EXPLORING MALARIA DISEASE BIOMARKERS: AN INTEGRATED APPROACH USING NETWORK TOPOLOGY AND ENRICHMENT ANALYSIS

Nana Oktaviana

Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Bogor, West Java, Indonesia
nanaoktaviana@apps.ipb.ac.id

Ananda Herlina Lutfitasari

Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Bogor, West Java, Indonesia
anandaherlina@apps.ipb.ac.id

Wisnu Ananta Kusuma*

Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University
Tropical Biopharmaca Research Center, IPB University
Bogor, West Java, Indonesia
ananta@apps.ipb.ac.id

Abstract

Malaria poses a substantial global public health threat, with 247 million reported cases and 619,000 deaths in 2021 alone, according to the World Health Organization. Identifying biomarkers is crucial for effective clinical intervention. This study focused on exploring protein-protein interaction (PPI) networks to identify significant proteins associated with malaria. Using RNA-seq gene expression data from the Gene Expression Omnibus, 92 unique malaria-related genes were filtered and subjected to network topology and enrichment analysis. Centrality analysis identified MMP9, LCN2, LTF, SPP1, COL6A1, MMP8, SDC1, TEK, COL17A1, and CEACAM8 as the top ten proteins with the highest centrality. Subsequent enrichment analysis highlighted SDC1 as a prominent malarial biomarker. This integrated approach contributes to a deeper understanding of malaria and holds the potential for informing targeted therapeutic developments, emphasizing the importance of advancing our knowledge in combating this life-threatening disease.

***Keywords*: Biomarkers; enrichment analysis; malaria; network topology.**

1. Introduction

Malaria is a highly serious and potentially fatal global health problem. According to the latest report from the World Health Organization (WHO) in the 2022 World Malaria Report, malaria cases reached 247 million in 2021, with a death toll of 619,000 [WHO (2022)]. In the same year, sub-Saharan Africa contributed nearly 95% of all malaria cases and approximately 96% of all deaths caused by this disease. Almost 80% of these deaths occur in children under the age of 5 years [Oladipo *et al*. (2012)]. On the other hand, the South-East Asia Region reported approximately 5.4 million malaria cases, contributing 2.1% to the total global malaria cases, while the Eastern Mediterranean Region recorded 6.2 million cases, contributing 2.5% to the total global malaria cases [WHO (2022)].

Malaria is an illness caused by protozoa that spreads through *Anopheles mosquitoes* [White *et al*. (2014)]. The causative agent for malaria is a tiny protozoan organism found within the *Plasmodium* species group, which includes multiple subspecies [Talapko *et al*. (2019)]. Among the five parasite species that cause malaria,

Plasmodium falciparum is responsible for the majority of malaria cases worldwide. One of the main challenges in managing malaria is the identification of reliable biomarkers.

According to Hulka *et al*. (1991), biomarkers are alterations occurring at the cellular, biochemical, or molecular level that are detectable in biological samples and signify biological, pathogenic, or therapeutic reactions. Biomarkers are specific biochemical molecules found in the body of individuals infected with malarial parasites. These molecules serve as indicators or markers for identifying the presence of malaria infection within an individual. Biomarkers play a crucial role not only in managing diseases but also in devising strategies before the occurrence of the disease, especially in the context of asymptomatic malaria [Jain *et al*. (2014)]. Identifying reliable and specific biomarkers for malaria can aid in timely and efficient diagnosis, enabling more effective management.

Network analysis offers a comprehensive method for comprehending the complexity of diseases, where nodes can symbolize various biological elements, such as genes, proteins, and metabolites, and edges can represent gene co-expression, physical protein-protein interactions, and other biological relationships in biological contexts [Alfano *et al*. (2023)]. Network topology refers to the distribution or organization of various network components, encompassing the relationships between these components, such as the connections and interfaces they utilize [Wu and Buyya (2015)]. Mistry *et al*. (2017) stated that centrality measures can be employed as features representing the influence of a node in interaction networks.

Enrichment analysis is widely employed to functionally analyze extensive gene lists identified using high-throughput technologies, such as expression microarrays [Machado *et al*. (2013)]. Functional enrichment analysis utilized Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) resources [Fan and Hu (2022)]. Gene Ontology (GO) is an established bioinformatics instrument employed to delineate the attributes of gene products in various species. It encompasses three distinct dimensions for characterizing protein function: molecular function, cellular components, and biological processes [Thomas (2017)]. According to Kanehisa and Goto (2000), the Kyoto Encyclopedia of Genes and Genomes (KEGG) is a system in which functional assignments involve connecting a group of genes within the genome to a network of interacting molecules in the cell, such as a pathway or a complex.

In line with Duffy *et al*. (2021) research, this study employed whole blood RNA sequencing, gene set enrichment analysis, and temporal modeling to identify potential markers for individual vaccine responses and comprehend the systemic immune responses to PfRAS vaccination. In this study, we investigated the immune responses in whole blood following a malaria challenge in participants administered the P. falciparum RAS vaccine. This study aimed to explore potential malarial disease biomarkers using an integrated approach involving network topology and enrichment analysis. This approach is expected to aid the identification of significant proteins associated with malaria. Hence, this research holds highly relevant objectives in global efforts to address the burden of malaria, enhance clinical management, and develop more effective therapies.Authors are encouraged to have their contribution checked for grammar. Abbreviations are allowed but should be spelt out in full when first used. Integers ten and below are to be spelt out. Italicize foreign language phrases (e.g. Latin, French).

