Transcriptional Profiles Analysis of COVID-19 and Malaria Patients Reveals Potential Biomarkers in Children

Nzungize Lambert [1 4] *, Jonas A. Kengne-Ouafo[1], Wesonga Makokha Rissy[2 5], Umuhoza Diane [3], Ken Murithi [6], Peter Kimani [6], Olaitan I. Awe [7], Allissa Dillman [8]

[1] Liverpool School of Tropical Medicine Research Unit, Centre for Research in Infectious Diseases (CRID), P.O. Box 13591, Yaoundé, Cameroon.

[2] African Institute of biomedical science and technology (AiBST), Wilkins Hospital, Block C, Corner J. Tongogara and R. Tangwena, Harare, Zimbabwe.

[3] Rwanda Agriculture and Animal Resources Board (RAB), P.O BOX 5016, Huye, Rwanda.

[4] Synthetic Biology and Omics Data Center, SynbioRwanda.

[5] Chinhoyi University of Technology (CUT), P.BAG 7724, Chinhoyi, Zimbabwe

[6]International Centre of Insect Physiology and Ecology (ICIPE) P.O. Box 30772-00100 Nairobi, Kenya.

[7]University of Ibadan, Ibadan, Oyo State, Nigeria

[8]National Institutes of Health, Bethesda, MD, U.S.A.

* Corresponding author: Nzungize Lambert, Email: nzulapa@outlook.com

Abstract

The clinical presentation overlap between malaria and COVID-19 poses special challenges for rapid diagnosis in febrile children. In this study, we collected RNA-seq data of children with malaria and COVID-19 infection from the public databases as raw data in fastq format paired end files. A group of six, five and two biological replicates of malaria, COVID-19 and healthy donors respectively were used for the study. We conducted differential gene expression analysis to visualize differences in the expression profiles. Using edgeR, we explored particularly the expressed genes in different phenotype groups relative to the healthy samples where 1084 genes and 2495 genes were differentially expressed in the malaria samples and COVID-19 samples respectively. Highly expressed genes in the COVID-19 samples were associated with biological processes such as cell division (CCDC124) and SLC12A5-AS1 a lncRNA gene associated with NK-cell while in the malaria samples were associated with biological processes such as immune response (CTSL), T cell activation (RSAD2) and proteolysis (LAP3). By comparing both malaria and COVID-19, the overlaps of 62 differentially expressed genes patterns were identified. Among the shared genes, the hemoglobin complexes and lipid mediators are highly expressed. We found six genes such as CYB5R3, RSAD2, ALOX15, HBQ1, HBM and PNPLA2 associated with malaria and COVID-19 diseases in children, which can be further validated as potential biomarkers. Our study provided new insights for further investigation of the biological pattern in hosts with malaria and COVID-19 coinfection.

Keywords: Malaria, SARS-CoV-2, COVID-19, children, RNA seq data, transcriptomic, gene expression, biomarker.

Introduction

The clinical diagnosis and distinction between malaria infection and COVID-19 in children presenting with malaria symptoms at a health care facility is a challenge to clinicians due to their overlapping symptoms. This causes a potential risk of misdiagnosis and in turn inappropriate treatment, therapy provision or untimely preventable death. The age structure and demography play a key role in COVID-19 mortality, where death tends more in elders than children [1].

Malaria is an opportunistic parasitic infection documented to be the leading cause of mortality and morbidity globally [2]. In 2020, due to the disruption to service because of the COVID-19 pandemic, the malaria case incidence increased in Africa and counted about 95% of cases [3]. Sub-Saharan Africa bears the highest burden of the disease, with *Plasmodium falciparum* contributing to the most severe form of the disease. Over the past two decades, the aversion of 1.5 billion malaria cases and 7.6 million malaria related deaths, has been greatly attributed to the use of Long-lasting Insecticidal nets (LLINs), Indoor residual spraying (IRS), and Artesunate combination therapy (ACT) as rolled out by existing national malaria control programs (NMCPs) [3]. Nonetheless, the recent WHO report still documents 229 million new infections and 409000 deaths globally as of 2019, with the highest observed mortality occurring in children under the age of five [4-6]. In this vulnerable population, malaria is quite severe and leads to the majority of most hospital admissions.

The sudden emergence of COVID-19 caused by a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in Wuhan, China in 2019, and the subsequent spread globally, is one such case [7]. To date no specific treatment while most of the information about COVID-19 viral infection has been published in different renowned journals [8, 9]. The nature of the disease coupled with overlapping symptoms brought with it a lot of confusion, especially in Sub-Saharan Africa where the burden of malaria alongside other infectious diseases is high. Recently, a lot of clinical studies have documented a strong relationship between severe malaria infection and SARS-CoV-2 in adults across the globe [10-12]. The interaction of the host and pathogen, either malaria or COVID-19 infection, defines the diagnosis of the pathogen.

Over the years, standard laboratory techniques used in the study of the role of genes in disease development include Northern blotting [13, 14], which allowed the study of gene expression via RNA detection [15]; Quantitative PCR [16, 17], contribute for the detection, characterization and quantification of RNA transcripts; and Microarray analyses, used to simultaneously detection of the expression levels of multiple genes at a time. Nonetheless, their limitation is the requirement of prior knowledge of genes, transcripts and the availability of a limited number of probes. Recently, RNA seq gene expression profiling using Next-generation sequencing (NGS) has supplemented microarrays as the preferred method for transcriptome wide identification of differentially expressed genes [18, 19]. Like other NGS platforms, this technique allows for massive high throughput sequencing, identification of novel transcriptomes, and the ability to perform single nucleotide resolution.

The advances in bioinformatics technology have enabled subsequent downstream analyses of the sequencing platform outputs. This includes the provision of high-quality visual outputs using qualitative and quantitative data that clearly describe what is being observed at the transcriptomic level [20-22]. These are inclusive of principal component analysis plots (PCA), Sample-to-sample distance plots, Dispersion estimate plots, Histogram of P-values plots, and MA plots. Recently, enrichment analysis tool kits provided visualization of overexpressed genes and transcripts in the treatment and affected groups, including the pathways associated with the observed differences [23-25]. This study provides a tool that utilizes RNA seq data for the identification of a biomarker that can help in the accurate and timely diagnosis of COVID-19 infections in children presenting with severe malaria symptoms and vice versa (Table 1).

The devastating impact of infectious disease outbreaks and pandemics on health systems could be overwhelming, especially when there is an overlap in clinical presentations with other disease conditions, for instance cases of malaria and COVID-19 (Table 2). We hypothesized that there can be similar biomarker signatures based on the children's immune responses to the two diseases. Therefore, by using RNA seq datasets available on open access databases, we explored the relationship and further characterized the distinctiveness of each etiological presentation classification at the transcriptome level.

Table 1. Similarities in clinical presentation and pathophysiological between malaria and COVID-19 infection.

