CONSTRUCTION OF MICRO RNA BASED GENE CO-EXPRESSION NETWORK AND ANALYSIS OF MICRO RNA-TARGET RAB SUB-NETWORKS IN HUMAN RAB ASSOCIATED DISORDERS- A NETWORK BIOLOGY APPROACH

By

Sayantoni Chaudhuri

Supervised by Dr Rajeev Mishra

Master of Science in Bioinformatics

Banaras Hindu University, Varanasi-221005, India

July 2020

CONSTRUCTION OF MICRO RNA BASED GENE CO-EXPRESSION NETWORK AND ANALYSIS OF MICRO RNA-TARGET RAB SUB-NETWORKS IN HUMAN RAB ASSOCIATED DISORDERS- A NETWORK BIOLOGY APPROACH

A senior thesis submitted in partial fulfilment of the requirements for the award of the degree of

Master of Science in Bioinformatics

By

Sayantoni Chaudhuri

Supervised by Dr Rajeev Mishra

Banaras Hindu University, Varanasi-221005, India

July 2020

Certificate

This is to certify that the research work embodied in this dissertation titled ***"Construction of microRNA based gene co-expression networks and analysis of microRNA-target Rab sub-networks in human Rab-associated disorders- a network biology approach"*** has been carried out by **Ms. Sayantoni Chaudhuri** under the supervision of **Dr. Rajeev Mishra** as a partial fulfillment of the M.Sc. degree in Bioinformatics at the Department of Bioinformatics, Banaras Hindu University. This work is original and no part of this thesis has been submitted for the award of any other degree to any other university.

Dr Rajeev Mishra (supervisor)

Coordinator and Assistant Professor

Department of Bioinformatics

Varanasi-221005

July 2020

Declaration

I hereby declare that this thesis work contained herein is my own, except where stated explicitly otherwise, in the text, in the form of references, adapted text, or adapted diagrams. This work is original and no part of this thesis has been submitted for the award of any other degree to any other university.

Sayantoni Chaudhuri

Varanasi-221005

July 2020

Dedicated to Science

Acknowledgments

Although my name appears on the cover of this dissertation, a great many people have contributed directly as well as indirectly to its production. I'd like to thank each of them because the work in itself has been quite an experience. For starters, this work would not come to life without the support of my supervisor Dr. Rajeev Mishra, Co-ordinator and Assistant professor, Department of Bioinformatics. Since the inception of this work, he has been a constant support and I take this opportunity to express my heartfelt gratitude to him for his exemplary guidance and encouragement throughout. It has indeed been an honor to work under his guidance.

I would like to thank my parents and friends and all those who were a part of my life, who stood by me in these two years and also through the course of my project. Special thanks to Pooja, for sticking through. This project would not have been complete without your support and wishes. In general, I would like to thank all those who love me, and those who hate me as well (since hate is a form of concern, so thanks for keeping me in your mind!).

Since this work is an extension of Ms. Nikita Poddar, my senior, I would like to thank her, especially because, without her, this work would not have existed. Also, I would like to thank all research scholars for the several interactive discussions that have played a major role in shaping up of this work. I would also like to thank the staff members for providing me with all department facilities and their co-operation.

In general, I would like to thank each of you who contributed to the making of this project and for making my life an adventure in these two years. That being said, life has been fun of course, and I've had the opportunity to savor every bit of it, philosophically and through science.

Foreword

It is a great opportunity for me to submit this booklet in the accomplishment of a degree in Bioinformatics, subject to limitation of time, a global pandemic towards the end, and effort - it's really never enough. However, the journey had been memorable, mostly because a fancy degree in an emerging field had been at the back of my mind all this while. This thesis had been written down assuming a crowd comprising essentially biologists, however, I have made efforts to minimalize scientific jargon, and explaining them in brief wherever necessary, since it occurred to me that readers might perhaps not be biology graduates, may not be in touch with the finest details of the descriptors of this thesis or just happen to come across this work and started reading it purely out of interest (I encourage you to do so, in that case!). Following that, I have color-coded figures and diagrams if a quick reference is required anywhere.

I have broken down the entire project into 15 chapters-starting from an introduction that explains what are we going to talk about in this project, and why, finishing off with a list of all the books, articles, reviews, research papers, and manuscripts I have drawn references from. If you think it's a bit too detailed, or you don't have much time in your hands, and if you do actually remember a lot of biology- you may as well skip to the results section where I have laid out the results I have obtained from all my experiments and calculations. I'm sure that will be an interesting read as well!

Abstract

With the increasing complexity of interaction data being generated, diseases can better be understood in the context of network principles. This paper is an attempt to study the effect of aberrant expression and mutational consequences of the Rab GTPase family, on various Rab-associated disorders, and how the impact of defective genes can branch out along the links of the network and alter the activity of others via network biology. In modern-day biomedical research, multiple types of small RNAs like microRNAs have emerged to play a substantial role in biomarker discovery and potential therapeutic targets. Since the role of miRNAs in developmental and pathological processes is substantial, and given that miRNAs are found both in intracellular and extracellular level, miRNA-based studies can help us understand the pathobiology of Rab associated disorders, interconnecting these disease-causing Rabs to the miRNAs that regulate them at a cellular level.

Keywords

Rabs, network biology, miRNA, Rab associated diseases, enrichment analysis, network analysis

List of Figures

List of tables

List of formulae

$$Rab\ occupancy = \frac{Number\ of\ Rabs\ including\ background\ Rabs\ each\ disease}{Total\ affiliated\ genes\ in\ the\ same}\ X\ 100$$

$$C_i = deg(i)$$

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

$$C_i = \frac{2e}{k(k-1)}$$

$$C_{clo} = \frac{1}{\sum dist_{ij}}$$

$$C_{bet(i)} = \frac{\sigma_{xy}(i)}{\sigma_{xy}}$$

List of abbreviations

mRNA	messenger RNA
TAP	Tandem Affinity Purification
NMR	Nuclear Magnetic Resonance
Y2H	Yeast Two-Hybrid
RNA	Ribonucleic acid
miRNA	microRNA
AD	Alzheimer's disease
PD	Parkinson's disease
FTD	frontotemporal dementia
GTP	Guanine Triphosphate
GDP	Guanine Diphosphate
GEO	Gene Expression Omnibus
GDA	Gene-Disease Association
VDA	Variant-Disease Association
KEGG	Kyoto Encyclopedia of Genes and Genomes
CDS	Coding Sequence
UTR	Untranslated regions
qRT-PCR	Quantitative Reverse Transcription PCR
RNA-seq	RNA Sequencing
DSI	Disease Specificity Index
RAB	Ras associated in binding
RAD	Rab associated diseases
EI	Evidence Index
PMIDs	PubMed IDs
GEFs	Guanine nucleotide Exchange Factors
GAPs	GTPase Activating Proteins
GDIs	Guanine nucleotide Dissociation Inhibitors
REP	Rab Escort Protein
cAMP	cyclic Adenosine Monophosphate
Pol	Polymerase
pri-miRNA	Primary miRNA
dsRNA	double-stranded RNA
bp	base pairs
RISC	RNA-induced Silencing Complex
Et al	et alia
graphML	graph Modelling Language
Fig.	Figure
BCC	Bi-connected Components
FDR	False Discovery Rate
DPI	Disease Pleiotropy Index
p(LI)	probability(Loss of function Intolerant)
GDAS	Gene-Disease Association Score

Table of Contents

1. Introduction

Cellular components seldom act as single entities; rather they exert their overall functions by interactions with other cellular components [1]. This is evidenced by the analysis of protein annotations, which reveals that similarly annotated proteins often have similar functions [2]. Cellular functions can, therefore, be understood as an outcome of tightly regulated inter and intracellular connectivity. These interactions can range from local interactions to interactions spanning across different cells altogether, and have been broadly studied as protein-protein interaction networks, gene regulatory networks, signaling and transcription-regulatory networks, and metabolic pathways. In a human cell comprising roughly ~25000 protein-coding genes, ~ 1000 metabolites, and an increasing number of proteins and functional mRNAs, we can imagine ~1,00,000 interactions [3] and perhaps even more, with ~2000 human disorders [4]. Given the fact that most cellular components are parts of such interactions, it can be hypothesized that a disease phenotype is seldom a result of modifications or alterations of discrete genetic elements, but a complex 'interplay' of genes, proteins and metabolites, and several other elements [3].

The bulk of research so far has been deriving protein-protein interaction data from model organisms with the help of various techniques. In vitro techniques include TAP tagging, affinity chromatography, co-immunoprecipitation, microarrays, protein fragment complementation assays, and finally, X-ray crystallography and NMR spectroscopy to allow visualization of protein structures with their interacting partners at atomic levels. Popular in vivo techniques are Y2H, bioluminescence resonance energy transfer, fluorescence resonance energy transfer, and bimolecular fluorescence complementation [5]. Despite being successful, some of these popular high throughput methods have shown discrepancies, mostly due to the generation of false positives [2]. Nevertheless, with the growing enormous data being generated, which includes a substantial rise in human-specific molecular interaction data, [3] the past decade has attempted to pair it up with network methods to further unravel protein functions, disease-gene and disease-disease association [6]. MiRNAs are a group of small non-coding RNAs, ~22 nucleotides in length, functioning as the guide molecules in RNA silencing [7], by forming base pairs with target mRNAs. The current model of miRNA biogenesis gives that miRNAs belong to class II genes and are transcribed by RNA polymerase II as long primary transcripts [8]. MiRNAs have been reported to be found at an intracellular level as well as circulating in extracellular fluids like blood, urine, saliva, peritoneal fluid, amniotic fluid, bronchial lavage, cerebrospinal fluid, and tears [9], [10], [11] thus serving as potential biomarkers and therapeutic targets in several cancers [12] including exosomal miRNAs in prostate cancer [13], circulating miRNAs in various neurodegenerative

diseases like Alzheimer's disease (AD), Parkinson's disease (PD) and frontotemporal dementia (FTD) [14], cardiovascular diseases like myocardial infarction [15], [16] and, monitoring the progress of early stages of diabetic nephropathy [17] In addition, miRNAs are highly stable, making them ideal biomarkers in modern-day medical research [18], [19] [20] [21].

Rab GTPases belong to the Ras superfamily of small GTPases and play a central role in vesicle budding, motility, and fusion [22]. Rab GTPases impart their functions by switching between two interconvertible states, the GDP bound 'inactive' and the GTP-bound 'active' form, allowing them to be involved in a wide range of cellular functions [23] due to their ability to act as scaffolds for recruitment of various effector molecules [24]. Recent analyses indicate ~60 different Rabs [25], and further studies have shown 70 Rabs, subdivided into 44 families [26]. Evidences of a few more Rabs were proposed, making a total of ~72 Rabs to date [27]. Our study has been made including 72 Rabs. Rabs and their associated effector molecules are shown to be dysregulated in various human diseases like Griscelli Syndrome, Charcot–Marie–Tooth type 2B neuropathy, Choroideremia, X-linked mental retardation, various cancers, and neurodegenerative diseases.

This study specifically deals with a wide number of Rab associated diseases, tabulated, later on, intending to predict how miRNAs, targeting Rab molecules alter their expression. Most human and animal miRNAs are evolutionarily conserved and possess conserved interactions with mRNAs [28]. For instance, miR-5100, has been shown to suppress tumor growth in lung cancer by targeting Rab6 [29], miR-15b-5p has been reported to target Rab1a in hepatocellular carcinoma [30]. Our study involves the prediction of all miRNAs targets for Rabs in various Rab-associated diseases and a computational network biology approach to analyze them. The diseases have been grouped into 12 classes namely: cancers, neurodegenerative diseases, musculoskeletal disorders, infectious diseases, immunological disorders, hematological diseases, endocrine and reproductive disorders, endocrine diseases, endocrine and metabolic diseases, congenital malformation, chromosomal abnormalities, and cardiovascular diseases (Elucidated in later sections) and miRNA based gene co-expression network models have been proposed for the same, to elucidate how miRNAs target genes, and specifically members of the Rab family of GTPases.

2. Dissertation outline

1. *Retrieval of all Rab associated diseases:*
   Eighty-seven diseases were found out that were associated with various members of the RAB family, using a comprehensive literature mining as well as database search.
2. *Prediction of the interaction of the particular Rab with other co-expressed Rab:*
   Co-expression data, originally collected from the GEO by the geneMANIA was taken, and Rab-Rab interactions for each Rabs were retrieved, i.e, Rabs that are linked with each other across similar conditions in gene expression studies.
3. *Prediction of genes involved in each disease through Malacards database:*
   For each disease, a list of affiliated genes was prepared using data obtained from Malacards database.
4. *Construction of miRNA-gene co-expression network for each Rab-associated diseases using miRNet:*
   miRNet, an integrated platform for linking genes, targets, and functions was used to construct the miRNA-target Rab co-expression networks for each Rab associated disease.
5. *Construction of miRNA-Rab disease networks for each Rab-associated diseases using miRNet:*
   Similarly, as above, miRNet was also used to construct the miRNA-Rab networks for each Rab associated disease.
6. *Prediction of miRNAs that target each Rab:*
   The miRNAs that target each of these sets of Rab involved in each of the eighty-seven Rab associated diseases were found out and verified using various databases and reported.
7. *Validation of the miRNAs targeting various Rabs in Rab-associated diseases by tool miRNet:*
   Using miRNA lists obtained at step 6, we retrieve the diseases that were originally Rab associated.
8. *miRNA enrichment sent analysis:*
   Using the miRNA-sets for each group of diseases, we perform an enrichment set analysis.

3. Materials and methods

The databases, tools, and software used in the research have been listed. Seven databases have been chosen for screening miRNAs for the validation step, and network construction and all network analysis were carried out using powerful tools like Cytoscape, Gephi, yED and, Network analyzer. *Table T1* describes in short all the databases, tools, and software used.

A. DATABASES

1. *DisGeNET:* DisGeNET is a storehouse of a huge collection of genes and variants associated with human diseases (*Piñero et al., 2019; Piñero et al., 2016; Piñero et al., 2015*). DisGeNET integrates data from various sources: curated repositories, GWAS catalogs, animal models as well as data extracted from scientific literature. DisGeNET contains ~628,685 gene-disease associations (GDAs) between ~17,549 genes and ~24,166 diseases, disorders, traits, and clinical or abnormal human phenotypes, and ~210,498 variant-disease associations (VDAs), between ~117,337 variants and ~10,358 diseases, traits, and phenotypes [31].

2. *KEGG:* KEGG DISEASE is a collection of diseases provided by the KEGG platform. Each KEGG entry is identified by a unique H number and tabulates items like disease description, genes (genetic factors), environmental factors, pathogens, and therapeutic drugs [32].

3. *Malacards:* MalaCards, a comprehensive human disease platform, stores diseases systematically in the form of 'disease cards'. Each such card contains various linked information, as well as annotations. MalaCards currently has ~17000 human disease entries. Malacards links to another platform, GeneCards [33].

4. *MicroT:* DIANA-microT-CDS is the 5th version of the microT algorithm. DIANA-microT-CDS. Has been written to identify miRNA targets both in 3′ untranslated region (3′UTR) as well as in coding sequences (CDS). This algorithm can detect miRNA targets in mRNA sequences of *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster,* and *C.elegans* [34], [35].

5. *miRanda:* miranda retrieves target sites for miRNAs in genomic sequence. Potential target sites are identified using a two-step strategy: First, a dynamic

programming local alignment is carried out between the query miRNA sequence and the reference sequence. The second phase of the algorithm takes high-scoring alignments detected from phase 1 and estimates the thermodynamic stability of RNA duplexes based on these alignments [36].

6. *miRBase:* The miRBase database is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references, and annotation. All sequence and annotation data are also available for download. The miRBase Registry provides miRNA gene hunters with unique names for novel miRNA genes before publication of results [37], [38], [39].

7. *miRecords:* miRecords is used to retrieve miRNA–target interactions in animals. Its entries are manually curated, and it stores only experimentally verified miRNA–target interactions. In addition, it provides te used with documentation of experimental support for each interaction. miRecords currently lists ~1135 validated miRNA–target interactions between ~301 miRNAs and ~902 target genes in seven animal species [40].

8. *picTar:* PicTar is an algorithm for the identification of microRNA targets. This searchable website provides details (3' UTR alignments with predicted sites, and other annotations) regarding microRNA target predictions in vertebrates, *Drosophila*, nematodes, and human microRNA targets. The data it generates is that of co-expression data, i.e miRNAs and mRNAs expressed in the same tissue). PicTar is a project of the Rajewsky lab at NYU's Center for Comparative Functional Genomics and the Max Delbruck Centrum, Berlin [41], [42].

9. *TargetScan:* TargetScan predicts biological targets of miRNAs by searching for the presence of conserved 8mer, 7mer, and 6mer sites that match the seed region of each miRNA [43].

B. SOFTWARE

10. *Chimera:* UCSF Chimera is an interactive visualization and analysis software used by bioinformaticians, especially to view 3D structures of molecules. In addition, it provides a range of functionalities like the creation of attractive density maps, trajectories, and sequence alignments. The most widely used functions are coloring, calculations of molecular properties, and different view layouts [44].

11. *Cytoscape:* Cytoscape is a popular open-source software for visualizing molecular interaction networks. It's other salient features are viewing large networks and analyzing them, using tools available along with the Cytoscape software platform. Analysis includes annotations, gene expression profiles, enrichment analysis and so on. Apps are available for network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases [45].

12. *Gephi:* Gephi is a tool for data analysts and scientists keen to explore and understand graphs. With Gephi, the user interacts with the representation; manipulate the structures, shapes and colors to reveal hidden patterns. The goal is to help data analysts to make hypothesis, intuitively discover patterns, and isolate structure singularities or faults during data sourcing. It is a complementary tool to traditional statistics, as visual thinking with interactive interfaces is now recognized to facilitate reasoning [46].

13. *Network Analyzer:* Analyzer, a Cytoscape plugin computes a comprehensive set of topological parameters for undirected and directed networks, including the number of nodes, edges and connected components, network diameter, radius and clustering coefficient, as well as the characteristic path length, charts for topological coefficients, betweenness, and closeness, distributions of degrees, neighborhood connectiveness, average clustering coefficients, shortest path lengths, number of shared neighbours and stress centrality [47].

14. *yED Graph Editor-yworks:* yEd is a powerful desktop application that can be used to quickly and effectively generate high-quality diagrams. It can be used to create diagrams manually, or import external data for analysis. Automatic layout algorithms arrange even large data sets. yEd is freely available and runs on all major platforms: Windows, Unix/Linux, and macOS [48].

C. TOOLS

15. *geneMANIA:* The GeneMANIA web server has been designed for analyzing genes, and often, a list of genes. Given a query list, GeneMANIA extends the list with functionally similar genes that it identifies using available genomics and proteomics data. GeneMANIA also reports weights that indicate the predictive value of each selected data set for the query. Six organisms are currently supported. Users can select arbitrary subsets of the data sets associated with an organism to perform their analyses and can upload their own data sets to analyze. The high accuracy of the GeneMANIA prediction algorithm, an intuitive user interface and a large database make GeneMANIA a useful tool for any biologist [49].

16. *miRNet:* miRNet is a user-friendly, high-performance, visual analytics tool to assist researchers in understanding miRNAs, their targets and functions through a network-based approach. The key features of miRNet include: (a) A comprehensive collection of miRNA functional annotations based on the integration of data from 11 miRNA databases on miRNA interactions with genes, small molecules, long noncoding RNAs (lncRNAs), epigenetic modifiers and disease associations. (b) An intuitive interface to allow users to start with various queries of interest. (c) Support for differential expression analysis for common datasets generated from miRNA functional studies including qRT-PCR, microarray or RNA-seq. (d) various built-in functions to help network creation, refinement and customization. (e) A powerful, high-performance and fully-featured network visualization system based on standard web technology, seamlessly integrated with enrichment analysis support [50].

17. *TAM 2.0:* TAM 2.0 is the updated web server of the previously published miRNA set enrichment analysis tool, TAM in 2010. Through manual curation of over 9,000 papers, a more than two-fold growth of reference miRNA sets has been achieved in comparison with previous TAM. MiRNAs were grouped into six categories of miRNA sets: miRNA-family sets, miRNA cluster sets, miRNA-disease, miRNA-function sets, miRNA-TF sets and tissue specificity sets. Compared with the previous version of TAM, new functions for miRNA set query and result visualization are also enabled in the TAM 2.0. In all, TAM 2.0 provides a tool to mine the functional and disease implication behind miRNAs of interests [51].

SL	Databases	Available at	Citations
1	DisGeNET	https://www.disgenet.org/	540
2	KEGG Disease	https://www.genome.jp/kegg/disease/	4013
3	MalaCards	https://www.malacards.org/	179
4	microT	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index	645
5	miRanda	http://www.microrna.org/microrna/getDownloads.do	NA
6	miRBase	http://www.mirbase.org/	4527
7	miRecords	http://c1.accurascience.com/miRecords/	1254
8	picTar	https://pictar.mdc-berlin.de/	4897
9	TargetScan	http://www.targetscan.org/vert_72/	5514

SL	Tools	Available at	Citations
1	geneMANIA	https://genemania.org/	1755
2	miRNet	https://www.mirnet.ca/	150
3	TAM 2.0	http://www.lirmed.com/tam2/	18
4	BEG	http://bioinformatics.psb.ugent.be/webtools/Venn/	NA

SL	Softwares	Available at	Citations
1	Cytoscape	https://cytoscape.org/	17561
2	Gephi	https://gephi.org/	5625
3	Network Analyzer	https://med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php	1064
4	yED Graph Editor	https://www.yworks.com/products/yed	134

Table T1: *List of databases, software, plugins and tools used in research*

D. DATA COLLECTION

1. *Screening of Rab-associated diseases:*
   Comprehensive literature search, multiple hits using keywords "Rab-associated diseases", "Rab proteins" were performed and screened carefully. Also, a database retrieval of Rab associated diseases was performed to collect all Rab-associated diseases. *(*Refer: *Supplementary Table: S1)*

2. *Screening of genes and selection of Rabs:*
   Using database search, top affiliated genes for each disease were tabulated, based on their significance, from Malacards. The Rabs involved in diseases were retrieved from Disgenet, based on disease specificity index (DSI), gene-disease Association (GDA) score, evidence index (EI), and number of PMIDs supporting associations. (Refer: *Supplementary Table S3*)

3. *Generating Rab-Rab interactions:*
   A set of 72 Rab GTPases were used, after careful observation of their roles in diseases. Using each of the 72 Rabs as input, all the co-expressed Rabs were retrieved, and for each disease, a cumulate set of all the Rabs involved were built and tabulated.

4. *Selection of miRNAs for validation:*

MiRNAs that are known to target Rabs have been recorded from seven different databases and carefully recorded. (Refer: *Supplementary Table S2*)

E. METHODOLOGY

Rab associated diseases were listed and screened for their associated Rabs. This was followed by retrieving all the co-expressed Rabs in each case. Genes involved in each disease were also retrieved. Using all these above data as inputs, miRNA-target gene co-expression networks were built and analyzed. From the networks, Rab-subnetworks, i.e, Rabs, and their corresponding miRNAs were extracted and analyzed. For validation, a reverse network approach was used and retrieved miRNAs were used as inputs to retrieve the original disease. The networks were visualized in Gephi and reconstructed with only those miRNAs that target/alters Rabs in the diseases. These miRNAs were grouped into 8 sets and used for miRNA set enrichment analysis.

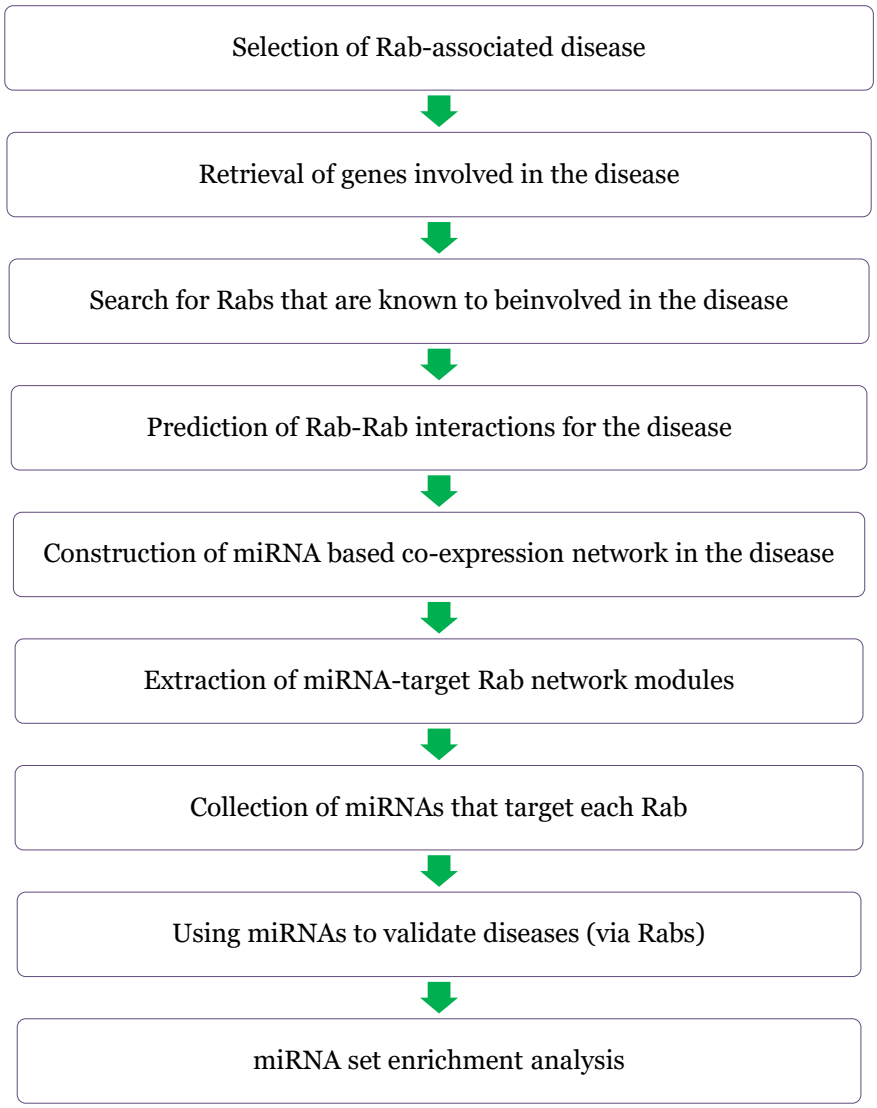


Fig. 1: *A detailed flowchart of the methodology used in the research*

The Rab-GTPase family

A. INTRODUCTION

Studies of the human genome reveals ~60 different Rabs in humans [52] [53], however, chunks of the human genome are still left to be sequenced and annotated. Further studies revealed ~70 different Rabs, distributed into 44 families [54]. Evidences of few more Rab were discovered leading to the investigation of ~72 Rabs to date. Rab genes are found widely distributed over the human chromosomes. In eukaryotes, transport of lipids and proteins between distinct membrane-bounded organelles is tightly regulated and typically occurs through transport vesicles, and Rab GTPases ('Ras-related in brain' [55]), belonging to the Ras superfamily of small GTPases, have emerged as central regulators of vesicle budding, motility and fusion [52]. All Rabs are known to exhibit two interconvertible conformational states: the GDP bound 'inactive' and the GTP-bound 'active' form. The flipping between these two states allows these proteins to perform, or to be involved in, the regulation of a wide variety of cellular functions [56]. Rab GTPases function by the effector protein that interacts with a particular GTP-bound active state of Rab and lead to a final effect. Multiple Rabs are known to interact this way, forming scaffolds allowing these interactions to span across classes of proteins and create different levels of cross talk even with receptor-mediated signalling pathways. Thus, as shown in *Fig. F2*, three main regulatory proteins are reported to modulate the guanine nucleotide-binding status of Rab: the guanine nucleotide exchange factors (GEFs), GTPase activating proteins (GAPs) and guanine nucleotide dissociation inhibitors (GDIs) [54].

