Network Analysis to Identify Potential Molecular Links in Parkinson's Disease and Type II Diabetes

Abstract

Parkinson's Disease (PD) is the world's most common neurodegenerative movement disorder. While its symptoms such as bradykinesia have been well-characterized, its pathophysiology remains obscure. Consequently, this lack of understanding of its mechanisms has prevented the development of an effective cure. In recent years, increasing epidemiological and clinical studies have suggested that PD and Type II Diabetes (T2D) may share similarities in their mechanisms. This provides researchers the opportunity to study and elucidate PD's mechanisms through the lens of T2D. However due to the perplexing nature of cellular dynamics, it remains a challenge to fully apprehend the cellular perturbations caused by multifactorial diseases such as PD and T2D. To this endeavour, researchers have devised powerful computational biology tools, which can mirror the intricate complexity of the cell. One of the more popular methods is to construct a protein-protein interaction (PPI) network and apply algorithms on the network that simulate a particular disease condition. These algorithms subsequently rank the network proteins in a list according to the probability of the protein being associated with the disease. In this paper, we chose to build a PPI network using data from STRING database and utilize the random-walk-with-restart (RWR) algorithm to simulate PD and T2D separately. The resulting lists were then overlapped to identify any proteins that may be potential molecular links in both diseases. A mitochondrial protein, Miro, was identified to be a top contender. Wet-lab experiments were then carried out to validate if Miro is similarly affected in PD and T2D. However, due to technical limitations, we were unable to complete these experiments. Nonetheless, if proven right, it can aid in illuminating PD's mechanisms and formulating a curative treatment for PD. In a larger context, the network constructed can also serve as a formidable tool in elucidating the mechanisms of other diseases.

Contents

1. Introduction

1.1 Parkinson's Disease

Parkinson's Disease (PD) is the world's most common neurodegenerative movement disorder, affecting approximately 10 million people worldwide [1]. Clinically, the disease is first recognised through its four cardinal motor hallmarks of resting tremors, bradykinesia (slowed movement), rigidity, and postural instability [2]. These movement dysfunctions develop progressively, eventuating in severe physical disability. In later stages of the disease, non-motor symptoms such as mood disorders and dementia also begin to dominate the clinical picture, further impairing the patient's quality of life [3].

From a histological perspective, PD is characterised by the loss of dopamine (DA)-producing neurons in a region of the brain known as the substantia nigra. These DA neurons primarily associate with another brain region known as the striatum, which is a critical component of the motor and reward systems. It is the depletion of dopamine along this nigrostriatal pathway that results in the impairment of movement control that we observe in PD patients [4]. While multiple lines of evidence suggest that factors such as impaired mitochondrial homeostasis, perturbed alpha-synuclein proteostasis and oxidative stress may account for this loss of DA neurons, it is still unclear whether these are causal factors or the consequences of DA depletion[4] and the exact pathogenesis remains elusive.

The treatment landscape of PD has evolved greatly over the years and now includes newer treatment strategies such as Deep Brain Stimulation and restorative therapy [5]. However, pharmacological substitutions of dopamine through oral ingestions of the dopamine-precursor levodopa remains the mainstay of PD treatment due to its highly efficacious and non-invasive nature [6]. Nevertheless, all current treatments are merely symptomatic and the efficacy of levodopa-based therapy is only limited to the short term; contributing to treatment-resistant motor and non-motor dysfunctions in the long run [6]. Therefore, PD remains as a chronic and incurable disease that poses a substantial social and economic burden on the individual patient and society. Compounding this situation is the

fact that the biggest risk factor for the development of PD is age, with 95% of all clinical cases being diagnosed in patients who are above the age of 60 - a burgeoning problem to an ageing global population [7]. There is thus an urgent need to develop a curative PD treatment. However, the obscurity of PD's mechanisms hinders the development of such a cure. Hence, we will first need to elucidate the mechanisms that underlie the pathophysiology of PD.

1.2 Type II Diabetes and PD

While PD has long been linked to the development of many other health conditions such as cardiovascular and musculoskeletal diseases, there has been an increasing scientific interest in the biological and clinical implications of the comorbidity of PD and Type II Diabetes (T2D) [8]. T2D is a chronic metabolic disease that comprises pancreatic beta-cell dysfunction and insulin resistance in target cells. This interest in their comorbidity has arisen from a growing body of epidemiological evidence that shows an increase in the risk of T2D patients developing PD, and vice versa [9–11]. Clinical trials using therapeutic drugs designed for the treatment of T2D in PD patients such as Exenatide have also been shown to alleviate PD symptoms to a limited extent [12]. These suggest that the comorbidity of the two diseases may stem from dysregulations in shared biological pathways. Therefore investigating these shared pathways may offer a novel perspective on the pathophysiology of PD. An effective way to map the interwoven pathways in T2D and PD is via the network-based approach, which is discussed below.

1.3 Network-based Approach

In recent years, the expanding arsenal of molecular interaction data has also spurred the development of powerful computational biology tools for the interpretation and integration of these data. These tools provide biologists with the opportunity to use available datasets to explore the complex molecular relationships that occur within and across cells. One way of analysing molecular data is to visualise these molecular interactions in the form of a graphical network. Network-based approaches can mirror and simplify

complex cellular interactions to aid in predicting the outcome of certain molecular perturbations within a cell. A popular choice of network is a protein-protein interaction (PPI) network, where the proteins are represented as nodes and the PPI as edges [13, 14]. In the context of investigating human diseases, PPI-based networks may offer potential disease biomarkers, identify disease-causing mutations and outline the progression of dysregulated pathways. This, in turn, leads to increased veracity in disease diagnosis and more potent drug therapies [15, 16]. In fact, PPI network analysis have been successfully applied to predict biomarkers and disease-causing genes across a wide range of diseases including breast cancer, T2D, and Alzheimer's Disease [17–21]. In light of the effectiveness of such analysis, we postulate that PPI network analysis can be used as a tool to identify potential molecular links in PD and T2D.

1.3.1 Databases for PPI Network

The choice of the data source for constructing the PPI network is crucial in establishing the veracity of the network. A recent study conducted by Ideker in 2018 demonstrated that different PPI networks perform differently in predicting disease genes; networks constructed from large databases such as STRING, GENEMANIA and GIANT have better accuracies in recovering disease genes [22]. Ideker concluded that the STRING network has the best performance in disease gene prediction of all the networks available. Hence, we decided to use the STRING database to construct our network.

STRING is a PPI database constructed using data acquired from a spectrum of interactions databases such as GWAS, KEGG, BioGRID, APID, HINT, IntAct and sources that include high-throughput experiments, text-mining from literature and co-expression data.

The interaction between proteins are scored from seven different channels – neighbourhood, fusion, co-occurrence, co-expression, experimental, database and text-mining. Each channel provides a confidence score for every individual PPI. The confidence score reflects the probability that the interaction occurs. The overall score for the interaction is aggre-

gated and is defined as such

$$CS(e) = 1000 * (1 - \prod_{i=1}^{n}(1 - (c(e_i)/1000)) \tag{1.1}$$

where e refers to an interaction, CS(e) refers to the confidence score of the interaction, c(e_i) refers to the confidence score of the interaction from a specific channel i and n refers to the total number of channels. Another variation of PPI is to replace confidence scores with indicator scores; if there exists an interaction, the score will be 1, otherwise, it will be 0. This provides a more unbiased approach as it removes the bias of the confidence scores.