	Malaria	**COVID-19**
Incubation period	Varies from 7 to 30 days (average 7-14 days).	Range from 2 to 14 days (average 5-6 days).
Pathophysiology	A cytokine storm triggers an exaggerated inflammatory response that may damage blood vessels, kidneys, the liver and lungs.	
	Acute respiratory distress syndrome (ARDS) from pulmonary thrombosis consequent to cytokine storm.	
Fatality	Fatal in children and pregnant women.	Mild in children and Fatal in elderly
Symptoms	Chest retraction	Trouble breathing
	Impaired taste	Loss of smell and taste
	Diarrhea, vomiting and Cough (frequent in children with malaria).	Diarrhea, vomiting and Cough.
	Headache, muscle or joint pain.	
	Fever, High temperature 38 °C or above	

Material and Methods

We used RNA seq data paired end reads produced by NGS platforms (Illumina HiSeq 2500 and Illumina NovaSeq 6000) from public repositories such as Gene Expression Omnibus (GEO) and European Nucleotide Archive (ENA). The reference genome were obtained from Open-Source databases such as ENSEMBL and GENCODE. All RNA seq dataset has been designed for transcriptomic analysis.

Data collection

The RNA-seq datasets were collected to study how the expression profiles under the host responses to pathogen either *Plasmodium falciparum* or SARS-CoV-2 in children (Fig.1). Specifically, our study examined the immune system responses based on the expression level profiles of patients with COVID-19 and malaria infection.

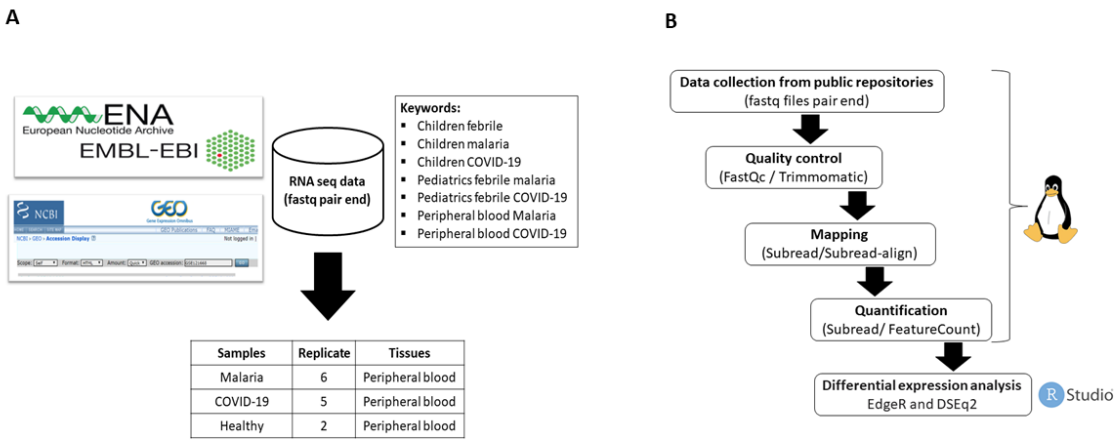


Figure 1. An illustration of the experimental design and RNA seq data analysis. **A)** A total of 13 RNA seq datasets of infected children were collected from public repositories, analyzed from three groups: healthy controls (2 samples); COVID-19 patients (5 samples); and malaria patients (6 samples). **B)** Downstream analysis. The workflow demonstrates the preprocessing of raw sequence data on the Linux platform, as well as downstream analysis of RNA-seq data using the R packages under the RStudio platform.

Data preprocessing and mapping

The raw data fastq files of paired end reads of between 100-150 bp read length were obtained from the public repositories with 13 samples, consisting of three groups of RNA seq data such as the group of COVID-19 samples with five biological replicates, a group of malaria samples with six biological replicates and a healthy group considered as control with two biological replicates.

The initial processing and quality assessment of the raw sequence reads was performed using FASTQC (v0.11.9, Babraham Bioinformatics, UK) tool for raw read processing to check for quality scores and identify good reads with default parameters. We checked for the percentage of the GC content of our raw sequences, the Phred score of our reads, per base quality score, per read-quality score, the overall quality score for the run, and made decisions on what parameters to adjust. Adapter trimming was then performed using the Trimmomatic (v0.39) [26]. tool to remove bad reads and low-quality reads. Clean RNA-seq reads were aligned to the human reference genome downloaded from the Gencode database (GRCh38.primary_assembly.genome.fa) and the annotation file was downloaded from the ENSEMBL database (Homo_sapiens.GRCh38.104.gtf). We indexed the reference genome with Subread/subread-buildindex tool and cleaned RNA seq reads that were aligned to the latest human reference genome using Subread/Sub-read align tool, therefore the output was in bam format. Alternatively, we performed transcript quantification using Subread/featureCounts tool and then generated a matrix as output for downstream analysis of differential gene expression using R packages (Fig.1).

Differential gene expression analysis

For differential expression analysis, we used edgeR (v3.34.0) a Bioconductor package in R software (v4.1.0) to determine the expression difference between groups. To ensure that all samples had a similar range and distribution of expression values, we performed a normalization process using a TMM (Trimmed Mean of M-values) method. The data normalized using the TMM method were quantitatively comparable between the COVID-19 group, the malaria group and the healthy group. We used PCA (Principal component analysis) to visualize the distribution of data between samples. In edgeR, Benjamini and Hochberg's method was used for estimating the adjusted p-values. In our study, by using a cut-off of 0.05 no significant expressed genes were detected upon COVID-19 samples and healthy samples. Therefore, to continue with the analysis, the adjusted p-values were relaxed to 0.1 to get the subjective significant expressed genes. We consider all genes with an adjusted *p* value (padj< 0.1) as significant (a fraction of 10% false positives acceptable) and differentially expressed genes with fold change (absolute log2) of 1 (abs(log2FoldChange) > 1). The DAVID-Functional Annotation Tool was used for functional enrichment analysis based on the differentially expressed genes between both groups

such as COVID-19 and malaria and the GOplot 1.0.2 package used for visualization. KMeans clustering was performed using the Integrated Differential Expression and Pathway analysis 95 (iDEP-95) platform (http://bioinformatics.sdstate.edu/idep/).

Table 2. Prevalence of children exposure to malaria and COVID-19 infections.

Condition	Prevalence *(%)*	Region	Ref.
Malaria	16	Global	[27]
COVID-19	18.9%	United States	[28]
Malaria and COVID-19 coinfection	11	Global	[29]

Results

RNA seq datasets characteristics

We used PCA to describe the clustering of the RNA seq data for six samples of malaria, five samples for COVID-19 and two samples for healthy as a control. The control and two diseased group samples showed the greatest difference. Here, each group was seen to cluster distinctively together. However, the COVID-19 seemed a bit scattered (Fig. 2B).

