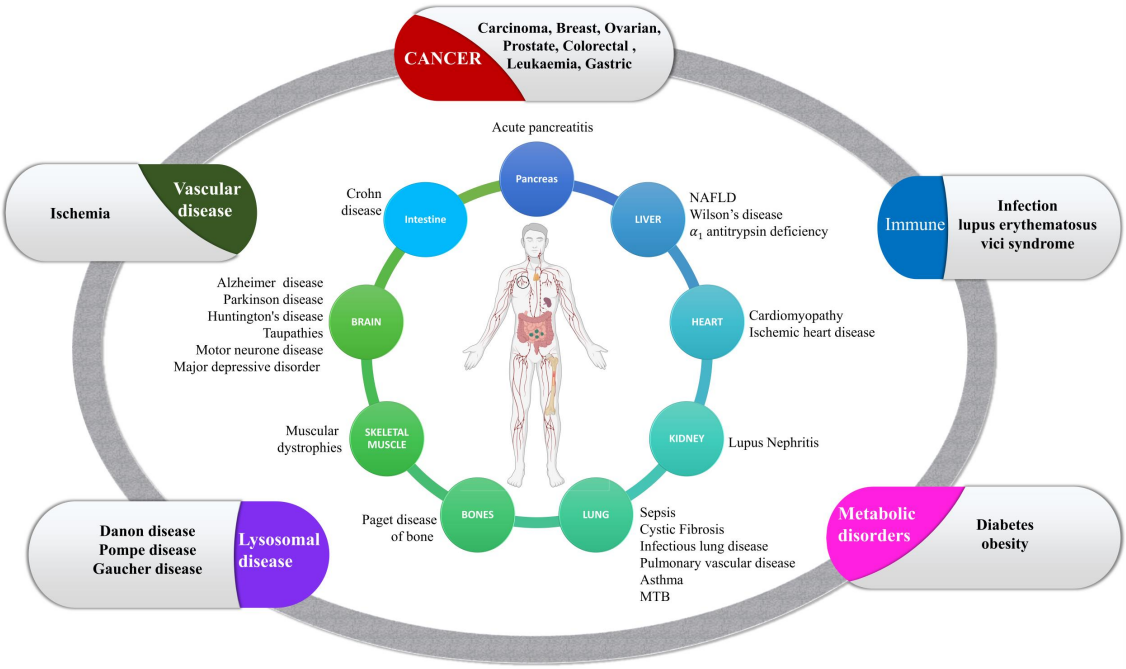


Figure 1.4: **Autophagy-related diseases.** These diseases can be categorised into two parts: organ-specific (shown in the inner circle) and multisystemic (shown in the outer circle). This figure shows the autophagy-disease interplay. The association with most human diseases, including varieties of cancer and immune disorders, has proved that autophagy is a quintessential process, and its manipulation can be targeted as a therapeutic strategy.

system-level impacts of the behaviours of large sets of individual genes or proteins. This necessitates the integration of an approach that, through its sophisticated tools and methods, can integrate quantitative technologies and extensive experimental measures to reveal the intracellular heterogeneity as well as the rapid adaptation of the cellular environment in response to stimuli that eventuate on timescales that cannot be described by evolutionary processes or clonal selection and this is where systems biology enters the autophagy scenario. With its potential to analyse a system both as a whole and in parts, systems biology unveils the curtains of the autophagy mechanism and helps it conquer the pesky ladder in the chasm of uncertainties.

1.2 Systems biology: The key to identify the wolves in the sheep's clothing

Comprehending at a systems level has always been a perennial theme in biological science. This comprehension extends beyond the assembly of genes and proteins. According to Hiroaki

Kitano [71], the interconnections between proteins and genes are just static road maps, and obtaining a complete inventory of them is equivalent to getting the parts of an aeroplane. They are necessary but cannot explain the system's fundamental complexity. For that part, it is an unmet need to examine how individual components interact during a process. These interactions are not random. There exist directions to avoid dejections, variations in interaction partners for task specificity, and flexibility to function within the cell, among cells and even between organisms. Biological systems are governed by specific rules, and systems biology acts as a hatchet to unveil these underlying principles [72]. Though Mihajlo Mesarovic is often credited with coining the term "systems biology" in 1968 [73, 74], many others believe that Ludwig von Bertalanffy, the "father of general systems theory," actually coined the word in the 1920s. Some of the crucial inventions aiding systems biology research are shown in **Figure 1.5**.

Every system possesses a hierarchical structure, and a systematic study of it helps to find how components are organised, viz., what lies in the core, and what remains on the periphery of the system. Again, these structures are interlinked together, where each lower level in the hierarchy creates the level immediately above (for example, cell to tissue, tissue to organ, organ to the organ system and so on) by means of some linkages. Systems biology is nothing but the study of both these structures and their linkages. By analysing systems, "layer by layer," systems biology allows for understanding why and how an event occurs, inevitably leading to 'what if' type questions and enabling predictions [75]. In other words, systems biology can grasp all the components of a biological system and, through quantification of these components and their relationships, it endeavours to provide a comprehensive model of the system. Two crucial pillars of systems biology are mathematical modelling an network analysis. The former looks into the association of a small number of proteins and the effect of their crosstalk on the systems dynamics while the latter investigates the entire system as a whole. **Figure 1.6** summarises systems biology, while the next section describes the two pillars in detail.

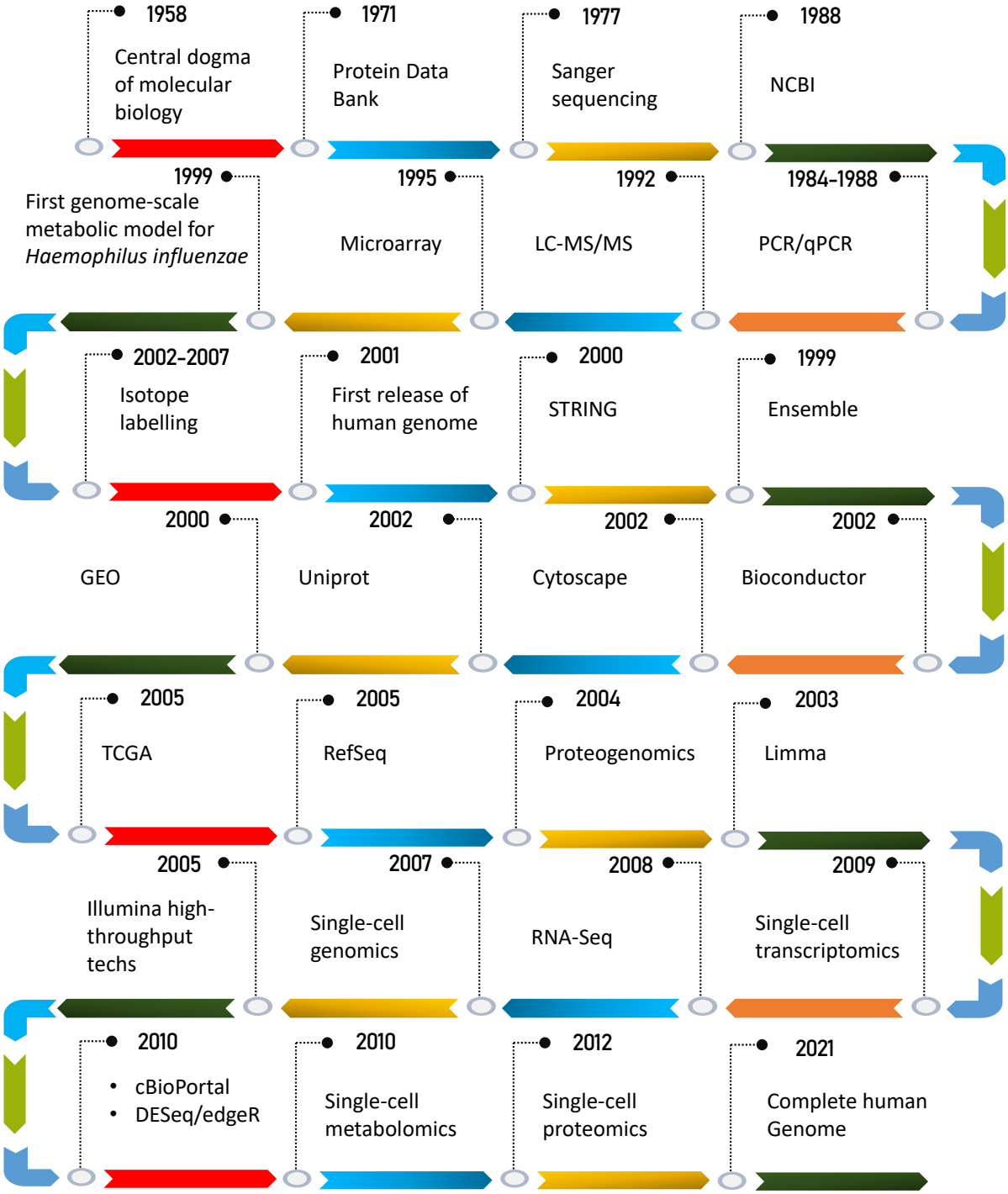


Figure 1.5: **Crucial inventions in systems biology.**

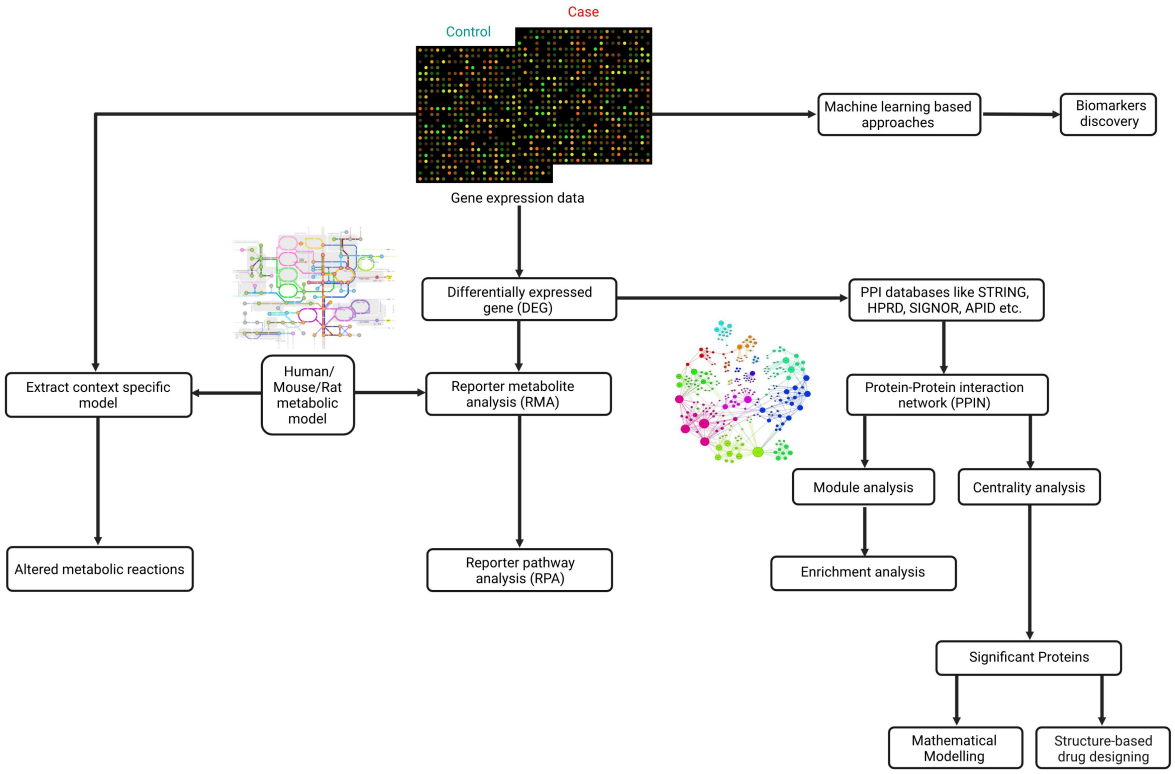


Figure 1.6: **An illustration of the systems biology approaches**. ML-based approaches can be used directly in the transcriptomic data of patients. The dataset can also be used to understand the metabolic alterations in the disease. By identifying differentially expressed genes and using various databases, protein-protein interaction network analysis can be carried out to identify the crucial proteins. These proteins can further be taken under the grasp of mathematical modelling studies to get more insight into the biological system. The structure-based drug-designing approach can also be applied to get more insight into these crucial proteins.

1.2.1 Mathematical modelling

1.2.1.1 Significance of mathematical models in studying biological systems

Biological systems are complex, and a mathematical model acts as a reflector of this complexity. The complexity of a biological system appears due to many reasons. One of them is the underlying hierarchy of the system that ranges from cells to the organism level. Each of the different hierarchy levels is dynamic. Even if they imitate regular and predictable behaviour, abrupt major and stochastic shifts may arise anytime, even for minute changes in the cellular environment. These uncertainties lead to complexities that are difficult to capture in experimental studies. Even if they do, it is challenging to grasp routes and patterns of the actual evolution of complexity. For example, to identify the functional role of a gene, scientists perform knockout experiments [76]. However, such experiments overemphasise the role and importance of a single gene and are not ideal for understanding the complex nature of the system. Mathematical models study these complexities of a system and portray an abstraction of reality. These models study the crucial genes/proteins capable of driving the system and the underlying dynamics of the association between the genes/proteins. These models have the audacity to identify crucial parameters capable of deciding the system's fate, which can be further proven by experimental validation.

The mathematical modelling is based on four crucial pillars where the first pillar is a literature survey of the system. The second pillar is the construction of the model, where the relationship between the model variables will be established using model parameters. This step is followed by the analysis of the model, and the last pillar is the validation part where the result of the model will be validated either by literature or by an experimental approach. If the model fails to deliver the appropriate output, the necessary changes will be implemented in the model until the desired outcome appears. Mathematical models are perfect examples of complexity and simplicity as they are ornamented with a set of equations, which are complex enough to replicate the properties of the system and concurrently simple enough to grab up the underlying phenomena of the system. Theoretically, these models can drive a system anywhere, but it has to follow some constraints in systems biology. For example, a negative concentration of a protein will make no sense, so as a species' negative population. Similarly, there must always be

an upper bound, be it the concentration of a protein inside a cell or the population of a species. But, although restricted to biological constraints, a mathematical model can help to find out the crucial parameters responsible for deciding the fate of the system. In other words, for a specific cellular process full of many regulatory patterns, mathematical modelling paves the way to pick the right one.

1.2.1.2 Different type of mathematical models

Differential equations are extensively employed to represent biological systems due to their effectiveness at capturing non-linear behaviour and illustrating interactions between several variables. Among the various models available to investigate biological processes, ordinary differential equations-based models are the most commonly used. These equations are of the form

$$\frac{dx_i}{dt} = f_i(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_k, t), i = 1 : n,$$

and they describe how individual variables x_i vary over time, λ_k are the model parameters describing different rate constants. When explicit dependence of the dependent variables on the independent variable is absent, the system is called autonomous. The functions, along with the parameters, beautifully portray the dynamics of the state variables. Such models usually describe the change in the dynamics of a variable (proteins/cells etc.) without considering factors like the noise, transversal and longitudinal diffusion of the model variables, etc.

Another type of mathematical model, called the delay differential equations (DDE) model, is used in representing biological phenomena. Time delay is an inherent property and occurs naturally in biological systems. The time delays in these models represent the duration of hidden processes between two major processes. The time between the infection of a cell by some pathogens and the subsequent production of new pathogens can be taken as an example. Another example is a susceptible population which requires a consistent period of time, termed the latent period of infection, to become infectious after coming into touch with an infected individual. This phenomenon is even reflected in the central dogma of biology, which states that the production of functional proteins is the result of a series of complex processes involving transcription, translation, and post-translational modifications. Due to the sequential

nature of protein synthesis, there is a delay between when RNA polymerase binds to promoter DNA and when fully functional proteins appear. Such phenomena are captured using the delay differential equation models. A typical delay differential equation is of the form

$$\frac{dX}{dt} = F(t, X(t-\tau_1), ..., X(t-\tau_n)); \tau_i, i = 1 : n \geq 0,$$

where $\tau_i, i = 1 : n$ are time delays. They are measurable and may be constant. Sometimes, the initial or boundary conditions may not be sufficient to predict the future state of a system. For such a scenario, it is indispensable to know how the system behaved in the early stages, and hence, delay differential equations play a vital role in understanding a biological system where the current state of some variables depends on the past states.

Both ODE and DDE are deterministic methods to map a biological system of equations. But, a biological system is always exposed to uncertainty that is not entirely understood. An approach to model such systems is by adding stochastic influence or noise. It is another inherent property of biological systems. Since biochemical kinetics at the single-cell level is inherently stochastic, stochastic models are required to adequately reflect the numerous sources of heterogeneity required for the realistic modelling of biological systems. Such models, however, are far more computationally intensive than deterministic models and significantly more difficult to fit into experimental data. Some stochastic behaviour examples are hormonal oscillations, respiration, blood pressure variations, cellular metabolism, etc. The general form of a stochastic differential equation (SDE) can be expressed as

$$dx_t = f(t, x_t)dt + G(t, x_t)dw_t,$$

or with the equivalent integral form

$$x_t = x_{t_0} + \int_{t_0}^{t} f(s, x_s)\, ds \; + \int_{t_0}^{t} G(s, x_s)\, dw_s,$$

with an initial value, x_{t_0}. Here, $f : [t_0, t] \times R^d \to R^d, G : [t_0, t] \prod R^d \to R^{d \times m}$ and $\{w_t\}_{t \in [t_0, t]}$ denote an m-dimensional Wiener process (Brownian motion). R^d and $R^{d \times m}$ are d-dimensional

and $d \times m$-dimensional Euclidian space, respectively.

Agent-based modelling (ABM) is an alternative approach that relies on a predefined logical programming language. In ABM, the system consists of interacting autonomous decision-making bodies known as agents. ODE modelling presupposes a homogeneous environment, while ABM is capable of simulating a transient and spatial evolution of a system in that each participant in the model is represented as an individual agent per its laws. One of the fundamental aspects of ABM is the occurrence of complex behaviour from a set of simple rules. It simulates the interactions between multiple independent agents and evaluates their effect on the overall system. It captures the emerging phenomena of a complex system from the perspective of its constituent components, making ABM a bottom-up approach [77]. The benefits of ABM include their flexibility, the natural way of description of the system, and the capability of capturing the emergent phenomena due to the interactions of individual entities [77]. Agent-based modelling facilitates both discrete and continuum mathematical modelling approaches. The study of tumour cell density, nutrient distribution, etc., comes within the radar of the continuum modelling approach, whereas cellular automation is a representation of discrete mathematical modelling. Agent-based models have been used extensively to explain biological phenomena in various biological systems. For example, a three-dimensional agent-based Voronoi-Delaunay hybrid model was developed by Schaller, and Meyer-Hermann [78], where reaction-diffusion equations depicted the spatiotemporal distribution of oxygen and glucose. Their study was an effort to test the hypothesized functional dependence of the absorption rates of glucose and oxygen, and to determine suitable mechanisms for necrosis induction. Another agent-based model was built by Engelberg et al. [79], where different spaces for tumour cells, oxygen, nutrient, and toxic inhibitors were considered. The goal of the study was to create a model consisting of separate cells that fairly represent the behaviour of an in vitro multicellular tumour spheroid.

Biological processes involve complex mechanisms with many pathways and molecules that change over time and space, and in the understanding of such systems, ABM would play a vital role. These models can also help with the mathematical portraiture of biological phenomena like the spatial and temporal requirement of autophagy-related protein to bacteria. However, ABM has certain drawbacks. For instance, it demands more details to be provided about the

system of interest, which may not always be reported in the literature. Another disadvantage of ABM is that it is more computationally expensive than partial differential equations (PDE) or ordinary differential equations.

Petri net is the creation of Carl Adam Petri in his doctoral dissertation [80]. It is constructed using two types of nodes, viz. places, depicted as circles, and transitions, represented as narrow black rectangles. In systems biology, places refer to chemical species such as metabolites, proteins, enzymes, DNA, RNA, etc., and transitions refer to chemical reactions such as activation, inhibition, phosphorylation, etc. Nodes are connected by arcs, which may only be directed from place to transition (input arcs) or transition to place (output arcs). A Petri net is always bipartite. The stoichiometry of a reaction is indicated by the weight of the arc. Although initially designed to model only discrete processes, improvements have been made in Petri nets to deal with a continuous process [81, 82] . Literature has witnessed many applications of Petri nets to different biochemical systems. For example, Koch et al. [83] built a Metabolic Petri net (where the places represent metabolites and the transitions represent the biochemical reactions between metabolites) consisting of 17 places and 27 transitions that qualitatively modelled the carbon metabolism in the potato tuber. Using this Petri net model as an example, the author has provided a method for model validation of metabolic networks using Petri net. Signal transduction pathways are commonly modelled with a set of ordinary differential equations, but unknown parameter estimation is a problem inherent in ODE modelling. To deal with this problem, Sackmann et al. [84] implemented the Petri net theory to model and analyse signal transduction pathways. The authors put forward a systematic model validation method for signal transduction pathways that depends only on the network structure. This method is then illustrated using the mating pheromone response pathway in *Saccharomyces cerevisiae*. Petri net is advantageous in the absence of quantitative data. So, in a field like autophagy, where a lot of pathways are involved, the Petri net model would play a vital role. However, it has the limitation that it will not capture the mechanism that one can obtain with the help of differential equation-based models.

1.2.1.3 Tools and packages

Different tools and packages have been built across multiple platforms (MATLAB, Python, R, etc.) to support mathematical modeling [85–91]. A structural diagram editor, Cell Designer [92], has also been developed to draw gene-regulatory and biochemical networks to make mathematical modelling a feasible approach in systems biology. CellML is an XML-based language designed to describe mathematical models in a machine-independent form suitable for sharing between different authors and archiving in a model repository [93].

1.2.1.4 Mathematical Preliminaries

In this section, some of the mathematical-modelling related prelliminaries are discussed.

Equilibria of Ordinary Differential Equation

Let us consider a system of differential equations,

$$\dot{x} = f(x); x \in R^n. \tag{1.1}$$

Here,

$$x \in \mathbb{R}$$

$$f = (f_1, f_2, ..., f_n)^T$$

and, $f_i = f_i(x_1, x_2, ..., x_n)$.

1. The initial value problem,
   $\dot{x} = f(x)$ with $x(0) = 0$, has a unique solution if the partial derivatives of $f_1, f_2, ..., f_n$ are C^1 functions.

2. A point $x_e \in \mathbb{R}$ is an equilibrium point of (1.1) if

$$\dot{x} = f(\bar{x_e}) = 0.$$

3. Jacobian Matrix: The Jacobian matrix, named after its developer, Carl Gustav Jacob Jacobi, contains all the partial derivatives of the first order of a vector-valued function. The Jacobian matrix of f at the equilibrium point $\bar{x}$, is the matrix of partial derivatives of 'f' evaluated at $\bar{x}$. It is given by:

$$J(\bar{x}) = \begin{vmatrix} \frac{\partial f_1(\bar{x})}{\partial x_1} & \frac{\partial f_1(\bar{x})}{\partial x_2} & ..... & \frac{\partial f_1(\bar{x})}{\partial x_n} \\ ..... & & & \\ \frac{\partial f_n(\bar{x})}{\partial x_1} & \frac{\partial f_n(\bar{x})}{\partial x_2} & ..... & \frac{\partial f_n(\bar{x})}{\partial x_n} \end{vmatrix}$$

or, simply, componentwise, it can be written as,

$$J_{ii}(\bar{x}) = \frac{\partial f_i(\bar{x})}{\partial x_j}.$$

Stability of Equilibrium Points

Let $\bar{x}(t)$ be any solution of the equation 1.1. Then, $\bar{x}(t)$ is stable if solutions starting near to $\bar{x}(t)$ at a given time remain close to $\bar{x}(t)$ for all times. It is asymptotically stable if nearby solutions converge to $\bar{x}(t)$ as $t \to \infty$. A solution which is not stable is said to be unstable [94].
Equilibrium points of a dynamical system can be categorised based onf the eigenvalues of the Jacobian matrix. If none of the eigenvalues of the Jacobian matrix has zero real part, then the equilibrium point is hyperbolic. If all eigenvalues have negative real parts, the equilibrium point is stable. If at least one has a positive real part, the equilibrium is an unstable node.

Sensitivity analysis

Sensitivity analysis (SA) refers to a wide range of mathematical methods to measure the extent to which model output variance may be attributable to model inputs [95]. Because the output behaviour of high-dimensional systems is frequently controlled by a small number of parameters, SA provides a method for isolating these parameters so that they can be the focus of subsequent studies. SA can be implemented both locally and globally. The former is the straightforward approach, in which each parameter is varied independently while the others

remain constant. This method has the disadvantage that it cannot examine the effect of the variation of all parameters simultaneously. These methods are informative in conditions where there does not exist much uncertainty in model inputs or if the interactions between the inputs are very few [96].

Global sensitivity analysis (GSA) methods take into account the variations in model outputs when input parameters are permitted to fluctuate simultaneously within specified ranges [95]. These methods are computationally expensive but provide more information than the local SA. Two of the most commonly used GSA methods are the Partial Rank Correlation Coefficient (PRCC) and the Extended Fourier Amplitude Sensitivity Test (EFAST). The former is a sampling-based while the latter is a variance-based method. In particular, PRCCs offer a measure of monotonicity following the removal of the linear effects of all variables but one. In comparison, eFAST yields fractional variance measurements attributable to individual as well as groups of variables [95]. Ideally, both these measures should be calculated for a comprehensive and insightful study.

1.2.1.5 Limitations of mathematical modeling

Despite being an excellent approach to study biological system dynamics, mathematical modelling possesses certain limitations and difficulties. These limitations must be taken into account in capturing the characteristics of a certain biological process with the help of mathematical modelling. Equations in a mathematical model contain parameters, and mathematical models are driven by these parameters. These parameters can be determined by experimental studies. However, many parameters still remain unknown because either the relevant experimental data is not available or the parameter values obtained in the literature are not from the system addressed by the model. For example, in a lung cancer model, the rate of degradation of beclin1 is a parameter, but in literature, this parameter value is reported in pancreatic cancer. Another difficulty in mathematical modelling is the different functioning times of various components of a pathway. For example, genetic regulatory processes are caused by metabolic reactions, but while the time taken by metabolic reactions is in seconds or minutes, the regulatory processes could occur for several hours or days. A mathematical model of a biological

system should always be abided by biological constraints. The findings of the model need to be validated according to the objectives of the model. Hence, a qualitative or quantitative association of model output and biological data is very much necessary. But quantitative experimental data on the time course of interaction between model variables is often very limited. Biological systems possess hierarchical layers (cells-tissue-organs etc.). To understand a system, it is necessary to understand the dynamics of each layer. However, it is hard to model the entire system as a whole as the model formulated would be non-computable. Hence, modelling is limited to studying the system in parts that necessitate the emergence of system biology approaches like network analysis, which can study the entire system as a whole.

1.2.2 Network biology

Biological systems can be portrayed as networks, and these networks depict the physical and spatial organisation of the organism. Systems biology employs a pragmatic approach to elucidate the emergent properties of such networks with the aim of quantitative explanation and to foresee the biological processes occurring at molecular, cellular, tissue, organ, and whole-body level. It focuses on a holistic analysis of biological networks of various processes and quests for the understanding of the extent to which the intermodular connectivity modulates a biological process.

Network analysis investigates the entire system as a whole. It is like a snapshot of the entire system at a particular time, where we can see all the nodes and their interactors. In systems biology, there are various types of the network depending on the nodes studied, such as protein-protein interaction (PPI) network, where nodes are proteins, and the edges are the interaction between them; metabolic network, where the nodes are metabolites and the edges are the reactions between them; gene regulatory networks, where nodes are genes and edges are the physical and/or regulatory relationships between the genes; ecological networks, where species are nodes and edges are the interactions which can be either trophic or symbiotic. In this thesis, we are mainly focused on PPI networks.

1.2.2.1 Protein-protein interaction network

Inter- and intracellular mechanism coordination is dependent on molecular interactions. In endorsing homeostasis, these interaction events are strictly controlled within a dynamic, interconnected terrain of molecular pathways that authorise cells to execute complex processes. Defects in these molecular pathways can lead to abnormal signalling and cellular malfunction. PPIs drive a significant portion of this subcellular communication, and hence, a comprehensive understanding of the PPI network is an unmet need for a better understanding of the molecular level alterations occurring due to the progression of disease or invasion of a pathogen.

In a PPI network, two proteins are connected by an edge if an association exists between the proteins, This association can be functional (activation/inhibition) or physical (direct binding). The former is an example of a directional network, while the latter represent an undirected network. The directionality captures the regulatory effect exerted by the source protein on the target protein. Identifying key proteins in an undirected network may lead to various false-positive results. For instance, when the mode of interactions for drug-disease relationships is absent, we cannot determine if a drug heals a disease or produces one as a side effect [97].

Irrespective of its size, a PPI network always possesses a small set of core nodes, which can modulate the fate of a biological system. Distinguishing these key proteins has proven to be a daunting task, further exacerbated by the intricacy of understanding how such proteins interact synergistically. In the conventional approach, defining such drivers relies only on the topology of the PPI networks and not on their context-specificity. Methods akin to this have the drawback that only the topological properties of PPI networks alone do not capture the whole landscape of the signalling complexity. Therefore, the derived driver proteins may not be sufficient to illuminate the complexity of the mechanism of disease progression. Identifying a target necessitates causal inferences about interacting partners, which must be augmented in specific contexts with knowledge about pathways, localisations, diseases, and biological processes. Numerous databases containing the multiple information of proteins have been created to fulfil the need for the aggregation of PPI data for a more informed insight into the mechanisms of cells and diseases. A comparison of these databases is shown in **Figure 1.7**.

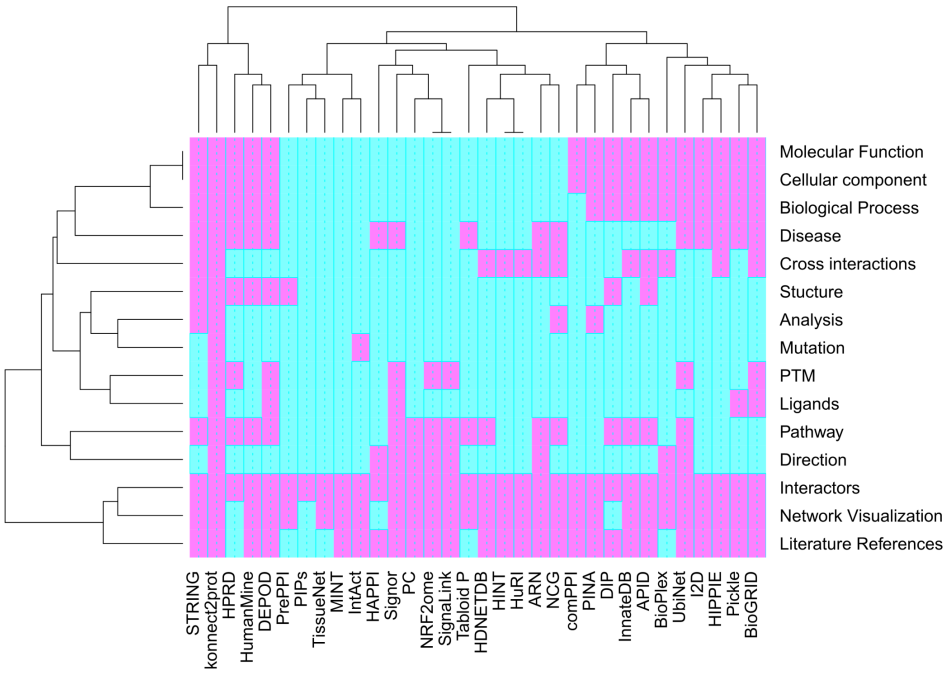


Figure 1.7: **The attribute-database heat map.** The absence or presence of an attribute in a database is represented by cyan, and violet colours, respectively.

1.2.3 Network biology glossary

Adjacency matrix:

Adjacency matrix defines the wiring diagram of a network. Let $G = (V, E)$ is a network with n nodes where V is the vertex set and E is the edge set. Then, the adjacency matrix A_{ij} is defined by

$$a_{ij} = \begin{cases} 1, & \text{if i and j are connected.} \\ 0, & \text{otherwise.} \end{cases}$$

For an unweighted undirected network, the adjacency matrix is symmetric, while in general, for an unweighted directed network, it is asymmetric. A directed graph gives a symmetric adjacency matrix if all the edges are bidirectional. In a weighted graph, the adjacency matrix values can be replaced by the weights of the corresponding edges.

Density:

The density ρ of a network $G = (V, E)$is defined as the fraction of total number of edges to total possible edges in the network. Mathematically,

$$\rho = \frac{|E|}{\frac{n(n-1)}{2}},$$

where n is the total number of nodes in the network.

Distance:

In a network, the distance between two nodes is the number of edges in the shortest path connecting them.

Diameter:

The diameter of a network is defined as the length of the "longest shortest path" between any pair of vertices. Therefore, it corresponds to the highest of all entries in the graph distance matrix.

Network module:

Modules are characterised by groupings of individual nodes that are strongly connected within and sparsely connected between [98]. Detecting modules is crucial for network exploitation since such substructures frequently correlate to essential functions. Elements of module M exhibit identical behaviour towards elements outside of M [99]. Thus, a module can be reduced to a single element without losing information about its neighbourhood and connectedness [99].

Components:

Networks may be both connected and disconnected. The former means that every node has at least one path to reach all the other nodes in the network. Otherwise, the network is referred to as disconnected. A disconnected network is comprised of numerous disjointly connected components, the largest of which is referred to as the huge component. In the former, every

node possesses at least one path to reach all the remaining nodes in the network. Otherwise, the network is called a disconnected network. A disconnected network consists of many disjoint connected components, and the largest among them is called the giant component. The robustness of a network is measured in terms of the size of the giant component.

Centrality measures:

Centrality measures identifies the most important nodes in a network. Since, the word 'important' is vague, it gives rise to many methods which have been used to find the important nodes of the system in their own way. Some of these methods include degree centrality, betweenness centrality, radiality, clustering coefficient, etc. Numerous novel algorithms have also been developed to find important nodes in a network. To better understand the core set of the nodes in a network, multiple centrality measures need to be used because a few nodes, which are central in one measure, may not be central in another analysis. Nodes central in multiple measures have a genuine capability to control the fate of a system.

Degree Centrality:

The calculation of degree centrality is one of the simplest. A node's degree centrality corresponds to its number of edges. The higher the degree, the more central the node is. In a graph $G = (V, E)$, where $|V| = n \neq 0$, the degree of a node v_i is defined as,

$$deg(v_i) = \sum_{j=1}^{n} a_{ij},$$

where

$$a_{ij} = \begin{cases} 1, & \text{if i and j are connected.} \\ 0, & \text{otherwise.} \end{cases}$$

The nodes with the highest degree are termed hubs. A PPI network possesses a small number of hubs and many poorly connected nodes. Numerous studies have demonstrated that the removal of a hub protein is more fatal than the deletion of a non-hub protein, a phenomenon called the centrality-lethality rule.

For an undirected number, degree is just the number of edges a node has. However, in a directed network, the degree of a node is the sum of the indegree and outdegree. The former measures the number of incoming edges to a node while the later measures the number of edges going awat from it, i.e., $indeg(v_i) = \sum_{j=1}^{n} A_{ij}$, $outdeg(v_i) = \sum_{i=1}^{n} A_{ij}$.

Betweenness centrality:

It evaluates the extent to which a vertex lies on the paths connecting other vertices. Mathematically, betweenness centrality (BC) of node v is given by,

$$BC(v) = \sum_{s \neq v \neq t} \frac{\rho_{st}(v)}{\rho_{st}}; s, v, t \in V,$$

where V is the vertex set, $\rho_{st}(v)$ is the number of shortest path between 's' and 'v' that passes through 'v' while ρ_{st} is the total number of shortest path between 's' and 't'.
Betweenness centrality ignores a node's degree. Instead, it investigates how much a node falls between other nodes. Therefore, a node with a low degree can acquire a high betweenness value. The largest to smallest possible betweenness value ratio in a network with n nodes is,

$$\frac{n^2 - n + 1}{2n - 1}.$$

For a large network, this value reduces to $\frac{n}{2}$. However, the betweenness centrality values in a network is often normalized by dividing it by the total number of node pairs (n^2), i.e.,

$$BC(v) = \frac{1}{n^2} \sum_{s \neq v \neq t} \frac{\rho_{st}(v)}{\rho_{st}}; s, v, t \in V.$$

Closeness centrality:

This centrality measure evaluates how close a node to other nodes in a network. For a node 'i', it is defined as

$$C(i) = \frac{n}{\sum_{j=1}^{n} d_{ij}},$$

where n is the total number of nodes in the network.

Despite being a natural centrality measure, it faces various obstacles. Frequently, the difference between the closeness values of two nodes can only be determined by observing the trailing digits. Also, the ratio between the largest and smallest closeness values in a typical network is five or less. Therefore, it is challenging to differentiate between the central and non-central nodes in the network.

Clustering coefficient:

The clustering coefficient (CC) measures a node's cliquishness or local connectivity. For example, let node 'i' is connected to two other nodes, 'j' and 'k'. Then, these three nodes will form a tuple. If 'j' and 'k' also interact, then the three nodes will create a triangle. The clustering coefficient for a node 'i' is defined as the ratio of observed triangles to all the possible triangles involving 'i', i.e.

$$CC(i) = \frac{\text{number of triangles involving i}}{\text{number of possible triangles involving i}}$$

$$\implies CC(i) = \frac{\sum_{j\neq i}\sum_{k\neq i,j} A_{ij}A_{jk}A_{ki}}{(\sum_{j\neq i} A_{ij})^2 - \sum_{j\neq i}(A_{ij})^2},$$

where A_{ij} is the adjacency matrix. It can be seen that $0 \leq CC(i) \leq 1$.

Bridging centrality:

In a graph, a bridge node is a node that connects densely connected components. The bridging centrality of a node is determined by multiplying its betweenness centrality (BC) and bridging coefficient (CB). CB(v) measures how well 'v' is positioned between the high-degree nodes and is defined as,

$$CB(v) = \frac{d(v)^{-1}}{\sum_{i\in N(v)} \frac{1}{d(i)}},$$

where d(v) is the degree of v, and N(v)={neighbors of node v}. Thus, the bridging centrality B(v) for node v is defined by:

$$B(v) = BC(v) \times CB(v)$$

$$\implies B(v) = \sum_{s \neq v \neq t} \frac{\rho_{st}(v)}{\rho_{st}} \times \frac{d(v)^{-1}}{\sum_{i \in N(v)} \frac{1}{d(i)}}.$$

Network controllability:

It evaluates the capacity of a single node to control a directed network. The control theory analysis gives the minimum number of nodes required to control the network. The time-invariant dynamics of a network of N nodes at a time can be expressed as [100],

$$\frac{dx}{dt} = Ax(t) + Bu(t). \tag{1.2}$$

Here, $x(t) = (x_1, x_2, ..., x_N)^T$ is the state vector of N nodes at time t, $A_{N \times N}$ is the adjacency matrix, $B_{N \times N}, (M \leq N)$ contains the nodes that are controlled by an outside controller, and $u(t) = (u_1(t), \ldots, u_M(t))^T$ is a time-dependent input vector which controls the network. The same signal $u_1(t)$ is capable of driving multiple nodes in the network. Identifying the minimal sets of nodes that can drive the network when steered by different signals is crucial to control the network. These nodes are termed driver nodes in the literature [100–102]. According to the controllability rank condition of Kalman [103], the system described by the equation 1.2 is controllable if the augmented matrix

$$C_{N \times NM} = (B, AB, A^2B, ...A^{N-1}B),$$

has full rank, i.e., $rank(C) = N$. Let $G = (V, E)$, $(|V| \neq 0)$, be a directed graph, i.e., $\forall e = (i, j) \in E$, $\exists$ a direction from 'i' to 'j'. Here 'i' is refereed as parent node, and 'j' as the child node. $M \subseteq N$ is called a matching set if $\nexists\ e_1, e_2\ (e_1 \neq e_2) \in M$ such that e_1 and e_2 share a common parent or child node. In a graph, a matching set of highest cardinality is called a maximum matching. If, $e_2 = (i, j) \in M$ then 'j' is called a matching node and the rest are called unmatched. Liu et al. [100] termed these unmatched nodes "driver nodes" and showed that they are sufficient to control the network. Nevertheless, the cardinality of multiple matching sets can be the same; consequently, a network can permit more than one maximal matching set. Consequently, detecting driven nodes is not unique, and several solutions may coexist.

Using this concept, the network nodes can be further categorised into critical, intermittent, and redundant categories [104]. The critical category contains the nodes appearing as driver nodes in all the matching sets, intermittent driver nodes appear in some but not all the matching sets, while the redundant category contains nodes that are not driver nodes. Numerous studies have been undertaken in an attempt to comprehend these driver nodes. Khazanchi et al. [105] compared driver nodes of four different PPI networks. They found that the driver nodes tend to be transcription factors and are enriched in first-degree neighbours of hubs. In addition, they demonstrated that the hubs are the lethal proteins in the network and that it is, therefore, preferable not to disturb the lethal hubs but rather the proteins close to the hubs. Badhwar et al. [106] used network controllability in the neuronal network of C. elegans. They found that driver neurons of C. elegans, were motor neurons located in the ventral nerve cord. Wu et al. [107] developed a methodology to identify the driver nodes in a network. The study was concerned with the states of disease biomolecules and biomolecules that cause adverse effects. Their goal was to make the states of disease-causing biomolecules healthy while minimising the state alterations of biomolecules that cause adverse effects. They discovered that the identified potential therapeutic targets are targets of approved medications or are consistent with previous research results, demonstrating the viability of the method.

Again, each node in the network can be divided into an indispensable (I), dispensable, and neutral node category, if its deletion, respectively, increases, decreases, or causes no impact on the minimal number of driver nodes required to control the network. These nodes are the most fragile nodes in the network [102] and are prone to mutations and are often targeted by viruses and drugs [101, 102].

Degree Distribution:

The degree distribution, P(k), is the probability that a node has precisely k links. It is calculated by dividing the number of nodes with degree k by the total number of nodes in the network.

Scale-free networks:

The notion that the vast majority or even every real-world network are scale-free is widely spread throughout scientific domains and network classes. A network is termed scale-free if the fraction of nodes with degree k obeys a power-law distribution, i.e., $P(k) \sim k^{-\gamma}$ where γ is called degree exponent, and it determines numerous properties of a system. For instance, for $2 < \gamma < 3$, a hierarchy of hubs is seen, while if $\gamma > 3$, hubs are irrelevant in the network. Scale-free networks exhibit a high level of resilience against random node failures, but are vulnerable to the failures of hubs.

Influential spreaders:

In several disciplines, a spreading process is a common and spontaneous event [108]. Influential spreaders (IS) in a complex network operate as maximisers or controllers of a spreading process. To increase the flow of information, for instance, an IS operates as a maximiser [109], whereas as a controller, they can manage epidemics or reduce bogus news in a social system [110]. Identifying these spreaders is usually divided into two categories, individual or multiple. The former ranks the nodes in the network according to their influence, while the latter identifies the minimum number of nodes to achieve maximum collective influence. Centrality measures like degree, betweenness, closeness, eigenvalue, etc. come in the first category, while Voterank [111], optimal percolation method [112], etc., come in the second category.

Coreness:

It is a methodical approach to determining a protein's local and global significance. It indicates whether the protein is associated with a densely connected region of the network or with its periphery. Additionally, it demonstrates how influential a node is at disseminating information throughout a network.

Co-expression analysis:

For a comprehensive understanding of the complicated interconnections in biological processes, approaches that can grasp the relationships between the genes involved are unmet needs.

To overcome this challenge, biological networks have been employed as a framework for representing and analysing gene-gene association. There are numerous ways accessible to systematically comprehend these relationships. Among them, one of the crucial methods is Gene Co-expression analysis. It is also a method for determining the roles of unidentified genes and their correlations with diseases. In such a network, the nodes represent the genes, and edges represent their correlation strength. Despite the fact that the underlying concept is a thorough understanding of the synergistic interaction between genes, the conclusion of such an analysis can vary depending on the context. For instance, after constricting the coexpression matrix, one may identify the genes that show similar expressions in a set of samples, or, using a guilt-by-association method, co-expression analysis can help to determine the function of unknown genes. The correlation strength between the genes across the conditions may vary. The genes which show a higher correlation in one condition (say |Pearson correlation coecient| ≥ 0.7) may show a lower correlation in another condition (say |Pearson correlation coecient| ≤ 0.3). Differential co-expression analysis addresses these facts and tries to find crucial genes across the conditions. This analysis relies on the premise that genes whose behaviour changes in relation to a substantial number of neighbours across conditions are more apt to be prospective targets or biomarkers.

1.2.4 Tools and software

Many packages across various platforms have also been used to perform the network-based study. We have enlisted a few useful and most used packages in **Table 1.3**. Proper visualisation of data is crucial for understanding the biological network. Frequently, the sheer quantity and variability of data pose a difficulty for visualisation. Many network visualisation tools and software are available in literature [113–120], most of which translate data onto two-dimensional graphs to depict their relationships. However, when thousands of nodes and connections must be evaluated and shown, the user-friendliness of many of these technologies reaches its limit. Four of the most widely used visualisation and analysis software are Cytoscape [121], Gephi [122], Tulip [123], and Pajek [124]. However, due to its user-friendliness and incorporation of numerous plugins, Cytoscape has established itself as the big Banyan tree in the realm of

network visualisations. **Table 1.4** contains some useful Cytoscape plugins used in network analysis, visualisations, and enrichment analysis.

Table 1.3: **Various useful packages and software for the network biological studies.**

Sr. No.	Package/software	Platform	Description	Ref.
1	dplyr	R	It is a powerful R package that facilitates the manipulation,cleaning, and summarizing unstructured data. It comes with many functions that perform widely used data manipulation operations.	[125]
2	ggplot2	R	An excellent data visualization package.	[126]
3	Bioconductor	R	The Bioconductor project is a collaborative effort to create computational biology and bioinformatics extensible packages and software. It uses the R programming platform and is open source and open development.	[127]
4	mlr	R	An R package to perform machine learning tasks.	[128]
5	limma	R	This R package is used for the analysis of gene expression data.	[129]
6	WGCNA	R	WCGNA is a popular R analytical package to constructs a gene co-expression network and identify modules.	[130]

Table 1.3 continued from previous page

Sr. No.	Package/software	Platform	Description	Ref.
7	biomaRt	R	This open-source R package incorporates easy and user-friendly functions to capture all genomic data or data for selected proteomes, genomes, coding sequences, and annotation files contained in the databases hosted by the National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EMBL-EBI)	[131]
8	DESeq2	R	DESeq2 is a widely used method for differential expression analysis of count data.	[132]
9	CFinder	-	CFinder is a stand-alone application that locates overlapping groups of densely interconnected nodes in a network with the aid of the clique percolation method	[133]
10	Enrichr	-	This is a useful web-based and mobile software application to perform gene enrichment analysis and is facilitated by various interactive visualization approaches to display enrichment results.	[134]
11	PyPathway	Python	PyPathway is free and open-source python package that performs functional enrichment analysis, network modelling, and network visualization.	[135]

Table 1.3 continued from previous page

Sr. No.	Package/software	Platform	Description	Ref.
12	Cytoscape	-	This software is designed for the visualisation of large-scale networks. Along with basic network analysis measures, this app comes with various plugins which are useful for finding clusters and modules, pathway enrichment, etc.	[121]
14	Metascape	-	It is a web portal that provides functional enrichment,interactome analysis and gene annotation.	[136]
15	PIANO	R	It is an R package to perform gene set analysis.	[137]
16	DiffCoEx	R	A method to identify gene co expression differences between multiple conditions.	[138]
17	ComBat	R	A package for correcting batch effects in datasets with a known batch covariate.	[139]
18	BioNetStat	R	A tool for comparison of two or more networks simultaneously.	[140]
19	CentiServer	R, Web-based portal	It can perform centrality analysis. Currently, it can perform 403 different types of centrality analysis.	[141]
20	DAVID	-	A web-server that can perform functional enrichment, gene ID-conversion etc.	[142]

Table 1.3 continued from previous page

Sr. No.	Package/software	Platform	Description	Ref.
21	GSEA	R, Software	It can perform rank-based gene set enrichment.	[143]

1.3 Omics technologies

The addition of the term "omics" to a molecular term denotes a thorough, or global, examination of a collection of molecules. The emergence of omics technologies such as genomics, epigenomics, transcriptomics, proteomics, metabolomics, microbiomics, etc., have embraced new possibilities to study a biological system to an extraordinary detailed level. Genomics is the study of the genome of an organism, epigenomics aims at exploring global epigenetic changes that offer crucial insights into mechanisms and function of gene regulation across several genes in a cell or organism. Transcriptomics relies on the qualitative and quantitative genome-wide study of RNA levels, while proteomics facilitates the study of the whole proteome of an organism [158]. Mass spectrometry-based proteomics is an indispensable approach to delineating protein expression, protein-protein interactions, subcellular localisation, and post-translational modifications. Similarly, metabolomics is the large-scale study of metabolites within cells, biofluids, tissues, or organisms [159]. The study of the microorganism in a given community comes under the focus of microbiomics [158]. Throughout the times, new dimensions have been added to omics, such as lipidomics, nutrigenomics, etc. The advent of these technologies has opened up new avenues for studying biological systems at an unprecedented level of detail. In such studies, the characteristics and quantity of a specific type of molecule in samples are quantified, and the patterns and/or relationships between the sample attributes are investigated([160]. Omics techniques yield massive amounts of multidimensional data that can be analysed using new informatics methodologies and traditional statistical methods. While systems theories, such as network analysis and machine learning, are well-suited for analysing

Table 1.4: **List of useful Cytoscape plugins**. All these plugins are freely available at the Cytoscape app store (https://apps.cytoscape.org/).

Sr. No	Plugin	Description	References
1	BiNGO	Quantifies GO terms that have been overrepresented in the network and portrays them as a network of relevant GO terms.	[144]
2	Mosaic and Cerebral	These two are visualization plugins for Cytoscape and can compartmentalize the genes/proteins in a network according to their subcellular localization.	[145, 146]
3	PathLinker	This package reconstructs signaling pathways from protein interaction networks.	[147]
4	CytoNCA	Perform centrality analysis of weighted and unweighted networks.	[148]
5	ClueGO	It helps to create and visualize a functionally grouped network of terms/pathways.	[149]
6	GeneMANIA	Uses public databases to import interaction networks from a list of genes with their annotations and putative functions.	[150]
7	BiNoM	It helps to access and analyze pathways.	[151]
8	PiNGO	Helps to locate candidate genes in a network that are linked with user-defined target GO terms.	[152]
9	MCODE	Create clusters in a given network based on the topology to identify densely connected regions.	[153]
10	ConsensusPathDBplugin	Retrieves interaction evidence for a given pair of genes or proteins	[154]
11	AgilentLiteratureSearch	Curates scientific literature to find publications associated with the search term and to create an interaction network based on the search result.	[155]
12	jActiveModules	Detects clusters where nodes show significant changes in expression levels.	[156]
13	cytoHubba	By using various topological algorithms, this Cytoscape plugin can predict and find important nodes and subnetworks in a given network.	[157]

these data, they must be used with a working knowledge of the relevant biological and computational theories. Systems biology addresses the issues raised by the complex organisation of biological processes by applying these methodologies to omics data. Integrating different network-based and artificial intelligence-based approaches to omics data contributes to identifying helpful markers of disease progression. For instance, numerous research has used blood transcriptome data to develop classification models capable of discriminating between samples from TB patients and controls within the cohort [161–163]. Burel et al. identified a CD4 T cell immune signature of LTBI by combined cell population transcriptomics and single-cell protein-profiling techniques [164]. The model provided novel insights into the phenotype of TB-specific CD4 T cells. Taking blood serum from individuals with active and latent TB, Cao et al. identified three potential serum biomarkers that can distinguish between these two types of TB [165]. In many cohort studies, it has been reported that the plasma proteomes are different in LTBI, TB, and HC cohorts and hence can be used as indicators to differentiate the three types of individuals. Based on this, Sun et al. performed a label-free quantitative proteomics analysis to identify plasma biomarkers that can discriminate pulmonary TB from active TB [166]. Personalised medicine addresses the notion that "we are all alike, yet unique." It is a fairy beacon of hope and clarity that burns brightly and shines jubilantly on getting patients on the appropriate medication, and that too in a shorter time. It is a notion that has the potential to change medical interventions by offering effective, individualised therapy methods based on an individual's genetic, epigenomic, and proteomic profile while taking into account the patient's unique circumstances. In other words, such technologies built a specific molecular window that allows peeping through the discrepancies between the genomic profiles of diseased and healthy individuals. However, this field is still in its infantry and requires the integration of more sophisticated tools and methods.

1.4 Tracing the footsteps of autophagy in computational biology[2]

The advent of systems biology is opportune for drug development. As the time and cost required to bring new pharmaceuticals to market continue to increase, it is crucial to expedite efforts to identify the most promising candidates as quickly as feasible. This necessitates a deeper comprehension of disease-related pathways to appropriately assess medication specificity or detect and identify unintended or undesirable effects. In addition to addressing progressively more complex, multivariate forms of the disease in target populations, the field of drug discovery faces the additional challenge of confounding underlying mechanisms. As the ultimate objective of pharmacotherapy is to modulate cellular biochemical function to induce physiological change or assure repair, the cell provides a preliminary level of systems abstraction for drug development. Even though cellular components are frequently functionally defined in isolation, they execute their intracellular and intercellular roles within a complex network of interactions. In these contexts, mathematical modelling studies also do not stay behind. The large-scale studies have the limitation that they can not take into account the system's situational changes, which include stochasticity, inherent delay, etc. These models can portray an abstraction of reality and theoretically take the system from any state to any desired state. That is why, together or individually, mathematical modelling and network biology can improve drug development endeavours, new target identification, and delimitation of off-target effects, ideally leading to preventative strategies and empowering individualised solutions. The process of autophagy, as previously mentioned, is a quintessential biological process which has been proven to be the cause or effect of a myriad of diseases. With the *in vivo* and *in vitro* studies of autophagy which have explored many novel discoveries, systems biology, with the potential of decrypting the system's complexity both as a whole and in part, has significantly emerged and made tremendous contributions to the field of autophagy. For example, using network biology processes, the core proteins in the autophagy process in a disease can be identified. Mathe-

[2]The bulk of this section is taken verbatim from the published article: Sarmah, Dipanka Tanu, Nandadulal Bairagi, and Samrat Chatterjee. **"Tracing the footsteps of autophagy in computational biology."** Briefings in Bioinformatics 22.4 (2021): bbaa286.

matical modelling can be done on this set of proteins to may identify potential parameters that otherwise could not be explained by network analysis alone. This methodology is shown in **Figure 1.8**.

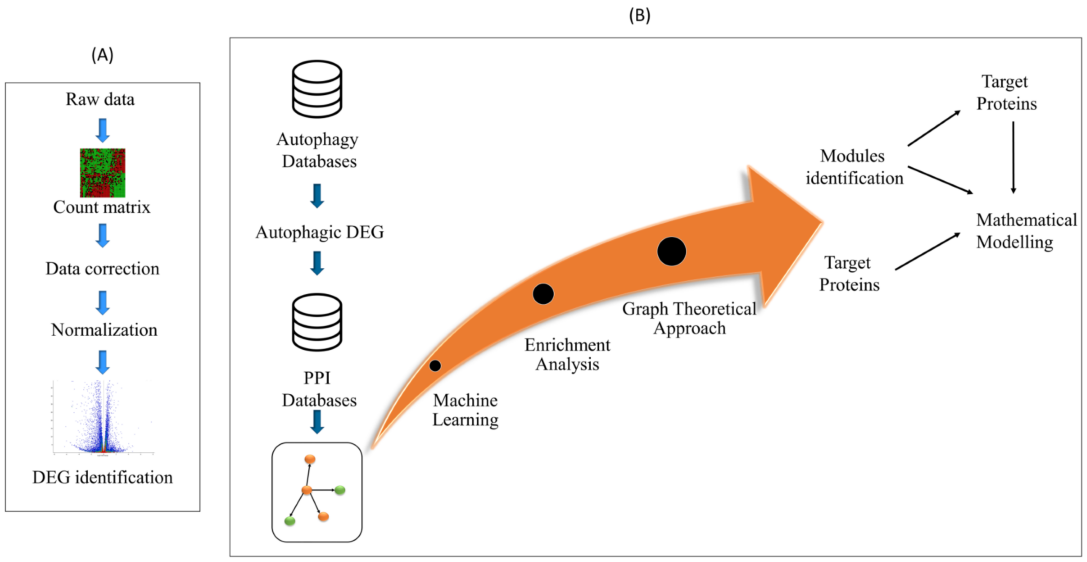


Figure 1.8: **The journey of proteins from being inside the system to the arms of mathematical modeling.** The palate (A) shows the processing of data. The raw gene expression data to study a particular disease is first corrected, including steps like dealing with the null values and the outliers. The data is then normalized, and the differentially expressed genes (DEG) are calculated. The palate (B) shows the autophagy specific study of the disease. The autophagic genes are first obtained from an autophagy database from where the differentially expressed autophagy genes are selected. Using PPI databases, a PPI network of the DEGs (which may be entirely autophagic DEGs or a mixture of autophagic and not-autophagic DEGs) is constructed. In the figure, the green color denotes the autophagic, and the orange color represents the non-autophagic genes. Implementing machine learning approaches, graph theoretical approaches, or enrichment analysis (pathway analysis, disease analysis, or gene ontology analysis), the significant modules or target proteins from the network are extracted. In the first case, the proteins driving the module can further be identified. Finally, the implementation of mathematical modeling approaches can explore the dynamics and underlying mechanism of the target proteins or the module.

Addressing this associations, in this section, we have encapsulated the overview of autophagy in computational biology explored via mathematical modeling and network analysis along with comprehensive insights about these approaches and their applications in the exploration of the autophagy process at various levels (molecules, cells, tissues). We have delineated several well-established methods such as mathematical models based on different types of differential equations, Petri net, agent-based models, enrichment analysis, and centrality analysis

to capture the dynamical behavior or the collective influencers in the network. Further, we have enlisted the available autophagic databases and the related resources and their feature selection and epitomized some conventional software and tools used for visualization and analysis in computational biology.

1.4.1 Mathematical modeling for autophagy

The process of autophagy consists of 5 steps, and all these stages are easily observable [167]. Different steps in the autophagy pathway may exert a different effect on the system. However, the biochemical reactions in autophagy are mostly nonlinear, i.e., a minute change in any of its stages will not necessarily exert a proportional effect throughout the system. Mathematical modelling endorses simplified abstractions and approximations to identify the steps of autophagy that are responsible for a particular behaviour in the system. Moreover, the constant shift in the behaviour of the system exerts randomness in the autophagy process. Mathematical modelling of autophagy keeps track of these factors and allows the researchers to investigate the dynamics of the system following any environmental conditions that may arise due to various external or internal perturbations or signals. Autophagy is a bridge between cell survival and cell death. Depending on certain extracellular or intracellular signalling, the process of autophagy may decide cell fate. At the single-cell level, these events may be mutually exclusive, indicating that cell death and cell survival events are different attractors of the system. Mathematical modelling can be done to understand the crosstalk between these two events using attractors, fixed points, and limit cycle concepts. Cell types differ in their response to autophagy stimuli. Addressing this cell-to-cell variability, various therapies have targeted autophagy manipulation in cancer therapy [67–69]. Mathematical modelling can help in planning and predicting the parameters that [168, 169] could be targeted and its outcome on the cell population. For example, autophagy helps in tumour cell survival under various stress conditions [170]. On the contrary, increased autophagy may lead to excessive cellular degradation and, thus, may initiate cell death [50]. A mathematical model can perfectly utilise these conditions to identify the biological parameters that increase the autophagy process in disease conditions so that the tumour cell gets less benefit from the basal level of autophagy and cell death initi-

ates. Thus, mathematical modelling can be used in a very effective way to decipher the process of autophagy and its role in various diseases or conditions. We have discussed below some of the modelling work done in the autophagy process to get an idea of the applicability of mathematical modelling in understanding the autophagy process.

