INFERRING GENE REGULATORY NETWORK FOR THE SPOROGENIC STAGE OF *PLASMODIUM FALCIPARUM* LIFE CYCLE USING GRAPH NEURAL NETWORKS AND scRNA-SEQ DATASET

ONYEMAECHI, PROMISE
(22PBF02398)

AUGUST, 2024

INFERRING GENE REGULATORY NETWORK FOR SPOROGENIC STAGE OF *PLASMODIUM FALCIPARUM* LIFE CYCLE USING GRAPH NEURAL NETWORKS AND scRNA-SEQ DATASET

BY

ONYEMAECHI, PROMISE
(22PBF02398)
B. Sc. (Hons) Plant Biology and Biotechnology, University of Benin, Benin City.

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF MASTER OF SCIENCE (M.Sc.) DEGREE IN BIOINFORMATICS DEPARTMENT OF COMPUTER AND INFORMATION SCIENCES, COVENANT UNIVERSITY, OTA, OGUN STATE, NIGERIA.

AUGUST, 2024

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 Bioinformatics in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.

Miss Adefunke F. Oyinloye
(Secretary, School of Postgraduate Studies) **Signature and Date**

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

DECLARATION

I, ONYEMAECHI, PROMISE (22PBF02398) declare that this research was carried out by me under the supervision of Prof. Marion Adebiyi of the Department of Computer, Landmark University, Omu-Aran, Nigeria. I attest that this dissertation has not been submitted 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.

ONYEMAECHI, PROMISE **Signature and Date**

CERTIFICATION

This is to certify that this dissertation titled "**INFERRING GENE REGULATORY NETWORK FOR THE SPOROGENIC STAGE OF *PLASMODIUM FALCIPARUM* LIFE CYCLE USING GRAPH NEURAL NETWORKS AND scRNA-SEQ DATASET**" is original research earned out by **ONYEMAECHI, PROMISE (22PBF02398)** in the Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Prof. Marion Adebiyi. We have examined and found this work acceptable as part of the requirements for the award of Master of Science (M.Sc.) in Bioinformatics.

Prof Marion O. Adebiyi
(Supervisor) **Signature and Date**

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

Prof. Olusegun Folorunso
(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 unending supply of grace, and strength given to me during my master's degree programme. Furthermore, to biomedical and computational researchers Who paved the way, providing platforms for insightful biomedical investigations.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to Almighty God for His unwavering guidance and blessings to this point. Special thanks to my supervisor, Prof. Marion Adebiyi, for her invaluable guidance, support, and mentorship in this research journey. I am also deeply appreciative of the Head of Department, Prof. O. Oladipupo, and the Postgraduate Coordinator, Dr. Itunuoluwa Isewon, for their invaluable input and organization. I extend my sincere appreciation to CApIC-ACE for supporting my studentship as well as the members of the faculty, staff, and students of the Department, for their contributions and support in making this proposal possible.

TABLE OF CONTENTS

CONTENTS **PAGES**

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN: Convolutional Neural Network

CGN: Graph Convolutional Neural Network

GNN: Graph Neural Network

GRN: Gene Regulatory Network

GWAS: Genome-Wide Association Studies

ITNs: Insecticide-Treated Bed Nets

KO: Knock Out

NGS: Next-Generation Sequencing

PCA: Principal Component Analysis

RBC: Red Blood Cell

RNA-seq: RNA Sequencing

RNN: Recurrent Neural Network (deep learning model)

scRNA-seq: Single-Cell RNA Sequencing

UMAP: Uniform Manifold Approximation and Projection

WHO: World Health Organization

ABSTRACT

Malaria, primarily caused by the parasite *Plasmodium falciparum*, remains one of the most severe infectious diseases globally, particularly in sub-Saharan Africa, where it leads to significant morbidity and mortality. A deep understanding of the molecular mechanisms that govern *P. falciparum* infection, especially during the sporogonic stage of the parasite’s life cycle, is crucial for controlling malaria and identifying new therapeutic targets. In this context, gene regulatory networks (GRNs) provide a valuable framework for exploring the gene interactions that drive the parasite's life cycle and its response to environmental changes. This research aims to the GRNs for the sporogenic stage of *P. Falciparum* life cycle using Graph Neural Networks (GNNs), a deep learning technique applied to single-cell RNA sequencing (scRNA-seq) data. With the advent of scRNA-seq, gene expression can now be studied at the individual cell level. This is particularly important in the life cycle of *P. falciparum*, whose sporogonic stage is heterogeneous and involves different cell types throughout its life cycle. The study began with the preprocessing of scRNA-seq data, followed by clustering the cells to identify distinct stages of the parasite. A correlation-based co-expression network was then constructed to capture the relationships between genes within these clusters. Graph convolutional neural networks were employed to reconstruct the GRNs, leveraging their ability to learn the interactions between genes in the network. The performance of the inferred GRNs was evaluated using the AUC, AUPRC ratio, and EPR metrics, with average values of 0.6756, 1.51, and 2.48, respectively. This research not only enhances the understanding of the sporogonic stage of the life cycle of *P. falciparum*'s gene regulatory mechanisms but also identifies potential master regulators and genes that play central roles in the parasite's survival and pathogenicity. Four master regulators were identified for four selected clusters of genes associated with the life cycle of the sporogonic stage, based on the centrality scores of genes in the network. Among these master regulators, PF3D7_0304500, PF3D7_1357200, PF3D7_0822300, and PF3D7_0212300 stood out as the most critical genes within their respective clusters.

Keywords: *Plasmodium falciparum, malaria, Graph Neural Network, scRNA seq, Gene regulatory network, master regulator.*