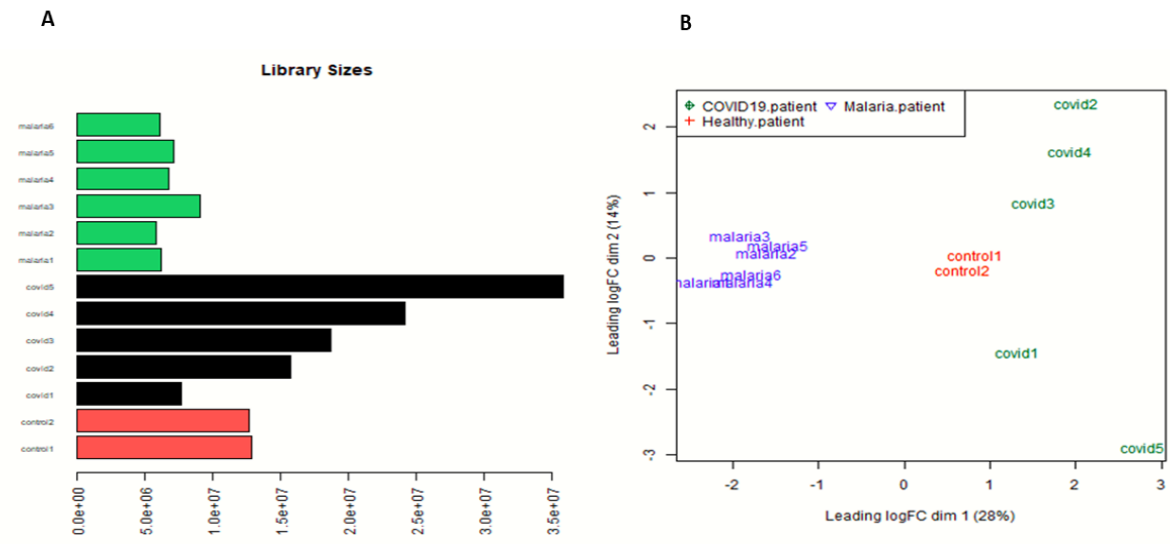


Figure 2. Distribution of RNA seq data between samples. **A)** Bar chart illustrating the library size of total reads that aligned to the human genome. The three groups represented by malaria (green), COVID-19 (black) and control (red). **B)** Principal-component analysis (PCA) of children transcriptome from PBMC of different conditions: children infected with COVID-19, children with malaria infection, and healthy children. COVID-19 samples and malaria samples were clearly separated. However, within the COVID-19 samples were relatively scattered, indicating poor consistency.

Correlation between patients with malaria and COVID-19 infection

We explore the correlation in immune system response between patients with malaria and COVID-19 infection. The heatmap depicting the levels of differentially genes in each patient as indicated by relative intensity (Fig. 3).

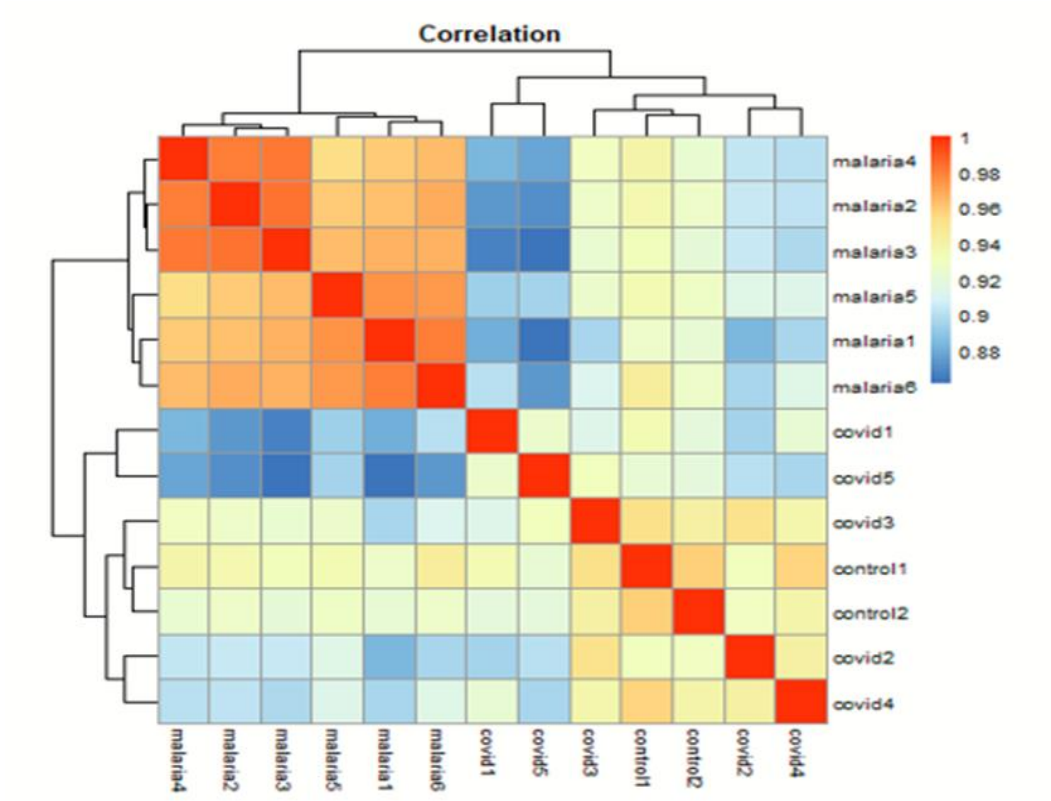


Figure 3. Heatmap comparing the relative expression levels of genes differentially expressed between children with of COVID-19 and malaria infection versus the healthy donor.

Gene expression among children with COVID-19 diseases

To identify differentially expressed genes between children with malaria and COVID-19 disease we used a volcano plot as covariates, and we considered a fold change adjusted *p*-value (padj) value < 0.1 as significant. The volcano plot highlights differences in gene expression with the key genes involved in the host responses. There we observed an upregulation of key genes involved in the host responses, such as SLC12A5-AS1, RAP1GAP, TCF3, STB1, CCDC124, IFI27, and CD177 (Fig. 4).

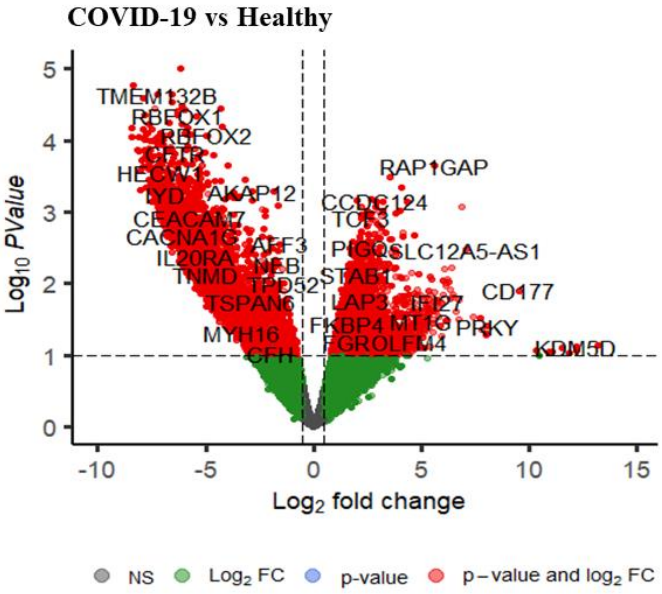


Figure 4. Volcano plot for comparison between COVID-19 infected individuals and healthy. The X-axis shows $\log_2$ fold change (positive values are up regulated relative to healthy. The Y-axis shows the −log10 of BH adjusted p-value (padj) value. The horizontal dashed line marks P= 1%, and the vertical dashed lines indicate two-fold expression difference among conditions. The differentially expressed genes are indicated in red according to a P<.01, green points represent genes not passing the threshold, and grey points represent genes with no significant difference.