1.3.2 Algorithm for Disease Genes Prioritization

Conventionally, disease gene prioritization is achieved with algorithms extrapolated from graph theory. These algorithms can classified into three main categories: linkage methods, clustering methods and network propagation methods [23]. Linkage methods assume that the proteins directly interacting with disease-associated proteins are also likely to be associated with the disease [17]. Clustering methods, on the other hand, presume that proteins involved in the same disease have a high likelihood to cluster together in a network [24]. Such clusters are usually referred to as disease modules and constructed using topological features such as nodes' degrees and functional features such as edges' weight. Last but not least, network propagation methods takes into account all proteins in the PPI network and assign each protein a probability that it is associated in the disease [18]. The probabilities are formulated with a starting set of proteins that are known to be associated with the disease of interest.

Different methodologies have differential predictive capabilities. A comparative study conducted by Navlakha and Kingsford in 2010 indicated a network propagation based method known as the random-walk-with-restart (RWR) algorithm has the highest accuracy in predicting disease genes amongst the various methodologies [25]. Hence, we adopted RWR as our main algorithm for disease gene prioritization.

RWR is an extremely powerful tool for disease gene prioritization and the basis for the algorithm is derived from the main algorithm behind Google's webpage ranking, Personalized PageRank [26]. The underlying concept emulates a diffusion process: each node of interest is "filled" with a certain amount of fluid, and the fluid slowly diffuses out into the network, allocating each node a probability to be affected by the fluid. These probabilities are proportionate to the structure of the network and edge weights.

The RWR algorithm requires the network to be represented in an adjacency matrix, where the row i and column j represent protein i and protein j respectively. The element in row i and row j is the confidence score of the interaction between protein i and j. The matrix is defined as such:

$$W : w_{i,j} = w_{j,i} = \begin{cases} CS(e_{i,j}) & ;\ i \neq j \\ 0 & ;\ i = j \end{cases}$$

where W is the adjacency matrix, $w_{i,j}$ is the element in row i and column j and $CS(e_{i,j})$ is the confidence score of the interaction between protein i and j.

There are numerous variations for the random walk with restart (RWR) algorithms[26]. The general form of the algorithm is defined as the equation below:

$$p_t = (1 - \alpha)W' p_{t-1} + \alpha p_0 \tag{1.2}$$

where W' is the normalized version of the adjacency matrix, W, from (1.2), p_t is a vector representing the probability of the proteins or the respective gene being associated with the disease at time t and α is the restart parameter. At time t, the vector p_t describes the probability distribution for each gene to be associated with the query disease q. This vector is defined as such:

$$p_t = (p_1, p_2, ..., p_n)$$

where n is the number of proteins in the network and for all $i = 1, ...n$, p_i is the probability of protein at position i being associated with q. At time $t = 0$, p_t or p_0 depicts a

distribution that gives equal probability to proteins that are known to be associated with q:

$$p_i = \begin{cases} \frac{1}{n_q} & ; \textit{protein at position } i \textit{ is associated with } q \\ 0 & ; \textit{protein at position } i \textit{ is not associated with } q \end{cases}$$

where n_q is the number of proteins that are known to be associated with q.

The rationale behind the normalization of the matrix, W, is to either reinforce or diminish the significance of interaction between protein i and j. For the RWR algorithm, there are two variations of W'[26]. The first version is a column-normalized adjacency matrix or the random-walk normalized Laplacian matrix:

$$W' = WD^{-1} \tag{1.3}$$

where D is a diagonal matrix where each diagonal entry refers to the degree information, d_i of protein at row i and column i. Degree information is the number of edges or interactions incident to the protein. The second variation is known as the symmetric normalized Laplacian matrix. Unlike the former version, the matrix is normalized with respect to both the row and column:

$$W' = D^{-1/2}WD^{-1/2} \tag{1.4}$$

Both versions (1.3) and (1.4) of W' have slightly different properties; the latter giving higher significance to the degree information. In a PPI network, higher significance in degree information translates to increased significance of the specificity of the interaction between the two proteins.

The α in (1.2) is the restart parameter and it ranges from 0 to 1. The restart parameter reinforces the significance of the starting seeds of random walk, in this case, p_0. The higher the restart parameter, the more influence the starting seeds have over the random walk. In the context of disease gene prioritization, the restart parameter ensures that the known disease genes remain the main driving forces in the disease.

The equation (1.2) is recursively computed until the the random walk reaches a steady state, which is measured by the Euclidean norm of $p_t - p_{t-1}$. The Euclidean norm represents the magnitude of the vector, p_t. If the Euclidean norm is below a certain threshold, the random walk is determined to be at a steady state and there is no significant difference between p_t and p_{t-1}. Once the random walk is at a steady state, the values in p_t represents the probabilities of each gene being associated with the disease of interest.

1.4 Validation of *in silico* prediction

Wet laboratory-based analyses will be used to assess the veracity of the predictions generated through our network and RWR algorithm. In a similiar vein to the strategy used in our *in silico* analysis, two independent models of PD and T2D will be generated *in vitro* using an SH-SY5Y cell line. SH-SY5Y is a human-derived neuroblastoma cell line that is often used to develop models for PD research due to its human origins, catecholaminergic properties, and ease of maintenance [27].

To generate a PD model, the SH-SY5Y cells will be treated with rotenone. Rotenone is a naturally-derived pesticide that acts as a strong inhibitor of the mitochondrial respiratory chain complex I [28]. In animals, the inhibition of cellular respiratory processes by rotenone leads to the induction of many key behavioral and pathological features of PD, including the degeneration of DA neurons, movement impairments, and accumulation of alpha-synuclein [28]. All these have led to the establishment of rotenone-treatment as a reliable *in vitro* model for PD. For this reason, we intend to subject healthy SH-SY5Y cells to dosage-dependent rotenone treatments for the establishment of our PD model.

Since T2D is characterized by dysfunctions in insulin signalling, we will model T2D by disrupting insulin signalling in SH-SY5Y cells [29]. In insulin signalling, the insulin receptor substrate 1 (IRS-1) is a transmembrane tyrosine kinase that is essential for mediating the intracellular signalling cascades that occur after the binding of insulin to the insulin receptors on its target cell [30]. IRS-1 is therefore a critical element that determines the biological actions of insulin, and changes in its structure or expression

levels are often associated with the development of T2D [31]. Therefore, to establish an *in vitro* T2D model, we conducted an siRNA-mediated knockdown of IRS-1. The resulting attenuated levels of insulin-stimulated AKT phosphorylation and glucose uptake will mimick the state of insulin resistance that is observed in T2D.

After establishing stable *in vitro* models of PD and T2D, the expression levels of the top proteins generated from our *in silico* analysis will then be analyzed by means of western blotting procedures and mRNA analysis. We expect to observe parallels in the changes in expression levels of our proteins of interest across both disease conditions, which would be indicative of potential shared dysregulation pathways in their pathogeneses.

2. Material & Methods

2.1 Data mining and generation of PPI network

PPI data is downloaded from STRING version 10 database (https:/ /string-db.org/). Data used is extracted from *Homo sapiens* species. The PPI data is processed using Python.

For this paper, we decide to omit neighbourhood, fusion, co-occurrence, co-expression and text-mining channels to normalize the scores. This is to improve the credibility of the confidence score by basing the score on experimental validation and other databases: the experiment and databases channels respectively. The score is re-calculated using (1.1). The score is cut off at the threshold of 0.4, identical to the threshold used in STRING database; if the score is lower than 0.4, it will be regarded as 0. The data is then used to construct a PPI network with the proteins as nodes, interactions as edges and the newly formulated confidence scores as the edge weights.