2. Materials and Methods

2.1. *Data of malaria*

The Gene Expression Omnibus (data obtained from the Gene Expression Omnibus (GEO). GEO) is a public functional genomics data repository (https://www.ncbi.nlm.nih.gov/geo/) by GEO accession GSE192757. The dataset is about Immunization via mosquito bite with radiation-attenuated *Plasmodium falciparum* sporozoites (IMRAS). This study aimed to evaluate the safety, tolerability, and biomarkers of protection of healthy malaria-naïve adults. Radiation-attenuated *Plasmodium falciparum* sporozoites (PfRAS) was used as the intervention agent. Samples will receive mosquito bites from *Anopheles stephensi* mosquitoes infected with *Plasmodium falciparum* sporozoites (PfRAS) for true immunization, or non-infected for mock immunization. This includes whole blood RNA sequencing data and analysis of gene expression responses to identify potential biomarkers for individual vaccine responses after PfRAS vaccination.

2.2. *Methods*

A search of Gene Expression Omnibus (GEO) for a dataset includes the search terms, ‘malaria’ or ‘*Plasmodium falciparum*,’ and ‘homo sapiens.’ Data on malaria will be obtained from differentially expressed genes (DEGs) and analyzed using bioinformatics methods. The first method used GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to identify the DEGs. Access the GEO2R tool by clicking the text “Analyze with GEO2R.” To identify upregulated and downregulated genes, the samples were defined into two groups “true immunized” and “mock immunized” by highlighting relevant samples. The samples are highlighted in group color. Genes analyzed using the threshold $\log2(FC)>1$ and $p < 0.05$ were statistically significant. Volcano plots display the statistical significance of the difference relative to the magnitude of the difference for

each gene in the comparison, usually through the negative base-10 log and base-2 log fold-change, respectively [McDarmaid (2019)]

Search tools for the retrieval of interacting genes using the STRING database (https://string-db.org/) (version 12.0), which integrates known and predicted PPIs, can be used to predict functional protein interactions. The STRING database aims to collect, score, and integrate all publicly available sources of protein–protein interaction information and to complement these with computational predictions [Szklarczyk *et al.* (2019)]. The significance genes were analyzed in multiple proteins and co-expression as well as species limited to "homo sapiens" and a protein-protein interaction (PPI) score by using the required score of medium confidence (0.4).

Cytoscape software, version 3.10.1, was used to visualize the PPI network. Hub genes were screened using Cytoscape according to their centrality. All PPI networks were analyzed using Cytoscape (https://cytoscape.org/). Cytoscape is an app specially designed to calculate centrality indices used to identify the most important nodes in a network [Scardoni (2015)]. CytoNCA provides multiple centrality calculations for both weighted and unweighted networks, various forms of visualization analysis, and quantitatively evaluates the computation results [Tang (2015)]. Centrality provides an estimate of the importance of a node or edge to the connectivity or information flow of a network. The eight centrality measures in CytoNCA were: Betweenness centrality, closeness centrality, degree centrality, eigenvector centrality, Local Average Connectivity-based Centrality, Network Centrality, Subgraph Centrality, and Information Centrality [Tang (2015)]. The choice of centrality measurement depends on the type of genes considered important in the pathway [Gu *et al.* (2012)]. This study utilized four of them to identify important proteins, including Betweenness Centrality, Closeness Centrality, Degree, and EigenVector. Overall centrality was carried out as an optimal linear combination of the four measurements. Suppose S is the covariance matrix of a protein data matrix with dimensions $N \times 4$, where N is the number of protein data points and 4 represents the centrality values utilized in this study. Let $v = (v_1, v_2, v_3, v_4)$ be the eigenvector of S, obtained from the highest eigenvalue. The equation for determining the overall centrality value can be computed as follows:

$$OC(i) = v_1 BC(i) + v_2 CC(i) + v_3 DC(i) + v_4 EC(i) \quad (1)$$

The results of the hub genes were analyzed using Overall Centrality by Rstudio and short from the largest result of Overall Centrality.

Enrichment analysis is more effective for pathways in which multiple genes have strong biological signals [Raimand (2019)]. Biological pathways are sets of the top 10 genes from the three bioinformatics methods that were subjected to Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes pathway analysis using Enrichr (https://maayanlab.cloud/Enrichr/). Enrichr is a popular gene set enrichment analysis web server search engine that contains hundreds of thousands of annotated gene sets [Evangelista (2023)]. Gene Ontology considers three main aspects for explaining protein functions, including molecular function, cellular component, and biological process.