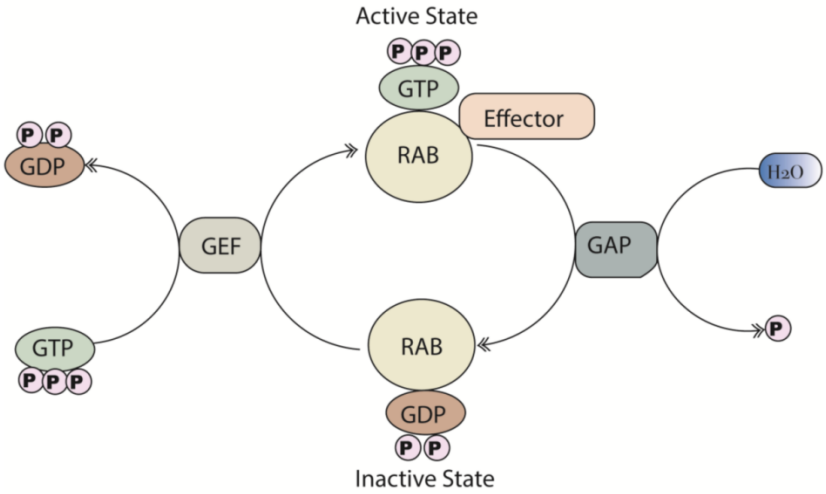


***Fig. F2**: Schematic representation of the Rab GTPase cycle. The activation and inactivation of Rab GTPases are regulated by GEFs and GAPs, respectively [59].*

B. PHYLOGENY

Rab GTPases perhaps are the products of gene duplication, evidenced by the several subfamilies of closely related Rab GTPases called isoforms, with 75-95% sequence identity as well as overlapping functions (*Fig. F4*) [53]. In general, Rab GTPases differ mostly in their carboxyl termini, as implicated in subcellular targeting [57], whereas regions involved in guanine-nucleotide binding are most conserved. Furthermore, alternative splicing has been reported in mammalian Rab genes [58]. *Fig. F3* depicts a phylogenetic analysis of some Rabs.

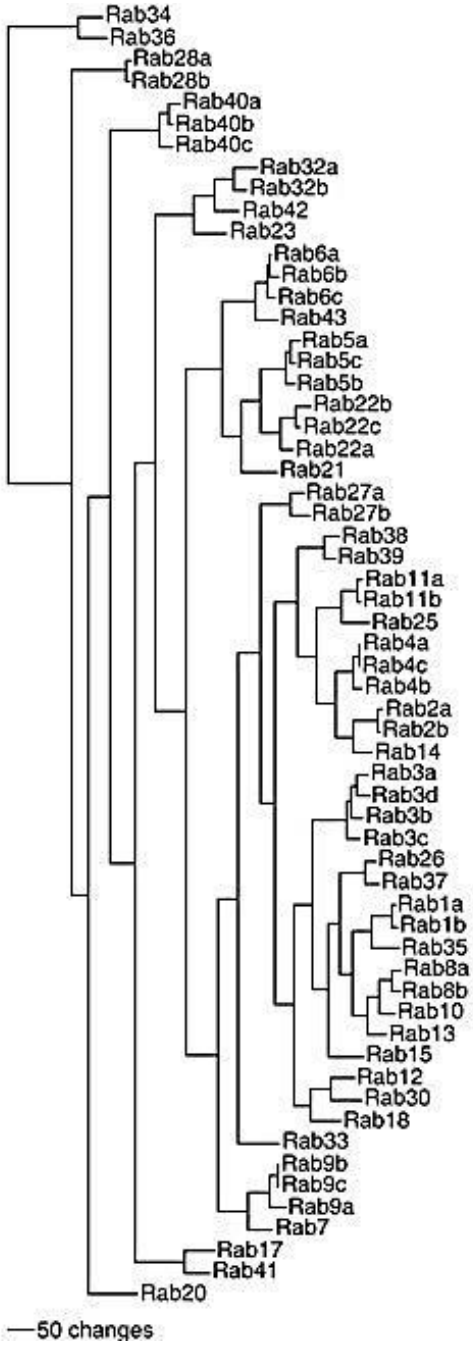


***Fig. F3**: Phylogenetic tree of human Rab GTPases.*

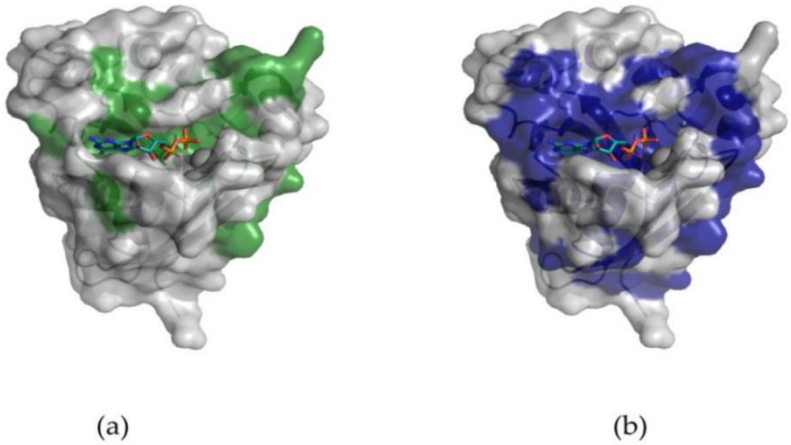


***Fig. F4**: (a) The surface representation of Rab1a crystal structure (green) from Homo sapiens (PDB ID: 4FML) shows residues that are fully conserved across all human Rab proteins. (b) Residues that are 100% conserved across Rab1a proteins from multiple species (blue). Adapted from [8]*

C. STRUCTURE

High conservation in the nucleotide-binding pocket of Rab reflects the main biochemical function- hydrolysis of guanosine triphosphate (GTP) to guanosine diphosphate (GDP) allowing Rab proteins to act as molecular "on/off" switches [60]. As schematically represented in *Fig. F1*, guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) regulate this cycle of activation and deactivation in the following way- GEFs catalyze the exchange of GDP for a GTP molecule, thereby activating the GTPases [61] and, conversely, GAPs inactivate Rabs by providing a catalytic group to accelerate the slow intrinsic GTP hydrolysis rate of the Rab-GTPases. Rabs comprise five distinct stretches of amino acids, [62], also known as RabF regions. These are known to cluster in and around the switch regions I and II. In addition, four more regions (RabF regions, shown in dark blue) have been identified that can fairly differentiate among subfamilies of Rab GTPases [57], [62]. The RabF regions lie on two different surfaces of the GTPases allowing specific binding of downstream effector molecules.

D. LOCALIZATION

Some Rab are expressed ubiquitously in human tissues, others are tissue-specific. Rabs are generally found localized in the cytosolic face of cells. Their distinct membrane localization depends on the post-translational modification of a cysteine motif at the carboxyl terminus, with one or two highly hydrophobic geranylgeranyl groups [63], which requires the initial recognition of a newly synthesized Rab protein by a Rab escort protein (REP), which presents the Rab protein to the geranylgeranyl transferase. REP functions as a chaperone that keeps the hydrophobic, geranyl-geranylated Rab soluble, thus delivering it to the appropriate membrane [64]. Rabs

act as scaffolds, selectively recruiting specific effectors [65]. A subset of 13 Rab proteins is known to be utilized in the endocytic pathway [66].

E. ALTERED RABS IN DISEASES

Rab proteins and its associated effector molecules are shown to be altered in human disease. Most often, they comprise a loss of function germline mutation. Rab protein overexpression or aberrant activation, triggered by somatic mutation or altered signaling, also underlies several disease states [66]. Loss of function mutations in Rab proteins, Rab regulatory molecules, or Rab effectors have been known to cause diseases like Griscelli syndrome type 2, Charcot–Marie–Tooth type 2B neuropathy, Choroideremia and X-linked mental retardation. Overexpression of endocytic Rab proteins are known in a range of human diseases like thyroid, vascular, and lung diseases, as well as cancers. Some important diseases include thyroid carcinoma, prostate carcinoma and cardiomyopathy.

(a) Griscelli syndrome type 2

It is a rare autosomal recessive disorder [67], with immune impairment and increased susceptibility to infections, characterized by partial albinism, resulting from the accumulation of melanosomes in melanocytes. The underlying causes of this disease are three missense mutations in highly conserved residues and numerous microdeletions or larger deletions in Rab27a [68] which is a critically important protein for the transport and release of melanosomes [69].

(b) Charcot–Marie–Tooth type 2B neuropathy

It is characterized by multiple phenotypic defects like sensory and motor neuron impairment, distal muscle weakness and atrophy, and ulcerations, caused primarily due to missense mutations in Rab7 where either a C to T transition leads to the substitution of Leu129 for Phe or a G to A transition results in mutation of Val162 to Met [70], given that Val162 residue is highly conserved in all species and Leu129 is localized adjacent to the GTP-binding domain of Rab7 making them critical for GTP binding.

(c) *Choroideremia and X-linked mental retardation*

Genetic defects in Rabs have been reports in choroideremia, X-linked mental retardation, and kidney disease in tuberous sclerosis [71], [72], [73]. Choroideremia and X-linked mental retardation result from germline mutations in regulatory factors that impact the membrane association of Rab proteins. Choroideremia is a form of retinal degeneration caused due to loss of retinal epithelium, choroids, and retinal photoreceptor cells, and characterized by progressive blindness in individuals [74]. Mutations in Rab escort protein1 (REP1), are known to be involved [75].

(d) Thyroid carcinoma

Tumors of the thyroid are reported to be associated with elevated levels of Rab5a and Rab7. Increased expression of Rab5a and Rab7 increases the rate of thyroglobulin endocytosis and processing in response to elevated cAMP [76].

(e) Cardiomyopathy

Upregulation of Rab1a, Rab4, and Rab6 has been reported in cardiomyopathy [77].

(f) Prostate carcinoma

The overexpression of a Rab regulatory factor decreases Rab activation, resulting in prostate carcinoma. PRC17, containing a GTPase-activating domain that interacts with Rab5 has been demonstrated to be highly upregulated in metastatic prostate tumors in mouse fibroblast 3T. These results suggest that the GAP activity of PRC17 is responsible for its oncogenic activity and thereby also suggests that the upregulation of GAP activity might alter Rab5 in human prostate disease [78].

4. MicroRNA and its biogenesis

A. INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs playing a major role in RNA silencing. MiRNAs have emerged to be involved in many developmental and pathological processes in animals. The biogenesis of miRNAs is under tight temporal and spatial control, and their aberrant expression has been linked to developmental abnormalities and human diseases, including cancer and cardiovascular disorders. In animals, miRNAs are ~22 nucleotides in length and generally produced by two RNase III proteins named Drosha and Dicer. MiRNA biogenesis is regulated at multiple levels, including transcriptional control of the miRNA itself. It is then processed by Drosha and Dicer in the nucleus and cytoplasm, respectively; modified by RNA editing, RNA methylation, uridylation and adenylation, argonaute loading and finally, RNA decay [79]. Often, several miRNA loci are close to each other, constituting a polycistronic transcription unit [80].

B. TRANSCRIPTION

MiRNA genes are transcribed by RNA polymerase II (Pol II), and the long primary transcript has a local hairpin structure where miRNA sequences are located. MiRNA transcription is basically controlled by RNA Pol II-associated transcription factors and epigenetic regulators [80], [81], [82]. RNA Pol III transcribes some endogenous miRNA-like small RNAs [83]. Transcription factors are also known to regulate miRNA expression [84], [85] whereas, epigenetic control, such as DNA methylation and histone modifications also contribute to miRNA gene regulation [86].

C. NUCLEAR PROCESSING

Following transcription, the primary miRNA (pri-miRNA) undergoes several steps of maturation, starting with Drosha, by cropping the stem–loop to release a small hairpin-shaped RNA of ~65 nucleotides in length, also called the pre-miRNA) [87]. Along with its essential cofactor DGCR8, Drosha now forms a complex called Microprocessor [88], cleaving the hairpin at ~11 bp away from the 'basal' junction between single-stranded RNA and dsRNA, and ~22 bp away from the 'apical' junction linked to the terminal loop [89], [90].

D. NUCLEAR EXPORT

After Drosha processing, the pre-miRNA is exported into the cytoplasm for further maturation process. Shortly after that, a specific protein, Exportin 5 (EXP5) forms a transport complex with GTP-binding nuclear protein RAN-GTP and a pre-miRNA [91], [92], [93]. Following translocation through the nuclear pore complex, GTP is hydrolyzed, resulting in the disassembly of the complex and the release of the pre-miRNA into the cytosol. Crystal studies of this complex shows that EXP5–Ran-GTP forms a 'baseball mitt'-like structure into which the pre-miRNA stem is positioned, allowing the interaction of the pre-miRNA stem with the positively charged inner surface. A tunnel-like structure at the bottom of the mitt-like structure strongly interacts with the two-nucleotide-long 3′ overhang of the pre-miRNA [94]. This tunnel-like structure is consistent with the results from earlier biochemical analyses: EXP5 recognizes a dsRNA stem of >14 bp in length together with a short 3′ overhang ~1–8 nucleotides in length [92], [95], [96].

E. CYTOPLASMIC PRE-MIRNA PROCESSING

Once released into the cytoplasm, the pre-miRNA is cleaved by Dicer near the terminal loop, releasing a small RNA duplex [97]. Dicer binds to pre-miRNA with a preference for a two-nucleotide-long 3′ overhang, initially generated by Drosha. In general, Dicer cleavage sites are located at a fixed distance from the 3′ end of the terminus of dsRNAs [98]. This distance is typically 21–25 nucleotides in length depending on the species and the type of Dicer. In mammals and flies, an additional mechanism to determine the cleavage site of pre-miRNA is present; Dicer binds to the 5′ phosphorylated end of the pre-miRNA chopping ~ 22 nucleotides away from the 5′ end [99]. The 5′ end binding occurs when the end is thermodynamically unstable, but not when the end is strongly paired (such as through G·C base pairs).

F. RISC FORMATION

A small RNA duplex generated by Dicer is loaded onto an AGO protein to form an effector complex called the RNA-induced silencing complex (RISC). RISC assembly involves two steps [100]: following Dicer processing, RNA duplexes are preferentially loaded onto particular types of AGO proteins [101]. The pre-RISC now removes the passenger strand to generate a mature RISC. Slicing-competent AGO proteins cleaves the passenger strand if the duplex is matched at the centre [102]. Removal of the cleaved passenger strand is facilitated by the endonuclease C3PO, a multimeric complex of translin and translin-associated protein X (TRAX) [103].

5. *The role of Rab in Rab associated diseases*

In this chapter, we have made a comprehensive analysis by tabulating diseases that are also known to be associated with various members of the Rab family of GTPases. As already discussed, we have so far gathered a fair idea of how Rab GTPases work inside a cell and how alterations or modifications may lead to various abnormalities. For this step, we have used DisGeNET, a database of genes and variants of human diseases, and tabulated the diseases along with the Rab reported to be associated with them as shown in *Table T2*. The list has been prepared based on the following parameters of DisGeNET: DSI (Disease specificity Index), GDA Score (gene-disease Association), and EI (Evidence Index). *Supplementary Table S3* comprises a list of the top affiliated genes associated with each disease, collected from MalaCards, a database of human maladies and their annotations based originally on GeneCards database of human genes. The Rab-Rab co-expression and interaction data for each Rab has been collected from geneMANIA (as illustrated in *Fig. F5*). Using this data, cumulative Rab-Rab interaction data has been prepared and tabulated in *Table T3*. We have now successfully associated the diseases to their associated Rab, background Rab and other affiliating genes involved, theoretically. In further studies, we will associate these Rab with the miRNAs that target them, theoretically again, and propose miRNA-target Rab networks. A schematic representation of the workflow is shown below. To quantify (somewhat), the role of Rab in each Rab-associated disease, we have calculated Rab occupancy using simple mathematics:

$$Rab\ occupancy = \frac{Number\ of\ Rabs\ including\ background\ RABs\ each\ disease}{Total\ number\ of\ affiliated\ genes\ in\ the\ same}\ X\ 100$$

The diseases have been group into 12 sets as follows, for simplicity purposes: Cancers (25), neurodegenerative diseases (21), musculoskeletal disorders (2), congenital malformation (10), cardiovascular disorders (3), endocrine and metabolic disorders (8), immunological disorders (5), endocrine diseases (5), infectious diseases (3), hematological disorders (1), endocrine and reproductive diseases (1) and chromosomal abnormalities (1). *Table T4* shows the data sheet of all diseases used to calculate Rab occupancies.

Cancers	NDDs	Musculoskeletal disorders	Congenital malformation
1. Cellular Ependymoma	1. Charcot-Marie-Tooth disease	1. Osteoporosis	1. Martsolf syndrome
2. Acute erythroblastic Leukemia	2. Alzheimer disease	2. Duchenne muscular dystrophy	2. Craniosynostoses
3. Neuroblastoma	3. Familial dementia		3. Smith-McCort dysplasia

4. Breast Carcinoma 5. leukemia 6. Liver carcinoma 7. Prostate carcinoma 8. Tongue Carcinoma 9. Colorectal Carcinoma 10. Cervical cancer 11. Carcinoma of lung 12. Renal Cell Carcinoma 13. Pancreatic carcinoma 14. Ovarian Carcinoma 15. Non-Small Cell Lung Carcinoma 16. Renal carcinoma 17. Thyroid carcinoma 18. Skin Carcinogenesis 19. Colon Carcinoma 20. Osteosarcoma 21. Carcinoma of bladder 22. Stomach Carcinoma 23. Esophageal carcinoma 24. T-Cell Lymphoma 25. Gastrointestinal stromal tumors	4. Parkinson disease 5. Autistic spectrum disorder 6. Ataxia telangiectasia 7. Amyotrophic lateral sclerosis (ALS) 8. Cone-rod dystrophy and cone dystrophy 9. Hereditary sensory and autonomic neuropathy 10. Familial amyloidosis 11. Marinesco-Sjogren syndrome 12. Multiple sclerosis 13. Rett syndrome 14. Deafness and myopia 15. Duchenne muscular dystrophy 16. Choroideremia 17. Mental retardation, X-linked, syndromic, Martin-Probst type 18. Mental retardation, X-linked 19. StRabismus 20. Intellectual disability 21. Paroxysmal extreme pain disorder	4. Postaxial polydactyly 5. Pulmonary valvular stenosis 6. Brachydactyly 7. Syndactyly 8. Ciliopathies 9. Micrognathism 10. Acrocephalopolysyndactyly type II	

Cardiovascular diseases	Endocrine and metabolic disorders	Immunological disorders	Endocrine disorders
1. Congestive heart failure 2. Coronary artery disease 3. Myocardial ischemia	1. Nonalcoholic fatty liver disease 2. Type 2 diabetes mellitus 3. Genetic obesity 4. Cystic fibrosis	1. Systemic lupus erythematosus 2. Griscelli syndrome, type 2 3. Ataxia telangiectasia 4. Asthma	1. Acromegaly 2. Genetic obesity 3. Polycystic ovary syndrome 4. Nonalcoholic fatty liver disease

	5. Oculocutaneous albinism 6. Hermansky-Pudlak syndrome 7. Lipidosis 8. hypercholesterolemia	5. Rheumatoid arthritis	5. Type 2 diabetes mellitus

Infectious diseases	Hematological disorders	Endocrine and reproductive disorders	Chromosomal abnormalities
1. Yellow fever 2. Lyme disease 3. Saint Louis encephalitis	1. Iron-refractory iron deficiency anemia	1. Polycystic ovary syndrome	1. Down syndrome

Table T2: *List of diseases studied*

SL	Disease	Rabs involved	Interacting (background) Rabs	Rabs involved in the network
1	Charcot-Marie-Tooth disease, axonal, Type 2b (disorder)	Rab7a	Rab1a, Rab5a	Rab7a, Rab1a, Rab5a, Rab7b
		Rab7b	-	
2	Osteoporosis	Rab7b	-	Rab7b
3	Pain disorder	Rab7a	Rab1a, Rab5a	Rab7a, Rab1a, Rab5a
4	Sclerocystic Ovaries	Rab2a	Rab1a	Rab2a, Rab1a
5	Myocardial Ischemia	Rab7a	Rab1a, Rab5a	Rab7a, Rab1a, Rab5a, Rab1b, Rab8a
		Rab1b	Rab8a	
		Rab5a	Rab7a	
6	Polycystic Ovary Syndrome	Rab2a	Rab1a	Rab1a, Rab2a
7	Cellular Ependymoma	Rab3a	Rab3b	Rab3a, Rab3b
8	Acute coronary Syndrome	Rab7b	-	Rab7b
9	Rheumatoid arthritis	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b, Rab27a
		Rab27a	-	
10	Colorectal cancer	Rab5c	Rab43	Rab5c, Rab43
11	Lupus Erythematosus, Systemic	Rab4a	-	Rab4a
12	Alzheimer's disease	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37, Rab4a, Rab3a, Rab3b, Rab6a, Rab11b, Rab7a, Rab1a, Rab11a, Rab1b, Rab38
		Rab4a	-	
		Rab3a	Rab3b	

		Rab6a	Rab11b	
		Rab7a	Rab1a, Rab5a	
		Rab11a	Rab1b, Rab6a	
		Rab38	-	
13	Acute Erythroblastic Leukemia	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b
14	Down Syndrome	Rab5a	Rab7a, Rab37	Rab5a, Rab37, Rab7a
15	Neuroblastoma	Rab6b	-	Rab6b, Rab7a, Rab1a, Rab5a, Rab7b, Rab1b, Rab8a, Rab11b, Rab40b
		Rab7a	Rab1a, Rab5a	
		Rab7b	-	
		Rab1b	Rab8a, Rab11b	
		Rab40b	-	
16	Dementia	Rab3a	Rab3b	Rab3a, Rab3b
17	Breast carcinoma	Rab6c	Rab5b, Rab10, Rab14	Rab6c, Rab1b, Rab8a, Rab3d, Rab1a, Rab2a, Rab3a, Rab5a, Rab6a, Rab11a, Rab31, Rab25, Rab27b, Rab27a, Rab22a, Rab40aL, Rab35, Rab21, Rab40c, Rab5b, Rab10, Rab14, Rab11b, Rab7a, Rab3b, Rab37, Rab5c
		Rab1b	Rab8a, Rab11b	
		Rab8a	Rab1b, Rab11b	
		Rab3d	-	
		Rab1a	Rab2a, Rab7a, Rab11a	
		Rab2a	Rab1a	
		Rab3a	Rab3b	
		Rab5a	Rab7a, Rab37	
		Rab6a	Rab11b	
		Rab11a	Rab1a	
		Rab31	-	
		Rab25	Rab5c	
		Rab27b	-	
		Rab27a	-	
		Rab22a	-	
		Rab40aL	-	
		Rab35	Rab5c	
		Rab21	-	
		Rab40c	-	

18	Leukemia	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b, Rab7a, Rab1a, Rab5a, Rab5c, Rab43, Rab4a, Rab27a, Rab40b
		Rab7a	Rab1a, Rab5a	
		Rab5c	Rab43	
		Rab4a	-	
		Rab27a	-	
		Rab40b	-	
19	Liver carcinoma	Rab5a	Rab7a, Rab37	Rab5a, Rab8a, Rab17, Rab18, Rab10, Rab25, Rab34, Rab27a, Rab31, Rab21, Rab23, Rab24, Rab27b, Rab40b,Rab7a, Rab37, Rab1b, Rab11b, Rab1a, Rab5c
		Rab8a	Rab1b, Rab11b	
		Rab17	-	
		Rab18	Rab1a	
		Rab10	-	
		Rab25	Rab5c	
		Rab34	-	
		Rab27a	-	
		Rab31	-	
		Rab21	-	
		Rab23		
		Rab24	-	
		Rab27b	-	
		Rab40b	-	
20	Yellow Fever	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b
21	Prostate carcinoma	Rab1a	Rab2a, Rab11a, Rab7a	Rab1a, Rab3b, Rab8a, Rab5a, Rab6a, Rab27a, Rab27b, Rab2a, Rab11a, Rab7a, Rab3a, Rab1b, Rab11b, Rab37
		Rab3b	Rab3a	
		Rab8a	Rab1b, Rab11b	
		Rab5a	Rab7a, Rab37	
		Rab6a	Rab11b	
		Rab27a	-	
		Rab27b	-	
22	Tongue carcinoma	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a
23	Parkinson disease	Rab3b	Rab3a	Rab3b, Rab3a, Rab1a, Rab2a, Rab7a, Rab11a, Rab6a, Rab11b, Rab29, Rab39b
		Rab1a	Rab2a, Rab7a, Rab11a	