2.2 Network Analysis with RWR

Disease genes for PD and T2D are predicted using RWR. The known candidate disease genes for PD and T2D are extracted from literature review [32–34] and shown in Table 2.1.

Table 2.1: List of candidate genes that are known to be associated with PD and T2D

Genes associated with PD	Genes associated with T2D
SNCA	IRS1
PARKIN	THADA
PINK1	GCKR
DJ-1	PPARG
LRRK2	ADCY5
ATP13A2	ADAMTS9
HTRA2	SLC30A8
PLA2G6	CDKN2A
FBXO7	TLE4
VPS35	KCNQ1
	CENTD2
	OASL

The algorithm is performed on network with varying parameters of the equation (1); the version of W', the restart parameter and the type of scores. The threshold for the Euclidean norm is set at 10^{-6}.

For each disease, RWR generates a list of genes with their corresponding probabilities to be associated with the disease of interest. The list is sorted in a descending order and a gene's rank is denoted by its position on the list; the gene with the highest probability will have rank 1. To identify potential genes that may play a part in both PD and T2D, we combined the ranks of the genes in T2D and PD. The smaller the combined rank, the more probable the gene is a potential candidate that is involved in both diseases. All of the above is coded in Python.

2.3 Antibodies and Reagents

Antibodies used were as follows: mouse anti-RHOT1 (HPA010687, Sigma), mouse anti-RHOT2 (SAB1408162, Sigma), rabbit anti-IRS1 (SAB4300483, Sigma), rabbit anti-caspase 3 (9662, Cell Signalling Technology), rabbit anti-cleaved caspase-3 (9669, Cell Signalling Technology), Anti-Mouse IgG (A9044, Sigma), and Anti-Rabbit IgG (A0545, Sigma); Chemicals/reagents used were as follows: ON-TARGETplus SMARTpool siRNAa (Dharmacon), DharmaFECT transfection reagent (Dharmacon), ReadyMix Taq PCR Reaction Mix (Sigma), Gel-Ready 100bp DNA Ladder (Sigma), All Blue Precision Plus Protein Standards (Bio-Rad), Human IRS-1 siRNA SMARTpool (Sigma), RNeasy Mini Kit (Qiagen), 10mM Rotenone (Sigma), 100% Dimethyl sulfoxide (DMSO) (Sigma).

2.4 Culturing SH-SY5Y neuroblastoma cell line

SH-SY5Y Neuroblastoma cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% Fetal Bovine Serum and 0.05 U/ml Pen-strep. Cells were incubated at 37℃, 5% CO_2 and passaged every 3-4 days.

2.5 Establishing Disease Models

A set of twelve healthy SH-SY5Y cell cultures were selected for the generation of disease models. Of these, half were used for the establishment of PD models and another half for the establishment of T2D models. The following cell treatment conditions were used to generate the PD model: Untreated, 0.1% DMSO-treated, 100 nM Rotenone-treated, 200 nM Rotenone-treated, 500 nM Rotenone-treated, and 1000 nM Rotenone-treated. Cells were incubated at 37℃ for 24 hours before lysis via sonication. To generate a T2D model, chemically-synthesized human IRS-1 siRNA SMARTpool and non-targeting siRNA controls were purchased from Dharmacon. Cells were treated under the following conditions: DharmaFECT transfection reagent-treated, 25 nM non-targeting siRNA-treated, 50 nM non-targeting siRNA-treated, 25 nM sequence-specific siRNA-treated, and 50 nM sequence-specific siRNA-treated. Cells were incubated at 37℃ for 48 hours before lysis via sonication. Prior to sonication, DMEM was replaced at the 24-hour time mark.

2.6 Western Blot Analysis

Cells were washed in ice-cold Phosphate Buffered Saline (PBS) and lysed in 1X lysis buffer [1.35 µl 1X PBS, 0.15 ml 10% Sodiym Dodecyl Sulfate (SDS), 1.5 µl AP, 15 µl phosphatase inhibitor 2 (Sigma), 15 µl phosphatase inhibitor 3 (Sigma), 7.5 µl Phenylmethylsulfonyl Fluoride (PMSF)]. Cells were scraped into clean tubes and lysed using a sonicator (30% pulse, 10 seconds each, 4 repeats). After centrifugation at 3,000 rpm, the supernatant was collected (cytoplasmic fraction) and remaining crude nuclei pellet further homogenised in 1X lysis buffer and broken by sonication.

For western blot analysis, an equivalent volume of 2X SDS loading dye was added was added to each tube of tissue lysate. The resulting mixture was boiled for five minutes for protein denaturation. Lysates of rotenone-treated samples were analysed through western blotting with mouse anti-RHOT1 (1:500 dilution ratio), mouse anti-RHOT2 (1:1000 dilution ratio), rabbit anti-caspase 3 (1:1000 dilution ratio), and rabbit anti-cleaved caspase-3 (1:1000 dilution ratio) primary antibodies. IRS-1 siRNA-treated cell lysates were probed

with rabbit anti-IRS1 antibodies (1:500 dilution ratio). Anti-Mouse IgG (1:3000 dilution ratio) was used as the secondary antibody for RHOT1 and RHOT2 blots, and Anti-Rabbit IgG IgG (1:3000 dilution ratio) as the secondary antibody for IRS-1, caspase-3 and cleaved caspase-3 blots. Actin was used as the loading control.

2.7 RNA Isolation

Total RNA was extracted from treated SH-SY5Y cells using the RNeasy Mini Kit, as per the manufacturer's protocol (Qiagen). Quantification of RNA was done using the NanoDrop spectrophotometer, and percentage purity was checked by measuring the A260/A280 absorbance ratios. Isolated RNA samples were stored at -30℃.

2.8 Reverse Transcription - Polymerase Chain Reaction (RT-PCR)

To prepare an RNA/Primer Mixture, 1 ng-5 g of isolated RNA samples was added to tubes containing 1 µl 10 mM dNTP mix, 1 µl 0.5 g/ µl oligo(dT) primer, and topped up to 10 µl with DEPC-treated water. Tubes were incubated at 65℃ for five minutes before placing on ice for one minute. To prepare a 2X RT Reaction Mix, components were added to a 1.5 ml tube in the following order: 10 µl 10X RT buffer, 20 µl 25 mM MgCl2, 10 µl 0.1 M DTT, and 5 µl 40 U/ µl RNaseOUT. Following which, 9 µl of 2X RT Reaction Mix was added to each RNA/Primer Mixture and contents were collected by brief centrifugation before incubation at room temperature for five minutes. 1 µl SuperScript II RT was then added to each tube.

RHOT1, RHOT2, IRS-1 and Actin DNA were cloned through PCR amplification of the cDNA molecules obtained from Reverse Transcription. The following primers were used: IRS1 forward primer (5-CAGAATGAAGACCTAAATGACC-3), IRS1 reverse primer (5-ATGCATCGTACC ATCCTACTG-3), RHOT1 forward primer (5-GTAGATTACTCAGA AGCAGAAC-3), RHOT1 reverse primer (5-GAATCCATCGACTTGTTACC-3), RHOT2 forward primer (5-AGAAGTACTTGATCCTCTGTG-3), RHOT2 reverse primer (5-CAT

GTAATGGTGCTTGTAGAC-3), Actin forward primer (5'-AGGCCAACCGCGAGAAG ATG-3'), Actin reverse primer (5'-TACCCCTCGTAGATGGGCAC-3'). Cycling conditions comprised: thirty-second strand denaturation process at 95℃, followed by 25 cycles of 95℃ for thirty seconds, 48℃ for thirty seconds, and 72℃ for one minute. After the 25th cycle, the thermocycler was cooled to 72℃ for one minute, before being held at 16℃. From each PCR reaction tube, 10 µl of PCR amplicon was obtained and loaded in a lane of a 1.8% agarose gel with SYBRsafe. 5 µl of 100bp DNA ladder was loaded for comparison. The gel ran in 1x TAE Buffer at 100V for forty minutes. Results were visualized using ChemiDoc.