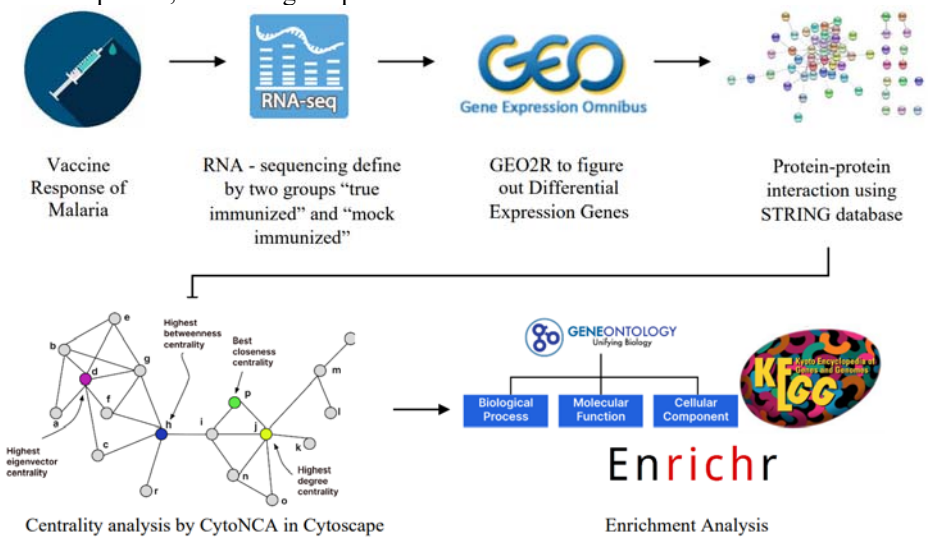

Fig. 1. Diagrammatic representation of the research.

3. Result and Discussion

3.1. *Differentially expressed genes*

The search identified GEO datasets and the selected dataset is GEO accession 'GSE192757'. The overall design of this data was a comparison of RNA-seq results from two cohorts, each with two treatments, 'true-

immunization' and 'mock-immunization.' Mosquitoes designated for 'mock-immunization' were raised, handled, and irradiated in the same manner as those for 'true-immunization'. The only difference was that they were fed blood cultures not infected with *Plasmodium falciparum* [Hickey *et al*. (2020)]. Differentially expressed gene (DEG) tools perform statistical tests based on quantifying expressed genes derived from the computational analyses of raw RNA-seq reads to determine which genes have a statistically significant difference [McDarmaid *et al*. (2018)]. GEO2R was used to compare the two groups of samples to identify genes that were differentially expressed across experimental conditions. software (https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html). The resulting data from DEGs using GEO2R produced 169 RNA-seq, which will be further analyzed to predict significant target proteins.

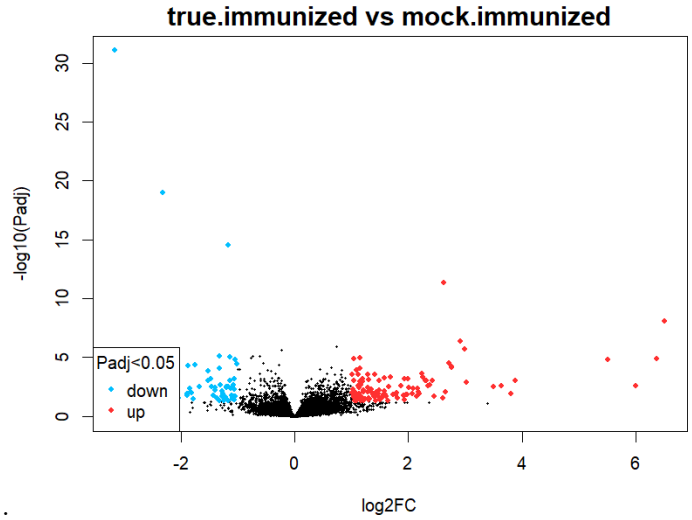

Fig. 2. Volcano plot of the differentially expressed genes between true immunized vs mock immunized

The dataset contains 117 upregulated DEGs (red) and 52 downregulated DEGs (blue). A volcano plot displaying statistical significance (-log10 P-value) versus magnitude of change (log2 fold change) was used to visualize the differentially expressed genes, and the highlighted genes were significantly differentially expressed at a default adjusted p-value cut-off of 0.05.

3.2. *Protein-protein interaction network and visualization of genes*

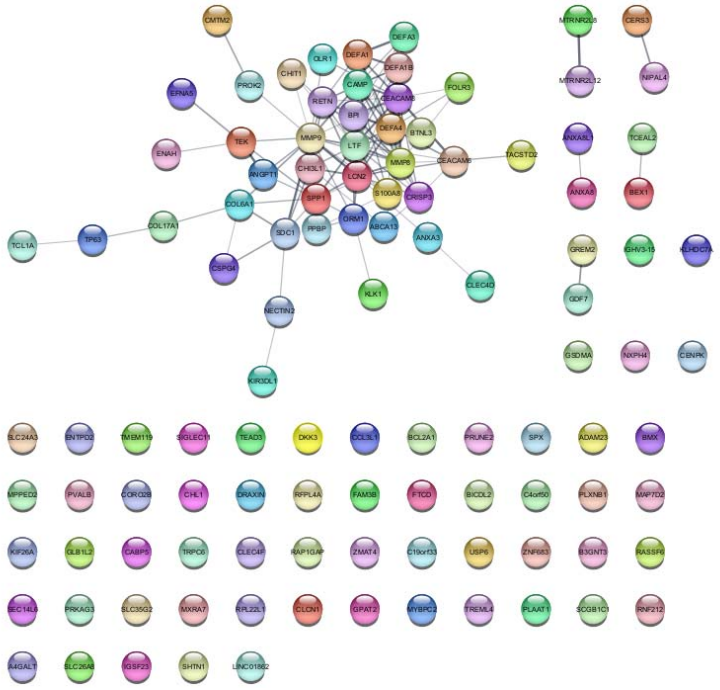

Fig. 3. Protein-protein interaction network in STRING database

This score is scaled between zero and one and provides an estimate of STRING's confidence in whether a proposed association is biologically meaningful, given all the contributing evidence [Szklarczyk *et al.* (2021)]. The resulting page shows the protein-protein interaction network for the protein of interest, with nodes

representing proteins and edges representing interactions between them. The scores are based on different types of evidence, including gene neighborhood, gene fusion, gene co-occurrence, gene co-expression, experiments/biochemistry, annotated pathways, and text mining [Szklarczyk *et al.* (2023)]. Several genes have duplicate proteins that are not detected in Homo sapiens; therefore, these protein targets are deleted, resulting in 110 unique proteins.