		Rab6a	Rab11b	
		Rab29	-	
		Rab39b	-	
24	Autistic disorder	Rab3a	Rab3b	Rab3a, Rab3b, Rab2a, Rab1a ,Rab39b
		Rab2a	Rab1a	
		Rab39b	-	
25	Congestive heart failure	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a
26	Colorectal carcinoma	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a, Rab27a, Rab27b, Rab22a, Rab38
		Rab27a	-	
		Rab27b	-	
		Rab22a	-	
		Rab38	-	
27	Ataxia Telangiectasia	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a
28	Asthma	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a, Rab5a, Rab37, Rab27a
		Rab5a	Rab7a, Rab37	
		Rab27a	-	
		Rab11a	-	
29	Amyotrophic Lateral Sclerosis	Rab1a	Rab2a, Rab7a, Rab11a	Rab1a, Rab2a, Rab7a, Rab11a, Rab5a, Rab37, Rab29
		Rab5a	Rab7a, Rab37	
		Rab11a	Rab1a	
		Rab29	-	
30	Ciliopathies	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b
31	Photoreceptor degeneration	Rab8a	Rab1b, Rab11b	Rab8a, Rab1b, Rab11b
32	Cervical cancer	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37, Rab40b
		Rab40b	-	
33	Mild cognitive disorder	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37
34	Carcinoma of lung	Rab5a	Rab7a, Rab37	Rab5a, Rab25, Rab40b, Rab7a, Rab37, Rab5c
		Rab25	Rab5c	
		Rab40b	-	
35	Hereditary Sensory autonomic Neuropathy, Type 1	Rab7a	Rab1a, Rab5a	Rab1a, Rab5a, Rab7a

36	Choroideremia	Rab7a	Rab1a, Rab5a	Rab1a, Rab7a, Rab5a
37	Amyloidosis	Rab7a	Rab1a, Rab5a	Rab7a, Rab11a, Rab1a, Rab5a
		Rab11a	Rab1a	
38	Sjogren's Syndrome	Rab4a	-	Rab4a
39	Multiple Sclerosis	Rab4a	-	Rab4a
40	Renal cell carcinoma	Rab5a	Rab7a	Rab5a, Rab7a, Rab37
		Rab37	-	
41	Pancreatic carcinoma	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37, Rab20, Rab1a, Rab11a
		Rab20	-	
		Rab11a	Rab1a	
42	Ovarian carcinoma	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37, Rab25, Rab5c, Rab22a, Rab35
		Rab25	Rab5c	
		Rab22a	-	
		Rab35	Rab5c	
43	Lyme disease	Rab5a	Rab7a, Rab37	Rab5a, Rab7a, Rab37, Rab22a
		Rab22a	-	
44	Encephalitis, St. Louis	Rab4a	-	Rab4a
45	Rett Syndrome	Rab14	-	Rab14
46	Non-Small cell Lung carcinoma	Rab14	-	Rab14, Rab18, Rab1a, Rab37
		Rab18	Rab1a	
		Rab37	-	
47	Renal carcinoma	Rab14	-	Rab14, Rab38, Rab40b
		Rab38	-	
		Rab40b	-	
48	Acromegaly	Rab18	Rab1a	Rab1a, Rab18
49	Martsolf syndrome	Rab18	Rab1a	Rab1a, Rab18
50	Obesity	Rab18	Rab1a	Rab1a, Rab18, Rab23, Rab21
		Rab23	-	
		Rab21	-	
51	Thyroid carcinoma	Rab11a	Rab1a	Rab11a, Rab1a, Rab23, Rab40b
		Rab23	-	

		Rab40b	-	
52	Skin carcinogenesis	Rab11a	Rab1a	Rab11a, Rab1a, Rab23
		Rab23	-	
53	Hypercholesterolemia	Rab9a	Rab11a	Rab9a, Rab11a
54	Lipoidosis	Rab9a	Rab11a	Rab9a, Rab11a
55	Griscelli Syndrome, Type 2	Rab27a	-	Rab27a
56	Acrocephalopolysyndactyly type 2	Rab23	-	Rab37
57	Craniosynostosis	Rab23	-	Rab23
58	Smith-Mccort dysplasia 2	Rab33b	-	Rab33b
59	Intellectual disability	Rab23	-	Rab23, Rab27a
		Rab27a	-	
60	Gastrointestinal Stromal Tumors	Rab23	-	Rab23
61	Polydactyly	Rab23	-	Rab23
62	Coronary artery disease	Rab37	-	Rab37
63	Micrognathism	Rab23	-	Rab23
64	Pulmonary Stenosis	Rab23	-	Rab23
65	Brachydactyly	Rab23	-	Rab23
66	Syndactyly of fingers and toes	Rab23	-	Rab23
67	Severe myopia	Rab28	-	Rab28
68	Colon carcinoma	Rab25	Rab5c	Rab25, Rab5c
69	Albinoidism, Oculocutaneous, autosomal dominant	Rab27a	-	Rab27a, Rab38
		Rab38	-	
70	Osteosarcoma	Rab22a	-	Rab22a, Rab27b
		Rab27b	-	
71	Carcinoma of bladder	Rab25	Rab5c	Rab25, Rab5c, Rab27a, Rab27b
		Rab27a	-	
		Rab27b	-	
72	Cystic Fibrosis	Rab27a	-	Rab27a, Rab40b
		Rab40b	-	
73	Refractory anemias	Rab27a	-	Rab27a
74	Stomach carcinoma	Rab23	-	Rab23, Rab32, Rab40c

		Rab32	-	
		Rab40c	-	
75	Dermatitis, atopic	Rab31	-	Rab31
76	Hermanski-Pudlak Syndrome	Rab32	-	Rab32, Rab38
		Rab38	-	
77	Marinesco-Sjogren syndrome	Rab32	-	Rab32
78	Albinism	Rab27a	-	Rab27a, Rab38
		Rab38	-	
79	Waisman syndrome	Rab39b	-	Rab39b
80	Martin-Probst deafness-Mental Retardation Syndrome	Rab40l	-	Rab40al
81	Mental Retardation, X-Linked	Rab39b	-	Rab39b, Rab40al
		Rab40al	-	
82	Strabismus	Rab39b	-	Rab39b
83	Esophageal carcinoma	Rab40b	-	Rab40b
84	Muscular dystrophy, Duchenne	Rab40al	-	Rab40al, Rab40c
		Rab40c	-	
85	Fatty Liver	Rab40b	-	Rab40b
86	Diabetes	Rab38	-	Rab38
87	T-cell Lymphoma	Rab40b	-	Rab40b

Table T3: *Diseases with their associated Rabs and interacting Rabs*

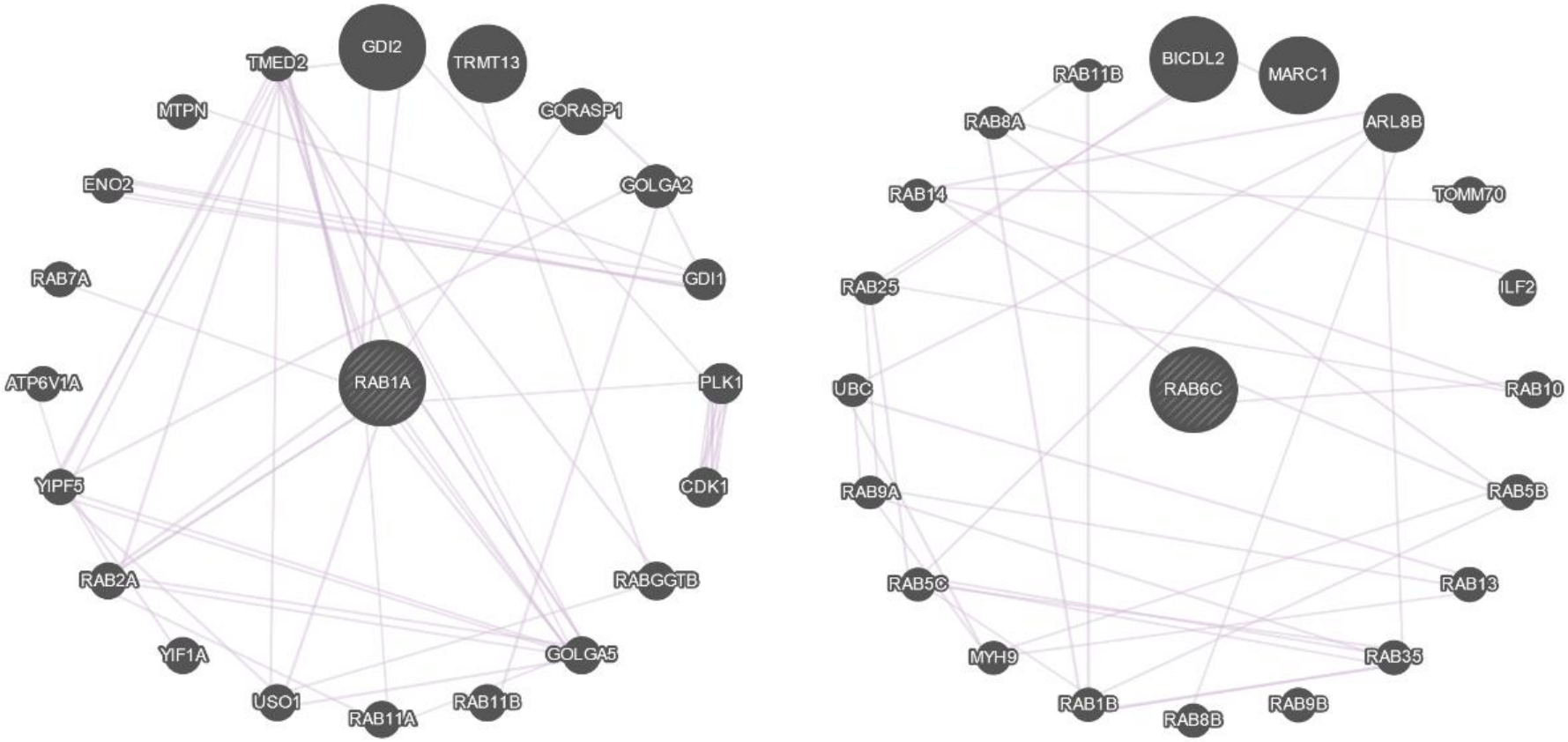


Fig. F5: *The Cytoscape GENEMANIA plugin used to investigate the Rabs that are co-expressed with each of the 72 Rabs in our study. To illustrate, interaction networks, arranged in a circular layout have been shown for Rab1a and Rab 6c. Similarly, all such interactions were recorded. Rab 2a, Rab11a, and Rab7a are co-expressed with Rab 1a. Rab 5b, Rab10, Rab14 are co-expressed with Rab6c. Co-expression has been shown with an existing edge between two Rabs (represented as nodes*

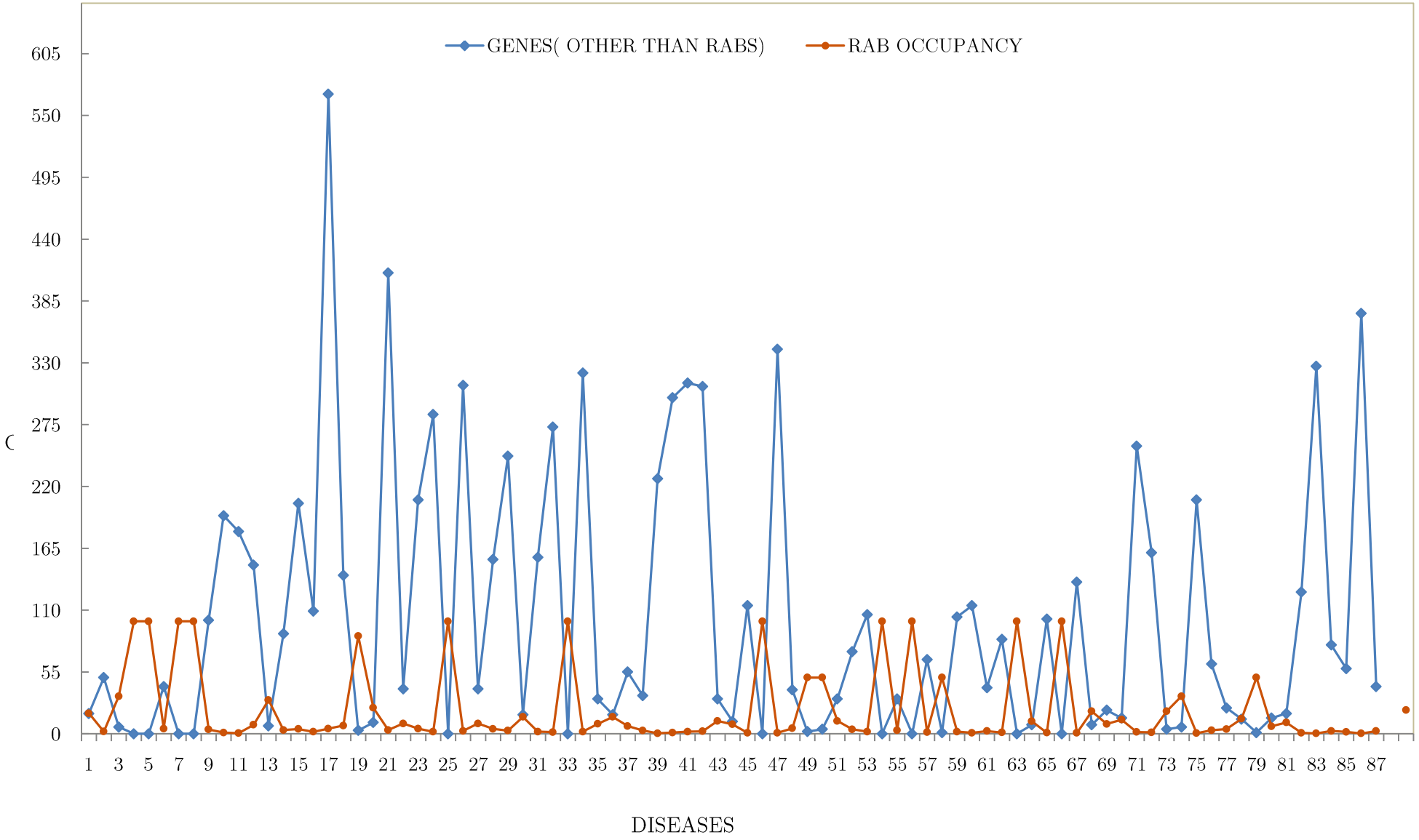
Disease	Genes other than Rabs	Rabs	Total genes (nodes)	Rab occupancy (Percent)
Charcot-Marie-Tooth disease	18	4	22	18.18181818
Osteoporosis	50	1	51	1.960784314
Pain Disorder	6	3	9	33.33333333
Sclerocystic Ovaries	0	1	1	100
Myocardial Ischemia	0	5	5	100
Polycystic Ovary Syndrome	42	2	44	4.545454545
Cellular Ependymoma	0	2	2	100
Acute Coronary Syndrome	0	1	1	100
Rheumatoid Arthritis	101	4	105	3.80952381
Colorectal Cancer	194	2	196	1.020408163
Lupus Erythematosus, Systemic	180	1	181	0.552486188
Alzheimer's Disease	150	13	163	7.975460123
Acute Erythroblastic Leukemia	7	3	10	30
Down Syndrome	89	3	92	3.260869565
Neuroblastoma	205	9	214	4.205607477
Dementia	109	2	111	1.801801802
Breast Carcinoma	569	27	596	4.530201342
leukemia	141	11	152	7.236842105
Liver carcinoma	3	20	23	86.95652174
Yellow Fever	10	3	13	23.07692308
Prostate carcinoma	410	14	424	3.301886792
Tongue Carcinoma	40	4	44	9.090909091
Parkinson Disease	208	10	218	4.587155963
Autistic Disorder	284	5	289	1.730103806
Congestive heart failure	0	4	4	100
Colorectal Carcinoma	310	8	318	2.51572327
Ataxia Telangiectasia	40	4	44	9.090909091
Asthma	155	7	162	4.320987654
Amyotrophic Lateral Sclerosis	247	7	254	2.755905512
Ciliopathies	17	3	20	15
Photoreceptor degeneration	157	3	160	1.875
cervical cancer	273	4	277	1.444043321
Mild cognitive disorder	0	3	3	100
Carcinoma of lung	321	6	327	1.834862385
Hereditary Sensory Autonomic Neuropathy, Type 1	31	3	34	8.823529412
Choroideremia	17	3	20	15
Amyloidosis	55	4	59	6.779661017
Sjogren's Syndrome	34	1	35	2.857142857
Multiple Sclerosis	227	1	228	0.438596491
Renal Cell Carcinoma	299	3	302	0.993377483
Pancreatic carcinoma	312	6	318	1.886792453
Ovarian Carcinoma	309	7	316	2.215189873
Lyme Disease	31	4	35	11.42857143

Encephalitis, St. Louis	11	1	12	8.333333333
Rett Syndrome	114	1	115	0.869565217
Non-Small Cell Lung Carcinoma	0	4	4	100
Renal carcinoma	342	3	345	0.869565217
Acromegaly	39	2	41	4.87804878
Martsolf syndrome	2	2	4	50
Obesity	4	4	8	50
Thyroid carcinoma	31	4	35	11.42857143
Skin Carcinogenesis	73	3	76	3.947368421
Hypercholesterolemia	106	2	108	1.851851852
Lipoidosis	0	2	2	100
Griscelli Syndrome, type 2	31	1	32	3.125
Acrocephalopolysyndactyly type 2	0	1	1	100
Craniosynostosis	66	1	67	1.492537313
Smith-Mccort dysplasia 2	1	1	2	50
Intellectual Disability	104	2	106	1.886792453
Gastrointestinal Stromal Tumors	114	1	115	0.869565217
Polydactyly	41	1	42	2.380952381
Coronary Artery Disease	84	1	85	1.176470588
Micrognathism	0	1	1	100
Pulmonary Stenosis	8	1	9	11.11111111
Brachydactyly	102	1	103	0.970873786
Syndactyly of fingers and toes	0	1	1	100
Severe myopia	135	1	136	0.735294118
Colon Carcinoma	8	2	10	20
Albinoidism, oculocutaneous	21	2	23	8.695652174
Osteosarcoma	14	2	16	12.5
Carcinoma of bladder	256	4	260	1.538461538
Cystic Fibrosis	161	2	163	1.226993865
Refractory anemias	4	1	5	20
Stomach Carcinoma	6	3	9	33.33333333
Dermatitis, Atopic	208	1	209	0.4784689
Hermanski-Pudlak Syndrome	62	2	64	3.125
Marinesco-Sjogren syndrome	23	1	24	4.166666667
Albinism	13	2	15	13.33333333
Waisman syndrome	1	1	2	50
Martin-Probst Deafness-Mental Retardation	14	1	15	6.666666667
Mental Retardation, X-Linked	18	2	20	10
Strabismus	126	1	127	0.787401575
Esophageal carcinoma	327	1	328	0.304878049
Muscular Dystrophy, Duchenne	79	2	81	2.469135802
Fatty Liver	58	1	59	1.694915254
Diabetes	374	1	375	0.266666667
T-Cell Lymphoma	42	1	43	2.325581395

Table T4: *Calculation of Rab occupancy*

We have shown the overall Rab occupancy in each of the 12 sets with highlighted data points as shown in *Fig. F6*. From these subplots, we can gather a fair picturesque idea of the number of Rab involved in each disease (Represented as integers 1, 2, 3, and so on). This is important because of obvious reasons-to characterize how important the Rab family of GTPases are in each disease group and to what extent a modification or alteration in them can carry the errors forward traveling through neighbors and causing disease symptoms. The following inferences can be logically drawn from the above occupancy graphs: Cellular ependymoma, liver carcinoma, non-small cell lung carcinoma, acute erythroblastic leukemia and stomach carcinoma had the highest Rab occupancies of 100, 86.9565, 100, 30 and 33.33 (percent) respectively among cancers. Charcot-Marie-Tooth disease, Choroideremia, and paroxysmal extreme pain disorder had the highest Rab occupancies of 18.1818, 15, and 33.3333 (percent) respectively, among neurodegenerative disorders. Martsolf syndrome, Smith-McCort dysplasia, Syndactyly, micrognathism, and acrocephalopolysyndactyly type 2 had the highest Rab occupancies of 50, 50, 100, 100 and 100 (percent) respectively among congenital disorders. Congestive heart failure and myocardial ischemia had Rab occupancies of 100 and 100 (percent) respectively among cardiovascular disorders. Genetic obesity and lipidosis had Rab occupancies of 50 and 100 (percent) respectively among endocrine and metabolic disorders. Ataxia-telangiectasia had the highest Rab occupancy of 9.0909% among immunological disorders. Genetic obesity had Rab occupancy of 50% among endocrine diseases. Yellow fever had Rab occupancy of 23.0769% among infectious diseases. Refractory anemia had Rab occupancy of 20% among hematological disorders. This accounted for an average of 21.1409% Rab occupancy in Rab-associated diseases.

Fig. F6: *Rab occupancy line-graph (Diseases numbered from 1-87)*



7. MiRNA-target gene co-expression network

A. OVERVIEW

An estimated 30% of all mammalian genes are known to be regulated by miRNAs, each regulating many mRNAs, and each mRNA being targeted by various miRNAs. MiRNAs tend to act via the downregulation of their gene targets, in an inverse correlation relationship [104]. Thus, it is evident how miRNAs involve in various ways, target various genes, altering and modifying them in cells. However, owing to the huge human interactome comprising a daunting ~25000 protein-coding genes, ~ 1000 metabolites, and an increasing number of proteins and functional mRNAs, we can imagine ~1,00,000 interactions [3] and perhaps even more, with ~2000 human disorders. This data is increasingly huge as well as complex. Also, as established by Barabasi et al, cellular components seldom act as single entities, rather involve groups of interacting proteins, each with their definite function, leading to a final function. It is therefore only justified to take into consideration all the interacting components of the component in question. Mathematically, such interaction networks can be written down in the form of a graph G= (V, E) where V is the set of vertices or nodes and E is the set of edges. A biological network, therefore, assumes each cellular component as nodes and the interaction between them as edges.

B. GENERATION OF NETWORKS

In our thesis work, we have built miRNA-target gene network models for each disease (listed previously). We have proposed our network models as undirected assuming genes are targeted by miRNAs in all cases. All networks were generated by feeding affiliating genes, Rabs, and background Rabs involved in each disease to the MiRNet software suite. The resulting networks were downloaded as graphML and SVG files (viewable in Gephi, yED Graph Editor, Cytoscape) and saved for further work. (Described in the next chapter). MiRNA based gene regulatory network comprises the following- miRNAs, genes, and PPIs. In *Fig. F7*, we have illustrated two such network models of Down Syndrome and Lupus Erythematosus, Systemic respectively. The Rabs and their corresponding interactions have been highlighted in pink, showing the portions of the networks solely occupied by Rabs and the miRNAs that target them.

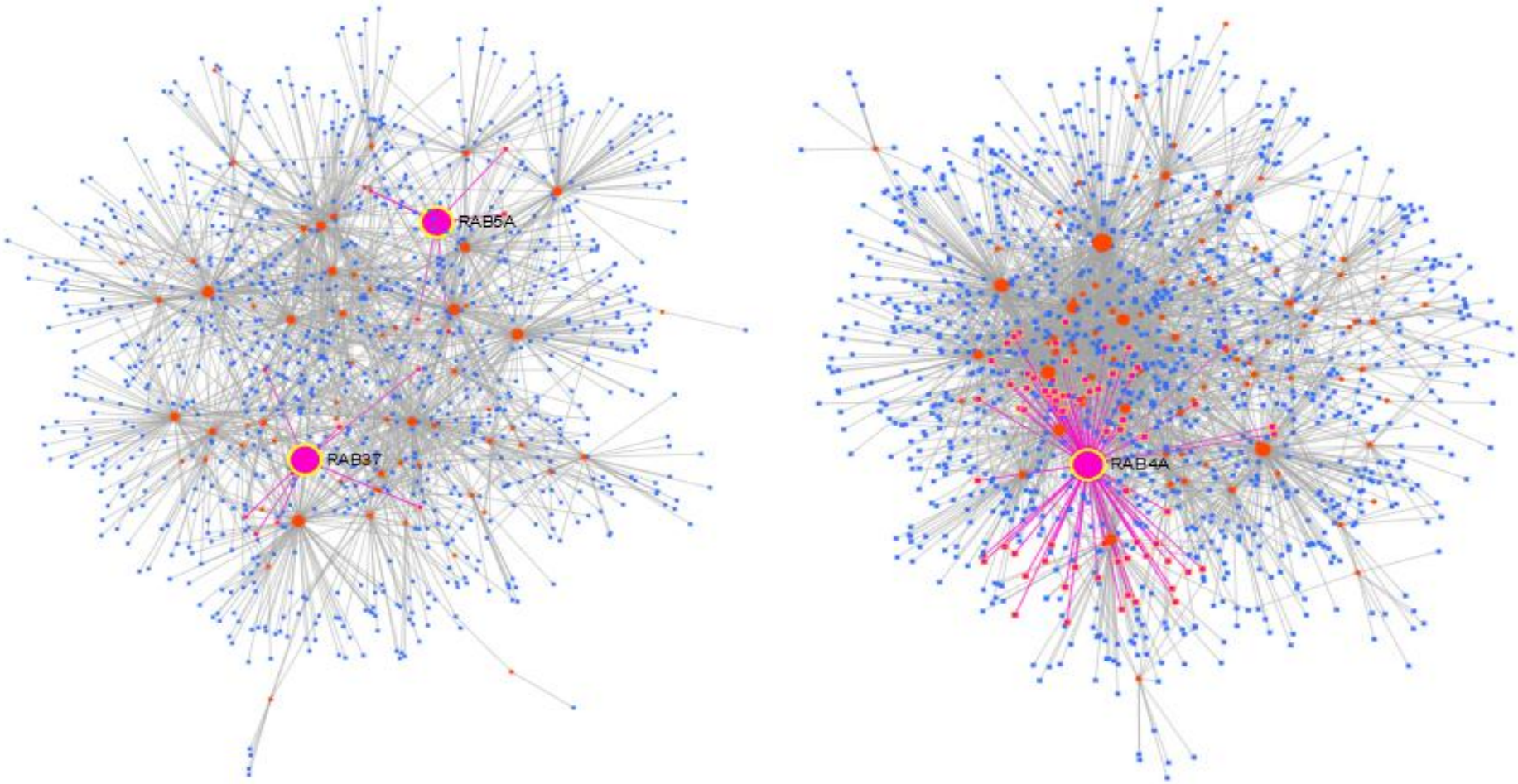
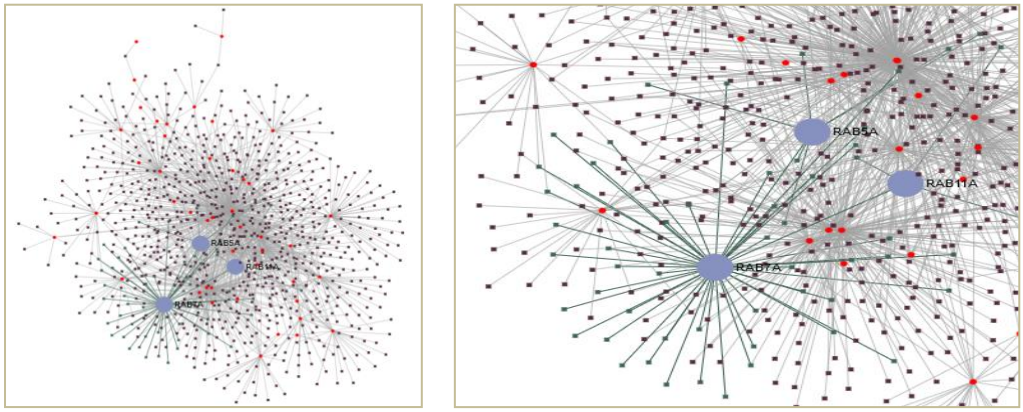


Fig. F7: *A force-atlas layout of miRNA- target gene co-expression networks of (a) Down Syndrome (b) Lupus Erythematosus, Systemic; generated using MiRNet. Nodes representing genes have been colored Red, miRNAs represented by small blue squares, and Rabs in enlarged pink circles. The network interactions (represented by edges) made by the Rabs in each of these diseases are highlighted in pink.*

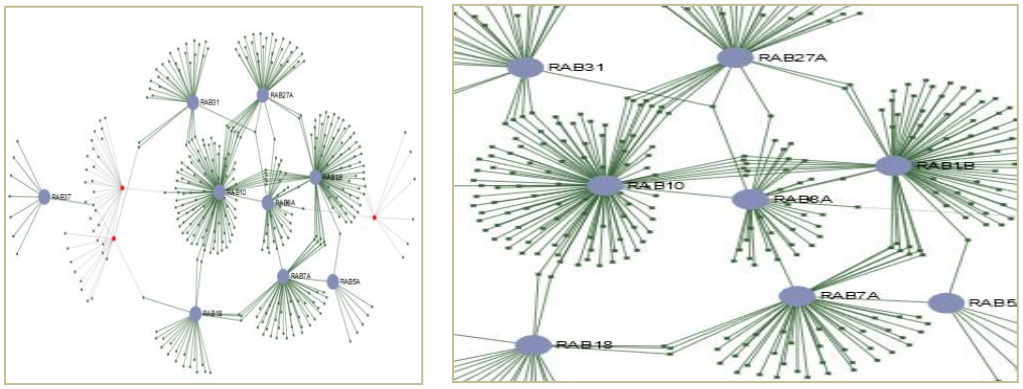
C. HUBS

For disease networks, it has been hypothesized that non-essential genes tend to spread along the network periphery, avoiding hub regions, and essential genes tend to cluster together occupying hubs [32]. In this analysis it has been observed that most Rabs occupied more or less central positions, thus emphasizing their central roles in contributing to disease phenotypes. Rabs being members of a large family also tend to cluster together in large networks involving a group of co-expressed Rabs. In *Fig. F8* we have shown miRNA-target gene networks for some diseases, with Rab-Rab interactions and miRNA-target gene interactions highlighted in green. Rabs have been found to interact closely with each other in most networks.