2.9 Statistical Analysis

Statistical significance for all quantitative data obtained was analysed using a two sample t-test.

3. Results

3.1 Network Construction

Data mining of STRING database extracted 19,576 proteins and 1,585,802 interactions initially. After filtering out data from channels other than the experiments and databases channels, 792,678 interactions remain. The data is used to construct an adjacency matrix. Figure 3.1 shows the visualization of the matrix.

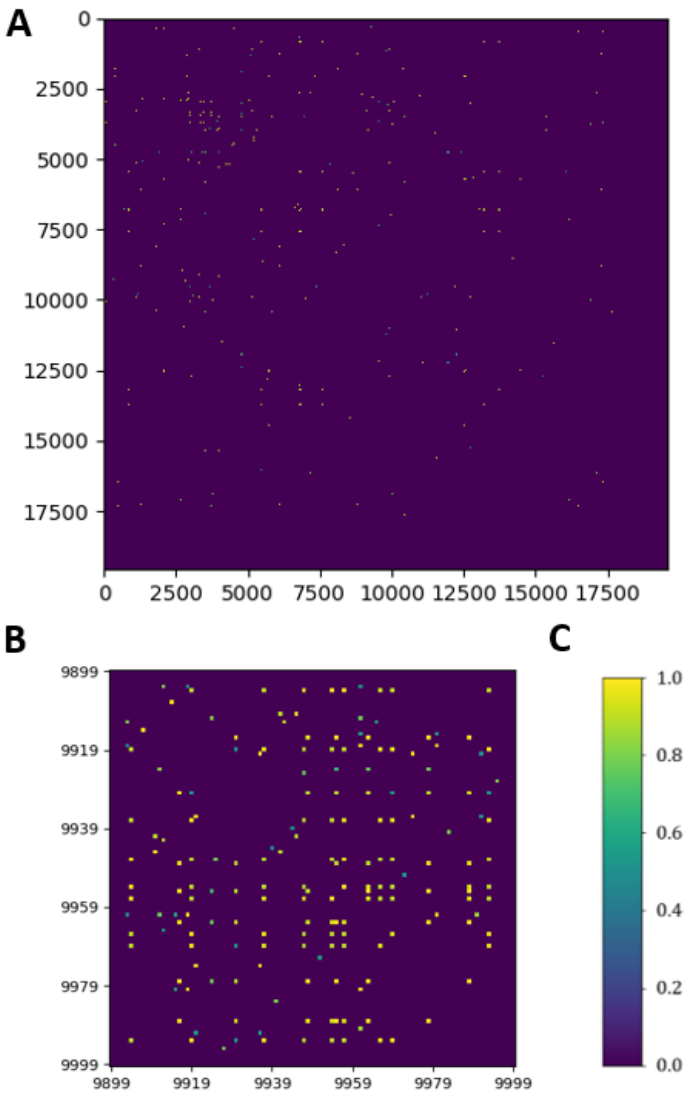

Figure 3.1: Visualization of PPI network adjacency matrix

A is the visualization of the full PPI network adjacency matrix constructed from the STRING database. The matrix is 19576 by 19576, where each row i and column i represent a protein i and the element in row i and column j holds the confidence score of the interaction between protein i and protein j. **B** is a magnification of the **A** from row 9899 to 9999 and column 9899 to 9999. **C** shows the color representation of probability depicted in the adjacency matrix.

Preliminary analysis of the network reveals that there is at least one direct interaction for every protein associated with either PD or T2D. Figure 3.2 portrays visual representation of subsets of PPI network that are directly associated with PD and T2D candidate genes.

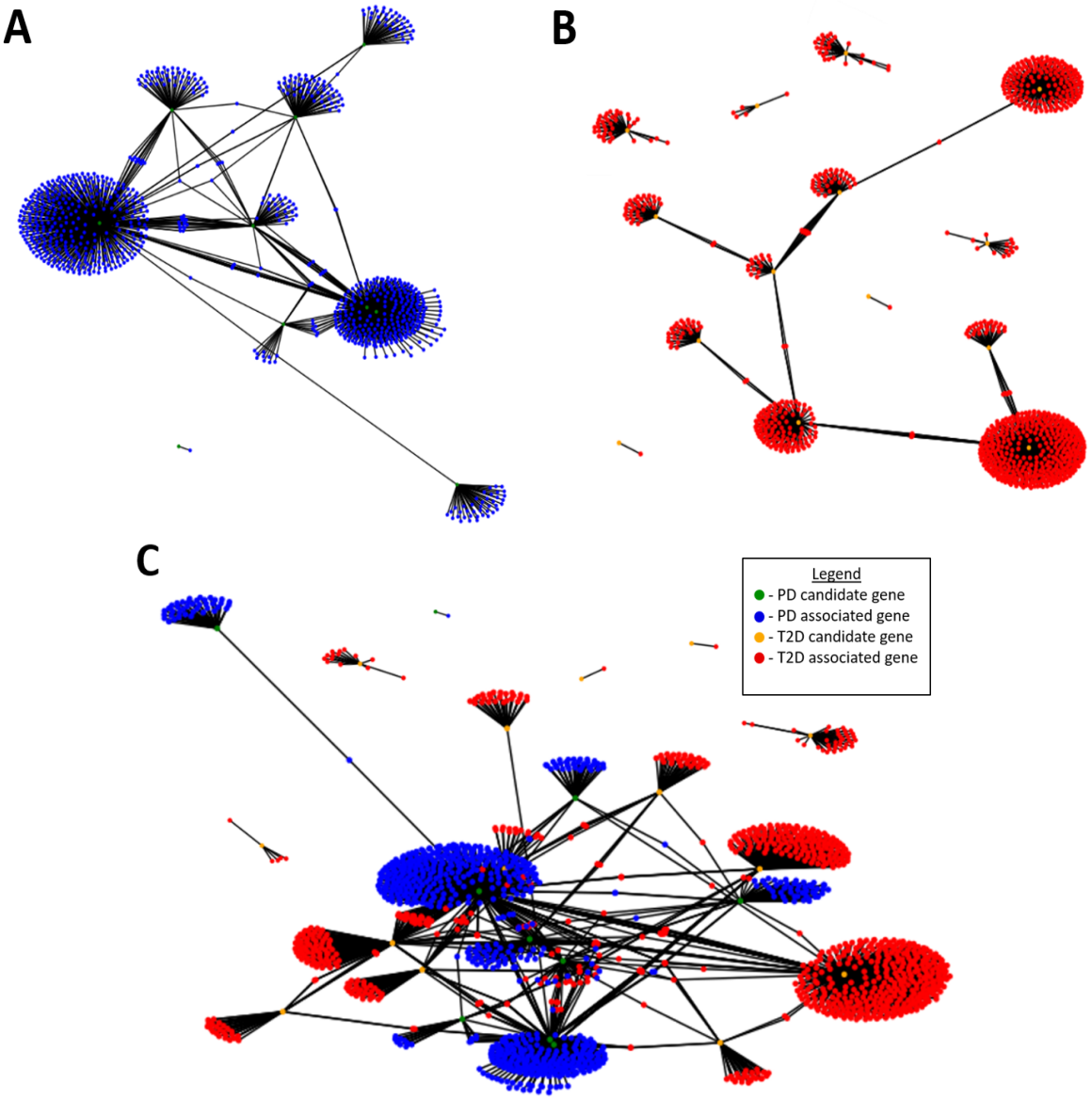

Figure 3.2: Visualization of PD and T2D subsets of the PPI network

A is the visualization of the subset of the PPI network that is directly associated with PD candidate genes. Green nodes and blue nodes represent the seeds and their direct interactors respectively. **B** is the visualization of the subset of the PPI network that is directly associated with T2D candidate genes. Orange nodes and red nodes represent the seeds and their direct interactors respectively. **C** is the visualization of the union between A and **B**.