3.3. *CytoNCA in Cytoscpae*

Cytoscape can be used to perform network analysis and visualization, including the analysis of differentially expressed genes. From the STRING Database, Cytoscape allows users to map the attributes of the nodes and edges to properties such as color and edge width. Centrality measures can be calculated using R and the top genes from the overall centrality. There were 52 protein nodes with 135 interaction edges identified. The next step will involve centrality analysis to identify essential protein PPI networks [Tang *et al*. (2015)].

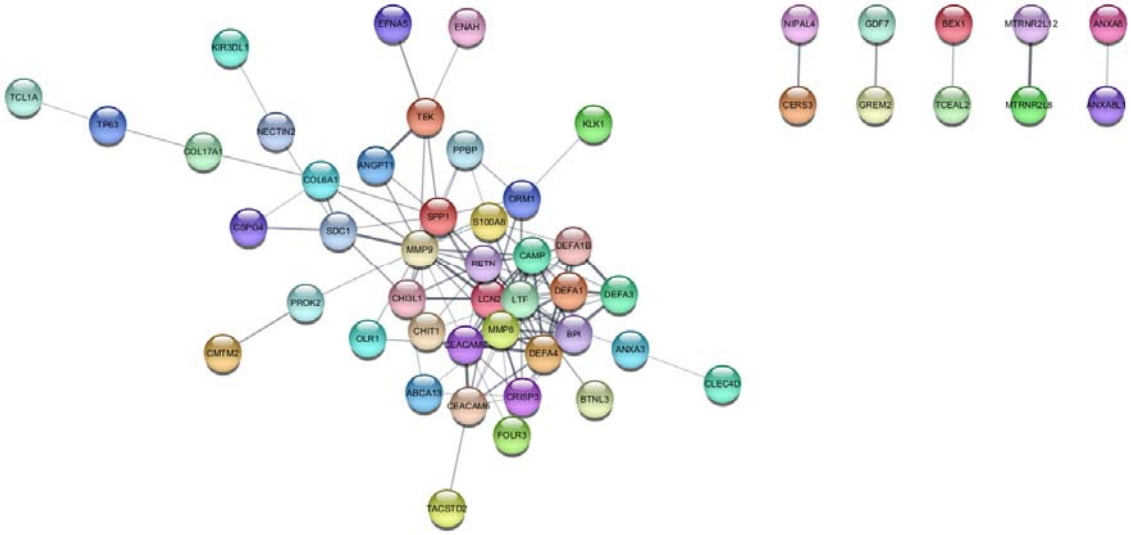

Fig. 4. Protein-protein interaction network in cytoscape.

Cytoscape showed protein-protein interactions (Figure 4). Nodes are colored in blue and edges are sized based on the number of genes shared by connected pathways. This study utilized four centrality measures from CytoNCA to identify proteins.

3.4. *Overall centrality with PCA*

The overall centrality was obtained through an optimal linear combination derived from eigenvectors with the highest eigenvalue based on PCA results [Gan and Djauhari (2012)]. Overall centrality offers a more comprehensive understanding of a node's importance within a network, considering both direct connections and the potential to act as a bridge between other nodes. Although degree centrality and betweenness centrality are valuable measures in network analysis, overall centrality provides a more holistic view of a node's importance by considering multiple factors. Optimality criteria were applied using PCA on a 52×4 data matrix representing 52 proteins and their standardized scores across the four measurements (Table 1).

Protein	BC	CC	DC	EC	OC
MMP9	5.1816	0.7118	2.4358	1.3557	15.8327
LCN2	1.7998	0.6647	2.2456	2.0945	6.6780
LTF	1.5369	0.6705	2.6259	2.2787	6.2203
SPP1	1.5867	0.6647	1.1045	0.5379	5.3883
COL6A1	1.7852	0.5522	-0.0366	-0.4516	5.1191
MMP8	0.8975	0.6531	2.2456	2.1321	4.2666
SDC1	1.2578	0.5576	0.1536	-0.3569	3.8376
TEK	0.8935	0.5199	-0.0366	-0.4666	2.7210
COL17A1	0.8755	0.3671	-0.6071	-0.7851	2.2073
CEACAM8	0.1685	0.6301	1.4849	1.5944	1.8306
ORM1	0.2348	0.5796	0.1536	0.0847	1.1310
CEACAM6	0.1949	0.5145	0.3438	0.5687	1.1007
CHI3L1	0.0033	0.6019	0.3438	0.3313	0.6505
RETN	-0.2384	0.6188	1.2947	1.5807	0.6201
CAMP	-0.2844	0.6188	1.2947	1.6765	0.4981
PROK2	0.1949	0.4987	-0.6071	-0.6230	0.4884
ANXA3	0.1949	0.4570	-0.6071	-0.5573	0.4594

Table 1. Centrality measures.