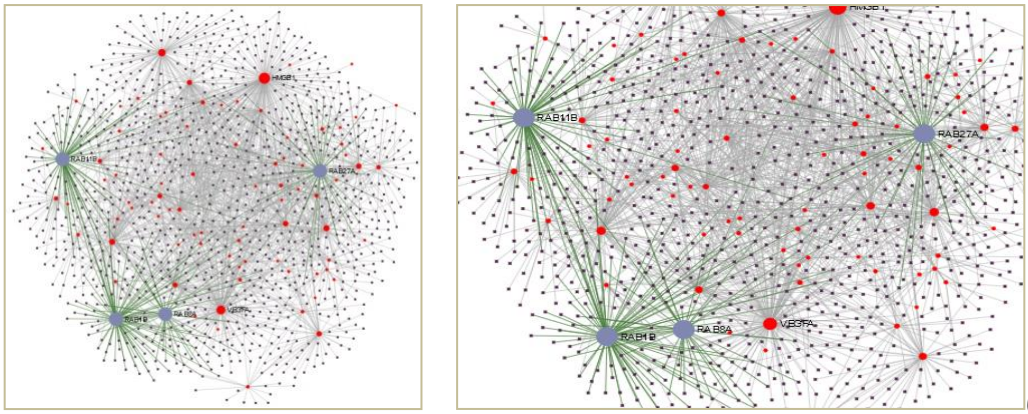


(A)

(B)



(C)



(D)

Fig. F8: *(A) Amyloidosis (B) Alzheimer's Disease (C) Liver carcinoma (D) Rheumatoid arthritis. miRNA-gene co-expression networks*

D. THE SHARED GENE HYPOTHESIS

Few diseases have been observed to share common genes, which in turn suggest that the diseases might have a common genetic origin. In addition, this proves that diseases symptoms are twice as likely to develop if that disease shares gene/genes with a primary disease [32]. For instance, breast carcinoma and ovarian carcinoma share 671 common genes, which indicate that the diseases might have a common genetic origin. In addition, the four Rabs (Rab22a, Rab25, Rab35, Rab5a) are reported to be involved in ovarian carcinoma have also been found in breast cancer.

E. DEGREE, DEGREE DISTRIBUTION AND SCALE-FREE PROPERTY

Most biological networks are known to be "scale-free", and governed by the Power law, $P(k) \sim k^{-\gamma}$ where γ is the degree component and $\sim$ indicates proportionality. In the majority of biological networks, we see this property which essentially means that if we plot the degree distribution of a scale-free network in a logarithmic scale, it fits in a power-law, with only a small number of nodes with high degree and a large number of nodes with a low degree [32]. Degree centrality is given by $C_i = \deg(i)$. Our results show that all miRNA- target gene networks follow the Power law. *Fig.F9* shows two network samples from our study exhibiting a scale-free topology. Shown below are the degree distributions of a few disease networks that also validate the power law and therefore make them scale-free.

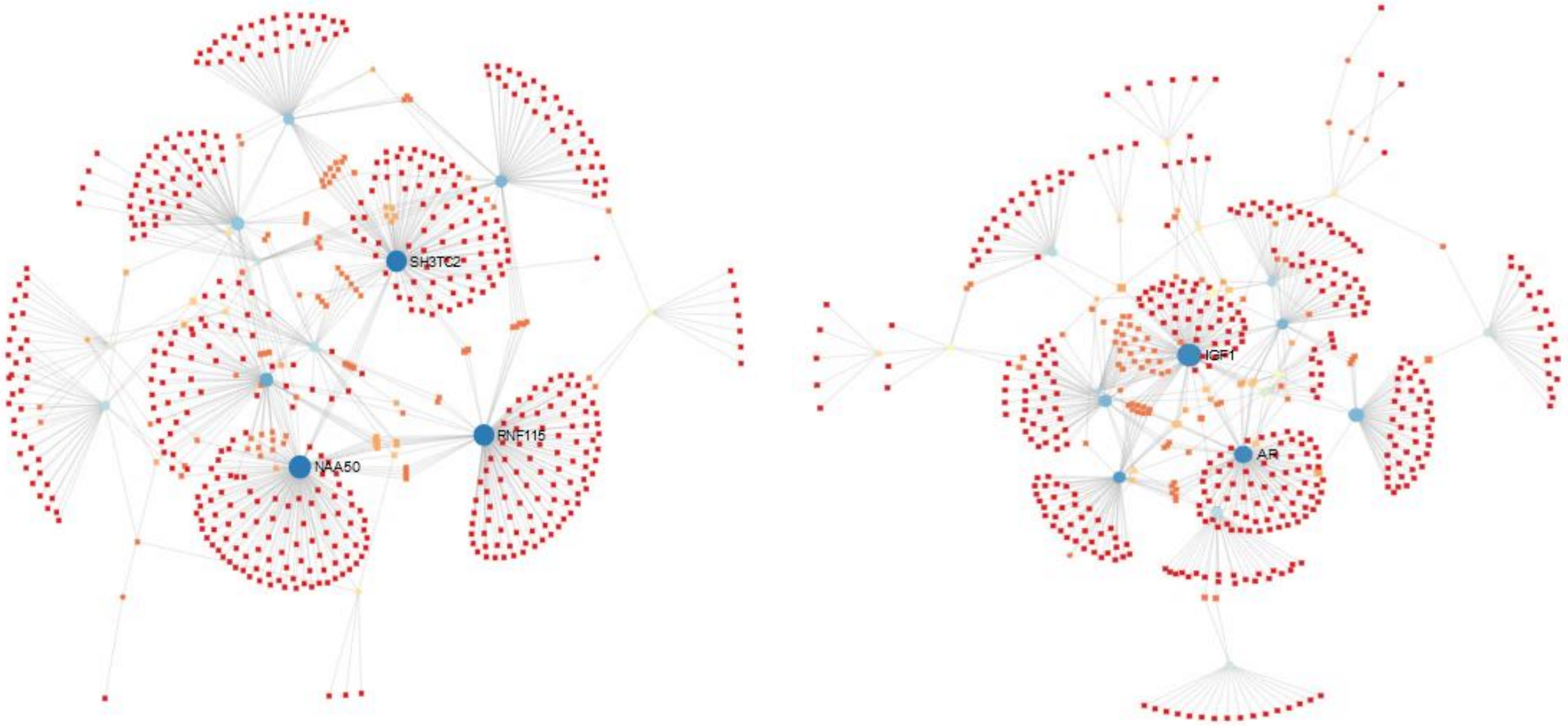


Fig. F9: *Scale-free property: (A) Charcot-Marie-Tooth Disease, Axonal, Type 2B (B) Polycystic ovary syndrome*

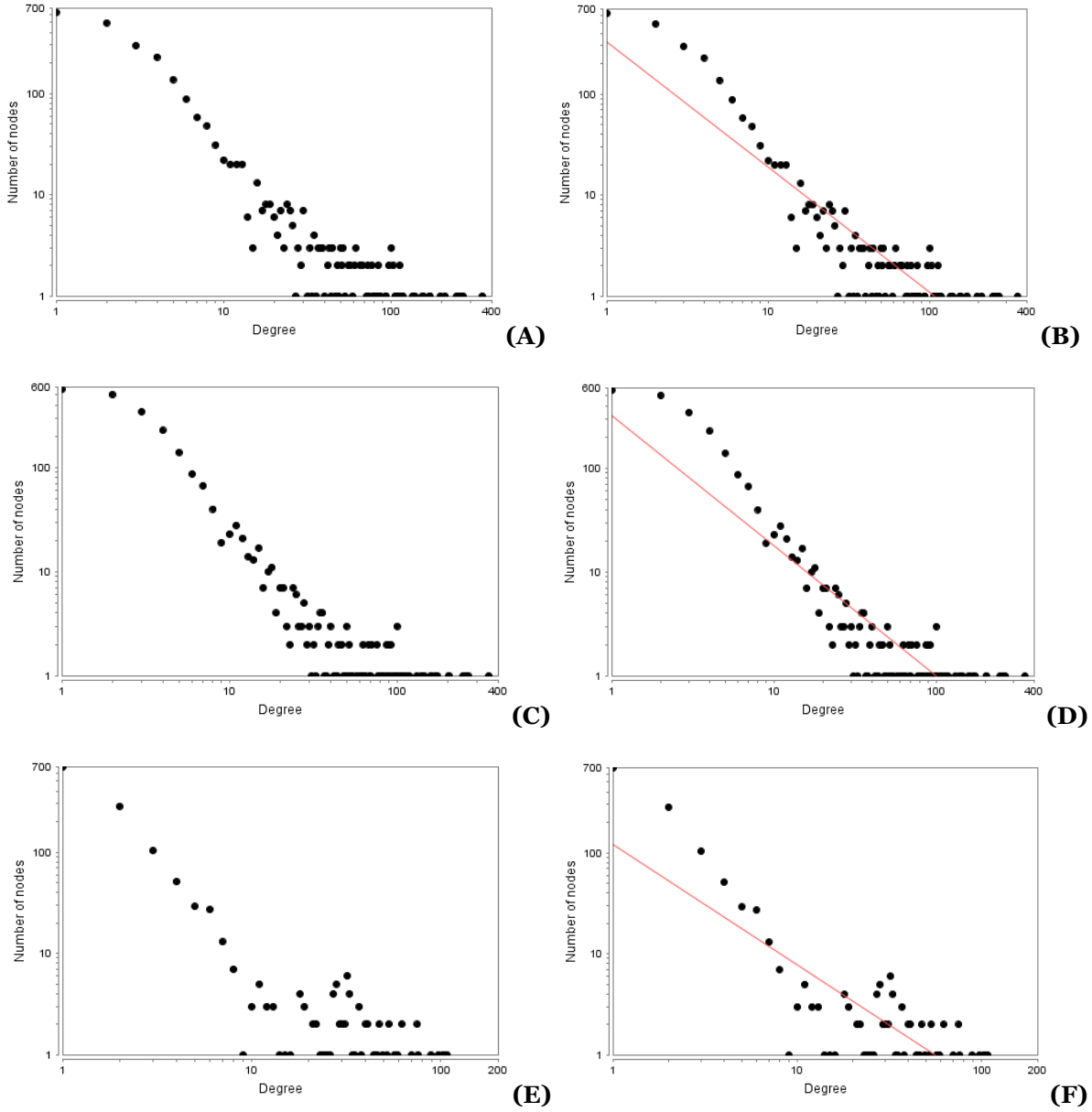


Fig. F10: *Degree distribution of miRNA-gene co-expression networks of (A) Pancreatic carcinoma (B) Power-law fitted degree distribution of (A) (C) Cervical cancer (D) Power-law fitted degree distribution of (C) (E) Cystic Fibrosis (F) Power-law fitted degree distribution of (E)*

However, it had been found that real systems observe a degree distribution that rarely follows a pure power law. For most real systems, P(k) has a slightly different shape thus deviating from the power-law behavior with a low-degree saturation and a high-degree cutoff and so is also the case with most biological networks. Given below is a degree distribution plot of the network of cervical cancer with a low degree saturation and a high degree cutoff, in a log-log, linear binning scale as shown in *Fig. F11*. The presence of these deviations might imply that a network might not be truly scale-free or the presence of other phenomena that are still not understood.

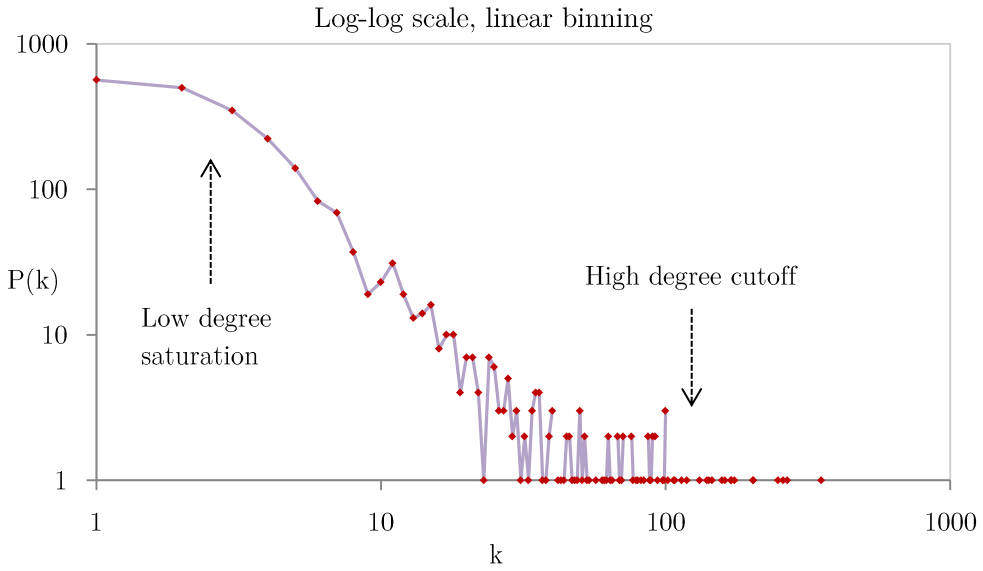


Fig. F11: *Deviations from Power-law property: Low degree saturation and high degree cut-off in biological networks*

F. DELETION ANALYSIS

Scatter-plots have been generated after eliminating the Rab involved in the diseases and reconstructing the networks. In *Fig. F12*, we have performed a computational deletion analysis by deleting Rabs and corresponding interactions. The degree distribution scatter-plots after deletion have been shown in each case, in comparison to their original degree distributions. The highlighted areas in the plots show few changes owing to loss of links made by Rabs, however, as in scale-free networks, these networks have mostly remained invariant and stable. But, invariability does not cover for a network's vulnerability to targeted attack: if important links are broken off, the network loses connections and forms a set of isolated sub-networks. In our study, although the networks did not lose their overall connectivity, some very important links involved in crucial pathways might have been broken, thus imparting disease symptoms. This is in agreement with the fact that random mutations in the genome did not affect the overall topology of the S. cerevisiae protein network as demonstrated by Barabasi et al [33].

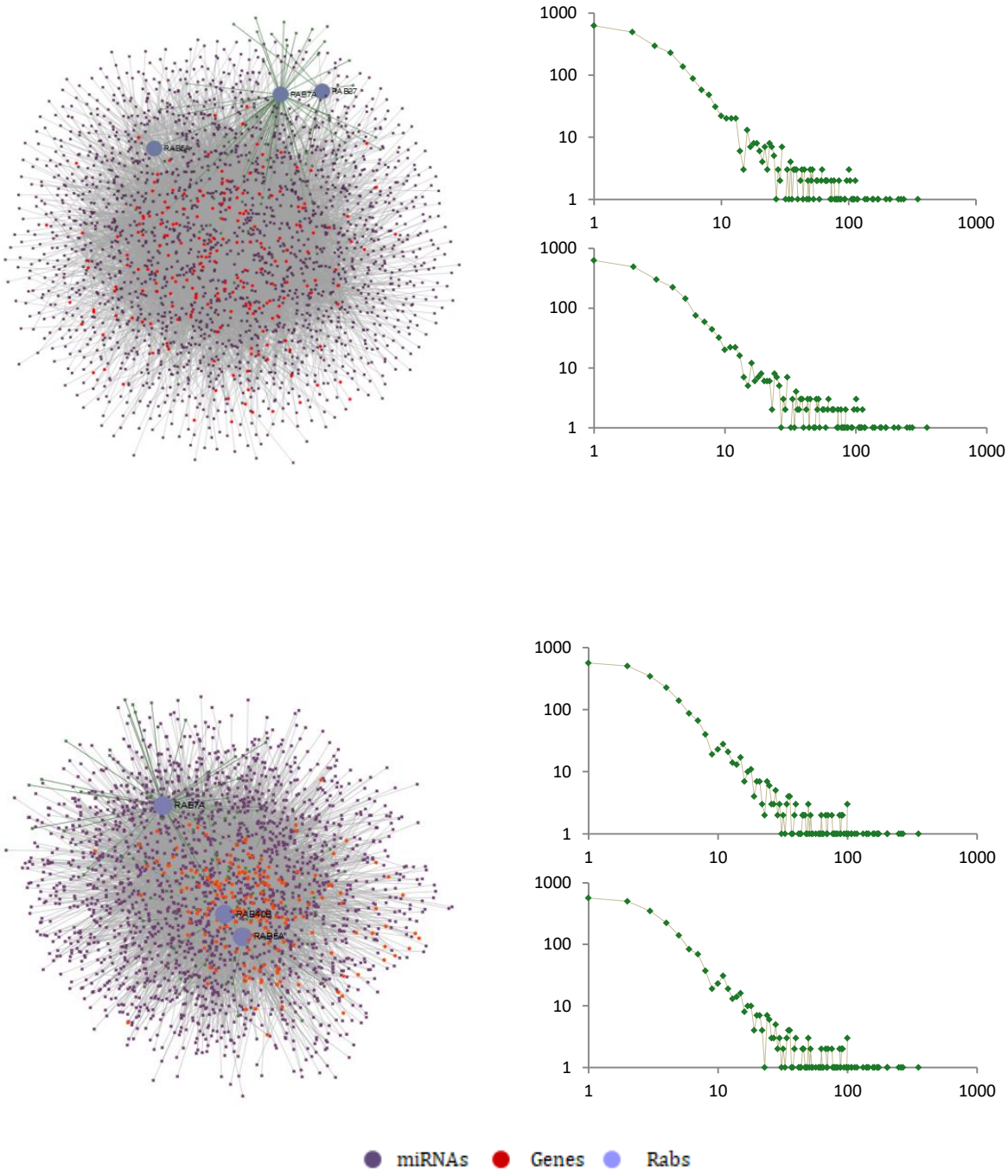


Fig. F12: *For each of the given networks, a degree distribution scatter-plot is shown (top) and a second scatter-plot showing the degree distribution after computational deletion of Rabs (A) Pancreatic Carcinoma (B) Cervical cancer*

G. NETWORK CENTRALITIES

1. Clustering coefficient

The clustering coefficient C_i shows whether a network or a node tends to form clusters or tightly connected communities. Mathematically, it is defined as:

$$C_i = \frac{2e}{k(k-1)}$$

where k is the degree and e is the number of edges between these k neighbors. Nodes have a clustering coefficient closer to 1 tend to form clusters or tight associations in a network. To characterize the network as a whole, the average clustering coefficient is generally considered, as well as average clustering coefficient C(k) over the node degree k. *Supplementary Table S4 and S5* lists the Average clustering coefficient plots for all the diseases in our study. In most of the plots, we find a general trend: for nodes with higher degrees, the average clustering coefficient is low and for nodes with low degrees, the average clustering coefficient is high. Some examples of clustering coefficient are shown below in *Fig. 13.*

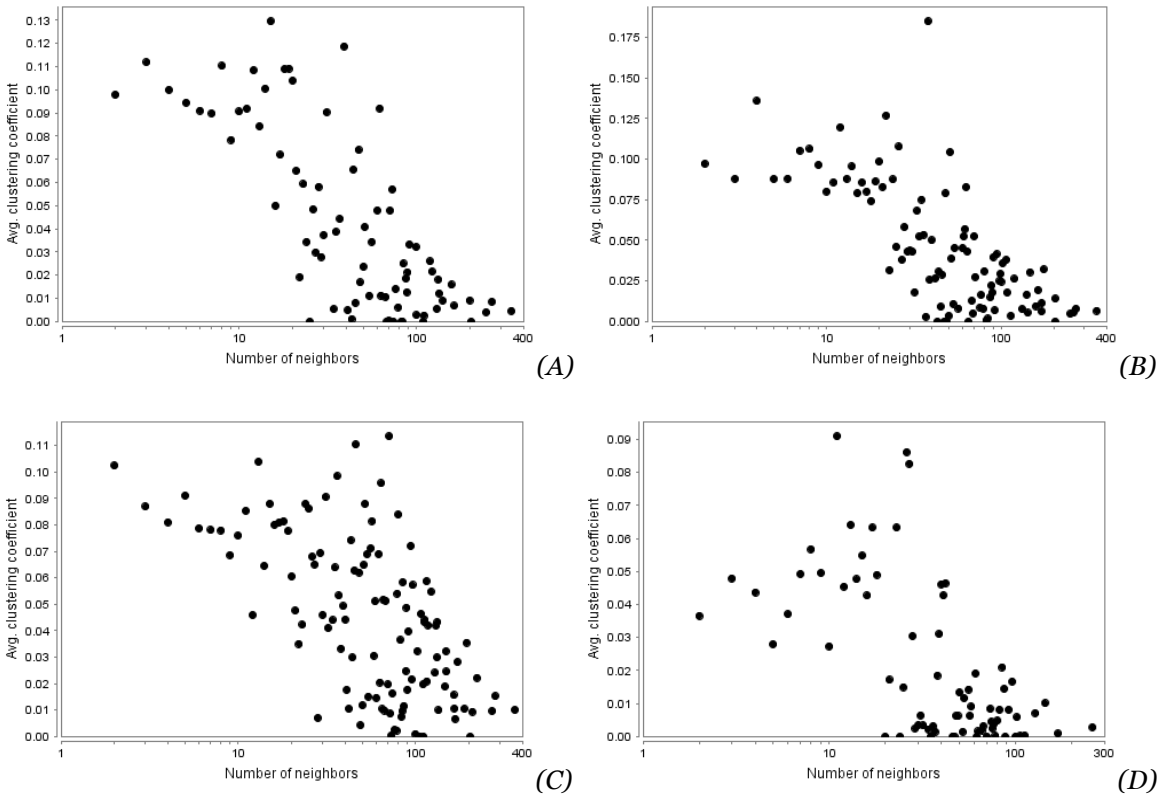


Fig. F13: *Average clustering coefficient plots for (A) bladder carcinoma (B) Cervical cancer (C) Breast carcinoma (D) Leukemia*

2. *Closeness centrality*

The closeness centrality is a measure to detect important nodes that can communicate quickly with other nodes in a network. For a graph G= (V, E), it is defined as

$$C_{clo} = \frac{1}{\sum dist_{ij}}$$

Some examples of closeness centrality plots observed are shown below in *Fig. F14.*

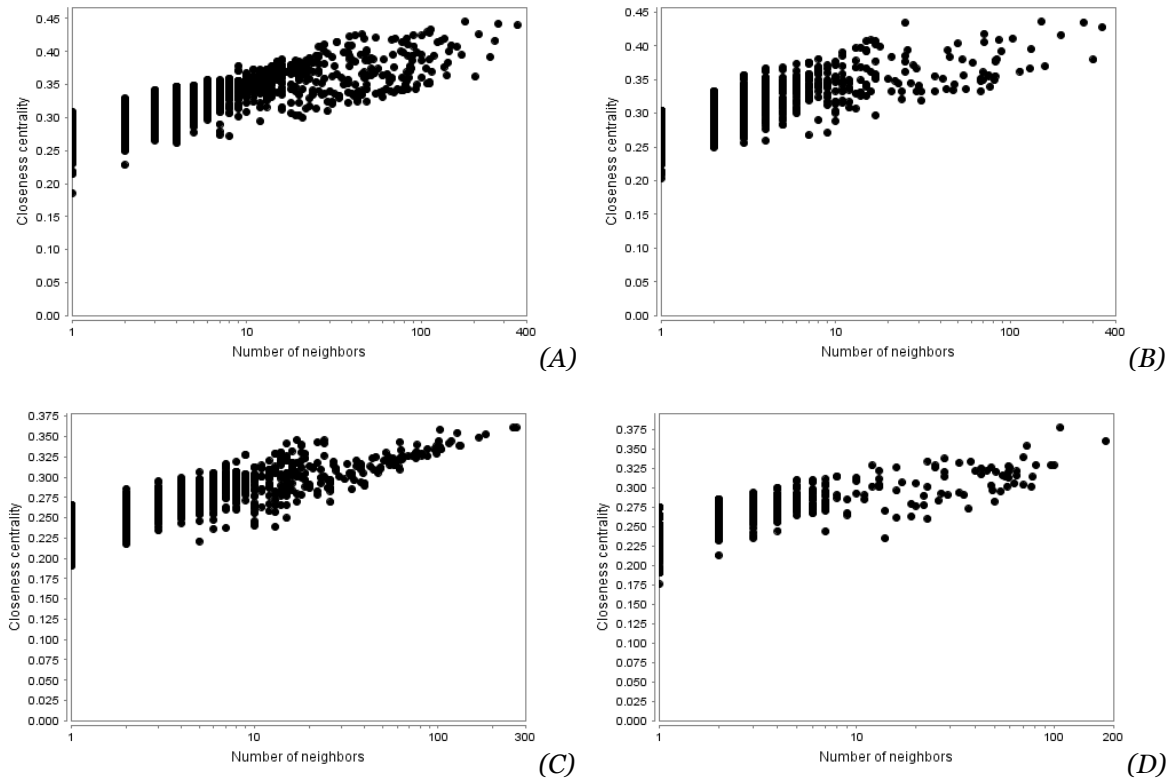


Fig. F14: *Closeness centrality plots for (A) Prostate carcinoma (B) Colorectal carcinoma (C) Autistic disorder (D) Alzheimer's disease*

3. *Betweenness centrality*

Betweenness centrality shows that nodes intermediate between neighbors are ranked higher. It is calculated as:

$$C_{bet(i)} = \frac{\sigma_{xy}(i)}{\sigma_{xy}}$$

Where $\sigma_{xy}$is the total number of shortest paths from node x to node y and $\sigma_{xy}(i)$ is the number of those paths that pass through node i. Nodes with high centrality in biological networks have been shown to play an important role in the network. *Fig. F15* illustrates few plots showing Betweenness centrality distributions.