1.4.1.1 Differential equations based models for autophagy

In 1975, Deter et al. [171] formulated the first mathematical model to delineate glucagon-induced autophagy in rat liver. This primitive study was based on experimental observations, collision theory and chemical kinetics and mainly focused on studying the population of telolysosomes, autophagosomes and autolysosomes in rat liver. Thereafter, various studies on autophagy have incorporated different types of mathematical models, viz. ODE, DDE, and SDE. The widely used ODE-based models are the simplest to study the process of autophagy. These models are entrenched in the assertion that the system considered is well-mixed and there are sufficient numbers of components so that their numbers can be considered as continuous quantities. Understanding the steady-state, stability, and other qualitative behaviour of a model will unveil the system's underlying mechanism. For example, response to cellular starvation is an intrinsic property of autophagy and was mathematically addressed by Jin et al. [172]. They classified the cells into normal phase and autophagic phase, and by taking nutrition as the third variable, a logistic type (three-dimensional) model of the yeast cell population was constructed and analysed. The model considered in this example has one unstable trivial equilibrium point when the nutrient concentration in the input flux and nutrient loss rate by output flux is constant and a locally asymptotically stable positive equilibrium point when the system is considered without autophagy. The model analysis concluded that an efficient autophagy level might be adequate to sustain a population during a long duration of starvation. However, the author did not incorporate any molecular regulation in their study. A hybrid model consisting of cell population dynamics and molecular regulation could have provided a better insight into cell fate regulation by autophagy. Addressing this issue, the same group later developed a hybrid model [173] to understand the molecular regulation and population dynamics of yeast by incorporating molecular level interactions, the amino acid exchange between cells, and cell behaviour.

ODE models are also built to predict optimal drug schedules to control autophagy. Shirin et al. [174] formulated a nonlinear ODE model to predict optimal drug schedules to control autophagy. Focussing on four autophagosome production influencers and their specific inhibitors, the model figured out various drug pairs that are more effective when taken together. Mathematical models can qualitatively estimate the protein levels capable of deregulating homeostasis, like Ouzounoglou et al. [175] formulated a model to understand the dynamics of Alpha-synuclein (ASYN) in Parkinson's disease.

Autophagy and apoptosis pathways are closely regulated, and some proteins, which regulate autophagy, can also regulate apoptosis [176, 177]. Hence, proper knowledge of autophagy and apoptosis interconnections may help stop or promote fatal cell decisions. Kapuy et al. [178] studied beclin1-mediated autophagy and caspases-mediated apoptosis by forming an ODE model. The model was built to address the B-cell lymphoma 2 (BCL2)-Beclin1-caspases minimal network. They have also considered the effect of stress on autophagy by taking it as a bifurcation input. Based on the observation, it was suggested that the autophagy apoptosis transition is adjudicated by a bistable switch and, depending upon the intensity and duration of stress levels, sequential activation of cellular response can be initiated by a combination of BCL2-dependent regulation and feedback loops between Beclin1 and caspases. Various other models have also been built on understanding the autophagy-apoptosis interplay [179, 180]. A key feature of autophagy is that it also plays a role in unfolded protein response (UPR). Cyto-protective or cyto-destructive UPR gets activated by anti-oestrogens or other drug therapies. Autophagy assists in the cyto-protective role of UPR, while the cyto-destructive role contributes to apoptosis [181]. Addressing these, a mathematical model of autophagy, apoptosis, and UPR was proposed to understand the interactions that accomplish anti-oestrogen resistance and the effects of GRP78 on both sensitive and resistant breast cancer cells [181]. The model provides a clear picture of interactions of autophagy, apoptosis, and UPR to produce both sensitivity and resistance to antioestrogen therapy under various conditions. The time delay associated with any biological process is not facilitated by ODE-based models. This is mainly addressed by DDE. These models address the time lags between biological processes and thus offer a better portrayal of biological systems. Time lag plays a vital role in autophagy,

as in many biological processes. Various studies have implemented DDE-based mathematical modelling to understand the hidden mechanisms in the autophagy process. For example, in autophagy, the formation of autolysosome follows autophagosome formation indicating a time delay. Han et al. [182] formulated an eight-dimensional (8D) model using the delay to study the behaviour of both resident (normal) and abnormal proteins along with the formation of autophagosomes and autolysosomes, the intracellular concentration of Adenosine triphosphate (ATP), and amino acids. The study showed that intracellular levels of autophagosomes and autolysosomes display an oscillatory behaviour. The same group later formed another mathematical model to explore the role of autophagy in the protein/organelle quality control when exposed to different physiological perturbations [183] and further extended their study to Alzheimer's disease [184]. ODE-based models do not consider the effect of noise, which is an inherent property in many dynamical systems. This property is addressed by the SDE models, as done by Martin et al. [182], who studied autophagy vesicle dynamics in a single cell. They used live-cell fluorescent microscopy to measure the synthesis and lysosomal turnover of autophagic vesicles (AV). The data was used to build a 4-dimensional ODE model, followed by a 23-dimensional SDE model for the accurate prediction of autophagic vesicle dynamics in a cell. The SDE model has implemented a sequence of biochemical and physiological steps in the autophagic pathway from PtdIns3KC3 activation through LC3 conjugation that comprises the nucleation of the phagophore, maturation of the autophagic vesicle and lysosomal degradation. The mechanistic model was a better portrayal of the autophagy dynamics in a cell. For example, correlating with the experimental data, the SDE model captured a time lag in the production of AV in response to treatment initiation, but no such behaviour could be achieved with the deterministic model. The SDE model was also capable of accurately predicting that an 80% decrease in ATG9 content would result in a corresponding reduction in vesicle synthesis rate. It also stated the correlation between AV size and LC3 levels across single cells. The study can be taken as an example to quote that although ODE models are less complicated and can portray biological behaviour, SDE models are a better illustrator of biological phenomena.

1.4.1.2 Agent-based models for autophagy

The applications of agent-based models to study autophagy are very few. The creation, movement, fusion, and deterioration of autophagy pathway vesicles are dynamic both temporally and spatially. To delineate the spatio-temporal aspects of autophagy regulation and its dynamic behaviour, Borlin et al. [185] have constructed an agent-based model using the NetLogo ABM platform. The first agent is the phagophore, which grows and matures to form the second agent autophagosome, which then fuses with the third agent lysosome to generate the last agent autolysosome. The newly formed autolysosomes can then either fuse with lysosomes, autophagosomes, or other autolysosomes to grow. They inferred spontaneous motion for phagophores and autolysosomes to simulate organelle movements, while autophagosomes and lysosomes travel directly towards or directly away from the nucleus to replicate their active transport along the cytoskeleton, at a pace that is independent of its size. The key parameters of the model were fitted with an iterative method using a genetic algorithm and a predefined fitness function. The model, integrated with high-resolution fluorescence microscopy data, could successfully reproduce the short-term and long-term behaviour and cell-to-cell variability.

1.4.1.3 Petri net

Minimal literature is available on the use of Petri net in the study of autophagy. Jennifer et al. [186] studied the Salmonella xenophagy in epithelial cells by designing a Petri net model. The model includes all biochemically proven and published processes of Salmonella xenophagy in epithelial cells and comprises 61 places (proteins/ macromolecular complexes/ organisms/ signals) and 184 arcs. The model consists of 16 T-invariants describing biological subpathways in steady-state and represents the fundamental dynamics of the system. The author has implemented *in silico* knockouts of specific proteins to investigate the model behavior and the corresponding biological effect.

1.4.2 Network biology based approach for autophagy

1.4.2.1 Omics and autophagy

The integrated method of the omics strategies and network biology enable a better understanding of the autophagy process. There are studies that incorporate a large-scale multi-omics approach to study the broad framework of autophagy and its association with other biological processes. These studies have deciphered the role of autophagy in host-pathogen interactions, tumor growth, various cancers, nervous systems etc. [187, 188]. Considering that "omics"-based studies are a pivotal area of current research to provide a more systematic view of biological processes, these approaches have driven our insights into the regulation of autophagy.

1.4.2.2 Network analysis for autophagy

Throughout the decades, the advancement of high-performing data collection technology has resulted in a large number of autophagy-related data. Network analysis approaches have been implemented in these data to delineate the association of autophagy with various diseases and biological processes. Network analysis also helps uncover the organizing principles of diseases and identifies the potential targets accountable for the disease pathogenesis.

Network analysis is well supported by autophagic databases, which play a crucial role in delineating the role of autophagy in various diseases. Various studies have been done by the implementation of the specific autophagic information obtained from these databases [188–190]. Lin et al. [190] carried out a comprehensive study of autophagy-related genes (ATG) and associated noncoding RNAs and transcription factors to investigate the association of autophagy with digestive system tumours (DST). The Cancer Genome Atlas database was used to get the digestive tumour transcription details. The autophagy genes were extracted from the Human autophagy modulator database. The study, facilitated by WGCNA, crosstalk connection, pivot analysis, and functional analysis, revealed that the autophagic genes control the pathogenesis of digestive system tumours and highlighted the potential role of autophagy in the treatment of DST. Wang et al. [188] constructed a disease autophagy network where disease genes were taken from online mendelian inheritance in man (OMIM) [191] and autophagic

genes were extracted from the human autophagy database (http://www.autophagy.lu/), the autophagy database [192], and the autophagy regulatory network database [193]. The autophagy genes were observed to act as a bridge between diseases and were found to be topologically important in the disease-autophagy network.

Network-based studies often facilitate the identification of hubs and modules. Modularity is an essential property of a network. It refers to the organization of nodes in clusters. Module-based analyses can contribute to a deeper understanding of biological systems. Hub proteins are also crucial in maintaining the global network structure. A study carried out by Durocher and co-researchers [194] elucidates the gene network in the peripheral blood transcriptome associated with human intracerebral haemorrhage. Using the WGCNA package in R, they identified the hubs and the modules in the network, and used ingenuity pathway analysis (IPA) and the DAVID Bioinformatics Database [195] to find the associated pathways and processes. Various studies [196–198] have performed a network-based analysis on autophagy by using the dataset obtained from the Gene Expression Omnibus (GEO) repository [199]. After following the preliminary analysis, the WGCNA package in R has been used to identify significant modules and hubs in the network [197, 198, 200, 201]. Although network analysis approaches have been applied extensively to study autophagy, methods like network stability, control theory, percolation, etc., are yet to be integrated to study the autophagy process. Given the importance of these methods, their implications will surely help identify novel targets and pathways related to autophagy. The lack of sufficient temporal data to understand a disease progression has also limited the network-based study of autophagy processes. Nonetheless, with time the data is growing, and we believe in the coming years, we will have enough data to make better and more accurate predictions.

1.4.3 Artificial Intelligence (AI) associated research of autophagy

As in many other biological processes, AI-based approaches have also been incorporated into the field of autophagy. In a recent study, Zhaoyue et al. [202] applied machine learning (ML) techniques to classify renal cell carcinoma (RCC) subtypes using autophagy proteins. The expression data of the key autophagy proteins in renal cell carcinoma was measured by im-

munohistochemical images. The data was then normalised with mean and standard deviation. K-Nearest Neighbor (KNN) algorithm was applied to the normalised data for classification. Their study identified the basal level of autophagy as a potential measurement for discrimination of renal cell carcinoma. In an early work by Janos and co-researchers [203], an image analysis pipeline was developed using the support vector machine (SVM) for the determination of novel selective pharmacological inducers of autophagy in human cancer cell lines. A variety of software incorporating a broad range of machine learning algorithms has been developed recently. For example, Serrano et al. [204] have used the software Scikit-learn [205] to study the effect of mRNA alterations of some autophagic genes, one proapoptotic gene, and one anti-apoptotic gene in HIV-infected patients effectively treated with combined antiretroviral therapy (cART).

In the past two decades, the pharmacological modulation of autophagy has gathered a great deal of attraction. The process of autophagy gets manipulated by various modulators. ML-based methods can be blended to study the mechanism of actions of these autophagy modulators to gain knowledge on various factors that include side effects, drug repurposing, and development of novel polypharmacological strategies [206]. AI approaches are powerful tools that associate important molecular changes with an observed phenomenon. However, these approaches remain silent on the underlying mechanism for such observations. To capture the possible mechanism, we need to take help from differential equation-based models.

1.4.4 Databases with the information related to autophagy

Biological databases play a central role in systems biological studies. They offer the opportunity to access a wide variety of biologically relevant data, which include protein-protein interaction information, disease-protein association information, microarray, next-generation sequencing, protein localization, post-translational modification, the structural details of a protein or compound, and pathways associated with proteins, etc. However, databases containing exclusively autophagic information are very few. In **Table 1.5**, we have enlisted eleven most used databases in autophagy. These databases contain various information like disease associations, pathways, the specific effect on autophagy, etc. In **Figure 1.9**, we have compared the

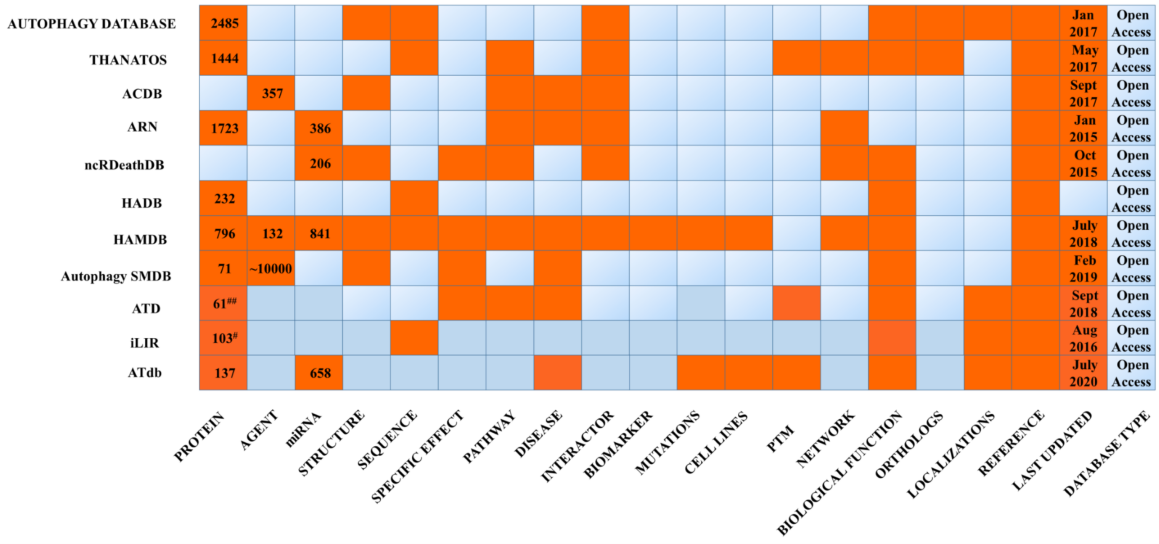


Figure 1.9: **Comparison of some of the well-explored autophagy databases in literature.** The columns contain the name of the databases, and features are placed in the row. The orange colour means that the particular feature is present in the database, and the blue colour means it is absent. Here, the agent feature includes drugs, chemicals, and small molecules. Only autophagy-related proteins are considered for the ARN database.

features of these databases.

Table 1.5: **Some of the most used databases in autophagy.** The features of these databases are shown in **Figure 1.9**.

Sr.no	Name	Full form	URL	Ref.
1	HAMDb	Human Autophagy Modulator Database	http://hamdb.scbdd.com	[207]
2	ARN	Autophagy Regulatory Network	http://autophagyregulation.org/	[193]
3	Autophagy database	Autophagy database	http://www.tanpaku.org/autophagy	[192]
4	ncRDeathDB	The noncoding RNA (ncRNA)-associated cell death database	http://www.rna-society.org/ncrdeathdb	[208]
5	ACDB	Autophagic compound database	http://www.acdbliulab.com	[209]
6	THANATOS	Autophagy, Necrosis, Apoptosis OrchestratorS database	http://thanatos.biocuckoo.org	[210]
7	HADb	Human Autophagy Database	http://www.autophagy.lu/	-
8	AutophagySMDB	Autophagy Small Molecule Database	http://www.autophagysmdb.org	[211]
9	ATD	Autophagy To Disease	http://auto2disease.nwsuaflmz.com	[212]
10	iLIR	In silico identification of functional LC3 Interacting Region Motifs database	https://ilir.warwick.ac.uk	[213]
11	ATdb	Autophagy and Tumor Database	http://www.bigzju.com/ATdb	[214]

HAMDB contained the most autophagic proteins among these databases. To grasp how these proteins function, it is necessary to comprehend their relationship with one another. In other words, we must construct and analyse the network of their interactions. We first extracted 551 autophagy-related genes from the HAMDB database to create the autophagy interactome. We used STRING [215] and SIGNOR [216] databases for information regarding their interactions. The former compiles information regarding both predicted and experimental evidence, while the latter details the functional interactions between proteins. As STRING also facilitates text mining, a confidence score of 900 was used to build the network. In both of these networks, the edges containing at least one autophagic protein were considered. We discovered that the STRING network had more autophagic proteins than the Signor network. This indicates that the functional information between these proteins has not yet been curated to its fullest extent. This also caused the STRING network to be more interactive than its counterpart **Figure 1.10**. The networks are shown in **Figure 1.11 and 1.12**.

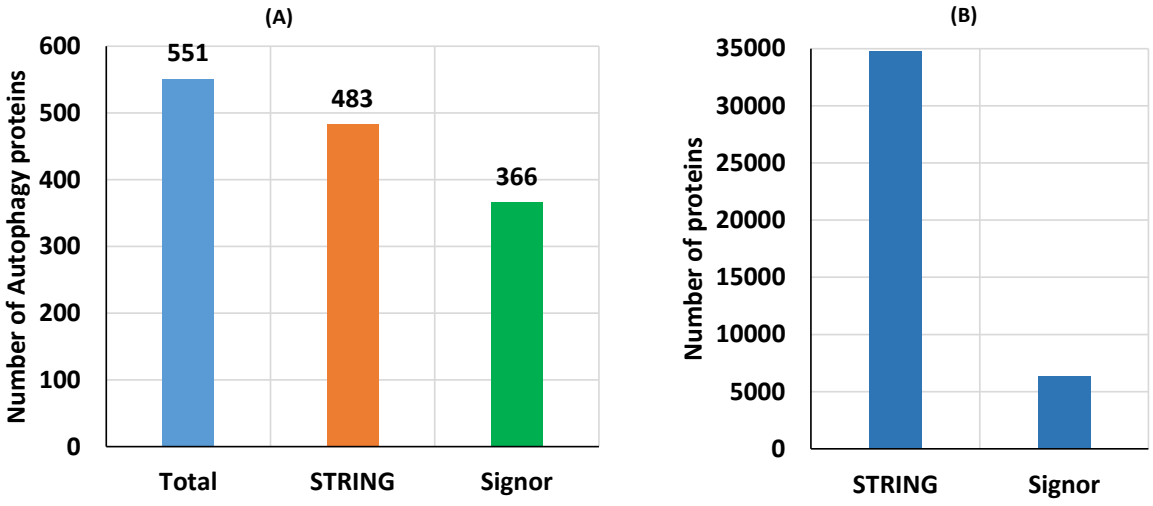


Figure 1.10: **Number of autophagic proteins and interactions in STRING and Signor database.** A) Number of proteins. B) Number of interactions.

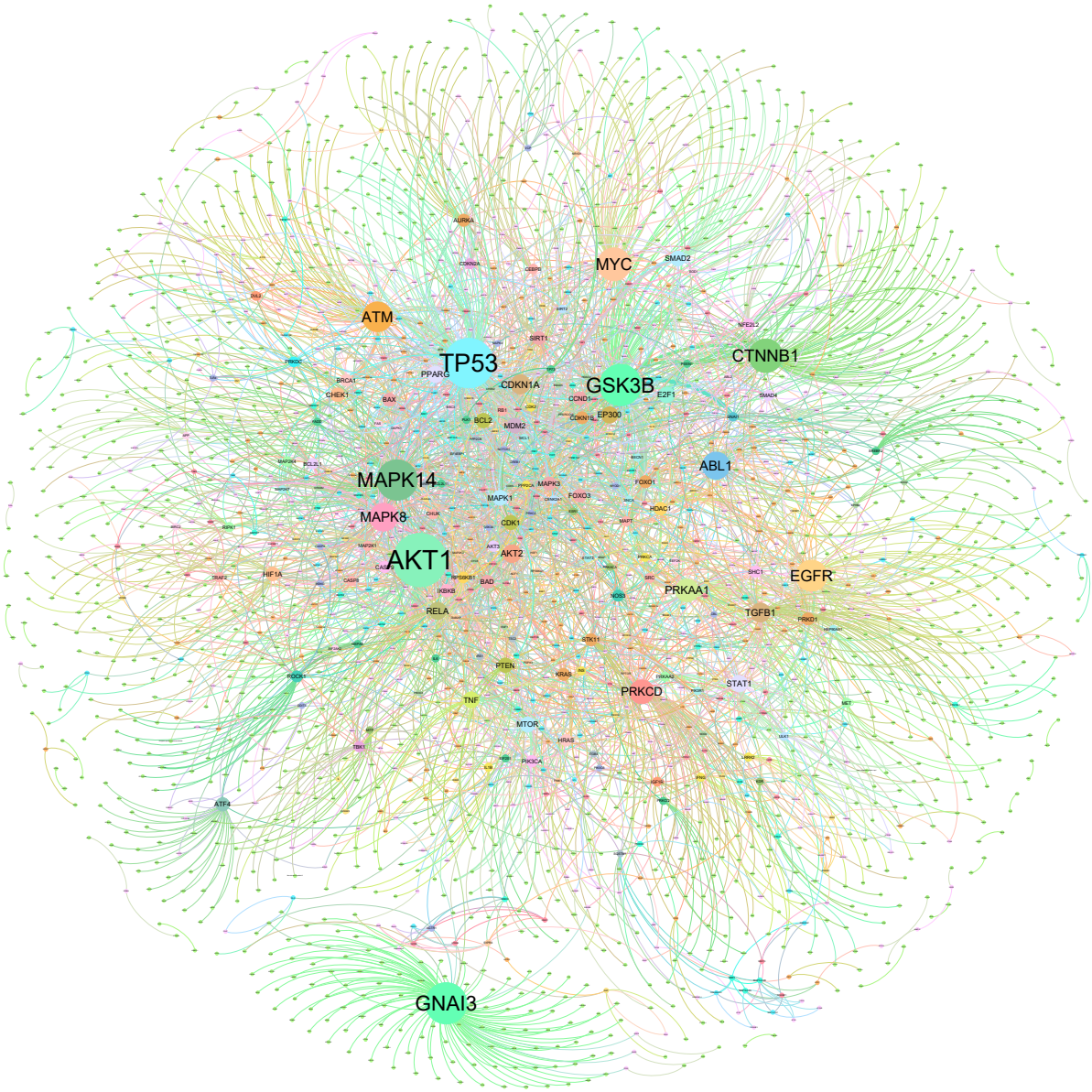


Figure 1.11: **Autophagy interactome constructed using SIGNOR.** Here, the nodes are sized and colored according to their degree.

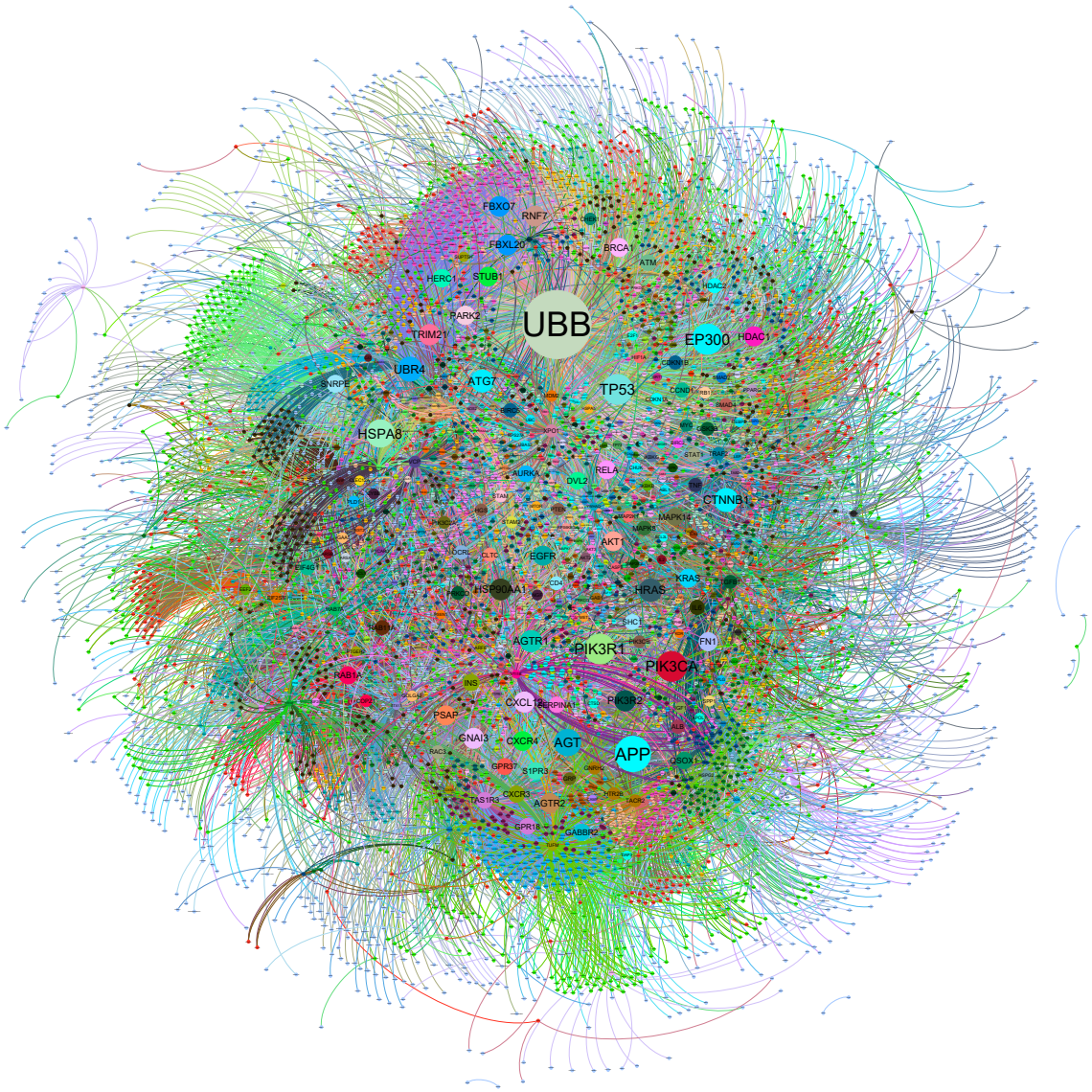


Figure 1.12: **Autophagy interactome constructed using STRING.** Here, the nodes are sized and colored according to their degree.

Table 1.6: **Top 20 hub proteins in the autophagy network constructed using the Signor database.**

Name	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	In-degree	Neighborhood Connectivity	Out-degree	Total Degree
GSK3B	3.175408426	42.34641026	0.314920119	0.006864989	51	11.88695652	286	337
TP53	3.75322442	66.04662936	0.266437572	0.006352802	234	10.37762238	87	321
AKT1	3.00171969	66.69242166	0.333142366	0.004135649	83	9.102564103	219	302
MAPK14	3.073086844	39.38906864	0.325405708	0.005689001	71	10.12389381	177	248
ATM	3.190885641	28.60482708	0.313392617	0.010526316	26	10.80263158	173	199
EGFR	3.999140155	36.27201305	0.250053752	0.005709877	81	8.802469136	98	179
MAPK8	3.627687016	17.79660553	0.275657739	0.008544087	44	11.42857143	128	172
CTNNB1	4.349957008	21.16809061	0.22988733	0.005487122	126	9.305263158	27	153
ABL1	3.492691316	38.6789882	0.286312161	0.006438632	28	8.183098592	117	145
PRKCD	3.427343078	11.05973485	0.291771199	0.005456349	11	10.71875	131	142
GNAI3	1	0.570726378	1	0	123	1.266666667	4	127
MTOR	3.695614789	22.24446747	0.270590973	0.016098485	28	13.75757576	94	122
MYC	4.37661221	33.00901004	0.22848723	0.006894791	58	8.752808989	63	121
IKBKB	4.558899398	8.241234062	0.219351188	0.011396011	52	11.18518519	61	113
AKT2	3.270851247	11.91587683	0.30573081	0.015942029	23	20.06521739	81	104
PRKAA1	3.689595873	10.31743343	0.271032393	0.005737705	7	10.04918033	96	103
MAPT	0	0	0	0.026190476	100	17.42857143	0	100
BAD	4.731728289	9.64888711	0.211339269	0.014112903	86	21.53125	7	93
FOXO3	6.000859845	1.874892418	0.166642785	0.022177419	74	20.53125	11	85
SMAD2	6.13155632	3.183732833	0.163090731	0.005882353	77	7.171428571	6	83

To find the topologically strong proteins in both the networks, we opt for some basic centrality measures: average shortest path, betweenness, closeness, clustering co-efficient, and degree centrality. We identified the top 20 hubs in the network which are provided in **Table 1.6**-**1.7**. TP53, CTNNB1, and GNAI3 were common hubs in both networks, echoing the fact that they are the three most significant proteins in the autophagy interactome.

1.4.5 One single process and various computational approaches: which door to choose?

The Mathematical modelling and network analysis approaches can grasp the underlying dynamics and topology of any biological system. We have summarised the applications of mathematical and computational biology tools to study autophagy with differential environmental conditions (**Figure 1.13**). Nevertheless, the complexity and the choice of the approach can vary from system to system, depending on the perspective of the study. From initiation to degradation, the process of autophagy comes under the influence of many proteins and stresses. Taking a few or all of them together, a mathematical model helps to understand how the dynamics of these sets of proteins influence the progression of autophagy by taking a deterministic approach.

Table 1.7: **Top 20 hub proteins in the autophagy network constructed using the STRING database.**

Name	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	Degree	Neighborhood Connectivity
UBB	2.276239	0.19432585	0.439321209	0.011478494	1129	20.6705
APP	2.676824	0.048394323	0.373577096	0.034493657	563	27.20426
EP300	2.49169	0.062551069	0.401334074	0.017183758	478	23.84728
TP53	2.487227	0.068664942	0.402054201	0.019924693	470	27.72553
PIK3R1	2.614651	0.025442495	0.382460271	0.033349735	469	30.9403
PIK3CA	2.578024	0.028200107	0.387893983	0.0345009	466	32.34979
HSPA8	2.626039	0.069486163	0.380801688	0.018712542	425	19.45412
AGT	2.897199	0.012404202	0.345160948	0.051182874	406	29.46305
UBR4	2.878424	0.019843949	0.347412318	0.040383248	368	22.30978
CTNNB1	2.594183	0.039263504	0.385477843	0.01886428	353	28.94618
HRAS	2.542167	0.026544531	0.393365216	0.03498178	340	35.70882
ATG7	2.822099	0.011061642	0.354346166	0.049386736	333	27.88889
AGTR1	2.697599	0.015938434	0.370699983	0.058679467	320	36.5875
HSP90AA1	2.551554	0.047490903	0.391917973	0.020623413	313	32.76038
TRIM21	2.981533	0.006950559	0.335397956	0.048301725	306	23.62745
PIK3R2	2.862881	0.007819009	0.3492985	0.047188163	303	29.63366
CXCL12	2.895199	0.006840072	0.345399458	0.067518125	299	33.8796
GNAI3	2.852878	0.015238235	0.35052325	0.065789769	299	32.16388
RNF7	3.031856	0.002052966	0.329830973	0.053284523	298	23.26174
FBXL20	3.032318	0.002001357	0.329780755	0.054306468	295	23.4678

These models can predict cellular fate through autophagy by using a suitable set of parameters and a core set of autophagy modulators. They can also be used to study the randomness in the process of autophagy occurring due to the variability of the stress and frequent changes in the cell's energy requirements. Agent-based models can range from continuous to discrete based on the requirement. Petri nets facilitate both the qualitative and quantitative models and hence can be used to model the involvement of autophagy in cellular biochemical reactions.

On the other hand, network biology can be used to identify crucial autophagy-related proteins responsible for the progression of diseases. Different sets of targets will be obtained for the same disease owing to the method applied, which will further require biological validation. For example, if the intention is to select only the most connected proteins, the proper method will be to measure the degree centrality. But, if the goal is to find the proteins that can disperse information very effectively, closeness centrality would be the best approach to consider. Contrary to the analysis of the topology of the system by network analysis, enrichment analysis focuses on extracting the pathways, localisation, and functions of the proteins present in the disease network. These pathways can then further be studied by constructing an autophagy-specific

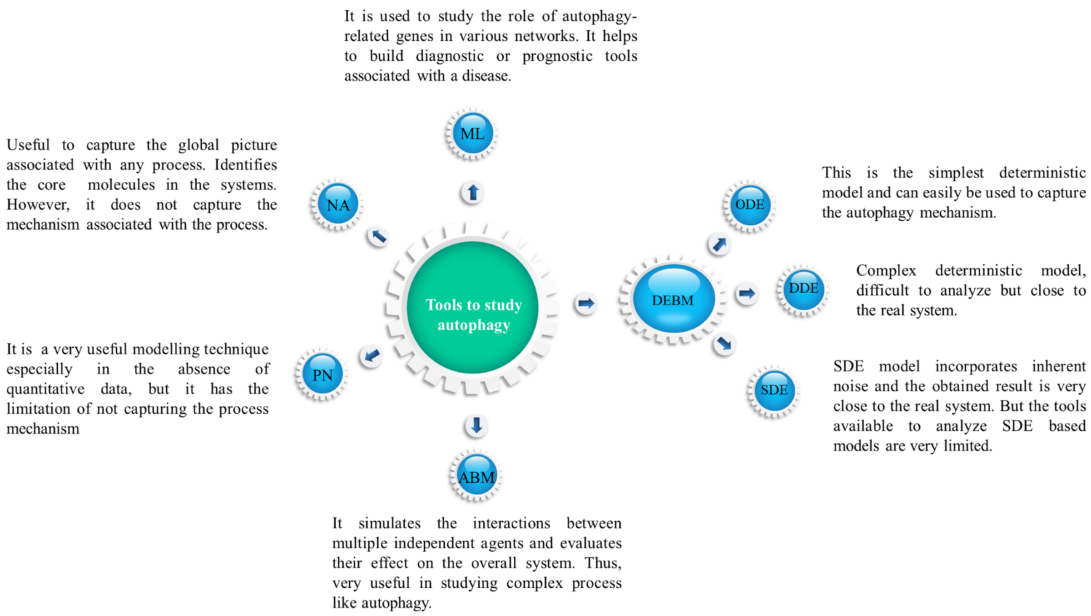


Figure 1.13: **Application of mathematical and computational biology tools to study autophagy in different environmental conditions.** Abbreviations: DEBM-Differential equation-based mathematical models; ABM-Agent-based model; PN- Petri net; NA- Network analysis; ML- Machine learning.

PPI network to detect influential proteins in that pathway.

1.5 Scope and objectives of the thesis

Proteins being the dominant molecules that carry out cellular functions and mediate numerous pathways and processes, understanding their perturbations in disease is an unmet need. Again, identifying crucial autophagic targets and regulatory mechanisms is essential for designing effective therapeutic strategies for diseases such as diabetes, cancer, and nonalcoholic steatohepatitis (NASH), as these diseases are significantly regulated by autophagy. One of the severe complications of diabetes is diabetic retinopathy. Autophagy has been demonstrated to play a dual role in this disease. In the early phase, it aids in the reduction of cellular stress. However, in the latter stage, when the stress is severe, it aids in cell death. Nevertheless, the essential autophagy-related genes that can prevent the onset or progression of the disease have not yet been found. It is crucial to identify these proteins since the mechanisms that drive dis-

ease development also affect autophagy.

Multifactorial diseases like NAFLD are modulated by perturbations at both the gene and the metabolite levels. Although numerous studies have been conducted, both in silico and experimental, studies that address perturbations at both these levels are very scarce. In addition, from a computational standpoint, NASH lacks a comprehensive method that simultaneously studies the molecular perturbations in NASH and determines whether genes with classification and controllability power can be a possible target.

Similarly, cancer and autophagy have always been a topic of discussion. Beclin1 is a quintessential protein in autophagosome formation and plays a crucial role in multiple autophagy-related proteins, including cancer. But, none of the proposed therapeutic methods has addressed the dynamics between DNA damage, p53 and beclin1.

To address these challenges, the broad objectives of the thesis can be stated as follows,

1. Investigation of the autophagic protein perturbations in disease using network biology approaches.

2. Investigation of the whole system protein perturbations in disease using network biology approaches.

3. To develop a mathematical model to get a mechanistic insight of some autophagic genes in cell proliferation and cell death

1.6 Thesis layout

The specific aim of this thesis involves a comprehensive analysis of the protein perturbations in diseases with an emphasis on the association of autophagy. For this purpose, mathematical modelling, clinical data and various PPI, as well as disease databases, are used in this study. Overall, this thesis is organised into six chapters containing the information explained below:

Chapter 1: This chapter addresses the “What’s, Why’s, and How’s” of autophagy. Following this, it addresses how mathematical modelling and systems biology-based applications are used to uncover the underlying mechanism of this process in multiple diseases. This chapter also

provides the history of autophagy and systems biology and various tools and databases used to study them, together or alone.

Chapter 2: The autophagy process is associated with diabetic retinopathy. However, the autophagy-related proteins targeting which can prevent or initiate the disease are yet to be identified. Addressing this, this chapter describes a multi-layer relatedness (MLR) approach to determine the relatedness of autophagic and Diabetic retinopathic (DR) proteins by incorporating both expression and prior-knowledge-based similarities. The goal of this study is fourfold: 1) identification of the topologically significant novel disease-related candidate autophagic proteins (CAP) in a PPI network constructed using prior knowledge-based information, 2) evaluation of the significance of these proteins in a gene co-expression network, and 3) in a differentially-expressed gene (DEG) network and 4) investigating the proximity of CAPs to the known disease-related proteins in the DEG network. The proteins identified through our methodology can influence the DR interactome in various layers of heterogeneity of clinical manifestations. In short, in this chapter, we have used the autophagy-DR protein interaction network to identify autophagy-related targets in diabetic retinopathy

Chapter 3: Addressing the perturbations in Non-alcoholic steatohepatitis (NASH) at both the gene and metabolic level, in this chapter, we have used a de novo methodology to identify the potential targets in NASH. We constructed a PPI network containing differentially expressed genes and significant metabolic genes associated with NASH. We applied the structural controllability in this network and identified three indispensable proteins capable of initiating a disease to a healthy transition in NASH. Interestingly, we observed one of the three identified targets to be autophagy-related, which echoes the predominance of autophagy in the NASH paradigm. However, to get a global view of the role of protein perturbation in a disease progression, we need to understand the importance of the relation of a protein with the known disease genes. In the next chapter, we captured the significance of proteins in terms of their proximity to the disease genes.

Chapter 4: A major challenge in the computational way of solving the conundrum of a disease system is to track down the effect of the potential recovery options on different layers of molecular understanding. The proteins identified from the PPI networks may not play a crucial role in

the metabolic level. The metabolic networks identify crucial reactions and the genes involved in them. However, such a network cannot provide information about how the gene products, i.e., proteins, will behave in conjunction. Also, the information about forming clusters, spreaders, role in information processing, etc., cannot be told from metabolic network analysis. Machine learning (ML) classifies proteins and can identify the nodes with the best predictive capability in the network. However, this approach alone can say nothing about metabolic adaptation, which is imperative for cell homeostasis following a physiological change. But, integrating with the ML approach in a study, the network analysis methods can be used to get more in-depth knowledge of the disease systems. Addressing all these, in this chapter, we applied a random walk restart multilayer approach to identify the proteins which remain in close proximity with the disease proteins, possess the classification capability, can control the network and eventually can alter the metabolic landscape. Here also, we found that the potential targets obtained using this methodology could affect the autophagy process. Though this network analysis gives us a global scenario of proteins and their effects on disease as a whole, it is an unmet need to understand how their modulation affects the disease dynamics, which demands the application of mathematical modelling-based studies.

Chapter 5: Finally, we have applied a mathematical model-based study to get mechanistic insight into some autophagic genes in cell proliferation and cell death. Reducing metabolic stress and increasing nutrient availability through the breakdown of cellular organelles and unfolded proteins are two ways autophagy fosters cancer growth and progression. This process is associated with DNA damage, a critical factor in cancer initiation. The guardian of the genome, p53, plays a crucial role in the repair of DNA and induces apoptosis if the damaged DNA can not be repaired. Hence, when p53 loses this ability, abnormal cell growth and, thereby, cancer initiates. Addressing these facts, this chapter describes a seven-dimensional non-autonomous ODE model to investigate the complex interplay between DNA damage, p53, autophagy, and lung cancer. The study aims to highlight the potential factors or parameters and propose that autophagic cell death mediated by perturbation of these parameters over a specified range is the way forward in lung cancer research.

Chapter 6: This chapter concludes the thesis by providing a brief summary of the work, its

contribution to the discipline of autophagy, and future directions.

2

Identification of critical autophagy-related proteins in diabetic retinopathy: A multi-dimensional computational study[1]

2.1 Introduction

Diabetes mellitus (DM) is a global epidemic that is associated with a high rate of morbidity, affecting over 415 million adults worldwide, and this number is only expected to rise to 642 million by 2040 [217]. It is a complex disease that causes both acute and chronic micro-and macrovascular complications. Patients with diabetes frequently develop ophthalmic complica-

[1]The bulk of this chapter has been communicated for possible publication.

tions such as corneal dysplasia, glaucoma, iris neovascularisation, cataracts, and neuropathies. However, the most prevalent and potentially lethal of these complications is diabetic retinopathy (DR). It occurs when damaged blood vessels in the retina allow fluid to leak into the macula, the area of the eye responsible for sharp central vision, resulting in blurred vision and eventual blindness [218]. Like many other slow-progressive diseases, DR patients remain asymptomatic at the early stages of the disease. However, once the disease reaches an advanced stage, various symptoms begin to appear that include microaneurysms, venous loops, and venous bleeding, dot and blot hemorrhages, flame-shaped hemorrhages, retinal edema, hard exudates, macular edema, etc. [219]. Blood pressure control and glycemia intervention can reduce the development risk and progression of DR at an early stage [220]. However, the underlying molecular mechanisms of DR are still not precise [221]. As the global burden of DR continues to increase year after year, it has become an unmet necessity to unravel the mystery and capture critical genes capable of regulating this nefarious disease.

As with many other diseases, DR is also susceptible to cellular degradation facilitated by autophagy. Autophagy has been shown in the literature to play a dual role in the initiation and development of DR [222]. During the early stage of the disease it protects the cells by inhibiting ER stress. However, under conditions of extreme stress, autophagy loses its protective function and contributes to the death of pericytes, an abnormality that may lead to DR [223]. Furthermore, autophagic dysfunction is implicated in the etiology of DR as an early occurrence [224]. Under hypoxic conditions, autophagy promotes angiogenesis in the retinal pigment epithelium (RPE) cells by increasing the amount of vascular endothelial growth factor (VEGF) [225]. By mediating the protective functions of retinal ganglion cells, autophagy can attenuate neurodegeneration in DR [226]. Thus, recognizing the molecular and pathophysiological aspects behind this autophagic process may aid in developing prophylactic or therapeutic approaches for DR, necessitating a computational analysis of the DR-autophagy interactome to identify the most significant molecules involved in this crosstalk.

There is a dearth of computational studies on DR. Gopalakrishnan et al. [227] studied the topology of DR proteins. However, his study was focused on the five previously experimentally identified DR genes and hence lacked a detailed insight into the entire interactome. Wang et al.

[228] studied the impact of Chinese medicine Taohang Siwu against DR, where they conducted a protein-protein interaction (PPI) network analysis by combining the targets of Taohang Siwu with the DR-related genes. Safei et al. [229] performed the protein-protein interaction (PPI) network analysis to monitor the protein alterations in rats following laser treatment. However, none of them focused on the involvement of autophagy in the disease. Recently, some computational work was performed to study the association of autophagy with DR. Gao et al. [230] performed a microarray analysis by isolating the total RNAs obtained from the retinas of diabetic mice. They identified four proteins, Bcl2, Gabarapl2, Atg4c, and Atg16L1, associated with the cell death pathways. Finally, through the qRT-PCR analysis, they found Atg16L1 to be significantly upregulated and concluded it to be a novel biomarker of DR. In another study, Wang et al. [231] identified 23 potential DR autophagy-related genes using a PPI network. They concluded that the downregulation of MAPK3 plays an important role in developing DR by regulating autophagy.

The computational analysis of a biological system requires a thorough grasp of the relatedness between the genes or their products. This relatedness can be investigated in three major ways: A) The prior knowledge-based investigation. Here, a PPI network is constructed using the information extracted from various databases. B) The expression-based similarities between the genes. It is commonly used to evaluate the conditional relatedness of the coexpression level between a pair of genes under a given situation. C) Combining these two aspects, a hybrid type of investigation can also be defined, which, based on statistical measures such as fold change, p-value, false discovery rate, etc., first extracts the genes that are significantly expressed on a certain condition, and then use the PPI databases to construct a network. Although type 'A' identifies the crucial proteins, they are extracted based on global relatedness and do not consider their expression values. The robustness of such information also needs to be improved [232]. Type 'B' covers these expectations, but genes with similar co-expression may not always have related functions [233]. Again, because these associations are based on specific data, they may not always be global. Type 'C', or the hybrid measure, covers both perspectives but may suffer from various aspects such as quantitative cut-off or the inherent noise in the data. To gain a complete understanding of a disease, these three types of investigations should be conducted

simultaneously to identify the set of proteins capable of driving the system without being influenced by global relatedness, expression similarity, or biasedness of certain quantitative cut-off criteria.

The computational studies in DR are limited to understanding only one of the above three perspectives. DR is associated with hyperglycemia, oxidative stress, hypoxia, endoplasmic reticulum (ER) stress, and nutrient starvation, all of which are associated with autophagic flux activation. However, the computational effort has been made so far to investigate the DR-autophagy association are very few. Hence, it won't be an embellishment to assert that there remains a lack of holistic study of DR that covers a broad spectrum of disease severity. Again, understanding the molecular and pathogenic mechanisms underpinning the autophagic process may help develop DR prevention or treatment strategies. Therefore, in this study, to uncover novel autophagy-related proteins involved in disease pathogenesis, we have employed a multi-layer relatedness (MLR) approach. The objective of MLR is to determine the relatedness of autophagic and DR genes by incorporating both expression and prior-knowledge-based similarities. The goal of this study is fourfold: 1) identification of the topologically significant novel disease-related candidate autophagic proteins (CAP) in a PPI network constructed using prior knowledge-based information, 2) evaluation of the significance of these proteins in a gene co-expression network, and 3) in a differentially-expressed gene (DEG) network and 4) investigating the proximity of CAPs to the known disease-related proteins in the DEG network. The proteins identified through our methodology can influence the DR interactome in various layers of heterogeneity of clinical manifestations.

2.2 Methodology

The method developed in this paper can be divided into four categories. i) The construction of the prior knowledge-base network and identification of candidate autophagic proteins (CAPs). ii) Evaluating the importance of the CAPs in a gene co-expression network. iii) Evaluating the importance of the CAPs in a differentially expressed gene (DEG) based PPI network. iv) Using a network propagation theory to identify the disease-associated genes. The methodology

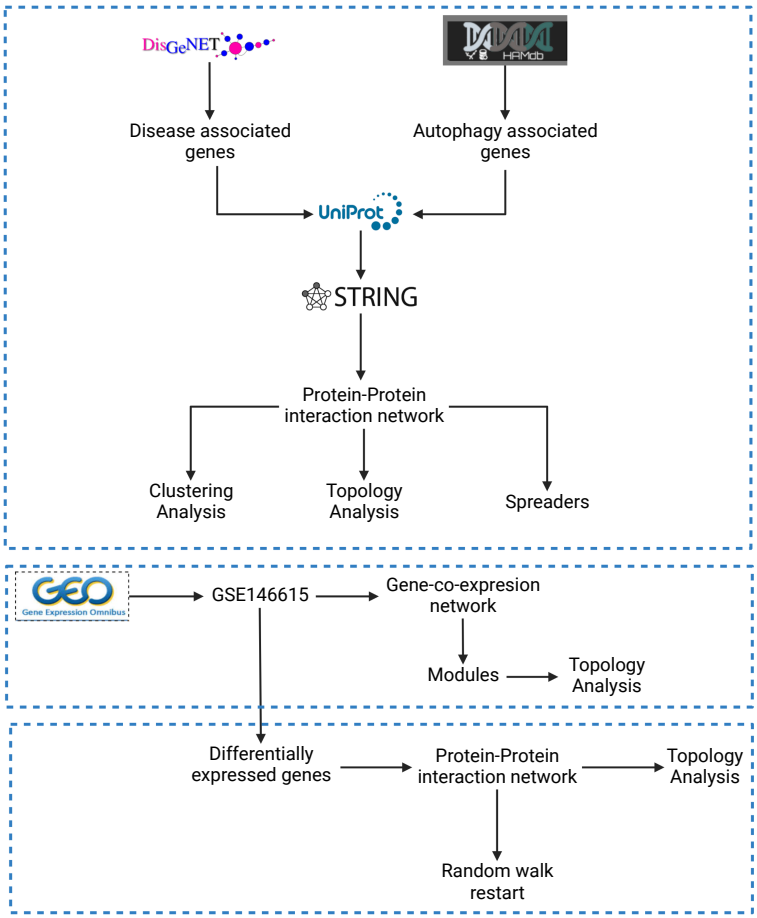


Figure 2.1: **The workflow used in this study.** The first dotted box showed the construction and analysis of the DR-autophagy interactome, the second and third boxes represented the integration of clinical data to study the gene co-expression and differentially expressed gene network.

is shown in **Figure 2.1**.

2.2.1 Extracting disease and autophagy-related genes

The DR related genes (DRGs) are extracted from the DisGeNET [234] database. Among the autophagy databases available in the literature, HAMDB [207] was found to be more information enriched [66], and hence all the autophagy-related genes (ARGs) are collected using this database.

2.2.2 Protein-protein interaction network construction

The DRGs and ARGs are converted to proteins using the UniProt database [235] and are hereafter referred to as DR-related proteins (DRPs) and autophagy-related proteins (ARPs), respec-

tively. To construct a PPI network of the selected proteins, we took the help of the Search Tool for Retrieval of Interacting Genes/Proteins (STRING) database [215]. The resulting protein-protein interaction (PPI) network contains interactions between proteins on both a physical and functional level. Each of the interactions in STRING is assigned a confidence score. The higher the score, the more probable the proteins will interact. We took the highest confidence score of 900 to construct a highly significant network.

We first constructed a network by taking the ARPs and the DRPs. However, the network was found to be disintegrated into different disconnected components. To circumvent this, we need to identify a set of minimal nodes to connect these components. To ensure unbiasedness, we used all the first neighborhoods of the ARPs, and DRPs and created one giant connected network. Few other studies have also reported the use of first neighbours to construct a PPI network [236, 237].

2.2.3 Methods for the analysis of PPI networks

To identify the core set of nodes in the network, we took the help of three classical centrality measures: degree, betweenness, and closeness. The degree centrality measures the connectivity of the proteins in the PPI network. The proteins with higher connectivity (usually those with a degree twice the average degree) are known as hub proteins. The betweenness centrality measures how much information passes through a specific node in the network. The proteins with a high level of betweenness are termed bottlenecks. On the other hand, closeness centrality quantifies how close a node is to other nodes in the network. The top 5% scoring proteins (TSPs) were chosen for each of the three centralities, and the common proteins between them were extracted. These TSPs are topologically significant and serve as the network's backbone. Next, from the TSPs, we identify the autophagic proteins (henceforth referred to as candidate autophagic proteins, CAPs that were not previously identified as DR-associated in DisGeNET [234]. Mathematically, these CAPs can be represented as

$$CAP = ARP \cap (TSP \setminus DRP). \tag{2.1}$$

Coreness is a global property of a network and examines the capability of a node to belong to a highly connected cluster. It also tells how much influential a node is in terms of the propagation of information throughout the network. To identify the influential spreaders in the network, we opt for the k-core decomposition method. Let $G(V,E)$ is an undirected and unweighted graph where V is the vertex set, and E is the set of the edges. A subset G_2 of G_1 is defined as k-core of G_1, if it is a maximal subgraph of G_1, in which all nodes have degree at least k. The k-core analysis enables the detection of interesting structural properties that are not captured by many other network topological measures. The central cores contain more strongly connected vertices and have many distinct paths connecting them [238].

2.2.4 Clinical data

From the Gene Expression Omnibus (GEO) [239] database, we selected the dataset GSE146615 [240] which contains the information on lymphoblastoid cell lines of seven healthy, seven diabetic individuals without DR, and eight diabetic individuals with DR. The dataset was quantile normalized and contained three biological replicates per individual and treatment while three individuals had five biological replicates.

2.2.5 Data pre-processing

The probe to gene mapping was done using the Illumina HumanHT-12 V4.0 expression beadchip platform. If a gene had any null values across the samples, it was removed from the study. We next used the *'filloutliers'* function in MATLAB to detect outliers by the *'median'* value of the gene across the samples and replaced them using the nearest value of the gene across the samples.

2.2.6 Weighted Gene co-expression network analysis (WGCNA)

To investigate the relatedness of the crtitical genes in DR, we opt for a gene co-expression analysis using WGCNA [130] which identifies the critical modules in the network. Modules are defined as groupings of genes that share similar expression patterns and are frequently

functionally connected and co-regulated. It is conceivable that such coordinated gene activities contribute significantly to the complexity of biological processes and pathways. Among the three set of cohorts, we considered the healthy group and the diabetic individuals with DR group.

2.2.6.1 Parameters in WGCNA

Here, in our analysis using WGCNA, we have used maxBlockSize of 20,000, minModuleSize of 30, and mergeCutHeight of 0.3. We have considered the disease status as clinical trait information. Here, if a sample comes from a healthy individual, it receives a score of 0, and if it comes from a diseased individual, it receives a score of 1. The module-trait associations were estimated by assessing the influence between the module signature and the phenotype (clinical traits), allowing for easy identification of highly correlated modules with the phenotype. Finally, for each module, a PPI network is constructed using the STRING database, and the influential nodes are identified by measuring the topological properties.

2.2.7 Identification and analysis of differentially expressed genes

The co-expression analysis identifies genes that show a coordinated expression patterns across the group of samples. However, such networks typically do not convey causal information or distinguish between regulatory and regulated genes [241]. To mitigate this, we opt to find differentially expressed genes (DEGs) in the dataset GSE146615. For this, we first removed all the genes which possessed any null value across the samples, and the remaining genes were then carried out for further analysis. Finally, we have taken the log fold change cut-off of ± 1 to identify the DEGs in the network.

2.3 Results

2.3.1 Prior knowledge-base investigation of the DR-autophagy interactome

We identified 645 DRGs from DisGeNET and 551 ARGs from HAMdb. The UniProt mapping of these genes has reduced these numbers to 573 DRPs and 482 ARPs, respectively. The network consisting of these proteins and their immediate neighbors had 7856 proteins and 583928 interactions (**Figure 2.2**). The topological properties of the top ten hubs are shown in (**Figure 2.3A**). Among these ten hubs, UBB and APP were ARPs. The top 10 biological processes associated with the proteins in the network are shown in (**Figure 2.3B**). The top 5% of nodes (rounded to 393 nodes) in each of the degree, betweenness, and closeness centrality categories were extracted, and 104 TSPs were identified (refer to Section 2.2.3). The overlap between the ARPs, DRPs, and the TSPs are shown in **Figure 2.3C**.

2.3.2 Candidate autophagic proteins (CAPs) control the network

Using equation (2.1), we identified 11 CAPs among the TSPs: APP, ATG7, GNAI3, HDAC1, HSP90AA1, HSPA8, KRAS, PIK3R1, TP53, UBB, and UBR4. Investigating their topological significance, we found that UBB is the third most connected node in the network while APP, PIK3R1, and TP53 are ranked among the top fifty hubs in the network. Similarly, all the CAPs were bottlenecks in the network, with UBB and TP53 being ranked fifth and seventh, respectively. UBB was ranked fifth in the closeness category, while five other proteins, HSP90AA1, HSPA8, KRAS, PIK3R1, and TP53, were ranked among the top fifty. These findings imply that CAPs play a critical role in regulating the DR interactome.

We found that three CAPs, ATG7, UBB, and UBR4, were present in the innermost cores, while the rest were present in the top ten innermost cores. This indicates that these nodes can disseminate information to a wider portion of the network. The topological properties of the CAPs are shown in **Figure 2.5A**, while the enriched biological processes are shown in **Figure 2.5B**. These proteins are enriched with cell death-related pathways, autophagy-related pathways, protein phosphorylation and modification, and cytokine-related pathways. All these processes are

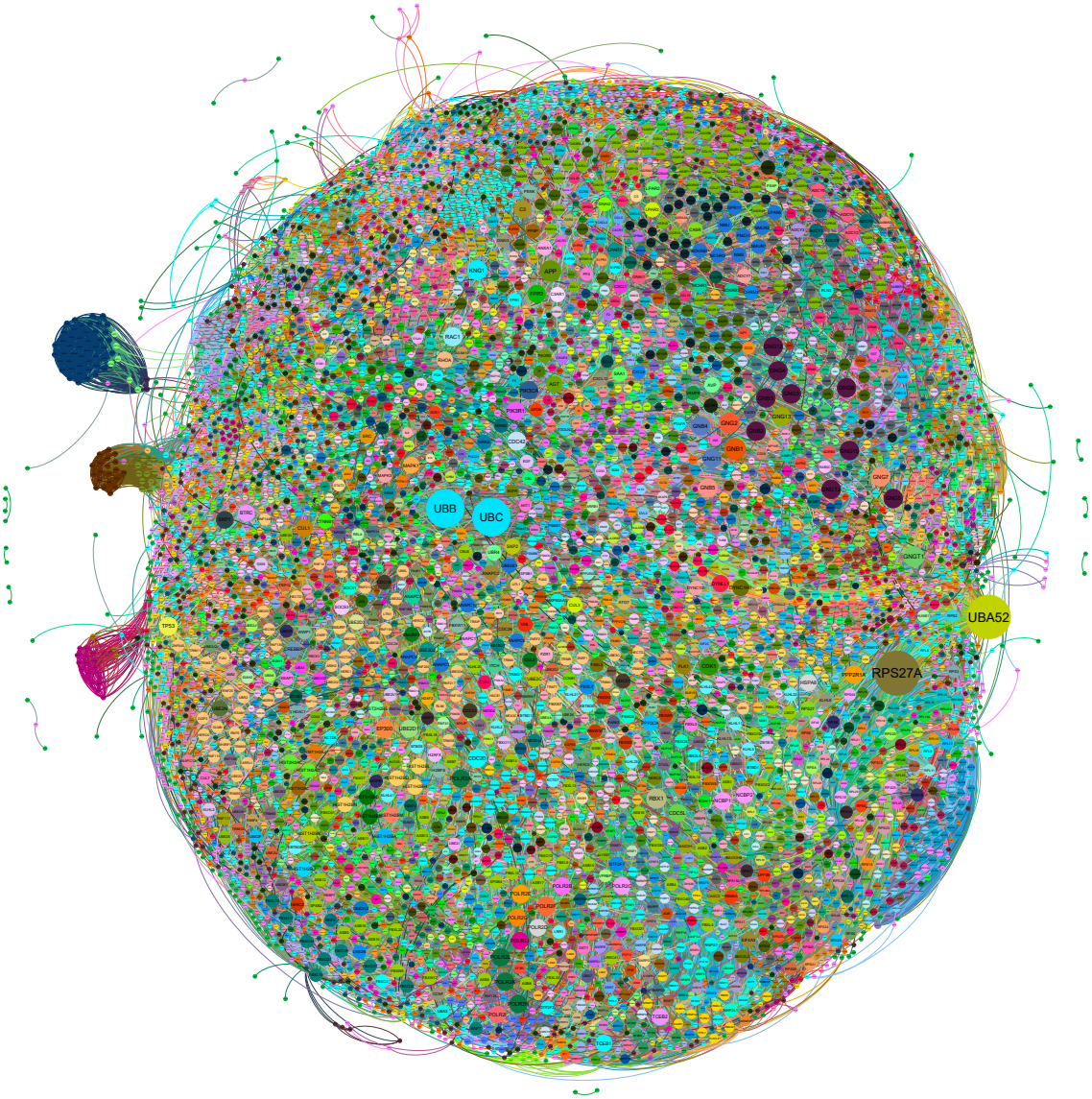


Figure 2.2: **The DR-autophagy interactome.** Here the nodes are sized and colored according to their degree.