Gene expression among children infected with malaria

The volcano plots highlight differences in gene expression between children infected with malaria and healthy children. There we observed an upregulation of key genes involved in the host responses, such as CTSL, RN7SKP255, RSAD2, NBN, MTHFD2, and LAP3 in children infected with malaria (Fig. 5).

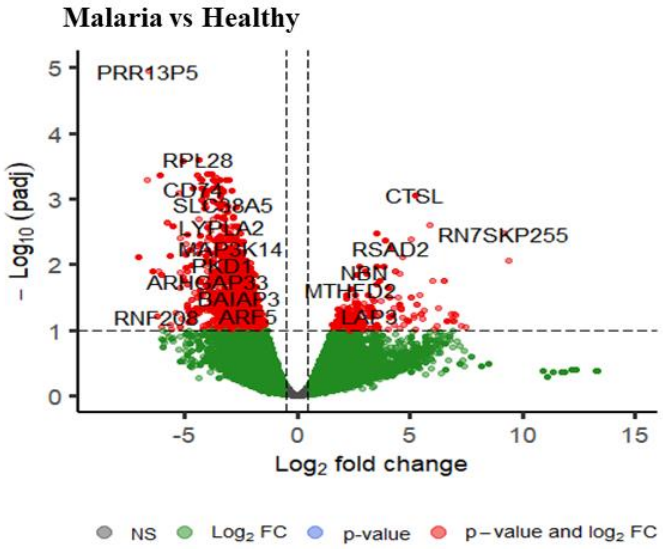


Figure 5. Volcano plot for comparison between malaria infected individuals and healthy. The X-axis shows $\log_2$ fold change (positive values are up regulated relative to healthy. The Y-axis shows the −log10 of BH adjusted p-value (padj) value. The horizontal dashed line marks P= 1%, and the vertical dashed lines indicate two-fold expression difference among conditions. The differentially expressed genes are indicated in red (padj, 0.01 & $\log_2$ FC>1). Red points indicate upregulated genes, green points represent without significantly different expression, and grey points represent genes with no significant difference.

The host transcriptional response between COVID-19 and malaria infection

Genes are more highly expressed in malaria group versus COVID-19 group such as ENPP2, CTSL, TNIP3, TWF1P1, SAMD9L are highly expressed in malaria and ARF5, MAP3K14, SPPL2B, DVL2, and MYL9 are highly expressed in COVID-19 samples (Fig. 6).

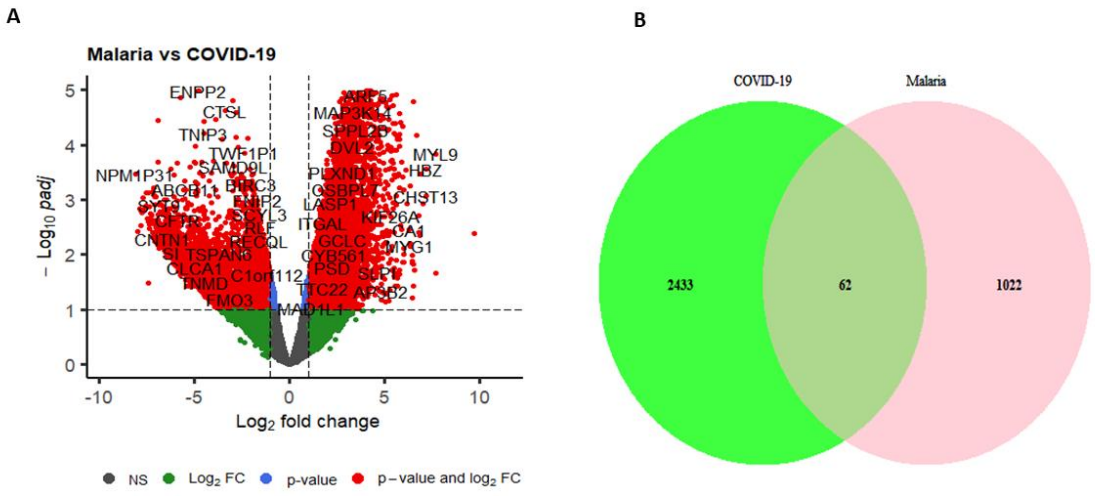


Figure 6. Immune system response in children with malaria infection and COVID-19 disease. A) Volcano plot for comparing between malaria and COVID-19 infected group samples. The X-axis shows $\log_2$ fold change (positive values are up regulated relative to malaria. The Y-axis shows the −log10 of BH adjusted p-value (padj) value. The horizontal dashed line marks P= 1%, and the vertical dashed lines indicate two-fold expression difference among conditions. The differentially expressed genes are indicated in red (padj, 0.01 & $\log_2$ FC>1). Red points indicate upregulated genes, green points represent without significantly different expression, and grey points represent genes with no significant difference. B) Venn diagram showing the shared, and unique numbers of differentially expressed genes in COVID-19 group and malaria group together. The overlaps of expression pattern of 62 genes differentially expressed were identified (padj< 0.1, |log2FC| > 1).

Heatmap of the overlap genes between malaria and COVID-19 patients

The number of differentially expressed genes (DEGs) for each infection and the overlap between immune responses of malaria patient and COVID-19 patient are presented in relative intensity of differentially expressed genes (Fig. 7).

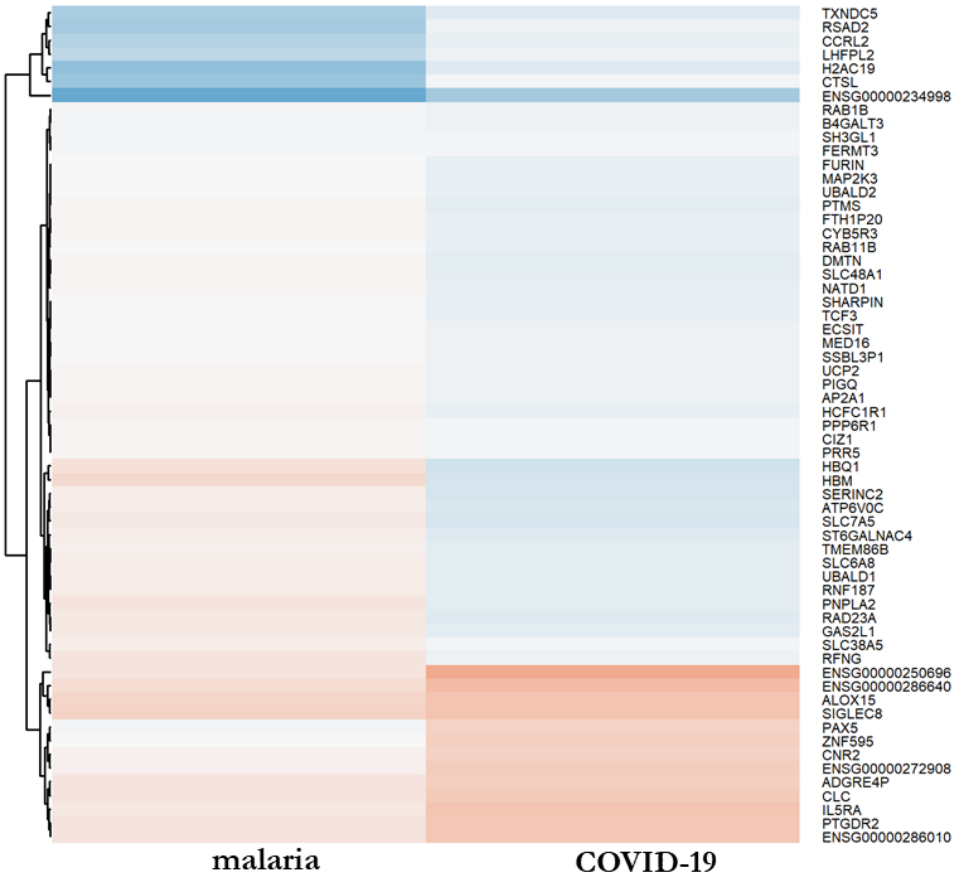


Figure *7.* Number of common DEGs per infection. Heat map of 62 genes overlapped differentially expressed from malaria and COVID-19. The columns represent individual patient per infection. Rows indicate immune response genes with significant differences in expression among both infections. Colors in the figure from red to blue indicate the level of gene expression from high to low.