Table 1 (Continued)

Protein	BC	CC	DC	EC	OC
DEFA4	-0.3312	0.5468	1.4849	1.6830	0.4412
NECTIN2	0.1949	0.3622	-0.6071	-0.7763	0.3911
TP63	0.1949	0.1926	-0.6071	-0.8152	0.2711
BPI	-0.4857	0.5306	1.1045	1.5124	-0.2202
DEFA1B	-0.4879	0.5252	0.9143	1.3016	-0.3498
DEFA1	-0.5027	0.5252	0.9143	1.3808	-0.3888
S100A8	-0.3885	0.5907	0.3438	0.6266	-0.3991
CRISP3	-0.4267	0.5199	0.5340	0.7903	-0.4311
DEFA3	-0.5215	0.4934	0.3438	0.7505	-0.8213
PPBP	-0.5089	0.5145	-0.2268	-0.2947	-1.1357
CHIT1	-0.5215	0.5468	-0.4169	-0.1365	-1.2641
ABCA13	-0.5047	0.4725	-0.4169	-0.3103	-1.2729
ANGPT1	-0.5215	0.5092	-0.4169	-0.4718	-1.2927
FOLR3	-0.5215	0.4519	-0.4169	-0.1200	-1.3309
OLR1	-0.5215	0.5306	-0.6071	-0.4.101	-1.3959
CSPG4	-0.5215	0.3720	-0.6071	-0.7438	-1.5100
BTNL3	-0.5215	0.4519	-0.7973	-0.5432	-1.5710
KLK1	-0.5215	0.3720	-0.7973	-0.7380	-1.6287
EFNA5	-0.5215	0.3192	-0.7973	-0.7869	-1.6662
ENAH	-0.5215	0.3192	-0.7973	-0.7869	-1.6662
TACSTD2	-0.5215	0.3145	-0.7973	-0.6945	-1.6689
CMTM2	-0.5215	0.3005	-0.7973	-0.8010	-1.6795
CLEC4D	-0.5215	0.2637	-0.7973	-0.7951	-1.7055
KIR3DL1	-0.5215	0.1797	-0.7973	-0.8144	-1.7649
TCL1A	-0.5215	0.0285	-0.7973	-0.8180	-1.8716
ANXA8	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
ANXA8L1	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
BEX1	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
CERS3	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
GDF7	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
GREM2	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
MTRNR2L12	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
MTRNR2L8	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
NIPAL4	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118
TCEAL2	-0.5215	-2.0114	-0.7973	-0.8183	-3.3118

The key elements of the initial eigenvector γ were $v_1 = 0.41$, $v_2 = 0.41$, $v_3 = 0.58$, and $v_4 = 0.56$. The values represent the relative significance of each centrality metric. We plugged in these values into the equation for calculating the overall centrality, which can be expressed as follows:

$$OC(i) = 0.41BC(i) + 0.41CC(i) + 0.58DC(i) + 0.56EC(i)$$

In this case, degree centrality is the most crucial measurement of covariance structure among these measurements. This is followed by eigenvector centrality, betweenness centrality, and closeness centrality in that order. The top ten proteins with the highest overall centrality values, namely MMP9, LCN2, LTF, SPP1, COL6A1, MMP8, SDC1, TEK, CEACAM8, and COL17A1, were further analyzed using enrichment analysis.

The search identified GEO datasets and the selected dataset is GEO accession 'GSE192757'. The overall design of this data was a comparison of RNA-seq results from two cohorts, each with two treatments, 'true-immunization' and 'mock-immunization.' Mosquitoes designated for 'mock-immunization' were raised, handled, and irradiated in the same manner as those for 'true-immunization'. The only difference was that they were fed blood cultures not infected with *Plasmodium falciparum* [Hickey *et al*. (2020)]. Differentially expressed gene (DEG) tools perform statistical tests based on quantifying expressed genes derived from the computational analyses of raw RNA-seq reads to determine which genes have a statistically significant difference [McDarmaid *et al*. (2018)]. GEO2R was used to compare the two groups of samples to identify genes that were differentially expressed across experimental conditions. software (https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html). The resulting data from DEGs using GEO2R produced 169 RNA-seq, which will be further analyzed to predict significant target proteins.

3.5. *Gene Ontology and KEGG Pathway*

As shown in Figure 5 (a), the results of gene ontology analysis were visualized using bar plots and displayed the top 10 enrichment terms for each aspect of gene ontology. In the context of GO Biological Process terms, based on the results of previous studies, there is no direct evidence of a correlation between the listed biological process terms and malaria. The search results encompass studies on gene expression profiling in peripheral malaria infection, the inner membrane complex at various stages of the malaria parasite, the cellular

and molecular basis for malaria parasite invasion of human red blood cells, the genome sequence of the human malaria parasite *Plasmodium falciparum*, and vector biology, ecology, and control of malaria. While some of these studies mentioned cellular differentiation, extracellular matrix organization, and inflammatory responses, none of them specifically mentioned the listed biological process terms in relation to malaria. Therefore, further research is needed to determine whether there is a correlation between these biological process terms and malaria.

In the context of Gene Ontology Molecular Function terms, Pena and Plampona (2022) discussed heme oxygenase-1, carbon monoxide, and malaria and highlighted the involvement of different transition metals, including iron, in heme metabolism and the pathogenesis of malaria. This implies that the binding of transition metal ions may be significant in the investigation of malaria.

In the context of GO Cellular Component terms, based on the search results, there was no direct evidence of a correlation between the listed cellular component terms and malaria. The search results include studies on the inner membrane complex in the multiple stages of the malaria parasite, the cellular and molecular basis for malaria parasite invasion of human red blood cells, and the substrate-binding site of an iron-detoxifying membrane transporter from *Plasmodium falciparum*. While some of these studies mentioned cellular structures and compartments, none of them specifically mentioned the listed cellular component terms in relation to malaria. Therefore, further research is needed to determine whether there is a correlation between these cellular component terms and malaria.