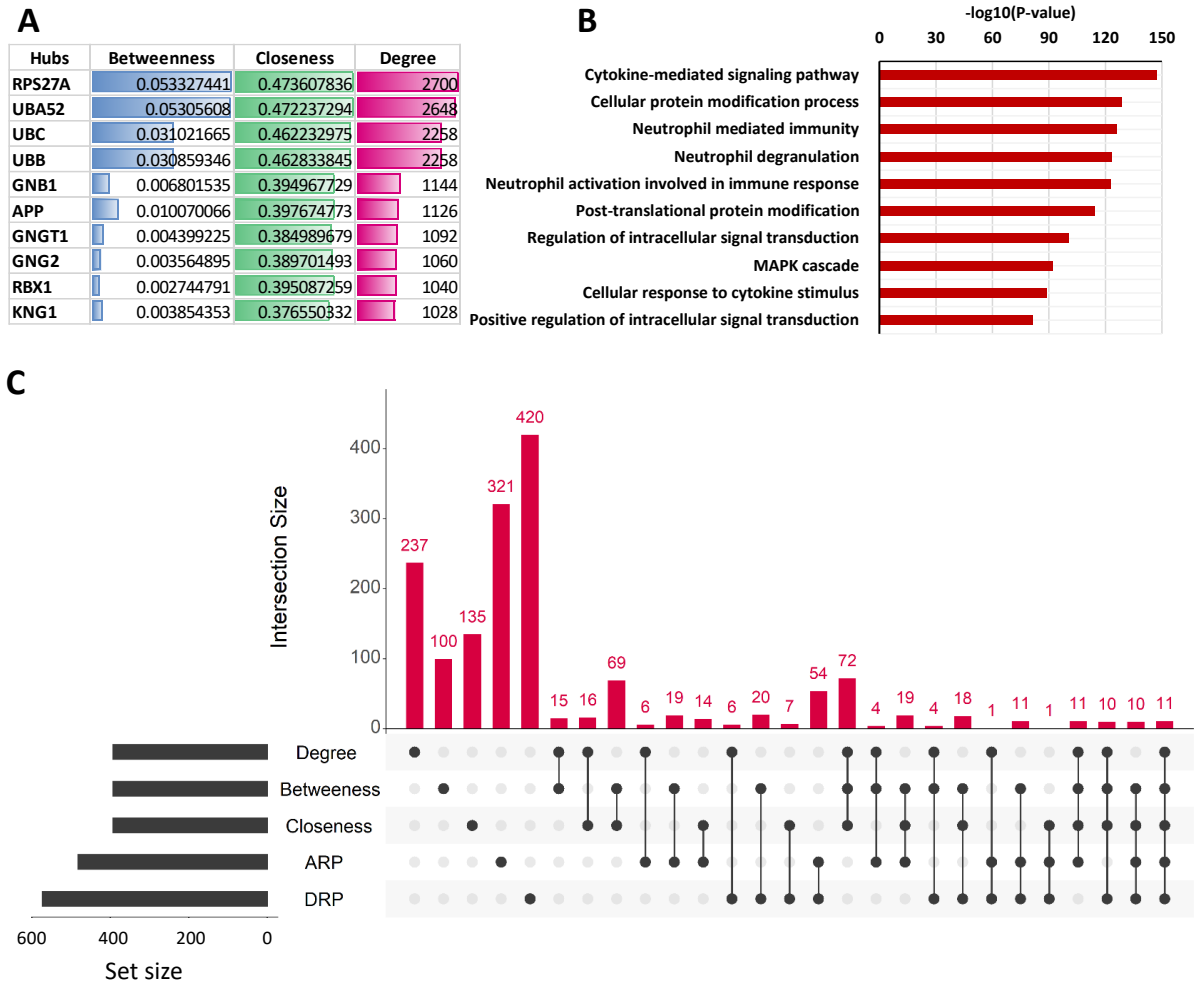


Figure 2.3: **Analysis of the autophagy-DR PPI network.** (A) The topological properties of the hubs in the network. A majority of these proteins also had high betweenness and closeness centrality. (B) Top 10 enriched biological processes associated with the proteins in the network. The enrichment analysis is done using Enrichr [242]. (C) The UpSet plot of the overlap between ARPs, DRPs, and the top 5% proteins in the degree, betweenness, and closness centralities. It can be seen that the ARPs were enriched in each of the categories.

found to play a crucial role in DR [175, 243, 244]. Clustering analysis facilitates the elevation of network nodes to a more comprehensive level by subdividing them into a smaller number of clusters. We used the Molecular Complex Detection (MCODE) [153] plugin in Cytoscape [121] to perform clustering analysis. The selection parameters for the MCODE analysis were set as follows: MCODE scores $\geq$5, degree cut-off = 2, node-score cut off = 0.2 and k-core = 2. From the obtained clusters, we took a MCODE cut-off of 10 and found 27 significant clusters. APP, ATG7, GNAI3, UBB and UBR4 were found in the topmost cluster. HSPA8 and PIK3R1 were present in the 2nd cluster, HSP90AA1 on the 5th cluster, KRAS and TP53 on the 8th cluster and HDAC on the 12th cluster. The top 20 mcode clusters are shown in **Figure 2.4.**

It is worth noting that the specificity of the CAPs is fourfold: (i) they are derived from a network containing the disease and autophagic proteins, (ii) they are among the topologically strong proteins in the network, (iii) they were not previously reported to be associated with DR, and (iv) their enriched pathways are also reported to play a role in disease physiology. We believe that all these CAPs, or at least some of them, will also play a critical role in a gene-gene association network that is constructed on the basis of expression similarities. To validate this, we took a clinical dataset from GEO and proceed to gene co-expression analysis.

2.3.3 Co-expression network analysis

The expression values of 18972 protein-coding genes across the 15 samples of the dataset GSE146615 were taken to WGCNA for co-expression analysis. All the samples were found to be well clustered, and no outliers were detected (**Figure 2.6A**). In this study, the power of $\beta = 12$ was selected as the soft-thresholding parameter to ensure a scale-free network (**Figure 2.6B**). We have constructed a signed co-expression network, and using the average linkage hierarchical clustering, a total of twelve modules were identified (**Figure 2.6C**). Except for the grey, each colour represents a gene module. The grey color indicates genes that are not module-assignable. Correlations between these co-expression modules and clinical traits were quantified using the correlations between module eigengenes and clinical traits. The groups of correlated eigengenes are identified using the eigengene dendrogram and heatmap (**Figure 2.6D**). The associations between the modules and the clinical trait were quantified using the

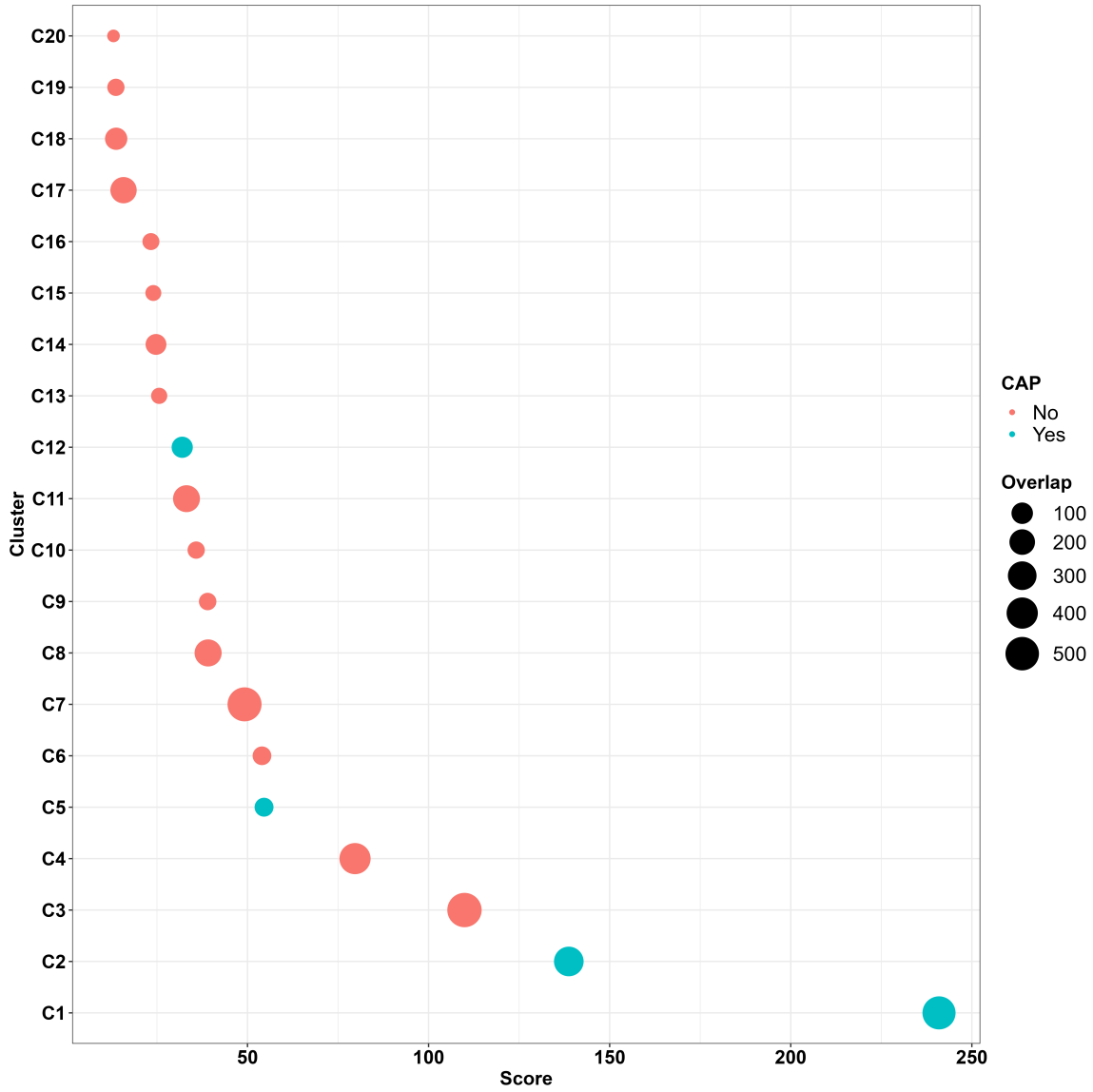


Figure 2.4: **The top 20 Mcode clusters.** The clusters which contain the CAPs are shown in cyan color.

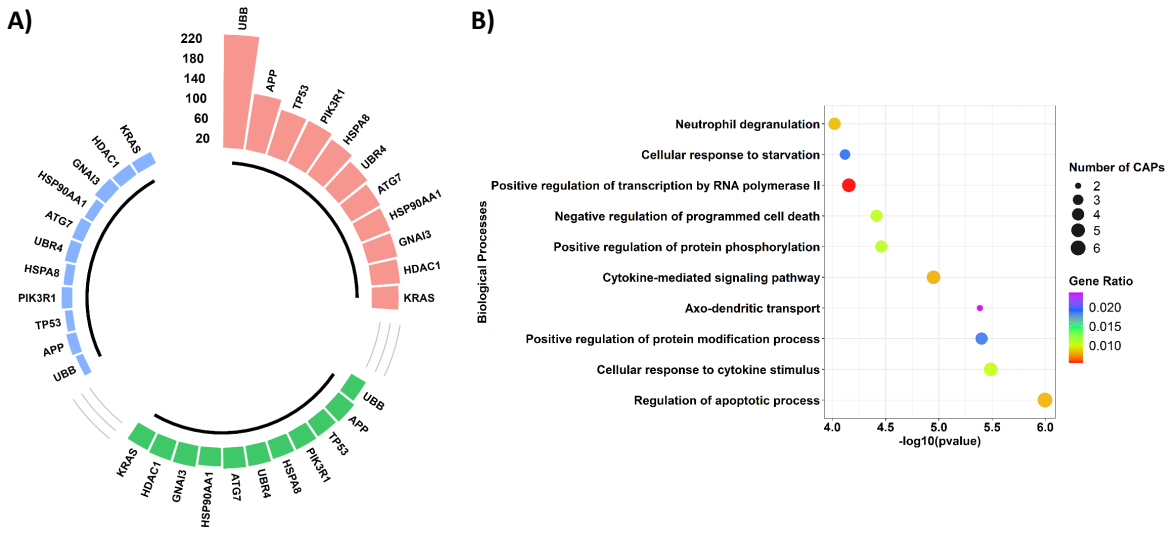
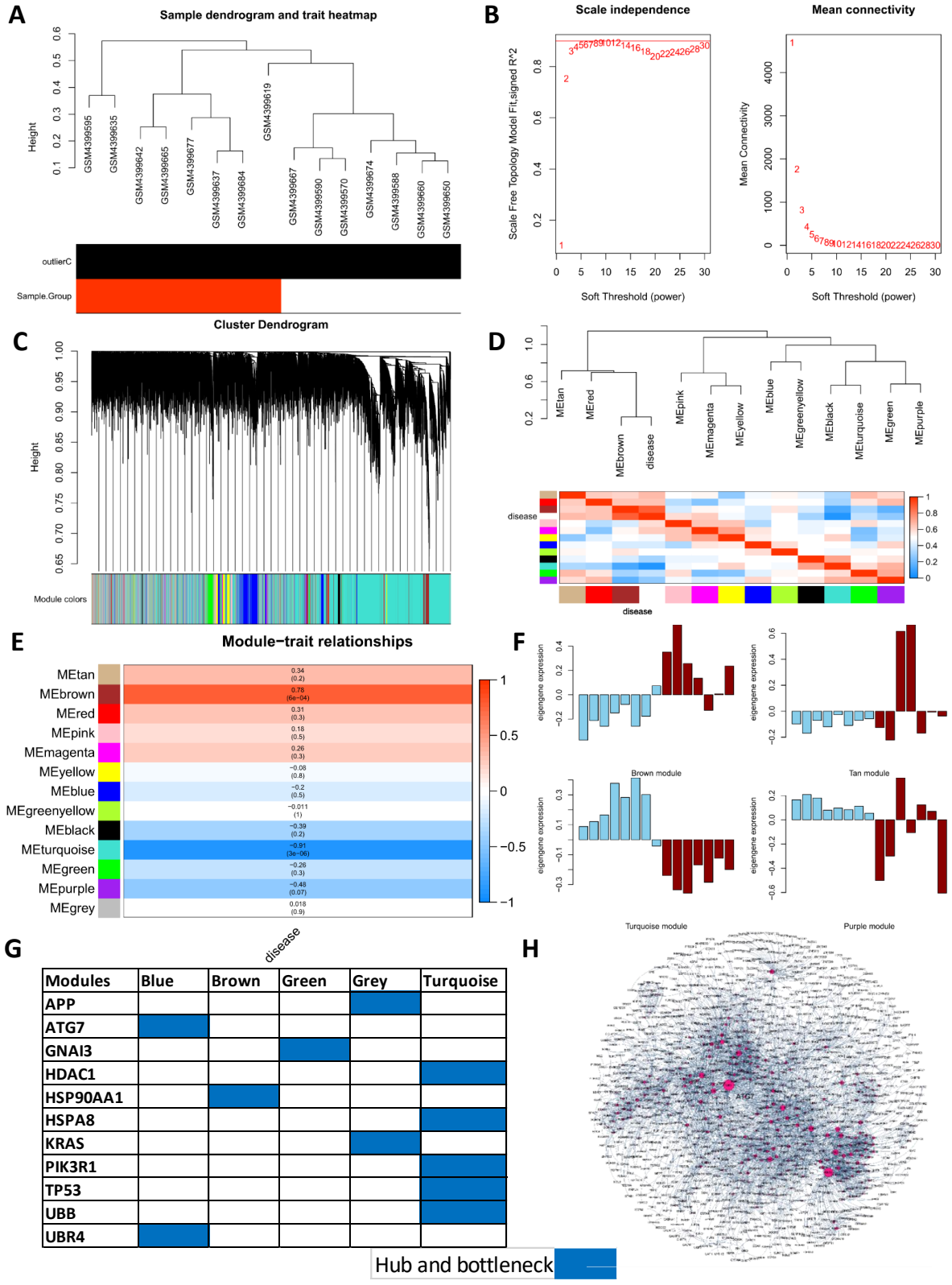


Figure 2.5: **Topological properties and the functional enrichment of the CAPs.** (A) The topological properties. To facilitate visualization, the original betweenness and closeness centrality values for each protein are transformed to the -log10 scale and then multiplied by ten. On the other hand, each protein's degree centrality value is reduced by tenfold. (B) The top ten biological processes associated with the CAPs. The enrichment analysis is done using Enrichr. The nodes sized are proportional to the quantity of candidate autophagic proteins in the process, while the colors represent the gene ratio.

correlation between the eigengenes and the clinical traits (**Figure 2.6E**).

The results indicated that the brown module strongly correlates with disease status whereas the turquoise module has a strong negative correlation. Mild positive correlations were seen with the tan and red modules, while black and purple modules showed mild negative correlations with disease. The expression of the eigengenes across the samples (**Figure 2.6F**) demonstrates that the genes in the brown module are significantly over-expressed, while the genes in the turquoise module are significantly down-regulated in the DR conditions. All the other modules showed a highly heterogeneous behavior in the eigengene expression throughout the samples. Among the thirteen modules, five (including the grey) contained at least one CAPs. To see the impact of these CAPs, we constructed a PPI network of these modules by mapping each of them to N_{900}. Because a module network may encode a pathway or a protein complex, these unique networks are beneficial. Interestingly, in all of the modules, the CAPs were among the hubs and bottlenecks (**Figure 2.6G**). This justifies that they are among the critical proteins in the co-expression modules and hence, are worthwhile to be further evaluated. An example of one such module is shown in **Figure 2.6(H)**.



Modules	Blue	Brown	Green	Grey	Turquoise
APP				■	
ATG7	■				
GNAI3			■		
HDAC1					■
HSP90AA1		■			
HSPA8					■
KRAS				■	
PIK3R1					■
TP53					■
UBB					■
UBR4	■				

Figure 2.6: **WGCNA analysis.** (A) Clustering dendrogram of the 15 samples. The Red denotes the normal while the white color represents the disease group. (B) The image on the left illustrates the soft-thresholding power versus the scale-free fit index. The right panel shows the mean connectivity versus soft-thresholding power. (C) Modularization of the co-expressed genes based on the topological overlap matrix. A total of 13 modules (including the grey module) were identified. (D) The eigengene dendrogram and heatmap identify groups of correlated eigengenes. (E) The module-trait relationships. Each row corresponds to a module eigengene, and the values in the column represent the correlation with the disease status. (F) The eigengene expression throughout the samples for the brown, tan, turquoise, and purple modules. (G) The modules which contain CAPs. (H) The PPI network of the blue module which contains the CAP ATG7.

2.3.4 Differentially expressed genes (DEG) network analysis

The differential gene expression analysis of the dataset GSE146115 revealed 6457 up-and 1115 down-regulated genes. We found that two of our CAPs, PIK3R1, and APP, were respectively up and down-regulated in the dataset **Figure 2.7A-B**. To investigate the alterations at the gene level, the KEGG enrichment analysis using the package clusterProfiler [245] was performed (**Figure 2.7**C). The analysis revealed that the PI3K-AKT pathway was the most enriched pathway, followed by neuroactive ligand-pathway interaction. Recent investigations have discovered that the PI3K/Akt/mTOR proteins are highly expressed in the retinal tissue of diabetic rats [246], justifying the predominant enrichment of the DEGs in this pathway. The prominence of neuroactive ligand-pathway interaction pathway on the enriched list suggests that neuroprotective factors may be depleted, and retinal neurodegeneration may occur during diabetes.

To construct an undirected PPI network, we mapped the DEGs to the N_{900} network. However, the constructed network was found to consist of many separate components. We used the STRING database to connect them and extract the fewest possible nodes. These additional nodes are referred to as mediators. We found that 417 mediators were required to connect the DEG network (**Figure 2.7D**). Notably, we found that two of our CAPs, TP53 and HSP90AA1, function as mediator proteins. This indicates that, despite their relatively low levels of expression, these proteins are critical for propagating information throughout the network. The final network contained 4466 nodes and 107832 edges (**Figure 2.7E**). Interestingly, all four CAPs were found to be hubs and bottlenecks in the network. The topological properties of the CAPs are provided in the **Table 2.1**.

Table 2.1: **Topological properties of the CAPs.**

CAP	Betweenness	Closeness	Degree
TP53	0.051276521	0.360080645	388
APP	0.016289037	0.338206332	464
HSP90AA1	0.026229137	0.367247903	256
PIK3R1	0.019253502	0.349866792	392

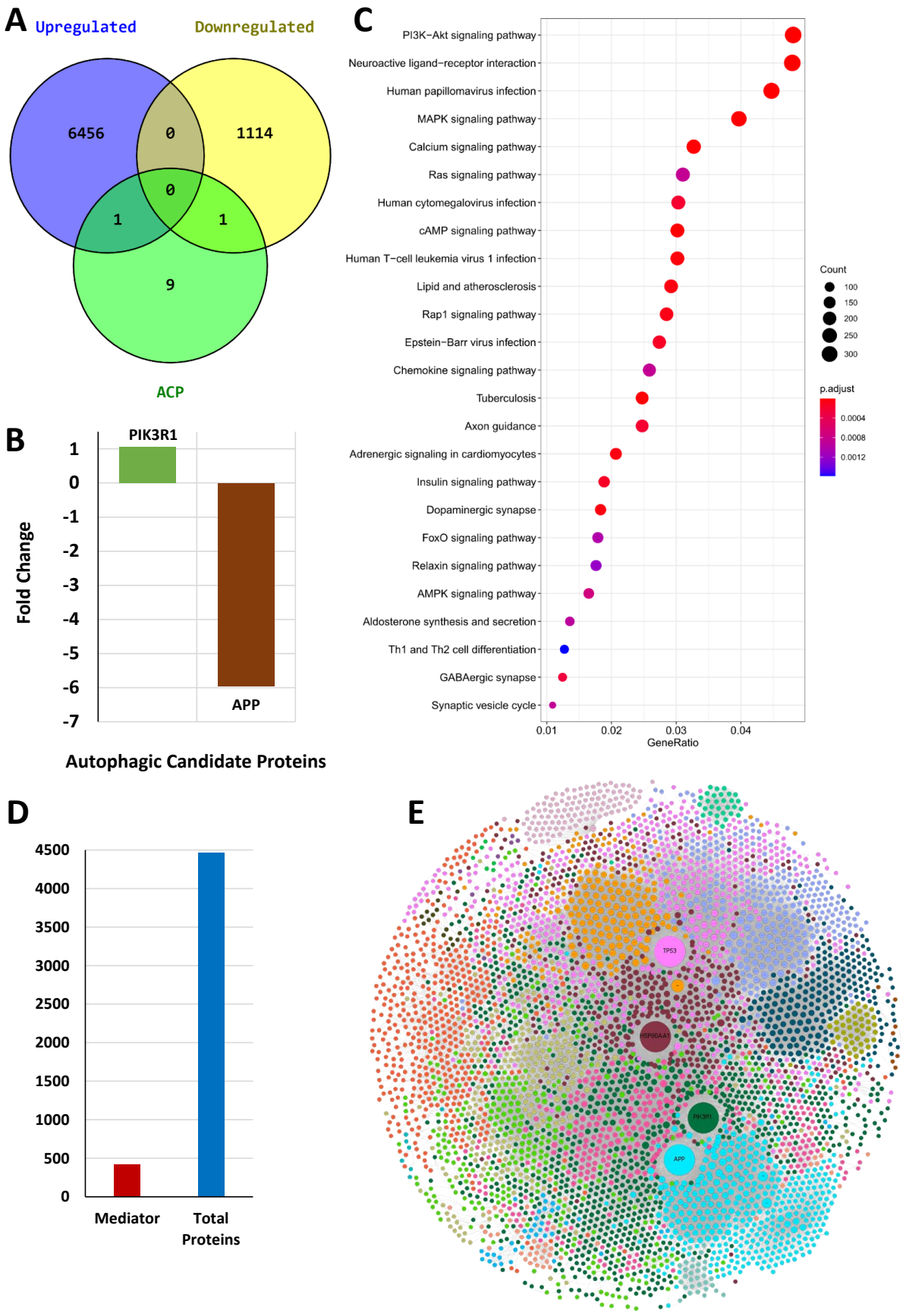


Figure 2.7: **Construction of the DEG network.** (A) The numbers of CAPs among the DEGs. (B) Expression values of the two CAPs. C) The KEGG pathway enrichment of the DEGs. (D) Total nodes and the number of mediators in the network. (E) The DEG network. Here, the nodes are colored according to their modularity index. The large labeled nodes are the four CAPs present in the network.

Further, to see which proteins these CAPs interact with, we opt for a first neighborhood analysis. The most interactors were found for APP, followed by PIK3R1 and TP53. On the other hand, PIK3R1 had the highest density of disease genes in its first neighborhood, whereas TP53 had the least percentage (**Figure 2.8A**). The presence of the most number of disease genes in the PIK3R1 interactome may be a result of its interaction with the PI3K/AKT pathway, which is critical for the progression of DR. The overlap between the neighbors of these proteins are shown in a UpSet plot (**Figure 2.8B**). As can be seen, these proteins have no common neighbors. A pairwise analysis revealed that APP and PIK3R1 shared the most neighbors, whereas APP and HSP90AA1 had no common neighbors. The interactors of these four proteins are shown in **Figure 2.8C-F**. To establish the relationship of the network proteins with the disease proteins, we opt for the random walk with restart (RWR) algorithm. The seed nodes for the analysis are identified using the equation, $S = DRP \cap V$, where V is the set of DEGs in the network. Each node in the network was assigned a score denoting its possibility of being a novel DR protein using the RWR algorithm. The higher the score, the more likely the protein is DR-related. We found that PIK3R1 is the top-ranked protein in the list while TP53 obtained the third rank. The remaining two proteins APP and HSP90AA1 were ranked at 21^{st} and 24^{th} positions. Thus, these CAPs are deemed to be in close proximity with the disease-related proteins in DR.

2.3.5 Discussion

Diabetic retinopathy (DR) prevalence has reached epidemic proportions [247]. It is a slowly progressive disease that arises from the complications of diabetes and in the last two decades, has become a global burden [221]. Unfortunately, currently available treatments for DR are invasive, less effective, and focus primarily on the chronic stages of the disease, with a slight improvement in vision repair [218]. The literature review established that the factors associated with DR, such as hyperglycemia, hypoxia, oxidative stress, ER stress, and nutrient deprivation, are all strongly related to the activation of autophagic flux. However, the critical autophagic genes required for the initiation and development of DR remain unknown.

This study developed a multi-level relatedness approach to find the novel autophagy-related

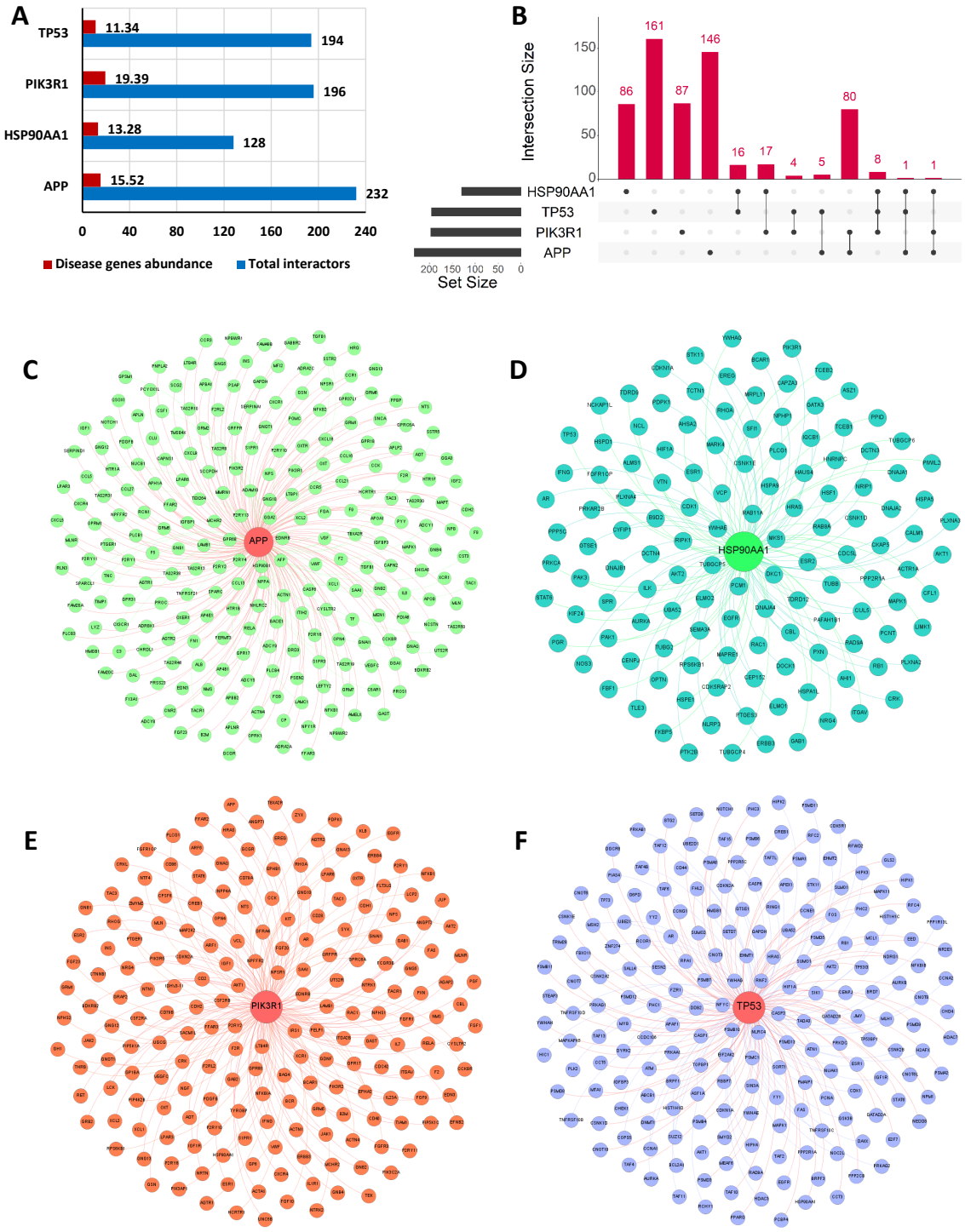


Figure 2.8: **First neighborhood analysis of the CAPs in the DEG network.** (A) The number of interactors of each CAPs. (B) The UpSet plot of the overlap between the neighbors of these CAPs. (C-F) The first neighborhood network of the CAPs.

proteins capable of regulating the DR interactome. We embraced a topological viewpoint to measure the vitality of the nodes in the network, which mainly included the degree, betweenness, closeness, and coreness of the nodes. We also performed a random walk restart analysis to explore the functional association of identified novel proteins with the already known DR proteins. MLR is a novel technique for determining the influence of biological molecules (genes or their progeny, proteins in our study) across many biological layers. Identifying crucial proteins in a network constructed from prior knowledge base information aids MLR in capturing the global relatedness of proteins. We used the high confidence network of STRING database to perform this analysis. The hub proteins in this network were found to be associated with development and progression of DR. For example, the suppression of cytokines, which are the key regulators of inflammations, have been shown to protect retinal capillaries from pathological alterations in animal models [248]. In the retinas of diabetic animals and patients, several physiologic and molecular alterations compatible with the role of inflammation have been observed. In animal models of diabetes, it has been demonstrated that inhibiting these inflammatory changes affects the development of retinal abnormalities [249]. While such relatedness may not always translate into significant biological activity, it can be aggregated with biological data in a high-throughput manner to do integrated analysis, such as creating the target landscape of the disease. Hence, this global influence is then converged to a clinical dataset to quantify the relevance of these proteins depending on their expression-based similarity. This relevance is quantified on two levels: first, at the gene level, using WGCNA, which enables a better understanding of which regulators may be driving transcriptional profiles during disease progression. The second level includes the investigation of this relevance in a network constructed using the DEGs in the clinical dataset. Leveraging this pipeline, we finally identified four proteins, TP53, HSAP90AA1, APP, and PIK3R1. However, APP was placed in the grey module in the co-expression analysis. We, therefore, focus on the remaining three CAPs and their associations with DR.

These three proteins are connected with several crucial proteins that regulate multiple characteristics of DR development. For example, TP53 interacts with CDK, which prevents angiogenesis by triggering cell cycle arrest and apoptosis [250]. HMGB1 is another interactor of

TP53 that mediates diabetes-induced damage in retinal pericytes [251] and Muller cells [252]. Similarly, PIK3R1 interacts with JAK1, whose inhibition ameliorates the blood-retinal barrier dysfunction [253]. We extracted the first neighbourhood of the three CAPs from N_{900} and performed enrichment analysis using Enrichr [242]. We found several ways these proteins could regulate the DR interactome exist. Il-1β play a significant role in DR as it expedites apoptosis of retinal capillary cells through activation of NF-κB [254]. The neighbourhoods of the three CAPs, TP53 and PIK3R1 are enriched with the IL-1β signalling pathway. There are mounting evidences that diabetic retinopathy is highly associated with the alterations in the Wnt-signaling pathways [255]. Enrichment of the biological process in the TP53 neighbourhood highlights their importance in DR pathogenesis. In all neighborhoods, we found that cytokine modulation is an enriched process. The literature has mentioned that the concentrations of various cytokines increase with the DR severity. The role of angiogenesis and its key mediator, VEGF, has been the most researched aspect of DR. Several anti-VEGF therapies have been proposed in DR. The neighbourhood of two CAPs, HSP90AA1 and PIK3R1, are found to be enriched with the VEGF signalling pathway, echoing their importance. Thus, we conclude that these proteins play a crucial role in the DR, and their modulation will affect the disease pathology. Their relation with DR-associated processes is summarized in **Table 2.2**.

Nevertheless, the association of proteins with essential pathways does not always fully justify the protein-disease association. In order to gain a complete understanding, it is necessary to examine their association with disease hallmarks. In DR, such characteristics include angiogenesis, death of retinal pigment epithelial cells, pericyte cells etc. Angiogenesis is mainly mediated by the vascular epithelial growth factor, VEGF. Its concentration is reported to be significantly correlated with the DR severity. Therefore, evaluation of VEGF expression has become a common method for determining angiogenesis and, by extension, the influence of a protein in DR. The primary function of RPE cells is to maintain retinal homeostasis by a series of secretory factors [267]. The loss of RPE cells has been implicated in the pathogenesis of DR [268]. Pericytes are essential components of the retina, and its microvasculature [269]. They have crucial roles in angiogenesis, vascular remodelling, regression, stabilization, and the formation and maintenance of the blood-brain barrier (BBB) and blood-retinal barrier

Table 2.2: **Association of the CAPs with DR.** The biological process enrichment of the neighbourhood of the CAPs is performed using Enrichr [242], and the DR-associated significant processes are selected based on their p-value $\leq$0.005. The literature evidence for DR-associated processes are given in the last column.

CAP	Process	P-value	Reference
TP53	Il-1 Signalling Pathway	3.50E-17	[254]
	Wnt Signalling Pathway	1.56E-16	[255]
	Cytokine Mediated Pathway	2.54E-12	[244]
	Histone Modification	1.74E-07	[256]
	Response To Oxidative Stress	3.82E-06	[257]
	Positive Regulation Of Blood Vessel Endothelial Cell Migration	0.001219	[258]
	Response To Insulin	0.001386	[259]
	Nik/NF-κB Signaling	1.29E-20	[260]
	Cellular Response To Hypoxia	3.56E-21	[259]
HSP90AA1	Vascular Endothelial Growth Factor Receptor Signaling Pathway	7.66E-18	[261]
	Regulation Of Apoptotic Process	2.11E-09	[262]
	Cytokine-Mediated Signaling Pathway	1.00E-07	[244]
	Positive Regulation Of Blood Vessel Endothelial Cell Migration	6.30E-07	[258]
	Response To Insulin	1.73E-05	[259]
	Response To Reactive Oxygen Species	3.91E-05	[263]
	Regulation Of Canonical Wnt Signaling Pathway	0.001288	[255]
	Regulation Of Nik/NF-κB Signaling	0.001789	[260]
PIK3R1	Cytokine-Mediated Signaling Pathway	1.56E-21	[244]
	Inflammatory Response	1.39E-13	
	Vascular Endothelial Growth Factor Receptor Signaling Pathway	2.22E-12	[261]
	Insulin Receptor Signaling Pathway	1.34E-10	[259]
	Positive Regulation Of Angiogenesis	2.26E-07	[264]
	Regulation Of Nik/NF-κB Signaling	1.21E-06	[260]
	Cellular Response To Interleukin-6	7.20E-06	
	Response To Insulin	1.93E-05	[259]
	Response To IL-1	2.25E-05	[254]
	IL-12-Mediated Signaling Pathway	8.70E-05	[265]
	Cellular Response To Lectin	9.79E-04	[266]

Table 2.3: **Experimental evidence from available literature.**

gene name	DR hallmark	*in vitro*	reference
TP53	Retinal pigment epithelial cell	High glucose induces p53 level in RPE cells and induces cell death.	[270]
		Knockdown of TP53 Inhibits RPE Apoptosis	[271]
	Pericyte cell	Increased O-GlcNAcylation of p53 leads to pericyte loss and microvascular dysfunction	[272]
	Angiogenesis	p53 rapidly induces VEGF transcription upon hypoxia exposure.	[273]
HSP90AA1	Angiogenesis	HSP90AA1 correlates with the upregulation of proteins in the VEGF pathway.	[274]

(BRB). Their demise results in the regression of the retinal microvasculature, which results in fluid leakage, leukocyte adherence to the vasculature, and hypoxia in the injured area. This eventually paves the way for DR. Various studies have shown the loss of ratinal ganglion cells following the progression of diabetes. Two of our proposed targets, TP53 and HSP90AA1, were experimentally observed to be associated with the DR hallmarks. These associations are shown in **Table 2.3**. Thus, these experimental observation collected from literature brings credibility to the results, which experimental biologist can explore further in understanding DR progression and prevention.

2.4 Conclusion

In this study, we have addressed the crosstalk between autophagy and DR through a multilayer relatedness approach. The analysis leads to the identification of three novel autophagy-related proteins, which can modulate the progression and pathogenesis of DR. The specificity of these proteins are six-fold. i) They are autophagy-related proteins, ii) not reported to be associated with DR, iii) topologically sound in the autophagy-DR interactome, iv) they are present in crucial modules in the co-expression analysis, v) important proteins in the DEG network,

and vi) possess a high-functional association with the DR associated proteins. To summarize, these proteins are critical to regulate the DR interactome and shed light on previously unknown aspects of the disease.

3

De novo analysis of a protein-protein interaction network reveals potential targets in NASH[1]

In Chapter 2, our study on understanding protein perturbation was limited to addressing the interplay of autophagy and disease proteins. However, since we are focussed on only the autophagy perspective, some of the crucial information about the progression of the disease remain incarcerated within the perturbations of proteins which were not associated with autophagy. Therefore, for a comprehensive understanding, it is an unmet need to study all the perturbations inside a disease system. Addressing this, in this chapter, we have tried to identify

[1]The bulk of this chapter has been communicated for possible publication.

potential targets in non-alcoholic steatohepatitis by investigating the disease perturbations at both the protein and metabolic levels. Intriguingly, here also we found that one of the identified targets is autophagy-related, which underlies the importance of this quintessential process.

3.1 Introduction

Non-alcoholic fatty liver disease (NAFLD) is an umbrella term for a spectrum of liver diseases defined by the aggregation of triglycerides in the liver, without other causes such as medications and excessive alcohol intake, or certain heritable conditions [275]. There is currently no commonly agreed criterion of significant alcohol use at which fatty liver disease can be deemed alcohol-related; nevertheless, threshold amounts of 10-20 g of alcohol per day for women and 20-40 g per day for males have been documented [276–278]. It is strongly associated with metabolic comorbidities, including obesity, hyperlipidemia, type 2 diabetes and metabolic syndrome. Unabated, its prevalence, estimated at 25% [279], is predicted to increase as a result of the significant global surge in factors such as obesity, ageing, T2DM, etc. The initial stage of the disease is the non-alcoholic fatty liver (NAFL), distinguished by steatosis of the liver, encompassing more than 5% of parenchyma, with no signs of hepatocyte damage [280]. Non-alcoholic steatohepatitis (NASH), characterised by steatosis, lobular inflammation, and hepatocellular ballooning, is the second stage of this continuum. If not appropriately treated, it may lead to cirrhosis and hepatocellular carcinoma. It is a slowly progressive disease and often remains clinically discerned, leading to late detections, thereby curbing the therapeutic options and contributing to poor outcomes. Although it is known that the accumulation of lipids is the key to steatosis, the molecular mechanism that governs the transition from steatosis to NASH is yet to be elucidated. Furthermore, despite decades of research, no drug has been approved by the FDA, and liver biopsy remains the gold standard for diagnosing NASH. An overview of NASH is shown in **Figure 3.1**. The multicenter Nonalcoholic Clinical Research Network (CRN) developed a scoring tool for the histological features of NAFLD (the NAFLD activity score, NAS) to measure the morphological changes during therapeutic trials [281]. The three primary characteristics of NAFLD, steatosis, hepatocellular ballooning, and lobular inflamma-

tion, are assessed in this scoring system. The former and the latter score ranges from 0-3, while the middle ranges from 0-2. Based on the assessment of these scores, each patient gets a score. A liver condition is called Not-NASH or NAFL if $1 < NAS < 3$ and NASH if $NAS \geq 5$. However, this scoring system does not consider fibrosis accumulation. Addressing this, another scoring method, called steatosis, activity (hepatocellular ballong+ lobular inflammation), fibrosis (SAF), was developed [282]. In this scoring system, steatosis ranges from 0-3, while activity, as well as fibrosis score, ranges from 0-4.

NASH is a consequence of various metabolic alterations [285] in which excess triglyceride (TG) synthesis is seen through the de novo lipogenesis process [286]. Also, its progression is associated with distinct metabolic network-level changes, most notably, disruption in the mitochondrial metabolism, de-novo lipogenesis, and gluconeogenesis [287]. Besides the metabolic alterations, pathways like inflammation, fibrosis, apoptosis, etc., also contribute to the disease progression,. In other words, the quest to identify plausible potential targets should consider both the protein and metabolic level alterations. In this context, protein-protein interaction (PPI) network analysis can be applied to identify a core set of proteins that are capable of governing the disease system and can identify the potential targets that influence all the aforementioned pathways. PPI network serves as the basis for the signalling circuitry of an organism, which governs cellular response to external and genetic inputs. Understanding this architecture may enhance the prediction of gene function and the cellular response to numerous diseases and disorders.

Irrespective of its size, a PPI network always possesses a small set of core nodes, which can modulate the fate of a biological system. Distinguishing these proteins has proven to be daunting, further exacerbated by the intricacy of understanding how such proteins interact synergistically. In literature, numerous methods exist that have tried to extract the core set of proteins from the network. These methods vary from basic or simplest to more advanced or complex architectures. However, in NASH, most of the works using the PPI network mainly focus on identifying only the hub genes from the differentially expressed genes (DEG) network constructed from various transcriptomic data [288–295]. Karbalaei et al. [296] tried to identify the common proteins between inflammatory bowel disease and NASH by using a systems bi-

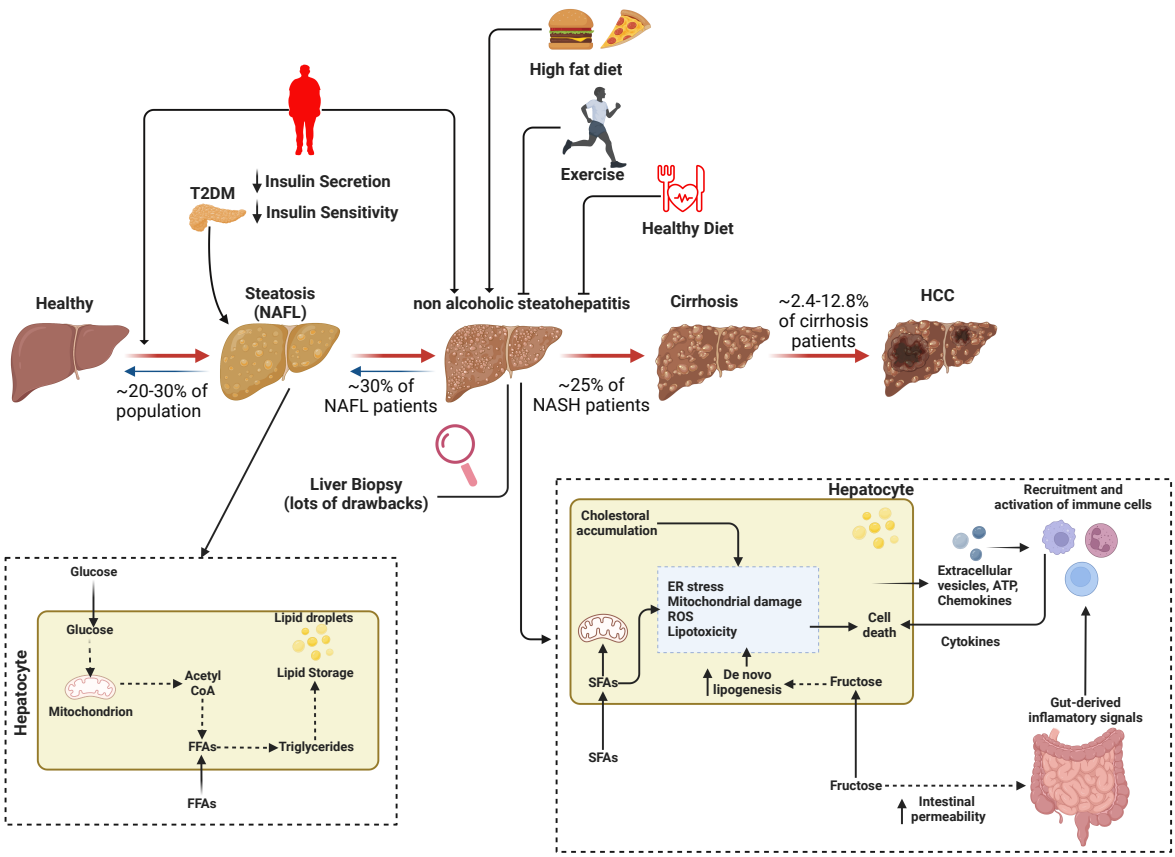


Figure 3.1: **Overview of the current understanding of NAFLD and the factors involved with progression and pathogenesis of NASH.** NAFLD is a progressive disease that commences with simple steatosis (non-alcoholic fatty liver, NAFL) and can advance to non-alcoholic steatohepatitis (NASH), which is characterised by steatosis, inflammation, and fibrosis. NASH can develop into cirrhosis and, in some instances, hepatocellular cancer (HCC). Obesity, T2DM, high-fat diet, etc., play a crucial role in the progression of NAFLD. Currently, no drug is approved for NASH, and physical activity, controlled diet etc., are recommended for possible disease reversibility. Liver biopsy despite having some drawbacks, remains the gold standard for diagnosing NASH at the moment. NAFL is characterised by the buildup of triglycerides in hepatocytes via de novo lipogenesis. This process is driven by the absorption of glucose and free fatty acids (FFAs) and their integration into lipid-synthesis pathways. Development of NASH is also aided by multiple stress like Endoplasmic reticulum (ER) stress, ROS formation, mitochondrial damage, etc. The origins of these stresses are multifaceted. For instance, increased fructose absorption or accumulation of ER cholesterol leads to de novo lipogenesis, which results in the accumulation of saturated fatty acids (SFA). In addition, fructose can activate hepatic immune cells via gut-derived inflammatory mediators. Consequently, it can cause liver inflammation and cell death. This sets off a series of events, such as the release of ATP, chemokines, and extracellular vesicles, which reinforce inflammatory processes and the formation of fibrosis. This figure is modified from [283] and [284].

ology perspective. They first extracted the disease-associated genes from DisGeNET, and the common genes were used to construct a PPI network, where they identified the hubs and bottlenecks and termed them as the key proteins in both diseases. Asadzadeh-Aghdaee et al. [289] constructed a PPI network of NASH by using the 'disease search' plugin of the string database in Cytoscape [121]. Based on three parameters, disease score, hub, and bottleneck, they identified the top 10 proteins in the network. These proteins were further searched for association with crucial biological processes and pathways, and finally, five key proteins were identified. Jiang et al. [291] used a clinical dataset GSE89632 which contained information on 19 NASH, 20 NAFL, and 24 healthy control (HC). Focussing on the NASH and HC, they constructed a PPI network of the DEGs using the STRING database. The hubs in the network were identified, and their immune infiltration analysis was further performed. Ye et al. [292] merged three clinical microarray datasets, GSE48452, GSE63067, and GSE89632 and identified the DEGs. A PPI network of these DEGs was then constructed, and the hub proteins were identified. Feng et al. [290] used multiple clinical datasets, and the common DEGs among them were used to construct a PPI network where they identified the hub genes and carried out their survivability analysis on a hepatocellular carcinoma dataset [297–301]. To explore the potential mechanism of GANLU powder (GLP) in the treatment of NASH, Gao et al. [299] constructed and identified the hub proteins in a PPI network of the targets of GLP and the nash-related targets. The study was further assisted by molecular docking analysis, where they investigated the capability of direct interaction of the hubs with the bioactive compounds of GLP and based on this, they claimed that GLP might treat NASH by regulating AKT1. However, all these studies are quite one-dimensional. Most of these studies are constructed in a very small network; hence, the global overview of the system is missing. Moreover, identifying only the hubs (very few studies have additionally considered the betweenness or closeness) does not necessarily justify the key proteins in the disease system.

Among several methods, a reliable perspective of studying such networks is the implications of control theory, which investigates how to manipulate a dynamical system's behaviour. One of the ultimate objectives of analysing a network is to regulate its behaviour or state. Obtaining the capability to control biological networks' behaviour entails altering the phenotypes of

biological systems as needed, which is crucial for treating diseases or any unwanted cellular abnormalities. In the 1960s, Kalman developed the concepts of controllability that are now the foundations of modern control theory. Each node in the network has its own state variable (for instance, concentrations of proteins at a certain time). Due to the complex interaction between proteins, the state of one node can affect another, which in turn can modify the state of another protein, and so on. Controllability quantifies the ability to manoeuvre a network's state space by controlling a minimal number of nodes. Recent years have witnessed the emergence of control theory as a mathematical framework for exploring complicated dynamic networks. However, no study has been undertaken on network controllability in NASH.

The fundamental disadvantage of only PPI-based studies is that they do not account for changes in metabolic flux level and, as a result, are unable to investigate perturbations in essential metabolic pathways that drive the progression of the disease. However, the efficient analysis of the PPI network can provide novel candidates for metabolic channelling. Again, the core set of proteins, which can be the potential drug targets, should be capable of reversing the disease-associated gene signatures. An established approach to check this endeavour is using the connectivity map database (CMap), a large-scale perturbagens network that contains the transcriptomic profiles of numerous cultivated cell lines treated with various chemical and genetic reagents [302]. It is a platform that helps to get the functional relationships between genes, perturbagens, and diseases. The query tool of CMap takes the list of upregulated and downregulated genes and provides a connectivity score to perturbagens, mainly based on the similarity between the query gene set and reference gene set. This score ranges from -100 to +100. The higher the positive score, the higher the correlation between the query set and the reference set of genes. Similarly, a negative score means that the induction of that particular perturbagen causes an opposite gene expression profile to the query gene set. An overview of CMap is shown in **Figure 3.2**.

Hence, to better understand the feasibility of the targets, their role in the disease system should be adequately investigated, which brings the collaborative effort of the context-specific molecular networks into the scenario. Probing these context-specific networks is probably the only way to make sense of the cellular anomalies during disease progression. However, like

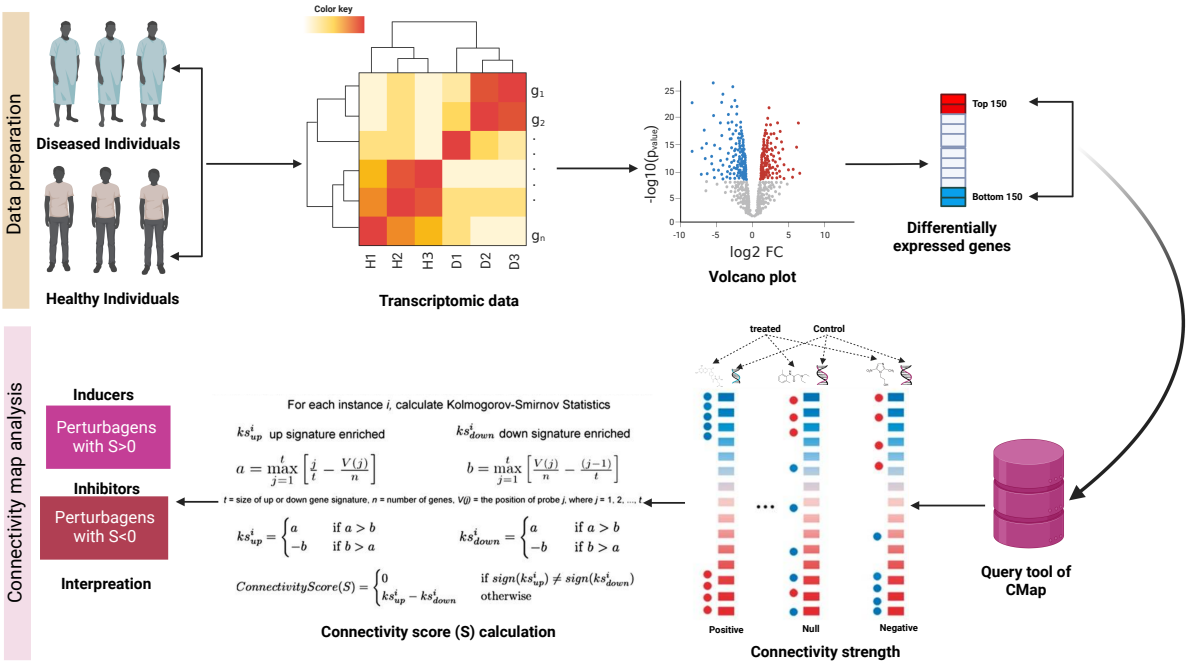


Figure 3.2: **The Mechanistic understanding of CMap.** CMap's reference database comprises gene-expression profiles acquired from treating human cultured cells with a large range of perturbagens. The query tool of CMap takes the top 150 upregulated and downregulated DEGs. Hence, from the transcriptomic data, the DEGs need to be identified. After obtaining the query signature, pattern-matching algorithms provide a score to every reference profile based on the direction and magnitude of enrichment with the query signature. Perturbagens are then ranked by the "connectivity score (S)" calculated using the Kolmogorov-Smirnov statistic. In general, a perturbagen is termed an inducer if $S > 0$ and an inhibitor if $S < 0$.

several other diseases, NASH is yet to come under the grasp of such collaborative studies.

Taking all these as our motivation, we endeavour to study NASH at both the protein and metabolic levels to identify a set of potential therapeutic targets. By taking a clinical dataset on NAFLD, we simultaneously investigate the gene and the metabolic level alterations to find the crucial genes that can initiate reversibility towards control. We further integrated these genes into a directional PPI network of the DEGs and applied control theory analysis to get the most fragile nodes in the interactome. The finally identified proteins are capable of inducing reverse gene expression, metabolic transformation towards control, and also affect network controllability. We then analysed their knockdown gene expression profile to tailor their effect on steatosis, inflammation, fibrosis, and cell death, the four clinical traits of NAFLD.

3.2 Materials and Methods

3.2.1 Data acquisition and pre-processing

To get the patient-specific transcriptomic data, we used the keywords: 'Non-alcoholic' and 'Human' in the Gene Expression Omnibus (GEO) database [239]. We further filtered out the datasets based on the following exclusion criterion, i) The datasets that contained samples of other diseases (such as HCC, HIV) or were infected with viruses (such as HBV, etc.) or were treated with interventions (such as dietary intervention) were eliminated. ii) All in-vitro models were eliminated. iii) Also eliminated are datasets that lack adequate stage categorisation descriptions. iv) From the remaining datasets, we chose only those containing RNA-Seq data. This is owing to the fact that RNA-Seq data distinguishes more differentially regulated transcripts, splice variants, and noncoding transcripts, hence shedding more light on numerous biomedical and biological topics. Finally, we received the datasets GSE130970 and GSE126848. The latter dataset was chosen based on its sample size. The control group of GSE126848 comprises 14 healthy normal-weight subjects with body mass index (BMI) 18.5–25 kg/m^2 and 12 overweight subjects with BMI 30 – 40 kg/m^2 [303]. They recruited 31 NAFLD patients, and the severity of the disease was measured using steatosis, activity, and fibrosis (SAF) score [282], which resulted in 15 NAFL and 16 NASH patients. We used the

'calculateTPM' package of R to normalise (TPM) the data of GSE126848. After normalisation, outliers of the dataset were replaced using the 'filloutliers' function in MATLAB, which detects the outliers and replaces them with the median one. This function defines outliers as elements more than three scaled median absolute deviations (MAD) from the median. We also excluded the genes with missing values in at least one sample.

3.2.2 Identifying DEGs in the clinical dataset

DEGs are expressed in all samples, and their behaviour varies in tandem with the categories. To identify them, a two-sample t-test was performed to calculate the p-values for two different clinical groups using the 'mattest' function in MATLAB. The 'mafdr' function was applied to evaluate the Benjamini, and Hochberg FDR adjusted p-values. Finally, DEGs were calculated using the cutoff values for FDR adjusted p-value ≤ 0.05 and $|\log(\text{fold change})| \geq 1.2$ [304]. The identified genes were further mapped to the UniProt [235], and only the reviewed proteins were selected. All the analyses are carried out with respect to control.