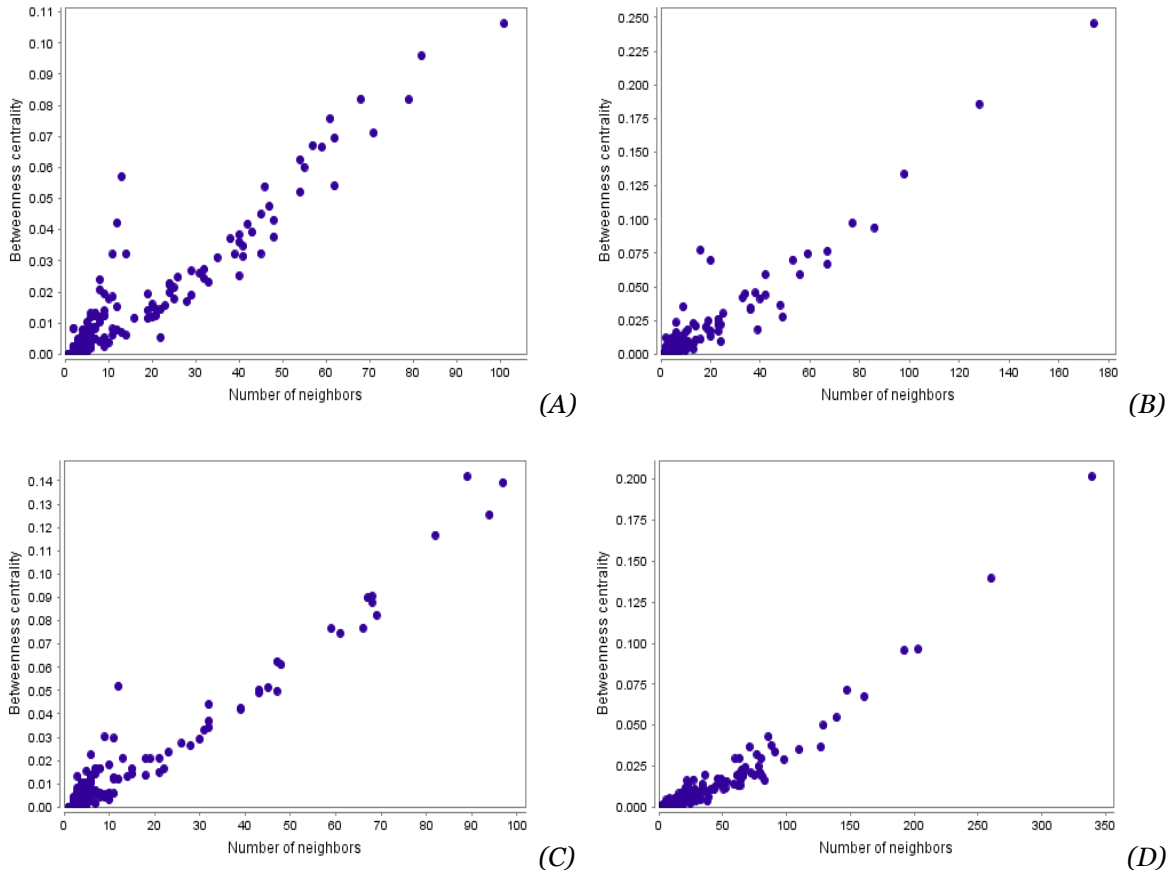


Fig. F15: *Betweenness centrality plots for (A) Rett syndrome (B) Rheumatoid arthritis (C) Down syndrome (D) Neuroblastoma*

4. *Power law fit*

In most of the plots, we find a general trend: for nodes with higher degrees, the average clustering coefficient is low and for nodes with low degrees, the average clustering coefficient is high. The closeness centrality is a measure to detect important nodes that can communicate quickly with other nodes in a network. Betweenness centrality shows that nodes intermediate between neighbors are ranked higher. Nodes with high centrality in biological networks have been shown to play an important role in the network. A trend analysis was performed and the curve of the form $y = ax^b$ gave the best fit on the clustering coefficient, closeness centrality, and betweenness centrality plots. *Fig. F16, F17* and *F18* shows scatter-plots for the above centralities in few disease networks.

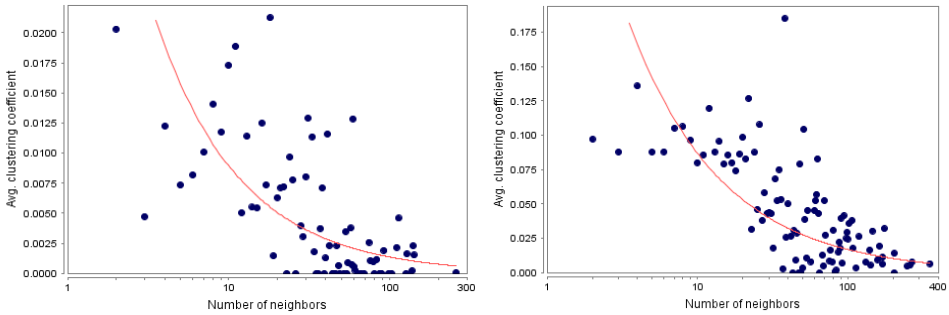


Fig. F16: *Power-law fitted clustering coefficient plots of (A) Amyotrophic Lateral Sclerosis (B)Cervical cancer*

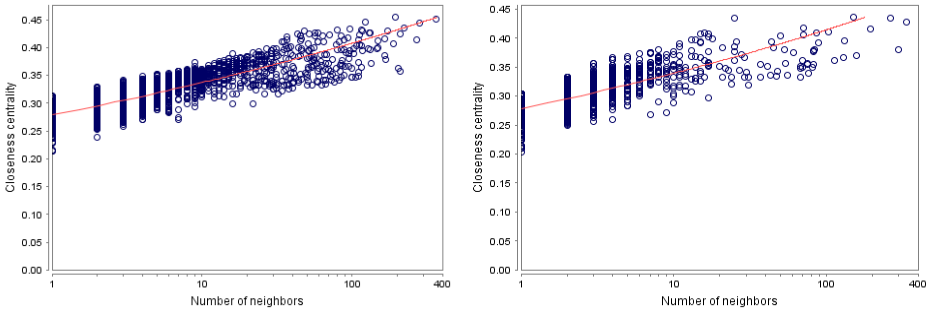


Fig. F17: *Power-law fitted closeness centrality plots of (A) Breast carcinoma (B)Colorectal carcinoma*

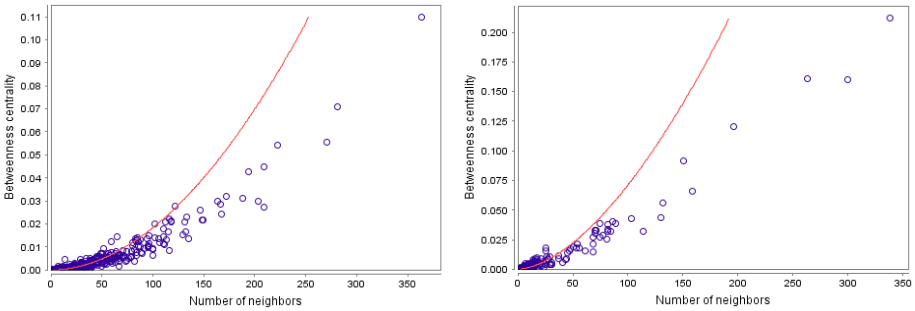


Fig. F18: *Power-law fitted betweenness centrality plots of (A) Breast carcinoma (B)Colorectal carcinoma*

8. Extraction and analysis of miRNA–target Rab sub-networks

A. OVERVIEW

For the past decade, the study of biological networks and analyzing them have become increasingly important because of the increasingly complex data that is being generated. Network visualization is still a fundamental approach in understanding of biological networks- therefore this chapter deals with miRNA-target Rab networks for Rab-associated diseases and understand how the Rab family of small GTPases, are affected.

B. EXTRACTION OF MIRNA-TARGET RAB SUB-NETWORKS

Rab subnetworks for each disease were extracted using Gephi's data laboratory and reconstructed in yED Graph Editor in various tidy layouts. Some examples are shown below in *Fig. F19, Fig, F20,* and *Fig. F21.* For the extracted networks, Rabs are highlighted in orange and miRNAs in blue.

C. NETWORK ANALYSIS

The extracted networks were analyses using the Network Analyzer plugin of Cytoscape and the results recorded in *Table T5.*

D. GRAPH LAYOUTS

(a) *Organic layout:* Organic layout is generally preferred for undirected graphs [105]. This algorithm identifies connected subnetworks and uses three physical forces – repulsion, springs and network energy – to spread nodes apart and stretch out links to form a tree-like structure in such a way that the sum of the forces emitted by the nodes and the edges reaches a (local) minimum [106], thus producing clear representations of complex networks.

(b) *Circular Layout:* This layout shows interconnected ring and star topologies and emphasizes group and tree structures within a network. This algorithm redraws the graph by introducing clusters, where the nodes are placed on the

circumference of an embedding circle, and edges are represented as single straight lines.

(c) *BCC Compact Layout:* In this layout, each partition represents a bi-connected component of the graph, which means that nodes are reachable by two edge-disjoint paths. Nodes belonging to more than one bi-connected component are assigned exclusively to one partition.

(d) *BCC Isolated Layout:* In this layout, nodes belonging to more than one bi-connected component are assigned an isolated partition.

(e) *Single Cycle Layout:* All nodes are represented as a single partition.

(f) *Tree Layout:* A tree is a graph that contains no undirected cyclic edge path. Nodes are placed side by side and the incident edges are routed in a bus-style fashion.

(g) *Balloon Layout:* Balloon layout requires a tree or a collection of trees as its input and the output will be a tree rooted in a quasi-radial style, called the balloon style. This layout is generally preferred for representing undirected, dense, or huge trees with a high number of nodes on a single hierarchy level.

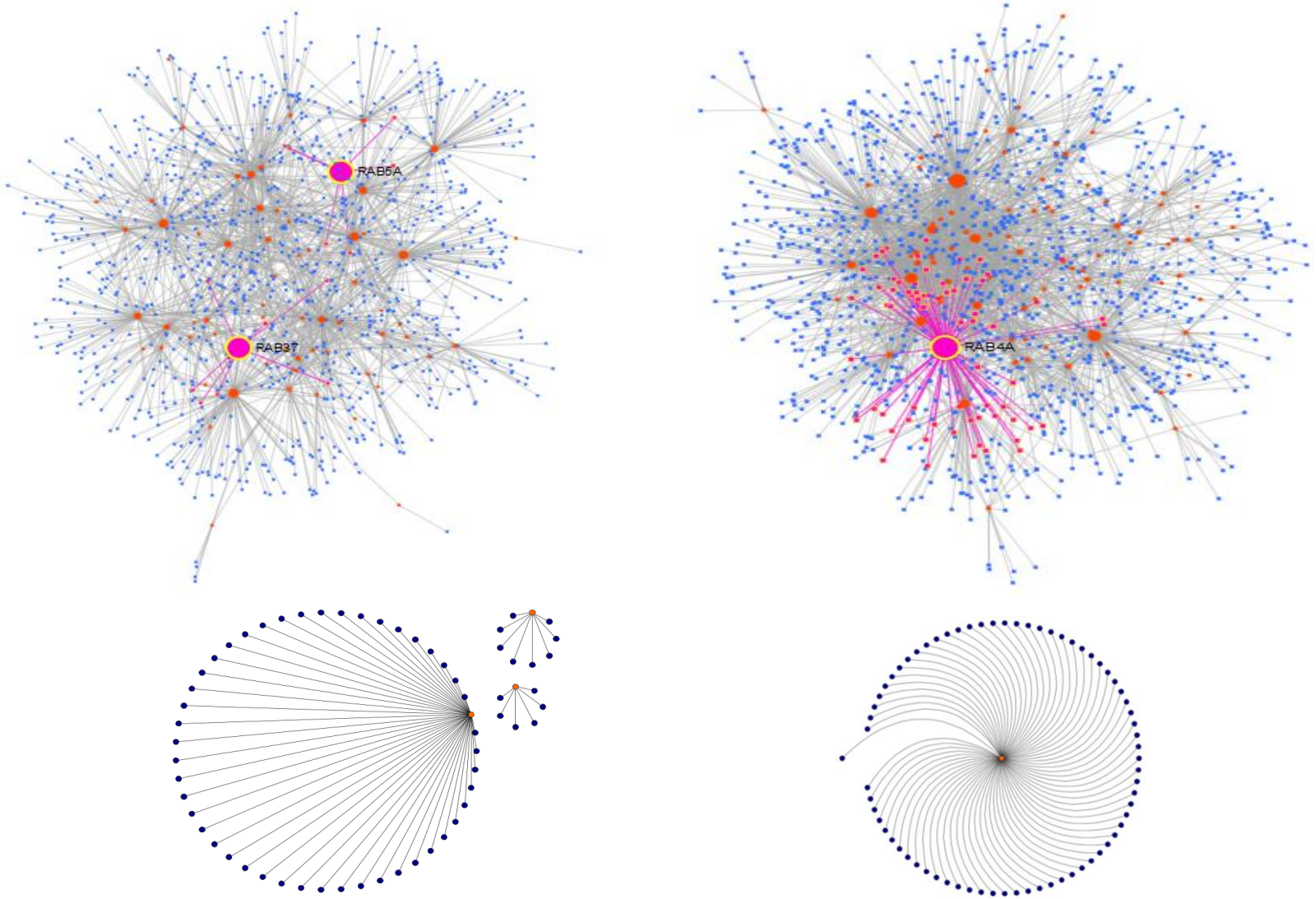


Fig. F19: *Extracted miRNA-Rab sub-networks of (a)Down Syndrome (b) Lupus Erythematosus, Systemic from their respective miRNA- target gene co-expression networks. Rab5a, Rab37, Rab7a in Down Syndrome, and Rab4a in Lupus Erythematosus, Systemic; have been extracted by reconstructing the whole network in yED Graph editor and viewed in Circular, single cycle and Circular, BCC compact layouts respectively. (miRNA marked in blue, Rabs in orange)*

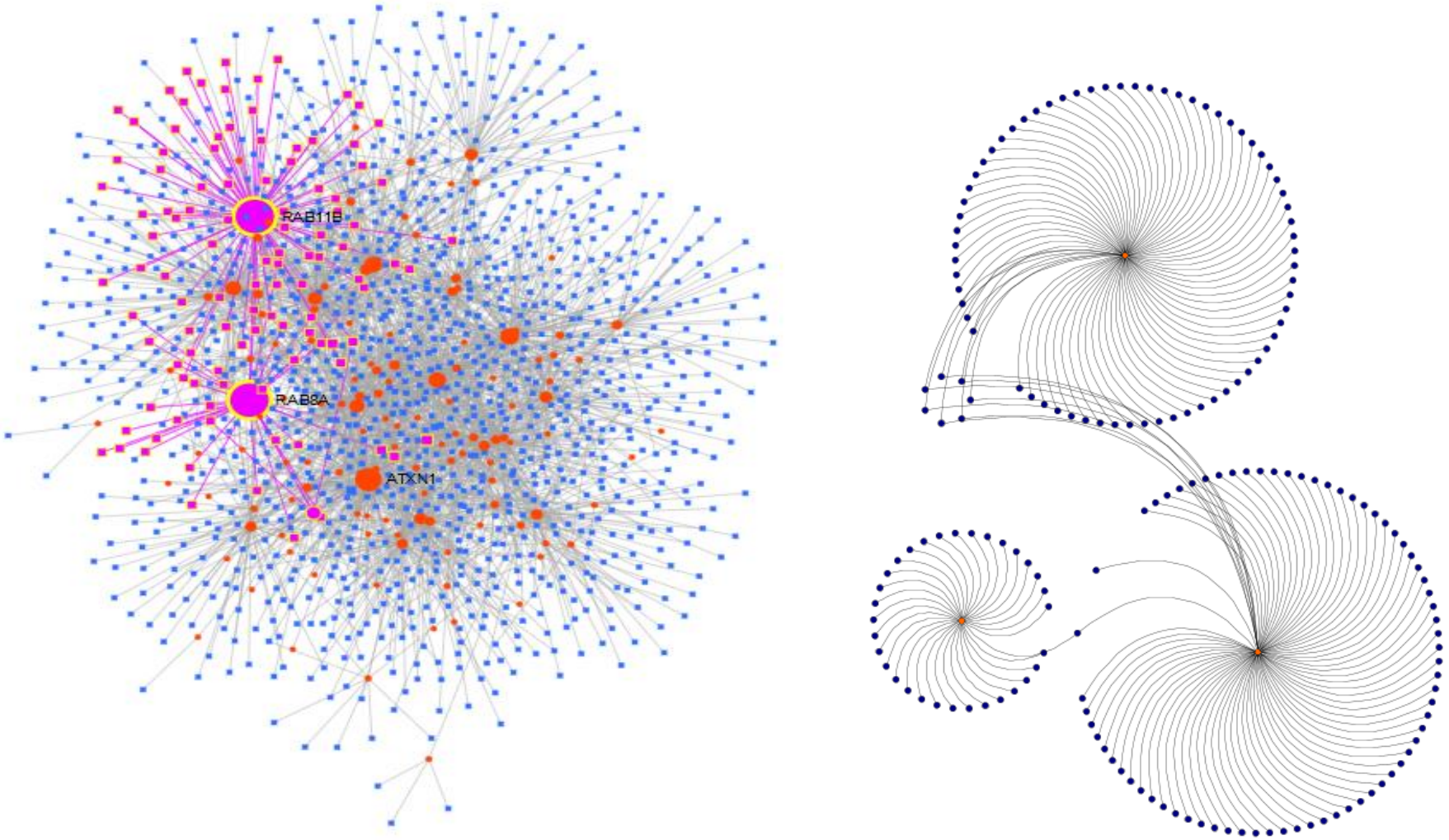


Fig. F20: *Photoreceptor degeneration miRNA-target gene co-expression network (left) generated in MiRNet, with its extracted Rab sub-network comprising Rab8a, Rab1b, and Rab11b along with their target miRNAs (right) reconstructed using yED Graph editor in a circular, BCC-isolated layout.*

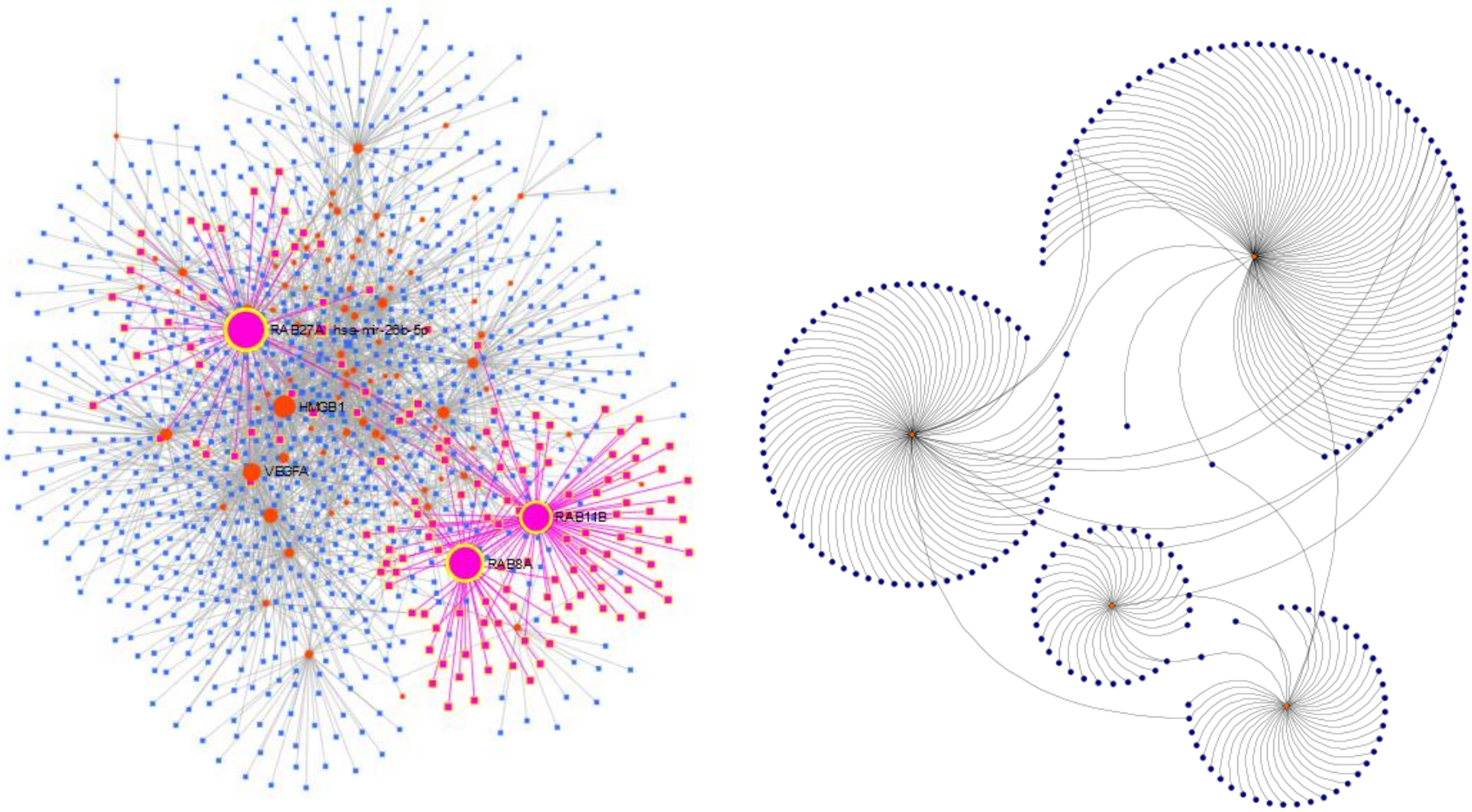


Fig. F21: *Rheumatoid arthritis miRNA-target gene co-expression network (left) generated in MiRNet, with its extracted Rab sub-network comprising Rab8a, Rab1b, Rab11b, Rab27a along with their target miRNAs (right) reconstructed using yED Graph editor in a tree-balloon layout*

RAB NETWORK ANALYSIS- CYTOSCAPE

Disease	*Node count*	*Edge count*	*Connected components*	*Network diameter*	*network density*	*Average number of neighbors*	*Network heterogeneity*	*Network centralization*
Charcot-Marie-Toothdisease,	105	101	4	2	0.018	1.924	3.204	0.432
Osteoporosis	4	3	1	2	0.5	1.5	0.577	1
Pain Disorder	101	98	3	2	0.019	1.941	3.237	0.449
Sclerocystic Ovaries	73	73	1	4	0.028	2	3.001	0.628
Myocardial Ischemia	200	208	1	6	0.01	2.08	3.509	0.38
Polycystic Ovary Syndrome	73	73	1	4	0.028	2	3.001	0.628
Cellular Ependymoma	69	67	2	2	0.029	1.942	3.535	0.849
Acute Coronary Syndrome	3	0	2	0.5	1.5	0.577	1	
Rheumatoid Arthritis	216	225	1	6	0.01	2.083	3.803	0.352
Colorectal Cancer	135	135	1	4	0.015	2	4.477	0.742
Lupus Erythematosus, Systemic	71	70	1	2	0.028	1.972	4.124	1
Alzheimer's Disease	357	411	2	8	0.006	2.303	3.492	0.211
Acute Erythroblastic Leukemia	180	185	1	6	0.011	2.056	3.986	0.423
Down Syndrome	63	60	3	2	0.031	1.905	2.993	0.735
Neuroblastoma	7	6	1	2	0.286	1.714	1.021	1
Dementia	69	67	2	2	0.029	1.942	3.535	0.849
Breast Carcinoma	788	1029	1	8	0.003	2.612	3.864	0.229
leukemia	487	558	1	6	0.005	2.292	3.633	0.202
Liver carcinoma	634	766	1	8	0.004	2.416	3.538	0.164
Yellow Fever	180	185	1	6	0.011	2.056	3.986	0.423
Prostate carcinoma	382	445	1	8	0.006	2.33	3.221	0.197
Tongue Carcinoma	122	122	2	6	0.017	2	3.074	0.37
Parkinson Disease	275	298	2	10	0.008	2.167	3.256	0.268
Autistic Disorder	155	160	2	6	0.013	2.065	3.067	0.368
Congestive heart failure	122	122	2	6	0.017	2	3.074	0.37
Colorectal Carcinoma	135	135	1	4	0.015	2	4.477	0.742
Ataxia Telangiectasia	122	122	2	6	0.017	2	3.074	0.37
Asthma	194	206	3	8	0.011	2.124	2.819	0.23
Amyotrophic Lateral Sclerosis	139	137	4	8	0.014	1.971	2.943	0.234