Functional enrichment of overlap genes

Gene ontology consists of three parts such as cellular components, biological processes, and molecular functions. The top two showed significant e enrichment are hemoglobin complex and lipid particle among the 62 genes overlapped between malaria and COVID-19 patient (Fig. 8). It is worth noting that those genes were to a large extent upregulated in covid patients as opposed to malaria ones showed reduced expression. These genes were found to be significantly associated with cellular components such hemoglobin complex and lipid particles.

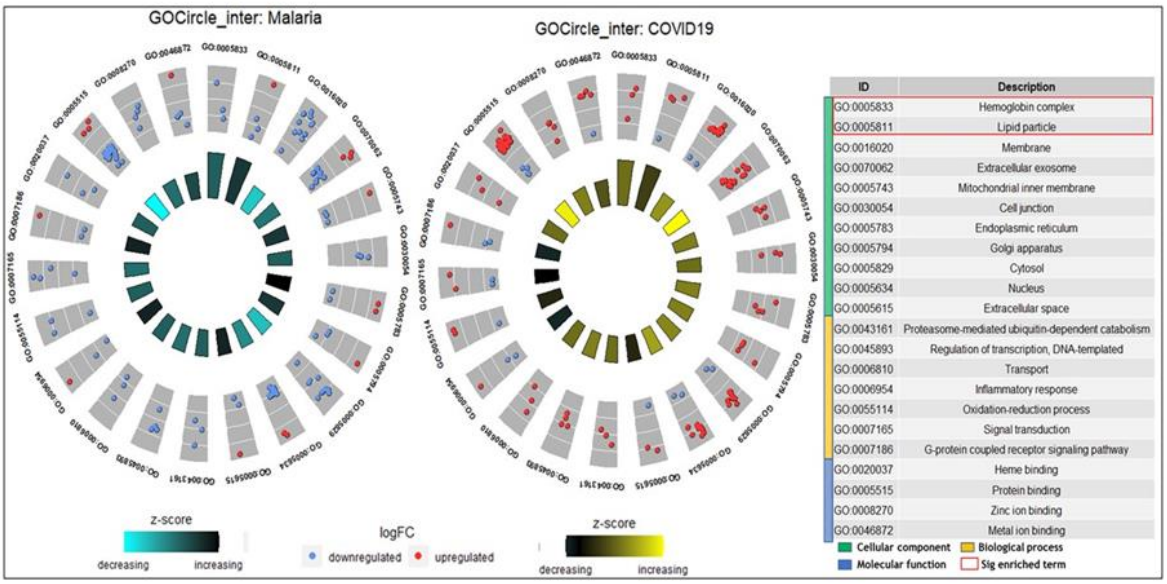


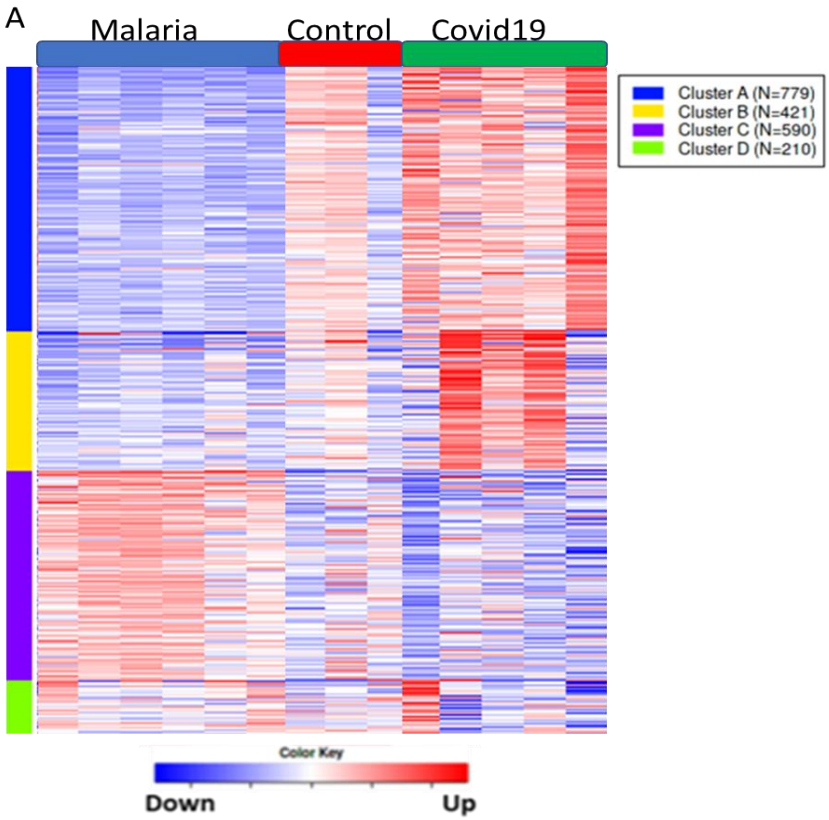
Figure 8**.** Gene ontology enrichment of overlapping gene (up-regulated and downregulated) between COVID-19 group and malaria group.

Table 3. The list and functions of shared genes associated with significant GO terms.