3.2 Network Analysis with RWR

Table 3.1: Results generated with random-walk Laplacian matrix and confidence scores; top 5 genes

$W' = WD^{-1}$					
$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 0.9$	
Gene	Rank	Gene	Rank	Gene	Rank
MK01 HUMAN	122	MK01 HUMAN	133	MIRO1 HUMAN	121
RL40 HUMAN	132	MIRO1 HUMAN	150	MIRO2 HUMAN	121
RS27A HUMAN	136	MIRO2 HUMAN	154	MK01 HUMAN	132
MIRO1 HUMAN	153	RL40 HUMAN	166	RL40 HUMAN	181
MIRO2 HUMAN	156	RS27A HUMAN	168	RS27A HUMAN	191

Table 3.2: Results generated with symmetric Laplacian matrix and confidence scores; top 5 genes

$W' = D^{-1/2}WD^{-1/2}$					
$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 0.9$	
Gene	Rank	Gene	Rank	Gene	Rank
MIRO1 HUMAN	293	MIRO1 HUMAN	294	MIRO1 HUMAN	293
MIRO2 HUMAN	306	MIRO2 HUMAN	304	XIAP HUMAN	301
XIAP HUMAN	325	XIAP HUMAN	309	MIRO2 HUMAN	305
KPCD HUMAN	372	KPCD HUMAN	339	KPCD HUMAN	320
KPCE HUMAN	415	KPCE HUMAN	376	KPCE HUMAN	365

Table 3.3: Results generated with random-walk Laplacian matrix and indicator scores; top 5 genes

$W' = WD^{-1}$					
$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 0.9$	
Gene	Rank	Gene	Rank	Gene	Rank
RS27A HUMAN	85	CALM HUMAN	113	CALM HUMAN	114
RL40 HUMAN	85	MKO1 HUMAN	145	MIRO1 HUMAN	116
CALM HUMAN	138	RL40 HUMAN	156	MIRO2 HUMAN	116
MK01 HUMAN	143	RS27A HUMAN	157	RS27A HUMAN	142
UBC HUMAN	152	MIRO1 HUMAN	161	RL40 HUMAN	143

Table 3.4: Results generated with symmetric Laplacian matrix and indicator scores; top 5 genes

$W' = D^{-1/2}WD^{-1/2}$					
$\alpha = 0.5$		$\alpha = 0.75$		$\alpha = 0.9$	
Gene	Rank	Gene	Rank	Gene	Rank
MIRO1 HUMAN	276	MIRO1 HUMAN	267	MIRO1 HUMAN	268
MIRO2 HUMAN	291	MIRO2 HUMAN	283	MIRO2 HUMAN	280
CALM HUMAN	311	CALM HUMAN	317	CALM HUMAN	307
KPCD HUMAN	398	KPCD HUMAN	367	KPCD HUMAN	351
XIAP HUMAN	407	XIAP HUMAN	388	KPCE HUMAN	393

Results from the RWR analysis using confidence scores (Tables 3.1 and 3.2) show that there are 2 consistent genes that are within the top 5 predicted genes associated with both PD and T2D: MIRO1 and MIRO2. Confidence scores are replaced with indicator scores and the new lists of predicted genes are generated using RWR analysis (Tables 3.3 and 3.4). With the exceptions of RWR using random-walk Laplacian matrix at restart parameter 0.5 and 0.75, both MIRO1 and MIRO2 are observed in the top 5 predicted genes to be associated with both PD and T2D. Further analysis reveals that MIRO1 and MIRO2 are the 6^{th} and 7^{th} predicted genes in the RWR analysis with random-walk Laplacian matrix at $\alpha = 0.5$. Similarly, MIRO2 is the 6^{th} predicted gene in the RWR analysis with random-walk Laplacian matrix at $\alpha = 0.75$. These cumulative results suggest that MIRO1 and MIRO2 are top contenders to be potential key players in PD and T2D based on the STRING database.

3.3 Miro expression in PD and T2D Models

3.3.1 PD Model

Caspase 3 and β-actin antibodies were used to probe lysates of SH-SY5Y cells grown for 24 hours under the following conditions: untreated, DMSO-treated (which acts as a control), as well as 100 nM, 200 nM, 500 nM and 100nM of rotenone (Figure 3.3A). ImageJ software was used to analyze the intensities of the bands obtained (Figure 3.3B).

The same aforementioned samples were probed for Miro1 and β-actin which yielded an immunoblot with high background noise (Figure 3.3C). There are two areas with significantly higher intensities in the lanes with 100nM and 200nM rotenone-treated samples. However, they do not correspond to the expected molecular size of 71 kDa for Miro1. Instead, both correspond to a size of less than 50 kDa. Hence, the immunoblot did not contain any band indicating Miro1 expression in these samples. As Miro1 expression could not be obtained from immunoblotting, Miro1 mRNA analysis was considered instead. RT-PCR and gel electrophoresis was conducted using RNA isolated from these samples (Figure 3.3D). The normalized intensities of the bands were then obtained via

ImageJ analysis (Figure 3.3E).