Ciliopathies	180	185	1	6	0.011	2.056	3.986	0.423
Photoreceptor degeneration	180	185	1	6	0.011	2.056	3.986	0.423
cervical cancer	94	90	4	2	0.021	1.915	2.881	0.484
Mild cognitive disorder	63	60	3	2	0.031	1.905	2.993	0.735
Carcinoma of lung	205	204	3	6	0.01	1.99	3.991	0.485
Hereditary Sensory Autonomic Neuropathy, Type 1	101	98	3	2	0.019	1.941	3.237	0.449
Choroideremia	101	98	3	2	0.019	1.941	3.237	0.449
Amyloidosis	105	101	4	2	0.018	1.924	3.204	0.432
Sjogren's Syndrome	71	70	1	2	0.028	1.972	4.124	1
Multiple Sclerosis	71	70	1	2	0.028	1.972	4.124	1
Renal Cell Carcinoma	63	60	3	2	0.031	1.905	2.993	0.735
Pancreatic carcinoma	113	109	4	4	0.017	1.929	3.096	0.401
Ovarian Carcinoma	245	245	3	6	0.008	2	3.974	0.405
Lyme Disease	123	119	4	2	0.016	1.935	3.41	0.475
Encephalitis, St. Louis	71	70	1	2	0.028	1.972	4.124	1
Rett Syndrome	50	49	1	2	0.04	1.96	3.429	1
Non-Small Cell Lung Carcinoma	123	125	1	6	0.017	2.002	3.041	0.391
Renal carcinoma	83	82	1	6	0.024	1.976	3.079	0.588
Acromegaly	68	68	1	4	0.03	2	2.97	0.667
Martsolf syndrome	68	68	1	4	0.03	2	2.97	0.667
Obesity	171	181	1	4	0.012	2.117	3.411	0.416
Thyroid carcinoma	113	120	2	4	0.019	2.124	2.92	0.339
Skin Carcinogenesis	89	90	2	4	0.023	2.022	3.118	0.511
Hypercholesterolemia	35	33	2	2	0.055	1.886	2.563	0.877
Lipoidosis	35	33	2	2	0.055	1.886	2.563	0.877
Griscelli Syndrome, Type 2	41	40	1	2	0.049	1.951	3.083	1
Acrocephalopolysyndactyly type 2	42	41	1	2	0.048	1.952	3.123	1
Craniosynostosis	42	41	1	2	0.048	1.952	3.123	1
Smith-Mccort Dysplasia 2	54	53	1	2	0.037	1.963	3.571	1
Intellectual Disability	79	81	1	4	0.026	2.051	3.024	0.512
Gastrointestinal Stromal Tumors	42	41	1	2	0.048	1.952	3.123	1
Polydactyly	42	41	1	2	0.048	1.952	3.123	1
Coronary Artery Disease	9	8	1	2	0.222	1.778	1.237	1
Micrognathism	42	41	1	2	0.048	1.952	3.123	1
Pulmonary Stenosis	42	41	1	2	0.048	1.952	3.123	1
Brachydactyly	42	41	1	2	0.048	1.952	3.123	1
Syndactyly of fingers and toes	42	41	1	2	0.048	1.952	3.123	1
Severe myopia	3	2						

Colon Carcinoma	125	122	3	2	0.016	1.952	4.612	0.804
Albinoidism, Oculocutaneous	44	43	1	4	0.045	1.955	2.973	0.927
Osteosarcoma	65	63	2	2	0.03	1.938	3.684	0.92
Carcinoma of bladder	166	166	2	6	0.012	2	4.181	0.601
Cystic Fibrosis	62	70	1	4	0.037	2.258	2.682	0.639
Refractory anemias	41	40	1	2	0.049	1.951	3.083	1
Stomach Carcinoma	191	194	1	6	0.011	2.031	4.517	0.612
Dermatitis, Atopic	31	30	1	2	0.065	1.953	2.647	1
Hermanski-Pudlak Syndrome	121	120	1	4	0.017	1.983	5.294	0.975
Marinesco-Sjogren syndrome	118	117	1	2	0.017	1.983	5.362	1
Albinism	44	43	1	4	0.045	1.955	2.973	0.927
Waisman syndrome	21	20	1	2	0.095	1.905	2.124	1
Martin-Probst Deafness-Mental Retardation Syndrome	21	20	1	2	0.095	1.905	2.124	1
Mental Retardation, X-Linked	21	20	1	2	0.095	1.905	2.124	1
Strabismus	21	20	1	2	0.095	1.905	2.124	1
Esophageal carcinoma	31	30	1	2	0.065	1.935	2.647	1
Muscular Dystrophy, Duchenne	37	36	1	2	0.054	1.946	2.917	1
Fatty Liver	31	30	1	2	0.065	1.935	2.647	1
Diabetes	4	3	1	2	0.5	1.5	0.577	1
T-Cell Lymphoma	31	30	1	2	0.065	1.935	2.647	1

Table T5: *miRNA-target Rab subnetwork analysis using Cytoscape Network Analyzer*

9. Validation

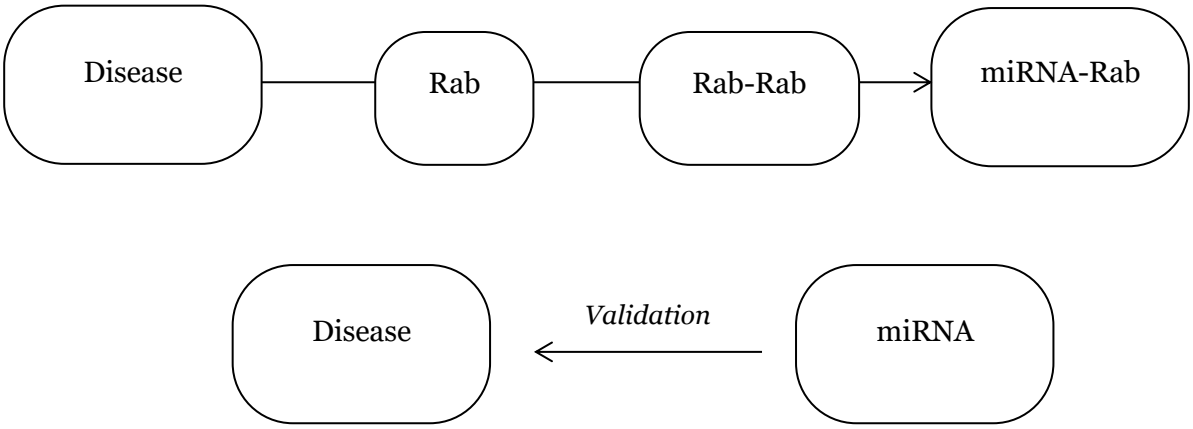


Fig. F22: *(A) Schematic representation of the workflow to create a miRNA-target RAB network. (B) Schematic representation of a reverse network to validate the above.*

Following our work protocol, we have proposed miRNA-target Rab networks for all the Rab associated diseases mentioned, but it is essential to note that the networks are a representation of the possible ways each disease can be caused by way of miRNA acting upon its target Rab. For example, in the undirected network proposed for Parkinson's disease, we have a total of 275 nodes and 298 edges, comprising 10 nodes for Rab and 265 for miRNAs that target these Rab. A comprehensive literature search, reverse network approach and enrichment analysis revealed that only three miRNAs act through Rab proteins to cause Parkinson's disease, namely, hsa-mir-214, hsa-mir-221, and hsa-mir-30e.

In other words, the rest 265 miRNAs could potentially play a similar role as the rest three but have not yet been recorded. Hence, this chapter deals with all the disease networks proposed (Disease to miRNA) in this study and validating them through a reverse network analysis (miRNA to Disease). For simplicity purposes, the diseases have been grouped into the following nine sets and further investigated: Set-1: Cardiovascular diseases, Set-2: Cancers, Set-3: Neurodegenerative disorders, Set-4: Endocrine and metabolic disorders, Set-5: Endocrine disorders, Set-6: Chromosomal abnormalities, Set-7: Immunological disorders, Set-8: Musculoskeletal disorders and Set-9: Hematological disorders. As discussed in the previous chapter, Rab-miRNA networks have been built following the protocol shown above, using Rab-Rab

interaction data and enriching them with their miRNA targets using miRNet, a miRNA-centric network visual analytics platform. A similar, but reverse approach has been used to obtain the miRNAs that have been recorded in databases and literature only. Furthermore, these lists of miRNAs have been used to carry out further analysis.

To validate, using the protocol shown above (miRNA to disease), we prepared a summary table comprising the Rab family of GTPases and their miRNA targets. The miRNA targets for each Rab have been collected and validated from at least two of the following databases: MiRBase, miRanda, TargetScan, microT, miRBase, picTar, and miRecords. Using these Rab and their target miRNAs as our input for each disease, we looked for diseases associated with these miRNA sets. To reconstruct the network, we used Gephi-0.9.2 to rule out the miRNAs that were not shown to cause disease through Rab alterations or regulations (but perhaps are involved). *Table T6* records the miRNAs that have been obtained for each group of diseases. (Set 1 through Set 9). These extracted miRNA sets comprise miRNA that has been verified to cause diseases through making alterations in Rab proteins, hence are an important part of this study and carried forward for further analysis. We list the number of miRNAs obtained in each set through validation as follows:

Set-1 (Cardiovascular diseases): 6, Set-2 (Cancers): 113, Set-3 (Neurodegenerative disorders): 50, Set-4 (Endocrine and metabolic disorders): 6, Set-5 (Endocrine disorders): 4, Set-6 (Chromosomal abnormalities): 1, Set-7 (Immunological disorders): 2, Set-8 (Musculoskeletal disorders): 2. Set-9 (Hematological disorders): 4

Set 1

hsa-let-7b
hsa-let-7c
hsa-let-7e
hsa-let-7g
hsa-let-7i
hsa-mir-101-3p
hsa-mir-103a-3p
hsa-mir-106a
hsa-mir-106b
hsa-mir-107
hsa-mir-130a
hsa-mir-130b
hsa-mir-134
hsa-mir-135b
hsa-mir-137
hsa-mir-141
hsa-mir-143
hsa-mir-144
hsa-mir-145
hsa-mir-148a
hsa-mir-148b
hsa-mir-149
hsa-mir-152
hsa-mir-155
hsa-mir-15a
hsa-mir-17
hsa-mir-17-5p
hsa-mir-181a-5p
hsa-mir-181b-5p
hsa-mir-181c
hsa-mir-181d
hsa-mir-182
hsa-mir-185
hsa-mir-186
hsa-mir-18a
hsa-mir-18b
hsa-mir-190a-5p
hsa-mir-192
hsa-mir-19a
hsa-mir-19b-3p
hsa-mir-200a
hsa-mir-200b
hsa-mir-200c
hsa-mir-204
hsa-mir-206
hsa-mir-20a
hsa-mir-21
hsa-mir-210
hsa-mir-211
hsa-mir-212
hsa-mir-214
hsa-mir-215
hsa-mir-216a
hsa-mir-216b
hsa-mir-221
hsa-mir-222
hsa-mir-23a
hsa-mir-23b
hsa-mir-25
hsa-mir-26b
hsa-mir-27b
hsa-mir-29a
hsa-mir-29b-3p
hsa-mir-29c
hsa-mir-301a
hsa-mir-301b
hsa-mir-302a
hsa-mir-302b
hsa-mir-302c
hsa-mir-302d
hsa-mir-30a
hsa-mir-30b
hsa-mir-30d
hsa-mir-30e
hsa-mir-31
hsa-mir-32
hsa-mir-338
hsa-mir-340
hsa-mir-34a
hsa-mir-34b
hsa-mir-34c
hsa-mir-363
hsa-mir-367
hsa-mir-372
hsa-mir-373
hsa-mir-374a
hsa-mir-374b
hsa-mir-380
hsa-mir-381
hsa-mir-383
hsa-mir-410
hsa-mir-429
hsa-mir-433
hsa-mir-454
hsa-mir-494
hsa-mir-495
hsa-mir-497
hsa-mir-498
hsa-mir-506
hsa-mir-519d
hsa-mir-570
hsa-mir-572
hsa-mir-622
hsa-mir-627
hsa-mir-646
hsa-mir-663a
hsa-mir-7-5p
hsa-mir-92b
hsa-mir-93
hsa-mir-939
hsa-mir-944
hsa-mir-96
hsa-mir-98

Set 2

hsa-mir-101-3p
hsa-mir-124-3p
hsa-mir-130a
hsa-mir-135b
hsa-mir-141
hsa-mir-144
hsa-mir-149
hsa-mir-152
hsa-mir-155
hsa-mir-181c
hsa-mir-186
hsa-mir-192
hsa-mir-19a
hsa-mir-19b-3p
hsa-mir-204
hsa-mir-206
hsa-mir-211
hsa-mir-212
hsa-mir-214
hsa-mir-215
hsa-mir-216a
hsa-mir-216b
hsa-mir-221
hsa-mir-222
hsa-mir-29b-3p
hsa-mir-29c
hsa-mir-300
hsa-mir-302b
hsa-mir-30a
hsa-mir-30b
hsa-mir-30d
hsa-mir-30e
hsa-mir-32
hsa-mir-340
hsa-mir-381
hsa-mir-410
hsa-mir-429
hsa-mir-448
hsa-mir-454
hsa-mir-494
hsa-mir-541
hsa-mir-544a
hsa-mir-549a-3p
hsa-mir-634
hsa-mir-661
hsa-mir-802
hsa-mir-892a
hsa-mir-939
hsa-mir-944
hsa-mir-9-5p

Set 3

hsa-mir-29c
hsa-mir-29b-3p

Set 4

hsa-mir-21
hsa-mir-216a

Set 5

hsa-mir-31
hsa-mir-182
hsa-mir-134
hsa-mir-340

Set 6

hsa-mir-130a
hsa-mir-200b
hsa-mir-222
hsa-mir-26b

Set 7

hsa-mir-130a
hsa-mir-200b
hsa-mir-222
hsa-mir-26b
hsa-mir-31

Set 8

hsa-mir-802

Set 9

hsa-mir-21
hsa-mir-212
hsa-mir-130a
hsa-mir-155
hsa-mir-30e
hsa-mir-340

Table T6: *List of miRNAs in each disease set*

10. MiRNA set enrichment analysis

A miRNA set enrichment analysis was conducted using TAM 2.0 webserver to investigate the functional and disease associations of individual sets of microRNAs under investigation (Set-1, Set-2, Set-3, Set-3, Set-4, Set-5, Set-6, Set-7, Set-8 and Set-9). Cardiovascular diseases are one of the leading causes of morbidity and mortality in the world. Recent years have seen an increase in the study of miRNAs as potential biomarkers with high sensitivity for early diagnosis [108]. Altered levels of circulating miRNAs have been found in acute coronary syndrome, stable coronary artery disease, and heart failure [109]. Our analysis revealed the following diseases/conditions which are also included in our original study, associated with the top cardiovascular diseases caused by various defects in Rab proteins: Coronary Atherosclerosis (FDR= 0.0549), Heart failure (FDR=1.02E-03), Myocardial-Ischemic-Reperfusion injury (FDR=0.0577). At least six miRNAs were found to be associated with heart failure (hsa-mir-340, hsa-mir-130a, hsa-mir-155, hsa-mir-30e, hsa-mir-212, hsa-mir-21), two with Coronary Atherosclerosis (hsa-mir-155, hsa-mir-21) and two with Myocardial Ischemic- Reperfusion Injury (hsa-mir-155, hsa-mir-21). Also, a study on various other Rab associated cardiovascular diseases revealed that four miRNAs were widely involved: hsa-mir-340, hsa-mir-130a, hsa-mir-21, hsa-mir-155. To summarize, the following miRNAs that target various members of the Rab family of GTPases have been found to be involved in Rab-associated cardiovascular diseases: hsa-mir-155, hsa-mir-21, hsa-mir-34, hsa-mir-130a, hsa-mir-30e, hsa-mir-212. *Fig. F23* represents a graphical comparison between the data from the enriched miRNAs set and the data obtained using a systems biology approach, which also validates that these miRNAs are perhaps widely involved in Rab-associated cardiovascular diseases, with heart failure related to all the six microRNAs found.

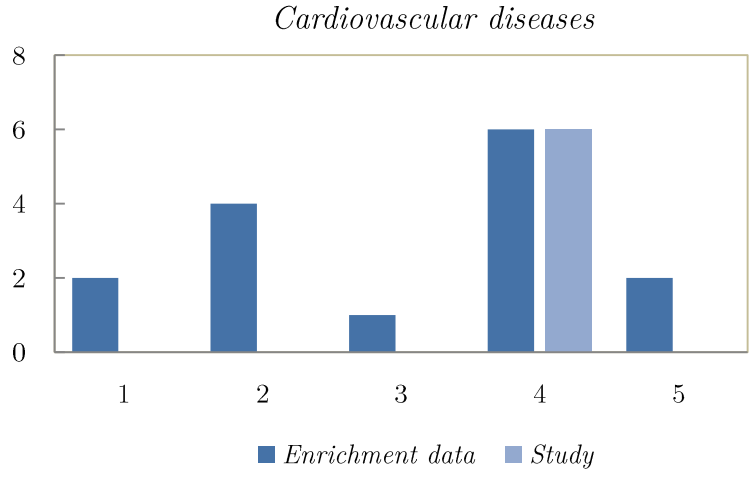


***Fig. F23**: Comparison of datasets in cardiovascular diseases*

Aberrant miRNA expressions have been reported in B-cell chronic lymphocytic leukemia [110]. Several cancers like liver, pancreas, esophagus, colon, stomach have been associated with both upregulated and downregulated miRNAs [111]. Lung and Breast carcinomas have also been reported to be associated with aberrant miRNA profiles [112]. Using the miRNAs from Set-2, an enrichment analysis was performed and it was reported that these miRNAs were involved in the following types of cancers: Breast carcinoma (FDR=1.93E-25), Colon carcinoma (FDR=1.21E-23), Lung Carcinoma (FDR=1.57E-14), Non-small-cell lung carcinoma (FDR=3.10E-18), Pancreatic Carcinoma (FDR=2.42E-19), Prostrate carcinoma (FDR=7.58E-19), Renal cell carcinoma (FDR=3.36E-15), Bladder carcinoma (FDR=3.38E-11), Gastrointestinal neoplasms (FDR=1.28E-17), Cervical carcinoma (FDR=1.97E-14), Thyroid carcinoma (FDR=1.07E-04), Gall bladder carcinoma (FDR=6.19E-04), Leukemia (FDR=2.53E-05), Skin Carcinoma (FDR=6.28E-05), Gastrointestinal carcinoma (FDR=0.4882), T-cell Lymphoma (FDR=5.64E-07) and clear-cell ovarian carcinoma (FDR=0.371). Among the set of miRNAs used, the following were found to be associated with a wide number of cancers: hsa-let-7b, hsa-let-7c, hsa-let-7e, hsa-let-7g, hsa-mir-101-1, hsa-mir-101-2, hsa-mir-106a, hsa-mir-106b, hsa-mir-107, hsa-mir-130a, hsa-mir-130b, hsa-mir-137, hsa-mir-141, hsa-mir-143, hsa-mir-144, hsa-mir-145, hsa-mir-148a, hsa-mir-148b, hsa-mir-152, hsa-mir-155, hsa-mir-15a, hsa-mir-17, hsa-mir-181a-1, hsa-mir-181a-2, hsa-mir-182, hsa-mir-18a, hsa-mir-19a, hsa-mir-200a, hsa-mir-200b, hsa-mir-200c, hsa-mir-204, hsa-mir-206, hsa-mir-20a, hsa-mir-21, hsa-mir-210, hsa-mir-212. Breast carcinoma, Colon carcinoma, prostrate carcinoma and non-small-cell carcinoma were found to be associated with 52, 50, 43 and 40 miRNAs respectively.

The two sets of data obtained from Enrichment and original study were compared and 43 miRNAs were reported to be common between them. In addition to these, several other miRNAs that target members of the Rab family of GTPase and may be involved in Rab-associated cancers were found: hsa-mir-17-5p, hsa-mir-181a-5p, hsa-mir-181b-5p, hsa-mir-181a-1, hsa-mir-181a-2, hsa-mir-181b-1, hsa-mir-181b-2, hsa-mir-190a, hsa-mir-19b-1, hsa-mir-19b-2, hsa-mir-190a-5p, hsa-mir-19b-3p, hsa-mir-214, hsa-mir-215, hsa-mir-216a, hsa-mir-216b, hsa-mir-221, hsa-mir-222, hsa-mir-23a, hsa-mir-23b, hsa-mir-25, hsa-mir-26b, hsa-mir-27b, hsa-mir-29a, hsa-mir-29b-3p, hsa-mir-29c, hsa-mir-301a, hsa-mir-301b, hsa-mir-302a, hsa-mir-302b, hsa-mir-302c, hsa-mir-302d, hsa-mir-30a, hsa-mir-30b, hsa-mir-30d, hsa-mir-30e, hsa-mir-31, hsa-mir-32, hsa-mir-338, hsa-mir-340, hsa-mir-34a, hsa-mir-34b, hsa-mir-34c, hsa-mir-363, hsa-mir-367, hsa-mir-372, hsa-mir-373, hsa-mir-374a, hsa-mir-374b, hsa-mir-380, hsa-mir-381, hsa-mir-383, hsa-mir-410, hsa-mir-429, hsa-mir-433, hsa-mir-454, hsa-mir-494, hsa-mir-495, hsa-mir-497, hsa-mir-498, hsa-mir-506, hsa-mir-519d, hsa-mir-570, hsa-mir-572, hsa-mir-622, hsa-mir-627, hsa-mir-646, hsa-mir-663a, hsa-mir-7-5p, hsa-mir-92b, hsa-mir-93, hsa-mir-939, hsa-mir-944, hsa-mir-96, hsa-mir-98. *Fig. F24* shows a comparison between the two data sets which shows that breast carcinoma, colon carcinoma, prostrate carcinoma and non-small-cell carcinoma might be associated with comparatively more miRNAs than other cancers are.

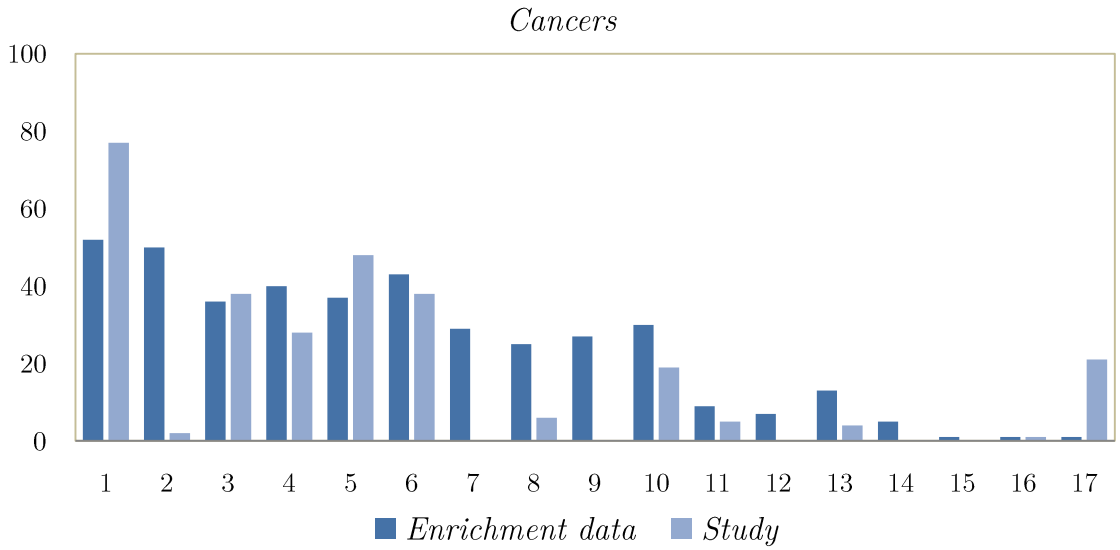


Fig. F24: *Comparision of datasets in cancers*

Dysregulation of miRNAs has been implicated in a growing number of NDDs including Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS) [113]. Similar enrichment analysis was performed for Set-3 and it was found that these miRNAs were involved in four major neurodegenerative diseases namely Alzheimer's Disease (FDR=2.26E-05), Amyotrophic Lateral Sclerosis (FDR=0.0358), Parkinson's Disease (FDR=3.07E-06) and Autism Spectrum disorder(FDR=1) Enrichment study revealed that has-mir-9-2 targeted the Rab GTPase to cause disorders in at least three Rab-associated neurodegenerative diseases namely Alzheimer's Disease, Amyotrophic lateral sclerosis, and Parkinson's disease while has-mir-212 was found to target Rab in causing Autism Spectrum Disease. We also found out that 17 such miRNAs were found to be associated with Alzheimer's disease and 12 associated with Parkinson's disease, making them an important part of this study. To compare data obtained from the two sets, we plotted a bar graph shown in *Fig. F25* to validate the number of miRNAs we obtained through both the data sets. The following 12 miRNAs were found common in both the sets: hsa-mir-141, hsa-mir-144, hsa-mir-155, hsa-mir-181c, hsa-mir-186, hsa-mir-206, hsa-mir-212, hsa-mir-214, hsa-mir-221, hsa-mir-222, hsa-mir-29c, hsa-mir-30a.