ENSEMBL_GENE_ID	Gene Name	Function	FC Malaria	FC Covid19	Ref.
ENSG00000086506	Hemoglobin subunit theta 1(HBQ1)	O_2 transport in human fetal erythroid tissue, unknown function in adult.	-4.25	4.15	[30]
ENSG00000100243	Cytochrome b5 reductase 3(CYB5R3)	Oxidation and reduction reactions, such as the reduction of methemoglobin to hemoglobin.	-2.26	2.27	[31]
ENSG00000206177	Hemoglobin subunit mu (HBM)	O_2 transport but not highly expressed and functional in adults.	-4.94	3.74	[32]
ENSG00000134321	Radical S-adenosyl methionine domain containing 2 (RSAD2)	Plays a major role in the cell antiviral state induced by type I and type II interferon. Promotes TLR7 and TLR9-dependent production of IFN-beta production in plasmacytoid dendritic cells (pDCs) by facilitating 'Lys-63'-linked ubiquitination of IRAK1 by TRAF6. Plays a role in CD4+ T-cells activation and differentiation. Facilitates T-cell receptor (TCR)-mediated GATA3 activation and optimal T-helper 2 (Th2) cytokine production by modulating NFKB1 and JUNB activities. Can inhibit secretion of soluble proteins.	4.16	2.09	[33]
ENSG00000161905	Arachidonate 15-lipoxygenase (ALOX15)	Functions to generate specific phospholipid (PL) oxidation products crucial for orchestrating the nonimmunogenic removal of apoptotic cells (ACs) as well as synthesizing precursor lipids required for production of specialized pro-resolving mediators (SPMs) that facilitate inflammation resolution	-5.48	-3.01	[34]
ENSG00000177666	Patatin like phospholipase domain containing 2 (PNPLA2)	Lipid metabolism,	-3.72	2.84	[35]

KMeans Clustering and Pathway analysis

To further confirm the differential gene expression between malaria and COVID-19 patients, Kmeans clustering was carried out. This analysis identified four main gene clusters (A-D) with cluster A and B genes having higher expression in COVID-19 patients as opposed to cluster C and D genes which had higher expression in malaria condition instead (Fig. 9). Pathway analysis revealed that cluster A genes were associated with significant enrichment of immune-related biological processes such as neutrophil and granulocyte activation and degranulation, cell-mediated immune response and vesicle-mediated transport. Interestingly, cluster B genes were associated with hematopeosis, gas and oxygen transport confirming the gene ontology analysis done with the common 62 genes differentially expressed in malaria and covid conditions. Genes in cluster C genes are likely to favour innate immune processes, inflammation and immune regulation while Cluster D genes could be playing important role in mitotic cell division including chromatin assembly and chromosome organization.



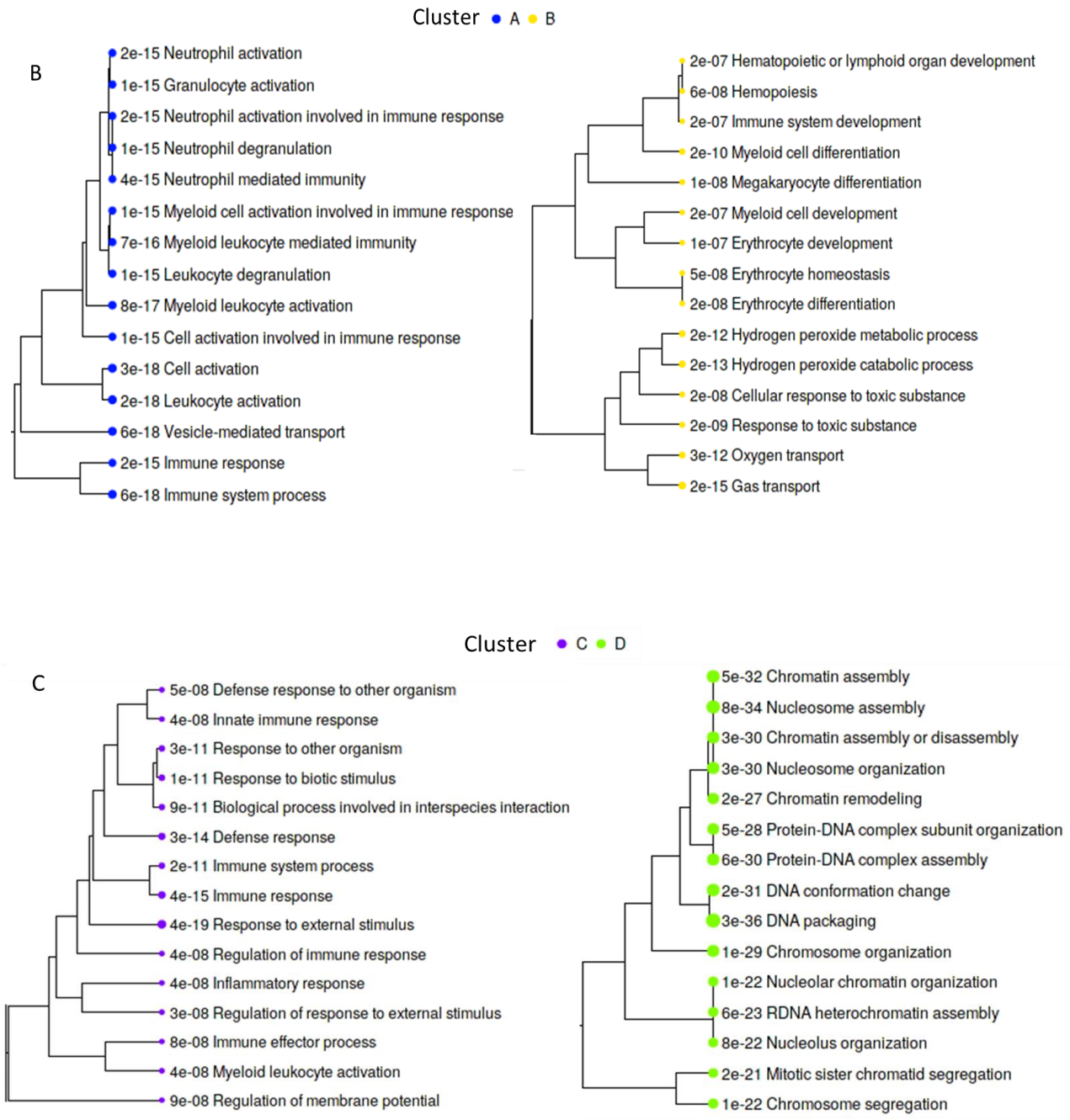


Figure 9. Kmeans clustering and associated pathway analysis. A) Heatmap showing clusters of genes differentially expressed in malaria, covid and healthy control. B and C) Pathway analysis, dendograms showing biological processes significantly enriched per gene cluster. This analysis was performed on the Integrated Differential Expression and Pathway analysis (iDEP-95) platform.

Discussion

In this study, we have used an RNA-seq approach to identify the transcriptional pattern of blood from infected children either with malaria or COVID-19 infection. This is useful because blood may be used to identify biological signatures between hosts that interact to pathogen and shorten the treatment [36]. As, we aim to determine whether peripheral blood transcriptomic profiles are associated with the overlapping symptoms either malaria or COVID-19 infection in febrile children. We initially explored the sample's diversity through PCA (principal component analysis) along a diagonal axis of 14% and PC2 of 28% (Fig. 2).

A total of 3,579 genes were differentially expressed in our RNA seq datasets. The comparison of COVID-19 samples relative to healthy samples we found that about 2495 genes differentially expressed such as SLC12A5-AS1, ENSG00000234998, GPS2P1, RCCD1-AS1 and RAP1GAP (Fig. 4). The high expression of SLC12A5-AS1 is a lncRNA gene associated with NK-cell in the lung and can respond to viral infections [37]. We also found the upregulation of ENSG00000234998 as a novel transcript belonging to the class of lncRNA, associated with CD64, which is the receptors of SARS-CoV-2 [38]. These receptors have been reported as biomarkers for early viral infections. We detected a significant expression of RAP1GAP (GTPase activating protein) which is involved in the viral invasion to change the epithelial morphogenesis and intercellular tight junction formation [39].