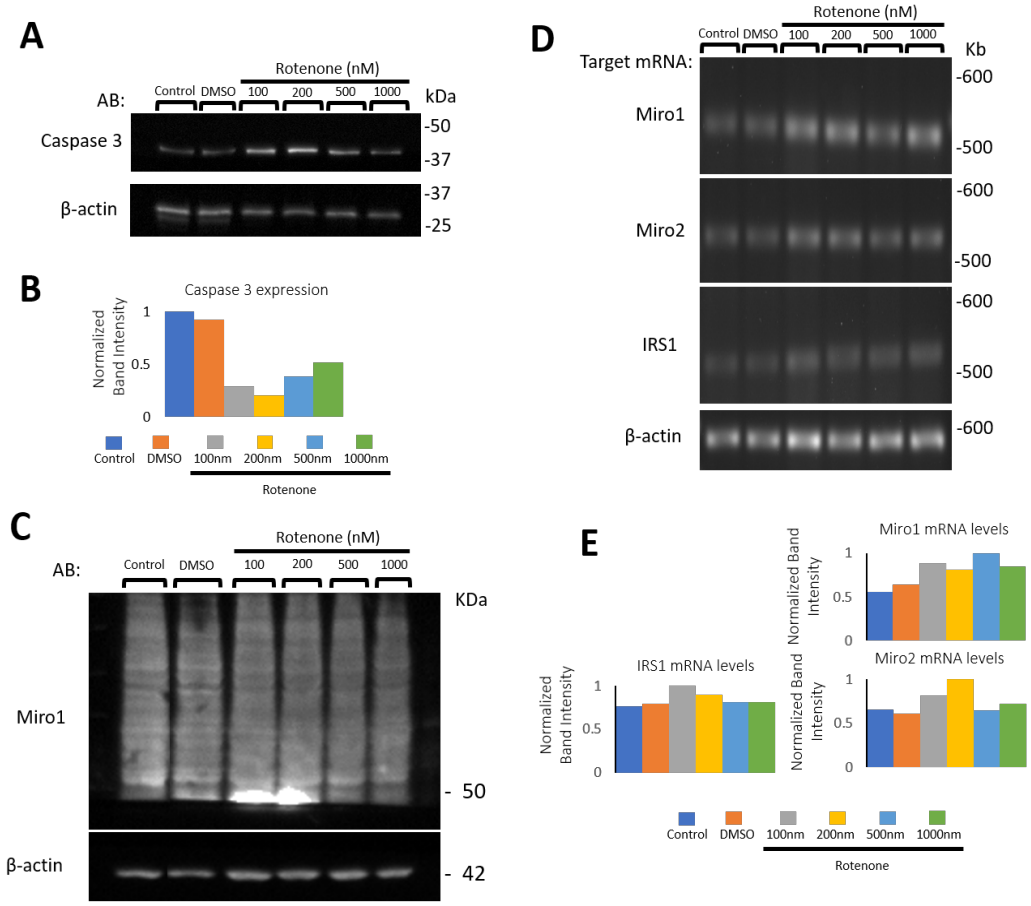

Figure 3.3: Wet lab validation of Miro expression in PD Model

A is the immunoblot of the lysates of rotenone-treated SH-SY5Y cells that were then probed with Caspase 3 and β - actin antibodies. **B** is a bar graph of normalized intensities of bands in **A**. **C** is the immunoblot of the aforementioned samples that were probed with Miro1 and β - actin antibodies. **D** is the gel electrophoresis result which reflects the Miro1, Miro2 and IRS1 mRNA levels in the same samples. **E** is the bar graph of the normalized intensities of bands in **D**.

3.3.2 T2D Model

To find out if the IRS1 knockdown via siRNA transfection was successful, IRS1, Miro1 and Miro2 mRNA levels were tested for in the lysates of SH-SY5Y cells treated with the following conditions: transfection reagents (which acts as a control), 25 nM siRNA, 50 nM siRNA, 25 nM non-targeting siRNA and 50 nM non-targeting siRNA. RT-PCR and subsequently agarose gel electrophoresis was conducted with RNA isolated from the aforementioned samples (Figure 3.4A). As a replicate had been carried out, an error bar was included for the band intensity analysis of all the samples except for the 25 nM siRNA-treated sample. A two sample t-test was conducted in order to determine if the

band intensities significantly varied from one another. There was no statistical difference observed between the band intensities of control and the following samples: 50 nM siRNA treated sample (p=0.272), 25 nM non-targeting siRNA treated sample (p=0.332) and 50 nM non-targeting siRNA treated sample (p=0.487). On the other hand, it was found that there was a significant difference in band intensities for IRS1 between the control and 25 nM siRNA treated sample (p=0.0146).

Conversely, there was no significant difference in band intensities for Miro1 between the control and the following samples: 25 nM siRNA treated sample (p=0.264), 50 nM siRNA treated sample (p=0.324), 25 nM non-targeting siRNA treated sample (p=0.487), 50 nM non-targeting siRNA treated sample (p=0.0596) (Figure 3.4B). The β-actin mRNA level for the samples was also tested for as a loading control. The absence of significant differences can also be observed in band intensities for Miro2 between the control and the following samples: 25 nM siRNA treated sample (p=0.303), 50 nM siRNA treated sample (p=0.316), 25 nM non-targeting siRNA treated sample (p=0.0711), 50 nM non-targeting siRNA treated sample (p=0.237) (Figure 3.4B).

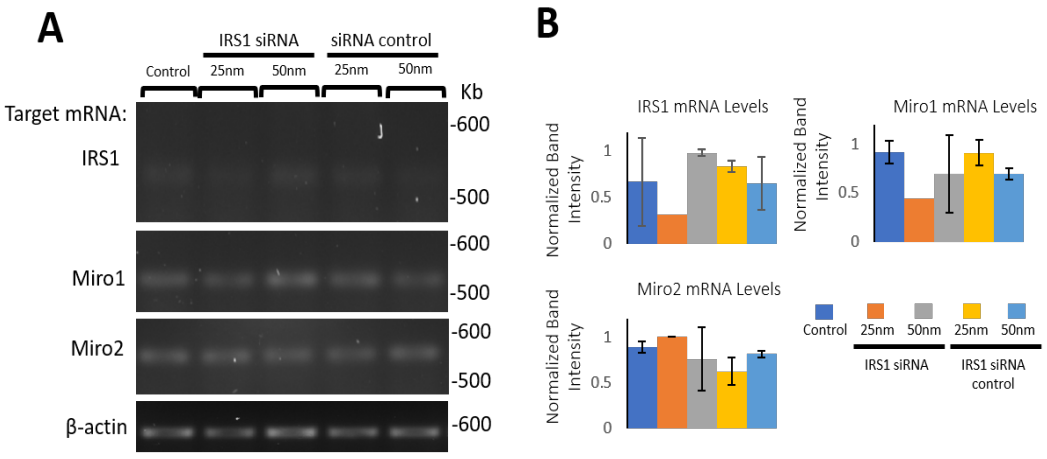

Figure 3.4: Wet lab validation of Miro expression in T2D Model

A is the gel electrophoresis result which reflects the IRS1, Miro1, Miro2 and β- actin mRNA levels in the IRS1 knocked-down SH-SY5Y cells. **B** is the bar graph of normalized intensities of bands in **A**.

4. Discussion

4.1 Miro in PD and T2D

From the RWR network analysis, the algorithm predicted Miro1 and Miro2 to play major roles in both PD and T2D. Majority of the other genes within the top 5 predictions such as MK01, RL40, RS27A, XIAP, KPCD, CALM and KPCE code for proteins that are intimately involved in cellular death processes. Since both PD and T2D ultimately result in cell death, the presence of these genes in the top 5 predictions is expected. On the other hand, The Miro1 and Miro2 code for proteins that are implicated in mitochondrial transport. Miro1 and Miro2 proteins are isoforms of the same protein Miro, an outer mitochondrial membrane protein that facilitates mitochondrial transport by anchoring it to microtubule motors.

Miro plays an important function in healthy neurons for mitochondrial homeostasis as it aids in the provision of healthy mitochondria when needed and the removal of unhealthy mitochondria [35]. Miro has also been associated with several of the candidate genes used in our PD network such as Parkin and PINK1. Parkin and PINK1 work in conjunction for the degradation of Miro which is crucial in mitophagy, the removal of mitochondria. When mitochondria are dysfunctional, they produce reactive oxygen species (ROS) such as superoxide radicals. The accumulation of this ROS is toxic to the neurons as it induces apoptosis [35]. Hence, it is imperative for the neuron to remove these damaged mitochondria. However, when Miro is attached to the mitochondrion, the mitochondrion is mobile and is therefore unable to undergo mitophagy. Parkin and PINK1 help to dissociate Miro from the mitochondrion and immobilize the mitochondrion, allowing the autophagosomes and lysosomes to begin degradation [36].

Since one of the hallmarks of PD is mitochondrial dysfunction, it is not surprising that a mitochondrial protein such as Miro is predicted to be associated with PD [37]. In a 2016 study, it was shown that impairment in Miro degradation is observed in different types of PD [35]. This functional impairment is expected in certain types of PD, where Parkin and PINK1 are mutated, but surprising in other types of PD with normal Parkin

and PINK1 [35]. While the cause for the impairment of Miro degradation is unknown, the results of this study suggest that the impairment of Miro degradation may be an important aspect in the progression of PD.