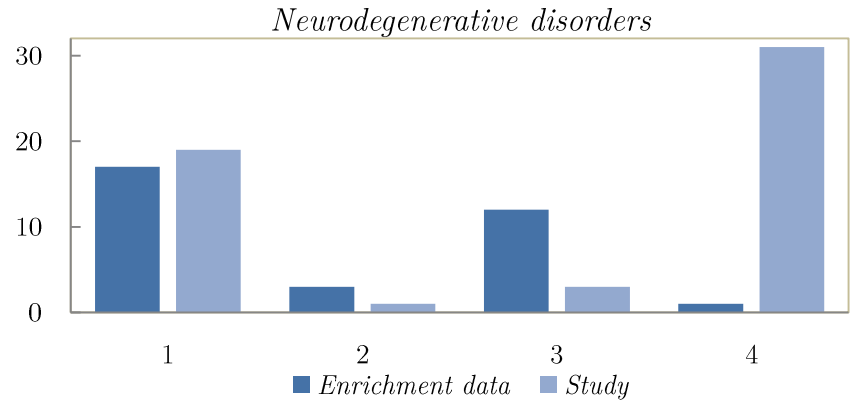


Fig. F25: *Comparison of datasets in neurodegenerative disorders*

Endocrine and metabolic disorders: In diabetes and metabolic disorders, many circulating miRNAs have been reported [114]. Analysis of Set-4 miRNAs revealed that they were found to target Rab that were associated with Fatty liver disease (FDR=0.2341), Diabetes Mellitus, Type 2 (FDR=0.6156), Obesity (FDR=0.2311) and Cystic Fibrosis(FDR=0.4241) hsa-mir-200b was found to target Rab involved in type 2 Diabetes Mellitus as well as Obesity, making its role substantial in our study. *Fig. F26* shows a comparison between the enrichment data set and the study data set. It was found that at least two miRNAs were associated with Obesity and one miRNA each in Fatty liver disease, Type 2 diabetes mellitus, and Cystic fibrosis. Few other Rab associated endocrine and metabolic diseases were investigated, namely Lipidosis, Oculocutaneous Albinism, Hermansky-Pudlak's syndrome, and Hypercholesterolemia but no significant miRNA targets were found for the same.

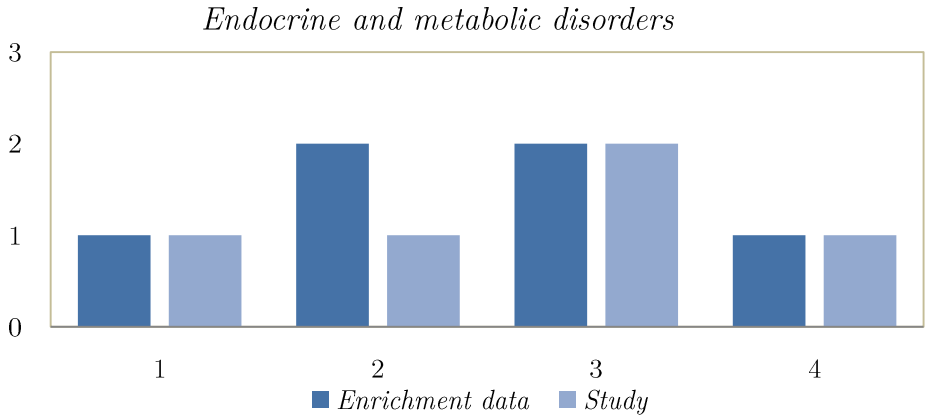


Fig. F26: *Comparison of datasets in endocrine and metabolic disorders*

Analyzing Set-5 miRNAs, we found that they were implicated in endocrine disorders like Diabetes mellitus (FDR= 0.1452), Fatty liver(FDR= 0.2571), and Ovarian neoplasm(FDR= 0.3833) associated with three, one and two miRNAs respectively. Hsa-mir-130a targeted Rab in all three diseases, making it an important regulator in Rab-associated Endocrine disorders. Hsa-mir-200b, hsa-mir-222, and hsa-mir-26b were also found to target Rab and cause disease. *Fig. F27* shows a comparison

between the Enrichment and study data sets showing that at least one miRNA is involved in Rab-associated Endocrine diseases.

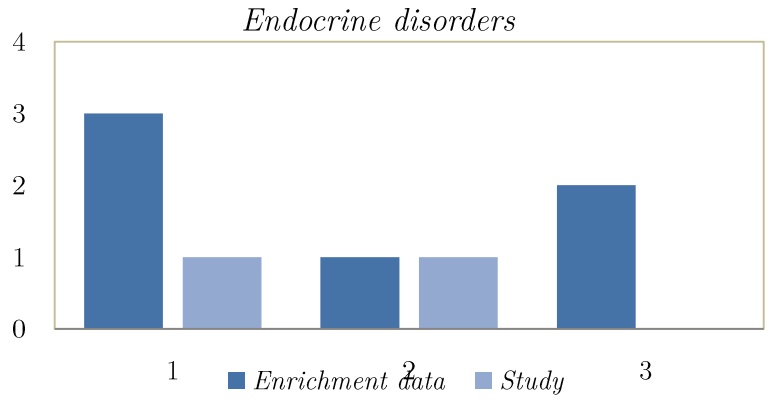


Fig. F27: *Comparison of datasets in endocrine disorders*

From the enrichment analysis of Set-6 miRNAs, we found out that hsa-mir-802 targeted Rab proteins in Rab-associated chromosomal disorder, Down Syndrome (FDR=1). *Fig. F28* shows a comparison between enrichment and study data sets which validates that Down Syndrome maybe be associated with at least one miRNA that targets Rab to cause it.

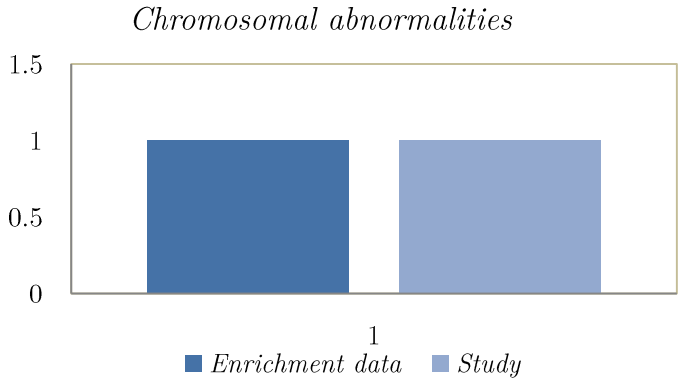


Fig. F28: *Comparison of datasets in chromosomal abnormalities*

Recent work has linked dysregulated miRNA expression with some SLE characteristics [115] and sepsis, an inflammatory response [116]. An enrichment analysis of Set-7 miRNAs was carried out and it was found out that Asthma (FDR=0.4073) and Rheumatoid Arthritis (FDR=0.4959) were Rab-associated diseases with hsa-mir-21 as targets in both. hsa-mir-216a was also reported to target Rab proteins implicated in Asthma. Also, few other Rab-associated Immunological disorders were reported, namely Systemic lupus erythematosus, Griscelli Syndrome type 2, and Ataxia telangiectasia, however, no significant miRNA that targets Rab proteins in these diseases were found. *Fig. F29* shows a comparison between the enrichment and study data sets to validate that at least one miRNA plays an important role in Rab associated immunological disorders.

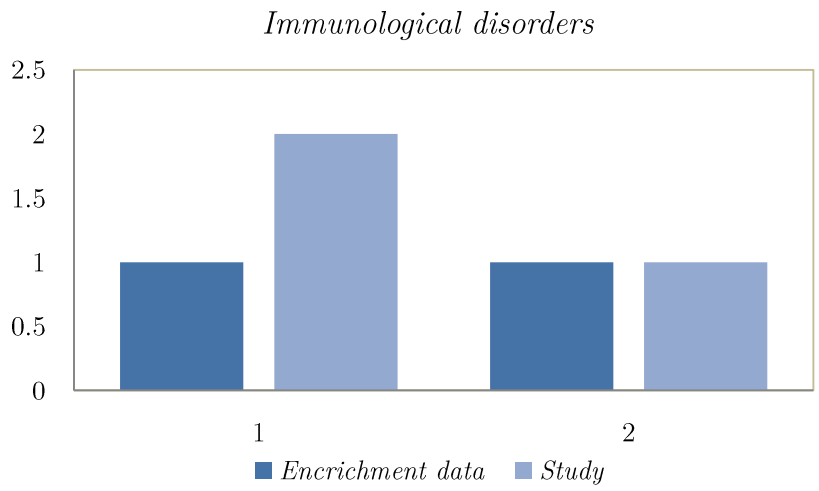


Fig. F29: *Comparison of data in Immunological disorders*

Altered miRNA profiles have been reported in Duchenne Muscular Dystrophy patients [117]. From the enrichment analysis of Set-8 miRNAs, we found out that Muscular Dystrophy (FDR= 0.7456) is associated with hsa-mir-29c. In addition, hsa-mir-29b-3p was also implicated to target Rab proteins to cause Muscular Dystrophy. *Fig. F30* represents a comparison between the two data sets validating that at least one miRNA plays a role in targeting Rab proteins to cause Rab-associated Musculoskeletal disorders.

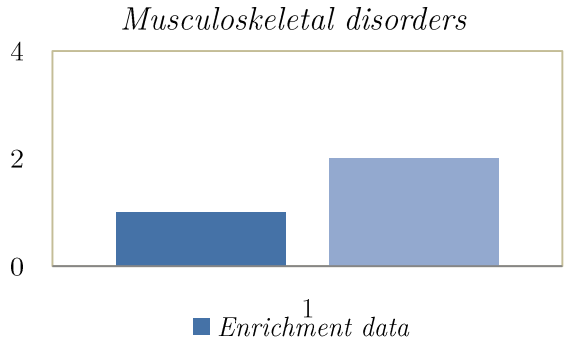


Fig. F30: *Comparison of data in Musculoskeletal disorders*

Summary *Table T7* below lists in detail all the miRNAs obtained in each case:

Disease group/Set#	#	Disease name	Enrichment count	miRNAs	Study	miRNAs
(Set-1)	1	Coronary Atherosclerosis	2	hsa-mir-155, hsa-mir-21		
	2	Coronary Heart Diseases	4	hsa-mir-340, hsa-mir-130a, hsa-mir-21, hsa-mir-155		
	3	Heart Diseases [unspecific]	1	hsa-mir-21		

	4	Heart Failure	6	hsa-mir-340, hsa-mir-130a, hsa-mir-155, hsa-mir-30e, hsa-mir-212, hsa-mir-21	6	hsa-mir-21, hsa-mir-212, hsa-mir-130a, hsa-mir-155, hsa-mir-30e, hsa-mir-340
	5	Myocardial Ischemic-Reperfusion Injury	2	hsa-mir-155, hsa-mir-21		
(Set-2)	1	Carcinoma, Breast	52	hsa-mir-18b, hsa-mir-181a-1, hsa-mir-135b, hsa-mir-18a, hsa-mir-130b, hsa-mir-200c, hsa-mir-210, hsa-mir-155, hsa-let-7g, hsa-mir-101-2, hsa-mir-19b-2, hsa-mir-181d, hsa-let-7c, hsa-mir-20a, hsa-mir-106a, hsa-mir-212, hsa-mir-200b, hsa-mir-107, hsa-mir-144, hsa-let-7b, hsa-mir-181c, hsa-mir-134, hsa-mir-190a, hsa-mir-206, hsa-mir-182, hsa-mir-204, hsa-mir-15a, hsa-mir-185, hsa-mir-106b, hsa-mir-145, hsa-mir-143, hsa-mir-149, hsa-let-7e, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2, hsa-mir-103a-2, hsa-mir-101-1, hsa-mir-130a, hsa-mir-137, hsa-mir-181b-1, hsa-mir-181a-2, hsa-mir-19b-1, hsa-mir-103a-1, hsa-mir-192, hsa-let-7i, hsa-mir-200a, hsa-mir-148a, hsa-mir-17, hsa-mir-148b, hsa-mir-152, hsa-mir-19a	77	hsa-let-7b, hsa-let-7c, hsa-let-7e, hsa-mir-15a, hsa-mir-17, hsa-mir-19a, hsa-mir-20a, hsa-mir-2, hsa-mir-23a, hsa-mir-26b, hsa-mir-30a, hsa-mir-31, hsa-mir-93, hsa-mir-96, hsa-mir-98, hsa-mir-106a, hsa-mir-148a, hsa-mir-30d, hsa-mir-181c, hsa-mir-182 hsa-mir-204, hsa-mir-210, hsa-mir-211, hsa-mir-212, hsa-mir-214, hsa-mir-215, hsa-mir-216a, hsa-mir-221, hsa-mir-222, hsa-let-7g, hsa-let-7i, hsa-mir-23b, hsa-mir-30b, hsa-mir-130a, hsa-mir-141, hsa-mir-144, hsa-mir-152, hsa-mir-134, hsa-mir-149, hsa-mir-185, hsa-mir-186, hsa-mir-200c, hsa-mir-155, hsa-mir-106b, hsa-mir-29c, hsa-mir-200a, hsa-mir-302a, hsa-mir-301a, hsa-mir-130b, hsa-mir-30e, hsa-mir-302b, hsa-mir-302c, hsa-mir-302d hsa-mir-372, hsa-mir-373, hsa-mir-374a, hsa-mir-380, hsa-mir-381, hsa-mir-383, hsa-mir-340, hsa-mir-135b, hsa-mir-148b, hsa-mir-410, hsa-mir-181d, hsa-mir-519d, hsa-mir-939, hsa-mir-17-5p hsa-mir-107, hsa-mir-206, hsa-mir-429, hsa-mir-301b, hsa-mir-216b, hsa-mir-498, hsa-mir-29b-3p, hsa-mir-181a-5p, hsa-mir-181b-5p, hsa-mir-190a-5p
	2	Carcinoma, Colon	50	hsa-mir-186, hsa-mir-107, hsa-mir-103a-2, hsa-mir-200c, hsa-mir-18b, hsa-	2	hsa-mir-18a, hsa-mir-374b

				mir-181a-1, hsa-mir-101-1, hsa-mir-130a, hsa-mir-137, hsa-mir-210, hsa-mir-181b-1, hsa-mir-144, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-19b-1, hsa-mir-135b, hsa-mir-134, hsa-mir-103a-1, hsa-mir-206, hsa-mir-18a, hsa-mir-182, hsa-mir-204, hsa-mir-15a,hsa-mir-192, hsa-let-7g, hsa-mir-185, hsa-mir-106b, hsa-mir-101-2, hsa-mir-19b-2, hsa-mir-145, hsa-mir-143, hsa-let-7i, hsa-mir-200a, hsa-let-7c, hsa-mir-20a, hsa-mir-106a, hsa-mir-148a,hsa-mir-17, hsa-mir-149, hsa-let-7e, hsa-mir-130b, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-211, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2, hsa-mir-19a		
	3	Carcinoma, Lung	36	hsa-mir-107, hsa-mir-200c, hsa-mir-181a-1, hsa-mir-101-1, hsa-mir-137, hsa-mir-144, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-19b-1, hsa-mir-134, hsa-mir-206, hsa-mir-182, hsa-mir-15a, hsa-mir-192, hsa-let-7g, hsa-mir-185, hsa-mir-101-2, hsa-mir-19b-2, hsa-mir-145, hsa-mir-143, hsa-mir-200a, hsa-let-7c, hsa-mir-106a, hsa-mir-148a,hsa-mir-17, hsa-mir-149, hsa-mir-212, hsa-let-7e, hsa-mir-130b, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-21, hsa-mir-141, hsa-mir-19a	38	hsa-mir-18a, hsa-mir-25, hsa-mir-30a, hsa-mir-32, hsa-mir-96, hsa-mir-192, hsa-mir-30d, hsa-mir-182, hsa-mir-204, hsa-mir-211, hsa-mir-200b, hsa-mir-30b, hsa-mir-130a, hsa-mir-141, hsa-mir-143, hsa-mir-144, hsa-mir-200c, hsa-mir-155, hsa-mir-200a, hsa-mir-302a hsa-mir-34b, hsa-mir-301a, hsa-mir-130b, hsa-mir-30e, hsa-mir-302b, hsa-mir-302c, hsa-mir-302d, hsa-mir-372, hsa-mir-373, hsa-mir-18b, hsa-mir-433, hsa-mir-92b, hsa-mir-454, hsa-mir-374b, hsa-mir-107, hsa-mir-206, hsa-mir-429, hsa-mir-570
	4	Carcinoma, Lung. Non-Small-Cell	40	hsa-mir-186, hsa-mir-107, hsa-mir-200c, hsa-mir-181a-1, hsa-mir-101-1, hsa-mir-130a, hsa-mir-137, hsa-mir-210, hsa-mir-	28	hsa-mir-25, hsa-mir-26b, hsa-mir-29a, hsa-mir-98, hsa-mir-148a, hsa-mir-204, hsa-mir-212, hsa-mir-214, hsa-

				181b-1, hsa-mir-144, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-19b-1, hsa-mir-135b, hsa-mir-134, hsa-mir-18a, hsa-mir-182, hsa-mir-204, hsa-mir-15a, hsa-mir-192, hsa-let-7g,hsa-mir-185, hsa-mir-101-2, hsa-mir-19b-2, hsa-mir-145, hsa-mir-143, hsa-mir-200a, hsa-let-7c, hsa-mir-148a, hsa-mir-17, hsa-mir-149, hsa-mir-212, hsa-let-7e, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2		mir-221, hsa-mir-222, hsa-mir-200b, hsa-mir-27b, hsa-mir-130a, hsa-mir-145, hsa-mir-152, hsa-mir-185, hsa-mir-200c, hsa-mir-155, hsa-mir-29c, hsa-mir-367, hsa-mir-340, hsa-mir-338, hsa-mir-495, hsa-mir-497, hsa-mir-92b, hsa-mir-454, hsa-mir-429, hsa-mir-216b
	5	Carcinoma, Pancreatic	37	hsa-mir-200c, hsa-mir-181a-1, hsa-mir-101-1, hsa-mir-210, hsa-mir-181b-1, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-135b, hsa-mir-206, hsa-mir-18a, hsa-mir-182, hsa-mir-204, hsa-mir-15a, hsa-mir-192, hsa-let-7g, hsa-mir-101-2, hsa-mir-145, hsa-mir-181d, hsa-mir-143, hsa-mir-200a, hsa-let-7c, hsa-mir-20a, hsa-mir-106a, hsa-mir-148a,hsa-mir-17,hsa-mir-212, hsa-let-7e, hsa-mir-130b, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-211, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2, hsa-mir-19a	33	hsa-mir-96, hsa-mir-192, hsa-mir-30d, hsa-mir-181c, hsa-mir-204, hsa-mir-211, hsa-mir-212, hsa-mir-216a, hsa-mir-221, hsa-mir-222, hsa-mir-200b, hsa-mir-30b, hsa-mir-130a, hsa-mir-141, hsa-mir-143, hsa-mir-144, hsa-mir-200c, hsa-mir-155, hsa-mir-200a, hsa-mir-302a, hsa-mir-301a hsa-mir-130b, hsa-mir-30e, hsa-mir-302b, hsa-mir-302c, hsa-mir-302d hsa-mir-372, hsa-mir-373, hsa-mir-340, hsa-mir-181d, hsa-mir-206, hsa-mir-301b, hsa-mir-181a-5p
	6	Carcinoma, Prostate	43	hsa-mir-103a-2, hsa-mir-200c, hsa-mir-181a-1, hsa-mir-101-1, hsa-mir-130a, hsa-mir-210, hsa-mir-181b-1, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-19b-1, hsa-mir-135b, hsa-mir-181c, hsa-mir-134, hsa-mir-190a, hsa-mir-103a-1, hsa-mir-18a, hsa-mir-182, hsa-mir-204, hsa-mir-15a, hsa-let-7g, hsa-mir-185, hsa-mir-106b, hsa-mir-101-2, hsa-mir-19b-2, hsa-mir-145, hsa-mir-181d, hsa-mir-143, hsa-mir-200a, hsa-let-7c, hsa-mir-20a, hsa-mir-	38	hsa-mir-19a, hsa-mir-21, hsa-mir-23a, hsa-mir-30a, hsa-mir-31, hsa-mir-30d, hsa-mir-182, hsa-mir-204, hsa-mir-210, hsa-mir-211, hsa-mir-212, hsa-mir-19a, hsa-mir-22, hsa-mir-23a, hsa-mir-30a, hsa-mir-32, hsa-mir-30d, hsa-mir-130a, hsa-mir-141, hsa-mir-144, hsa-mir-134, hsa-mir-186, hsa-mir-155, hsa-mir-200a, hsa-mir-302a, hsa-mir-301a, hsa-mir-130b, hsa-mir-30e, hsa-mir-302b, hsa-mir-302c,

				148a, hsa-mir-17, hsa-mir-149,hsa-mir-212, hsa-let-7e, hsa-mir-130b, hsa-mir-200b, hsa-mir-152, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2, hsa-mir-19a		hsa-mir-302d, hsa-mir-372, hsa-mir-383, hsa-mir-340, hsa-mir-206, hsa-mir-19b-3p, hsa-mir-101-3p, hsa-mir-190a-5p
	7	Carcinoma, Renal Cell	29	hsa-mir-200c, hsa-mir-181a-1, hsa-mir-210, hsa-mir-144, hsa-mir-155, hsa-mir-181a-2, hsa-let-7b, hsa-mir-134, hsa-mir-206, hsa-mir-182, hsa-mir-204, hsa-mir-15a, hsa-let-7g, hsa-mir-185, hsa-mir-106b, hsa-mir-145, hsa-mir-143, hsa-let-7i, hsa-mir-200a, hsa-let-7c, hsa-mir-106a, hsa-mir-148a, hsa-mir-17, hsa-let-7e, hsa-mir-148b, hsa-mir-200b, hsa-mir-21, hsa-mir-141, hsa-mir-19a	0	
	8	Carcinoma, Bladder	25	hsa-mir-186, hsa-mir-200c, hsa-mir-101-1, hsa-mir-130a, hsa-mir-137, hsa-mir-210, hsa-mir-144, hsa-mir-155, hsa-mir-182, hsa-mir-192, hsa-mir-106b, hsa-mir-101-2, hsa-mir-145, hsa-mir-143, hsa-mir-200a, hsa-let-7c, hsa-mir-20a, hsa-mir-106a, hsa-mir-148a, hsa-mir-130b, hsa-mir-200b, hsa-mir-152, hsa-mir-21, hsa-mir-141, hsa-mir-19a	6	hsa-mir-31, hsa-mir-182, hsa-mir-130b, hsa-mir-374b, hsa-mir-101-3p, hsa-mir-7-5p
	9	Gastrointestinal Neoplasms	27	hsa-mir-107, hsa-mir-103a-2, hsa-mir-106a, hsa-mir-200c, hsa-mir-148a, hsa-mir-18b, hsa-mir-20a, hsa-mir-106b, hsa-mir-181c, hsa-mir-17, hsa-mir-212, hsa-mir-130b, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-103a-1, hsa-mir-145, hsa-mir-18a, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2, hsa-mir-137, hsa-mir-210, hsa-mir-181b-1, hsa-mir-19a, hsa-mir-155, hsa-let-7g	0	
	10	Carcinoma, Cervical	30	hsa-mir-107, hsa-mir-181a-1, hsa-mir-101-1, hsa-mir-130a, hsa-mir-181b-1, hsa-mir-155, hsa-mir-181a-2, hsa-mir-135b, hsa-mir-	19	hsa-mir-25, hsa-mir-30a, hsa-mir-96, hsa-mir-30d, hsa-mir-182, hsa-mir-211, hsa-mir-200b, hsa-mir-30b,

				181c, hsa-mir-206, hsa-mir-18a, hsa-mir-182, hsa-mir-15a, hsa-mir-106b hsa-mir-101-2, hsa-mir-145, hsa-mir-181d, hsa-mir-143, hsa-mir-200a, hsa-mir-20a, hsa-mir-148a, hsa-mir-17, hsa-mir-212, hsa-mir-148b, hsa-mir-200b, hsa-mir-152, hsa-mir-211, hsa-mir-21, hsa-mir-141, hsa-mir-181b-2		hsa-mir-130a, hsa-mir-143, hsa-mir-200a, hsa-mir-301a, hsa-mir-130b, hsa-mir-30e hsa-mir-373, hsa-mir-494, hsa-mir-497, hsa-mir-107, hsa-mir-206
	11	Carcinoma, Thyroid	9	hsa-mir-200a, hsa-mir-148a, hsa-mir-106b, hsa-mir-149, hsa-mir-145, hsa-mir-21, hsa-mir-204, hsa-mir-144, hsa-mir-155	5	hsa-mir-96, hsa-mir-182, hsa-mir-204, hsa-mir-155, hsa-mir-34b
	12	Carcinoma, Gallbladder	7	hsa-mir-181a-1, hsa-mir-130a, hsa-mir-145, hsa-mir-21, hsa-mir-182, hsa-mir-181a-2, hsa-mir-143	0	
	13	Leukemia	13	hsa-mir-19b-1, hsa-mir-20a, hsa-mir-181a-1, hsa-mir-181c, hsa-mir-17, hsa-mir-19b-2, hsa-mir-18a, hsa-mir-21, hsa-mir-15a, hsa-mir-19a, hsa-mir-155, hsa-mir-181a-2, hsa-mir-143	4	hsa-mir-18a, hsa-mir-143, hsa-mir-155, hsa-mir-663a
	14	Carcinoma, Skin	5	hsa-mir-148a, hsa-mir-200a, hsa-mir-200b, hsa-mir-200c, hsa-mir-21	0	
	15	Carcinoma, Gastrointestinal	1	hsa-mir-149	0	
	16	Lymphoma, T-Cell, Cutaneous	1	hsa-mir-155	1	hsa-mir-204
	17	Carcinoma, Ovarian, Clear Cell	1	hsa-mir-21	21	hsa-mir-30a, hsa-mir-148a, hsa-mir-200b, hsa-mir-152, hsa-mir-200c, hsa-mir-200a, hsa-mir-302c, hsa-mir-494, hsa-mir-497, hsa-mir-627, hsa-mir-206, hsa-mir-429, hsa-mir-572, hsa-mir-622, hsa-mir-137
(Set-3)	1	Alzheimer's Disease	17	hsa-mir-186, hsa-mir-9-1, hsa-mir-124-2, hsa-mir-29b-1, hsa-mir-101-1, hsa-mir-9-2, hsa-mir-124-3, hsa-mir-155, hsa-mir-181c, hsa-mir-29b-2, hsa-mir-206, hsa-mir-222, hsa-mir-29c, hsa-mir-101-2, hsa-mir-212, hsa-mir-124-1, hsa-mir-9-3	19	

	2	Amyotrophic Lateral Sclerosis	3	hsa-mir-9-2, hsa-mir-206, hsa-mir-141	1	
	3	Parkinson's Disease	12	hsa-mir-19b-1, hsa-mir-29c, hsa-mir-9-1, hsa-mir-221, hsa-mir-19b-2, hsa-mir-29b-2, hsa-mir-30a, hsa-mir-29b-1, hsa-mir-214, hsa-mir-9-3, hsa-mir-9-2, hsa-mir-144	3	
	4	Autism Spectrum Disorder	1	hsa-mir-212	31	
(Set-4)	1	Fatty Liver [unspecific]	1	hsa-mir-130a	1	
	2	Diabetes Mellitus, Type 2	2	hsa-mir-222, hsa-mir-200b	1	
	3	Obesity	2	hsa-mir-26b, hsa-mir-200b	2	
	4	Cystic Fibrosis	1	hsa-mir-31	1	
(Set-5)	1	Diabetes Mellitus	3	hsa-mir-26b, hsa-mir-200b, hsa-mir-130a	1	
	2	Fatty Liver [unspecific]	1	hsa-mir-130a	1	
	3	Ovarian Neoplasms	2	hsa-mir-130a, hsa-mir-200b	0	
(Set-6)	1	Down Syndrome	1	hsa-mir-802	1	
(Set-7)	1	Asthma	1	hsa-mir-21	2	
	2	Rheumatoid Arthritis	1	hsa-mir-29c	1	
(Set-8)	1	Muscular Dystrophy	1	hsa-mir-29c	2	

Table T7: *miRNA enrichment set analysis: summary table*

11. Results and discussion

1. We have made a comprehensive analysis by tabulating 87 diseases that are also known to be associated with various members of the Rab family of GTPases. For this step, we have used DisGeNET, a database of genes and variants of human diseases, and tabulated the diseases along with the Rab reported to be associated with them, background Rabs and other affiliating genes involved. Rab occupancy for each disease sets was estimated which accounted for an average of 21.1409% Rab occupancy in Rab-associated diseases.