3.2.3 Gene set enrichment analysis of the DEGs

We used the log fold change values of gene expression from each clinical group to perform the gene set enrichment analysis (GSEA) [143]. This analysis was carried out in R using the 'fgsea' package [305]. For geneset/pathway annotation, we used the KEGG [306] and Reactome [307] subcategories from the "Canonical pathways" category of MSigDB (version 7.4.1) database [308].

3.2.4 Identifying possible candidate proteins using CMap

The query tool of CMap takes the list of up-and down-regulated genes and provides a connectivity score (ranges from -100 to +100) to perturbagens, mainly based on the similarity between the query gene set and the reference gene set. The higher the positive score, the higher is the correlation between the query set and the reference set of genes. Similarly, a negative score means that the induction of that particular perturbagen causes an opposite gene expression pro-

file to the query gene set. We hypothesized that the perturbagens with a negative connectivity score might possess a potential role in abrupting the propagation of disease. For this purpose, we selected the perturbagens with a connectivity score $\leq$ -90. To generate a list of potential candidate proteins (C_p), we considered only the gene knockdown among the three forms of perturbagens available in the CMap database.

3.2.5 Construction of the directed PPI network (DPN)

A crucial attribute of a PPI network is directionality [101, 102]. In a PPI network, it refers to the functional relationships between the proteins, which captures the regulatory effect exerted by the source protein on the target protein. The identification of key proteins in an undirected network may lead to various false-positive results [309]. For instance, when the mode of interactions for drug-disease relationships is absent, we cannot determine if a drug heals a disease or produces one as a side effect [97]. We have extracted the human interactome information from the STRING [215] and the SIGNOR 2.0 [216]. We selected only the functional (inhibition and activation) relationships from the former and filter out the interactions where the directionality is applicable. We also extract the directional network provided by Arunachalam et al. [310].

3.2.6 Construction of the DPN for each category

For each category, the nodes of the network consisted of three groups. The first and second groups consisted of the category-specific DEGs and C_p, respectively. The third group consists of the metabolic genes having positive TS scores (metabolic candidate genes, MCGs) obtained from the *in silico* single gene knockdown analysis. These nodes are then projected to the DPN to construct a directional network. However, each of these networks consisted of many disjoint components. To connect them, we took the help of the DPN and found a minimum number of nodes to connect the components for each category. These nodes are termed mediators. However, some components were still disconnected and were then removed from the study.

3.2.7 Identification of the influential nodes in the network

Irrespective of its size, the fate of a network always hinges on a small set of nodes. To identify these nodes in our study, we have opted for network controllability analysis. For details please refer to Chapter 1, Section 1.2.3.

3.2.8 Gene knockdown profile extraction using the CMap database

We extracted the LINCS L1000 dataset from the CMap database. The LINCS L1000 dataset consists of five separate levels of dataset ranging from original, processed, to z-score data. The level 1 dataset contains raw fluorescent intensity values from Luminex 1000 platform, the level 2 dataset provides the gene expression values for the 978 landmark genes obtained after deconvolution. The level 3 dataset includes the normalised gene expression values of the landmark genes and 11,350 additional genes which are estimated using the normalized expression values of the landmark genes. The level 4 dataset contains Z-scores for each gene based on the Level 3 data with respect to the entire plate population. In general, the L1000 experiments are carried out in 3 or more biological replicates. The level 5 data of the LINCS database contains the consensus replicate signature by applying the moderated z-score (MODZ) procedure. Here, we have opted for the level 5 dataset in our study as it is more robust (39), and biological discovery is more likely to be achieved using this level of data. It contains the consensus replicate signature by applying the MODZ procedure. We selected the HEPG2 cell line, the consensus gene signature of treated genes, and their untreated control vectors and obtained 3341 treated and corresponding 533 control conditions for 12328 genes. The Z difference score for each gene is then measured using the equation

$$Z^{ij}_{diff} = Z^{i}_{treated} - Z^{j}_{control}, i = 1:3341, j = 1:533. \tag{3.1}$$

Here, Z^{ij}_{diff} is the Z difference score for treated condition i and control condition j.

$Z^{i}_{treated}$ is the Z score of a gene corresponding to treated condition i.

$Z^{j}_{control}$ is the Z score of a gene corresponding to control condition j.

We then identified the up-and downregulated genes for which $Z^{ij}_{diff} \geq 1.5$ or $Z^{ij}_{diff} \leq -1.5$

satisfies at least 60% of the control conditions, respectively.

3.3 Results

3.3.1 Molecular alterations associated with NAFLD

The selected dataset, GSE126848, included RNA-Seq liver biopsy data of the 15 NAFL, 16 NASH, and 26 control individuals. The dataset is first Transcripts Per Million (TPM) normalised. Differential gene expression analysis on this dataset revealed 5468 and 5672 DEGs in the NAFL and NASH categories, respectively (**Figure 3.3A**). Again, being two consecutive stages of the NAFLD spectrum, we found an overlap between the DEGs of these two categories (**Figure 3.3B**). A significant number of metabolic genes were also found to be perturbed in these two stages (**Figure 3.3C**).

3.3.2 Functional enrichment of the differentially expressed genes

To find the up-and downregulated pathways in both the categories, we performed GSEA on the DEGs (**Figure 3.3D**). We found that the collagen-associated pathways, autophagy, apoptosis pathways, oxidative phosphorylation, DNA damage, and hypoxia-related pathways were upregulated in both groups. Collagen deposition, increased cholesterol biosynthesis, and apoptosis of liver cells are linchpins of the NAFLD progression. The upregulation of these indicates the deterioration of the liver in individuals. Hypoxia enhances cellular lipid deposition and upregulates genes involved in lipogenesis, lipid uptake, and lipid droplet formation, according to several *in vitro* and *in vivo* studies [311, 312]. Similarly, some DNA damage checkpoint proteins are found to promote apoptosis and fibrosis in NAFLD [313]. The upregulation of the base excision repair pathway indicates elevated DNA repair activity, a feat often found to correlate with fatty liver [312]. The increase in inflammation can be observed by the upregulation of the interleukin-1 signalling and oxidative phosphorylation pathway [314, 315]. However, quite uncharacteristically, we found various inflammation-related pathways among the downregulated pathways. However, this may be due to the reason that none of the NAFL and NASH individuals was seen to have a high inflammation score [303]. Other than that, we have seen the

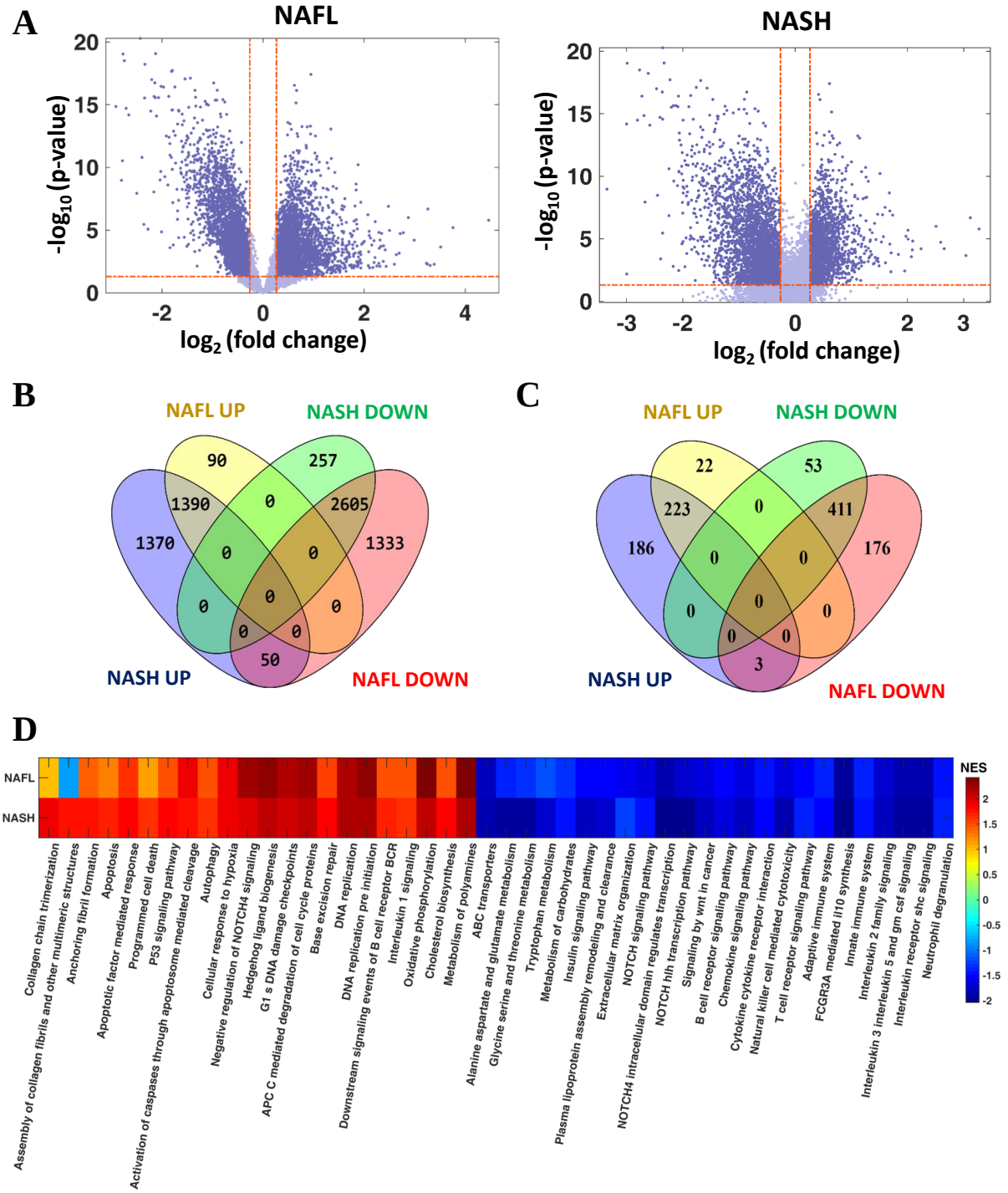


Figure 3.3: **The transcriptome landscape of NAFLD.** A) Differential expression analysis of genes in the NAFL and NASH group compared to the control. The blue colour in the volcano plot represents the genes that are differentially expressed. B) Venn diagram representation of the differentially expressed genes (DEGs) in the NAFL and NASH group. From the figure, it is evident that there exists an overlap between the DEGs of each group. C) Venn diagram representation of the altered metabolic genes. D) The normalised enrichment scores (NES) obtained from the gene set enrichment analysis (GSEA).

downregulation of carbohydrate metabolism and insulin signalling, both correlate with disease progression in NAFLD. On the other hand, metabolic flux level analysis by integrating the gene expression data on *iHepatocytes2322* [316] followed by flux variability analysis (FVA) also revealed the alterations in various crucial metabolic reactions [317]. These include carbohydrate metabolism, fatty acid oxidation, fatty acid activation, desaturation, and beta-oxidation etc.

3.3.3 Candidate genes capable of metabolic transformation or reverse gene expression

The disease-associated alterations were further used to determine potential recovery options that can revert the system back towards a healthy state. At the metabolic stage, we have systematically carried out a 90% gene knockdown to sort the metabolic genes capable of reversing the disease flux state towards the control and identified 93 genes in the NASH and 114 genes in NAFL category [317]. These genes, hereafter referred to as metabolic candidate genes (MCG), are therefore selected for further evaluation.

At the gene level, we have used the Connectivity Map database (CMap) [302] to obtain the set of genes whose knockdown may result in a reverse gene expression profile of the DEGs. CMap is a large-scale drug perturbation network piloted by the transcriptomic profiles of varieties of cultivated cell lines treated with various chemical compounds. Here, for the HEPG2 cell line, we considered the genes whose knockdown results in a connectivity score of -90 or less when queried with our topmost (top 150 up-and downregulated) DEGs. These genes might be crucial in the NAFLD landscape because their knockdown demonstrates reversal effects on gene expression in NAFLD. These are termed candidate proteins (C_p), and their numbers in each category are shown in **Figure 3.4A**. As can be observed, C_p are largely distinct across the two groups. The ten common C_p, were found to take part in endothelial cell migration, protein modification, fatty acid biosynthesis, etc. Overall, the C_p for both categories is enriched with cell death-related processes. While the NAFL group was enriched with autophagy-related processes, the NASH group was found to be enriched in cytokine-mediated signalling pathways (**Figure 3.4B-C**). The variability observed in molecular alterations at the gene, and metabolic level reflects that organising principles are fundamental to a given biological scale. However,

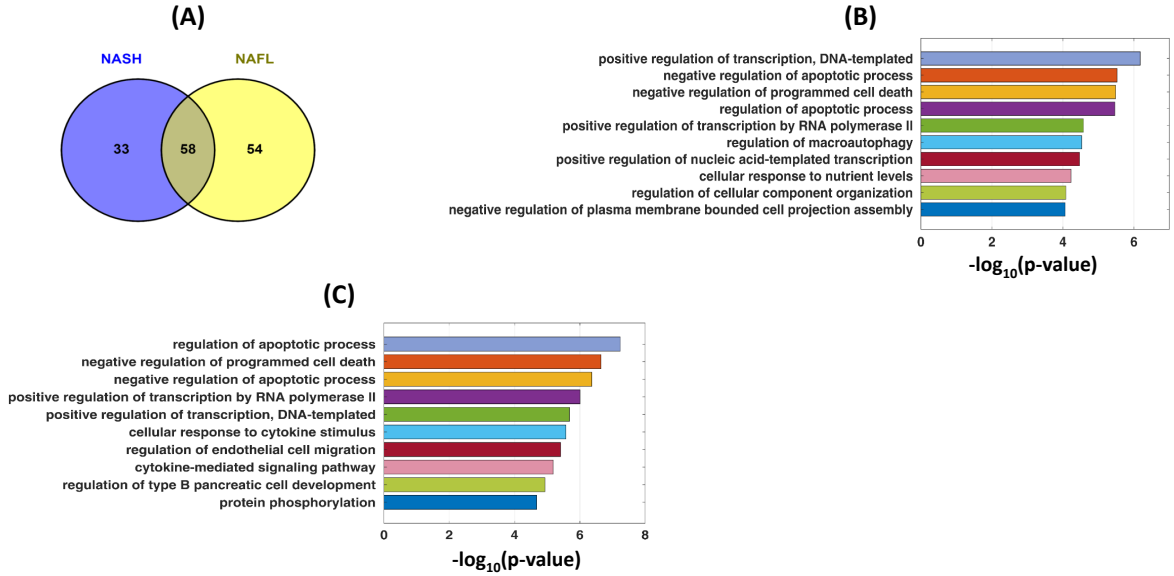


Figure 3.4: **Connectivity map (CMap) analysis**. A) Venn diagram representations of the common proteins in the connectivity map analysis. B-C) Top ten biological processes enriched with the candidate genes for NAFL and NASH group, respectively.

the investigations at these individual levels were insufficient to shed light on the delinquent proteins that clandestinely influence the crucial biological pathways and processes. Thus, such studies fail to capture the global influence of any perturbed protein to get the actual efficacy of the proposed target. To overcome this, we extend our individual-level analysis to a collaborative study to identify the proteins capable of disease regulation at both the metabolic and protein levels.

3.3.4 Construction and analysis of a PPI network bridging metabolic with genomic level identified crucial proteins

The intricate interactions of proteins regulate the changes in both the transcriptomic and metabolic networks. Understanding these interactions enables decoding the complicated association between proteins and disease-related abnormalities. To fully realise this potential, however, the interactions between the proteins must be directed as the orientation of a PPI network gives more insight into the signalling pathways, disease progression, drug development, and treatment combinations, to name but a few important applications [318].

The constructed directional PPI network (DPN) contained 8673 nodes and 60546 edges. The

average degree of the network was 13.9620. the highest degree, indegree, and outdegree of the network were 506, 250, and 377, respectively. This network is shown in **Figure 3.5**.

In the next step, we collected the DEGs, C_P, and the MCGs for both the NAFL and NASH groups and integrated them into the DPN. Not surprisingly, each consisted of one giant component and many small disconnected components. On the other hand, being a part of the network is futile unless it is linked to the giant component. This is due to the fact that a disconnected component would be unable to participate in the system's information dissemination. We, therefore, identified the minimum number of nodes required to connect the components and finally obtained a connected directional network for NAFL and NASH (**Figures 3.6A-B**) with different sizes and orders (**Figures 3.6C-D**). Both of them obey the power law of the degree distribution, indicating the scale-free nature of the network. The network centrality properties like betweenness, closeness, clustering coefficient, indegree, outdegree, and the total degree of both networks are calculated. The degree centrality ranking of the proteins for the two networks was found to be almost identical (Pearson correlation coefficient: 0.9923), and the top two proteins for both networks were TP53 and 14-3-3 protein zeta/delta. The histogram plot of the degree-rank differences of the proteins has revealed that only 19 proteins had a larger rank in the NASH network as compared to the NAFL network, while the rank of 14 proteins remained the same (**Figure 3.6E**). With a few exceptions, the betweenness rank follows the same trend as degree ranks (Pearson correlation coefficient: 0.9266). As with degree centrality, the top two proteins in this category were also TP53 and 14-3-3 protein zeta/delta. The rank correlations for clustering and closeness centrality within these two networks were 0.8440 and 0.9043, respectively.

We then used structural controllability theory [100] to identify the minimum number of driver nodes capable of controlling the whole network. The nodes were then graded into indispensable (I), dispensable, and neutral categories based on the number of driver nodes. We found that 45% and 44% of nodes are drivers and 18.67% and 20.59% are indispensable, respectively, in the NAFL and NASH categories (**Figure 3.6F**). The degree distribution plots of these categories show that the indispensable nodes and the neutral nodes tend to be the hubs while the dispensable nodes are low-degree nodes in the network (**Figure3.7**). We use the equation

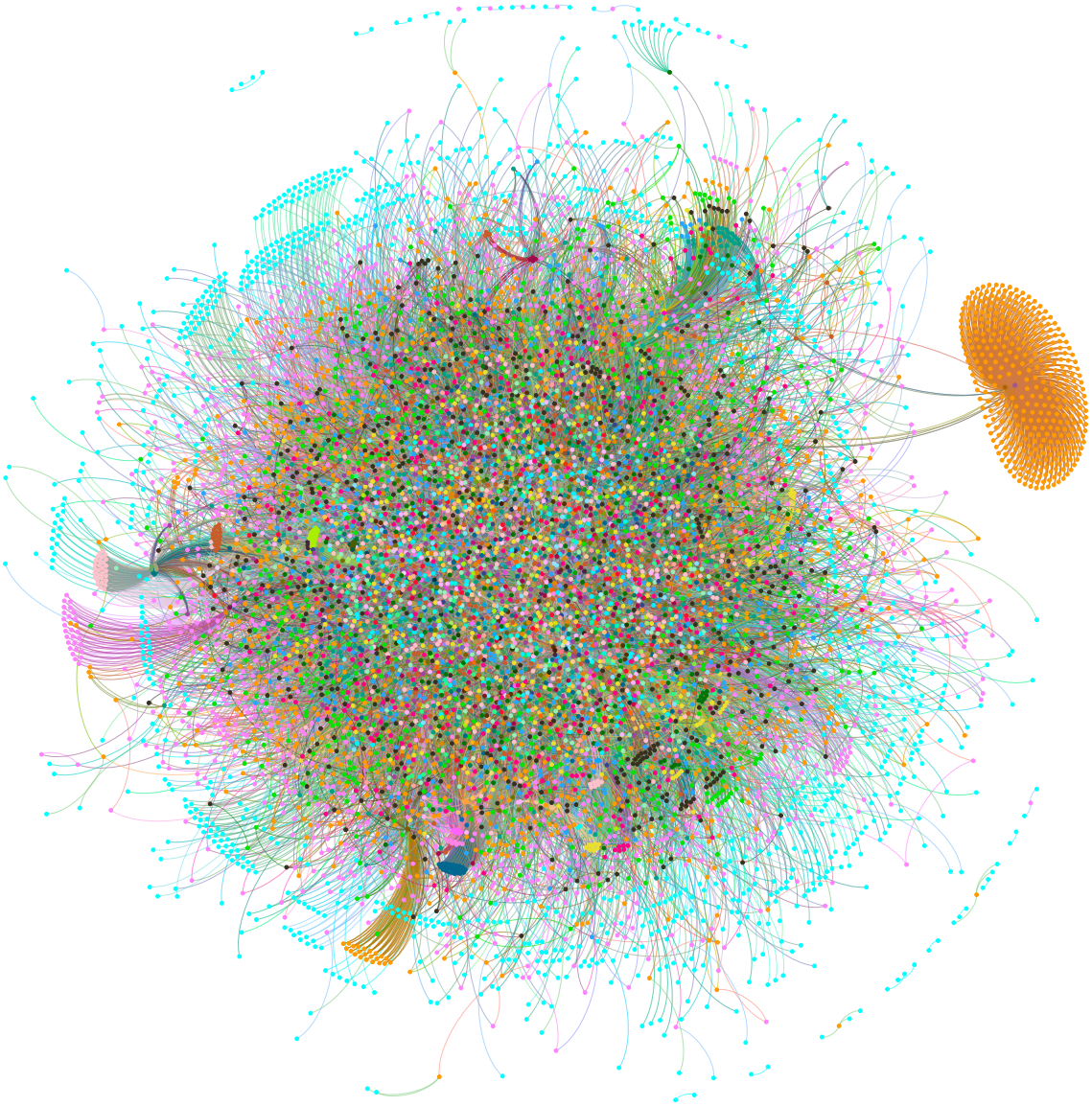


Figure 3.5: **The constructed directional network.** It is constructed using STRING (version 11.0) [215] , SIGNOR 2.0 [216], and the directional network provided by Arunachalam et al. [310]. We selected only the functional (inhibition and activation) relationships from the former and filtered out the interactions where the directionality is applicable.

$IC_P = I \bigcap C_P$ to find the set of indispensable candidate proteins (IC_p) in the network, where I = set of indispensable proteins and C_p = set of candidate proteins. IC_p contains 15 and 29 proteins in the NAFL and NASH categories, respectively (**Figure 3.6G**). These are the most crucial proteins in the network because they not only generate reverse gene expression profiles but also influence network controllability [101]. By investigating the involvement of these proteins in NASH, we discovered that they are critical in the development of the disease. For example, the IC_p obtained in the NASH category are enriched in the insulin resistance pathway (PPP1CC, PRKAA2, AKT1, and FOXO1), in the apoptosis pathway (CASP3, AKT1, CYCS, and RIPK1), glucagon signalling pathway (PRKAA2, AKT1, and FOXO1) and Toll-like receptor signalling pathway (AKT1, IL12A, and RIPK1). Insulin resistance is a necessary condition for the development of NASH [319]; apoptosis leads to death of the hepatocellular cells [320]; the glucagon pathway, on the other hand, regulates lipid metabolism [321].

3.3.5 Potential therapeutic targets in NAFLD

Diseases are driven by the perturbations in gene expressions, which tailor a cascade of events that has a direct influence on the proteins and metabolic fluxes. Although IC_p contains more significant proteins than the others in the network, an aura of eeriness surrounds their prospective roles as targets in their respective categories. This can be mitigated by performing an *in silico* knockdown and quantifying this effect in terms of the transformation from the disease state to the healthy one. For this, we extracted the gene knockdown expression data of the HEPG2 cell line from the CMap database and revealed the up-and down-regulated genes following the knockdown of each C_p using a Z-score difference cutoff of 1.5. These profiles were then used to capture the metabolic flux level transition from the disease state to the target state. Next, using the metabolic transformation algorithm, we identified the IC_ps with positive transformation score (TS) [317] and were deemed as potential targets because they can induce reverse gene expression to the DEGs, affect the controllability of the network, and revert the disease flux state towards control.

The three obtained targets for NASH were BAG6, CYCS, and CASP3. We portrayed the significance of these targets in the NASH landscape through their interactors in the NASH network.

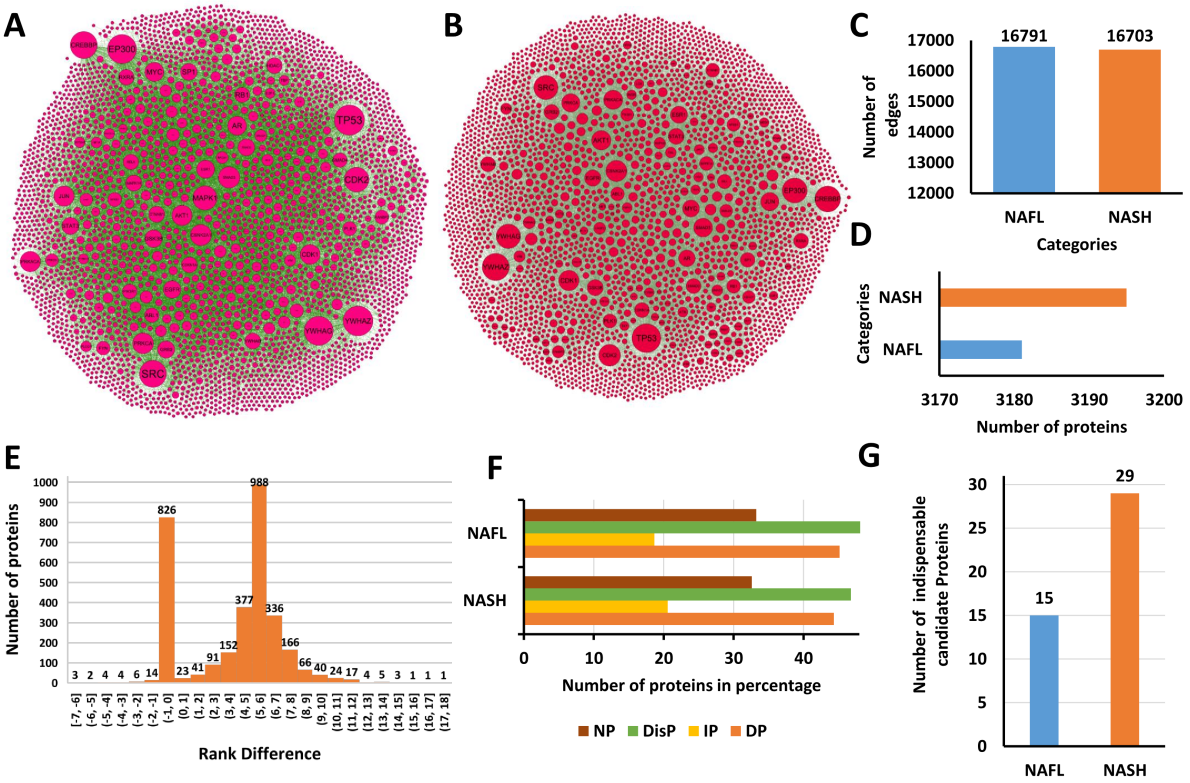


Figure 3.6: **Protein-protein interaction network analysis.** A-B) The PPI network of the NAFL and NASH groups, respectively. Here the greater the node size, the larger its degree. C-D) The size and order of the two networks. The NAFL network consists of 3181 nodes and 16791 interactions, while the NASH network consists of 3195 nodes and 16703 edges. E) The histogram plot shows the degree rank difference between the common proteins in the NAFL and NASH network. F) The abundance of driver nodes (DN), indispensable (IP), dispensable (DisP), and neutral (NP) nodes in the two networks. G) Indispensable candidate proteins (IC_p) in the NAFL and NASH group.

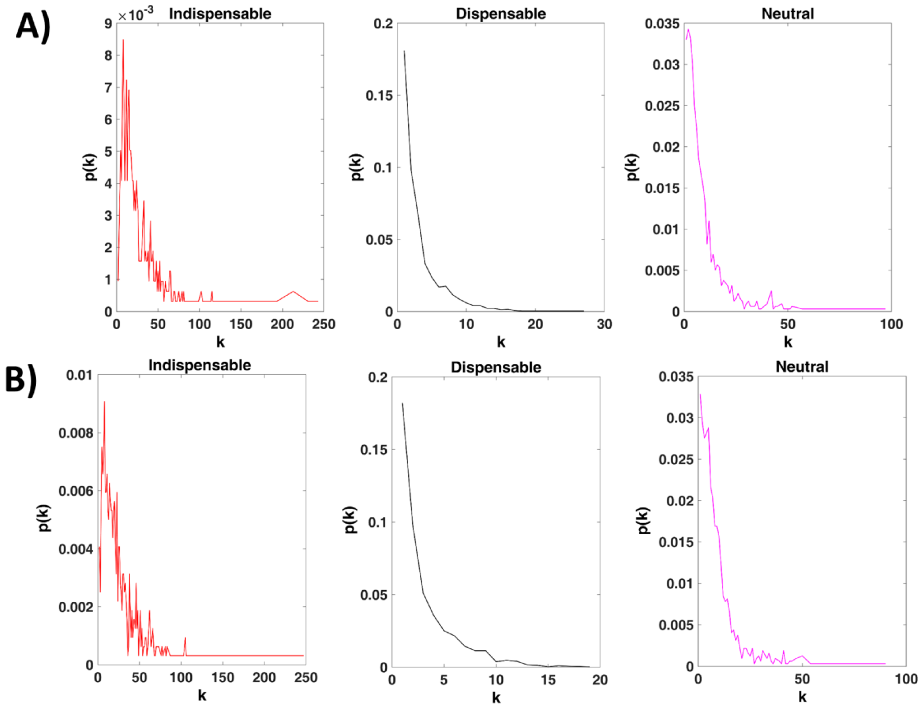


Figure 3.7: **Degree distribution plots of the indispensable, dispensable and neutral nodes.** A) NAFL network. B) NASH network. It is seen that in both networks, the indispensable nodes are high-degree nodes.

Among these three proteins, CASP3 has the maximum number of interactors, followed by BAG6 and CYCS (**Figure 3.8A**). The PPI networks of these proteins are provided in **Figure 3.8C-E**. Topological analysis of these targets in the NASH network has revealed that CASP3 was the more topologically enriched protein with a relatively higher rank than the other two in all the previously mentioned centrality measures. Its interactors include genes that play a critical role in the pathogenesis of NASH. For instance, CFLAR inhibits JNK phosphorylation, thereby ameliorating the clinical characteristics of NASH [322]. Another crucial interactor is TNFR1, whose inhibition is reported as a promising perspective for NAFLD treatment [323]. Similarly, TRAF6 in the BAG6 and TP53 in the CYCS interactome can be named among many crucial genes in the NAFLD landscape [324, 325]. Although CYCS was found to have a low rank in all the centralities, TP53, the node with the highest degree and betweenness centrality in the network, is an immediate neighbour of CYCS. This suggests that through TP53, CYCS can exert its influence on the other proteins in the network. The results gathered from various clinical studies have shown that NAFLD activation affects the immune system, leading to immune infiltration. NASH has a chronic inflammatory phenotype and has been demonstrated to be related to numerous immune cells. Interestingly, we found that 26.25%, 15.38%, and 11.76% interactors of CASP3, BAG6, and CYCS, respectively, are immune-related genes. This suggests that the identified targets are also associated with immune cell infiltration.

Coreness is a methodical approach to determining a protein's local and global significance. It indicates whether the protein is associated with a densely connected region of the network or with its periphery. Additionally, it demonstrates how influential a node is in disseminating information throughout a network. Its biological importance has been established in various investigations [326, 327]. We have used the k-core decomposition algorithm [12] to measure the coreness of a node in the network. It subdivides the network into multiple layers where the outside layers indicate the network's periphery, while the inner layers with larger k values reflect the network's densely connected core. The k-core analysis of the NASH PPI network revealed that it consists of 8 cores, and the three targets of NASH, CASP3, BAG6, and CYCS, are placed in 7^{th}, 4^{th}, and 6^{th} core respectively. The higher coreness indicates that although the possible target proteins are not hubs in the NASH PPI network, they are capable of spreading

the transcription signal to other associated proteins.

The first neighbourhood analysis of these targets in the NASH network showed that they do not share a common first neighbour. It means that there does not exist one single drug or compound whose implement can affect all these three targets simultaneously. However, at an individual level, BAG6 and CASP3 are connected by ATN1 and CDKN1A; and CASP3 and CYCS are connected by BID and CASP9 (**Figure 3.8F**). Nevertheless, all three targets have less connectivity (i.e., total degree, **Table 3.1**, particularly BAG6 and CYCS in the NASH network, implying that targeting them will have very little off-target effects [328]. The biological processes associated with the proteins of each network are shown in **Figure 3.8F-H**. The topological properties of all the potential targets are provided in **Table 3.1**.

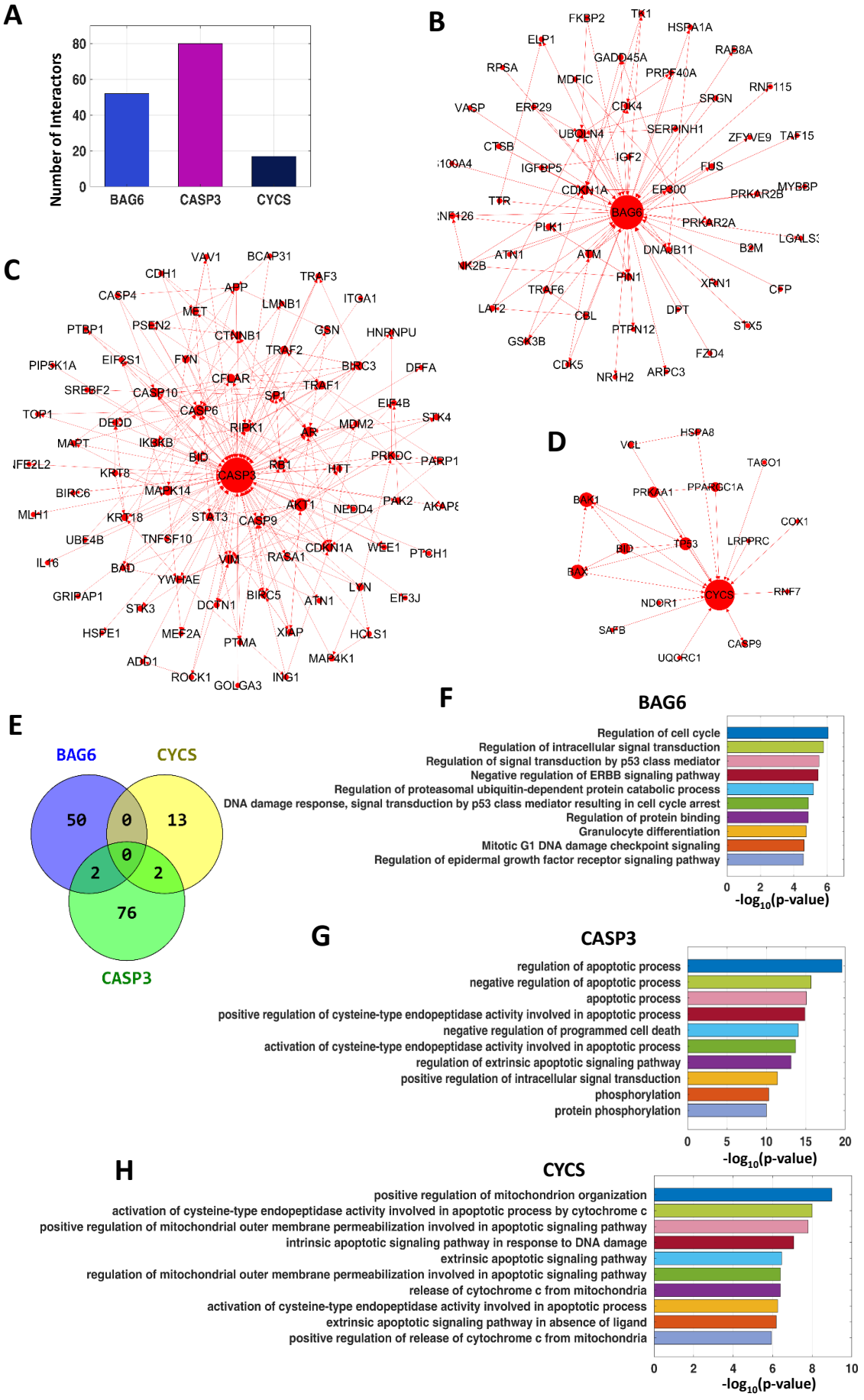


Figure 3.8: **The potential targets in NAFL and NASH.** A) The targets in the NASH category and their number of interactors in the network. B-D) The interactors of the three targets of NASH in the network. E) Venn diagram of the interactors of the targets. F-H) The top 10 enriched biological processes of BAG6, CYCS, and CASP3, respectively.

Table 3.1: **The potential targets of NASH and NAFL and their topological characteristics in the respective networks.** The information about subcellular localisation is taken from the UniProt database.

Category	Proteins	Indegree	Outdegree	Betweenness	Closeness	Clustering Coefficient	Subcellular localization	Total degree
NASH	BAG6	17	34	5.725298	0.318627	0.012549	Cytosol, nucleus, extracellular exosome	51
	CYCS	14	3	2.086636	0.215301	0.045833	Mitochondrion intermembrane space	17
	CASP3	39	48	10.7608	0.311951	0.027102	Cytoplasm	87
NAFL	ACVR1	12	20	4.478068	0.299641	0.039683	Membrane	32
	BMP4	6	5	0.538888	0.215454	0.013889	Extracellular matrix	11
	CASP3	35	45	9.703389	0.31803	0.025235	Cytoplasm	80
	FGFR3	3	13	1.53019	0.287682	0.38462	Endoplasmic reticulum, cell membrane cytoplasmic vesicle, Secreted	16
	GABPB1	6	5	0.981823	0.217806	0.066667	Nucleus	11
	PLAU	8	4	0.854707	0.226706	0.106061	Secreted	12
	PPP1CC	11	14	1.669648	0.300754	0.139493	Mitochondrion, microtubule organizing center, nucleus, nucleolus, nucleoplasm, nucleus speckle, cytoplasm, kinetochore, cleavage furrow	25
	VDAC1	6	4	0.186174	0.243246	0.041667	Cell membrane, multi-pass membrane protein, mitochondrion outer membrane, membrane raft	10

3.3.6 Capturing the effects of the potential targets on the disease-associated perturbed pathways

Often, silencing a gene causes havoc in the genome, resulting in perturbations in the expression levels of specific genes. This is because these proteins share several similar pathways, processes, and functions. Thus, an interpretation of the pathways and ontological properties of these genes aids in the comprehension of the mechanistic understanding of the knocked-down gene. To accomplish this, we retrieved such profiles following the knockdown of each potential target and evaluated their effect on NASH (**Figure 3.9A**). The numbers of affected genes following the knockdown of the potential targets for both categories are shown in **Figure 3.9B-C**. The knockdown of CASP3 and PLAU affected the most number of genes in the NASH and NAFL categories, respectively. The knockdown profiles of these targets revealed that silencing CYCS and CAPS3 improves inflammation, fibrosis, and apoptotic pathways, whereas silencing BAG6 ameliorates the latter two (**Table 3.2**). Further, using GSMM, we evaluated their knockdown effect on the metabolic landscape and found that our proposed targets can revert 66% of the altered reactions found in fatty acid oxidation pathways [317]. The knockdown effect of the NAFL targets is provided in **Table 3.3**. Here also, we found that the knockdown of the NAFL targets can increase the flux rates of the reactions involved in mitochondrial fatty acid beta-oxidation, echoing their significance in the metabolic level.

3.4 Discussion

Diseases are nothing but aberrations to the normal interplay between proteins. These aberrations rise due to their perturbations following the progression of a disease. To gain a comprehensive grasp on the underlying mechanism of the progression of a disease and thereby identify the potential targets, it is an unmet need to capture and analyse these perturbations.

In this chapter, we have developed a methodology focused on capturing these perturbations in both the gene and metabolic levels. The former aids in the identification of genes whose expression significantly changes following disease progression, while the latter helps identify altered

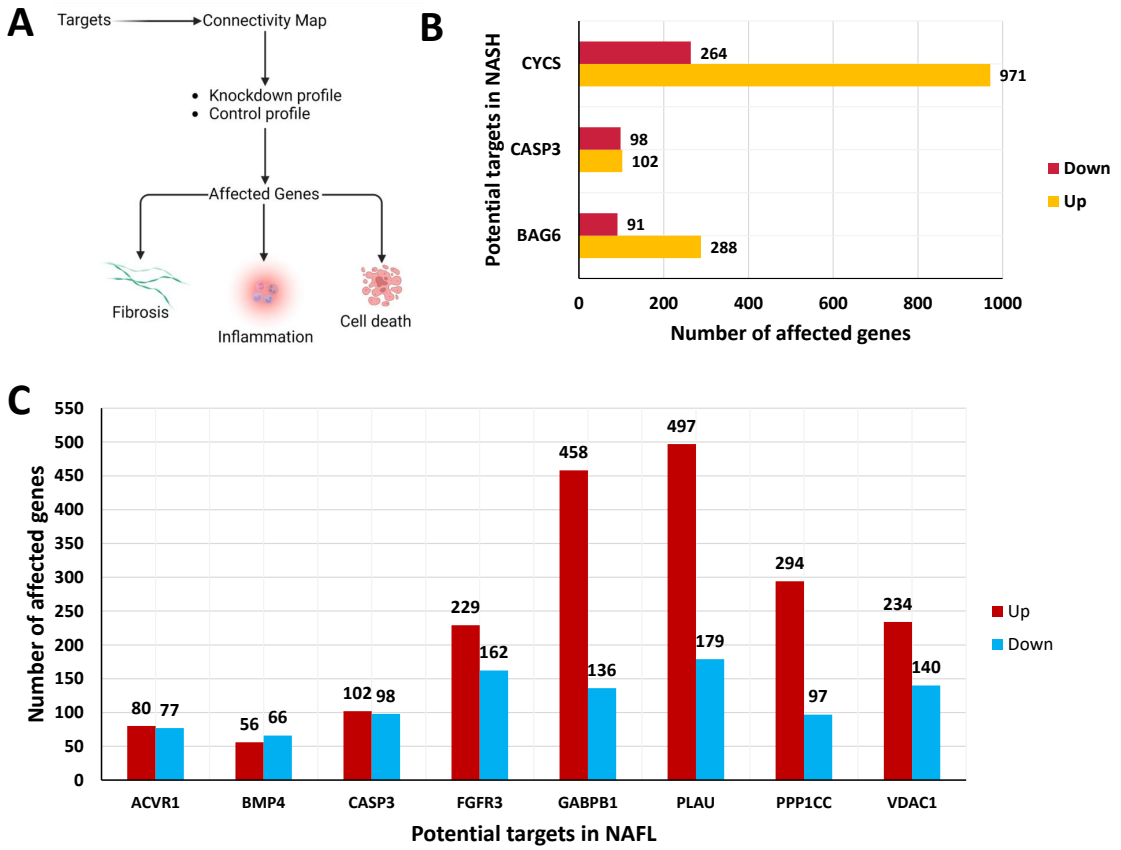


Figure 3.9: **Mechanistic understanding of the potential targets.**A) The schematic diagram for exploring the potential targets. The process starts with extracting the gene knockdown profile of each target from the CMap database and ends with identifying changes in the disease-specific traits such as hepatic cell death, inflammation, and fibrosis. B-C) The bar plot of the numbers of affected genes following the knockdown of the potential targets in the NAFL and NASH groups, respectively. The total number of affected genes is shown at the top of each bar.

Table 3.2: **The potential targets in NASH and their therapeutic effect**. At the metabolic level, their knockdown was found to revert 66% of the altered reactions found in fatty acid oxidation pathways.

Targets	Fibrosis	Inflammation	Cell death
BAG6	Downregulates FN1 and DAG1, and upregulates MMP1.	Not found	Downregulates TNFRSF21.
CASP3	Downregulates FN1 and upregulates WIF1.	Downregulates S100A4.	Inhibiting CASP3 protects against hepatocellular damage and cell death [329].
CYCS	Downregulates Galectin1.	Decreases expression of cytokine IL32.	CYCS release from mitochondria to the cytosol is a key process in the initiation of apoptosis [330].

metabolites and reactions. The methodology is general and can be applied to any disease. We have applied this methodology to NASH to identify potential targets in this multifactorial disease.

NASH is a progressive multifactorial disease, which initiates from a benign NAFL and may move to severe cirrhosis and liver failure [275]. It is currently the leading cause of liver transplants worldwide, and unabated, the numbers will continue to climb. However, despite decades of research, the molecular mechanism driving NASH has not been unravelled, making the disease exceedingly challenging to treat. Consequently, seeking potential therapeutic targets for NASH has become a top priority.

Proteins govern the biological processes, pathways, and molecular functions inside the cell. They work in conjunction with other proteins to accomplish various cellular processes. The causal PPI networks hold an upper hand against the physical interactions as these networks provide the additional dimension of directionality and thus help gain knowledge about the information flow in the network. On the other hand, metabolic networks can explore variations in the metabolism that emerge across the whole histological range of disease, examining the changes that occur as benign to severe stages progress. Thus, there is a high demand for identifying an ideal target, which not only regulates the disease interactome and generates a reverse

Table 3.3: **The potential targets in NAFL and their mechanistic understanding.** In the metabolic level, we found that the knockdown of the NAFL targets can increase the flux rates of the reactions involved in mitochondrial fatty acid beta-oxidation.

Targets	Inflammation	Cell death	Fibrosis
ACVR1	Downregulates SPON2 and C3.	Downregulates C3.	Not found
BMP4	Upregulates transcription factor FOXO4 and downregulates C3.	Upregulates transcription factor FOXO4 and downregulates C3.	Not found
CASP3	Downregulates S100A4.	CASP3 inhibition protects against hepatocellular damage and cell death [329].	Downregulates FN1 and upregulates WIF1.
FGFR3	Downregulates SPON2 and VEGFA, and upregulates HMOX2.	Upregulates HMOX2.	Downregulates HIF1A and VEGFA.
GABPB1	Downregulates C3.	Downregulates C3.	Downregulates FN1, PLOD3, COL1A1, and LAMB1.
PLAU	Not found	Not found	Not found
PPP1CC	Not found	Not found	Downregulates galectin1 and CCN2.
VDAC1	Downregulates PREP.	Downregulates keratin 8.	Upregulates COX-2 and downregulates NME1.

gene expression profile but is also capable of exerting a significant influence on the altered metabolic pathways under diseased conditions.

In this context, the collaborative study of metabolic and PPI networks opens a door to capture the effects of disease progression in multiple layers and offers more realistic solutions to understand the disease mechanism. Here, we have developed a pipeline that can be used to identify targets in any *in silico* studies through a better understanding of disease mechanisms. The pipeline finds the significant elements of the genomic and metabolic level alterations and, based on the former, identifies the candidate proteins (C_P), which can initiate a reverse gene expression profile. To demonstrate our hypothesis that these proteins may contain potential targets, we constructed a PPI network with nodes consisting of DEGs, C_P, and metabolic genes with a positive transformation score (TS). The control theory algorithm identified the driver nodes from the network that eventually resulted in indispensable nodes, which are the most fragile nodes in the network [102]. These nodes are prone to mutations and are often targeted by viruses and drugs [101, 102]. We identified the indispensable candidate proteins (IC_P), the common nodes between the C_P and the indispensable proteins (I), and checked their knockdown effects on the disease-specific GSMMs. The calculated TS values were then used to get a quantitative marker of their effectiveness. Finally, IC_P with positive TS scores were deemed as potential targets because they can induce reverse gene expression to the DEGs, affect the controllability of the network, and revert the disease flux state towards the control.

Like many other diseases, NASH is also driven by certain characteristics such as deranged lipid accumulation (steatosis), hepatic cell death, inflammation, and fibrosis [331]. The potential targets of NASH, regardless of the method by which they are found, should be capable of positively affecting these traits. We have identified three potential targets for NASH, CASP3, BAG6, and CYCS. CASP3 is a member of the caspases family, which are critical mediators of the inflammatory response and apoptosis, and contribute significantly to cellular and organismal homeostasis [332]. Genetically modified mice with loss of CASP3 activity are found to be resistant to diet-induced NASH [329]. The analysis of the knockdown gene expression profile of CASP3 revealed that its inhibition could exert a four-dimensional effect on NASH by modulating all the aforementioned traits. First, its knockdown effect on hepatic steatosis can

be seen from the increased fluxes for the altered (down-regulated) reactions involved in fatty acid activation and mitochondrial beta-oxidation pathways. Secondly, its knockdown downregulates a fibrogenic gene, FN1, which plays a crucial role during liver fibrosis [333]. Also, by upregulating WIF1, which is an inhibitor of the Wnt/β-catenin signaling pathway, a therapeutic target for treating liver fibrosis [334], the CASP3 knockdown can again regulate fibrosis development. Thirdly, its knockdown can regulate the inflammatory process by downregulating the inflammation inducer S100A4. The role of CASP3 as a hepatic cell death inducer, on the other hand, is very well explored [329]. These facts hence support that CASP3 could be an attractive therapeutic target for NASH treatment.

CYCS is a small soluble heme protein found abundantly in the inner mitochondrial membrane. The release of CYCS from mitochondria to the cytosol is a critical process in initiating the intrinsic and extrinsic pathways of apoptosis [330]. Also, its translocation into the extracellular space induces inflammation [335]. Its role in NASH is, however, yet to be properly explored. We found that, like CASP3 and BAG6, CYCS knockdown can affect the crucial reactions in the lipid accumulation process. Its silencing downregulates IL32, a NAFLD-related hepatic cytokine, modulating the hepatic inflammation process. Tailoring CYCS to hepatic fibrosis, we found that its knockdown downregulated Galectin 1, which can ameliorate fibrosis by inducing apoptosis to HSCs [336]. Thus, like CASP3, CYCS knockdown also possesses a four-dimensional effect on the development and progression of NASH.

BAG6 (BAT3/Scythe) is a ubiquitin-like protein involved in a myriad of non-related physiological and pathological processes, including apoptosis, antigen presentation, immunological pathways, and T-cell response. It is cleaved in the cytosol by CASP3 in response to intrinsic or extrinsic apoptosis, yielding a C-terminal fragment of BAG6 that induces apoptosis [337]. Again, the exosomal BAG6 activates the NK cells while the soluble BAG6 inhibits the NK cell cytotoxicity [337]. However, its role in NASH is yet to be explored. We linked its knockdown effects to the traits mentioned above to investigate BAG6 as a potential candidate for NASH recovery. Increased extracellular matrix deposition, such as collagen types I and III, plays a role in hepatic fibrosis [338]. Collagens are primarily synthesised by hepatic stellate cells (HSCs). MMP1, an upregulated gene upon BAG6 silencing, can attenuate hepatic fibrosis via collagen

type I and III breakdowns. Additionally, it has been shown that its gene delivery produces fewer HSCs [338]. Also, the knockdown of BAG6 downregulated the fibrogenic gene FN1. Thus, by the modulation of MMP1 and FN1, BAG6 knockdown attenuates hepatic fibrosis. Its knockdown also downregulated an apoptotic gene, TNFRSF21 and hence played a role in hepatic cell death. Finally, we observed that, like CASP3, its knockdown also increases the fluxes of the altered reactions involved in fatty acid activation and mitochondrial beta-oxidation pathways. These facts showed that BAG6 knockdown exerts a three-dimensional beneficial effect on NASH and could be considered a potential therapeutic target.

On the otherhand, NAFL is characterised by the buildup of hepatic lipids in excess quantities. This unregulated accumulation further increases lipotoxicity, which stimulates inflammation, hepatocyte death, and fibrosis, paving the way for the development of NASH. We discovered that eliminating each of the possible NAFL targets improved hepatic steatosis. Some of them, including CASP3, VDAC1, GABPB1, and FGFR3, were seen to ameliorate inflammation, hepatic cell death, and liver fibrosis. ACVR1 and BMP4 were seen to improve inflammation and hepatic cell death, while PPP1CC only improve the fibrosis process. As a result, targeting these proteins might prevent the transition from NAFL to the NASH stage.

Interestingly, by peering through the identified targets, we noticed the dominance of autophagy in NASH. Among the three potential targets, BAG6 is found to be an autophagy-related protein. It is essential for basal autophagy in mice embryos and for basal and starvation-induced autophagy in wild-type and BAG6 -/- mouse embryonic fibroblasts. Its absence reduces autophagosomes in cells and thereby reduces the autophagic flux, echoing its essentiality in forming autophagosomes. It is shown to modulate autophagy by affecting the intracellular localisation of EP300 [339]. Moreover, the cleaved N-terminal BAG6 (located in cytosol) interacts with both the LC3B-I and the unprocessed form of LC3B to suppress autophagy [340].

3.5 Conclusion

Our strategy of leveraging and interconnecting the context-specific molecular networks is the first of its kind to study NASH, where we identified three potential targets to control NASH

and eight targets to control NAFL. These potential targets exert their effects at both the gene and metabolic levels and reverse disease-associated gene signatures. We have shown that the knockdown of these potential therapeutic targets affects several critical metabolic reactions involved in steatohepatitis, and genes involved in inflammation and fibrosis development. This proposed methodology lays out a pragmatic framework for identifying potential therapeutic targets with a higher probability of success. It will save tremendous resources and time during the drug discovery process and be used as a general pipeline to identify targets in any *in silico* studies.

4

A data-driven multilayer approach for identification of potential therapeutic targets in non-alcoholic steatohepatitis[1]

An intriguing fact about proteins is that the proteins associated with a disease always remain in close proximity, and only a few of them gets identified as pathogenic. However, the crosstalk between all these proteins, irrespective of whether they are identified as pathogenic or not, governs the development and progression of diseases. There are many diseases in which various potential target proteins are identified *in vitro* or *in vivo* and are taken to clinical phases. Still, none of them eventually crosses all the clinical trial phases. Therefore, it won't be an

[1]The bulk of this chapter has been communicated for possible publication.

embellishment to assert that the existing disease proteins fail to capture the actual mechanism governing the disease progression. In this chapter, using NASH as an example, we have developed an approach that integrates whole-system protein perturbation, finds proteins that remain in close proximity to known disease proteins and then identify potential targets among them.

4.1 Introduction

A major challenge in the computational way of solving the conundrum of a disease system is to track down the effect of the protein perturbations on different layers of molecular understanding. The proteins identified from the protein-protein interaction (PPI) networks may not play a crucial role in metabolic level. The metabolic networks identify crucial reactions and the genes involved in them. However, such a network cannot provide information about how the gene products, i.e., proteins, will behave in conjunction. Also, the information about forming clusters, spreaders, role in information processing, etc., cannot be told from metabolic network analysis. Machine learning (ML), on the other hand, classifies proteins and can identify the nodes with the best predictive capability in the network. However, this approach alone can say nothing about metabolic adaptation, which is imperative for cell homeostasis following a physiological change. But, integrating with the ML approach in a study, the network analysis methods can be used to get more in-depth knowledge of the disease systems. Various studies have used machine-learning algorithm to investigate the classification capabilities of the hubs or other topologically strong proteins obtained from PPI analysis to identify important subset among them [341, 342]. Thus, an approach combining these three methods would identify proteins that are topologically important, have a strong predictive capability, and are able to influence metabolites which are crucial in the progression of NASH. In other words, a joint effort of these three approaches would enhance the probability of getting more viable disease targets.

The computational efforts in NASH, are primarily focussed on only one of the above three methods. The metabolic level investigations have considerably contributed to a deeper understanding of the condition, whereas the PPI-based research impede a thorough analysis of the

system (as mentioned in Chapter 4). In the metabolic level, one of the pioner work was done by Mardinglo et al. [316]. They developed a liver-specific genome scale metabolic model (GSMM) *iHepatocytes2322*, and used it to investigate the transcriptomic data obtained from NAFLD indivudals. According to their findings, the development of treatment strategies based on the augmentation of endogenous serine and AKG levels may cure the fundamental aetiology of NASH. In a subsequent investigation, it was demonstrated that plasma levels of serine and glycine are decreased in NASH patients, indicating a serine shortage in these patients [343]. ML-based studies are used to uncover non-invasive biomarkers and generate diagnostic scores to differentiate NAFL from NASH [344, 345]. For instance, Amaro et al. [341] developed an ML-based approach to liver histology assessment that characterises disease heterogeneity and severity and quantifies the treatment response in NASH. Perakakis et al. [346] used a variety of ML algorithms to find unique combinations of glycans, lipids, and hormonal variables to accurately diagnose the presence of NASH, NAFL, or normal status. However, because these studies [346, 347] are limited to biomarker detection, the aptness of machine learning to discover targets remains less explored.

With the advent of large-scale PPI networks, the application of graph-theory-based methods for their analysis, in an effort to glean insights into the information they carry about cellular function. These techniques take advantage of the propensity for functionally-related proteins to reside in the same network neighbourhood. Specifically, network-based guilt-by-association techniques have been utilised extensively to uncover new disease-associated genes. Random walk with restart (RWR), which was initially developed for internet search engines, is an effective guilt-by-association method [348]. It simulates the behaviour of an internet user, who, depending on his needs, can navigate between web pages using the accessible hyperlinks. Therefore, a few pages will be accessed more frequently than the rest during his internet session. RWR is a state-of-the-art method in computational biology to identify candidate disease genes /proteins. Here, a set of nodes, most preferably a set of disease nodes, are taken as seed(s), and the remaining nodes in the network are then ranked according to their proximity to these seed nodes.

In this study, we have developed a novel systematic approach that could be used to predict

potential therapeutic targets that can control the disease PPI network and metabolic landscape. The methodology starts with RWR, followed by a combination of classification approaches to identify candidate proteins that classify the disease condition and hence are hypothesised to be strongly associated with the disease. By analysing their controlling capability, the driver proteins are then identified. Finally, these driver proteins are used to investigate the altered metabolic landscape observed in NASH and identify possible targets for the development of effective NASH therapy regimens with likely minimum adverse effects. We have termed this multi-layer approach as a random walk restart multilayer approach (RWRMLA) and applied this to identify potential targets for NASH. The methodology used in this study is shown in **Figure 4.1**.

4.2 Materials and Methods

4.2.1 Construction of the undirected liver-specific protein-protein interaction network

We first constructed an undirected PPI network using the STRING database [215], which had 19385 nodes and 11938498 interactions. In this network, we kept only those interactions whose confidence score was greater than or equal to 900. The network contains 11749 proteins and 245760 interactions. The network was made liver-specific by integrating it with the nodes reported to be expressed in the liver as per the human protein atlas [349]. This liver-specific network (hereafter referred to as N_{900}) had 10,118 proteins and 2,09,828 interactions.