Alternatively, by comparing the malaria samples relative to the healthy samples we identified 1084 genes differentially expressed such as CTSL, and RSAD2 (Fig. 5). We also identified the upregulation of the CTSL gene, which is cathepsin L It belongs to the lysosomal enzyme cathepsin family and was previously reported as an enzyme participating in hemoglobin degradation in the food vacuole of *Plasmodium falciparum* trophozoites [40]. Interestingly by comparing both group samples COVID-19 and malaria (Fig. 6), the highly expressed gene was associated with biological processes such as in immune response (ENPP2, MAP3K14, and SPPL2B), in MyD88-independent toll-like receptor signaling pathway (TNIP3), and platelet aggregation (MYL9).

Kmeans clustering and gene ontology showed that inflammation, innate immunity and cell-mediated immunity to be enriched in children with both covid and malaria infections. The Corona virus and Plasmodium parasite are both intracellular pathogens which are mainly fought

by cell-mediated arm of the immune system, including the innate one [41-43]. Biological processes associated with hematopoesis, oxygen and gas transport were found significantly enriched in covid patients. Covid is known to be associated with cough and difficult breathing which may have as corollary, low oxygen levels in blood and tissue hence, the enhancement of hematopoetic processes and; gas and oxygen transport. Among the common 62 genes highly differentially expressed (Table 3), we identified hemoglobin subunits to be upregulated in children with covid. Hemoglobin is a key play in oxygen transport in the system. CYB5R3 a cytochrome b5 reductase 3, which transfers electrons from NADH to cytochrome b5 in red blood cells, was also found upregulated in children with COVID-19. In malaria, the mutation of CYB5R3 has been shown to be associated to the risk of developing severe anemia in children [44, 45]. The high expression of RSAD2 gene, an interferon-inducible gene, reported to be involved in the innate immune response against viruses [46] was upregulated in both malaria and covid conditions. This could be due to the intracellular nature of both pathogens. We identified the expression of ALOX15 (Arachidonate 15-lipoxygenase) gene which is involved in cytokine signaling pathway, inflammation mediated by chemokine in the presence of *Plasmodium falciparum* [47]. The upregulation of the PNPLA2 gene (patatin-like phospholipase domain containing 2) encodes ATGL (adipose triglyceride lipase). The enzyme is involved in triglyceride hydrolysis and lipid-droplet homeostasis which is linked to severe malaria in humans [48]. We noticed that among the shared genes, the hemoglobin complexes and lipid mediators are highly expressed between the COVID-19 samples and malaria samples (Fig. 7). Thus, we suggest that CYB5R3, RSAD2, ALOX15, HBQ1, HBM and PNPLA2 might be the protein-coding genes associated with malaria and COVID-19 diseases in children (Fig. 8 and Fig. 9).

Conclusion

We scrutinized the transcriptional responses associated between children infected with malaria and children with COVID-19 infection. Among 3579 differentially expressed profiled in both infectious diseases, the overlap substantially showed 62 genes associated with malaria and COVID-19 infection, which may either indicate linkage or predict the coinfection status. These common expressed genes (CYB5R3, RSAD2, ALOX15, HBQ1, HBM and PNPLA2) are potential biomarkers for differential diagnosis of children with malaria and COVID-19

coinfection. We recommended further research with the coinfection samples, particularly in regions with high malaria prevalence.

Limitations of the study and suggestions for further work

The predicted role of different genes (upregulated and downregulated) needs to be further investigated either in children or adults with data generated from scratch instead of using RNA seq data from public repositories which are produced in different high-throughput sequencing methods for diverse scientific goals. Another limitation was the missing data with coinfection of malaria with COVID-19 in children. This would have enabled us to generate distinctive transcriptomic profiles of responses in case both infections were present within one host.

Data acquisition and software availability

The RNA-seq data were collected from public databases to study children infected with malaria and COVID-19 infection. Only datasets with paired end reads were used in this study. The study examined the transcription profile responses of children with *Plasmodium falciparum* and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The datasets used for children with malaria are available on European Nucleotide Archive (ENA) through the study accession number PRJEB33892. The datasets used for children with COVID-19 infection are available on Gene Expression Omnibus (GEO) through accession number GSE178388. The reference genome used for indexing and mapping was downloaded from GENCODE. The annotation reference genome used for counting reads was downloaded from ENSEMBL genome database. The data files for this study, as well as the source code, Software tools, databases and command line used for preprocessing data under Linux and R scripts for downstream analysis, are available on Github. (https://github.com/omicscodeathon/RNA-seq-Malaria).

Author contributions: Conceptualization, N.L; J.A.K.O; Formal analysis, N.L; J.A.K.O; Investigation, N.L; J.A. K.O; W.M. R; K.M; U.D; P.K. Methodology, N.L; J.A.K.O. Resources, O.I.A; N.L. Data Curation, N.L; J.A.K.O. Project Administration, O.I.A. Resources, O.I.A; N.L. Supervision, O.I.A; N.L. Writing and original Draft Preparation, W.M.R; U.D; N.L; Writing–Review & Editing, N.L; J.A.K.O; W.M.R; K.M; U.D; P.K; O.I.A. Funding Acquisition, O.I.A; A.D. All authors have read and agreed to the published version of the manuscript.

Competing interests: No competing interests were disclosed.

Acknowledgements

This study was funded by the National Institutes of Health Office of Data Science Strategy (ODSS), in partnership with the African Society for Bioinformatics and Computational Biology (ASBCB). We are grateful for their support to provide High performance computing (HPC) in the cloud for RNA seq data analysis. Centre for Research in Infectious Diseases (CRID), for the onsite technical support.

References

1. Hussein, M.I.H., et al., *Malaria and COVID-19: unmasking their ties.* Malaria Journal, 2020. **19**(1): p. 1-10.
2. Liu, Q., et al., *Trends of the global, regional and national incidence of malaria in 204 countries from 1990 to 2019 and implications for malaria prevention.* J Travel Med, 2021. **28**(5).
3. WHO, *World malaria report 2021*. 2021.
4. Gaston, R.T. and S. Ramroop, *Prevalence of and factors associated with malaria in children under five years of age in Malawi, using malaria indicator survey data.* Heliyon, 2020. **6**(5): p. e03946.
5. Kamau, A., et al., *Malaria infection, disease and mortality among children and adults on the coast of Kenya.* Malar J, 2020. **19**(1): p. 210.
6. Mutombo, A.M., et al., *Severe malaria and death risk factors among children under 5 years at Jason Sendwe Hospital in Democratic Republic of Congo.* Pan Afr Med J, 2018. **29**: p. 184.
7. Kumar, S., D. Dwivedi, and M. Chopra, *COVID-19 Outbreak: Neurological Manifestations Beyond Cough and Fever.* CNS Neurol Disord Drug Targets, 2021. **20**(2): p. 101-104.
8. Liu, W.Y., et al., *Insights from the comparisons of SARS-CoV and COVID-19 outbreaks: The evidence-based experience of epidemic prevention in China.* Medicine (Baltimore), 2021. **100**(6): p. e24650.