2. Given that cellular components seldom act as single entities, rather act as interacting scaffolds, and that any biological interaction can be represented mathematically in the form of a graph G= (V, E) where V is the set of vertices or nodes and E is the set of edges, it is indicative that the problem in question can be addressed by constructing miRNA-target gene co-expression networks for Rab-associated diseases.

3. In all networks, most Rabs occupied central positions, showing that they play a major role in contributing to disease phenotypes. Since Rabs are members of a large family of GTPases comprising 72 known members to date, and given that Rabs are products of gene duplication, it is not surprising that Rabs involved in a disease tend to cluster together. It is therefore of certain interest to study these Rab sub-networks. Diseases sharing common genes hypothesize the presence of common genetic origin, which have been found in many cases in our study. Such genetic overlaps however have not been explicitly mentioned in this work since we are interested in network-based studies only.

4. Most biological networks are scale-free, and all networks in our study agree with the same. A scale-free topology essentially follows the Power-law and a degree distribution plot on a logarithmic scale shows that only a small number of nodes have a high degree and a large number of nodes have a low degree. However, real systems rarely follow a pure power law, and so is the case with miRNA-target gene networks. This deviation from the power law is marked by low degree saturation and a high degree cut-off in all of our networks. Whereas this phenomenon can be indicative of the fact a sub-population of biological networks are indeed not purely scale-free, not ignoring the fact the presence of latent phenomenon that is not clearly understood.

5. Computational deletion of Rabs and their corresponding interactions established that most networks have remained invariant and stable even with the loss of links. However, invariability does not cover for a network's vulnerability to targeted attack: if important links are broken off, the network loses connections and forms a set of isolated sub-networks. In our study, although the networks did not lose their overall connectivity, some very important links involved in crucial pathways might have been broken, thus imparting disease symptoms.

6. In most of the plots, we find a general trend: for nodes with higher degrees, the average clustering coefficient is low and for nodes with low degrees, the average clustering coefficient is high. A trend analysis was performed and the curve of the form $y = ax^b$ gave the best fit on the clustering coefficient, closeness centrality, and Betweenness centrality plots of the networks.

12. Supplementary data

Table S1: Summary table of Gene-disease association based on GDAS

For each Rab associated disease, the table comprises the following information: The eighty-seven diseases in question, Rab involved with each, DSI, DPI, p(LI), GDAS, EI, number of PubMed references found. The table also includes additional information like UniProt IDs for protein sequence and annotation data, as well as literature-based search information.

Table S2: Summary table of miRNA and their target Rabs

For each Rab, we have listed the miRNAs that target each Rab. Data has been collected from different sources and validated by at least two of the databases as listed: MiRBase, miRanda, TargetScan, microT, picTar, and miRecords. This list also contains the total number of miRNAs that target each Rab.

Table S3: Summary table of Disease and its affiliating genes

For each Rab-associated disease, the list comprises of the top genes associated with each disease. Information has been collected from MalaCards, an integrated database of human maladies and their annotations, modeled on the database GeneCards.

Table S4, S5: miRNA-target gene co-expression network analysis

For each Rab-associated disease, network centralities have been recorded as scatter plots in Table S4 and S5, the four network centralities being Degree distribution, Clustering coefficient, Betweenness centrality, and Closeness centrality.

Table S6: Rab occupancy

For each Rab-associated disease, Rab occupancy has been calculated using the data provided in the sheet.

Table S7-S14: miRNA set enrichment analysis data from TAM 2.0

The miRNA set enrichment data has been recorded for each group of diseases in sheets S7-S14

N.B-Supplementary data has been provided in the CD attached at the back

13. Key points and highlights

1. **Identification** of Rab associated diseases in a systematic approach, dividing them into categories.
2. Investigation of the miRNAs that **alter/target/modify Rabs** and may be involved in these diseases.
3. A network biology approach to **build miRNA-gene co-expression network models** for these diseases, and the extraction of Rab sub-networks to propose miRNA-Rab co-expression networks.
4. In large networks, Rabs were found to '**group**' together.
5. **Deletion analysis** proved that since most biological networks are scale-free, the deletion of Rabs did not affect the overall degree distribution.
6. However, few links that were **broken** were enough to cause disease.
7. Few diseases have been observed to share **common genes**, which in turn suggest that the diseases might have a common genetic origin, for example, breast carcinoma and ovarian carcinoma.

14. Frontiers

As miRNA-based research advances, a better understanding of how miRNAs are involved in alteration of/ modifying Rabs in Rab associated diseases will help researchers to develop miRNA-based therapies for several Rab-associated diseases, given that miRNAs can serve as potential biomarkers in biomedical research. Furthermore, a database of miRNAs that can act as potential biomarkers by targeting Rabs can be maintained, since no such databases exist to date.

15. Bibliography

[1] Barabasi, Albert-Laszlo, and Zoltan N. Oltvai. "Network biology: understanding the cell's functional organization." *Nature reviews genetics* 5.2 (2004): 101-113.

[2] Zhang, Aidong. *Protein interaction networks: computational analysis*. Cambridge University Press, 2009.

[3] Barabási, Albert-László, Natali Gulbahce, and Joseph Loscalzo. "Network medicine: a network-based approach to human disease." *Nature reviews genetics* 12.1 (2011): 56-68.

[4] Vidal, Marc, Michael E. Cusick, and Albert-László Barabási. "Interactome networks and human disease." *Cell* 144.6 (2011): 986-998.

[5] Rao, V. Srinivasa, et al. "Protein-protein interaction detection: methods and analysis." *International journal of proteomics* 2014 (2014).

[6] Liu, Wei, et al. "Integrative analysis of human protein, function and disease networks." *Scientific reports* 5 (2015): 14344.

[7] Ha, Minju, and V. Narry Kim. "Regulation of microRNA biogenesis." Nature reviews Molecular cell biology 15.8 (2014): 509-524

[8] Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J.23, 4051–4060 (2004).

[9] G. Williams, M. L. Uchimoto, N. Coult, D. World, and E. Beasley, “Body fluid mixtures: resolution using forensic microRNA analysis,” Forensic Science International: Genetics Supplement Series, vol. 4, no. 1, pp. e292–e293, 2013.

[10] A. Etheridge, I. Lee, L. Hood, D. Galas, and K. Wang, “Extracellular microRNA: a new source of biomarkers,” Mutation Research, vol. 717, no. 1-2, pp. 85–90, 2011

[11] J. F. Barger and P. S. Nana-Sinkam, “MicroRNA as tools and therapeutics in lung cancer,” Respiratory Medicine, vol. 109, no. 7, pp. 803–812, 2015

[12] Heneghan, Helen M., Nicola Miller, and Michael J. Kerin. "MiRNAs as biomarkers and therapeutic targets in cancer." Current opinion in pharmacology 10.5 (2010): 543-550.

[13] Hessvik, Nina Pettersen, Kirsten Sandvig, and Alicia Llorente. "Exosomal miRNAs as biomarkers for prostate cancer." Frontiers in genetics 4 (2013): 36.

[14] Grasso, Margherita, et al. "Circulating miRNAs as biomarkers for neurodegenerative disorders." Molecules 19.5 (2014): 6891-6910

[15] Dimmeler, Stefanie, and Andreas M. Zeiher. "Circulating microRNAs: novel biomarkers for cardiovascular diseases?." (2010): 2705-2707

[16] Goretti, Emeline, Daniel R. Wagner, and Yvan Devaux. "miRNAs as biomarkers of myocardial infarction: a step forward towards personalized medicine?." Trends in molecular medicine 20.12 (2014): 716-725

[17] Yang, Yeyi, et al. "Urine miRNAs: potential biomarkers for monitoring progression of early stages of diabetic nephropathy." Medical hypotheses 81.2 (2013): 274-278.

[18] MicroRNA isolation and stability in stored RNA samples, M.MrazK. MalinovaJ. MayerS. Pospisilova

[19] Mall, Christine, et al. "Stability of miRNA in human urine supports its biomarker potential." Biomarkers in medicine 7.4 (2013): 623-631

[20] Bovell, Liselle, et al. "miRNAs are stable in colorectal cancer archival tissue blocks." Frontiers in bioscience (Elite edition) 4 (2012): 1937-1940.

[21] Heneghan, Helen M., Nicola Miller, and Michael J. Kerin. "MiRNAs as biomarkers and therapeutic targets in cancer." Current opinion in pharmacology 10.5 (2010): 543-550

[22] Stenmark, Harald, and Vesa M. Olkkonen. "The Rab gtpase family." Genome biology 2.5 (2001): reviews3007-1

[23] Goud, Bruno, and Mary McCaffrey. "Small GTP-binding proteins and their role in transport." Current opinion in cell biology 3.4 (1991): 626-633

[24] Yi, Rui, et al. "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." Genes & development 17.24 (2003): 3011-3016

[25] Bock, Jason B., et al. "A genomic perspective on membrane compartment organization." Nature 409.6822 (2001): 839-841.

[26] D'Adamo, Patrizia, et al. "RAB GTPases and RAB-interacting proteins and their role in the control of cognitive functions." Neuroscience & Biobehavioral Reviews 46 (2014): 302-314

[27] Echard, Arnaud, et al. "Alternative splicing of the human Rab6A gene generates two close but functionally different isoforms." Molecular biology of the cell 11.11 (2000): 3819-3833.

[28] Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19:92–105

[29] Huang, H. et al. (2015) miR-5100 promotes tumor growth in lung cancer by targeting Rab6. Cancer Lett.362, 15–24

[30] Yang, Y. et al.(2015) miR-15b-5p induces endoplasmic reticulum stress and apoptosis in human hepatocellular carcinoma, both in Vitro and in vivo, by suppressing Rab1A. Oncotarget6, 16227–16238

[31] 1. Piñero, Janet, et al. "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants." *Nucleic acids research* (2016): gkw943.

[32] Minoru Kanehisa, Michihiro Araki, Susumu Goto, Masahiro Hattori, Mika Hirakawa, Masumi Itoh, Toshiaki Katayama, Shuichi Kawashima, Shujiro Okuda, Toshiaki Tokimatsu, Yoshihiro Yamanishi, KEGG for linking genomes to life and the environment, *Nucleic Acids Research*, Volume 36, Issue suppl_1, 1 January 2008, Pages D480–D484, https://doi.org/10.1093/nar/gkm882

[33] Noa Rappaport, Noam Nativ, Gil Stelzer, Michal Twik, Yaron Guan-Golan, Tsippi Iny Stein, Iris Bahir, Frida Belinky, C. Paul Morrey, Marilyn Safran, Doron Lancet, MalaCards: an integrated compendium for diseases and their annotation, *Database*, Volume 2013, 2013, bat018, https://doi.org/10.1093/database/bat018

[34] Paraskevopoulou, Maria D., et al. "DIANA-microT web server v5. 0: service integration into miRNA functional analysis workflows." *Nucleic acids research* 41.W1 (2013): W169-W173.

[35] Reczko M, Maragkakis M, Alexiou P, Grosse I, Hatzigeorgiou AG., "Functional microRNA targets in protein coding sequences.", Bioinformatics. , 2012 Jan 27.

[36] A.J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander, D.S. Marks; (2003) MicroRNA targets in Drosophila; Genome Biology 5(1):R1.

[37] Ana Kozomara, Maria Birgaoanu, Sam Griffiths-Jones, miRBase: from microRNA sequences to function, *Nucleic Acids Research*, Volume 47, Issue D1, 08 January 2019, Pages D155–D162, https://doi.org/10.1093/nar/gky1141

[38] Sam Griffiths-Jones, Russell J. Grocock, Stijn van Dongen, Alex Bateman, Anton J. Enright, miRBase: microRNA sequences, targets and gene nomenclature, *Nucleic Acids Research*, Volume 34, Issue suppl_1, 1 January 2006, Pages D140–D144, https://doi.org/10.1093/nar/gkj112

[39] Kozomara, Ana, and Sam Griffiths-Jones. "miRBase: integrating microRNA annotation and deep-sequencing data." *Nucleic acids research* 39.suppl_1 (2010): D152-D157.

[40] Xiao, Feifei, et al. "miRecords: an integrated resource for microRNA–target interactions." *Nucleic acids research* 37.suppl_1 (2009): D105-D110.

[41] Krek, A., Grün, D., Poy, M. *et al.* Combinatorial microRNA target predictions. *Nat Genet* **37,** 495–500 (2005). https://doi.org/10.1038/ng1536

[42] Chen, K., Rajewsky, N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* **38,** 1452–1456 (2006). https://doi.org/10.1038/ng1910

[43] Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife, 4:e05005*, (2015). eLife Lens view.

[44] Pettersen, Eric F., et al. "UCSF Chimera—a visualization system for exploratory research and analysis." *Journal of computational chemistry* 25.13 (2004): 1605-1612.

[45] Shannon, Paul, et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks." *Genome research* 13.11 (2003): 2498-2504.

[46] Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: an open source software for exploring and manipulating networks." *Third international AAAI conference on weblogs and social media*. 2009.

[47] Assenov, Yassen, et al. "Computing topological parameters of biological networks." *Bioinformatics* 24.2 (2008): 282-284.

[48] Wiese, Roland, Markus Eiglsperger, and Michael Kaufmann. "yfiles—visualization and automatic layout of graphs." *Graph Drawing Software*. Springer, Berlin, Heidelberg, 2004. 173-191.

[49] David Warde-Farley, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz, Chris Grouios, Farzana Kazi, Christian Tannus Lopes, Anson Maitland, Sara Mostafavi, Jason Montojo, Quentin Shao, George Wright, Gary D. Bader, Quaid Morris, The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, *Nucleic Acids Research*, Volume 38, Issue suppl_2, 1 July 2010, Pages W214–W220, https://doi.org/10.1093/nar/gkq537

[50] Fan, Yannan, et al. "miRNet-dissecting miRNA-target interactions and functional associations through network-based visual analysis." *Nucleic acids research* 44.W1 (2016): W135-W141

[51] Li, Jianwei, et al. "TAM 2.0: tool for MicroRNA set analysis." *Nucleic acids research* 46.W1 (2018): W180-W185.

[52] Stenmark, Harald, and Vesa M. Olkkonen. "The Rab gtpase family." *Genome biology* 2.5 (2001): reviews3007-1.

[53] Bock, Jason B., et al. "A genomic perspective on membrane compartment organization." *Nature* 409.6822 (2001): 839-841.

[54] D'Adamo, Patrizia, et al. "RAB GTPases and RAB-interacting proteins and their role in the control of cognitive functions." *Neuroscience & Biobehavioral Reviews* 46 (2014): 302-314

[55] Touchot, Nicolas, Pierre Chardin, and Armand Tavitian. "Four additional members of the ras gene superfamily isolated by an oligonucleotide strategy: molecular cloning of YPT-related cDNAs from a rat brain library." *Proceedings of the National Academy of Sciences* 84.23 (1987): 8210-8214.

[56] Goud, Bruno, and Mary McCaffrey. "Small GTP-binding proteins and their role in transport." *Current opinion in cell biology* 3.4 (1991): 626-633.

[57] Chavrier, Philippe, et al. "Hypervariable C-termmal domain of Rab proteins acts as a targeting signal." *Nature* 353.6346 (1991): 769-772.

[58] Echard, Arnaud, et al. "Alternative splicing of the human Rab6A gene generates two close but functionally different isoforms." *Molecular biology of the cell* 11.11 (2000): 3819-3833.

[59] Gopal Krishnan, Priya D., et al. "Rab GTPases: Emerging Oncogenes and Tumor Suppressive Regulators for the Editing of Survival Pathways in Cancer." *Cancers* 12.2 (2020): 259.

[60] Segev, Nava. "Ypt/Rab gtpases: regulators of protein trafficking." *Sci. STKE* 2001.100 (2001): re11-re11.

[61] Hutagalung, Alex H., and Peter J. Novick. "Role of Rab GTPases in membrane traffic and cell physiology." *Physiological reviews* 91.1 (2011): 119-149.

[62] Moore, Ian, Jeff Schell, and Klaus Palme. "Subclass-specific sequence motifs identified in Rab GTPases." *Trends in biochemical sciences* 20.1 (1995): 10-12.
Gabe Lee, Meng-Tse, Ashwini Mishra, and David G. Lambright. "Structural mechanisms for regulation of membrane traffic by Rab GTPases." *Traffic* 10.10 (2009): 1377-1389.

[63] Anant, Janmeet S., et al. "Mechanism of Rab geranylgeranylation: formation of the catalytic ternary complex." *Biochemistry* 37.36 (1998): 12559-12568.

[64] Alexandrov, K., et al. "Rab escort protein-1 is a multifunctional protein that accompanies newly prenylated Rab proteins to their target membranes." *The EMBO journal* 13.22 (1994): 5262-5273.

[65] Pei, Lin, et al. "PRC17, a novel oncogene encoding a Rab GTPase-activating protein, is amplified in prostate cancer." *Cancer research* 62.19 (2002): 5420-5424.

[66] Stein, Mary-Pat, Jianbo Dong, and Angela Wandinger-Ness. "Rab proteins and endocytic trafficking: potential targets for therapeutic intervention." *Advanced drug delivery reviews* 55.11 (2003): 1421-1437.

[67] Griscelli, Claude, et al. "A syndrome associating partial albinism and immunodeficiency." *The American journal of medicine* 65.4 (1978): 691-702.

[68] Ménasché, Gaël, et al. "Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome." *Nature genetics* 25.2 (2000): 173-176.

[69] Wu, Xufeng, et al. "Rab27a enables myosin Va-dependent melanosome capture by recruiting the myosin to the organelle." *Journal of Cell Science* 114.6 (2001): 1091-1100.

[70] Verhoeven, Kristien, et al. "Mutations in the small GTP-ase late endosomal protein RAB7 cause Charcot-Marie-Tooth type 2B neuropathy." *The American Journal of Human Genetics* 72.3 (2003): 722-727.

[71] Alory, Christelle, and William E. Balch. "Organization of the Rab-GDI/CHM superfamily: the functional basis for choroideremia disease." *Traffic* 2.8 (2001): 532-543.

[72] Bienvenu, Thierry, et al. "Non-specific X-linked semidominant mental retardation by mutations in a Rab GDP-dissociation inhibitor." *Human molecular genetics* 7.8 (1998): 1311-1315.

[73] Xiao, Guang-Hui, et al. "The tuberous sclerosis 2 gene product, tuberin, functions as a Rab5 GTPase activating protein (GAP) in modulating endocytosis." *Journal of Biological Chemistry* 272.10 (1997): 6097-6100.

[74] Pereira-Leal, José B., Alistair N. Hume, and Miguel C. Seabra. "Prenylation of Rab GTPases: molecular mechanisms and involvement in genetic disease." *FEBS letters* 498.2-3 (2001): 197-200.

[75] van den Hurk, Jose AJM, et al. "Molecular basis of choroideremia (CHM): mutations involving the Rab escort protein-1 (REP-1) gene." *Human mutation* 9.2 (1997): 110-117.

[76] Croizet-Berger, Karine, et al. "The endocytic catalysts, Rab5a and Rab7, are tandem regulators of thyroid hormone production." *Proceedings of the National Academy of Sciences* 99.12 (2002): 8277-8282.

[77] Wu, Guangyu, et al. "Increased myocardial Rab GTPase expression: a consequence and cause of cardiomyopathy." *Circulation research* 89.12 (2001): 1130-1137.

[78] Pei, Lin, et al. "PRC17, a novel oncogene encoding a Rab GTPase-activating protein, is amplified in prostate cancer." *Cancer research* 62.19 (2002): 5420-5424.

[79] Ha, Minju, and V. Narry Kim. "Regulation of microRNA biogenesis." *Nature reviews Molecular cell biology* 15.8 (2014): 509-524.

[80] Lee, Yoontae, et al. "MicroRNA maturation: stepwise processing and subcellular localization." *The EMBO journal* 21.17 (2002): 4663-4670.

[81] Cai, Xuezhong, Curt H. Hagedorn, and Bryan R. Cullen. "Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs." *Rna* 10.12 (2004): 1957-1966.

[82] Pfeffer, Sébastien, et al. "Identification of microRNAs of the herpesvirus family." *Nature methods* 2.4 (2005): 269-276.

[83] Babiarz, Joshua E., et al. "Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs." *Genes & development* 22.20 (2008): 2773-2785.

[84] Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).

[85] Krol, Jacek, Inga Loedige, and Witold Filipowicz. "The widespread regulation of microRNA biogenesis, function and decay." *Nature Reviews Genetics* 11.9 (2010): 597-610.

[86] Davis-Dusenbery, B. N. & Hata, A. Mechanisms of control of microRNA biogenesis. *J. Biochem.* **148**, 381–392 (2010).

[87] Lee, Yoontae, et al. "The nuclear RNase III Drosha initiates microRNA processing." *Nature* 425.6956 (2003): 415-419.

[88] Denli, Ahmet M., et al. "Processing of primary microRNAs by the Microprocessor complex." *Nature* 432.7014 (2004): 231-235.

[89] Han, Jinju, et al. "Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex." *cell* 125.5 (2006): 887-901.

[90] Zeng, Yan, Rui Yi, and Bryan R. Cullen. "Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha." *The EMBO journal* 24.1 (2005): 138-148.

[91] Bohnsack, Markus T., Kevin Czaplinski, and DIRK GÖRLICH. "Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs." *Rna* 10.2 (2004): 185-191.

[92] Lund, Elsebet, et al. "Nuclear export of microRNA precursors." *science* 303.5654 (2004): 95-98.

[93] Yi, Rui, et al. "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." *Genes & development* 17.24 (2003): 3011-3016.

[94] Okada, Chimari, et al. "A high-resolution structure of the pre-microRNA nuclear export machinery." *Science* 326.5957 (2009): 1275-1279.

[95] Gwizdek, C. et al. Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J. Biol. Chem.* **278**, 5505–5508 (2003).

[96] Zeng, Y. & Cullen, B. R. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res.* **32**, 4776–4785 (2004).

[97] Bernstein, Emily, et al. "Role for a bidentate ribonuclease in the initiation step of RNA interference." *Nature* 409.6818 (2001): 363-366.

[98] Zhang, Haidi, et al. "Single processing center models for human Dicer and bacterial RNase III." *Cell* 118.1 (2004): 57-68.

[99] Park, Jong-Eun, et al. "Dicer recognizes the 5′ end of RNA for efficient and accurate processing." *Nature* 475.7355 (2011): 201-205.

[100] Kawamata, Tomoko, and Yukihide Tomari. "Making risc." *Trends in biochemical sciences* 35.7 (2010): 368-376.

[101] Förstemann, Klaus, et al. "Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1." *Cell* 130.2 (2007): 287-297.

[102] Miyoshi, Keita, et al. "Slicer function of Drosophila Argonautes and its involvement in RISC formation." *Genes & development* 19.23 (2005): 2837-2848.

[103] Liu, Ying, et al. "C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation." *Science* 325.5941 (2009): 750-753.

[104] Da Silveira, Willian A., et al. "miRmapper: a tool for interpretation of miRNA–mRNA interaction networks." *Genes* 9.9 (2018): 458.

[105] https://yed.yworks.com/support/manual/index.html

[106] https://cambridge-intelligence.com/introducing-the-powerful-organic-graph-layout/

[107] https://www.visual-computing.org/2016/01/18/group-structures graphs/diss_cv02/https://www.visual-computing.org/2016/01/18/group-structures-graphs/diss_cv02/

[108] Lindahl B. Acute coronary syndrome - the present and future role of biomarkers. Clin Chem Lab Med. 2013

[109] Diagnosis, Prognosis and Therapeutic Role of Circulating miRNAs in Cardiovascular Diseases Ali Sheikh Md Sayed, MD

[110] Visone, Rosa, and Carlo M. Croce. "MiRNAs and cancer." The American journal of pathology 174.4 (2009): 1131-1138.

[111] Micro-RNAs in Gastrointestinal and Liver Disease, Rosa Visone, Fabio Petrocca, Carlo M. Croce

[112] MV Iorio, M Ferracin, MicroRNA gene expression deregulation in human breast cancer, 2005

[113] Causes and Consequences of MicroRNA Dysregulation in Neurodegenerative Diseases

[114] Circulating miRNAs as biomarkers for endocrine disorders, H.Butz

[115] miRNA in systemic lupus erythematosus, Gil Amarilyo

[116] Use of miRNAs as Biomarkers in Sepsis, Raluca Dumache

[117] miRNAs as serum biomarkers for Duchenne muscular dystrophy, Davide Cacchiarelli, Ivano Legnini, 2011