4.2.2 Random walk restart algorithm

Random walk with restart (RWR) algorithm [348] simulates a walker's passage from its current nodes to its neighbours in a network starting at several given seed nodes. It quantifies the proximity of a node to a set of seed nodes in a graph. The algorithm can be expressed as,

$$P_{t+1} = (1-r)MP_t + rP_0, \tag{4.1}$$

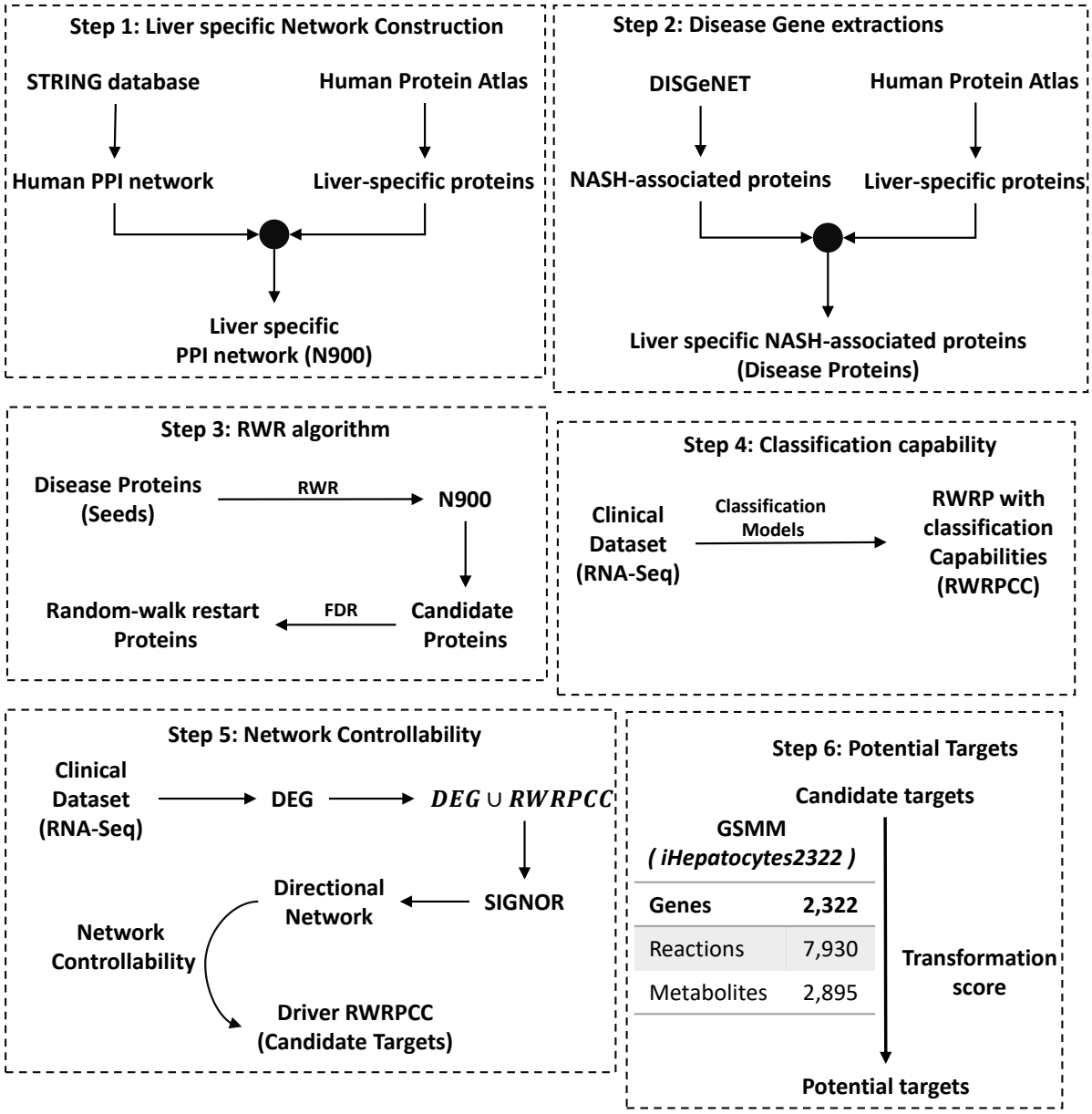


Figure 4.1: The workflow of the proposed method.

Here, P_0 is the vector of the initial probability distribution where only the seeds have values other than zero, the vector of the initial probability distribution where only the seeds have values other than zero. P_t is a vector whose 'i^{th} element represents the likelihood of getting the random walker at node 'i' at step 't'. M is the normalised adjacency matrix, 'r' is the global restart parameter. After multiple iterations, the difference between the vectors P_{t+1} and P_t becomes negligible and the elements in these vectors represent a proximity measure from every graph node to the seed(s). We have implemented RWR by taking the disease proteins as seeds and all the remaining nodes in the network as candidates. Each candidate node was assigned a score based on their proximity to the seed nodes. Proteins with a score greater than 0.00005 [350] were selected for further analysis.

4.2.3 Screening method to select RWRPs

To exclude the proteins that may arise due to topological biasedness, a filtering method as proposed in [351] was used. Here, 1000 sets were randomly produced, denoted by $S_1, S_2, ...,$ and S_{1000}, each with the same disease protein size. Using each of these S_i as a seed set, the RWR algorithm generates 1000 candidate sets (one for each S_i). The permutation false discovery rate was then calculated for each candidate protein 'p' as follows:

$$FDR(p) = \frac{\sum_{i=1}^{1000} f_i}{1000},$$

where $f_i = 1$ if the score of p is larger than the score computed by taking the disease proteins as seed nodes, or $f_i = 0$, otherwise. A candidate protein with a large permutation FDR is more likely to be a universal protein and less likely to be NASH-associated.

4.2.4 Data acquisition and pre-processing

To get the patient-specific transcriptomic data, we used the keywords: 'Non-alcoholic' and 'Human' in the Gene Expression Omnibus (GEO) database [239]. We further filtered out the datasets based on the following exclusion criterion, i) The datasets that contained samples of other diseases (such as HCC, HIV) or were infected with viruses (such as HBV, etc.) or were

Table 4.1: The comparison of two clinical datasets used in this study.

Dataset	Data Type	NASH definition	Normal individual	NASH individual
GSE162694	RNA-Seq	NAS	32	40
GSE126848	RNA-Seq	SAF score	26	16

treated with interventions (such as dietary intervention) were eliminated. ii) All *in vitro* models were eliminated. iii) Also eliminated are datasets that lack adequate stage categorisation descriptions. iv) From the remaining datasets, we chose only those containing RNA-Seq data. This is owing to the fact that RNA-Seq data distinguishes more differentially regulated transcripts, splice variants, and noncoding transcripts, hence shedding more light on numerous biomedical and biological topics. Finally, based on the sample size, we selected the datasets, GSE162694 [352] and GSE126848 [303]. A comparison of this two datasets is provided in **Table 4.1**. Two steps of data processing are performed in our study. In the first step, the genes with a raw count less than ten were removed using the FilterbyExpr function from the edgeR package [353, 354]. In the next step, the data was normalised using the trimmed mean of M values (TMM) method [355].

4.2.5 Selection of optimal feature from ML algorithm

There are numerous classification algorithms in machine learning, and each is biased toward its own objective function. So, redundant entities were removed using an ensemble strategy in six widely known algorithms: SVM, perceptron, decision tree, XGBoost, random forest and logistic regression. The recursive feature elimination (RFE) method has been used to identify the important features, which was initially implemented for SVM [356] and later has been implemented for other algorithms like random forest etc [357]. This method establishes a classification model utilizing all available features, ranks them by importance, and abandons the least important ones among them. The process of elimination goes on till the minimum number of features has been selected for maximum accuracy. This recursive process of eliminating the features is performed for 10-fold cross-validation for obtaining the reduced set. RFE has been used as a wrapper in all of these six algorithms to obtain minimum features in each algorithm.

4.2.6 Classification of the reduced RWR proteins

To build the machine learning (ML) models for assessing the classification capability, the RNA-Seq dataset GSE162694 [352] has been used where the gene represents the input features to be utilised in the algorithms. The SVM, perceptron, decision tree, XGBoost, random forest and logistic regression algorithms were compared to check the maximum possible prediction strength. For robust projection of results, a repeated 10-fold stratified cross-validation strategy has been incorporated [358]. The stratified k-fold cross-validation strategy randomly splits the disease and non-disease samples into k-equally proportioned subsets. Each time one subset was used as the testing dataset, and the rest of them were used as the training dataset. To further strengthen the ML models, a grid-based hyperparameter parameter search has been performed with cross-validation. The identified optimal parameters have been used to tune the ML algorithm for final prediction. The results of this binary classification problem also include accuracy, F1-score, precision, and recall along with AUROC.

4.2.7 Construction of the directional DEG network

To construct a directional network, the functional human PPI interactome is extracted from the SIGNOR database [216] and further filtered on the following criteria: i) both the interactors must be proteins, ii) the source protein should either upregulate or downregulate the target protein. Duplicated edges are then removed, and the final network (hereafter referred to as $N_{directional}$) thereby obtained was found to contain 4594 proteins and 11310 interactions. The DEGs and the RWRPCCs are next mapped to $N_{directional}$, and a subnetwork is constructed. The network thus constructed was found to be disconnected. Hence, $N_{directional}$ was used to obtain the minimum number of nodes which are required to connect the disconnected DEGs. These additional nodes are termed mediators. Some proteins, even after performing this step, remained disconnected and therefore removed from the network.

4.2.8 Prediction of gene knockdown effect in the metabolic network through genome-scale metabolic modelling

The gene knockdown profile of each gene was integrated in a functional genome-scale metabolic model (GSMM) for hepatocytes, *iHepatocytes2322* [316], to capture the knockdown effect in transforming the disease state back to the healthy state. To do so, the following preprocessing steps were performed: (1) Determining the baseline flux distribution of the disease (source) state (V^{ref}). To obtain the disease-specific GSMM, the average expression values of each metabolic gene was integrated into the *iHepatocytes2322* by applying the E-Flux method [359]. Additionally, a fasting condition was imposed as the liver biopsy samples were normally taken at the fasting state. During the fasting conditions, the liver uptakes gluconeogenic substrates (like lactate, glycerol, etc.), non-esterified fatty acids and amino acids. It produces glucose, very-low-density lipoprotein (VLDL), ketone bodies, and plasma proteins [343]. Hence, we selected lactate, glycerol, fatty acids, amino acids as input variables and glucose, VLDL, ketone bodies as output variables in the model. We also allowed the uptake of oxygen, phosphate, minerals, and the secretion of urea, H_2O, CO_2. The 'gpSampler' function implemented in the CobraToolbox 3.0 [360] was used for uniform sampling, and the mean of the different flux distributions was considered as V^{ref}.

(2) Analyzing the gene expression data of disease (source) and control (target) data to determine the changed and unchanged reactions of the model. A detailed Boolean gene-to-reaction mapping was employed to map the differentially expressed metabolic genes to reactions, representing the model's changed reactions under the disease state [361]. The average gene expression data of disease samples was used as a baseline and employed 2-fold up-or down-regulation on the expression values of the previously obtained up-and down-regulated genes, respectively, to predict the probable knockdown effect of each gene in disease situations (refer to Chapter 3, Section 3.2.8). This newly obtained probable knockdown specific gene expression data was integrated into the *iHepatocytes2322* using the E-Flux method, and the corresponding flux state V^{res} was predicted by applying the algorithm: Minimization of Metabolic Adjustment (MOMA) ([362]). MOMA basically minimises the Euclidean distance between the disease-specific flux distribution V^{ref} and the knockdown flux distribution V^{res}. Finally, a transforma-

tion score (TS) was assigned for each of the knockdown genes, similar to the scoring method proposed in the metabolic transformation algorithm (MTA) algorithm, using the relation,

$$TS = \frac{\sum_{i \in R_{success}} |V_i^{ref} - V_i^{res}| - \sum_{i \in R_{unsuccess}} |V_i^{ref} - V_i^{res}|}{\sum_{i \in R_s} |V_i^{ref} - V_i^{res}|}.$$

The changed relations are classified into two groups, success ($R_{success}$) and unsuccess ($R_{unsuccess}$) based on the transition of flux rates in the right direction, and the group (R_s) represents the unchanged reactions.

4.3 Results

4.3.1 Novel NASH-related proteins identified through the RWR algorithm

The proteins associated with a disease always remain in a close proximity, and only a few among them would have been previously identified as pathogenic [365]. Crosstalk between these proteins governs the development and progression of diseases. As no drugs have been approved for NASH, it won't be an embellishment to assert that the existing disease proteins fail to capture the actual mechanism governing the disease progression, necessitating the need to identify novel NASH-related proteins. To uncover them, we have used the random walk with restart (RWR) [348], which is capable of identifying novel proteins associated with disease development.

For this purpose, using the STRING database [366], a high-confidence undirected PPI network (with confidence score ≥ 900) containing 11,749 proteins and 2,45,760 interactions was first constructed. This general network was further filtered and made liver-specific by retaining only the nodes which are reported to be expressed in the liver as per the human protein atlas [349]. This liver-specific network (hereafter referred to as N_{900}) had 10,118 proteins and 2,09,828 interactions (**Figure 4.2A**). The final step before applying RWR requires a set of seed nodes in the form of NASH associated protein from the network. To get them, the DisGeNET database [367] was queried using the keyword 'Nonalcoholic Steatohepatitis', and 434 genes were extracted. These genes were mapped to the N_{900} network, and 336 proteins were obtained, which

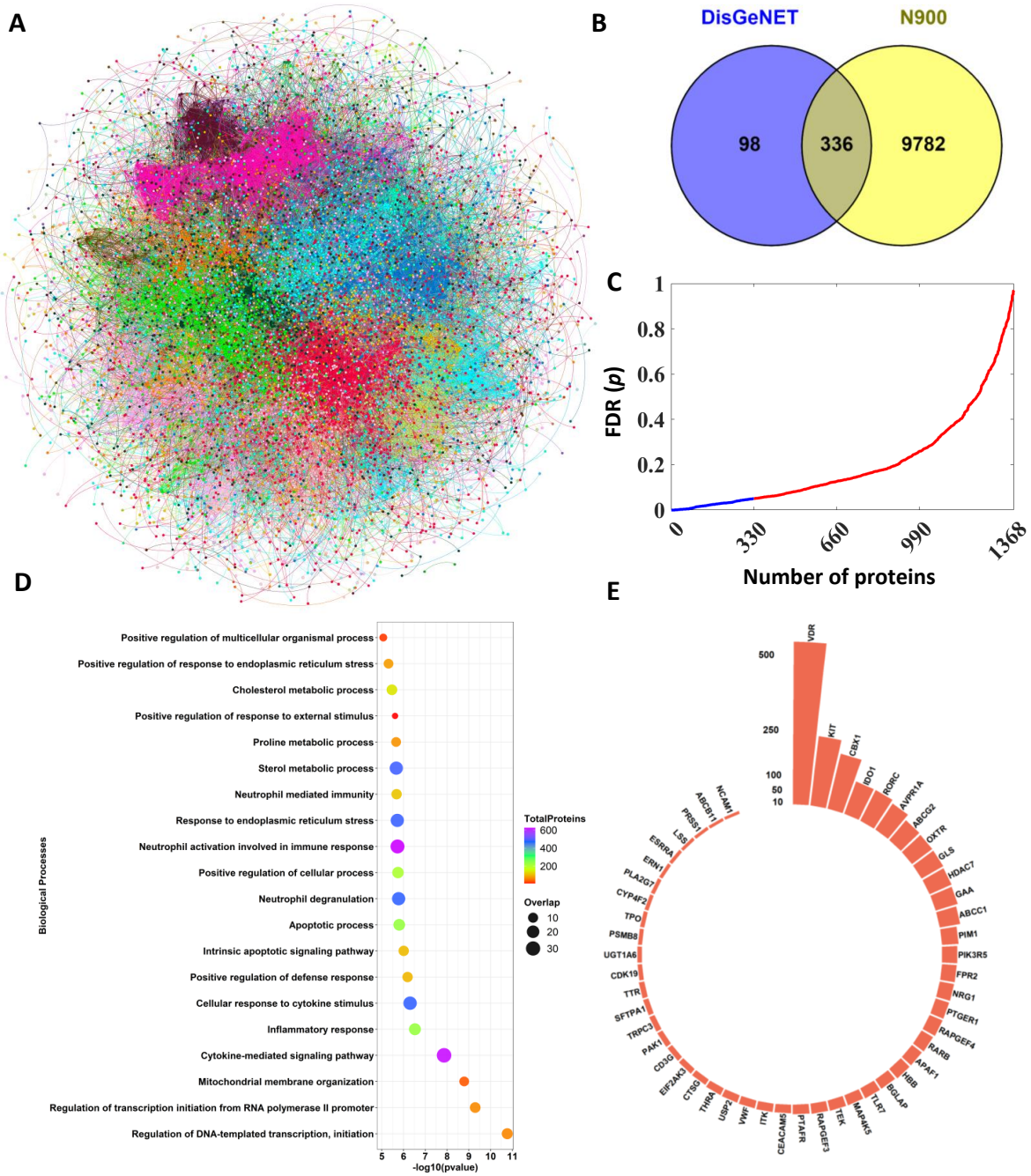


Figure 4.2: **Random walk with restart analysis.** A) The undirected liver-specific human PPI network, N_{900}, contains 10,118 proteins and 2,09,828 interactions. The nodes with the same modularity are given the same colour while the node size corresponds to its degree. B) Venn diagram representation of the proteins in the N_{900} and disease proteins. C) Permutation-FDR values of the proteins obtained from RWR. Here, an FDR cut-off value of 0.05 is used to select the 330 RWRPs shown in blue colour. D) GO-enrichment analysis of the RWRPs. The enrichment analysis is performed using Enrichr [363]. E) The number of drugs reported against these proteins as per the DGIdb [364]. Here, only the proteins which are associated with more than ten drugs are shown.

are referred to as 'disease proteins' (**Figure 4.2B**).

The RWR algorithm was applied to the network N_{900} with the 'disease proteins' as seeds and a score is assigned to each candidate node based on its proximity to the seed nodes. The proteins with a score greater than 0.00005 [350] were selected for further analysis. However, some of these proteins may only appear due to their higher topological significance. This structural biasedness is addressed by evaluating the relevance of each protein using a screening procedure [351] (refer to the Section 4.2.3), which results in 330 proteins (hereafter referred to as random walk restart proteins, RWRPs) (**Figure 4.2C**).

4.3.2 Topological significance of the RWRPs

We found that the average shortest path lengths of 340 RWRPs are less than the characteristic path length (4.0421) in the network. This tells that these proteins can rapidly spread information throughout the network. Again, 27 RWRPs appeared as hubs (degrees twice the average degree), suggesting that they can alter the stability of the disease network. The tendency of a node to form clusters is a crucial property in a network. Among the RWRPs, 129 possessed a higher clustering value than the average clustering value (0.4062) in the network. Finally, the investigation of the neighbourhood centrality revealed that 164 RWRPs have a greater neighbourhood centrality value than the average (50.7573), suggesting that these proteins are associated with high degree nodes in the network. The topological significance of these proteins is shown in **Figure 4.3**.

4.3.3 Functional enrichment of the RWRPs

The enrichment analysis elucidates the biological processes that must be explored to identify new therapeutic alternatives for disease [368]. The enriched biological processes (**Figure 4.2D**) associated with these RWRPs are cholesterol metabolism, sterol metabolic process, inflammatory response, cytokine-mediated signalling pathway etc., indicates that the RWRPs might play a significant role in the deterioration of healthy liver and the development of NASH. Next, they were mapped to the DGIdb database [364] and found that 135 (40.91%) of them were already in the clinical stages for different diseases (**Figure 4.2**). Among them, 51 are associated with

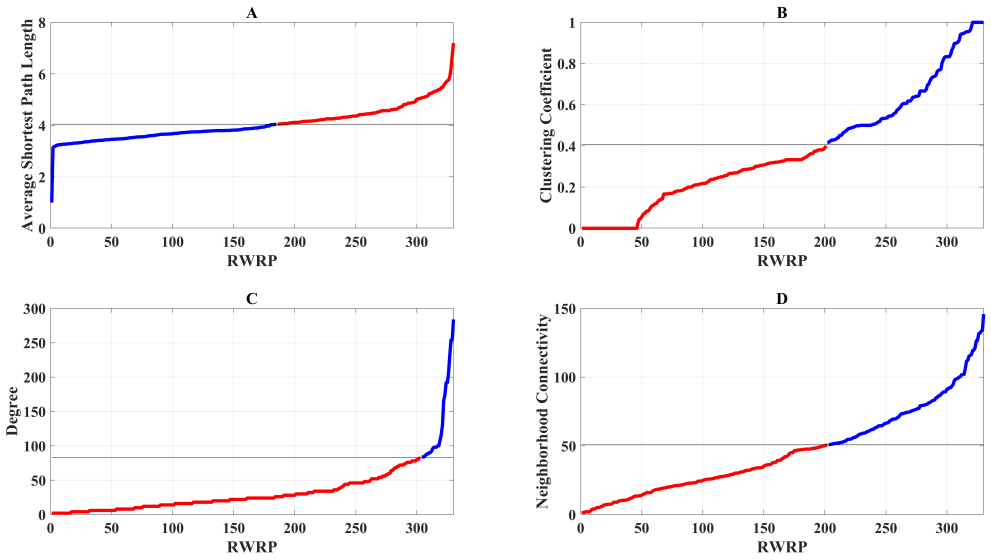


Figure 4.3: **Topological significance of the RWRPs across the four topological measures.** A) The average shortest path length, B) clustering coefficient, C) the total degree, and D) Neighborhood centrality of the RWRPs. The blue colour denotes that the RWRPs in this region are topologically significant.

more than ten drugs. This indicates that the RWRPs can also serve as a rapid and comprehensive strategy for identifying drug repurposing opportunities against NASH. Considering all this facts, it is reasonable to anticipate that the RWRP family is replete with potential targets for NASH.

However, the present study is keener in exploring and filtering these RWRPs further to identify targets that are better probable and robust. One such character to filter these RWRPs is its capability to distinguish disease.

4.3.4 Capturing RWRPs that are capable of disease classification

To delve into the classification capability of the RWRPs, a clinical dataset GSE162694 [352] was investigated, which includes RNA-Seq data of liver biopsy samples from 32 control and 40 NASH individuals. Following data filtration, 306 out of the 330 RWRPs were mapped with this dataset. The recursive feature elimination [356] technique is then applied as a wrapper with six prediction algorithms (SVM, perceptron, decision tree, XGBoost, random forest and logistic regression) using ten-fold cross-validation to obtain the best features for each classifier. The resulting feature sets from the algorithms were almost disjoint with a small overlap (**Figure**

4.4A). To select the optimal feature subset, the union of all the important features was taken for building the final feature set [369]. The new feature pool gives a more robust set of 199 features (hereafter referred to as RWRPCCs) and might enhance the predictive performance. Next, the collective predictability of these RWRPCCs was assessed to classify the clinical phenotype of control and NASH. The predictive performance was obtained in terms of accuracy, F1-score, area under the ROC curve (AUROC), Precision, and recall with repeated ten-fold cross-validation associated with the six algorithms . For the testing dataset, the obtained AUROC value for RWRPCCs ranged from 0.7 to 0.93 (**Figure 4.4B**), while the same for the 306 RWRPs ranged from 0.7 to 0.91 (**Figure 4.4C**). Whereas, on the training dataset, the obtained range for RWRPCCs is 0.89 to 1 (**Figure 4.5A**) and for the 306 RWRPs is 0.9 to 1 (**Figure 4.5B)**. Hence, the RWRPCCs showed a better classification capability than all the 306 RWRPs on the testing dataset. These RWRPCCs pushed the average AUROC value beyond 0.9 for four algorithms, namely SVM, KNN, decision tree, and logistic regression, while the remaining two algorithms exhibited a minor increase in AUROC value on the test dataset. The AUROC value above 0.9 was only achieved for all features in the random forest (0.91) and XGBoost (0.90). This suggests that the RWRPCCs successfully capture the inherent pattern in the expression data, which are exclusive to specific clinical labels.

The above classification value reports the maximum possible strength to distinguish the two clinical groups. However, these values were obtained using the default parameters of the ML algorithms and could be further enhanced by tuning hyperparameters like leaf size, number of estimators, solver etc. (**Table 4.2**). Using grid search with cross-validation, the optimal parameters set was obtained that had been further used to enhance the classification performance. The finely tuned models show the improved range of AUROC from 0.87 to 1 on the training dataset (**Figure 4.5C)** and 0.71 to 0.94 on the test dataset (**Figure 4.4D**). Along with enhancing the overall score range, it also improved AUROC in most of the prediction algorithms by $\geq$ 1% except SVM. So, the currently obtained models have the highest performance with a lesser possibility of overfitting.

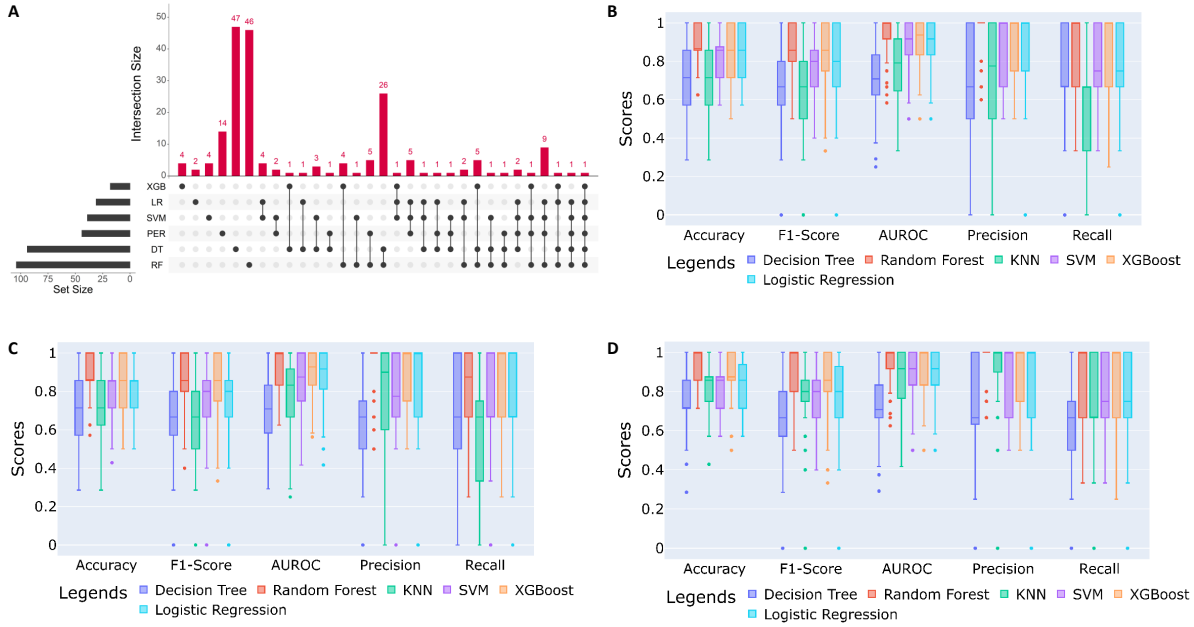


Figure 4.4: **Classification capability of the RWRPCCs.** A) Overlap between the important features. B) Model evaluation metrics of each algorithm in the test set using 199 RWRPCCs. C) Model evaluation metrics of each algorithm in the test set using 306 RWRPs. D) Model evaluation metrics of tuned algorithms in the test set using all 199 RWRPCCs.

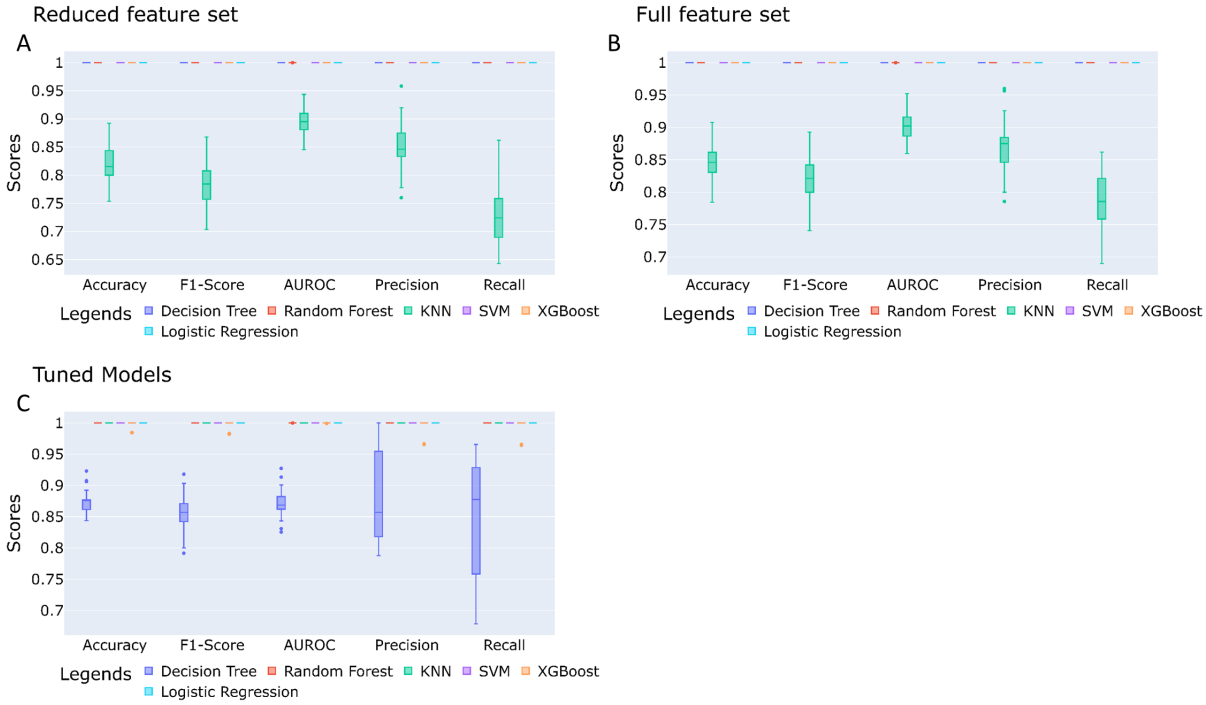


Figure 4.5: **Classification capability of the RWRPCCs (training set).** A) Model evaluation metrics of each algorithm in the train set using 199 RWRPCCs. B) Model evaluation metrics of each algorithm in the train set using 306 RWRPs. C) Model evaluation metrics of tuned algorithms in the train set using all 199 RWRPCCs.

Table 4.2: **Hyperparameters available in SKLearn library, python and their tuned value used in the study.**

1	Decision Tree	Criterion	Gini
		Max_depth	1
		Min_samples_leaf	1
		Min_samples_split	2
2	Random Forest	Max_features	sqrt
		N_estimators	1000
3	KNeighborsClassifier	Metric	Manhattan
		N_neighbors	7
		Weights	Distance
4	SVM	Kernel	Linear
5	XGBClassifier	Learning_rate	0.1
		Max_depth	7
		N_estimators	100
		Subsample	0.5
6	LogisticRegression	C	100
		Penalty	L2
		Solver	Lbfgs
		Max_iter	1000

So, the filtered list of proteins (RWRPCCs) are not only have the potential to be targetted but also have the ability to distinguish the disease. The accuracy was further confirmed with an independent clinical dataset, GSE126848 [303] (**Figure 4.6**). Now such proteins to be really considered for drug target should be capable of regulating the protein perturbations caused by the disease progression. This was investigated through the controllability paradigm of the RWRPCCs in the differentially expressed gene (DEG) network of NASH in the following section.

Figure 4.6: **Classification capability of the RWRPCCs in GSE126848.** A) Model evaluation metrics of each algorithm in the train set using 199 RWRPCC. B) Model evaluation metrics of each algorithm in the test set using all 199 RWRPCC.

4.3.5 Controllability paradigm of RWRPCCs in the differentially expressed gene (DEG) network

To identify the DEGs in the dataset GSE162694, we considered a fold change cut-off of 1.2 [370] and a false discovery rate (FDR) adjusted p-value cut-off of 0.05 [371]. We obtained 3451 up-and 1838 down-regulated genes denoting the perturbed protein profile in disease state. To be considered a target, a protein must be capable of affecting these perturbations. Network controllability is an effective method for identifying such proteins since it determines the smallest number of driver nodes necessary to control the complete network. To apply this algorithm we constructed a directional PPI network by integrating RWRPCCs and the DEGs to the functional human PPI interactome obtained from the SIGNOR database [216]. This directional PPI network (**Figure 4.7**, referred to as the DEG network since the majority of the nodes are DEGs) contained 575 DEGs and 50 RWRPCCs. To identify the RWRPCCs with control capability, the algorithm proposed by Guo et al. [372] was applied. It uses two predefined information, a set of target nodes and a set of constrained control nodes. Here, the RWRPCCs were taken as the control set, while the remaining DEGs of the network were taken as the target set. This process was repeated 1000 times, and the proteins which appeared at all the repetitions were considered as final driver proteins. Thus, the constrained controllability analysis finally identified 18 RWRPCCs that have the ability to control the DEG network. The fold change in gene expression of the driver RWRPCCs are shown in **Figure 4.8B**. The first neighbourhood PPI network of these proteins was abundant with various proteins which play a role in the development and

progression of NASH (**Figure 4.8C**). For example, SIRT5 increases the expression of G6PD, which is reported to increase oxidative and inflammatory response in adipose tissues of obese animals [373]. It also activates NFE2L2, which is shown to play a dual role in the NASH landscape [374]. ZEB1 inhibits FBP1, which deficiency is reported to disrupt the liver metabolism [375] and can cause fatty liver disease [376].

Thus, the network controllability method identified the proteins capable of changing the system's state. Given that the effects of these proteins are quantified in the DEG network, it won't be an embellishment to assert that they have the ability to regulate the perturbations caused by disease progression. Now, as a final evaluation of these short-listed proteins as potential targets, we set out to examine their effect on the metabolic landscape.

4.3.6 Effects of the driver RWRPCCs on the disease-associated metabolic landscape

NASH is a consequence of various metabolic alterations in which excess triglyceride (TG) synthesis and accumulations are seen. Hence, it demands further exploration of the identified driver RWRPCCs on the NASH metabolic state. GSMM is the widely used *in silico* method for understanding the disease-associated metabolic alterations [316, 370] and are also capable of identifying drug targets [361, 377]. Here, the DEGs were mapped into the liver-specific GSMM, *iHepatocytes2322* [316], and obtained 1285 and 1556 reactions associated with the up-and down-regulated genes, respectively. To capture the effect of the identified 18 driver RWRPCCs on the NASH metabolic state, the gene knockdown profiles of 3341 genes from the CMap database [378] were extracted, and the genes that were up-and down-regulated in response to the knockdown were identified (refer to Chapter 4, Section 3.2.8). Among these driver RWRPCCs, only 13 were found to be present in the CMap database. Amidst them, knockdown of PRKAR1A affected the most number of genes, followed by ETS1 (**Figure 4.9A-B**).

These profiles were then utilised to quantify the potentiality of transforming the NASH state into a healthy one. Based on the network perturbations, the transformation score (TS), as proposed by Yizhak et al. [361], was calculated for each of the 13 driver RWRPCCs. Basically,

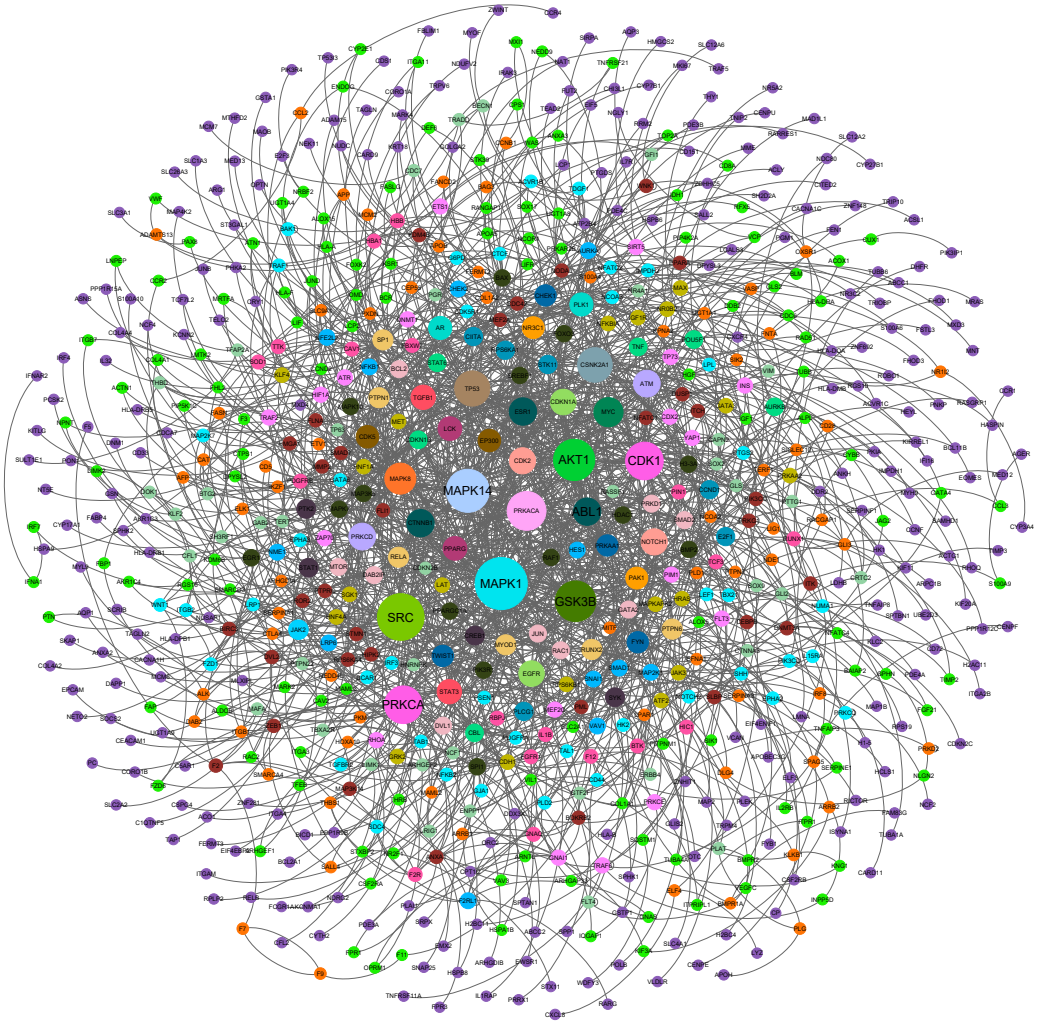


Figure 4.7: **The differentially expressed gene (DEG) network of GSE62694.** It contains 761 nodes and 1912 interactions.

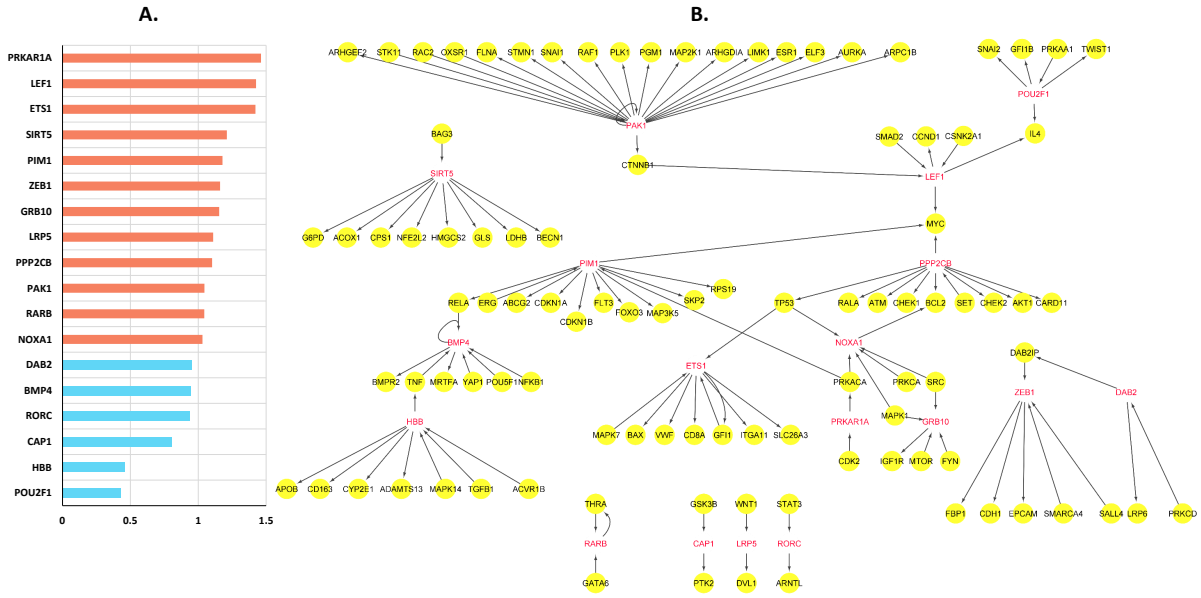


Figure 4.8: **The controllability analysis of the DEGs.** A) The expression fold change values of the driver RWRPCCs. The red and blue colour bar represent the up-and down-regulated proteins, respectively. B) The first neighbours of these proteins. The red-labelled nodes are the driver proteins of the network.

TS reflects to which extent a metabolic network perturbation may transform the disease state into a healthy flux state. A positive value indicates that the amount of successful changes in the flux state is more than the unsuccessful ones, and vice versa. Thus, RWRPCCs with positive TS were selected and finally obtained four potential targets: ETS1, LEF1, ZEB1, and PAK1. Next, the number of successfully transformed DEG-associated reactions were captured for all of these potential targets (**Figure 4.9C**). For example, the knockdown profile of PRKAR1A shows the highest effect on the DEG-associated reactions, followed by ETS1. The knockdown of the former can transform 88.63% (1139 reactions) and 15.61% (243 reactions) of the up-and down-regulated genes-associated reactions towards healthy ones, whereas knockdown of ETS1 shows the transformation of 85.21% and 15.55% reactions, respectively.

As per DGIdb, except for ZEB1, all three remaining proteins are clinically actionable [379]. Investigating their reported drugs in DRUGBANK [380], we found LEF1 is a target of Etacrynic acid [381], and PAK1 is a target of Fostamatinib [382]. No information on ETS1 and ZEB1 was found in the DRUGBANK database. For PAK1, one compound (CHEMBL3609372) is reported in CHEMBL [383] and three compounds (CID: 136590564, CID: 137125241, and CID: 138115195) in the PubChem database [384]. Any such information

for ETS1 in the aforementioned databases was not found. However, Jie et al. [385] have reported four novel small molecule inhibitors of ETS1. Again, investigating the literature, ZEB1 was found to be inhibited by Salinomycin [386]. Among these four potential targets, only ETS1 was found to be associated with hepatic fibrosis, as reported in the GWAS catalogue (study accession number GCST004938). To further see the association of these four potential targets with NASH, a co-expression analysis in NASH using the dataset GSE162694 was performed. All these four targets were found to be co-expressed with genes that govern various NASH-associated pathways. As an example, the co-expression partners of ETS1 that are associated with NASH-related pathways and processes are shown in **(Figure 4.9D)**. The neighbours of this protein in the N_{900} network also showed such abundance **(Figure 4.9E)**. Here, the redox related genes are taken from [387], fatty acid metabolism-related genes from the Virtual Metabolic Human database [388], autophagy-related genes from HAMDB [207], and ARN database [193], and the immune-related genes from [389].

4.3.7 ETS1 and autophagy

To find the affect of ETS1 on autophgay, we performed three separate analysis. We first look into its neighborhood in the N_{900} network and identified the abundance of autophagy protein in it. The same exercise is repeated for the DEG network. We also identified the numbers of up- and down regulated autophagy-related genes following the knockdown of ETS1. The result is shown in **Figure 4.10**. It is seen that the 25% of the ETS1 neighborhood are autophagy-related proteins. This justifies that the function of ETS1 highly affect autophagy process. **Figure 4.11** shows the neighborhood network and the effected genes.

Moreover, the knockdown of ETS1 is shown to increase autophagy in literature [390]. In their study, Zhang et al. had reported that, following the silencing of ETS1, the mRNA level of ATG5 and LC3II in pancreatic cancer cell lines increases as compared to the control.

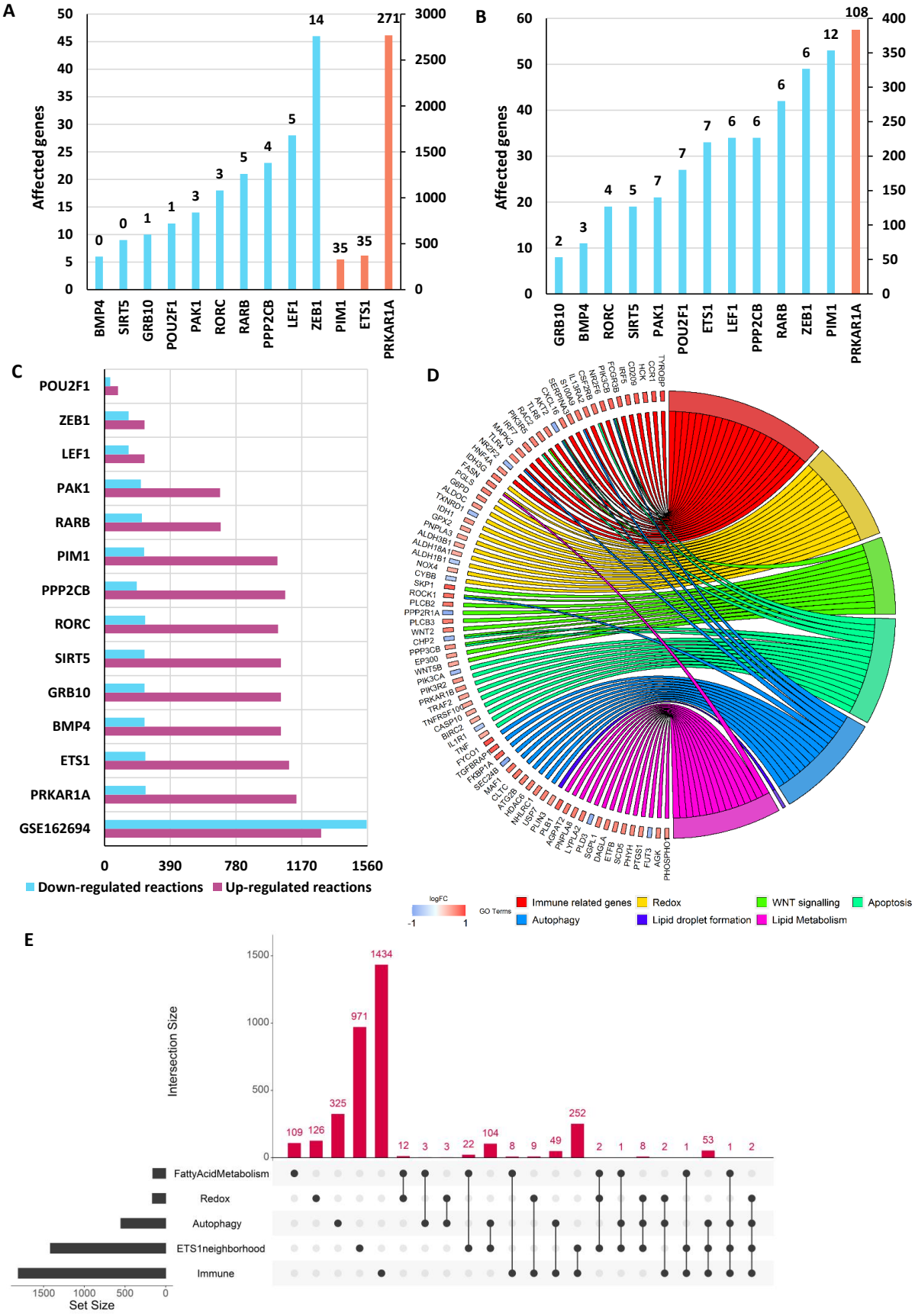


Figure 4.9: **Potential target identification in NASH.** A-B) The numbers of up-and down-regulated metabolic genes following the knockdown of the driver RWRP, respectively. For both the figures, the number above the bar is the abundance of metabolic genes. C) The number of up-and-downregulated reactions following the knockdown of each driver RWRPCC. D) The co-expression partners of ETS1 and their association with various pathways. E) The 2^{nd} neighbourhood of ETS1 in the N_{900} network. It can be seen that these proteins are abundant with various NASH-related pathways and processes.

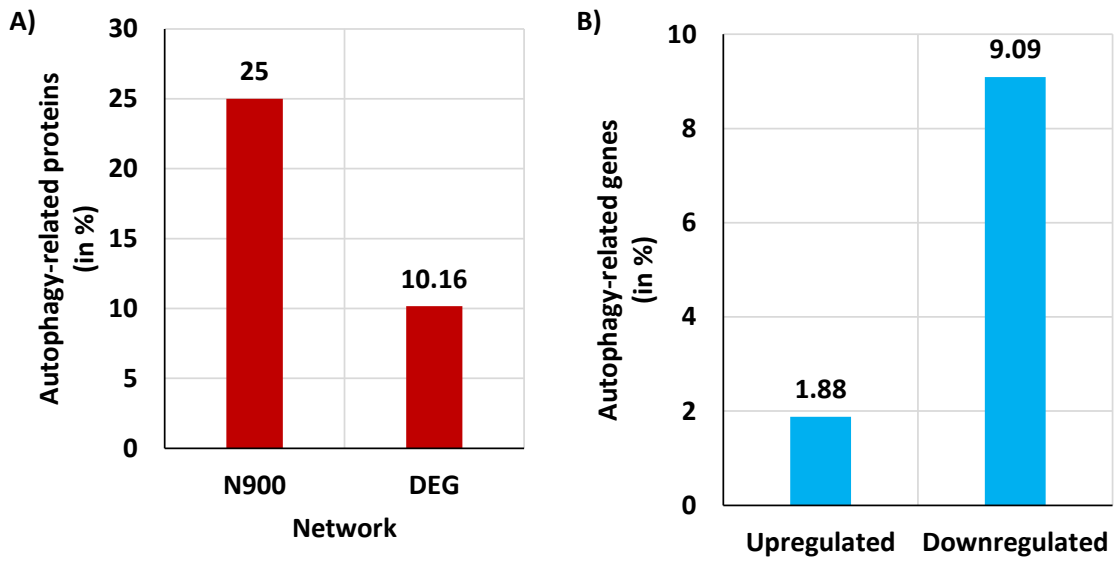


Figure 4.10: **Association of ETS1 and autophagy.** A) The figure shows the percentage of autophagy-related proteins in the neighborhood of ETS1 in the N_{900} and DEG network. B) The number (in %) of affected autophagic genes following the knockdown of ETS1.

4.4 Discussion

The present work and method contribute in measuring and controlling the alterations at different molecular levels. It also provides a systematic approach that will produce a minimum number of false positives.

Our methodology, the RWR multilayer approach (RWRMLA), identify the proteins that not only have the classification capability but also are capable of controlling the disease network. The identified proteins are the most crucial in the network and must be controlled if a transition from disease to a healthy system is sought. Finally, RWRMLA investigates the importance of these proteins in the metabolic landscape. Performing an in silico knockdown, it measures whether silencing a driver protein will initiate a disease to healthy transition or not. Since diseases result from aberrations at both protein and metabolic levels, the driver proteins that can show such behaviour will be termed as potential targets in a disease.

NASH is the inflammatory subtype of NAFLD and may lead to liver cirrhosis if not treated properly [391]. It is caused by complex interactions between metabolic and stress pathways in hepatocytes [392], triggered by chronically elevated lipid levels [393], and inflammatory processes mediated by multiple immune cell populations [283], collectively resulting in the

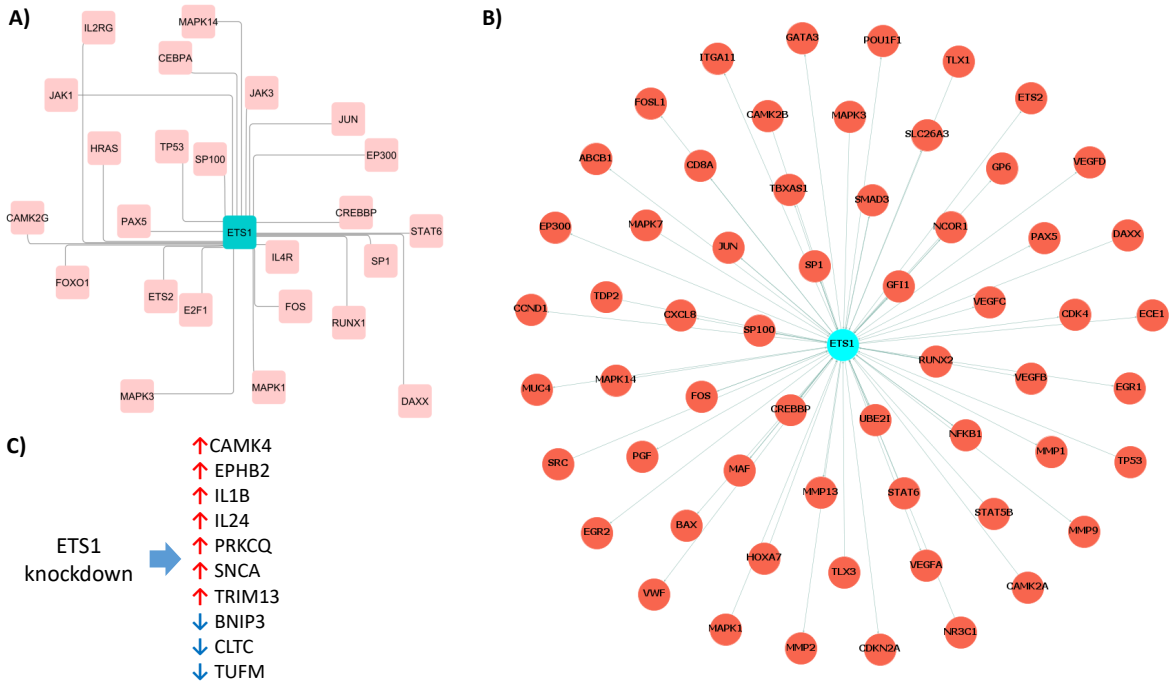


Figure 4.11: **Association of ETS1 and autophagy.** A) The neighborhood of ETS1 in the N_{900} network. B) The neighborhood of ETS1 in the DEG network. B) The affected autophagic genes following the knockdown of ETS1.

histological appearance of active steatohepatitis. Despite being studied for decades, no treatment for NASH has been developed.

We applied our methodology to identify new potential drug targets for NASH. To start, RWR is applied to a liver-specific PPI network, and a set of novel candidate proteins in NASH was identified. To eliminate any possible topological bias, this list was filtered using a screening process and obtained 330 proteins. The 330 proteins were further filtered using ML in a clinical dataset GSE162694, where 199 highly disease-associated proteins were identified. The disease associativity is defined based on their classification capability. The highest AUROC value achieved in this dataset was 0.93. The accuracy was further confirmed with an independent clinical dataset, GSE126848 [303] (**Figure 4.6**). Next, using network controllability, 22 driver proteins among these proteins were retrieved. Finally, GSMM was applied to investigate their ability to initiate a disease to a healthy transition and identified ETS1, PAK1, LEF1, and ZEB1 as potential targets. The four reasons to believe these proteins work as a target in NASH are: i) they are novel proteins associated with the known disease proteins, ii) they possess classification capability, iv) they can control the DEG network, and v) their silencing can initiate

a disease to healthy transition in NASH.

To substantiate our claim on the identified proteins as targets, we seek literature support and found that various studies have identified ETS1 as a promising target candidate [394–396]. It is a transcription factor and plays a role in regulating differentiation, proliferation, apoptosis, angiogenesis, migration, and cell metabolism. Its interacting partners, both at the protein and gene levels, are highly enriched with NASH-associated processes and pathways such as redox metabolism, immune pathway, autophagy and apoptosis, WNT-signalling pathway, development of steatosis, etc. The other target, LEF1, is also a transcription factor and acts as a mediator of the canonical Wnt signalling pathway by activating the Wnt-responsive genes in association with β-catenin [397]. Therefore, this protein can contribute to the development of fibrosis and activation of hepatic stellate cells (HSC), both of which are responsible for NASH development. Its first neighbours in the PPI network consisted of immune-related, autophagic, and apoptotic proteins. For instance, LEF1 is shown to upregulate MYC, whose overexpression activates HSC and promotes liver fibrosis [398]. It also upregulates CCND1, which is involved in both lipogenesis and gluconeogenesis in the liver [399]. The third target, ZEB1, is also a transcription factor which is shown to regulate the expression of its target genes by epigenetic mechanisms. Although it is widely reported for liver cancer, it can also play a crucial role in the development of NASH by inhibiting crucial proteins like FBP1 [400], whose loss is shown to disrupt liver metabolism [375]. Finally, PAK1 is a protein kinase and found to be associated with various NASH-associated proteins. For instance, it upregulates ELF3 [401], CTNNB1 [402], and PLK1 [403], which are reported to promote liver fibrosis [404]. These facts support that the identified targets might ameliorate NASH and initiate a disease to healthy transition and thus needs further experimental exploration.

RWRMLA allows itself to be adjusted in several ways to reduce its constraints and challenges and augment its accuracy and adaptability to a larger range of diseases. The first is fewer samples in the clinical dataset to perform ML-based analysis. The same, however, can be circumvented by performing the analysis in two or more different clinical datasets for the robustness of the analysis. The second limitation occurs in terms of the extraction of functional relationships between proteins. Signor [216] is an excellent database for extracting such in-

formation, but contains only 6584 proteins. However, this number can be improved through an extensive curation of literature and using various PPI direction prediction algorithms. The third limitation occurs in terms of getting the gene knockdown information. The only excellent source to obtain the same is the CMap database. However, it contains the information of only a few cell lines and a limited number of genes. Despite these limitations, this methodology lays forth a pragmatic framework for finding possible therapeutic targets with a greater probability of success and will save a substantial amount of time and resources during the drug discovery process.

4.5 Conclusion

Given the absence of effective therapeutic alternatives for NASH, it is not exaggerated to assert that the known disease proteins do not sufficiently represent the true mechanism of disease progression. For this purpose, we developed and applied a methodology RWRMLA on NASH, and identified four proteins as potential therapeutic candidates. RWRMLA is an efficacious generic methodology that can be applied to any metabolic disease to identify potential targets. Nevertheless, despite the fact that the proposed target proteins are implicated in potentially significant disease pathways, their efficacy must be confirmed by follow-up studies and assessed by experimental investigation.

5

The interplay between DNA damage and autophagy in lung cancer: A mathematical study[1]

The previous chapters used protein perturbations to identify large-scale data and potential targets. In all these chapters, we have seen the capability of autophagy to mitigate the progression and development of a disease. However, although these three studies have shown the significance of autophagy, they can not shed details on the mechanistic understanding of autophagy. For this purpose, in this chapter, we have developed a mathematical model to get mechanistic insight into some autophagic genes in cell proliferation and cell death in lung cancer.

[1]The bulk of this chapter has been published in Biosystems, 206 (2021) p.104443.

5.1 Introduction

Lung cancer is the global leader in cancer-related mortality in the population, with the second and fourth highest diagnoses among men and women, respectively [405]. While exposure to smoking is considered a major etiologic factor, non-smokers have also been diagnosed with lung cancer [406]. One of the challenges of lung cancer is its late detection. Therefore, the search for a more efficacious diagnosis and treatment of lung cancer has become obligatory. Various new methods, drugs, and pathways have been discovered to tackle lung cancer [407–409]. Although enormous advances have been seen in the diagnosis and treatment of lung cancer, several limitations have yet to be addressed, e.g., KRAS-mutated lung cancer shows no response to treatments targeting oncogenic mutations [410]. Resistance to molecular-targeted drugs and mechanisms has also been reported in the literature [410].

DNA damage leads to various diseases, including cancer, and often results from UV rays, oxidative stress, ionising radiation, and various genotoxic attacks. To protect DNA from such epochal events, the body maintains a strong cellular mechanism that acts as a saviour of DNA, and p53 plays an unequivocal role in this process. Due to its ability to repair DNA damage during stress, p53 is known as a protector of the genome. Under stress conditions, p53 activates and initiates cell cycle arrest, allowing the cell to correct potential defects. In addition to being an important presence in stress and nutritional response networks, the diverse activities of p53 are important in tumour suppression, metabolism, development, ageing, and neurodegeneration [411].