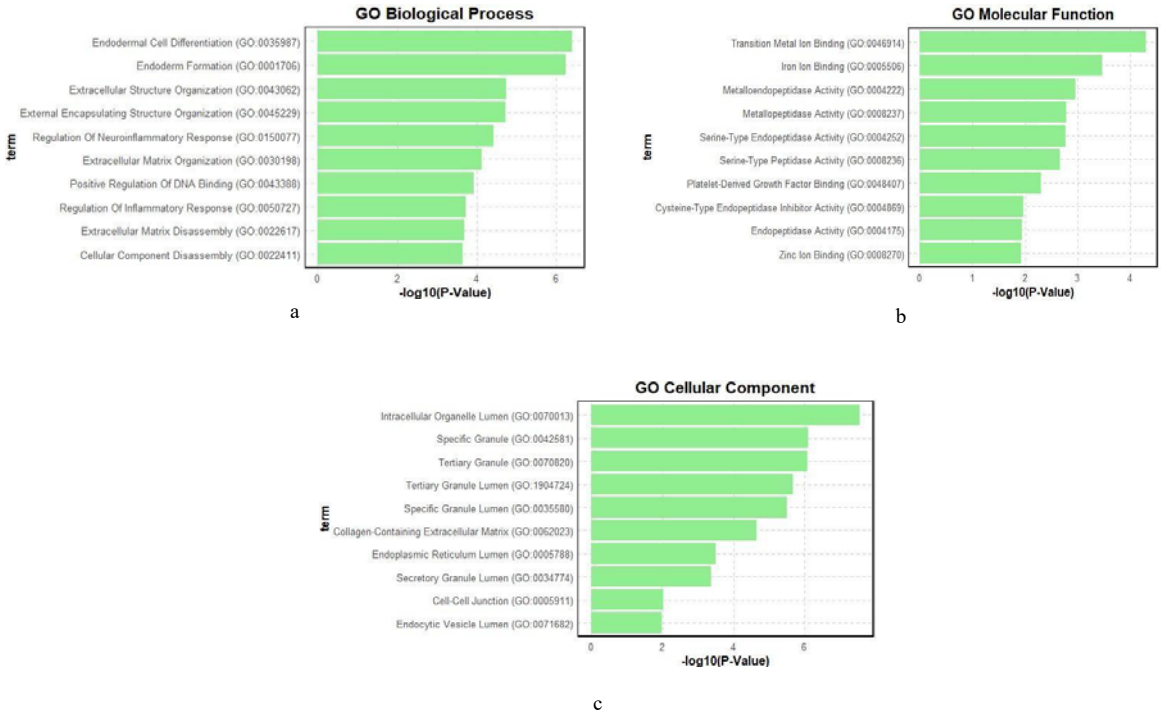

Fig. 5. Gene ontology (a) biological process, (b) molecular function, (c) cellular component.

As shown in Figure 6, the KEGG Pathway was sorted by the p-value. The results of KEGG pathway analysis can be further expanded to reveal gene sets related to malaria. This suggests that SDC1 can potentially serve as a biomarker for malaria. SDC1 is a genetic code responsible for syndecan-1, a proteoglycan with heparan sulfate that spans cell membranes and participates in processes such as cell adhesion, migration, and signaling [Agerbæk *et al.* (2019)]. While there is no direct proof of SDC1's involvement in malaria, some search findings hint at its potential role in the interaction between *Plasmodium falciparum* malaria parasites and host cells.

One study found that the seclusion of *Plasmodium falciparum* malarial parasites in the placenta is facilitated by the connection between VAR2CSA and chondroitin sulfate A on syndecan-1 (Ayres Pereira et al., 2016). This study suggests that SDC1 is modified by a unique chondroitin sulfate in the placenta, but not in other organs. Another study found that a glycosaminoglycan binding malaria protein, pl-CS, interacts with several proteins, including syndecan 1 (SDC1), in human melanoma cells [Salanti *et al.*, 2015]. This research indicates that pl-CS may serve as a potential indicator of melanoma progression. However, it is important to note that these computational findings still require further support from additional biological research to validate these findings.

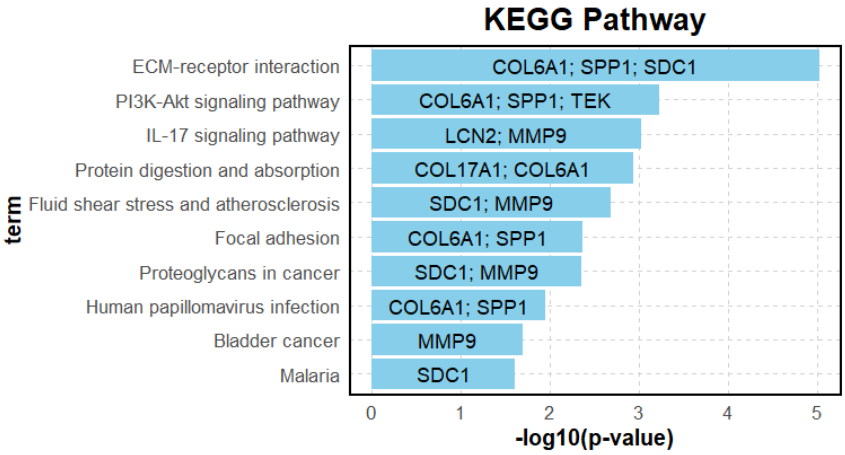

Fig. 6. KEGG Pathway.

