DISCOVERY OF HIDDEN PATHWAYS IN PROTEIN NETWORK FOR DIABETES THERAPEUTIC ADVANCES AND TREATMENT

OGBU, HENRY NWAGU

(19PCH02041)

B. Tech Information Technology, Federal University of Technology, Minna

SEPTEMBER, 2021

DISCOVERY OF HIDDEN PATHWAYS IN PROTEIN NETWORK FOR DIABETES THERAPEUTIC ADVANCES AND TREATMENT

By

OGBU, HENRY NWAGU

(19PCH02041)

B. Tech Information Technology, Federal University of Technology, Minna

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE (M.Sc) DEGREE IN MANAGEMENT INFORMATION SYSTEMS, DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY, OTA.

SEPTEMBER, 2021

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Sciences in Management and Information Systems in the Department of Computer and Information Systems, College of Science and Technology, Covenant University, Ota, Nigeria.

Mr John A. Philip
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I, **OGBU, HENRY NWAGU (19PCH04021)** declare that this research was carried out under the supervision of Prof. Victor C. Osamor of the Department of Computer and Information Systems, College of Science and Technology, Covenant University, Ota, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

OGBU, HENRY NWAGU **Signature and Date**

CERTIFICATION

We certify that this dissertation titled **DISCOVERY OF HIDDEN PATHWAY IN PROTEIN NETWORK FOR DIABETES THERAPEUTIC ADVANCES AND TREATMENT** is an original research work carried out by **OGBU, HENRY NWAGU (19PCH02041)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Prof. Victor C. Osamor. We have examined and found this work acceptable as part of the requirements for the award of Master of Science in Management and Information Systems.

Prof. Victor C. Osamor
(Supervisor) **Signature and Date**

Dr. Olufunke O. Oladipupo
(Head of Department) **Signature and Date**

Prof. Olusegun Folorunsho
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

I dedicate this project to God Almighty for His grace in my life during my programme. Also, I dedicate this work to the National Information Technology Development Agency (NITDA) for their support and sponsorship throughout my MSc. Programme.

ACKNOWLEDGEMENTS

All thanks to God Almighty for giving me the insight and enablement to embark on this project.

I like to thank my parents, Mr. & Mrs Ogbu, for their constant support and love. My regards also extend to all my lovely siblings for their good wishes and helpful advice. I also extend my appreciation to Ogbu Sunday, Victoria, Samuel, Daniel, Bright and Confidence for their love and care.

My acknowledgement also goes to the Chancellor of Covenant University, Dr David O. Oyedepo, for his commitment to raising a generation of leaders with a difference. I appreciate the entire management of Covenant University for the role they played during my programme. Special thanks to the office of the Dean School of Postgraduate Studies, Covenant University, for the assistance and timely response to all my requests.

Furthermore, I want to acknowledge the Head of Department, Computer and Information Science (CIS), Associate Professor Olufunke O. Oladipupo, for her impact and tutelage.

I also thank my supervisor, Prof. Victor C. Osamor, for his constant support, guidance and encouragement throughout the period of this work. He is indeed a mentor and a support system.

I also thank the PG Coordinator, Dr Aderonke Oni, for her patience and lessons in and outside the lecture halls.

Special thanks to the National Information Technology and Development Agency (NITDA) for sponsoring my programme at Covenant University.

Again, my acknowledgement also goes to Dr Rilwan Enesi for his love and the reference letter he gave me for this programme. Special thanks to Mr Japhet and his family for their hospitality and love. I equally thank Dr Ganiyu Shefiu and Dr Abdulhameed A. Kayode, both from the Federal University of Technology, Minna, for their support and encouragement all along.

I also want to thank Nwafor Stanley, Adetona Joshua, Franklyn Emmanuel, Shofadekan Ayobami and Oruma Samson for their moral support and kind words of encouragement. They have been wonderful colleagues all along. I also want to acknowledge Mr Joshua Nudegbesi for his constant prayers and advice.