Autophagy is an evolutionary conserved lysosomal degradation process to maintain cellular homeostasis. There are three main types of autophagy, viz. macroautophagy, microautophagy, and chaperone-mediated autophagy. Here, by autophagy, we will be referring to macroautophagy, the autophagic process in which substrates are sequestered within autophagosomes. This essential process is found at basal levels in most of the cells. It plays a crucial role in controlling cellular homeostasis and therefore is consistently boosted in stress situations for the adaptation and survival of cells. It engulfs unnecessary or misfolded proteins and organelles inside the cell during stress. However, despite being a quintessential process, autophagy must be strictly controlled [14, 66] as the excessive degradation of cytosolic components may lead

to autophagic cell death. Both proliferation and deterioration of autophagy have been found to be associated with the pathogenesis of cancer [412, 413].

In comparison to the experimental work on autophagy that has been published, mathematical work on autophagy is relatively scarce. Kapuy et al. [178] studied the autophagy apoptosis interplay using a simplified mathematical model and revealed the underneath bistable nature. Martin et al. [182] presented a computational model of autophagic vesicle dynamics in single cells, Jennifer et al. [186] formulated a Petri net model for xenophagy, while Borlin et al. [185] formulated an agent-based model to understand the spatiotemporal autophagy dynamics. Mathematical models also need to be developed to investigate autophagy's role in determining cell fate in lung cancer. To the best of our knowledge, no mathematical work has been done elucidating the p53-autophagy association and beclin1-induced cell death following DNA damage, although autophagy is triggered after DNA damage as it is associated with DNA damage response pathways [414]. Again, no study has been done on beclin1-induced cell death following DNA damage. This association of p53 and autophagic cell death is vital as the former acts whenever any threat comes to DNA [415, 416]. In contrast, the latter may act as a saviour of tissue homeostasis by getting rid of the culprit cell.

Following severe DNA damage, p53 is known to induce apoptosis by activating its downstream apoptotic regulators. However, the ability of p53 to induce apoptosis gets compromised in all cancers, including the lung, allowing the cancer cells to grow and proliferate. This necessitates the need for an alternative pathway to mitigate the progression of cancer. In this paper, we have formulated a seven-dimensional non-autonomous ODE model to investigate the complex interplay between DNA damage, p53, autophagy, and lung cancer. We aim to highlight the potential factors or parameters and propose that autophagic cell death mediated by perturbation of these parameters over a specified range is the way forward in lung cancer research. We have shown that if the ability of p53 to repair DNA damage gets compromised and can no more suppress cancer, the cancer growth can be mitigated by the modulation of beclin1-mediated autophagic cell death, and AMPK and BCL2 play vital roles in this restoration.

5.2 Model description

5.2.1 Biological background

To capture the role of autophagy and p53, we considered a circuit with p53, MDM2, MDMX, AMPK, mTOR, BCL2, and beclin1 (**Figure 5.1**). Following any exposure to DNA damage, cellular p53 accumulates to promote the DNA repair process to reduce the risk of breeding mutations [415, 416]. To maintain the low level of p53 in the absence of stress, p53 and MDM2, a principal cellular antagonist of p53, form a negative feedback loop where p53 induces MDM2 expression, which for its part, encourages p53 degradation and quells the cellular activity of p53 [417–419]. The primary inhibition of p53 by MDM2 is by the proteasome-mediated degradation of p53 through the E3 ubiquitination ligase activity [416]. It also shuts p53 out to the cytoplasm from the nucleus [417], binds to p53 to prevent it from interacting with transcriptional co-activators p53 [420], and blocks its interaction as well as interacts with other nuclear corepressors to inhibit the activity of p53 [421]. In comparison, p53 regulates the expression of MDM2 by binding to its promoter [422]. The p53-MDM2 interaction is boosted by another p53 regulator, MDMX. MDM2 possesses an exceedingly low half-life and thus remains largely ineffective in down-regulating p53 in the absence of MDMX. MDMX enables MDM2 protein stable enough to function at its full potential for p53 degradation [423]. MDMX binds to p53 in its transactivation domain and inhibits the transcriptional activity of p53 [424]. Like MDM2, MDMX also gets positively regulated by p53 [425]. Furthermore, following DNA damage, MDMX is degraded by MDM2 due to the post-translational modifications of MDMX so that p53 can respond accordingly [425]. However, although there exist other pathways that affect the p53-MDM2 interaction, to keep the model simple, we did not consider them explicitly. Beclin1 is a quintessential protein in the process of autophagy. It comes to the autophagy scenario during the initiation stage, where it takes part in the formation of the isolation membrane, which engulfs the cytoplasmic material to form the autophagosome [426]. It acts as a bottleneck protein in the autophagy network and hence remains at the crossroads of many autophagy-related genes, some in favour of and some against the autophagy process, while many acts as dual regulators. Among our model variables, AMPK comes in the first, mTOR, and BCL2 in the

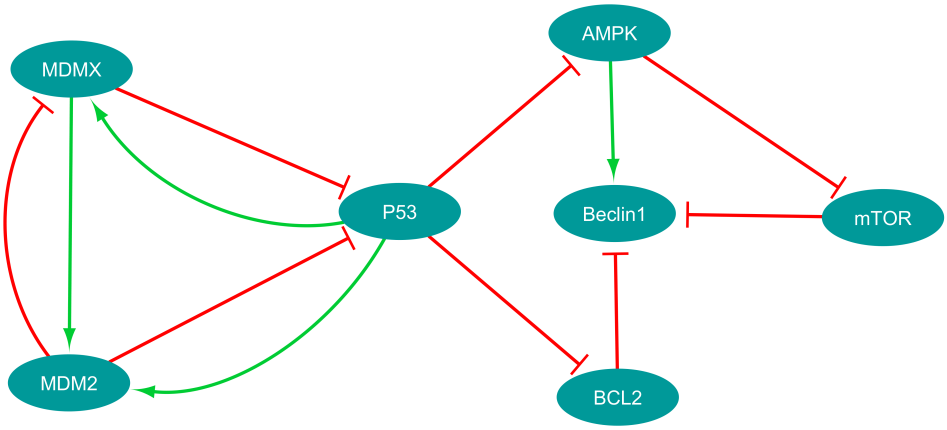


Figure 5.1: **Schematic diagram of the activation-inhibition mechanisms of P53-induced autophagy.** Here the green and red arrows represent activation and inhibition, respectively.

second, while p53 belongs to the third category. AMPK is the nutrient and energy sensor which catches the perturbation in the AMP: ATP, or ADP: ATP ratio and decides whether the system has sufficient energy (based on ATP concentration) or not and if not, it initiates the autophagy acceleration [427]. AMPK portrays a dual positive role for beclin1. It increases beclin1 activity by phosphorylation of beclin1 directly at the Thr388 site [427] and by negatively regulating the activity of mTOR, a negative regulator of autophagy [428]. *In vivo* studies have reported that the inhibition of mTOR increases beclin1 activity, indicating the negative regulation of beclin1 by mTOR [429]. Another negative regulator of autophagy is the ER-localized BCL2, which interacts with beclin1 to form beclin1: BCL2 complex and inhibits beclin1 activity [430]. Beclin1 should be released from this association for the initiation of autophagosome formation, which brings p53 into the autophagy scenario. p53 reduces the autophagy level by inhibiting AMPK [431], and by transcriptionally downregulating BCL2 expression, plays a positive role in autophagy [432].

5.2.2 Model formation

With the above biological background, we proposed the non-autonomous model (5.1), where the concentrations of p53, MDM2, MDMX, AMPK, mTOR, BCL2 and beclin1 are denoted by y_1, y_2, y_3, y_4, y_5, y_6 and y_7, respectively. In our proposed model, the p53-MDM2 interaction dynamics are extended versions of the model provided by Bar et al. [433]. They have mentioned the presence of an intermediary between p53 and MDM2, which is replaced in our model by

MDMX. Further complexity is added to their model by considering the activation of MDM2 by p53 and the inhibition of p53 by MDMX.

$$\begin{aligned}
\frac{dy_1}{dt} &= s_1 - \frac{d_1 y_1 y_2}{\theta + y_2} + \alpha e^{-\sigma t} y_1 - \frac{d_2 y_1 y_3}{\gamma + y_3} - a_{11} y_1, \\
\frac{dy_2}{dt} &= p_1 + \frac{p_2 y_3^{n_1}}{km_2^{n_1} + y_3^{n_1}} - d_3 y_2 + \frac{\alpha_1 y_1^{n_2}}{km_3^{n_2} + y_1^{n_2}}, \\
\frac{dy_3}{dt} &= \frac{c_1 e^{-\sigma t} y_1}{1 + c_2 y_1 y_2} - d_4 y_3, \\
\frac{dy_4}{dt} &= k_1 - \frac{d_6 y_1 y_4}{1 + y_1} - d_7 y_4, \\
\frac{dy_5}{dt} &= k_2 - d_8 y_4 y_5 - d_9 y_5, \\
\frac{dy_6}{dt} &= k_3 - d_{10} y_1 y_6 - d_{11} y_6, \\
\frac{dy_7}{dt} &= k_4 - d_{12} y_5 y_7 + a_1 y_4 - d_{13} y_6 y_7 - d_{14} y_7.
\end{aligned} \tag{5.1}$$

The variation in p53 is represented by the first equation of the system (5.1). The basal production and degradation rates of p53 are denoted by the terms s_1 and a_{11}, respectively. The parameter s_1 is dependent on various factors that induce p53 following DNA damage. As already mentioned, p53-MDM2, and p53-MDMX form two negative feedback loops and the effects of MDM2 and MDMX on p53 has been incorporated by two Hill-type functions $\frac{d_1 y_1 y_2}{\theta + y_2}$ and $\frac{d_2 y_1 y_3}{\gamma + y_3}$, respectively, where θ and γ are the half-saturation constants, d_1 and d_2 are the maximum degradation rates of p53 by MDM2 and MDMX, respectively. The Stress-dependent p53 activation term is taken from the reference [433] and is represented by the term $\alpha e^{-\sigma t}$, a decreasing function of time. The stress is maximum at $t = 0$ that gradually fades with the resolution rate of σ. The parameter α is the stress coefficient and attains the value 0 when there is no stress. The second equation represents the MDM2 dynamics. The term p_1 represents the p53-independent MDM2 transcription and translation rate. The second term, $\frac{p_2 y_3^{n_1}}{km_2^{n_1} + y_3^{n_1}}$ represents MDMX-induced MDM2 activation [423, 434] as described by Bar et al. [433], where km_2 is the Michelis constant, p_2 is the maximum production rate of Mdm2 and n_1 is Hill's coefficient in the production of MDM2 via MDMX. d_3 represents the basal degradation rate of

MDM2. The p53-dependent activation of MDM2 follows a Hill-type function [435, 436] and is represented by the term $\frac{\alpha_1 y_1^{n_2}}{km_3^{n_2} + y_1^{n_2}}$, where α_1 denotes the maximum production rate of MDM2 by p53, km_3 and n_2 denote the corresponding Michelis constant and Hill coefficient. The first term of the third equation was built following [433]. It is reported that MDM2 binds to p53 to reduce its transcriptional activity, causing a decrement in the production of MDMX [425]. Such MDM2-dependent inhibition in the production of MDMX is incorporated through the denominator function $(1 + c_2 y_1 y_2)$, where c_2 is the MDM2-dependent inhibition rate. The numerator shows the production of MDMX by p53 and stress with c_1 as its maximum production rate. The basal degradation rate of MDMX is denoted by the term d_4. The fourth equation represents the rate equation of AMPK. Its basal production and degradation follow the exponential law with k_1 and d_7 as the rate constants. p53 has an inhibitory effect on the AMPK level, as stated earlier. Such a sigmoidal inhibitory effect of p53 on AMPK is represented by the second term in this equation with d_6 as its maximum value. The basal production and degradation rates of AMPK are denoted by k_1 and d_7, respectively. The next equation describes mTOR dynamics. Here, the second and third terms represent AMPK-induced and basal degradation rates of mTOR, while the first term represents its basal production. d_8, d_9 and k_2 are the respective rate constants. In the sixth equation, the transcriptional down-regulation of BCL2 by p53 is denoted by the term $d_{10} y_1 y_6$ with rate constant d_{10}. The basal production rate of BCL2 is measured by the parameter k_3, while its exponential decay rate is denoted by d_{11}. The last equation represents beclin1 dynamics. The baseline production rate of beclin1 is k_4, and its exponential decay occurs with a rate constant d_{14}. $a_1 y_4$ is referred to as the phosphorylation rate of beclin1 by AMPK, where a_1 is the rate constant. On the other hand, $d_{13} y_6 y_7$ and $d_{12} y_5 y_7$ represent the inhibitory effect of BCL2 and mTOR on beclin1, where d_{13} and d_{12} are the respective rate constants.

5.2.3 Some preliminary analysis

The equations of system (5.1) represent the protein dynamics and hence it is of prime importance to show that the state variables $y_i, \mathrm{i} = 1:7$ are non-negative and bounded. This analysis cinches that the model is well-posed, and no anomaly is there in the realistic portrayal of the protein dynamics.

5.2.3.1 Positivity

Let $Y = (y_i)^T$, i $= 1:7 \in \mathrm{R}^7$ and $F(Y) = [F_i(Y))]^T$, i $= 1:7$. The system (5.1) can be written as

$$\dot{Y} = F(Y),$$

together with suitable initial conditions $Y(0) = Y_0 \in R_+^7$. It can be easily checked that whenever $Y_0 \in R_+^7$ with $Y_i = 0, i = 1:7$, then $F_i(Y|_{Y_i=0}) \geq 0$. Thus, from the lemma of Nagumo [437], any solution of the system (5.1) with $Y_0 \in R_+^7$, say $Y(t) = Y(t, Y_0)$, is such that $Y(t) \in R_+^7$ for all $t > 0$.

5.2.3.2 Boundedness

$$\dot{y}_3 = \frac{c_1 e^{-\sigma t} y_1}{1 + c_2 y_1 y_2} - d_4 y_3$$

$$\implies \dot{y}_3 + d_4 y_3 \leq c_1.$$

Therefore, the solution is bounded by the condition

$$y_3 \leq \frac{c_1}{d_4}.$$

Again,

$$\frac{dy_2}{dt} = p_1 + \frac{p_2 y_3^{n_1}}{k m_2^{n_1} + y_3^{n_1}} - d_3 y_2 + \frac{\alpha_1 y_1^{n_2}}{k m_3^{n_2} + y_1^{n_2}}$$

$$\implies \dot{y_2} + d_3 y_2 \leq p_1 + p_2^2 + \alpha_1$$

$$\implies y_2 \leq \frac{p_1 + p_2^2 + \alpha_1}{d_3}.$$

Finally,

$$\dot{y}_1 = s_1 - \frac{d_1 y_1 y_2}{\theta + y_2} + \alpha e^{-\sigma t} y_1 - \frac{d_2 y_1 y_3}{\gamma + y_3} - a_{11} y_1$$

$$\implies \dot{y}_1 \leq s_1 - \beta y_1 \implies y_1 \leq \frac{s_1}{\beta},$$

where $\beta = \min\left(\frac{d_1 y_2}{\theta + y_2} + \frac{d_2 y_1 y_3}{\gamma + y_3} + a_{11} - \alpha e^{-\sigma t}\right)$.

$$\dot{y_4} = k_1 - \frac{d_6 y_1 y_4}{1 + y_1} - d_7 y_4 \implies \dot{y_4} + \left(\frac{d_6 y_1}{1 + y_1} + d_7\right) y_4 \leq k_1$$

$$\implies y_4 \leq \frac{k_1}{\frac{d_6 y_1}{1 + y_1} + d_7} \implies y_4 \leq \frac{k_1(1 + y_1)}{d_6 y_1 + d_7(1 + y_1)}.$$

Since y_1 is bounded, therefore y_4 is bounded. Similarly, it can be shown that

$$y_5 \leq \frac{k_2}{d_8 y_4 + d_9},$$

$$y_6 \leq \frac{k_3}{(d_{11} + d_{10} y_1)},$$

$$y_7 \leq \frac{k_4 + a_1 y_4}{(d_{12} y_5 + d_{13} y_6 + d_{14})}.$$

Thus, the solution of the model with positive initial condition is bounded and the bounds are

1. $y_1 \leq \frac{s_1}{\beta}$, where $\beta = \min\left(\frac{d_1 y_2}{\theta + y_2} + \frac{d_2 y_3}{\gamma + y_3} + a_{11} - \alpha e^{-\sigma t}\right)$,
2. $y_2 \leq \frac{p_1 + p_2^2 + \alpha_1}{d_3}$,
3. $y_3 \leq \frac{c_1}{d_4}$,
4. $y_4 \leq \frac{k_1(1 + y_1)}{(d_6 + d_7) y_1 + d_7}$,
5. $y_5 \leq \frac{k_2}{d_8 y_4 + d_9}$,
6. $y_6 \leq \frac{k_3}{(d_{11} + d_{10} y_1)}$,
7. $y_7 \leq \frac{k_4 + a_1 y_4}{(d_{12} y_5 + d_{13} y_6 + d_{14})}$.

The nonlinear non-autonomous system (5.1) is too complicated for analytical solutions, and we, therefore, opted for its numerical solutions.

Table 5.1: **Description of the model parameters of the p53 module with their default values and reference.**

Sl. No.	Parameter	Description	Value	Reference
1	s_1	Basal production rate of p53	0.5 nM min^{-1}	[433]
2	d_1	Degradation rate of p53 by MDM2	2.1 min^{-1}	estimated
3	α	Stress coefficient	0.1099 min^{-1}	estimated
4	σ	Resolution rate of stress signal	0.0001 min^{-1}	[433]
5	d_2	Degradation rate of p53 by MDMX	0.000125	[433]
6	p_1	p53-independent MDM2 transcription and translation	0.00235 nM min^{-1}	[433]
7	p_2	MDMX dependent production rate of MDM2	0.003 nM min^{-1}	[433]
8	n_1	Hill's coefficient in the production of MDM2 via MDMX	50	[433]
9	km_2	Michealis constant in the production of MDM2 via MDMX	25 nM	[433]
10	d_3	Degradation rate of MDM2	0.05 nM min^{-1}	[433]
11	α_1	input dependent production rate of MDM2	0.001 nM min^{-1}	estimated
12	n_2	Hill coefficient in the production of MDM2 via p53	0.13	estimated
13	km_3	Michelis constant in the production of MDM2 via p53	4 nM	estimated
14	c_1	Input dependent production rate of MDMX	0.09 min^{-1}	[433]
15	c_2	Mdm2 dependent degradation of MDMX	0.01 nM^{-2}	[433]
16	d_4	Degradation rate of MDMX	0.09 min^{-1}	estimated
17	a_{11}	basal degradation rate of p53	0.0095 min^{-1}	estimated
18	θ	half maximal rate of MDM2	1.05 nM	estimated
19	γ	half-maximal rate of MDMX	1 nM	estimated

5.3 Numerical simulations

To find different types of dynamical behaviour that can be unveiled by the model (5.1) in various parameter regions, we solved the system numerically using the values of the parameter set given in **Table 5.1**. The parameter set so chosen depicts the experimentally observed oscillations of p53 as shown in [433]. Here, parameters are either taken from reference [433] or estimated to produce **Figure 5.2**. As shown in this figure, the normal level of p53 will imply a healthy system, and if in the no-repair region, it will indicate a lung-cancerous system. We will analyse the healthy system first to identify potential threats that could lead to the system deteriorating into a disease state, and then the diseased system to identify potential targets for restoring the system's healthiness. For the analysis of the whole system, we have considered the hypothetical values for the rest of the parameters. These parameter values are estimated to keep beclin1 in the normal range as given by [438]. To make the system parameter independent, later, we performed the parameter sensitivity analysis using the Partial Rank Correlation Coefficient (PRCC) method [95].

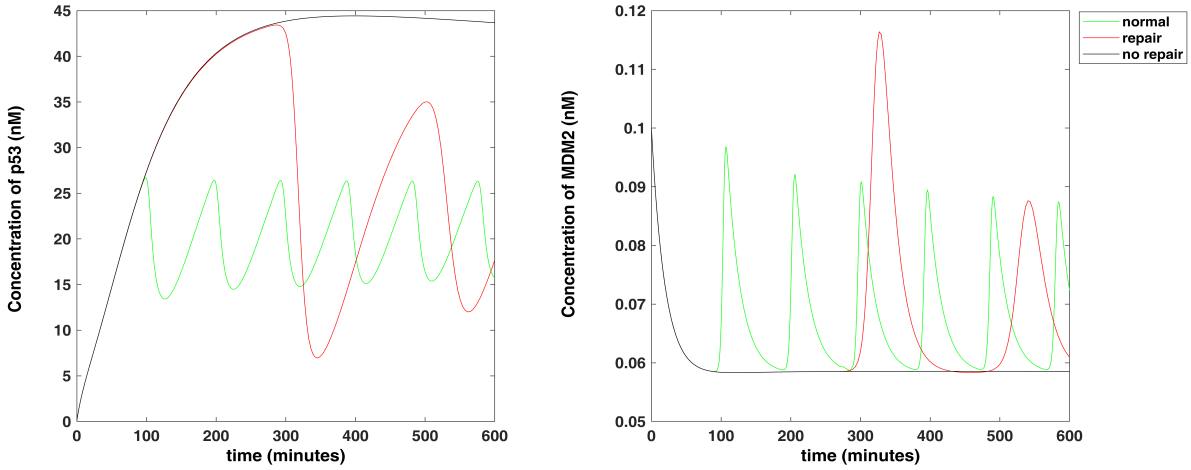


Figure 5.2: **The dynamics of p53.** The levels of p53 denote the state of DNA. The green curve denotes the p53 pulses when the system is in a healthy state and hereafter considered as the normal level of p53. The elevation of p53, denoted by the black curve, indicates the severe DNA damage and hereafter is considered as the no-repair phase of p53. The red curve denotes an intermediate phase of p53, where DNA damage is not that extreme and can be repaired. The right figure denotes the MDM2 level for three similar cases. The corresponding non-repairable case, however, indicates an almost undetectable level of MDM2, an opposite characteristic of p53. The time is calculated in minutes. The green curve is generated by using the parameter values of **Table 5.1**, the black curve is generated by keeping $c_1 = 0.004$ and the red curve is generated by keeping $c_1 = 0.004, d_4 = 0.004$. The figure shows the p53 dynamics as mentioned in [433]

Table 5.2: **Description of the hypothetical parameters to keep beclin1 in normal range as given in reference [438]**.

Sl. No.	Parameter	Description	Value
1	k_1	Basal production rate of AMPK	0.5 nM min^{-1}
2	d_6	p53 dependent inhibition rate of AMPK	0.17 $\text{nM}^{-1}\text{min}^{-1}$
3	d_7	Basal degradation rate of AMPK	0.0159 min^{-1}
4	k_2	Basal production rate of mTOR	0.5 nM min^{-1}
5	d_8	Degradation rate of mTOR by AMPK	$0.4\text{nM}^{-1}\text{min}^{-1}$
6	d_9	Basal degradation rate of mTOR	0.2min^{-1}
7	k_3	Basal production rate of BCL2	$0.09nM\text{min}^{-1}$
8	d_{10}	Degradation rate of BCL2 by p53	$0.0162\text{nM}^{-1}\text{min}^{-1}$
9	d_{11}	Degradation rate of BCL2	0.015min^{-1}
10	k_4	Basal production rate of beclin1	0.5 nM min^{-1}
11	a_1	Production of beclin1 by AMPK	0.00001 min^{-1}
12	d_{12}	mTOR dependent inhibition of beclin1	0.1 $\text{nM}^{-1}\text{min}^{-1}$
13	d_{13}	Association rate of beclin1:BCL2 complex	0.68 $\text{nM}^{-1}\text{min}^{-1}$
14	d_{14}	degradation rate of beclin1	0.145min^{-1}

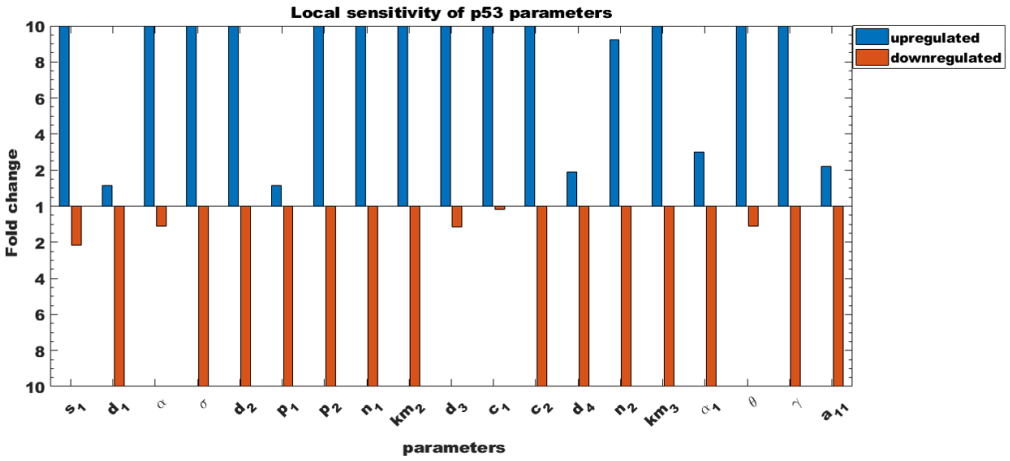


Figure 5.3: **Local sensitivity analysis of p53.** Each parameter was varied over a range of ± 10 fold from their defined values reported in Table 5.1, with other parameters fixed as in (**Table 5.1**). The bars represented the fold divergence from the base value, and its length indicated the ranges beyond which the dynamics of p53 shift from normal to no-repair state. The blue colour was used when the result was obtained by increasing the parameter value and the red color in the opposite case.

5.3.1 Parameter sensitivity analysis of p53 module

Often, the output behaviour of a high-dimensional system with numerous parameters is determined by a small number of parameters. Sensitivity analysis (SA) provides a way to distinguish these parameters so that they can be targeted by further studies. A local sensitivity analysis (LSA) for p53 was carried out by perturbing one parameter at a time over a range of ± 10 fold from its reference value (**Table 5.1**) while the remaining parameters were kept at their corresponding reference values. The local sensitivity measures how much disturbance is needed for a parameter to propel p53 from a normal to a no-repair region (**Figure 5.2**). The LSA results (**Figure 5.3**) show that the parameters s_1, α, d_3, c_1, and θ are sensitive when they are down-regulated and the parameters d_1, p_1, d_4, α_1, and a_{11} are sensitive when they are up-regulated, while the parameters $\sigma, d_2, p_2, n_1, km_2, c_2, n_2, km_3$, and γ are robust.

The result from LSA is acquired by keeping most of the parameters fixed. To check the robustness of the result under the uncertainties associated with the model's parameters, we performed a global sensitivity analysis (GSA), which allows simultaneous and uniform variations of model parameters from their basal values. GSA (**Figure 5.4**) for the healthy system is performed following Latin Hypercube Sampling (LHS), and Partial Ranked Correlation Co-

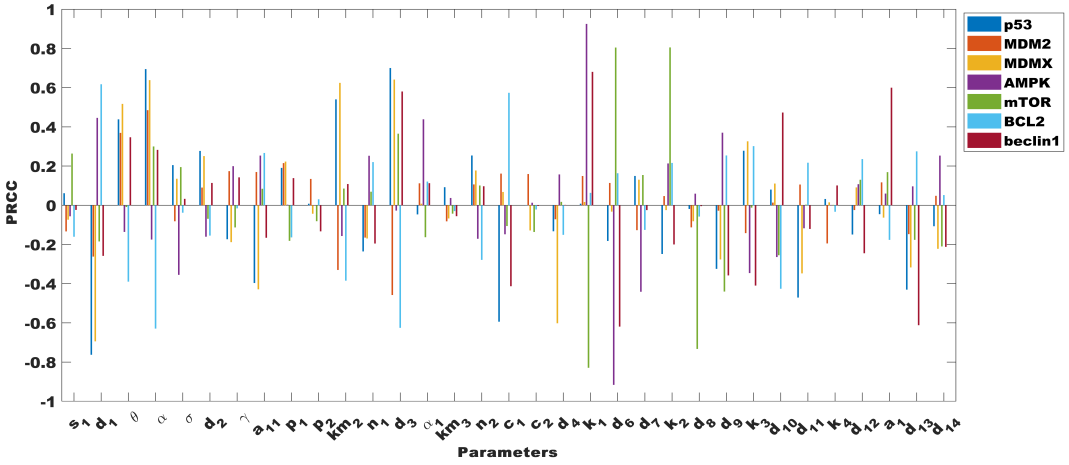


Figure 5.4: **Sensitivity analysis of model parameters using the LHS-PRCC sensitivity analysis.** We evaluated the PRCC value for each parameter related to each of the seven variables.

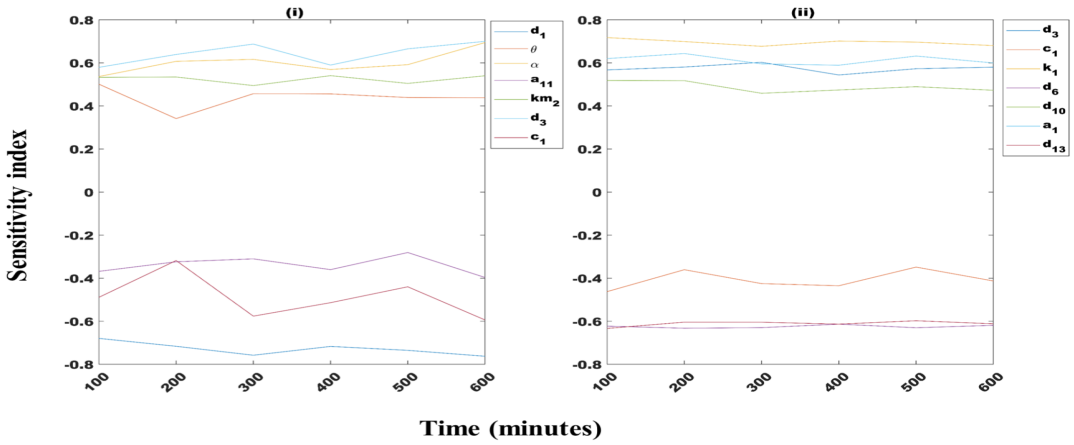


Figure 5.5: **Temporal sensitivity analysis of the sensitive parameters obtained from Global sensitivity analysis.** Figure (i) denotes the variation of sensitive parameters for p53, while figure (ii) denotes the variation of sensitive parameters for beclin1.

efficient (PRCC). The length of the bar represents the PRCC index of the parameter which illustrates the linearity of the parameters for the specific output variable. The results obtained from GSA were nearly identical to those obtained from LSA. We took a sensitivity index cutoff of ± 0.4 and found that $d_1, \theta, \alpha, d_3, a_{11}, km_2$, and c_1 were among the parameters with the most effect on p53. Among these parameters, km_2 was found to be robust in the LSA, while c_1 was found to be less sensitive. The other robust parameters from local sensitivity analysis were also found to be less significant in GSA. To find the parameters affecting the beclin1 dynamics, we took a ± 0.4 PRCC index cut off and found that the parameters $d_3, c_1, k_1, d_6, d_{10}, d_{13}$, and a_1 were the most significant parameters.

We performed a temporal sensitivity analysis to ascertain the variation in sensitivity of the GSA-identified sensitive parameters for p53 and beclin1. We have measured the sensitivity of these parameters every 100 minutes, and the results are shown in **Figure 5.5**. The results indicated that the parameters d_1, d_3, and α had a reduced sensitivity index at t=400 but increased for subsequent time points. The sensitivity of the parameter θ was lower at t=600 than at t=100, whereas the sensitivity of the parameter km_2 varied across time points. a_1 and c_1 become more sensitive at later time points. For beclin1, apart from d_3 all the parameters showed a decrease in sensitivity at t=600 than at t=100. The sensitivity of d_3 was also found to be low at t=400, which, however, increases for the later time points.

5.3.2 Recalibration of the parameters of the p53 module

The presence of normal p53 oscillations indicates the removal of any oncogenic hazard that leads to DNA damage [439]. The capability of p53 to suppress DNA damage can be boosted but also carries the threat of getting lowered by the variation of some parameters. The robust parameters obtained from the sensitivity analysis produce no change in the normal behaviour of p53. In other words, they maintained the goodness of the system until the 10-fold barrier. So in an attempt to enhance the aforementioned goodness of the system, we, therefore, need to increase the robustness of the sensitive parameters derived from LSA. We chose the robust parameters from LSA as target parameters and increased each of them one by one to 10 fold from their reference value (**Table 5.1**) and varied other parameters to measure the change in their robustness. We repeated the same exercise by decreasing the robust parameters to 10 fold. By removing the parameters with almost no changes (the difference between new sensitivity and old sensitivity with a cut off ± 0.08), the results are shown in **Figure 5.6**. It has been seen that the parameters showed no significant changes in their sensitivity level, although the robust parameters were fixed at a higher value from their reference value. This signifies that we can not enhance the efficiency of p53 to repair DNA damage by recalibrating the p53 module, and hence the system is moving towards cancer. Therefore, in case of p53 loses its DNA damage repair ability, the system demands for other mechanisms to forestall the progression of oncogenic activity.

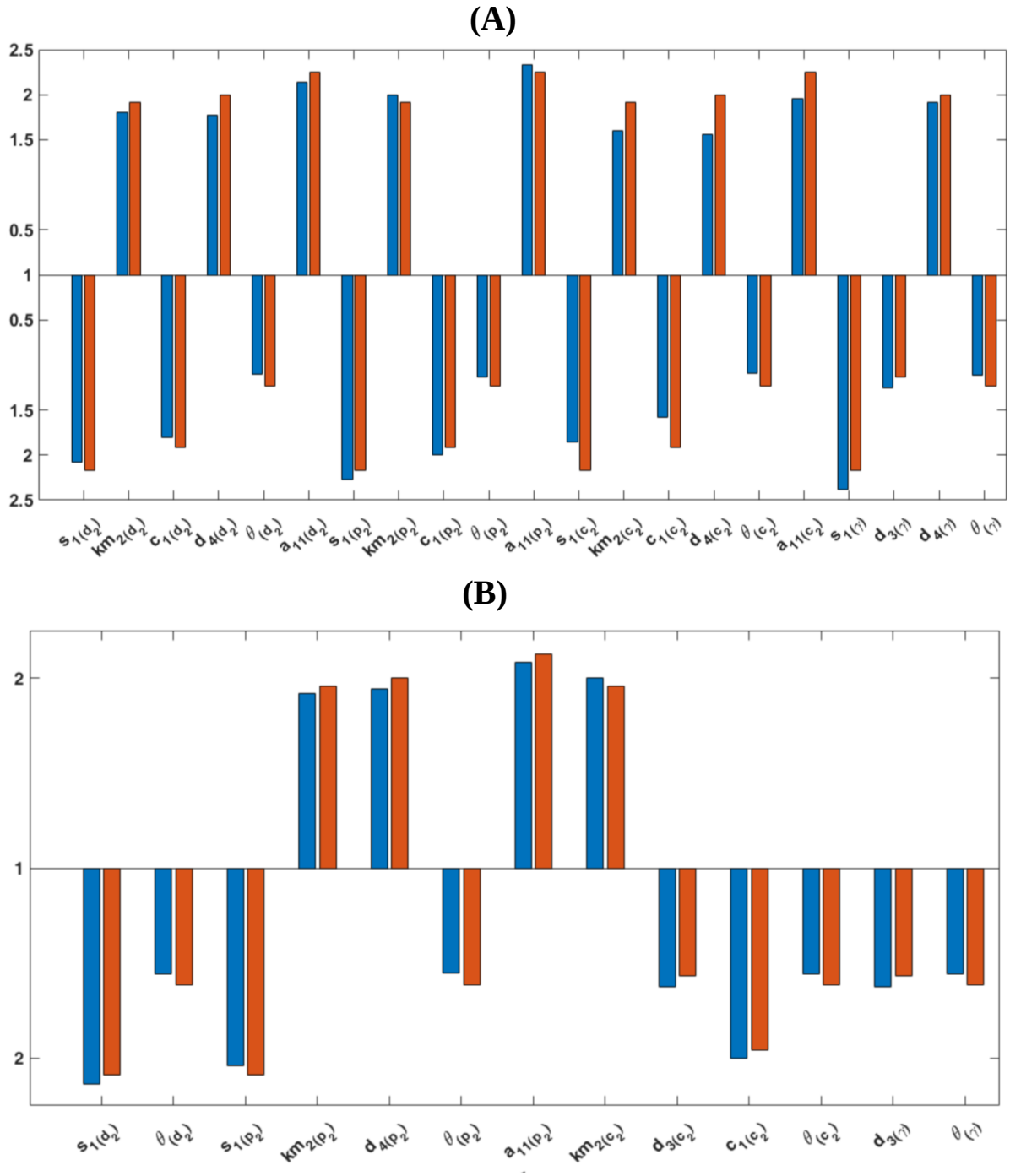


Figure 5.6: **Parameter sensitivity recalibration analysis.** (A) When the robust parameters are fixed at ten fold up, than their reference value **Table 5.1**. The blue colour represents the old sensitivity value, and red represents the new sensitivity value observed due to the change in the robust parameter values. (B) When the robust parameters are fixed at ten-fold down than their reference value **Table 5.1**. The color description is same as **Figure 5.6A.**

5.3.3 Restoration from disease to healthy state: the elucidation of p53-beclin1 analogy

Provided that the recalibration analysis of the p53 module parameters was insufficient to prevent the DNA damage and the oncogenic activity now creates anarchy to the tissue homeostasis, we trace the effect of p53 on the autophagy process via the autophagy inducer beclin1. The objective was to see if the progression of cancer can be mitigated by modulation of the beclin1-induced autophagic cell death when the system goes to the no-repair region.

In cancer, autophagy plays a dual role in the promotion and subjugation of tumours depending on the stage of development and the type of tumour [440]. The loss of beclin1 has been reported to be associated with hepatocarcinoma, lung, ovarian, breast, and testicular cancer [27, 189, 441]. DNA damage induces stress in the cell, to which the autophagy level gets increased. However, tumour cells use autophagy to withstand nutritional deficiencies and hypoxic conditions [442]. That being said, an abnormal rise in the autophagy level will eventually result in cell death. For that purpose, in our model, we have tried to find out possible mechanisms to propel beclin1 beyond the normal region.

5.3.3.1 Local sensitivity analysis of beclin1

We here carried out a local sensitivity analysis on beclin1 at the disease state (i.e., where p53 is abnormally high, and beclin1 is at the lung cancerous region as defined by [438]). We perturbed a parameter over a range of ± 10 fold from its reference value while keeping other parameters fixed. A parameter is called sensitive only if its perturbation can increase beclin1 beyond its normal region. We have found six such parameters and are shown in **Figure 5.7**. The bar length, i.e., the orange portion in the figure, represents the region to which a parameter can hold beclin1 in its normal region. If we vary these parameters within the orange region, beclin1 will remain in the normal region only. However, if a parameter is pushed beyond the orange region, the becin1 level will reach the excess region. For example, if we decrease the parameter d_6 by 6.0714 fold, the corresponding beclin1 level will enter the excess region. Similarly, if we increase the parameter k_1 by 3.98 fold, it will propel beclin1 to the excess region. It is seen from the figure that the association rate of BCL2:beclin1 complex, d_{13}, and the basal

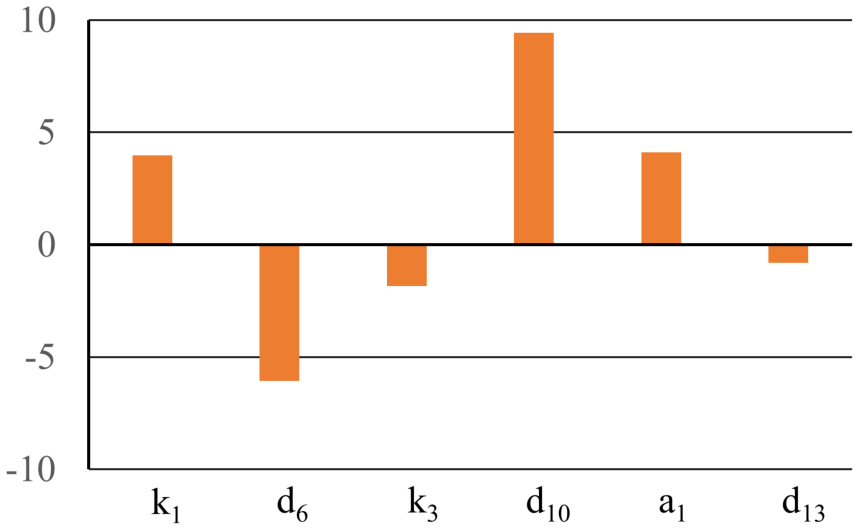


Figure 5.7: **The figure shows the fold change of the parameters at $c = 0.045$ for which beclin1 exceeds the normal level.** The orange portion denotes the region when beclin1 is in the normal region beyond which beclin1 reaches the excess region.

production rate of AMPK, k_3, require less than 2 fold variation from their reference value to increase the beclin1 level. Throughout the analysis, the p53 level was kept at the no-repair region by keeping $c_1 = 0.0450$.

5.3.3.2 Two-dimensional parameter spaces analysis

To further investigate the association of parameters in increasing beclin1 level following DNA damage, we performed a two-dimensional parameter spaces analysis (**Figure 5.8**). Here, we varied two parameters from their reference value while keeping the other parameters fixed. In their study, Liu et al.[438] have described the level of beclin1 in normal and lung cancer tissues. To see the effect of the variation of two parameters on beclin1, we have subdivided the parameter space into four regions. The first region contains those parameter values for which beclin1 remains below the lung cancerous region, the second region contains the parameter values for which beclin1 is in the lung cancer region, the third region contains the values for which beclin1 level exceeds the lung cancerous region but remain within the normal region. Finally, the last region for which beclin1 exceeds the normal region and will pave the way for

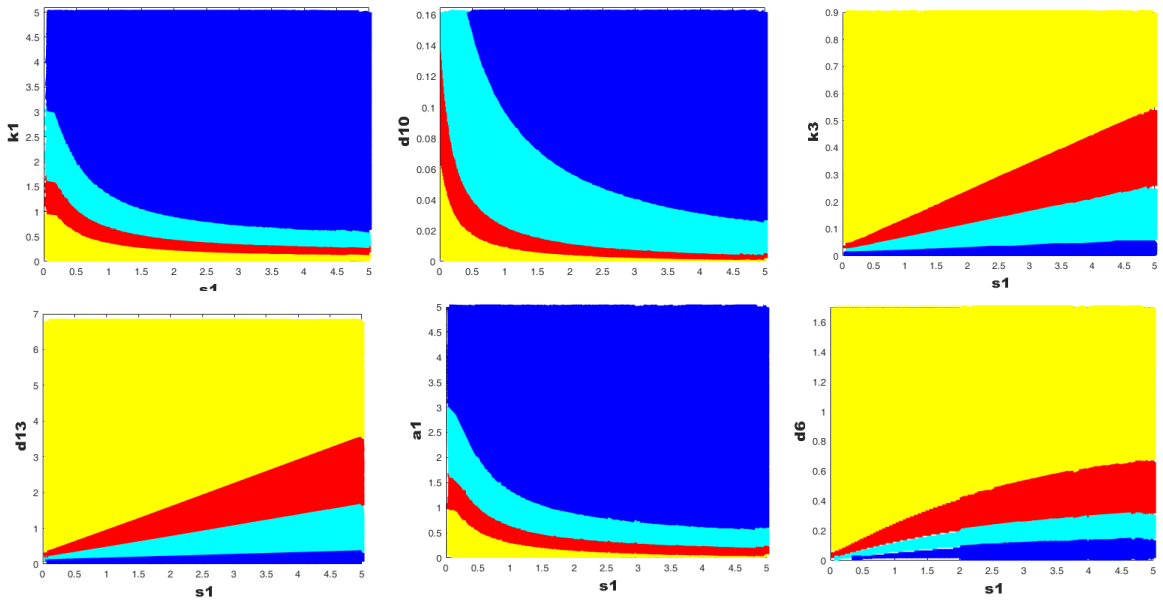


Figure 5.8: **The figure shows the change in steady-state beclin1 level upon variation of two parameters at a time**. Based on this steady-state value of beclin1 obtained after each simulation and the reference value of beclin1 in normal and lung cancer tissues as given in reference [438], we partitioned the region into four layers. According to Liu et al. [438], the red and cyan regions denote, respectively, the beclin1 level in lung cancer ($2.6 \leq beclin1 \leq 4.305$) and normal tissues ($4.306 \leq beclin1 \leq 7.875$). The yellow portion denotes a very low level of beclin1 ($beclin1 \leq 2.5$), while the blue portion indicates that the beclin1 level is more than its normal value ($beclin1 \geq 7.876$). Throughout all the simulations, the p53 level was kept at no-repair region.

autophagic cell death. The basal production of p53, s_1, was varied against all the parameters from **Figure 5.7**. The variation of different parameters produced different effects on the beclin1 dynamics. The variation of basal production of p53 (s_1) with the production of beclin1 by AMPK (a_1) showed the dominance of the latter. A similar effect was seen when s_1 was varied with the basal production rate of AMPK, k_1, and with the basal production of beclin1, k_4. Again, the variation of s_1 with the transcriptional degradation of BCL2 by p53 (d_6) showed that if the basal level of p53 increases, it will dissociate the BCL2:beclin1 interaction to induce autophagy. A similar effect was observed when s_1 is perturbed with the degradation rate of beclin1 by BCL2 (d_{13}). All these effects were captured by keeping p53 at the no-repair region.

5.4 Discussion

One of the leading causes of cancer deaths has been the rising mortality rate due to lung cancer. Given the substantial improvements in diagnosis and service, lung cancer prognosis remains low, as in most cases. It is detected during the later stages, which curb therapeutic options and contribute to poor outcomes. This demonstrates the importance of early diagnosis of lung cancer, leading researchers to identify biomarkers that could be used for early detection and enhance their understanding of lung cancer. In this paper, we have studied the interplay between DNA damage and autophagy in lung cancer and have proposed a mathematical model to identify potential parameters/factors that may be fruitful in restoring the system's healthiness.

DNA damage persists across a variety of endogenous and exogenous factors in all organisms and appears to play a prominent role in many biological processes, which eventually lead to cancer. Amidst the plethora of proteins that take part in the DNA damage response network, we have chosen the guardians of the genome, the most commonly mutated protein in cancer, p53, and its downstream regulators that eventually connect autophagy to the DNA damage response network. The concentration of p53 protein within the cell is tightly regulated by the dynamic duo of MDMX and MDM2. The interactions between the triplet of p53-MDM2-MDMX have been experimentally verified. The oscillatory responses of p53 due to the interaction with MDM2, which get boosted by the presence of MDMX, have been observed in literature [423]. This oscillation refers to the DNA damage, which results in the elevation of p53 levels due to post-translational modification of the p53 polypeptide [443]. This stimulates the DNA damage repair process, which upon completion, is followed by the degradation of p53 by MDM2 and MDMX. The constant DNA damage in the cell elevates the p53 level and, thus, the level of MDM2 and MDMX, and these activation-degradation scenarios give rise to p53 oscillations. Our model successfully portrayed this oscillatory behaviour of p53 (**Figure 5.2**). However, not all DNA damage is repairable. When the DNA gets severely damaged, the cell initiates cellular senescence, which results in an elevation of p53 level [444]. Our model has also addressed this phenomenon, seen by the black line in **Figure 5.2**. Our model has thus addressed and effectively substantiated the p53-DNA damage complex interaction.

Following any DNA damage, the p53 level upsurges, and it induces arrest within the G1

phase of the cell cycle [445]. Bar et al. [433] have stated that the severity of DNA damage can be correlated with the changes in p53 concentration with time. p53 acts as a tumour suppressor. It can repair DNA damage and prevent the path, which eventually leads to various types of diseases, including cancer. The ability of p53 to repair the DNA damage is expressed through the normal oscillations as shown by the green curve in **Figure 5.2**. Again, following DNA damage, the elevation of the p53 level has also been reported in literature [446]. Here, we have addressed the first condition as the normal p53 level and the latter as the no-repair level of p53. The intermediate phase, where the p53 level elevates but eventually returns to normal, indicating a repair of DNA, is considered a repair phase in our study. The appearance of pulses is an elucidation that DNA damages are being repaired in the system. We first search for the important factors that can exacerbate the DNA damage-suppressing capability of p53. This was obtained by implementing local and global sensitivity analysis on the p53 module. The global sensitivity analysis identified crucial parameters that are influential in the regulation of p53 dynamics, which include the degradation of p53 by MDM2 (d_1), half-maximal rate of MDM2 (θ), the stress coefficient α, basal degradation rate of p53 (a_{11}), Michaelis constant in the production of MDM2 via MDMX ($km2$), the degradation rate of MDM2 (d_3) and the production rate of MDMX (c_1). Our findings justify the experimentally observed strong effect of the MDM2 on p53. MDM2 retains an ascendancy in the regulation of p53 rates, which is observed by the presence of θ, d_3, and d_1 as key parameters. Again, due to the threat, they pose towards DNA, following any stresses, the level of p53 increases, which justifies the stress coefficient α as a crucial regulator of p53. The rate of p53 degradation (a_{11}) prevents it from increasing its level. Likewise, the effect of MDMX on p53 can not be overlooked because it directly or indirectly monitors and regulates the level of p53, as evidenced by the emergence of the sensitive parameters c_1 and km_2. The temporal sensitivity analysis (**Figure 5.5**) indicated that the degradation rate of p53 by MDM2 is the most sensitive parameter throughout the time points. It is also seen that the sensitivity index of these parameters varies with time. This may result from the complex interactions between p53, MDM2, and MDMX. From a biological perspective, this variance may be the result of the presence of numerous factors affecting the dynamics of p53 along with MDM2 and MDMX.

Again, an upward shift of the basal production level of p53, s_1, will increase the concentration of p53. However, each parameter's effect, along with s_1, needs to be strictly monitored to control their effect on p53 as it determines the cell fate. A transition from normal to no-repair region will indicate a menace to tissue homeostasis. In contrast, an opposite transition will indicate a possible recovery of the system. Addressing the former, we have used local sensitivity analysis to find out the range for each parameter beyond which the p53 level can shift to the no-repair region from the normal (**Figure 5.3**). Decrease in the stress coefficient, α, degradation of MDM2, d_3, stress-dependent production of MDMX, c_1, and half-maximal rate of Mdm2, θ, while an increase in the degradation rate of MDMX, d_4, was a crucial factor in driving p53 to no-repair region. MDMX ablation will shift p53 beyond its normal state. Literature has reported p53-dependent embryonic death in the absence of MDMX [447]. Since MDM2 acts as the principal antagonist of p53, a decrease in its level will surely uplift the p53 level. Our model suggests that three factors mediate this uplift, directly by increasing MDM2 degradation and decreasing its half-life and indirectly by decreasing MDMX production. Both of these results correlate with literature findings [425, 448].

The normal oscillations of p53 denote that the DNA damage is being repaired. To keep the system in a healthy state, it is, therefore, necessary to make sure that p53 remains in the normal region and its oscillations do not vanish. The aim of the sensitivity analysis was to foresee the existence of any parameters of the p53 module which can boost the efficiency of p53 to repair the DNA damage. Although we derived some robust parameters, there was a set of sensitive parameters capable of abolishing the p53 oscillations. The recalibration analysis was an effort to diminish this effect of sensitive parameters on p53. However, we observed that there is no significant effect on the robustness of the sensitive parameters, briefing the fact that p53 can no longer suppress cancer due to the loss of oscillations. This limitation of the p53 module necessitates the need to find an alternate pathway in case p53 loses its suppression capability, and it is now challenging to reacquire it through the recalibration of the parameters of the p53 module.

Evidence in the literature suggesting the dual role of autophagy in cancer is pervasive. The cancer cells confide in autophagy to compensate for high metabolic demand due to increased

cell proliferation [449]. Autophagy provides nutrition to cancer cells through the lysosomal recycling of intracellular nutrients. To restore homeostasis, this protective effect of autophagy on tumour cells must be annihilated, which can be attained by inducing an excess level of autophagy to pave the way to autophagic cell death. The expression of the key autophagy regulator beclin1 was found to be low in lung cancer [438]. To induce autophagic cell death, its level will have to be increased. The induction of autophagy following DNA damage was found to be DRAM-dependent [450], but to our knowledge, no study has linked beclin1-induced cell death with DNA damage. Here, we have hypothesised that the changes which happen following DNA damage due to lung cancer can be implemented in a beneficial way to get rid of the culprit cells. The global sensitivity analysis resulted in seven sensitive parameters of beclin1. The result obtained highlighted the importance carried by AMPK, p53, and BCL2 on the beclin1 dynamics. The temporal sensitivity analysis showed a decrease in the sensitivity at the last time points of all but one parameter (d_3). From a biological perspective, this variance of sensitivity may result from the presence of numerous pathways that always monitor the levels of beclin1. For example, the association rate of the BCL2:BECLIN1 complex negatively regulates the function of beclin1, and hence the autophagy level gets reduced. However, when the body needs nutrition, various factors like HMGB1 [451], BNIP3 [430] binds to BCL2 or beclin1 and inhibits this interaction. We performed a local sensitivity analysis on beclin1 by retaining p53 in the no-repair region and thus emulating a lung cancerous system, we quest for parameters that could increase beclin1 beyond its normal level that will eventually lead to autophagic cell death. We have identified six such parameters which, if perturbed beyond the range specified, will increase the beclin1 level (**Figure 5.7**). The AMPK dynamics and BCL2 dynamics were fruitful in altering the beclin1 level. The effect of p53 was also vital as both of its gateways to autophagy, i.e., the p53-mediated degradation rates of AMPK and BCL2 were found crucial in enhancing the beclin1 level. We next carried out a 2D-parameter space analysis to determine the effect of variations of parameters in combination to induce autophagic cell death. Since the literature has reported that DNA damage and autophagy go hand in hand, we vary all the six parameters of **Figure 5.7** with the basal production of p53 (**Figure 5.8**). We have observed that the p53 level plays a significant role in the increase of the beclin1 level. Following DNA

damage, the basal production rate of p53 increases as a part of the DNA damage response program. From the two-dimensional parameter space, we can see that, as the basal level of p53 increases, the parameters obtained from **Figure 5.7** required minimum variation from their base value than the previously found ranges to uplift the beclin1 level and hence, are capable of inducing autophagic cell death in earlier phases of lung cancer.

5.5 Conclusion

par In this paper, we have formulated a mathematical model to study p53-mediated autophagy in lung cancer. To acquire a comprehensive understanding of the association between p53 and autophagic cell death, specifically, how it converts the chaos generated by DNA damage into serendipity, was crucially important to decipher. Our study has provided the mechanism to understand this association by identifying the parametric range required to initiate cell death in lung cancer. We have also identified the parameters which are capable of sustaining the DNA damage repair efficiency of p53.

Even if a mathematical model can not grasp everything present in a real system, it can still put forward conditions, parameters, or factors that could possibly control the system. In that sense, we claim that the parametric range we have defined is sufficient to induce autophagic cell death and is capable of mitigating cancer progression. Some of our findings are consistent with the established literature. Taken together, we conclude that our study adds a new dimension to the DNA damage-p53-autophagy interplay in lung cancer in the scenario when the p53 level fails to suppress the DNA damage and sheds light on the possible way to induce an excessive level of beclin1 to pave the way for autophagic cell death.

6

Conclusions and future directions

6.1 Conclusions

The preponderance of cellular components exerts their roles via interactions with other cellular components positioned within the same cell, between cells, or even between organs. With ≈ 25,000 protein-coding genes and an undetermined number of different proteins, the inherent complexity of the emergent network, the PPI network, is intimidating. This inter- and intracellular interconnectedness suggests that the influence of a single genetic anomaly is not limited to the behaviour of the proteins that carry it but can propagate through the entire network and alter the activity of proteins that do not have the abnormality. Determining the phenotypic effects of a disease's characteristics requires a comprehension of these perturbations and, eventually, the complicated interaction network between proteins. The protein perturbations are crucial

in diseases and must be captured and decoded to comprehensively understand the disease progression, which is an unequivocal result.

Autophagy is a quintessential biological process that breaks down unwanted cellular constituents and thus plays a crucial role in maintaining cellular homeostasis. Due to its immense importance in the modulation of cellular fate, the process of autophagy remains at the crossroads of various cellular processes and signalling pathways. The tracking of signals that modulate autophagy, and genes, which have a role in the autophagy process, has encouraged the detection and controlling of the autophagy pathway. Any hindrance to either of them may lead to various diseases. In the last decade, there have been eminent progress in understanding the role of autophagy in different processes, but the underlying mechanism that leads to the observed phenomenon is still far from being captured. This is because autophagy is a very complex process and exhibits different behaviours depending upon the situation. To understand the underlying complexity, comprehensive knowledge of autophagy is important, necessitating the implications of sophisticated methodology to unravel this complexity.

The present thesis aims to investigate protein perturbations in diseases with an emphasis on the autophagy process using network biology and mathematical modelling approaches. As discussed in **Chapter 1**, systems biology has played a crucial role in unrevealing the association between autophagy and other biological processes and diseases by investigating their protein-protein interaction network. It has also helped provide mechanistic insight into the process of autophagy with the help of mathematical modelling approaches. Despite being studied for decades, there exist many diseases in which the perturbations of autophagy-related proteins were not properly investigated. In **Chapter 2**, we used the interplay between autophagy and diabetic retinopathy (DR) to identify novel autophagy-related targets in DR. We addressed this crosstalk through a multilayer relatedness approach. The analysis identified three novel autophagy-related proteins, TP53, PIK3R1, and HSP90AA1, which can modulate the progression and pathogenesis of DR. These proteins were significant in many aspects. Firstly, they were not previously reported as DR-associated. Secondly, they were autophagy-related, i.e., they can modulate the autophagy process, either increasing or decreasing it. Thirdly, they were hubs, high-betweenness, and high-closeness nodes in the autophagy-DR interactome. Fourthly,

in the co-expression analysis using WGCNA, where we investigated gene-gene relatedness, we found that these genes were present in the disease-related modules. Fifthly, when we examined the PPI network of the DEGs of DR in a clinical dataset, we found that some of these proteins are DEGs. Also, a few among them were found to act as meditators, the proteins acting as a bridge to connect the disconnected components. In the co-expression and DEG network, these proteins were also found to be hubs, echoing their topological importance. Finally, we applied the RWR algorithm by taking the known disease proteins as seed and found that PIK3R1 obtained the top, TP53 fifth, while HSP90AA1 received the 25th position in the RWR ranking. This suggests that these proteins remain in close proximity to the known disease protein in DR. To summarise, these proteins are critical to regulate the DR interactome and shed light on previously unknown aspects of the disease.

However, comprehending the global changes in disease and identifying its targets necessitates a full understanding of the entire system's perturbations compared to exploring a subset of the perturbations. To this purpose, in **Chapter 3**, we have applied a de novo methodology to identify potential targets in NASH. Our approach of harnessing and integrating context-specific molecular networks was the first of its genre. Using this, we found three possible targets for NASH and eight for NAFL. Interestingly, we found that among the three targets for NASH, one was associated with autophagy. These potential targets exert their effects at both the gene and metabolic levels and reverse disease-associated gene signatures. We have demonstrated that inhibiting these potential therapeutic targets impacts a number of crucial metabolic processes involved in steatohepatitis, as well as proteins engaged in inflammation and fibrosis development. This proposed methodology lays out a pragmatic framework for identifying potential therapeutic targets with a higher probability of success. It will save tremendous resources and time during the drug discovery process and be used as a general pipeline to identify targets in any *in silico* studies.