Conversely, Miro has yet to be investigated in T2D or insulin resistance. Since impairments in Miro degradation and insulin resistance are both observed in PD, this opens up the possibility of insulin resistance playing a role in the impairment of Miro degradation. Therefore, we chose to investigate if the development of insulin resistance, using a T2D model, has an effect on Miro degradation and parallel the results with a PD model. If proven right, this will uncover part of a novel and common molecular mechanism between the two diseases and aid in elucidating the enigmatic mechanisms of PD. To compare if Miro is similarly implicated in PD and T2D, it is necessary to observe a similar trend across increasing severities of both diseases.

4.2 PD Model

To mimic the varying severities of PD, SH-SY5Y cells were subjected to different concentrations of rotenone. Upon treatment, to confirm that rotenone was indeed acting on the cell cultures, caspase 3 and cleaved caspase 3 assays were carried out (Figure 3.3A). Caspase 3 is a protease which itself needs to be cleaved by another caspase protein to be activated. Upon activation, cleaved caspase 3 catalyzes the cleavage of key cellular proteins during apoptosis . Rotenone results in mitochondrial dysfunction and the accumulation of ROS that causes high levels of caspase-3-mediated apoptosis [38]. Hence, as the activation of caspase 3 is a hallmark of apoptosis, to test for rotenone-induced cell death, the levels of caspase 3 and cleaved caspase 3 were measured in the lysates of cells grown for 24 hours under the different conditions mentioned above. However, the immunoblot for cleaved caspase 3 was empty, hence it cannot be deduced if there was increased apoptosis in response to the rotenone treatment. The bands obtained from the caspase 3 immunoblot were still analyzed using ImageJ as the intensity of a band is directly proportional to the amount of Caspase 3 expression in a particular sample. Based on the normalized band intensities, the rotenone-treated samples had much lower Cas-

pase 3 levels compared to the untreated and DMSO-treated samples (Figure 3.3B). This was possibly due to the increased cleavage and activation of caspase 3. However, without knowing the cleaved caspase 3 levels, it remains inconclusive if the rotenone treatment had an effect on the cells and the PD model has not been established.

Nonetheless, the expression of Miro1 was still investigated via immunoblotting (Figure 3.3C). High background noise that was observed was likely due to the non-specific binding of the antibody to a different protein within the sample instead of Miro1. The same result was observed across replicates. Since western blotting for Miro1 was unsuccessful, Miro1, Miro2 and IRS1 mRNA levels were tested for in the rotenone-treated and control samples via RT-PCR and gel electrophoresis (Figure 3.3D). Since the degradation of Miro occurs on a protein level, no significant difference in mRNA levels across the different samples was expected. However, due to the lack of replicates, it cannot be concluded whether there was any statistically significant difference in the normalized band intensities (Figure 3.3E).

Overall, from the results shown in Figure 3.3, a cleaved caspase 3 assay must be repeated to show the successful treatment of rotenone. More replicates must also be carried out to conduct statistical tests to find out if there are significant differences in mRNA levels across the 6 samples. Furthermore, more optimization of the western blotting protocol could be done to obtain optimal Miro protein expression.

4.3 T2D Model

To establish the T2D model and observe a trend in the resultant Miro expression, IRS1 was knocked down in SH-SY5Y cells using 2 concentrations of IRS1 siRNA; 25 nM and 50 nM. These siRNA treated cell cultures were then compared with various controls. In particular, non-targeting siRNA-treated cells were used as a negative control as non-targeting siRNA does not target any known gene in the cells. This therefore shows that any changes in mRNA levels is specifically due to the effects of the sequence-specific IRS1 mRNA. To determine the success of the knockdown, the IRS1 mRNA levels in these cell

cultures were then tested for via RT-PCR and agarose gel electrophoresis. (Figure 3.4A). It can be seen that very faint bands were produced for all 5 samples for IRS1. One possible explanation might be due to weak binding of the primers. An analysis of the band intensities was then done using ImageJ and normalized against the intensities of their respective actin bands (Figure 3.4B). A two sample t-test was performed using this data to find out if there was any significant difference in band intensities across samples of varying conditions. The 25 nM siRNA treated sample was omitted from the statistical test due to the lack of a replicate. Surprisingly, the 50 nM siRNA-treated sample had a comparable band intensity to the controls, indicating that the IRS1 mRNA levels were similar across all these conditions except for the 25 nM siRNA-treated sample. Since the actin band are comparable, the difference in band intensities cannot be attributed to the unequal loading of samples. Hence, a possible reason for this difference could be the variable success of transfection in the 5 cell cultures. To improve transfection efficiency, instead of 2 days, a longer incubation time could be given for the cell cultures as it can take anytime between 48-96 hours for successful transfection to take place. There would also need to be further experimentation on the most optimal conditions for transfection of this particular siRNA in this cell line as the success of transfection can be influenced by many different factors. Along with IRS1, Miro1 and Miro2 mRNA levels were also tested in the T2D models.

From the results shown in Figure 3.4, it cannot be concluded whether the T2D model was indeed successful. For more accurate results, additional replicates must be done to find out if there was indeed insulin resistance in the T2D model and consequently if it had an influence on Miro in the SH-SY5Y cells.

5. Conclusion

Our study managed to provide a unique perspective in resolving the obscurity of PD's mechanism: by looking at PD through the lens of a related disease, T2D. Using network analysis on a PPI network we constructed with STRING database, we were able to isolate proteins that may play pivotal roles in the progression of PD and T2D. If these predictions are proven right, it will illuminate a novel pathway in the pathogenesis in PD and aid in developing more potent therapeutics.

Since the first association of insulin resistance with PD, the prognosis of insulin resistance in PD has remained inconclusive, owing to the vague understanding of PD. If an impairment in Miro degradation is observed in the T2D model, it would have contributed greatly to solving the cryptic puzzle of PD. In addition, if insulin resistance had been observed to cause the impairment in Miro degradation, there would be a strong implication that insulin resistance arises in the early stages of PD. Impairment in Miro degradation would then consequently impede mitochondrial homeostasis and cause the accumulation of ROS. A vicious cycle is thus perpetuated whereby dysfunctional mitochondria lead to the accumulation of ROS that results in more mitochondrial damage. Eventually, ROS toxicity results in cell death. Hence, the prognosis of insulin resistance would allow clinicians the opportunity to slow down or even prevent the disease by eliminating or alleviating insulin resistance.

We were unable to determine any conclusions on Miro in PD and T2D from our experiments due to time constraints and technical issues. Although we were unable to validate the network's predictions, the network promises an exciting prospect as a tool that can further our comprehension of many other diseases. The cell is intricately complicated, paralleling a complex machinery. When a disease condition arises, similar to the breakdown of a machine, it will be imperative to have an accurate apprehension of the assembly and wiring of the cell machinery. The network serves as a fundamental blueprint mirroring the interwoven tapestry within the cell and the network analysis will enable scientists to pinpoint the sources of perturbations for the disease condition. This positions the network as an extremely robust tool for identifying potential biomarkers and devising

effective treatments.

Despite the potential usefulness of the PPI network, it only offers one side of the story. The PPI network assumes the cell to be in a closed state but in reality, the cell is in an open state. Within the cell, interactions are not limited to only protein-protein interactions, other interactions such as RNA-protein and RNA-DNA interactions also exist. The cell also interacts with surrounding cells and is under the influence of environmental factors. To provide the full picture within a cell, the network should be integrated with other data such as genomic and environmental data, transforming a unidimensional approach to a multidimensional one. The PPI network is also limited by the incompleteness of the available data and the limitations of current analytic tools. Nonetheless, the network is the first step of many to building a complete image of the cellular complexity. Once achieved, it will serve as the basis for future network medicine [23].