9. Sharma, H., S. Singh, and S. Pathak, *Pathogenesis of COVID-19, Disease Outbreak: A Review.* Current Pharmaceutical Biotechnology, 2021.
10. Aborode, A.T., et al., *Fighting COVID-19 at the Expense of Malaria in Africa: The Consequences and Policy Options.* The American Journal of Tropical Medicine and Hygiene, 2021. **104**(1): p. 26.
11. Ajayi, I.O., O.O. Ajumobi, and C. Falade, *Malaria and COVID-19: commonalities, intersections and implications for sustaining malaria control.* The Pan African Medical Journal, 2020. **37**(Suppl 1).
12. Pusparani, A., J. Henrina, and A. Cahyadi, *Co-infection of COVID-19 and recurrent malaria.* The Journal of Infection in Developing Countries, 2021. **15**(05): p. 625-629.
13. Bhardwaj, A.R., et al., *Northern Blotting Technique for Detection and Expression Analysis of mRNAs and Small RNAs*, in *RNA Abundance Analysis*. 2021, Springer. p. 155-183.
14. Várallyay, E., J. Burgyán, and Z. Havelda, *MicroRNA detection by northern blotting using locked nucleic acid probes.* Nature protocols, 2008. **3**(2): p. 190-196.
15. Dallas, P.B., et al., *Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR–how well do they correlate?* BMC genomics, 2005. **6**(1): p. 1-10.
16. Jozefczuk, J. and J. Adjaye, *Quantitative real-time PCR-based analysis of gene expression.* Methods Enzymol, 2011. **500**: p. 99-109.
17. Schneider, S., et al., *Quantitative determination of p16 gene expression by RT-PCR.* Methods Mol Biol, 2004. **281**: p. 91-103.
18. Mantione, K.J., et al., *Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq.* Med Sci Monit Basic Res, 2014. **20**: p. 138-42.
19. Wang, D. and A.L. Karamyshev, *Next generation sequencing (NGS) application in multiparameter gene expression analysis*, in *Molecular Toxicology Protocols*. 2020, Springer. p. 17-34.
20. Creecy, J.P. and T. Conway, *Quantitative bacterial transcriptomics with RNA-seq.* Curr Opin Microbiol, 2015. **23**: p. 133-40.
21. Pei, X.T., X.W. Jiao, and Z.G. Ping, *[Visualization of transcriptome data].* Zhonghua Yu Fang Yi Xue Za Zhi, 2020. **54**(5): p. 586-592.

22. Ståhl, P.L., et al., *Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.* Science, 2016. **353**(6294): p. 78-82.
23. Chang, S., et al., *Pathway-based analysis for genome-wide association study data of bipolar disorder provides new insights for genetic study.* Protein & cell, 2015. **6**(12): p. 912-915.
24. He, K., et al., *Gene set enrichment analysis of pathways and transcription factors associated with diabetic retinopathy using a microarray dataset.* International Journal of Molecular Medicine, 2015. **36**(1): p. 103-112.
25. Inoue, K., *Feedback enrichment analysis for transcription factor-target genes in signaling pathways.* Biosystems, 2020. **198**: p. 104262.
26. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
27. WHO, *World malaria report. 2016.* 2016.
28. Pediatrics, A.A.o., *Children and COVID-19: State-Level Data Report.* 2022.
29. Wilairatana, P., et al., *Prevalence and characteristics of malaria among COVID-19 individuals: A systematic review, meta-analysis, and analysis of case reports.* PLoS Negl Trop Dis, 2021. **15**(10): p. e0009766.
30. Spritz, R.A., et al., *Complete nucleotide sequence of the human δ-globin gene.* Cell, 1980. **21**(3): p. 639-646.
31. Shunji, T., et al., *The organization and the complete nucleotide sequence of the human NADH-cytochrome b5 reductase gene.* Gene, 1989. **80**(2): p. 353-361.
32. Goh, S.-H., et al., *A newly discovered human α-globin gene.* Blood, 2005. **106**(4): p. 1466-1472.
33. Yuan, Y., et al., *Targeting UBE4A revives viperin protein in epithelium to enhance host antiviral defense.* Molecular cell, 2020. **77**(4): p. 734-747. e7.
34. Snodgrass, R.G. and B. Brüne, *Regulation and functions of 15-lipoxygenases in human macrophages.* Frontiers in pharmacology, 2019. **10**: p. 719.
35. De la Cruz-Color, L., et al., *Association of the PNPLA2, SCD1 and leptin expression with fat distribution in liver and adipose tissue from obese subjects.* Experimental and Clinical Endocrinology & Diabetes, 2020. **128**(11): p. 715-722.

36. Ramilo, O., et al., *Gene expression patterns in blood leukocytes discriminate patients with acute infections.* Blood, 2007. **109**(5): p. 2066-2077.

37. Maucourant, C., et al., *Natural killer cell immunotypes related to COVID-19 disease severity.* Science immunology, 2020. **5**(50): p. eabd6832.

38. Karawajczyk, M., et al., *High expression of neutrophil and monocyte CD64 with simultaneous lack of upregulation of adhesion receptors CD11b, CD162, CD15, CD65 on neutrophils in severe COVID-19.* Therapeutic advances in infectious disease, 2021. **8**: p. 20499361211034065.

39. Tian, W., et al., *Immune suppression in the early stage of COVID-19 disease.* Nature communications, 2020. **11**(1): p. 1-8.

40. Rosenthal, P.J., et al., *A malarial cysteine proteinase is necessary for hemoglobin degradation by Plasmodium falciparum.* The Journal of clinical investigation, 1988. **82**(5): p. 1560-1566.

41. *Immunity to Infection*, in *Primer to the Immune Response (Second Edition)*, T.W. Mak, M.E. Saunders, and B.D. Jett, Editors. 2014, Academic Cell: Boston. p. 295-332.

42. Varadé, J., S. Magadán, and Á. González-Fernández, *Human immunology and immunotherapy: main achievements and challenges.* Cellular & Molecular Immunology, 2021. **18**(4): p. 805-828.

43. Kengne-Ouafo, J.A., et al., *Immune Responses to the Sexual Stages of Plasmodium falciparum Parasites.* Frontiers in Immunology, 2019. **10**.

44. Gordeuk, V.R., et al., *Cytochrome B5 Reductase T116S Polymorphism Is Associated with Decreased Risk of Severe Anemia Among Zambian Children with Malaria.* Blood, 2010. **116**(21): p. 4233.

45. Shah, B.N., et al., *Host Genetic Factors in Glucose-6-Phosphate Dehydrogenase and Cytochrome B5 Reductase 3 Affect the Susceptibility of Developing Severe Malarial Anemia.* Blood, 2016. **128**(22): p. 2438.

46. Jang, J.-S., et al., *Rsad2 is necessary for mouse dendritic cell maturation via the IRF7-mediated signaling pathway.* Cell death & disease, 2018. **9**(8): p. 1-11.

47. Schwarzer, E., et al., *Malaria-parasitized erythrocytes and hemozoin nonenzymatically generate large amounts of hydroxy fatty acids that inhibit monocyte functions.* Blood, The Journal of the American Society of Hematology, 2003. **101**(2): p. 722-728.

48. Tripathi, J., et al., *A Novel Chemically Differentiated Mouse Embryonic Stem Cell-Based Model to Study Liver Stages of Plasmodium berghei.* Stem cell reports, 2020. **14**(6): p. 1123-1134.