An intriguing fact about proteins is that the proteins associated with a disease always remain in close proximity, and only a few of them gets identified as pathogenic. However, the crosstalk between all these proteins governs the development and progression of diseases. There are many diseases in which various potential target proteins are identified in vitro/ in vivo and are

taken to clinical phases. Still, none of them eventually crosses all the clinical trial phases. Therefore, it won't be an embellishment to assert that the existing disease proteins fail to capture the mechanism governing the disease progression. Addressing this and taking NASH as an example, in **Chapter 4**, we have incorporated a methodology, RWRMLA, which first identifies the proteins which remain in close proximity to the disease proteins; among them, it identifies the proteins that not only have the classification capability but also are capable of controlling the disease network and finally investigates the capability of these proteins to alter the metabolic landscape. A protein which passes through all these is termed the potential target of the disease. RWRMLA identified four proteins as potential therapeutic candidates in NASH. Here also, we found that the identified proteins are associated with the autophagy process. However, regardless of the identified target proteins implicated in potential significant disease pathways, their efficacy must be substantiated through follow-up investigations and evaluated through experimental inquiry. RWRMLA is an efficacious generic methodology and can be applied to any metabolic disease to identify potential targets.

In **Chapter 5**, we applied a mathematical modelling approach to get mechanistic insight into some autophagic genes in cell proliferation and cell death. Here, we have formulated a mathematical model to study p53-mediated autophagy in lung cancer. To acquire a comprehensive understanding of the association between p53 and autophagic cell death, specifically, how it converts the chaos generated by DNA damage into serendipity, was crucially important to decipher. Our study has provided the mechanism to understand this association by identifying the parametric range required to initiate cell death in lung cancer. We claimed that the parametric range we have defined is sufficient to induce autophagic cell death and can mitigate cancer progression. Some of our findings are consistent with the established literature. Taken together, we conclude that our study adds a new dimension to the DNA damage-p53-autophagy interplay in lung cancer when the p53 level fails to suppress the DNA damage and sheds light on the possible way to induce an excessive level of beclin1 to pave the way for autophagic cell death. Overall, the work presented in this thesis proposes several novel potential targets in NASH and diabetic retinopathy and establishes beclin1 as a potential therapeutic intervention in lung cancer using a mathematical model. We found the dominance of autophagic proteins in these

diseases and proposed that targeting these proteins will mitigate the disease progression. On the other hand, the methodologies developed in this thesis are general and can be applied to study any disease. In conclusion, the work given here can aid in formulating more effective treatment regimens and the drug development process.

6.2 Future scope

The work presented in the thesis can further be extended from various aspects of modelling and network analysis perspectives. In chapter 5, we have proposed a mathematical model to address the interplay between DNA damage, autophagy, p53, and lung cancer. Here, we showed that increasing beclin1 from its basal level could be a potential therapeutic intervention in lung cancer. The importance of beclin1, and thereby, autophagy, can be analysed using a large-scale clinical dataset of lung cancer and incorporating various network analysis approach. Additionally, using inherent properties, like delay and stochasticity, the complexity of the model can further be increased to get a more realistic portrayal of lung cancer. In chapters 2, 3, and 4, we used network biology approaches to identify potential therapeutic targets. The identified targets can further be evaluated using in-vitro or in-vivo studies to gain insight into their therapeutic efficacy. Moreover, to get an insight into how the interplay between these targets is governing the respective disease dynamics, mathematical modelling-based studies could be incorporated. The process of autophagy has been a prominent focus of research as it is still a puzzle with various missing pieces due to its complex mechanism in numerous biological processes and diseases. From a network analysis perspective, temporal analysis of genes is done in diseases like obesity, but none has been reported in autophagy. Such a study will help to decipher the change in the behaviour of autophagy-related genes with respect to time and identify potential drug targets. Again, using computational drug discovery approaches can also be incorporated to investigate crucial factors like druggability. This study can also be facilitated by the mathematical modelling of the identified targets. Such a pipeline-based study in autophagy is lacking and will surely help to provide fruitful insights into the autophagy process.

References

[1] Noboru Mizushima, Beth Levine, Ana Maria Cuervo, and Daniel J Klionsky. Autophagy fights disease through cellular self-digestion. *Nature*, 451(7182):1069–1075, 2008.

[2] G. M. Whitesides. The origins and the future of microfluidics. *Nature*, 442:368–373, 2006.

[3] Nicholas T Ktistakis. In praise of m. anselmier who first used the term "autophagie" in 1859, 2017.

[4] Sam L Clark Jr. Cellular differentiation in the kidneys of newborn mice studied with the electron microscope. *The Journal of Cell Biology*, 3(3):349–362, 1957.

[5] Alex B Novikoff, H Beaufay, and Christian de Duve. Electron microscopy of lysosome-rich fractions from rat liver. *The Journal of Biophysical and Biochemical Cytology*, 2(4):179, 1956.

[6] Antti U Arstila and Benjamin F Trump. Studies on cellular autophagocytosis. the formation of autophagic vacuoles in the liver after glucagon administration. *The American Journal of Pathology*, 53(5):687, 1968.

[7] Christian De Duve, Anthony VS De Reuck, Margaret P Cameron, et al. *Ciba Foundation Symposium: Lysosomes: Proceedings*. Churchill, 1963.

[8] U Pfeifer and P Strauss. Autophagic vacuoles in heart muscle and liver. a comparative morphometric study including circadian variations in meal-fed rats. *Journal of Molecular and Cellular cardiology*, 13(1):37–49, 1981.

[9] ULRICH Pfeifer and MONIKA Warmuth-Metz. Inhibition by insulin of cellular autophagy in proximal tubular cells of rat kidney. *American Journal of Physiology-Endocrinology and Metabolism*, 244(2):E109–E114, 1983.

[10] Glenn E Mortimore, Nancy J Hutson, and Cynthia A Surmacz. Quantitative correlation between proteolysis and macro-and microautophagy in mouse hepatocytes during starvation and refeeding. *Proceedings of the National Academy of Sciences*, 80(8):2179–2183, 1983.

[11] Russell L Deter, Pierre Baudhuin, and Christian de Duve. Participation of lysosomes in cellular autophagy induced in rat liver by glucagon. *The Journal of Cell biology*, 35(2):C11, 1967.

[12] Per O Seglen and Paul B Gordon. 3-methyladenine: specific inhibitor of autophagic/lysosomal protein degradation in isolated rat hepatocytes. *Proceedings of the National Academy of Sciences*, 79(6):1889–1892, 1982.

[13] J Fred Dice, Carlos D Walker, Betsy Byrne, and Alex Cardiel. General characteristics of protein degradation in diabetes and starvation. *Proceedings of the National Academy of Sciences*, 75(5):2093–2097, 1978.

[14] Kazuhiko Takeshige, Misuzu Baba, Shigeru Tsuboi, Takeshi Noda, and Yoshinori Ohsumi. Autophagy in yeast demonstrated with proteinase-deficient mutants and conditions for its induction. *The Journal of Cell Biology*, 119(2):301–311, 1992.

[15] Miki Tsukada and Yoshinori Ohsumi. Isolation and characterization of autophagy-defective mutants of saccharomyces cerevisiae. *FEBS Letters*, 333(1-2):169–174, 1993.

[16] M Thumm, R Egner, B Koch, M Schlumpberger, M Straub, M Veenhuis, and DH Wolf. Isolation of autophagocytosis mutants of saccharomyces cerevisiae. *FEBS letters*, 349(2):275–280, 1994.

[17] Misuzu Baba, Masako Osumi, Sidney V Scott, Daniel J Klionsky, and Yoshinori Ohsumi. Two distinct pathways for targeting proteins from the cytoplasm to the vacuole/lysosome. *The Journal of Cell Biology*, 139(7):1687–1695, 1997.

[18] Tanya M Harding, Kevin A Morano, Sidney V Scott, and Daniel J Klionsky. Isolation and characterization of yeast mutants in the cytoplasm to vacuole protein targeting pathway. *The Journal of Cell Biology*, 131(3):591–602, 1995.

[19] Hiroyuki Mukaiyama, Masahide Oku, Misuzu Baba, Takeshi Samizo, Adam T Hammond, Benjamin S Glick, Nobuo Kato, and Yasuyoshi Sakai. Paz2 and 13 other paz gene products regulate vacuolar engulfment of peroxisomes during micropexophagy. *Genes to Cells*, 7(1):75–90, 2002.

[20] Yasuyoshi Sakai, Antonius Koller, Linda K Rangell, Gilbert A Keller, and Suresh Subramani. Peroxisome degradation by microautophagy in pichia pastoris: identification of specific steps and morphological intermediates. *The Journal of Cell Biology*, 141(3):625–636, 1998.

[21] Weiping Yuan, Daniel L Tuttle, Yu-Jiang Shi, Gareth S Ralph, and WA Dunn. Glucose-induced microautophagy in pichia pastoris requires the alpha-subunit of phosphofructokinase. *Journal of Cell Science*, 110(16):1935–1945, 1997.

[22] Vladimir I Titorenko, Ineke Keizer, Wim Harder, and Marten Veenhuis. Isolation and characterization of mutants impaired in the selective degradation of peroxisomes in the yeast hansenula polymorpha. *Journal of Bacteriology*, 177(2):357–363, 1995.

[23] Daniel J Klionsky, James M Cregg, William A Dunn, Scott D Emr, Yasuyoshi Sakai, Ignacio V Sandoval, Andrei Sibirny, Suresh Subramani, Michael Thumm, Marten Veenhuis, et al. A unified nomenclature for yeast autophagy-related genes. *Developmental Cell*, 5(4):539–545, 2003.

[24] Edward FC Blommaart, Joost JFP Luiken, Pietjan JE Blommaart, George M van Woerkom, and Alfred J Meijer. Phosphorylation of ribosomal protein s6 is inhibitory for autophagy in isolated rat hepatocytes. *Journal of Biological Chemistry*, 270(5):2320–2326, 1995.

[25] Noboru Mizushima, Takeshi Noda, Tamotsu Yoshimori, Yae Tanaka, Tomoko Ishii, Michael D George, Daniel J Klionsky, Mariko Ohsumi, and Yoshinori Ohsumi. A protein conjugation system essential for autophagy. *Nature*, 395(6700):395–398, 1998.

[26] Yoshinobu Ichimura, Takayoshi Kirisako, Toshifumi Takao, Yoshinori Satomi, Yasutsugu Shimonishi, Naotada Ishihara, Noboru Mizushima, Isei Tanida, Eiki Kominami, Mariko Ohsumi, et al. A ubiquitin-like system mediates protein lipidation. *Nature*, 408(6811):488–492, 2000.

[27] Xiao Huan Liang, Saadiya Jackson, Matthew Seaman, Kristy Brown, Bettina Kempkes, Hanina Hibshoosh, and Beth Levine. Induction of autophagy and inhibition of tumorigenesis by beclin 1. *Nature*, 402(6762):672–676, 1999.

[28] Jochen Hampe, Andre Franke, Philip Rosenstiel, Andreas Till, Markus Teuber, Klaus Huse, Mario Albrecht, Gabriele Mayr, Francisco M De La Vega, Jason Briggs, et al. A genome-wide association scan of nonsynonymous snps identifies a susceptibility variant for crohn disease in atg16l1. *Nature Genetics*, 39(2):207–211, 2007.

[29] Alicia Meléndez, Zsolt Tallóczy, Matthew Seaman, Eeva-Liisa Eskelinen, David H Hall, and Beth Levine. Autophagy genes are essential for dauer development and life-span extension in c. elegans. *Science*, 301(5638):1387–1391, 2003.

[30] Brinda Ravikumar, Rainer Duden, and David C Rubinsztein. Aggregate-prone proteins with polyglutamine and polyalanine expansions are degraded by autophagy. *Human Molecular Genetics*, 11(9):1107–1117, 2002.

[31] Akio Kihara, Takeshi Noda, Naotada Ishihara, and Yoshinori Ohsumi. Two distinct vps34 phosphatidylinositol 3–kinase complexes function in autophagy and carboxypeptidase y sorting in saccharomyces cerevisiae. *The Journal of Cell biology*, 152(3):519–530, 2001.

[32] Carmine Settembre, Chiara Di Malta, Vinicia Assunta Polito, Moises Garcia Arencibia, Francesco Vetrini, Serkan Erdin, Serpil Uckac Erdin, Tuong Huynh, Diego Medina, Pasqualina Colella, et al. Tfeb links autophagy to lysosomal biogenesis. *Science*, 332(6036):1429–1433, 2011.

[33] Robbie Loewith, Estela Jacinto, Stephan Wullschleger, Anja Lorberg, José L Crespo, Débora Bonenfant, Wolfgang Oppliger, Paul Jenoe, and Michael N Hall. Two tor complexes, only one of which is rapamycin sensitive, have distinct roles in cell growth control. *Molecular Cell*, 10(3):457–468, 2002.

[34] Cristina Mammucari, Giulia Milan, Vanina Romanello, Eva Masiero, Ruediger Rudolf, Paola Del Piccolo, Steven J Burden, Raffaella Di Lisi, Claudia Sandri, Jinghui Zhao, et al. Foxo3 controls autophagy in skeletal muscle in vivo. *Cell Metabolism*, 6(6):458–471, 2007.

[35] Jinghui Zhao, Jeffrey J Brault, Andreas Schild, Peirang Cao, Marco Sandri, Stefano Schiaffino, Stewart H Lecker, and Alfred L Goldberg. Foxo3 coordinately activates protein degradation by the autophagic/lysosomal and proteasomal pathways in atrophying muscle cells. *Cell Metabolism*, 6(6):472–483, 2007.

[36] Rosalie E Lawrence and Roberto Zoncu. The lysosome as a cellular centre for signalling, metabolism and quality control. *Nature Cell Biology*, 21(2):133–142, 2019.

[37] Li Yu, Christina K McPhee, Lixin Zheng, Gonzalo A Mardones, Yueguang Rong, Junya Peng, Na Mi, Ying Zhao, Zhihua Liu, Fengyi Wan, et al. Termination of autophagy and reformation of lysosomes regulated by mtor. *Nature*, 465(7300):942–946, 2010.

[38] Joungmok Kim, Mondira Kundu, Benoit Viollet, and Kun-Liang Guan. Ampk and mtor regulate autophagy through direct phosphorylation of ulk1. *Nature Cell Biology*, 13(2):132–141, 2011.

[39] Maria Zachari and Ian G Ganley. The mammalian ulk1 complex and autophagy initiation. *Essays in Biochemistry*, 61(6):585–596, 2017.

[40] GM Fimia, A Stoykova, A Romagnoli, L Giunta, and S Di Bartolomeo. Nardacci 518 r, corazzari m, fuoco c, ucar a, schwartz p, gruss p, piacentini m, 519 chowdhury k, cecconi f. 2007. ambra1 regulates autophagy and development 520 of the nervous system. *Nature*, 447:1121–5.

[41] Eisuke Itakura, Chieko Kishi-Itakura, and Noboru Mizushima. The hairpin-type tail-anchored snare syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell*, 151(6):1256–1269, 2012.

[42] Zheng Wang, Guangyan Miao, Xue Xue, Xiangyang Guo, Chongzhen Yuan, Zhaoyu Wang, Gangming Zhang, Yingyu Chen, Du Feng, Junjie Hu, et al. The vici syndrome protein epg5 is a rab7 effector that determines the fusion specificity of autophagosomes with late endosomes/lysosomes. *Molecular Cell*, 63(5):781–795, 2016.

[43] Divya Khatter, Vivek B Raina, Devashish Dwivedi, Aastha Sindhwani, Surbhi Bahl, and Mahak Sharma. The small gtpase arl8b regulates assembly of the mammalian hops complex on lysosomes. *Journal of Cell Science*, 128(9):1746–1761, 2015.

[44] Peidu Jiang, Taki Nishimura, Yuriko Sakamaki, Eisuke Itakura, Tomohisa Hatta, Tohru Natsume, and Noboru Mizushima. The hops complex mediates autophagosome–lysosome fusion through interaction with syntaxin 17. *Molecular Biology of the Cell*, 25(8):1327–1337, 2014.

[45] Bernd A Schröder, Christian Wrocklage, Andrej Hasilik, and Paul Saftig. The proteome of lysosomes. *Proteomics*, 10(22):4053–4076, 2010.

[46] Willa Wen-You Yim, Yoshitaka Kurikawa, and Noboru Mizushima. An exploratory text analysis of the autophagy research field. *Autophagy*, 18(7):1648–1661, 2022.

[47] Shanshan He, Zhen Zhao, Yongfei Yang, Douglas O'Connell, and Xiaowei Zhang. Truncating mutation in the autophagy gene truncating mutation in the autophagy gene uvrag confers oncogenic properties and chemosensitivity in colorectal cancers. 2015.

[48] Yoshinori Takahashi, Domenico Coppola, Norimasa Matsushita, Hernani D Cualing, Mei Sun, Yuya Sato, Chengyu Liang, Jae U Jung, Jin Q Cheng, James J Mul, et al. Bif-1 interacts with beclin 1 through uvrag and regulates autophagy and tumorigenesis. *Nature Cell Biology*, 9(10):1142–1151, 2007.

[49] Min Sung Kim, Eun Goo Jeong, Chang Hyeok Ahn, Sung Soo Kim, Sug Hyung Lee, and Nam Jin Yoo. Frameshift mutation of uvrag, an autophagy-related gene, in gastric carcinomas with microsatellite instability. *Human Pathology*, 39(7):1059–1063, 2008.

[50] Dipanka Tanu Sarmah, Nandadulal Bairagi, and Samrat Chatterjee. The interplay between dna damage and autophagy in lung cancer: A mathematical study. *Biosystems*, 206:104443, 2021.

[51] Tao Luo, Jing Fu, An Xu, Bo Su, Yibing Ren, Ning Li, Junjie Zhu, Xiaofang Zhao, Rongyang Dai, Jie Cao, et al. Psmd10/gankyrin induces autophagy to promote tumor progression through cytoplasmic interaction with atg7 and nuclear transactivation of atg7 expression. *Autophagy*, 12(8):1355–1371, 2016.

[52] Mengjie Liu, Lili Jiang, Xiao Fu, Wenjuan Wang, Jiequn Ma, Tao Tian, Kejun Nan, and Xuan Liang. Cytoplasmic liver kinase b1 promotes the growth of human lung adenocarcinoma by enhancing autophagy. *Cancer science*, 109(10):3055–3067, 2018.

[53] Yuki Kawano and David E Cohen. Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease. *Journal of Gastroenterology*, 48(4):434–441, 2013.

[54] Mark J Czaja. Function of autophagy in nonalcoholic fatty liver disease. *Digestive Diseases and Sciences*, 61(5):1304–1313, 2016.

[55] Aaron Ciechanover and Yong Tae Kwon. Degradation of misfolded proteins in neurodegenerative diseases: therapeutic targets and strategies. *Experimental & Molecular medicine*, 47(3):e147–e147, 2015.

[56] Chie Ebato, Toyoyoshi Uchida, Masayuki Arakawa, Masaaki Komatsu, Takashi Ueno, Koji Komiya, Kosuke Azuma, Takahisa Hirose, Keiji Tanaka, Eiki Kominami, et al. Autophagy is important in islet homeostasis and compensatory increase of beta cell mass in response to high-fat diet. *Cell Metabolism*, 8(4):325–332, 2008.

[57] Yael Riahi, Jakob D Wikstrom, Etty Bachar-Wikstrom, Nava Polin, Hava Zucker, Myung-Shik Lee, Wenying Quan, Leena Haataja, Ming Liu, Peter Arvan, et al. Autophagy is a major regulator of beta cell insulin homeostasis. *Diabetologia*, 59(7):1480–1491, 2016.

[58] Jai-Sing Yang, Chi-Cheng Lu, Sheng-Chu Kuo, Yuan-Man Hsu, Shih-Chang Tsai, Shih-Yin Chen, Yng-Tay Chen, Ying-Ju Lin, Yu-Chuen Huang, Chao-Jung Chen, et al. Autophagy and its link to type ii diabetes mellitus. *Biomedicine*, 7(2), 2017.

[59] Yingling Jiang, Wei Huang, Jing Wang, Zhipeng Xu, Jieyu He, Xiaohong Lin, Zhiguang Zhou, and Jingjing Zhang. Metformin plays a dual role in min6 pancreatic β cell function through ampk-dependent autophagy. *International Journal of Biological Sciences*, 10(3):268, 2014.

[60] Ettore Bergamini. Autophagy: a cell repair mechanism that retards ageing and age-associated diseases and can be intensified pharmacologically. *Molecular Aspects of Medicine*, 27(5-6):403–410, 2006.

[61] Beth Levine and Daniel J Klionsky. Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Developmental cell*, 6(4):463–477, 2004.

[62] Shipeng Dang, Zhi-ming Yu, Chang-ying Zhang, Jie Zheng, Ku-lin Li, Ying Wu, Ling-ling Qian, Zhen-yu Yang, Xiao-rong Li, Yanyun Zhang, et al. Autophagy promotes apoptosis of mesenchymal stem cells under inflammatory microenvironment. *Stem Cell Research & Therapy*, 6(1):1–9, 2015.

[63] Mutian Zhang, Ling Su, Zhenna Xiao, Xianfang Liu, and Xiangguo Liu. Methyl jasmonate induces apoptosis and pro-apoptotic autophagy via the ros pathway in human non-small cell lung cancer. *American Journal of Cancer Research*, 6(2):187, 2016.

[64] Donna Denton, Bhupendra Shravage, Rachel Simin, Kathryn Mills, Deborah L Berry, Eric H Baehrecke, and Sharad Kumar. Autophagy, not apoptosis, is essential for midgut cell death in drosophila. *Current Biology*, 19(20):1741–1746, 2009.

[65] Ze-fang Chen, Yan-bo Li, Jun-yong Han, Jing Wang, Jia-jing Yin, Jing-bo Li, and Hui Tian. The double-edged effect of autophagy in pancreatic beta cells and diabetes. *Autophagy*, 7(1):12–16, 2011.

[66] Dipanka Tanu Sarmah, Nandadulal Bairagi, and Samrat Chatterjee. Tracing the footsteps of autophagy in computational biology. *Briefings in Bioinformatics*, 2020.

[67] Katherine L Cook, Ayesha N Shajahan, and Robert Clarke. Autophagy and endocrine resistance in breast cancer. *Expert review of anticancer therapy*, 11(8):1283–1294, 2011.

[68] Kirti Jain, Krishna S Paranandi, Savitha Sridharan, and Alakananda Basu. Autophagy in breast cancer and its implications for therapy. *American Journal of Cancer Research*, 3(3):251, 2013.

[69] Ya-ping Yang, Li-fang Hu, Hui-fen Zheng, Cheng-jie Mao, Wei-dong Hu, Kang-ping Xiong, Fen Wang, and Chun-feng Liu. Application and interpretation of current autophagy inhibitors and activators. *Acta Pharmacologica Sinica*, 34(5):625–635, 2013.

[70] X Sui, R Chen, Z Wang, Z Huang, N Kong, M Zhang, W Han, F Lou, J Yang, Q Zhang, et al. Autophagy and chemotherapy resistance: a promising therapeutic target for cancer treatment. *Cell death & disease*, 4(10):e838–e838, 2013.

[71] Hiroaki Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.

[72] Hiroaki Kitano. Computational systems biology. *Nature*, 420(6912):206–210, 2002.

[73] MD Mesarovic. General systems theory and biology–view of a theoretician. *General Systems Theory and Biology. Springer*, 1968.

[74] D Mesarovic, Mihajlo, SN Sreenath, and JD Keene. Search for organising principles: understanding in systems biology. *Systems biology*, 1(1):19–27, 2004.

[75] Nagasuma Chandra, Dhiraj Kumar, and Kanury Rao. Systems biology of tuberculosis. *Tuberculosis*, 91(5):487–496, 2011.

[76] Xiwen Xiong, Rongya Tao, Ronald A DePinho, and X Charlie Dong. Deletion of hepatic foxo1/3/4 genes in mice significantly impacts on glucose metabolism through downregulation of gluconeogenesis and upregulation of glycolysis. *PloS one*, 8(8):e74340, 2013.

[77] Eric Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl_3):7280–7287, 2002.

[78] Gernot Schaller and Michael Meyer-Hermann. Multicellular tumor spheroid in an off-lattice voronoi-delaunay cell model. *Physical Review E*, 71(5):051910, 2005.

[79] Jesse Engelberg, Suman Ganguli, and C Anthony Hunt. Agent-based simulations of in vitro multicellular tumor spheroid growth. *SIMULATION SERIES*, 38(2):141, 2006.

[80] Carl Adam Petri. Kommunikation mit automaten. 1962.

[81] Hiroshi Matsuno, Yukiko Tanaka, Hitoshi Aoshima, Mika Matsui, Satoru Miyano, et al. Biopathways representation and simulation on hybrid functional petri net. *In silico biology*, 3(3):389–404, 2003.

[82] Peter JE Goss and Jean Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Sciences*, 95(12):6750–6755, 1998.

[83] Ina Koch, Björn H Junker, and Monika Heiner. Application of petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–1226, 2005.

[84] Andrea Sackmann, Monika Heiner, and Ina Koch. Application of petri net based analysis techniques to signal transduction pathways. *BMC bioinformatics*, 7(1):1–17, 2006.

[85] Pawan Dhar, Tan Chee Meng, Sandeep Somani, Li Ye, Anand Sairam, Mandar Chitre, Zhu Hao, and Kishore Sakharkar. Cellware—a multi-algorithmic software for computational systems biology. *Bioinformatics*, 20(8):1319–1321, 2004.

[86] Robert W Smith, Rik P van Rosmalen, Vitor AP Martins dos Santos, and Christian Fleck. Dmpy: a python package for automated mathematical model construction of large-scale metabolic systems. *BMC Systems Biology*, 12(1):1–16, 2018.

[87] Carlos F Lopez, Jeremy L Muhlich, John A Bachman, and Peter K Sorger. Programming biological models in python using pysb. *Molecular Systems Biology*, 9(1):646, 2013.

[88] Henning Schmidt and Mats Jirstrand. Systems biology toolbox for matlab: a computational platform for research in systems biology. *Bioinformatics*, 22(4):514–515, 2006.

[89] Lingchong You, Apirak Hoonlor, and John Yin. Modeling biological systems using dynetica—a simulator of dynamic networks. *Bioinformatics*, 19(3):435–436, 2003.

[90] Pedro Mendes. Gepasi: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Bioinformatics*, 9(5):563–571, 1993.

[91] Thierry Emonet, Charles M Macal, Michael J North, Charles E Wickersham, and Philippe Cluzel. Agentcell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics*, 21(11):2714–2721, 2005.

[92] Akira Funahashi, Yukiko Matsuoka, Akiya Jouraku, Mineo Morohashi, Norihiro Kikuchi, and Hiroaki Kitano. Celldesigner 3.5: a versatile modeling tool for biochemical networks. *Proceedings of the IEEE*, 96(8):1254–1265, 2008.

[93] Autumn A Cuellar, Catherine M Lloyd, Poul F Nielsen, David P Bullivant, David P Nickerson, and Peter J Hunter. An overview of cellml 1.1, a biological model description language. *Simulation*, 79(12):740–747, 2003.

[94] Stephen Wiggins, Stephen Wiggins, and Martin Golubitsky. *Introduction to applied nonlinear dynamical systems and chaos*, volume 2. Springer, 2003.

[95] Simeone Marino, Ian B Hogue, Christian J Ray, and Denise E Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of theoretical biology*, 254(1):178–196, 2008.

[96] Andrea Saltelli and Paola Annoni. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517, 2010.

[97] Liang Yu, Jianbin Huang, Zhixin Ma, Jing Zhang, Yapeng Zou, and Lin Gao. Inferring drug-disease associations based on known protein complexes. *BMC medical genomics*, 8(2):1–13, 2015.

[98] Rui-Sheng Wang, Shihua Zhang, Xiang-Sun Zhang, and Luonan Chen. Identifying modules in complex networks by a graph-theoretical method and its application in protein interaction networks. In *International Conference on Intelligent Computing*, pages 1090–1101. Springer, 2007.

[99] Michel Habib, Lalla Mouatadid, Eric Sopena, and Mengchuan Zou. ({\alpha},{\beta})-modules in graphs. *arXiv preprint arXiv:2101.08881*, 2021.

[100] Yang-Yu Liu, Jean-Jacques Slotine, and Albert-László Barabási. Controllability of complex networks. *Nature*, 473(7346):167–173, 2011.

[101] Rajat Anand, Dipanka Tanu Sarmah, and Samrat Chatterjee. Extracting proteins involved in disease progression using temporally connected networks. *BMC Systems Biology*, 12(1):1–13, 2018.

[102] Arunachalam Vinayagam, Travis E Gibson, Ho-Joon Lee, Bahar Yilmazel, Charles Roesel, Yanhui Hu, Young Kwon, Amitabh Sharma, Yang-Yu Liu, Norbert Perrimon, et al. Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets. *Proceedings of the National Academy of Sciences*, 113(18):4976–4981, 2016.

[103] Rudolf Emil Kalman. Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(2):152–192, 1963.

[104] Jose C Nacher and Tatsuya Akutsu. Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control. *New Journal of Physics*, 14(7):073005, 2012.

[105] Rohan Khazanchi, Kathryn Dempsey, Ishwor Thapa, and Hesham Ali. On identifying and analyzing significant nodes in protein-protein interaction networks. In *2013 IEEE 13th International Conference on Data Mining Workshops*, pages 343–348. IEEE, 2013.

[106] Rahul Badhwar and Ganesh Bagler. Control of neuronal network in caenorhabditis elegans. *PloS one*, 10(9):e0139204, 2015.

[107] Lin Wu, Yichao Shen, Min Li, and Fang-Xiang Wu. Network output controllability-based method for drug target identification. *IEEE Transactions on Nanobioscience*, 14(2):184–191, 2015.

[108] Sergey N Dorogovtsev, Alexander V Goltsev, and José FF Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.

[109] Antonios Garas, Panos Argyrakis, Céline Rozenblat, Marco Tomassini, and Shlomo Havlin. Worldwide spreading of economic crisis. *New Journal of Physics*, 12(11):113043, 2010.

[110] Javier Borge-Holthoefer and Yamir Moreno. Absence of influential spreaders in rumor dynamics. *Physical Review E*, 85(2):026116, 2012.

[111] Jian-Xiong Zhang, Duan-Bing Chen, Qiang Dong, and Zhi-Dan Zhao. Identifying a set of influential spreaders in complex networks. *Scientific Reports*, 6(1):1–10, 2016.

[112] Alfredo Braunstein, Luca Dall'Asta, Guilhem Semerjian, and Lenka Zdeborová. Network dismantling. *Proceedings of the National Academy of Sciences*, 113(44):12368–12373, 2016.

[113] Nikos Bikakis, John Liagouris, Maria Krommyda, George Papastefanatos, and Timos Sellis. Graphvizdb: A scalable platform for interactive large graph visualization. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 1342–1345. IEEE, 2016.

[114] Jacob Köhler, Jan Baumbach, Jan Taubert, Michael Specht, Andre Skusa, Alexander Rüegg, Chris Rawlings, Paul Verrier, and Stephan Philippi. Graph-based analysis and visualization of experimental results with ondex. *Bioinformatics*, 22(11):1383–1390, 2006.

[115] Florian Iragne, Macha Nikolski, Bertrand Mathieu, David Auber, and David Sherman. Proviz: protein interaction visualization and exploration. *Bioinformatics*, 21(2):272–274, 2005.

[116] Zhenjun Hu, Jui-Hung Hung, Yan Wang, Yi-Chien Chang, Chia-Ling Huang, Matt Huyck, and Charles DeLisi. Visant 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Research*, 37(suppl_2):W115–W121, 2009.

[117] Emek Demir, Ozgun Babur, U Dogrusoz, A Gursoy, Gurkan Nisanci, Renguel Cetin-Atalay, and Mehmet Ozturk. Patika: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18(7):996–1003, 2002.

[118] Georgios A Pavlopoulos, Sean D Hooper, Alejandro Sifrim, Reinhard Schneider, and Jan Aerts. Medusa: A tool for exploring and clustering biological networks. *BMC Research Notes*, 4(1):1–6, 2011.

[119] Maria Secrier, Georgios A Pavlopoulos, Jan Aerts, and Reinhard Schneider. Arena3d: visualizing time-driven phenotypic differences in biological systems. *BMC Bioinformatics*, 13(1):1–11, 2012.

[120] Bobby-Joe Breitkreutz, Chris Stark, and Mike Tyers. Osprey: a network visualization system. *Genome biology*, 4(3):1–4, 2003.

[121] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

[122] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Web and Social media*, volume 3, pages 361–362, 2009.

[123] David Auber. Tulip—a huge graph visualization framework. In *Graph Drawing Software*, pages 105–126. Springer, 2004.

[124] Andrej Mrvar and Vladimir Batagelj. Analysis and visualization of large networks with program package pajek. *Complex Adaptive Systems Modeling*, 4(1):1–8, 2016.

[125] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2022. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr.

[126] David J Kahle and Hadley Wickham. ggmap: spatial visualization with ggplot2. *R J.*, 5(1):144, 2013.

[127] Robert C Gentleman, Vincent J Carey, Douglas M Bates, B Bolstad, M Dettling, S Dudoit, B Ellis, L Gautier, Y Ge, and J Gentry. others (2004). bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80.

[128] Bernd Bischl, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary M Jones. mlr: Machine learning in r. *The Journal of Machine Learning Research*, 17(1):5938–5942, 2016.

[129] Robert Gentleman, Vincent J Carey, Wolfgang Huber, Rafael A Irizarry, Sandrine Dudoit, et al. *Bioinformatics and computational biology solutions using R and Bioconductor*, volume 1. Springer, 2005.

[130] Peter Langfelder and Steve Horvath. Wgcna: an r package for weighted correlation network analysis. *BMC bioinformatics*, 9(1):1–13, 2008.

[131] Steffen Durinck, Paul T Spellman, Ewan Birney, and Wolfgang Huber. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols*, 4(8):1184–1191, 2009.

[132] MI Love, W Huber, and S Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2 (2014) genome biol., 15. *CrossRef][PubMed]*, page 550.

[133] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[134] Edward Y Chen, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark, and Avi Ma'ayan. Enrichr: interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(1):1–14, 2013.

[135] Yang Xu and Xiao-Chun Luo. Pypathway: python package for biological network analysis and visualization. *Journal of Computational Biology*, 25(5):499–504, 2018.

[136] Yingyao Zhou, Bin Zhou, Lars Pache, Max Chang, Alireza Hadj Khodabakhshi, Olga Tanaseichuk, Christopher Benner, and Sumit K Chanda. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications*, 10(1):1–10, 2019.

[137] Leif Väremo, Jens Nielsen, and Intawat Nookaew. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research*, 41(8):4378–4391, 2013.

[138] Bruno M Tesson, Rainer Breitling, and Ritsert C Jansen. Diffcoex: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics*, 11(1):1–9, 2010.

[139] Yuqing Zhang, Giovanni Parmigiani, and W Evan Johnson. Combat-seq: batch effect adjustment for rna-seq count data. *NAR genomics and bioinformatics*, 2(3):lqaa078, 2020.

[140] Vinícius Carvalho Jardim, Suzana de Siqueira Santos, Andre Fujita, and Marcos Silveira Buckeridge. Bionetstat: A tool for biological networks differential analysis. *Frontiers in genetics*, 10:594, 2019.

[141] Mahdi Jalili, Ali Salehzadeh-Yazdi, Yazdan Asgari, Seyed Shahriar Arab, Marjan Yaghmaie, Ardeshir Ghavamzadeh, and Kamran Alimoghaddam. Centiserver: a comprehensive resource, web-based application and r package for centrality analysis. *PloS one*, 10(11):e0143111, 2015.

[142] Da Wei Huang, Brad T Sherman, Qina Tan, Jack R Collins, W Gregory Alvord, Jean Roayaei, Robert Stephens, Michael W Baseler, H Clifford Lane, and Richard A Lempicki. The david gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome biology*, 8(9):1–16, 2007.

[143] Aravind Subramanian, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, Scott L Pomeroy, Todd R Golub, Eric S Lander, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005.

[144] Steven Maere, Karel Heymans, and Martin Kuiper. Bingo: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–3449, 2005.

[145] Chao Zhang, Kristina Hanspers, Allan Kuchinsky, Nathan Salomonis, Dong Xu, and Alexander R Pico. Mosaic: making biological sense of complex networks. *Bioinformatics*, 28(14):1943–1944, 2012.

[146] Aaron Barsky, Jennifer L Gardy, Robert EW Hancock, and Tamara Munzner. Cerebral: a cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics*, 23(8):1040–1042, 2007.

[147] Daniel P Gil, Jeffrey N Law, and TM Murali. The pathlinker app: connect the dots in protein interaction networks. *F1000Research*, 6, 2017.

[148] Yu Tang, Min Li, Jianxin Wang, Yi Pan, and Fang-Xiang Wu. Cytonca: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems*, 127:67–72, 2015.

[149] Gabriela Bindea, Bernhard Mlecnik, Hubert Hackl, Pornpimol Charoentong, Marie Tosolini, Amos Kirilovsky, Wolf-Herman Fridman, Franck Pagès, Zlatko Trajanoski, and Jérôme Galon. Cluego: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8):1091–1093, 2009.

[150] Jason Montojo, Khalid Zuberi, Harold Rodriguez, Farzana Kazi, George Wright, Sylva L Donaldson, Quaid Morris, and Gary D Bader. Genemania cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics*, 26(22):2927–2928, 2010.

[151] Andrei Zinovyev, Eric Viara, Laurence Calzone, and Emmanuel Barillot. Binom: a cytoscape plugin for manipulating and analyzing biological networks. *Bioinformatics*, 24(6):876–877, 2008.

[152] Michael Smoot, Keiichiro Ono, Trey Ideker, and Steven Maere. Pingo: a cytoscape plugin to find candidate genes in biological networks. *Bioinformatics*, 27(7):1030–1031, 2011.

[153] Gary D Bader and Christopher WV Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, 4(1):1–27, 2003.

[154] Atanas Kamburov, Konstantin Pentchev, Hanna Galicka, Christoph Wierling, Hans Lehrach, and Ralf Herwig. Consensuspathdb: toward a more complete picture of cell biology. *Nucleic Acids Research*, 39(suppl_1):D712–D717, 2011.

[155] Aditya Vailaya, Peter Bluvas, Robert Kincaid, Allan Kuchinsky, Michael Creech, and Annette Adler. An architecture for biological information extraction and representation. In *Proceedings of the 2004 ACM Symposium on Applied Computing*, pages 103–110, 2004.

[156] Trey Ideker, Owen Ozier, Benno Schwikowski, and Andrew F Siegel. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18(suppl_1):S233–S240, 2002.

[157] Chia-Hao Chin, Shu-Hwa Chen, Hsin-Hung Wu, Chin-Wen Ho, Ming-Tat Ko, and Chung-Yen Lin. cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, 8(4):1–7, 2014.

[158] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome biology*, 18(1):1–15, 2017.

[159] Masato Akiyama. Multi-omics study for interpretation of genome-wide association study. *Journal of Human Genetics*, 66(1):3–10, 2021.

[160] Ryo Yamada, Daigo Okada, Juan Wang, Tapati Basak, and Satoshi Koyama. Interpretation of omics data analyses. *Journal of human genetics*, 66(1):93–102, 2021.

[161] Jeroen Maertzdorf, Gayle McEwen, January Weiner 3rd, Song Tian, Eric Lader, Ulrich Schriek, Harriet Mayanja-Kizza, Martin Ota, John Kenneth, and Stefan HE Kaufmann. Concise gene signature for point-of-care classification of tuberculosis. *EMBO molecular medicine*, 8(2):86–95, 2016.

[162] Lilly M Verhagen, Aldert Zomer, Mailis Maes, Julian A Villalba, Berenice Del Nogal, Marc Eleveld, Sacha AFT van Hijum, Jacobus H de Waard, and Peter WM Hermans. A predictive signature gene set for discriminating active from latent tuberculosis in warao amerindian children. *BMC genomics*, 14(1):1–11, 2013.

[163] Nicholas D Walter, Mikaela A Miller, Joshua Vasquez, Marc Weiner, Adam Chapman, Melissa Engle, Michael Higgins, Amy M Quinones, Vanessa Rosselli, Elizabeth Canono, et al. Blood transcriptional biomarkers for active tuberculosis among patients in the united states: a case-control study with systematic cross-classifier evaluation. *Journal of clinical microbiology*, 54(2):274–282, 2016.

[164] Julie G Burel, Cecilia S Lindestam Arlehamn, Nabeela Khan, Grégory Seumois, Jason A Greenbaum, Randy Taplitz, Robert H Gilman, Mayuko Saito, Pandurangan Vijayanand, Alessandro Sette, et al. Transcriptomic analysis of cd4+ t cells reveals novel immune signatures of latent tuberculosis. *The Journal of Immunology*, 200(9):3283–3290, 2018.

[165] Meng Shao, Fang Wu, Jie Zhang, Jiangtao Dong, Hui Zhang, Xiaoling Liu, Su Liang, Jiangdong Wu, Le Zhang, Chunjun Zhang, et al. Screening of potential biomarkers for distinguishing between latent and active tuberculosis in children using bioinformatics analysis. *Medicine*, 100(5), 2021.

[166] Shu Hui Cao, Yan Qing Chen, SUN Yong, LIU Yang, Su Hua Zheng, Zhi Guo Zhang, and Chuan You Li. Screening of serum biomarkers for distinguishing between latent and active tuberculosis using proteome microarray. *Biomedical and Environmental Sciences*, 31(7):515–526, 2018.

[167] Ziyan Zhang, Rajat Singh, and Michael Aschner. Methods for the detection of autophagy in mammalian cells. *Current protocols in toxicology*, 69(1):20–12, 2016.

[168] Andrew Thorburn, Douglas H Thamm, and Daniel L Gustafson. Autophagy and cancer therapy. *Molecular pharmacology*, 85(6):830–838, 2014.

[169] Marta Pérez-Hernández, Alain Arias, David Martínez-García, Ricardo Pérez-Tomás, Roberto Quesada, and Vanessa Soto-Cerrato. Targeting autophagy for cancer treatment and tumor chemosensitization. *Cancers*, 11(10):1599, 2019.

[170] Eeva-Liisa Eskelinen. The dual role of autophagy in cancer. *Current Opinion in Pharmacology*, 11(4):294–300, 2011.

[171] RL Deter. Analog modeling of glucagon-induced autophagy in rat liver: I. conceptual and mathematical model of telolysosome-autophagosome-autolysosome interaction. *Experimental cell research*, 94(1):122–126, 1975.

[172] Huiqin Jin and Jinzhi Lei. A mathematical model of cell population dynamics with autophagy response to starvation. *Mathematical biosciences*, 258:1–10, 2014.

[173] Huiqin Jin and Jinzhi Lei. A hybrid model of molecular regulation and population dynamics for yeast autophagy. *Journal of Theoretical Biology*, 402:45–53, 2016.

[174] Afroza Shirin, Isaac S Klickstein, Song Feng, Yen Ting Lin, William S Hlavacek, and Francesco Sorrentino. Prediction of optimal drug schedules for controlling autophagy. *Scientific Reports*, 9(1):1–15, 2019.

[175] Türküler Özgümüs, Oksana Sulaieva, Ruchi Jain, Isabella Artner, and Valeriya Lyssenko. Starvation to glucose reprograms development of neurovascular unit in embryonic retinal cells. *Frontiers in cell and developmental biology*, 9, 2021.

[176] Andrew Thorburn. Apoptosis and autophagy: regulatory connections between two supposedly different processes. *Apoptosis*, 13(1):1–9, 2008.

[177] Diane Crighton, Simon Wilkinson, Jim O'Prey, Nelofer Syed, Paul Smith, Paul R Harrison, Milena Gasco, Ornella Garrone, Tim Crook, and Kevin M Ryan. Dram, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell*, 126(1):121–134, 2006.

[178] Orsolya Kapuy, PK Vinod, József Mandl, and Gábor Bánhegyi. A cellular stress-directed bistable switch controls the crosstalk between autophagy and apoptosis. *Molecular BioSystems*, 9(2):296–306, 2013.

[179] I Tavassoly, J Parmar, AN Shajahan-Haq, R Clarke, William T Baumann, and John J Tyson. Dynamic modeling of the interaction between autophagy and apoptosis in mammalian cells. *CPT: pharmacometrics & systems pharmacology*, 4(4):263–272, 2015.

[180] Bojie Yang, Quansheng Liu, and Yuanhong Bi. Autophagy and apoptosis are regulated by stress on bcl2 by ambra1 in the endoplasmic reticulum and mitochondria. *Theoretical Biology and Medical Modelling*, 16(1):1–9, 2019.

[181] Jignesh H Parmar, Katherine L Cook, Ayesha N Shajahan-Haq, Pamela AG Clarke, Iman Tavassoly, Robert Clarke, John J Tyson, and William T Baumann. Modelling the effect of grp78 on anti-oestrogen sensitivity and resistance in breast cancer. *Interface Focus*, 3(4):20130012, 2013.

[182] Katie R Martin, Dipak Barua, Audra L Kauffman, Laura M Westrate, Richard G Posner, William S Hlavacek, and Jeffrey P MacKeigan. Computational model for autophagic vesicle dynamics in single cells. *Autophagy*, 9(1):74–92, 2013.

[183] Kyungreem Han, Jinwoong Kim, and Mooyoung Choi. Computer simulations unveil the dynamics of autophagy and its implications for the cellular quality control. *Journal of Biological Systems*, 22(04):659–675, 2014.

[184] Kyungreem Han, Soon Ho Kim, and MooYoung Choi. Computational modeling of the effects of autophagy on amyloid-β peptide levels. *Theoretical Biology and Medical Modelling*, 17(1):1–16, 2020.

[185] Christoph S Börlin, Verena Lang, Anne Hamacher-Brady, and Nathan R Brady. Agent-based modeling of autophagy reveals emergent regulatory behavior of spatio-temporal autophagy dynamics. *Cell Communication and Signaling*, 12(1):1–16, 2014.

[186] Jennifer Scheidel, Leonie Amstein, Jörg Ackermann, Ivan Dikic, and Ina Koch. In silico knockout studies of xenophagic capturing of salmonella. *PLoS computational biology*, 12(12):e1005200, 2016.

[187] Guolin Lu, Dongdong Rao, Min Zhou, Liangpu Xu, Longxin Zhang, Sujing Zhang, and Yuping Wang. Autophagic network analysis of the dual effect of sevoflurane on neurons associated with gabarapl1 and 2. *BioMed Research International*, 2020, 2020.

[188] Ji-Ye Wang, Wei-Xuan Yao, Yun Wang, Yi-lei Fan, and Jian-Bing Wu. Network analysis reveals crosstalk between autophagy genes and disease genes. *Scientific Reports*, 7(1):1–9, 2017.

[189] Zhenyu Yue, Shengkan Jin, Chingwen Yang, Arnold J Levine, and Nathaniel Heintz. Beclin 1, an autophagy gene essential for early embryonic development, is a haploinsufficient tumor suppressor. *Proceedings of the National Academy of Sciences*, 100(25):15077–15082, 2003.

[190] Liewen Lin, Ligang Xia, Donge Tang, Yong Dai, and Wenbiao Chen. Analysis of autophagy-related genes and associated noncoding rnas and transcription factors in digestive system tumors. *Future Oncology*, 15(36):4141–4154, 2019.

[191] Joanna S Amberger, Carol A Bocchini, Alan F Scott, and Ada Hamosh. Omim. org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Research*, 47(D1):D1038–D1043, 2019.

[192] Keiichi Homma, Koji Suzuki, and Hideaki Sugawara. The autophagy database: an all-inclusive information resource on autophagy that provides nourishment for research. *Nucleic Acids Research*, 39(suppl_1):D986–D990, 2010.

[193] Dénes Türei, László Földvári-Nagy, Dávid Fazekas, Dezső Módos, János Kubisch, Tamás Kadlecsik, Amanda Demeter, Katalin Lenti, Péter Csermely, Tibor Vellai, et al. Autophagy regulatory network—a systems-level bioinformatics resource for studying the mechanism and regulation of autophagy. *Autophagy*, 11(1):155–165, 2015.

[194] Marc Durocher, Bradley P Ander, Glen Jickling, Farah Hamade, Heather Hull, Bodie Knepp, Da Zhi Liu, Xinhua Zhan, Anh Tran, Xiyuan Cheng, et al. Inflammatory, regulatory, and autophagy co-expression modules and hub genes underlie the peripheral immune response to human intracerebral hemorrhage. *Journal of neuroinflammation*, 16(1):1–21, 2019.

[195] Da Wei Huang, Brad T Sherman, and Richard A Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, 4(1):44–57, 2009.

[196] Fengfeng Wang, William Cho, Lawrence WC Chan, SC Wong, Nancy BY Tsui, Parco M Siu, Shea Ping Yip, and Benjamin YM Yung. Gene network exploration of crosstalk between apoptosis and autophagy in chronic myelogenous leukemia. *BioMed research international*, 2015, 2015.

[197] Mahmoud Ahmed, Jin Seok Hwang, Trang Huyen Lai, Sahib Zada, Huynh Quoc Nguyen, Trang Min Pham, Miyong Yun, and Deok Ryong Kim. Co-expression network analysis of ampk and autophagy gene products during adipocyte differentiation. *International Journal of Molecular Sciences*, 19(6):1808, 2018.

[198] Guangyin Zhang, Shixin Xu, Zhuo Yuan, and Li Shen. Weighted gene coexpression network analysis identifies specific modules and hub genes related to major depression. *Neuropsychiatric Disease and Treatment*, 2020.

[199] Tanya Barrett, Stephen E Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F Kim, Maxim Tomashevsky, Kimberly A Marshall, Katherine H Phillippy, Patti M Sherman, Michelle Holko, et al. Ncbi geo: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995, 2012.

[200] Jing Wang, Shanshan Cong, Han Wu, Yanan He, Xiaoli Liu, Liyuan Sun, Xibo Zhao, and Guangmei Zhang. Identification and analysis of potential autophagy-related biomarkers in endometriosis by wgcna. *Frontiers in Molecular Biosciences*, 8, 2021.

[201] Miao Xu, Tianxiang Ouyang, Kaiyang Lv, and Xiaorong Ma. Integrated wgcna and ppi network to screen hub genes signatures for infantile hemangioma. *Frontiers in Genetics*, 11:614195, 2021.

[202] Zhaoyue He, He Liu, Holger Moch, and Hans-Uwe Simon. Machine learning with autophagy-related proteins for discriminating renal cell carcinoma subtypes. *Scientific Reports*, 10(1):1–7, 2020.

[203] Janos Kriston-Vizi, Ching Aeng Lim, Peter Condron, Kelvin Chua, Martin Wasser, and Horst Flotow. An automated high-content screening image analysis pipeline for the identification of selective autophagic inducers in human cancer cell lines. *Journal of biomolecular screening*, 15(7):869–881, 2010.

[204] Amandine Serrano, Saïd El Haddad, Frédéric Moal, Thierry Prazuck, Eric Legac, Chloé Robin, Fabienne Brulé, Stéphane Charpentier, Thierry Normand, Alain Legrand, et al. Dysregulation of apoptosis and autophagy gene expression in peripheral blood mononuclear cells of efficiently treated hiv-infected patients. *Aids*, 32(12):1579–1587, 2018.

[205] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of Machine Learning Research*, 12:2825–2830, 2011.

[206] Qingya Shi, Fen Pei, Gary A Silverman, Stephen C Pak, David H Perlmutter, Bing Liu, and Ivet Bahar. Mechanisms of action of autophagy modulators dissected by quantitative systems pharmacology analysis. *International journal of molecular sciences*, 21(8):2855, 2020.

[207] Ning-Ning Wang, Jie Dong, Lin Zhang, Defang Ouyang, Yan Cheng, Alex F Chen, Ai-Ping Lu, and Dong-Sheng Cao. Hamdb: a database of human autophagy modulators

with specific pathway and disease information. *Journal of Cheminformatics*, 10(1):1–8, 2018.

[208] Deng Wu, Yan Huang, Juanjuan Kang, Kongning Li, Xiaoman Bi, Ting Zhang, Nana Jin, Yongfei Hu, Puwen Tan, Lu Zhang, et al. ncrdeathdb: A comprehensive bioinformatics resource for deciphering network organization of the ncrna-mediated cell death system. *Autophagy*, 11(10):1917–1926, 2015.

[209] Yiqi Deng, Lingjuan Zhu, Haoyang Cai, Guan Wang, and Bo Liu. Autophagic compound database: a resource connecting autophagy-modulating compounds, their potential targets and relevant diseases. *Cell Proliferation*, 51(3):e12403, 2018.

[210] Wankun Deng, Lili Ma, Ying Zhang, Jiaqi Zhou, Yongbo Wang, Zexian Liu, and Yu Xue. Thanatos: an integrative data resource of proteins and post-translational modifications in the regulation of autophagy. *Autophagy*, 14(2):296–310, 2018.

[211] Ravikanth Nanduri, Rashi Kalra, Ella Bhagyaraj, Anuja P Chacko, Nancy Ahuja, Drishti Tiwari, Sumit Kumar, Monika Jain, Raman Parkesh, and Pawan Gupta. Autophagysmdb: a curated database of small molecules that modulate protein targets regulating autophagy. *Autophagy*, 15(7):1280–1295, 2019.

[212] Wenjing Wang, Peng Zhang, Leijie Li, Zhaobin Chen, Weiyang Bai, Guiyou Liu, Liangcai Zhang, Haiyang Jia, Li Li, Yingcui Yu, et al. Atd: a comprehensive bioinformatics resource for deciphering the association of autophagy and diseases. *Database*, 2018, 2018.

[213] Anne-Claire Jacomin, Siva Samavedam, Vasilis Promponas, and Ioannis P Nezis. ilir database: A web resource for lir motif-containing proteins in eukaryotes. *Autophagy*, 12(10):1945–1953, 2016.

[214] Kelie Chen, Dexin Yang, Fan Zhao, Shengchao Wang, Yao Ye, Wenjie Sun, Haohua Lu, Zhi Ruan, Jinming Xu, Tianru Wang, et al. Autophagy and tumor database: Atdb, a novel database connecting autophagy and tumor. *Database*, 2020, 2020.

[215] Damian Szklarczyk, Annika L Gable, David Lyon, Alexander Junge, Stefan Wyder, Jaime Huerta-Cepas, Milan Simonovic, Nadezhda T Doncheva, John H Morris, Peer Bork, et al. String v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1):D607–D613, 2019.

[216] Livia Perfetto, Leonardo Briganti, Alberto Calderone, Andrea Cerquone Perpetuini, Marta Iannuccelli, Francesca Langone, Luana Licata, Milica Marinkovic, Anna Mattioni, Theodora Pavlidou, et al. Signor: a database of causal relationships between biological entities. *Nucleic Acids Research*, 44(D1):D548–D554, 2016.

[217] Ranjit Unnikrishnan, Ranjit Mohan Anjana, and Viswanathan Mohan. Diabetes mellitus and its complications in india. *Nature Reviews Endocrinology*, 12(6):357–370, 2016.

[218] LZ Heng, O Comyn, T Peto, C Tadros, E Ng, S Sivaprasad, and PG Hykin. Diabetic retinopathy: pathogenesis, clinical grading, management and future developments. *Diabetic Medicine*, 30(6):640–650, 2013.

[219] Jason Noble and Varun Chaudhary. Five things to know about... diabetic retinopathy. *CMAJ: Canadian Medical Association journal= journal de l'Association medicale canadienne*, 182(15):1646–1646, 2010.

[220] Diabetes Care. Standards of medical care in diabetes 2019. *Diabetes Care*, 42(Suppl 1):S124–138, 2019.

[221] Michelino Di Rosa, Gisella Distefano, Caterina Gagliano, Dario Rusciano, and Lucia Malaguarnera. Autophagy in diabetic retinopathy. *Current neuropharmacology*, 14(8):810–825, 2016.

[222] Qiaoyun Gong, Haiyan Wang, Ping Yu, Tianwei Qian, and Xun Xu. Protective or harmful: The dual roles of autophagy in diabetic retinopathy. *Frontiers in Medicine*, 8, 2021.

[223] Mingyuan Wu, Ying Chen, Kenneth Wilson, Alin Chirindel, Michael A Ihnat, Yongxin Yu, Michael E Boulton, Luke I Szweda, Jian-Xing Ma, and Timothy J Lyons. Intraretinal leakage and oxidation of ldl in diabetic retinopathy. *Investigative ophthalmology & visual science*, 49(6):2679–2685, 2008.

[224] Jacqueline M Lopes de Faria, Diego A Duarte, Chiara Montemurro, Alexandros Papadimitriou, Sílvio Roberto Consonni, and José B Lopes de Faria. Defective autophagy in diabetic retinopathy. *Investigative ophthalmology & visual science*, 57(10):4356–4366, 2016.

[225] Rong Li, Jun-Hui Du, Guo-Min Yao, Yang Yao, and Jin Zhang. Autophagy: a new mechanism for regulating vegf and pedf expression in retinal pigment epithelium cells. *International Journal of Ophthalmology*, 12(4):557, 2019.

[226] Rosario Amato, Elisabetta Catalani, Massimo Dal Monte, Maurizio Cammalleri, Ilaria Di Renzo, Cristiana Perrotta, Davide Cervia, and Giovanni Casini. Autophagy-mediated neuroprotection induced by octreotide in an ex vivo model of early diabetic retinopathy. *Pharmacological Research*, 128:167–178, 2018.

[227] Vidhya Gopalakrishnan, Arvind Nambiar, Sukanya Basu, and G Madhuvanthi. Ppi network analysis of diabetic retinopathy genes. *International Journal of Computational Biology and Drug Design*, 13(3):302–315, 2020.

[228] Lei Wang, Shuyan Li, Leilei Wang, Kai Lin, Jialun Du, Wanhong Miao, and Lei Zhang. Uncovering the protective mechanism of taohong siwu decoction against diabetic retinopathy via hif-1 signaling pathway based on network analysis and experimental validation. *BMC complementary medicine and therapies*, 20(1):1–16, 2020.

[229] Akram Safaei, Mostafa Rezaei Tavirani, Mona Zamanian Azodi, Alireza Lashay, Seyed Farzad Mohammadi, Mohamad Ghasemi Broumand, Ali Asghar Peyvandi, Farshad Okhovatian, Hassan Peyvandi, and Mohammad Rostami Nejad. Diabetic retinopathy and laser therapy in rats: A protein-protein interaction network analysis. *Journal of lasers in medical sciences*, 8(Suppl 1):S20, 2017.

[230] Xinxiao Gao, Yunhui Du, Wayne Bond Lau, Yu Li, Siquan Zhu, and Xin-Liang Ma. Atg16l1 as a novel biomarker and autophagy gene for diabetic retinopathy. *Journal of Diabetes Research*, 2021, 2021.

[231] Nan Wang, Linfeng Wei, Die Liu, Quyan Zhang, Xiaobo Xia, Lexi Ding, and Siqi Xiong. Identification and validation of autophagy-related genes in diabetic retinopathy. *Frontiers in Endocrinology*, 13, 2022.

[232] Juan I Fuxman Bass, Alos Diallo, Justin Nelson, Juan M Soto, Chad L Myers, and Albertha JM Walhout. Using networks to measure similarity between genes: association index selection. *Nature methods*, 10(12):1169–1176, 2013.

[233] Sahra Uygun, Cheng Peng, Melissa D Lehti-Shiu, Robert L Last, and Shin-Han Shiu. Utility and limitations of using gene expression data to identify functional associations. *PLoS computational biology*, 12(12):e1005244, 2016.

[234] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I Furlong. The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*, 48(D1):D845–D855, 2020.

[235] Uniprot: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489, 2021.