In conclusion, the "think globally, act locally" paradigm of the network method allows researchers to formulate a multifaceted strategy to understand a particular disease since it encapsulates and integrates all available data. Therefore, we believe that adopting PPI network analysis to understand perplexing diseases such as PD is a far more holistic and effective approach as opposed to traditional methods of looking at a single pathway.

Bibliography

[1] R. Abdullah, I. Basak, K. Patil, G. Alves, J. Larsen, and S. Mller. Parkinson's disease and age: The obvious but largely unexplored link. *Experimental Gerontology*, 68:33–38, 2015.

[2] K. Bhattacharyya. *Hallmarks of Clinical Aspects of Parkinson's Disease Through Centuries. International Review Of Neurobiology* . Elsevier, 2017.

[3] K. Rosqvist, P. Odin, P. Hagell, S. Iwarsson, M. Nilsson, and A. Storch. Dopaminergic effect on non-motor symptoms in late stage parkinsons disease. *Journal Of Parkinson's Disease*, 8(3):409–420, 2018.

[4] W. Dauer and S. Przedborski. Parkinson's disease. *Neuron*, 39(6):889–909, 2003.

[5] T. Yasuhara, M. Kameda, T. Sasaki, N. Tajiri, and I. Date. Cell therapy for parkinsons disease. *Cell Transplantation*, 26(9):1551–1559, 2017.

[6] P. LeWitt. Levodopa for the treatment of parkinson's disease. *New England Journal Of Medicine*, 359(23):2468–2476, 2008.

[7] X. Yin, J. Geller, and Y. Li. *Health Information Science 5th International Conference, HIS 2016, Shanghai, China, November 5-7, 2016, Proceedings.* Springer International Publishing, 2016.

[8] S. Chatterjee, K. Khunti, and M. Davies. Type 2 diabetes. *The Lancet*, 389(10085):2239–2251, 2017.

[9] X. Yue, H. Li, H. Yan, P. Zhang, L. Chang, and T. Li. Risk of parkinson disease in diabetes mellitus. *Medicine*, 95(18):3549, 2016.

[10] Q. Xu, Y. Park, X. Huang, A. Hollenbeck, A. Blair, and A. Schatzkin. et al. Diabetes and risk of parkinson's disease. *Diabetes Care*, 34(4):910–915, 2011.

[11] E. Schernhammer, J. Hansen, K. Rugbjerg, L. Wermuth, and B. Ritz. Diabetes and the risk of developing parkinson's disease in denmark. *Diabetes Care*, 34(5):1102–1108, 2011.

[12] I. Aviles-Olmos, J. Dickson, Z. Kefalopoulou, A. Djamshidian, P. Ell, and T. Soderlund. et al. Exenatide and the treatment of patients with parkinsons disease. *Journal of Clinical Investigation*, 123(6):2730–2736, 2013.

[13] J-P. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, and N. Li. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437:1173–1178, 2005.

[14] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. Brembeck, and H. Goehler. et al. A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, 122(6):957–968, 2005.

[15] A. Zanzoni, M. Soler-Lpez, and P. Aloy. A network medicine approach to human disease. *FEBS Letters*, 583(11):1759–1765, 2009.

[16] T. Pawson and R. Linding. Network medicine. *FEBS Letters*, 582(8):1266–1270, 2008.

[17] M. Oti. Predicting disease genes using protein-protein interactions. *Journal of Medical Genetics*, 43(8):691–698, 2006.

[18] S. Kohler, S. Bauer, D. Horn, and P. Robinson. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008.

[19] M. Liu, A. Liberzon, S. Kong, W. Lai, P. Park, and I. Kohane. et al. Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genetics*, 3(36):96, 2007.

[20] M. Pujana, J. Han, L. Starita, K. Stevens, M. Tewari, and J. Ahn. et al. Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genetics*, 39(11):1338–1349, 2007.

[21] O. Vanunu, O. Magger, E. Ruppin, and R. Shlomi, T.and Sharan. Associating genes

and protein complexes with disease via network propagation. *Plos Computational Biology*, 6(1), 2010.

[22] J. Huang, D. Carlin, M. Yu, W. Zhang, J. Kreisberg, P. Tamayo, and T. Ideker. Systematic evaluation of molecular networks for discovery of disease genes. *Cell Systems*, 6(4):484–495.e5., 2018.

[23] A. Barabsi, N. Gulbahce, and J. Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.

[24] K. Lage, E. Karlberg, Z. Strling, P. lason, A. Pedersen, and O. Rigina. et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, 25(3):309–316, 2007.

[25] S. Navlakha and C. Kingsford. The power of protein interaction networks for associating genes with diseases. *Bioinformatics*, 26(8):1057–1063, 2010.

[26] L. Cowen, T. Ideker, B. Raphael, and R. Sharan. Network propagation: a universal amplifier of genetic associations. *Nature Review Genetics*, 18(9):551–562, 2017.

[27] H. Xicoy, B. Wieringa, and G. Martens. The sh-sy5y cell line in parkinsons disease research: a systematic review. *Molecular Neurodegeneration*, 12(1), 2017.

[28] K. Newhouse. Rotenone-induced apoptosis is mediated by p38 and jnk map kinases in human dopaminergic sh-sy5y cells. *Toxicological Sciences*, 79(1):137–146, 2004.

[29] R. Haeusler and D. McGraw, T. amd Accili. Biochemical and cellular properties of insulin receptor signalling. *Nature Reviews Molecular Cell Biology*, 19(1):31–44, 2017.

[30] S. Waters and J. Pessin. Insulin receptor substrate 1 and 2 (irs1 and irs2): what a tangled web we weave. *Trends In Cell Biology*, 6(1):1–4, 1996.

[31] M. Caruso, D. Ma, Z. Msallaty, M. Lewis, B. Seyoum, and W. Al-janabi. et al. Increased interaction with insulin receptor substrate 1, a novel abnormality in insulin resistance and type 2 diabetes. *Diabetes*, 63(6):1993–1947, 2014.

[32] O. Ali. Genetics of type 2 diabetes. *World Journal of Diabetes*, 4(4):144, 2013.

[33] C. Klein and A. Westenberger. Genetics of parkinson's disease. *Cold Spring Harbor Perspectives In Medicine*, 2(1), 2012.

[34] G. Smushkin and A. Vella. Genetics of type 2 diabetes. *Current Opinion In Clinical Nutrition And Metabolic Care,*, 13(4):471–477, 2010.

[35] C. Hsieh, A. Shaltouki, A. Gonzalez, A. Bettencourt da Cruz, L. Burbulla, and E. St. Lawrence. et al. Functional impairment in miro degradation and mitophagy is a shared feature in familial and sporadic parkinsons disease. *Cell Stem Cell*, 19(6):709–724, 2016.

[36] X. Wang, D. Winter, G. Ashrafi, J. Schlehe, Y. Wong, and D. Selkoe. et al. Pink1 and parkin target miro for phosphorylation and degradation to arrest mitochondrial motility. *Cell*, 147(4):893–906, 2011.

[37] A. Bartels and K. Leenders. Parkinson's disease: The syndrome, the pathogenesis and pathophysiology. *Cortex*, 45(8):915–921, 2009.

[38] F. Ahmadi, D. Linseman, T. Grammatopoulos, S. Jones, R. Bouchard, and C Freed. et al. The pesticide rotenone induces caspase-3-mediated apoptosis in ventral mesencephalic dopaminergic neurons. *Journal Of Neurochemistry*, 87(4):914–921, 2004.