I am grateful to you all, and I love you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PPI	Protein-Protein Interaction
ORA	Over Representation Analysis
FCS	Functional Class Scoring
GSEA	Gene Set Enrichment Analysis
PIN	Protein Interaction Network
DEGs	Differentially Expressed Genes
T2D	Type 2 Diabetes
NCBI	National Centre for Biotechnology Information
GEO	Gene Expression Omnibus
LIMMA	Linear Model for Microarrays
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
AQR	aquarius intron-binding spliceosomal factor
PRPF19	pre-mRNA processing factor 19
XAB2	XPA binding protein 2
SNW1	SNW domain containing 1
SNRPF	small nuclear ribonucleoprotein polypeptide F
SNRPD2	small nuclear ribonucleoprotein D2 polypeptide
SNRPD1	small nuclear ribonucleoprotein D1 polypeptide
SNRPD3	small nuclear ribonucleoprotein D3 polypeptide
CDC5L	cell division cycle 5 like
EIF4A3	eukaryotic translation initiation factor 4A3
SNRNP40	small nuclear ribonucleoprotein U5 subunit 40
PPIL1	peptidylprolyl isomerase like 1
BUD31	BUD31 homolog
SYF2	SYF2 pre-mRNA splicing factor
SNRPA1	small nuclear ribonucleoprotein polypeptide A.'
SNRNP200	small nuclear ribonucleoprotein U5 subunit 200
RBM22	RNA binding motif protein 22
BCAS2	breast carcinoma amplified sequence 2
CDC40	cell division cycle 40
CRNKL1	crooked neck pre-mRNA splicing factor 1
MED4	mediator complex subunit 4
MED1	mediator complex subunit 1
SRSF7	serine and arginine-rich splicing factor 7
SRSF3	serine and arginine-rich splicing factor 3
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1
SRSF1	serine and arginine-rich splicing factor 1
HNRNPA1	heterogeneous nuclear ribonucleoprotein A1
SF3B3	splicing factor 3b subunit 3
CDK19	cyclin-dependent kinase 19

MED31	mediator complex subunit 31
MED28	mediator complex subunit 28
MED17	mediator complex subunit 17
MED10	mediator complex subunit 10
MED15	mediator complex subunit 15
MED26	mediator complex subunit 26
MED20	mediator complex subunit 20
MED9	mediator complex subunit 9
MED21	mediator complex subunit 21
MED8	mediator complex subunit 8
MED27	mediator complex subunit 27
MED11	mediator complex subunit 11
MED30	mediator complex subunit 30
MED29	mediator complex subunit 29
MED14	mediator complex subunit 14
MED16	mediator complex subunit 16
MED25	mediator complex subunit 25
MED19	mediator complex subunit 19
MED23	mediator complex subunit 23
MED24	mediator complex subunit 24
MED18	mediator complex subunit 18
MED13	mediator complex subunit 13
MED12	mediator complex subunit 12
DHX8	DEAH-box helicase 8
ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit
ERCC4	ERCC excision repair 4, endonuclease catalytic subunit
XPA	XPA, DNA damage recognition and repair factor
PRMT5	protein arginine methyltransferase 5
WDR77	WD repeat domain 77
RAD9A	RAD9 checkpoint clamp component A
HUS1	HUS1 checkpoint clamp component
RAD1	RAD1 checkpoint DNA exonuclease
MUS81	MUS81 structure-specific endonuclease subunit
GRK1	G protein-coupled receptor kinase 1
RHO	rhodopsin
SAG	S-antigen visual arrestin
AGO4	argonaute 4, RISC catalytic component
AP1G1	adaptor related protein complex 1 gamma 1 subunit
AP1S1	adaptor related protein complex 1 sigma 1 subunit
CDH13	cadherin 13
CDH9	cadherin 9

ABSTRACT

Diabetes is one of the world's deadliest diseases caused when the pancreas cannot produce the insulin required by the body to regulate the amount of sugar. Several attempts have been made to produce drugs that would be used to cure diabetes, but to no avail, and it has no cure as of today. Several experimental methods have been applied in the drug discovery process, but they are very slow, more expensive, and environmentally dependent. This study computationally modelled a protein-protein interaction network to identify pathways to diabetes disease that might be useful in the drug discovery process. This work was done with Cytoscape and Bioconductor package of R language. The differentially expressed genes (DEGs) were used to construct the protein-protein interaction network with STRING using k-means clustering. High confidence of 0.9 was used as a threshold for interacting proteins, and the network was further visualized and analysed for degree and betweenness centrality with centiscape, a plugin of Cytoscape 3.8.2. G: profiler was used to perform network enrichment so that similar genes were clustered together and a list of the most enriched pathways were found, and Cytoscape was used to discover, analyse, annotate and visualize the pathways associated with core genes in the diabetic network. The analysis from Cytoscape showed that Aquarius intron-binding spliceosomal factor, pre-mRNA processing factor 19 and XPA binding protein as the three genes that came out with the highest degree centrality score of 29 from the network interactions. However, the SNW domain containing 1 (SNW1) gene had the highest betweenness centrality score of approximately 1008 and a degree centrality of about 90% of the maximum score. Consequently, the significantly common pathways among all the involved genes as ranked by g: profiler using their adjusted p-value include mRNA splicing-major pathway, mRNA splicing, and processing capped intron-containing pre-mRNA pathways. Therefore, this study recommends that the significantly common pathways in the AQR, XAB2, PRPF19 and SNW1 genes be considered a possible drug target to seek solutions to diabetes Type 2.

Keywords: Diabetes Mellitus, Protein-Protein Interaction (PPI), Protein Interaction Network (PIN), Differentially Expressed Genes (DEGs), Pathway Enrichment Analysis.