[236] Xu-Dong Zou, Ke An, Yun-Dong Wu, and Zhi-Qiang Ye. Ppi network analyses of human wd40 protein family systematically reveal their tendency to assemble complexes and facilitate the complex predictions. *BMC Systems Biology*, 12(4):49–62, 2018.

[237] Athina I Amanatidou and George V Dedoussis. Construction and analysis of protein-protein interaction network of non-alcoholic fatty liver disease. *Computers in biology and medicine*, 131:104243, 2021.

[238] Arnold I Emerson, Simeon Andrews, Ikhlak Ahmed, Thasni KA Azis, and Joel A Malek. K-core decomposition of a protein domain co-occurrence network reveals lower cancer mutation rates for interior cores. *Journal of clinical bioinformatics*, 5(1):1–11, 2015.

[239] Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210, 2002.

[240] Andrew D Skol, Segun C Jung, Ana Marija Sokovic, Siquan Chen, Sarah Fazal, Olukayode Sosina, Poulami P Borkar, Amy Lin, Maria Sverdlov, Dingcai Cao, et al. Integration of genomics and transcriptomics predicts diabetic retinopathy susceptibility genes. *Elife*, 9:e59980, 2020.

[241] Sipko Van Dam, Urmo Vosa, Adriaan van der Graaf, Lude Franke, and Joao Pedro de Magalhaes. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in bioinformatics*, 19(4):575–592, 2018.

[242] Maxim V Kuleshov, Matthew R Jones, Andrew D Rouillard, Nicolas F Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L Jenkins, Kathleen M Jagodnik, Alexander Lachmann, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–W97, 2016.

[243] Emma M Lessieur, Haitao Liu, Aicha Saadane, Yunpeng Du, Jie Tang, Jianying Kiser, and Timothy S Kern. Neutrophil-derived proteases contribute to the pathogenesis of early diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, 62(13):7–7, 2021.

[244] Hailiang Wu, De-Kuang Hwang, Xudong Song, and Yong Tao. Association between aqueous cytokines and diabetic retinopathy stage. *Journal of Ophthalmology*, 2017, 2017.

[245] Tianzhi Wu, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, Lang Zhou, Wenli Tang, Li Zhan, et al. clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3):100141, 2021.

[246] Yue Wang, Yun-hong Lu, Chao Tang, Mei Xue, Xiao-yu Li, Yun-peng Chang, Ying Cheng, Ting Li, Xiao-chen Yu, Bei Sun, et al. Calcium dobesilate restores autophagy by inhibiting the vegf/pi3k/akt/mtor signaling pathway. *Frontiers in pharmacology*, page 886, 2019.

[247] Yingfeng Zheng, Mingguang He, and Nathan Congdon. The worldwide epidemic of diabetic retinopathy. *Indian journal of ophthalmology*, 60(5):428, 2012.

[248] Teresa Tsai, Sandra Kuehn, Nikolaos Tsiampalis, Minh-Khoa Vu, Vinodh Kakkassery, Gesa Stute, H Burkhard Dick, and Stephanie C Joachim. Anti-inflammatory cytokine and angiogenic factors levels in vitreous samples of diabetic retinopathy patients. *PLoS One*, 13(3):e0194603, 2018.

[249] Antonia M Joussen, Toshinori Murata, Akitaka Tsujikawa, Bernd Kirchhof, Sven-Erik Bursell, and Anthony P Adamis. Leukocyte-mediated endothelial cell injury and death in the diabetic retina. *The American journal of pathology*, 158(1):147–152, 2001.

[250] Xin Gao, Yuan Zhang, Rui Zhang, Zichan Zhao, Haorui Zhang, Jinhui Wu, Wei Shen, and Ming Zhong. Cyclin-dependent kinase 1 disruption inhibits angiogenesis by inducing cell cycle arrest and apoptosis. *Experimental and Therapeutic Medicine*, 18(4):3062–3070, 2019.

[251] Junghyun Kim, Chan-Sik Kim, Eunjin Sohn, and Jin Sook Kim. Cytoplasmic translocation of high-mobility group box-1 protein is induced by diabetes and high glucose in retinal pericytes. *Molecular Medicine Reports*, 14(4):3655–3661, 2016.

[252] Quresh Mohamed, Mark C Gillies, and Tien Y Wong. Management of diabetic retinopathy: a systematic review. *Jama*, 298(8):902–916, 2007.

[253] Eimear M Byrne, María Llorián-Salvador, Timothy J Lyons, Mei Chen, and Heping Xu. Tofacitinib ameliorates retinal vascular leakage in a murine model of diabetic retinopathy with type 2 diabetes. *International Journal of Molecular Sciences*, 22(21):11876, 2021.

[254] RA Kowluru and S Odenbach. Role of interleukin-1β in the pathogenesis of diabetic retinopathy. *British Journal of Ophthalmology*, 88(10):1343–1347, 2004.

[255] Kimberly A Drenser. Wnt signaling pathway in retinal vascularization. *Eye and Brain*, 8:141, 2016.

[256] Renu A Kowluru, Rakesh Radhakrishnan, and Ghulam Mohammad. Regulation of rac1 transcription by histone and dna methylation in diabetic retinopathy. *Scientific Reports*, 11(1):1–10, 2021.

[257] Qingzheng Kang and Chunxue Yang. Oxidative stress and diabetic retinopathy: Molecular mechanisms, pathogenetic role and therapeutic implications. *Redox Biology*, 37:101799, 2020.

[258] Qiong Huang and Nader Sheibani. High glucose promotes retinal endothelial cell migration through activation of src, pi3k/akt1/enos, and erks. *American Journal of Physiology-Cell Physiology*, 295(6):C1647–C1657, 2008.

[259] Aneliya Parvanova, Ilian Iliev, Marco Filipponi, Borislav D Dimitrov, Monica Vedovato, Antonio Tiengo, Roberto Trevisan, Giuseppe Remuzzi, and Piero Ruggenenti. Insulin resistance and proliferative retinopathy: a cross-sectional, case-control study in 115 patients with type 2 diabetes. *The Journal of Clinical Endocrinology & Metabolism*, 89(9):4371–4376, 2004.

[260] Ning Jiang, Xiao-Long Chen, Hong-Wei Yang, and Yu-Ru Ma. Effects of nuclear factor κb expression on retinal neovascularization and apoptosis in a diabetic retinopathy rat model. *International Journal of Ophthalmology*, 8(3):448, 2015.

[261] N Gupta, S Mansoor, A Sharma, A Sapkal, J Sheth, P Falatoonzadeh, BD Kuppermann, and MC Kenney. Diabetic retinopathy and vegf. *The Open Ophthalmology journal*, 7:4, 2013.

[262] Derrick J Feenstra, E Chepchumba Yego, and Susanne Mohr. Modes of retinal cell death in diabetic retinopathy. *Journal of clinical & Experimental Ophthalmology*, 4(5):298, 2013.

[263] Landon J Rohowetz, Jacob G Kraus, and Peter Koulen. Reactive oxygen species-mediated damage of retinal neurons: Drug development targets for therapies of chronic neurodegeneration of the retina. *International Journal of Molecular Sciences*, 19(11):3362, 2018.

[264] Talia N Crawford, D Virgil Alfaro III, John B Kerrison, Eric P Jablon, et al. Diabetic retinopathy and angiogenesis. *Current Diabetes Reviews*, 5(1):8–13, 2009.

[265] Antonela Gverović Antunica, Ksenija Karaman, Ljubo Znaor, Ada Sapunar, Vesna Buško, and Velibor Puzović. Il-12 concentrations in the aqueous humor and serum of diabetic retinopathy patients. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 250(6):815–821, 2012.

[266] CB Holt, IT Hoffmann-Petersen, TK Hansen, HH Parving, S Thiel, P Hovind, L Tarnow, P Rossing, and JA Østergaard. Association between severe diabetic retinopathy and lectin pathway proteins-an 18-year follow-up study with newly diagnosed type 1 diabetes patients. *Immunobiology*, 225(3):151939–151939, 2020.

[267] Murugeswari Ponnalagu, Murali Subramani, Chaitra Jayadev, Rohit Shetty, and Debashish Das. Retinal pigment epithelium-secretome: A diabetic retinopathy perspective. *Cytokine*, 95:126–135, 2017.

[268] Gao-Hua Liang, Yan-Ni Luo, Ri-Zhang Wei, Jia-Yang Yin, Zhi-Liang Qin, Li-Li Lu, and Wen-Hao Ma. Circznf532 knockdown protects retinal pigment epithelial cells against high glucose-induced apoptosis and pyroptosis by regulating the mir-20b-5p/stat3 axis. *Journal of Diabetes Investigation*, 13(5):781–795, 2022.

[269] David Ferland-McCollough, Sadie Slater, Jai Richard, Carlotta Reni, and Giuseppe Mangialardi. Pericytes, an overlooked player in vascular pathobiology. *Pharmacology & Therapeutics*, 171:30–42, 2017.

[270] Qian Li, Lei Pang, Hui Shi, Wei Yang, Xin Liu, Guanfang Su, and Yu Dong. High glucose concentration induces retinal endothelial cell apoptosis by activating p53 signaling pathway. *International journal of clinical and experimental pathology*, 11(5):2401, 2018.

[271] Sujoy Bhattacharya, Edward Chaum, Dianna A Johnson, and Leonard R Johnson. Age-related susceptibility to apoptosis in human retinal pigment epithelial cells is triggered by disruption of p53–mdm2 association. *Investigative Ophthalmology & Visual Science*, 53(13):8350–8366, 2012.

[272] Zafer Gurel, Balyn W Zaro, Matthew R Pratt, and Nader Sheibani. Identification of o-glcnac modification targets in mouse retinal pericytes: implication of p53 in pathogenesis of diabetic retinopathy. *PloS one*, 9(5):e95561, 2014.

[273] Morvarid Farhang Ghahremani, Steven Goossens, David Nittner, Xavier Bisteau, Sonia Bartunkova, Aleksandra Zwolinska, Paco Hulpiau, Katharina Haigh, Lieven Haenebalcke, Benjamin Drogat, et al. p53 promotes vegf expression and angiogenesis in the absence of an intact p21-rb pathway. *Cell Death & Differentiation*, 20(7):888–897, 2013.

[274] Haocheng Ao, Bingqian Liu, Haichun Li, and Lin Lu. Egr1 mediates retinal vascular dysfunction in diabetes mellitus via promoting p53 transcription. *Journal of cellular and molecular medicine*, 23(5):3345–3356, 2019.

[275] Adam C Sheka, Oyedele Adeyi, Julie Thompson, Bilal Hameed, Peter A Crawford, and Sayeed Ikramuddin. Nonalcoholic steatohepatitis: a review. *Jama*, 323(12):1175–1183, 2020.

[276] A Tannapfel, H Denk, HP Dienes, C Langner, P Schirmacher, M Trauner, and B Flott-Rahmel. Histopathological diagnosis of non-alcoholic and alcoholic fatty liver disease. *Pathologe*, 2010.

[277] Arun J Sanyal. Aga technical review on nonalcoholic fatty liver disease. *Gastroenterology*, 123(5):1705–1725, 2002.

[278] Geoffrey C Farrell and Claire Z Larter. Nonalcoholic fatty liver disease: from steatosis to cirrhosis. *Hepatology*, 43(S1):S99–S112, 2006.

[279] Zobair M Younossi, Aaron B Koenig, Dinan Abdelatif, Yousef Fazel, Linda Henry, and Mark Wymer. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology*, 64(1):73–84, 2016.

[280] Pushpjeet Kanwar and Kris V Kowdley. The metabolic syndrome and its influence on nonalcoholic steatohepatitis. *Clinics in Liver Disease*, 20(2):225–243, 2016.

[281] David E Kleiner, Elizabeth M Brunt, Mark Van Natta, Cynthia Behling, Melissa J Contos, Oscar W Cummings, Linda D Ferrell, Yao-Chang Liu, Michael S Torbenson, Aynur Unalp-Arida, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology*, 41(6):1313–1321, 2005.

[282] Pierre Bedossa, Christine Poitou, Nicolas Veyrie, Jean-Luc Bouillot, Arnaud Basdevant, Valerie Paradis, Joan Tordjman, and Karine Clement. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology*, 56(5):1751–1759, 2012.

[283] Thierry Huby and Emmanuel L Gautier. Immune cell-mediated features of non-alcoholic steatohepatitis. *Nature Reviews Immunology*, pages 1–15, 2021.

[284] Montgomery Blencowe, Tilan Karunanayake, Julian Wier, Neil Hsu, and Xia Yang. Network modeling approaches and applications to unravelling non-alcoholic fatty liver disease. *Genes*, 10(12):966, 2019.

[285] Jonathan C Cohen, Jay D Horton, and Helen H Hobbs. Human fatty liver disease: old questions and new insights. *Science*, 332(6037):1519–1523, 2011.

[286] Jukka Westerbacka, Anna Kotronen, Barbara A Fielding, John Wahren, Leanne Hodson, Julia Perttilä, Tuulikki Seppänen-Laakso, Tapani Suortti, Johanna Arola, Rolf Hultcrantz, et al. Splanchnic balance of free fatty acids, endocannabinoids, and lipids in subjects with nonalcoholic fatty liver disease. *Gastroenterology*, 139(6):1961–1971, 2010.

[287] Nishanth E Sunny, Elizabeth J Parks, Jeffrey D Browning, and Shawn C Burgess. Excessive hepatic mitochondrial tca cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. *Cell Metabolism*, 14(6):804–810, 2011.

[288] Ming-Jiang Liu, Hu Jin, Yu-Bing Chen, Jing-Jing Yu, Zhen-Ya Guo, Song-Qing He, and Yong-Lian Zeng. Screening of non-alcoholic steatohepatitis (nash)-related datasets and identification of nash-related genes. *International Journal of Clinical and Experimental Pathology*, 14(5):567, 2021.

[289] Hamid Asadzadeh-Aghdaee, Vahid Mansouri, Ali Asghar Peyvandi, Fathollah Moztarzadeh, Farshad Okhovatian, Farhad Lahmi, Reza Vafaee, and Mohammad Reza Zali.

Topological and functional analysis of nonalcoholic steatohepatitis through protein interaction mapping. *Gastroenterology and Hepatology from Bed to Bench*, 9(Suppl1):S23, 2016.

[290] Gong Feng, Xue-Ping Li, Chun-Yan Niu, Man-Ling Liu, Qin-qin Yan, Li-Ping Fan, Ya Li, Ke-Lin Zhang, Jie Gao, Mei-Rui Qian, et al. Bioinformatics analysis reveals novel core genes associated with nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Gene*, 742:144549, 2020.

[291] Zhen-yu Jiang, Yi Zhou, Lu Zhou, Shao-wei Li, and Bang-mao Wang. Identification of key genes and immune infiltrate in nonalcoholic steatohepatitis: A bioinformatic analysis. *BioMed Research International*, 2021, 2021.

[292] Jianzhong Ye, Yishuai Lin, Qing Wang, Yating Li, Yajie Zhao, Lijiang Chen, Qing Wu, Chunquan Xu, Cui Zhou, Yao Sun, et al. Integrated multichip analysis identifies potential key genes in the pathogenesis of nonalcoholic steatohepatitis. *Frontiers in Endocrinology*, 11:601745, 2020.

[293] Ruifeng Wang, Xiaobing Wang, and Liwei Zhuang. Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease. *Annals of Hepatology*, 15(2):190–199, 2016.

[294] Z Wang, Z Zhao, Y Xia, Z Cai, C Wang, Y Shen, R Liu, H Qin, J Jia, and G Yuan. Potential biomarkers in the fibrosis progression of nonalcoholic steatohepatitis (nash). *Journal of Endocrinological Investigation*, pages 1–14, 2022.

[295] Jiao Wang, Honghong Liu, Guijiao Xie, Wei Cai, and Jixiong Xu. Identification of hub genes and key pathways of dietary advanced glycation end products-induced nonalcoholic fatty liver disease by bioinformatics analysis and animal experiments. *Molecular Medicine Reports*, 21(2):685–694, 2020.

[296] Reza Karbalaei, Mehran Piran, Mostafa Rezaei-Tavirani, Hamid Asadzadeh-Aghdaei, and Mohammad Hossein Heidari. A systems biology analysis protein-protein interaction of nash and ibd based on comprehensive gene information. *Gastroenterology and Hepatology from Bed to Bench*, 10(3):194, 2017.

[297] Lei Li, Sheng-he Li, Jin-peng Jiang, Chang Liu, and Li-li Ji. Investigating pharmacological mechanisms of andrographolide on non-alcoholic steatohepatitis (nash): a bioinformatics approach of network pharmacology. *Chinese Herbal Medicines*, 13(3):342–350, 2021.

[298] Wenlong Sun, Panpan Liu, Bendong Yang, Meng Wang, Tianqi Wang, Wenbo Sun, Xudong Wang, Weilong Zheng, Xinhua Song, and Jingda Li. A network pharmacology

approach: Inhibition of the nf-κb signaling pathway contributes to the nash preventative effect of an oroxylum indicum seed extract in oleic acid-stimulated hepg2 cells and high-fat diet-fed rats. *Phytomedicine*, 88:153498, 2021.

[299] Rui Gao, Xiaobo Zhang, Zhen Zhou, Jiayi Sun, Xuehua Tang, Jialiang Li, Xin Zhou, and Tao Shen. Pharmacological mechanism of ganlu powder in the treatment of nash based on network pharmacology and molecular docking. *Disease Markers*, 2022, 2022.

[300] Wenjun Zhou, Ziye Zhu, Xiaoli Xiao, Chunlin Li, Li Zhang, Yanqi Dang, Guangbo Ge, Guang Ji, Mingzhe Zhu, and Hongxi Xu. Jiangzhi granule attenuates non-alcoholic steatohepatitis by suppressing tnf/nfκb signaling pathway-a study based on network pharmacology. *Biomedicine & Pharmacotherapy*, 143:112181, 2021.

[301] Fu Chen, Yong Zhou, Zhiyuan Wu, Yunze Li, Wenlong Zhou, and Yong Wang. Integrated analysis of key genes and pathways involved in nonalcoholic steatohepatitis improvement after roux-en-y gastric bypass surgery. *Frontiers in Endocrinology*, 11:611213, 2021.

[302] Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell*, 171(6):1437–1452, 2017.

[303] Malte P Suppli, Kristoffer TG Rigbolt, Sanne S Veidal, Sara Heebøll, Peter Lykke Eriksen, Mia Demant, Jonatan I Bagger, Jens Christian Nielsen, Denise Oró, Sebastian W Thrane, et al. Hepatic transcriptome signatures in patients with varying degrees of non-alcoholic fatty liver disease compared with healthy normal-weight individuals. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 316(4):G462–G472, 2019.

[304] Kai K Kummer, Theodora Kalpachidou, Michaela Kress, and Michiel Langeslag. Signatures of altered gene expression in dorsal root ganglia of a fabry disease mouse model. *Frontiers in Molecular Neuroscience*, 10:449, 2018.

[305] Gennady Korotkevich, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N Artyomov, and Alexey Sergushichev. Fast gene set enrichment analysis. *BioRxiv*, page 060012, 2021.

[306] Minoru Kanehisa and Susumu Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[307] David Croft, Gavin O'kelly, Guanming Wu, Robin Haw, Marc Gillespie, Lisa Matthews, Michael Caudy, Phani Garapati, Gopal Gopinath, Bijay Jassal, et al. Reactome: a

database of reactions, pathways and biological processes. *Nucleic Acids Research*, 39(suppl_1):D691–D697, 2010.

[308] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6):417–425, 2015.

[309] Hasun Yu, Sungji Choo, Junseok Park, Jinmyung Jung, Yeeok Kang, and Doheon Lee. Prediction of drugs having opposite effects on disease genes in a directed network. In *BMC Systems Biology*, volume 10, pages 17–25. Springer, 2016.

[310] Arunachalam Vinayagam, Ulrich Stelzl, Raphaele Foulle, Stephanie Plassmann, Martina Zenkner, Jan Timm, Heike E Assmus, Miguel A Andrade-Navarro, and Erich E Wanker. A directed protein interaction network for investigating intracellular signal transduction. *Science Signaling*, 4(189):rs8–rs8, 2011.

[311] Tomohiro Suzuki, Satoko Shinjo, Takatomo Arai, Mai Kanai, and Nobuhito Goda. Hypoxia and fatty liver. *World Journal of Gastroenterology: WJG*, 20(41):15087, 2014.

[312] Pi-Jung Hsiao, Tusty-Jiuan Hsieh, Kung-Kai Kuo, Wei-Wen Hung, Kun-Bow Tsai, Ching-Hsiu Yang, Ming-Lung Yu, and Shyi-Jang Shin. Pioglitazone retrieves hepatic antioxidant dna repair in a mice model of high fat diet. *BMC Molecular Biology*, 9(1):1–10, 2008.

[313] Erin K Daugherity, Gabriel Balmus, Ahmed Al Saei, Elizabeth S Moore, Delbert Abi Abdallah, Arlin B Rogers, Robert S Weiss, and Kirk J Maurer. The dna damage checkpoint protein atm promotes hepatocellular apoptosis and fibrosis in a mouse model of non-alcoholic fatty liver disease. *Cell Cycle*, 11(10):1918–1928, 2012.

[314] Icksoo Lee and Maik Hüttemann. Energy crisis: the role of oxidative phosphorylation in acute inflammation and sepsis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(9):1579–1586, 2014.

[315] Andreea-Manuela Mirea, Cees J Tack, Triantafyllos Chavakis, Leo AB Joosten, and Erik JM Toonen. Il-1 family cytokine pathways underlying nafld: towards new treatment strategies. *Trends in Molecular Medicine*, 24(5):458–471, 2018.

[316] Adil Mardinoglu, Rasmus Agren, Caroline Kampf, Anna Asplund, Mathias Uhlen, and Jens Nielsen. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nature communications*, 5(1):1–11, 2014.

[317] Dipanka Tanu Sarmah, Abhijit Paul, Umang Berry, Milan Surjit, Nandadulal Bairagi, and Samrat Chatterjee. De novo analysis of protein-protein interaction network reveals potential targets in nash. *(submitted)*.

[318] Dana Silverbush and Roded Sharan. A systematic approach to orient the human protein–protein interaction network. *Nature communications*, 10(1):1–9, 2019.

[319] Shivakumar Chitturi, Shehan Abeygunasekera, Geoffrey C Farrell, Jane Holmes-Walker, Jason M Hui, Caroline Fung, Rooshdiya Karim, Rita Lin, Dev Samarasinghe, Christopher Liddle, et al. Nash and insulin resistance: insulin hypersecretion and specific association with the insulin resistance syndrome. *Hepatology*, 35(2):373–379, 2002.

[320] Naim Alkhouri, Christine Carter-Kent, and Ariel E Feldstein. Apoptosis in nonalcoholic fatty liver disease: diagnostic and therapeutic implications. *Expert review of gastroenterology & hepatology*, 5(2):201–212, 2011.

[321] Katrine D Galsgaard, Jens Pedersen, Filip K Knop, Jens J Holst, and Nicolai J Wewer Albrechtsen. Glucagon receptor signaling and lipid metabolism. *Frontiers in physiology*, 10:413, 2019.

[322] Pi-Xiao Wang, Yan-Xiao Ji, Xiao-Jing Zhang, Ling-Ping Zhao, Zhen-Zhen Yan, Peng Zhang, Li-Jun Shen, Xia Yang, Jing Fang, Song Tian, et al. Targeting casp8 and fadd-like apoptosis regulator ameliorates nonalcoholic steatohepatitis in mice and nonhuman primates. *Nature Medicine*, 23(4):439–449, 2017.

[323] Franziska Wandrer, Stephanie Liebig, Silke Marhenke, Arndt Vogel, Katharina John, Michael P Manns, Andreas Teufel, Timo Itzel, Thomas Longerich, Olaf Maier, et al. Tnf-receptor-1 inhibition reduces liver steatosis, hepatocellular injury and fibrosis in nafld mice. *Cell Death & Disease*, 11(3):1–9, 2020.

[324] Zhi Zhang, Huiqing Wen, Bangjian Peng, Jun Weng, and Fanhong Zeng. Hfd-induced traf6 upregulation promotes liver cholesterol accumulation and fatty liver development via ezh2-mediated mir-429/pparα axis. *Molecular Therapy-Nucleic Acids*, 24:711–727, 2021.

[325] Zhibin Yan, Xiaokang Miao, Bangzhi Zhang, and Junqiu Xie. p53 as a double-edged sword in the progression of non-alcoholic fatty liver disease. *Life Sciences*, 215:64–72, 2018.

[326] Stefan Wuchty and Eivind Almaas. Peeling the yeast protein network. *Proteomics*, 5(2):444–449, 2005.

[327] Christopher M Yates and Michael JE Sternberg. Proteins and domains vary in their tolerance of non-synonymous single nucleotide polymorphisms (nssnps). *Journal of molecular biology*, 425(8):1274–1286, 2013.

[328] Dániel Bánky, Gábor Iván, and Vince Grolmusz. Equal opportunity for low-degree network nodes: a pagerank-based method for protein target identification in metabolic graphs. *PLoS One*, 8(1):e54204, 2013.

[329] Samjhana Thapaliya, Alexander Wree, Davide Povero, Maria Eugenia Inzaugarat, Michael Berk, Laura Dixon, Bettina G Papouchado, and Ariel E Feldstein. Caspase 3 inactivation protects against hepatic cell death and ameliorates fibrogenesis in a diet-induced nash model. *Digestive Diseases and Sciences*, 59(6):1197–1206, 2014.

[330] Miguel-Angel Barrios-Maya, Angélica Ruiz-Ramírez, Héctor Quezada, Carlos L Céspedes Acuña, and Mohammed El-Hafidi. Palmitoyl-coa effect on cytochrome c release, a key process of apoptosis, from liver mitochondria of rat with sucrose diet-induced obesity. *Food and Chemical Toxicology*, 154:112351, 2021.

[331] Kara Wegermann, Anna Mae Diehl, and Cynthia A Moylan. Disease pathways and molecular mechanisms of nonalcoholic steatohepatitis. *Clinical Liver Disease*, 11(4):87–91, 2018.

[332] Claire H Wilson and Sharad Kumar. Caspases in metabolic disease and their therapeutic potential. *Cell Death & Differentiation*, 25(6):1010–1024, 2018.

[333] Xiao-Ya Liu, Rui-Xia Liu, Fei Hou, Li-Jian Cui, Chun-Yun Li, Cheng Chi, Entong Yi, Yan Wen, and Cheng-Hong Yin. Fibronectin expression is critical for liver fibrogenesis in vivo and in vitro. *Molecular Medicine Reports*, 14(4):3669–3675, 2016.

[334] Koji Nishikawa, Yosuke Osawa, and Kiminori Kimura. Wnt/β-catenin signaling as a potential target for the treatment of liver cirrhosis using antifibrotic drugs. *International Journal of Molecular Sciences*, 19(10):3103, 2018.

[335] Theodoros Eleftheriadis, Georgios Pissas, Vassilios Liakopoulos, and Ioannis Stefanidis. Cytochrome c as a potentially clinical useful marker of mitochondrial and cellular damage. *Frontiers in Immunology*, 7:279, 2016.

[336] Zhi-Jun Jiang, Qing-Hua Shen, Hai-Yong Chen, Zhe Yang, Ming-Qi Shuai, and Shu-Sen Zheng. Galectin-1 gene silencing inhibits the activation and proliferation but induces the apoptosis of hepatic stellate cells from mice with liver fibrosis. *International Journal of Molecular Medicine*, 43(1):103–116, 2019.

[337] Janina Binici and Joachim Koch. Bag-6, a jack of all trades in health and disease. *Cellular and molecular life sciences*, 71(10):1829–1837, 2014.

[338] Yuji Iimuro, Toshihiro Nishio, Taisuke Morimoto, Takashi Nitta, Branko Stefanovic, Sung Kyu Choi, David A Brenner, and Yoshio Yamaoka. Delivery of matrix metalloproteinase-1 attenuates established liver fibrosis in the rat. *Gastroenterology*, 124(2):445–458, 2003.

[339] Salwa Sebti, Christine Prébois, Esther Pérez-Gracia, Chantal Bauvy, Fabienne Desmots, Nelly Pirot, Céline Gongora, Anne-Sophie Bach, Andrew V Hubberstey, Valérie Palissot, et al. Bag6/bat3 modulates autophagy by affecting ep300/p300 intracellular localization. *Autophagy*, 10(7):1341–1342, 2014.

[340] Yuanyuan Chu, Xingqi Dong, Yingjin Kang, Jingnan Liu, Tao Zhang, Cuiwei Yang, Zhangshun Wang, Wangchen Shen, Huanhuan Huo, Min Zhuang, et al. The chaperone bag6 regulates cellular homeostasis between autophagy and apoptosis by holding lc3b. *Iscience*, 23(11):101708, 2020.

[341] Leimarembi Devi Naorem, Mathavan Muthaiyan, and Amouda Venkatesan. Integrated network analysis and machine learning approach for the identification of key genes of triple-negative breast cancer. *Journal of Cellular Biochemistry*, 120(4):6154–6167, 2019.

[342] Md Rabiul Auwul, Md Rezanur Rahman, Esra Gov, Md Shahjaman, and Mohammad Ali Moni. Bioinformatics and machine learning approach identifies potential drug targets and pathways in covid-19. *Briefings in Bioinformatics*, 22(5):bbab120, 2021.

[343] Adil Mardinoglu, Elias Bjornson, Cheng Zhang, Martina Klevstig, Sanni Söderlund, Marcus Ståhlman, Martin Adiels, Antti Hakkarainen, Nina Lundbom, Murat Kilicarslan, et al. Personal model-assisted identification of nad+ and glutathione metabolism as intervention target in nafld. *Molecular Systems Biology*, 13(3):916, 2017.

[344] Matt Docherty, Stephane A Regnier, Gorana Capkun, Maria-Magdalena Balp, Qin Ye, Nico Janssens, Andreas Tietz, Jürgen Löffler, Jennifer Cai, Marcos C Pedrosa, et al. Development of a novel machine learning model to predict presence of nonalcoholic steatohepatitis. *Journal of the American Medical Informatics Association*, 28(6):1235–1241, 2021.

[345] Fabian Heinemann, Gerald Birk, and Birgit Stierstorfer. Deep learning enables pathologist-like scoring of nash models. *Scientific Reports*, 9(1):1–10, 2019.

[346] Nikolaos Perakakis, Stergios A Polyzos, Alireza Yazdani, Aleix Sala-Vila, Jannis Kountouras, Athanasios D Anastasilakis, and Christos S Mantzoros. Non-invasive diagnosis of non-alcoholic steatohepatitis and fibrosis with the use of omics and supervised learning: a proof of concept study. *Metabolism*, 101:154005, 2019.

[347] Rafael Garcia-Carretero, Luis Vigil-Medina, Oscar Barquero-Perez, and Javier Ramos-Lopez. Relevant features in nonalcoholic steatohepatitis determined using machine learning for feature selection. *Metabolic syndrome and related disorders*, 17(9):444–451, 2019.

[348] Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N Robinson. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008.

[349] Mathias Uhlén, Linn Fagerberg, Björn M Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt, Anna Asplund, et al. Tissue-based map of the human proteome. *Science*, 347(6220):1260419, 2015.

[350] Lin Li, YanShu Wang, Lifeng An, XiangYin Kong, and Tao Huang. A network-based method using a random walk with restart algorithm and screening tests to identify novel genes associated with menière's disease. *PLoS One*, 12(8):e0182592, 2017.

[351] ShaoPeng Wang, GuoHua Huang, Qinghua Hu, and Quan Zou. A network-based method for the identification of putative genes related to infertility. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1860(11):2716–2724, 2016.

[352] Lorena Pantano, George Agyapong, Yang Shen, Zhu Zhuo, Francesc Fernandez-Albert, Werner Rust, Dagmar Knebel, Jon Hill, Carine M Boustany-Kari, Julia F Doerner, et al. Molecular characterization and cell type composition deconvolution of fibrosis in nafld. *Scientific Reports*, 11(1):1–14, 2021.

[353] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.

[354] Mohan S Rao, Terry R Van Vleet, Rita Ciurlionis, Wayne R Buck, Scott W Mittelstadt, Eric AG Blomme, and Michael J Liguori. Comparison of rna-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. *Frontiers in genetics*, 9:636, 2019.

[355] Mark D Robinson and Alicia Oshlack. A scaling normalization method for differential expression analysis of rna-seq data. *Genome Biology*, 11(3):1–9, 2010.

[356] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[357] Burcu F Darst, Kristen C Malecki, and Corinne D Engelman. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, 19(1):1–6, 2018.

[358] Zahra Nematzadeh, Roliana Ibrahim, and Ali Selamat. Comparative studies on breast cancer classifications with k-fold cross validations using machine learning techniques. In *2015 10th Asian control conference (ASCC)*, pages 1–6. IEEE, 2015.

[359] Caroline Colijn, Aaron Brandes, Jeremy Zucker, Desmond S Lun, Brian Weiner, Maha R Farhat, Tan-Yun Cheng, D Branch Moody, Megan Murray, and James E Galagan. Interpreting expression data with metabolic flux models: predicting mycobacterium tuberculosis mycolic acid production. *PLoS Computational Biology*, 5(8):e1000489, 2009.

[360] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N Mendoza, Anne Richelle, Almut Heinken, Hulda S Haraldsdóttir, Jacek Wachowiak, Sarah M Keating, Vanja Vlasov, et al. Creation and analysis of biochemical constraint-based models using the cobra toolbox v. 3.0. *Nature Protocols*, 14(3):639–702, 2019.

[361] Keren Yizhak, Orshay Gabay, Haim Cohen, and Eytan Ruppin. Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nature communications*, 4(1):1–11, 2013.

[362] Daniel Segre, Dennis Vitkup, and George M Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, 2002.

[363] Zhuorui Xie, Allison Bailey, Maxim V Kuleshov, Daniel JB Clarke, John E Evangelista, Sherry L Jenkins, Alexander Lachmann, Megan L Wojciechowicz, Eryk Kropiwnicki, Kathleen M Jagodnik, et al. Gene set knowledge discovery with enrichr. *Current protocols*, 1(3):e90, 2021.

[364] Sharon L Freshour, Susanna Kiwala, Kelsy C Cotto, Adam C Coffman, Joshua F McMichael, Jonathan J Song, Malachi Griffith, Obi L Griffith, and Alex H Wagner. Integration of the drug–gene interaction database (dgidb 4.0) with open crowdsource efforts. *Nucleic Acids Research*, 49(D1):D1144–D1151, 2021.

[365] Alberto Valdeolivas, Laurent Tichit, Claire Navarro, Sophie Perrin, Gaelle Odelin, Nicolas Levy, Pierre Cau, Elisabeth Remy, and Anaïs Baudot. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics*, 35(3):497–505, 2019.

[366] Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, Sampo Pyysalo, Nadezhda T Doncheva, Marc Legeay, Tao Fang, Peer Bork, et al. The string database in 2021: customizable protein–protein networks, and functional

characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49(D1):D605–D612, 2021.

[367] Janet Piñero, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I Furlong. Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, page gkw943, 2016.

[368] Dipanka Tanu Sarmah, Nandadulal Bairagi, and Samrat Chatterjee. Tracing the footsteps of autophagy in computational biology. *Briefings in Bioinformatics*, 22(4):bbaa286, 2021.

[369] Anwar Ul Haq, Defu Zhang, He Peng, and Sami Ur Rahman. Combining multiple feature-ranking techniques and clustering of variables for feature selection. *IEEE Access*, 7:151482–151492, 2019.

[370] Abhijit Paul, Salman Azhar, Phonindra Nath Das, Nandadulal Bairagi, and Samrat Chatterjee. Elucidating the metabolic characteristics of pancreatic β-cells from patients with type 2 diabetes (t2d) using a genome-scale metabolic modeling. *Computers in Biology and Medicine*, 144:105365, 2022.

[371] Susan M Corley, Karen L MacKenzie, Annemiek Beverdam, Louise F Roddam, and Marc R Wilkins. Differentially expressed genes from rna-seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC genomics*, 18(1):1–13, 2017.

[372] Wei-Feng Guo, Shao-Wu Zhang, Ze-Gang Wei, Tao Zeng, Fei Liu, Jingsong Zhang, Fang-Xiang Wu, and Luonan Chen. Constrained target controllability of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(6):063402, 2017.

[373] Mira Ham, Joo-Won Lee, A Hyun Choi, Hagoon Jang, Goun Choi, Jiyoung Park, Chisayo Kozuka, Dorothy D Sears, Hiroaki Masuzaki, and Jae Bum Kim. Macrophage glucose-6-phosphate dehydrogenase stimulates proinflammatory responses with oxidative stress. *Molecular and cellular biology*, 33(12):2425–2435, 2013.

[374] Sílvia S Chambel, Andreia Santos-Gonçalves, and Tiago L Duarte. The dual role of nrf2 in nonalcoholic fatty liver disease: regulation of antioxidant defenses and hepatic lipid metabolism. *BioMed research international*, 2015, 2015.

[375] Fuming Li, Peiwei Huangyang, Michelle Burrows, Kathy Guo, Romain Riscal, Jason Godfrey, Kyoung Eun Lee, Nan Lin, Pearl Lee, Ian A Blair, et al. Fbp1 loss disrupts liver metabolism and promotes tumorigenesis through a hepatic stellate cell senescence secretome. *Nature cell biology*, 22(6):728–739, 2020.

[376] Magali Gorce, Elise Lebigot, Alina Arion, Anaïs Brassier, Aline Cano, Pascale De Lonlay, François Feillet, Claire Gay, François Labarthe, Marie-Cécile Nassogne, et al. Fructose-1, 6-bisphosphatase deficiency causes fatty liver disease and requires long-term hepatic follow-up. *Journal of Inherited Metabolic Disease*, 2021.

[377] Abhijit Paul, Rajat Anand, Sonali Porey Karmakar, Surender Rawat, Nandadulal Bairagi, and Samrat Chatterjee. Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. *Scientific Reports*, 11(1):1–13, 2021.

[378] Justin Lamb. The connectivity map: a new tool for biomedical research. *Nature reviews cancer*, 7(1):54–60, 2007.

[379] Alex H Wagner, Adam C Coffman, Benjamin J Ainscough, Nicholas C Spies, Zachary L Skidmore, Katie M Campbell, Kilannin Krysiak, Deng Pan, Joshua F McMichael, James M Eldred, et al. Dgidb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Research*, 44(D1):D1036–D1044, 2016.

[380] David S Wishart, Craig Knox, An Chi Guo, Savita Shrivastava, Murtaza Hassanali, Paul Stothard, Zhan Chang, and Jennifer Woolsey. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(suppl_1):D668–D672, 2006.

[381] Desheng Lu, Jerry X Liu, Tomoyuki Endo, Haowen Zhou, Shiyin Yao, Karl Willert, Ingo GH Schmidt-Wolf, Thomas J Kipps, and Dennis A Carson. Ethacrynic acid exhibits selective toxicity to chronic lymphocytic leukemia cells by inhibition of the wnt/β-catenin pathway. *PloS one*, 4(12):e8294, 2009.

[382] Michael G Rolf, Jon O Curwen, Margaret Veldman-Jones, Cath Eberlein, Jianyan Wang, Alex Harmer, Caroline J Hellawell, and Martin Braddock. In vitro pharmacological profiling of r406 identifies molecular targets underlying the clinical effects of fostamatinib. *Pharmacology research & perspectives*, 3(5):e00175, 2015.

[383] David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, et al. Chembl: towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1):D930–D940, 2019.

[384] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, et al. Pubchem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, 49(D1):D1388–D1395, 2021.

[385] Yamin Jie, Guijun Liu, E Mingyan, Ying Li, Guo Xu, Jingjing Guo, Yinyin Li, Guanghua Rong, Yongwu Li, and Anxin Gu. Novel small molecule inhibitors of the transcription factor ets-1 and their antitumor activity against hepatocellular carcinoma. *European Journal of Pharmacology*, 906:174214, 2021.

[386] E Sanchez-Tillo, L Fanlo, L Siles, S Montes-Moreno, A Moros, G Chiva-Blanch, R Estruch, A Martinez, D Colomer, Balázs Győrffy, et al. The emt activator zeb1 promotes tumor growth and determines differential response to chemotherapy in mantle cell lymphoma. *Cell Death & Differentiation*, 21(2):247–257, 2014.

[387] Rui Benfeitas, Gholamreza Bidkhori, Bani Mukhopadhyay, Martina Klevstig, Muhammad Arif, Cheng Zhang, Sunjae Lee, Resat Cinar, Jens Nielsen, Mathias Uhlen, et al. Characterization of heterogeneous redox responses in hepatocellular carcinoma patients using network analysis. *EBioMedicine*, 40:471–487, 2019.

[388] Alberto Noronha, Jennifer Modamio, Yohan Jarosz, Elisabeth Guerard, Nicolas Sompairac, German Preciat, Anna Dröfn Daníelsdóttir, Max Krecke, Diane Merten, Hulda S Haraldsdóttir, et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1):D614–D624, 2019.

[389] Lunxu Li, Shilin Xia, Xueying Shi, Xu Chen, and Dong Shang. The novel immune-related genes predict the prognosis of patients with hepatocellular carcinoma. *Scientific Reports*, 11(1):1–14, 2021.

[390] Xiu Zhang, Dan Wu, Mohanad Aldarouish, Xiaodong Yin, Chunyan Li, and Cailian Wang. Ets-1: A potential target of glycolysis for metabolic therapy by regulating glucose metabolism in pancreatic cancer. *International Journal of Oncology*, 50(1):232–240, 2017.

[391] ILKe Nalbantoglu and Elizabeth M Brunt. Role of liver biopsy in nonalcoholic fatty liver disease. *World journal of Gastroenterology: WJG*, 20(27):9026, 2014.

[392] Manu V Chakravarthy and Brent A Neuschwander-Tetri. The metabolic basis of nonalcoholic steatohepatitis. *Endocrinology, Diabetes & Metabolism*, 3(4):e00112, 2020.

[393] Klementina Fon Tacer and Damjana Rozman. Nonalcoholic fatty liver disease: focus on lipoprotein and lipid deregulation. *Journal of Lipids*, 2011, 2011.

[394] Osamu Tetsu and Frank McCormick. Ets-targeted therapy: can it substitute for mek inhibitors? *Clinical and Translational Medicine*, 6(1):1–9, 2017.

[395] J Kaur and K Tikoo. Ets1 identified as a novel molecular target of rna aptamer selected against metastatic cells for targeted delivery of nano-formulation. *Oncogene*, 34(41):5216–5228, 2015.

[396] Dechen Liu, Kai Wang, Kai Li, Rufeng Xu, Xiaoai Chang, Yunxia Zhu, Peng Sun, and Xiao Han. Ets-1 deficiency alleviates nonalcoholic steatohepatitis via weakening tgf-β1 signaling-mediated hepatocyte apoptosis. *Cell Death & Disease*, 10(6):1–14, 2019.

[397] Juan Galceran, Claudio Sustmann, Shu-Chi Hsu, Stephanie Folberth, and Rudolf Grosschedl. Lef1-mediated regulation of delta-like1 links wnt and notch signaling in somitogenesis. *Genes & Development*, 18(22):2718–2723, 2004.

[398] Yulia A Nevzorova, Wei Hu, Francisco J Cubero, Ute Haas, Julia Freimuth, Frank Tacke, Christian Trautwein, and Christian Liedtke. Overexpression of c-myc in hepatocytes promotes activation of hepatic stellate cells and facilitates the onset of liver fibrosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1832(10):1765–1775, 2013.

[399] Kelley G Núñez, Janet Gonzalez-Rosario, Paul T Thevenot, and Ari J Cohen. Cyclin d1 in the liver: Role of noncanonical signaling in liver steatosis and hormone regulation. *Ochsner Journal*, 17(1):56–65, 2017.

[400] Zhanyu Wang and Chenfang Dong. Gluconeogenesis in cancer: function and regulation of pepck, fbpase, and g6pase. *Trends in Cancer*, 5(1):30–45, 2019.

[401] Bramanandam Manavathi, Suresh K Rayala, and Rakesh Kumar. Phosphorylation-dependent regulation of stability and transforming potential of ets transcriptional factor ese-1 by p21-activated kinase 1. *Journal of Biological Chemistry*, 282(27):19820–19830, 2007.

[402] G Zhu, Y Wang, B Huang, J Liang, Y Ding, A Xu, and W Wu. A rac1/pak1 cascade controls β-catenin activation in colon cancer cells. *Oncogene*, 31(8):1001–1012, 2012.

[403] B Maroto, MB Ye, K Von Lohneysen, A Schnelzer, and UG Knaus. P21-activated kinase is required for mitotic progression and regulates plk1. *Oncogene*, 27(36):4900–4908, 2008.

[404] Anne Loft, Ana Jimena Alfaro, Søren Fisker Schmidt, Felix Boel Pedersen, Mike Krogh Terkelsen, Michele Puglia, Kan Kau Chow, Annette Feuchtinger, Maria Troullinaki, Adriano Maida, et al. Liver-fibrosis-activated transcriptional networks govern hepatocyte reprogramming and intra-hepatic communication. *Cell Metabolism*, 33(8):1685–1700, 2021.

[405] Christina Fitzmaurice, Christine Allen, Ryan M Barber, Lars Barregard, Zulfiqar A Bhutta, Hermann Brenner, Daniel J Dicker, Odgerel Chimed-Orchir, Rakhi Dandona, Lalit Dandona, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA oncology*, 3(4):524–548, 2017.

[406] Elisabeth Smolle and Martin Pichler. Non-smoking-associated lung cancer: A distinct entity in terms of tumor biology, patient characteristics and impact of hereditary cancer predisposition. *Cancers*, 11(2):204, 2019.

[407] Cecilia Zappa and Shaker A Mousa. Non-small cell lung cancer: current treatment and future advances. *Translational lung cancer research*, 5(3):288, 2016.

[408] John A Howington, Matthew G Blum, Andrew C Chang, Alex A Balekian, and Sudish C Murthy. Treatment of stage i and ii non-small cell lung cancer: Diagnosis and management of lung cancer: American college of chest physicians evidence-based clinical practice guidelines. *Chest*, 143(5):e278S–e313S, 2013.

[409] Arya Amini, Norman Yeh, Laurie E Gaspar, Brian Kavanagh, and Sana D Karam. Stereotactic body radiation therapy (sbrt) for lung cancer patients previously treated with conventional radiotherapy: a review. *Radiation oncology*, 9(1):210, 2014.

[410] Kenichi Suda and Tetsuya Mitsudomi. Successes and limitations of targeted cancer therapy in lung cancer. In *Successes and Limitations of Targeted Cancer Therapy*, volume 41, pages 62–77. Karger Publishers, 2014.

[411] Beth Levine and John Abrams. p53: The janus of autophagy? *Nature cell biology*, 10(6):637, 2008.

[412] Peter GH Clarke. Developmental cell death: morphological diversity and multiple mechanisms. *Anatomy and embryology*, 181(3):195–213, 1990.

[413] Guido Kroemer and Beth Levine. Autophagic cell death: the story of a misnomer. *Nature reviews Molecular cell biology*, 9(12):1004–1010, 2008.

[414] Aristides G Eliopoulos, Sophia Havaki, and Vassilis G Gorgoulis. Dna damage response and autophagy: a meaningful partnership. *Frontiers in Genetics*, 7:204, 2016.

[415] Slavica Tudzarova, Paul Mulholland, Ayona Dey, Kai Stoeber, Andrei L Okorokov, and Gareth H Williams. p53 controls cdc7 levels to reinforce g1 cell cycle arrest upon genotoxic stress. *Cell Cycle*, 15(21):2958–2972, 2016.

[416] Amanda S Coutts, Cassandra J Adams, and Nicholas B La Thangue. p53 ubiquitination by mdm2: a never ending tail? *DNA Repair*, 8(4):483–490, 2009.

[417] Ygal Haupt, Ruth Maya, Anat Kazaz, and Moshe Oren. Mdm2 promotes the rapid degradation of p53. *Nature*, 387(6630):296, 1997.

[418] Ute M Moll and Oleksi Petrenko. The mdm2-p53 interaction. *Molecular Cancer Research*, 1(14):1001–1008, 2003.

[419] Michael HG Kubbutat, Stephen N Jones, and Karen H Vousden. Regulation of p53 stability by mdm2. *Nature*, 387(6630):299–303, 1997.

[420] Jonathan D Oliner, Jennifer A Pietenpol, Sam Thiagalingam, Jeno Gyuris, Kenneth W Kinzler, and Bert Vogelstein. Oncoprotein mdm2 conceals the activation domain of tumour suppressor p53. *Nature*, 362(6423):857–860, 1993.

[421] Chuangui Wang, Alexey Ivanov, Lihong Chen, William J Fredericks, Ed Seto, Frank J Rauscher, and Jiandong Chen. Mdm2 interaction with nuclear corepressor kap1 contributes to p53 inactivation. *The EMBO Journal*, 24(18):3279–3290, 2005.

[422] Karen H Vousden and Carol Prives. Blinded by the light: the growing complexity of p53. *Cell*, 137(3):413–431, 2009.

[423] Jijie Gu, Hidehiko Kawai, Linghu Nie, Hiroyuki Kitao, Dmitri Wiederschain, Aart G Jochemsen, John Parant, Guillermina Lozano, and Zhi-Min Yuan. Mutual dependence of mdm2 and mdmx in their functional inactivation of p53. *Journal of Biological Chemistry*, 277(22):19251–19254, 2002.

[424] Stjepan Uldrijan, Willem-Jan Pannekoek, and Karen H Vousden. An essential function of the extreme c-terminus of mdm2 can be provided by mdmx. *The EMBO Journal*, 26(1):102–112, 2007.

[425] Miriam Shadfan, Vanessa Lopez-Pajares, and Zhi-Min Yuan. Mdm2 and mdmx: Alone and together in regulation of p53. *Translational Cancer Research*, 1(2):88, 2012.

[426] R Kang, HJ Zeh, MT Lotze, and D Tang. The beclin 1 network regulates autophagy and apoptosis. *Cell Death & Differentiation*, 18(4):571–580, 2011.

[427] Deyi Zhang, Wei Wang, Xiujie Sun, Daqian Xu, Chenyao Wang, Qian Zhang, Huafei Wang, Wenwen Luo, Yan Chen, Huaiyong Chen, et al. Ampk regulates autophagy by phosphorylating becn1 at threonine 388. *Autophagy*, 12(9):1447–1459, 2016.

[428] Minsu Jang, Rackhyun Park, Hyunju Kim, Sim Namkoong, Daum Jo, Yang Hoon Huh, Ik-Soon Jang, Jin I Lee, and Junsoo Park. Ampk contributes to autophagosome maturation and lysosomal fusion. *Scientific Reports*, 8(1):1–10, 2018.

[429] Xuefei Li, Xiaorong Hu, Jichun Wang, Weipan Xu, Chunfeng Yi, Ruisong Ma, and Hong Jiang. Inhibition of autophagy via activation of pi3k/akt/mtor pathway contributes to the protection of hesperidin against myocardial ischemia/reperfusion injury. *International Journal of Molecular Medicine*, 42(4):1917–1924, 2018.

[430] Rebecca T Marquez and Liang Xu. Bcl-2: Beclin 1 complex: multiple, mechanisms regulating autophagy/apoptosis toggle switch. *American Journal of Cancer Research*, 2(2):214, 2012.

[431] Bing Liu, Zoltán N Oltvai, Hülya Bayır, Gary A Silverman, Stephen C Pak, David H Perlmutter, and Ivet Bahar. Quantitative assessment of cell fate decision between autophagy and apoptosis. *Scientific Reports*, 7(1):1–14, 2017.

[432] MT Hemann and SW Lowe. The p53–bcl-2 connection, 2006.

[433] Ruth Lev Bar-Or, Ruth Maya, Lee A Segel, Uri Alon, Arnold J Levine, and Moshe Oren. Generation of oscillations by the p53-mdm2 feedback loop: a theoretical and experimental study. *Proceedings of the National Academy of Sciences*, 97(21):11250–11255, 2000.

[434] Robert Stad, Natalie A Little, Dimitris P Xirodimas, Ruth Frenk, Alex J van der Eb, David P Lane, Mark K Saville, and Aart G Jochemsen. Mdmx stabilizes p53 and mdm2 via two distinct mechanisms. *EMBO Reports*, 2(11):1029–1034, 2001.

[435] Lan Ma, John Wagner, John Jeremy Rice, Wenwei Hu, Arnold J Levine, and Gustavo A Stolovitzky. A plausible model for the digital response of p53 to dna damage. *Proceedings of the National Academy of Sciences*, 102(40):14266–14271, 2005.

[436] Ket Hing Chong, Sandhya Samarasinghe, and Don Kulasiri. Mathematical modelling of p53 basal dynamics and dna damage response. *Mathematical Biosciences*, 259:27–42, 2015.

[437] Mitio Nagumo. Über die lage der integralkurven gewöhnlicher differentialgleichungen. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 24:551–559, 1942.

[438] Quan Liu, Jianjun Wang, Yongcheng Pan, Lifeng Meng, and Xi Zhan. Expression of beclinl in non small cell lung cancer and its clinical significance. *Journal of Nanjing Medical University*, 22(2):121–124, 2008.

[439] Xiao-Peng Zhang, Feng Liu, and Wei Wang. Two-phase dynamics of p53 in the dna damage response. *Proceedings of the National Academy of Sciences*, 108(22):8990–8995, 2011.

[440] Fernanda Antunes, Adolfo Garcia Erustes, Angélica Jardim Costa, Ana Carolina Nascimento, Claudia Bincoletto, Rodrigo Portes Ureshino, Gustavo José Silva Pereira, and Soraya Soubhi Smaili. Autophagy and intermittent fasting: the connection for cancer therapy? *Clinics*, 73, 2018.

[441] Xueping Qu, Jie Yu, Govind Bhagat, Norihiko Furuya, Hanina Hibshoosh, Andrea Troxel, Jeffrey Rosen, Eeva-Liisa Eskelinen, Noboru Mizushima, Yoshinori Ohsumi, et al. Promotion of tumorigenesis by heterozygous disruption of the beclin 1 autophagy gene. *The Journal of Clinical Investigation*, 112(12):1809–1820, 2003.

[442] Dong-Wei Xu, Guan-Qing Zhang, Zong-Wei Wang, Xiao-Yin Xu, and Tong-Xiang Liu. Autophagy in tumorigenesis and cancer treatment. *Asian Pac J Cancer Prev*, 16(6):2167–2175, 2015.

[443] Michael B Kastan, Onyinye Onyekwere, David Sidransky, Bert Vogelstein, and Ruth W Craig. Participation of p53 protein in the cellular response to dna damage. *Cancer Research*, 51(23 Part 1):6304–6311, 1991.

[444] O Surova and B Zhivotovsky. Various modes of cell death induced by dna damage. *Oncogene*, 32(33):3789–3797, 2013.

[445] Linda J Ko and Carol Prives. p53: puzzle and paradigm. *Genes & Development*, 10(9):1054–1072, 1996.

[446] Alexander Loewer, Eric Batchelor, Giorgio Gaglia, and Galit Lahav. Basal dynamics of p53 reveal transcriptionally attenuated pulses in cycling cells. *Cell*, 142(1):89–100, 2010.

[447] Domenico Migliorini, Eros Lazzerini Denchi, Davide Danovi, Aart Jochemsen, Manuela Capillo, Alberto Gobbi, Kristian Helin, Pier Giuseppe Pelicci, and Jean-Christophe Marine. Mdm4 (mdmx) regulates p53-induced growth arrest and neuronal cell death during early embryonic mouse development. *Molecular and Cellular Biology*, 22(15):5527–5538, 2002.

[448] Dongsheng Pei, Yanping Zhang, and Junnian Zheng. Regulation of p53: a collaboration between mdm2 and mdmx. *Oncotarget*, 3(3):228, 2012.

[449] Eileen White. The role for autophagy in cancer. *The Journal of Clinical Investigation*, 125(1):42–46, 2015.

[450] Meruna Nagata, Satoko Arakawa, Hirofumi Yamaguchi, Satoru Torii, Hazuki Endo, Masatsune Tsujioka, Shinya Honda, Yuya Nishida, Akimitsu Konishi, and Shigeomi Shimizu. Dram1 regulates dna damage-induced alternative autophagy. *Cell stress*, 2(3):55, 2018.

[451] Daolin Tang, Rui Kang, Kristen M Livesey, Chun-Wei Cheh, Adam Farkas, Patricia Loughran, George Hoppe, Marco E Bianchi, Kevin J Tracey, Herbert J Zeh III, et al. Endogenous hmgb1 regulates autophagy. *Journal of Cell Biology*, 190(5):881–892, 2010.

List of publications related to thesis

Published

1. Sarmah DT, Bairagi N, Chatterjee S. Tracing the footsteps of autophagy in computational biology. *Briefings in Bioinformatics.* (2021).

2. Sarmah DT, Bairagi N, Chatterjee S. The interplay between DNA damage and autophagy in lung cancer: A mathematical study. *Biosystems* (2021): 104443.

Communicated

1. Sarmah DT, Bairagi N, Chatterjee S. Identification of critical autophagy-related proteins in diabetic retinopathy: A multi-dimensional computational study. *Gene* (revised version submitted)

2. Sarmah DT*, Paul A*, Bairagi N, Chatterjee S. De novo analysis of metabolic and protein interaction networks reveals potential targets in non-alcoholic steatohepatitis. (*Equal Contribution)

3. Sarmah DT, Kumar S, Paul A, Bairagi N, Chatterjee S. A data-driven multilayer approach for identification of potential therapeutic targets in non-alcoholic steatohepatitis.

List of Other Publications

Published

1. Anand, R., **Sarmah, D. T.,** & Chatterjee, S. (2018). Extracting proteins involved in disease progression using temporally connected networks. BMC systems biology, 12(1), 1-13.

2. Kumar, P., Soory, A., Mustfa, S. A., **Sarmah, D. T.**, Devvanshi, H., Chatterjee, S., ... & Srikanth, C. V. (2022). Bidirectional regulation between AP-1 and SUMOylation pathway genes modulates inflammatory signaling during Salmonella infection. Journal of Cell Science, 135(16), jcs260096.

Communicated

1. Kumar S*, **Sarmah D,T.***, Asthana S, Chatterjee S. konnect2prot: A web application to explore the protein properties in a functional protein-protein interaction network. (*Equal Contribution)

2. **Sarmah D,T.**, Parveen R, Kundu J, Chatterjee S. Latent tuberculosis and computational biology: A less-talked affair

3. Kumar S, **Sarmah D, T.**, Paul A, Chatterjee S. Differential co-expression analysis of glioma identifies potential targets responsible for the progression of low grade glioma to glioblastoma multiforme

4. **Sarmah D,T.***, Kumar S*, Bairagi N, Chatterjee S. Dissecting big RNA-Seq cancer data using machine learning to find disease-associated genes and the causal mechanism. (*Equal Contribution)

Conferences

Poster Presentations

1. Indian Science Congress, January 2019, at Lovely Professional University.
2. International conference on Multiscale simulation and mathematical modelling of complex biological systems, January 2019, at Jawaharlal Nehru University, New Delhi.
3. Interdisciplinary Science conference on big data and computational biology, October, 2019, at Jamia Milia Islamia, New Delhi.
4. One day symposium on Metabolic Associated Fatty Liver Disease, June 2022, at Translational Health Science and Technological Institute, Faridabad. (**Best poster award**)

Oral Talk

1. Oral presentation on "Extracting proteins involved in disease progression using temporally connected networks" at International conference on Multiscale simulation and mathematical modelling of complex biological systems, January 2019, at Jamia Milia Islamia, New Delhi.