4. Conclusion

The enrichment analysis results for the top 10 proteins, including MMP9, LCN2, LTF, SPP1, COL6A1, MMP8, SDC1, TEK, COL17A1, and CEACAM8, suggested a possible connection to malaria. Notably, SDC1 was the most significantly associated protein with the disease. This implies that SDC1 has the potential to be considered a candidate biomarker for malaria. These findings highlight a potential link between these proteins and malarial pathology.

Acknowledgment

This study was financially supported by the Tropical Biopharmaca Research Center. We are grateful for the funding and support provided by this research.

Conflict of Interest

The authors have no conflicts of interest to declare in this study.

References

[1] Agerbæk, M.; Bang-Christensen, S; Salanti, A. (2019): Fighting cancer using an oncofetal glycosaminoglycan-binding protein from Malaria Parasites. Trends in Parasitology, **35**(3), pp. 178–181. Available at: https://doi.org/10.1016/j.pt.2018.11.004.

[2] Alfano, C.; Farina, L.; Petti, M. (2023): Networks as biomarkers: uses and purposes. Genes, **14**(2). Available at: https://doi.org/10.3390/genes14020429.

[3] Ayres Pereira, M.; Mandel Clausen, T.; Pehrson, C.; Mao, Y.; Resende, M.; Daugaard, M.; Riis Kristensen, A.; Spliid, C.; Mathiesen, L.; E. Knudsen, L.; Damm, P.; G. Theander, T.; R. Hansson, S.; A. Nielsen, M.; Salanti, A. (2016): Placental sequestration of *Plasmodium falciparum* malaria parasites is mediated by the interaction between VAR2CSA and chondroitin sulfate a on syndecan-1. PLoS Pathogens, **12**(8), pp. 1–26. Available at: https://doi.org/10.1371/journal.ppat.1005831.

[4] Duffy, F. J.; Du, Y.; Carnes, J.; Epstein, J. E.; Hoffman, S. L.; Abdulla, S.; Jongo, S.; Mpina, M.; Daubenberger, C.; Aitchison, J. D.; Stuart, K. (2021): Early whole blood transcriptional responses to radiation-attenuated *Plasmodium falciparum* sporozoite vaccination in malaria naïve and malaria pre-exposed adult volunteers. Malaria Journal, **20**(1), pp. 1–13. Available at: https://doi.org/10.1186/s12936-021-03839-3.

[5] Evangelista, J. E.; Xie, Z.; Marino, G. B.; Nguyen, N.; Clarke, D. J. B.; and Ma'ayan, A. (2023): Enrichr-KG: Bridging enrichment analysis across multiple libraries. Nucleic Acids Research, **51**(W1), pp. W168–W179. Available at: https://doi.org/10.1093/nar/gkad393.

[6] Fan, S.; Hu, Y. (2022): Integrative analyses of biomarkers and pathways for heart failure. BMC Medical Genomics, **15**(1), pp. 1–18. Available at: https://doi.org/10.1186/s12920-022-01221-z.

[7] Gan, S.L.; Djauhari, M.A. (2012): An overall centrality measure : The case of U.S stock market. International Journal of Basic & Applied Science, **12**(06), pp. 99–104.

[8] Gu, Z.; Liu, J.; Cao, K.; Zhang, J.; Wang, J. (2012): Centrality-based pathway enrichment: a systematic approach for finding significant pathways dominated by key genes. BMC Systems Biology, **6**, pp. 56. Available at: https://doi.org/10.1186/1752-0509-6-56.

[9] Hickey, B.; Teneza-Mora, N.; Lumsden, J.; Reyes, S.; Sedegah, M.; Garver, L.; Hollingdale, M. R.; Banania, J. G.; Ganeshan, H.; Velasco, R.; … Epstein, J. E. (2020): IMRAS—A clinical trial of mosquito-bite immunization with live, radiation-attenuated P. falciparum sporozoites: Impact of immunization parameters on protective efficacy and generation of a repository of immunologic reagents. PLoS ONE, **15**(6), pp. 1–30. Available at: https://doi.org/10.1371/journal.pone.0233840

[10] Hulka, B. S. (1991): Epidemiological studies using biological markers: issues for epidemiologists. Cancer Epidemiology Biomarkers and Prevention, **1**(1), pp. 13–19.

[11] Jain, P.; Chakma, B.; Patra, S.; Goswami, P. (2014): Potential biomarkers and their applications for rapid and reliable detection of malaria. BioMed Research International, 2014, pp: 1-20. Available at: https://doi.org/10.1155/2014/852645.

[12] Machado, C. M.; Freitas, A. T.; Couto, F. M. (2013): Enrichment analysis applied to disease prognosis. Journal of Biomedical Semantics, **4**(1), pp: 1–17. https://doi.org/10.1186/2041-1480-4-21

[13] McDermaid, A.; Monier, B.; Zhao, J.; Liu, B.; Ma, Q. (2019): Interpretation of differential gene expression results of RNA-seq data: review and integration. Briefings in Bioinformatics, **20**(6), pp: 2044–2054. Available at: https://doi.org/10.1093/bib/bby067

[14] Mistry, D.; Wise, R. P.; Dickerson, J. A. (2017): DiffSLC: A graph centrality method to detect essential proteins of a protein-protein interaction network. PLoS ONE, **12**(11), pp. 1–25. Available at: https://doi.org/10.1371/journal.pone.0187091

[15] Oladipo, H. J.; Tajudeen, Y. A.; Oladunjoye, I. O.; Yusuff, S. I.; Yusuf, R. O.; Oluwaseyi, E. M.; AbdulBasit, M. O.; Adebisi, Y. A.; El-Sherbini, M. S. (2022): Increasing challenges of malaria control in sub-Saharan Africa: Priorities for public health research and policymakers. Annals of Medicine and Surgery, **81**, pp. 1-6. Available at: https://doi.org/10.1016/j.amsu.2022.104366
[16] Salanti, A.; Clausen, T. M.; Agerbæk, M. Ø.; Al Nakouzi, N.; Dahlbäck, M.; Oo, H. Z.; Lee, S.; Gustavsson, T.; Rich, J. R.; Hedberg, B. J.; Mao, Y.; Barington, L.; Pereira, M. A.; LoBello, J.; Endo, M.; Fazli, L.; Soden, J.; Wang, C. K.; … ; Daugaard, M. (2015): Targeting human cancer by a glycosaminoglycan binding malaria protein. cancer cell, **28**(4), pp. 500–514. Available at: https://doi.org/10.1016/j.ccell.2015.09.003
[17] Scardoni, G.; Tosadori, G.; Faizan, M.; Spoto, F.; Fabbri, F.; Laudanna, C. (2014): Biological network analysis with CentiScaPe: centralities and experimental dataset integration. F1000Research 2015, **3**, pp. 139. Available at: https://doi.org/10.12688/f1000research.4477.2
[18] Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. V. (2019): STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research, **47**(D1), pp. D607–D613. Available at: https://doi.org/10.1093/nar/gky1131
[19] Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; von Mering, C. (2021): The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic acids research, **49**(D1), pp. D605–D612. Available at: https://doi.org/10.1093/nar/gkaa1074
[20] Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P; Jensen, L. J.; von Mering, C. (2023): The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic acids research, **51**(D1), pp. D638–D646. Available at: https://doi.org/10.1093/nar/gkac1000
[21] Talapko, J.; Škrlec, I.; Alebić, T.; Jukić, M.; Včev, A. (2019): Malaria: The past and the present. Microorganisms, **7**(6), p. 179. Available at: https://doi.org/10.3390/microorganisms7060179
[22] Tang, Y.; Li, M.; Wang, J.; Pan, Y.; Wu, F. X. (2015): CytoNCA: a Cytoscape plugin for centrality analysis and evaluation of protein interaction networks. Biosystems, **127**, pp. 67–72. Available at: https://doi.org/10.1016/j.biosystems.2014.11.005
[23] Thomas, P.D. (2017): The Gene Ontology and the Meaning of Biological Function. The Gene Ontology Handbook, **1446**, pp. 15-24. Available at: https://doi.org/10.1007/978-1-4939-3743-1
[24] White, N. J.; Pukrittayakamee, S.; Hien, T. T.; Faiz, M. A.; Mokuolu, O. A.; Dondorp, A. M. (2014): Malaria. Lancet (London, England), **383**(9918), pp. 723–735. Available at: https://doi.org/10.1016/S0140-6736(13)60024-0
[25] World Health Organization. (2022): *World Malaria Report 2022*. Geneva, Switzerland: Author.
[26] Wu, C.; Buyya, R. (2015): Data Center Networks. In *Cloud Data Centers and Cost Modeling*. Elsevier Inc, Amsterdam. Available at: https://doi.org/10.1016/b978-0-12-801413-4.00013-1
[27] Yi, Y.; Fang, Y.; Wu, K., Liu, Y.; Zhang, W. (2020): Comprehensive gene and pathway analysis of cervical cancer progression. Oncology Letters, **19**(4), pp. 3316–3332. Available at: https://doi.org/10.3892/ol.2020.11439

Authors Profile

Nana Oktaviana is a final-year student majoring in Statistics and Data Science at the Department of Mathematics and Natural Sciences, IPB University. She has completed an internship in the field of bioinformatics at the Tropical Biofarmaka Research Center. Nana is passionate about applying statistical methods to solve real-world problems with a focus on data-driven research. She is currently exploring the intersection of statistics and life sciences to contribute to advancements in data-analysis methodologies.

Ananda Herlina Lutfitasari is a final-year student of Statistics and Data Science, at the Faculty of Mathematics and Natural Science, IPB University. Ananda has completed the bioinformatics internship at the Tropical Biofarmaka Research Center, IPB University. Ananda has applied statistical analysis to solve real problems. She is a data scientist who uses and processes raw data to discover interesting insights that help researchers make more informed decisions through visualization.

Wisnu Ananta Kusuma received his bachelor's and master's degrees from Bandung Institute of Technology, as well as his Ph.D. from Tokyo Institute of Technology, in 2012. He is currently an Associate Professor at the Department of Computer Science, IPB University. He also serves as Executive Secretary of Institute for International Research on Advanced Technology, IPB University, coordinator of Bioinformatics Working Group, Faculty of Mathematics and Natural Sciences, IPB University, and coordinator of Bioinformatics and High Performance Computing Research Group, Advanced Research Laboratory, IPB University. He has been the author of more than 60 articles and has reviewed for international journals. His current research interests include machine learning, high-performance computing, and bioinformatics research.