Metabolomics of *Plasmodium knowlesi* malaria patients in Malaysia

499 Progress Report after 1 Semester

Abstract

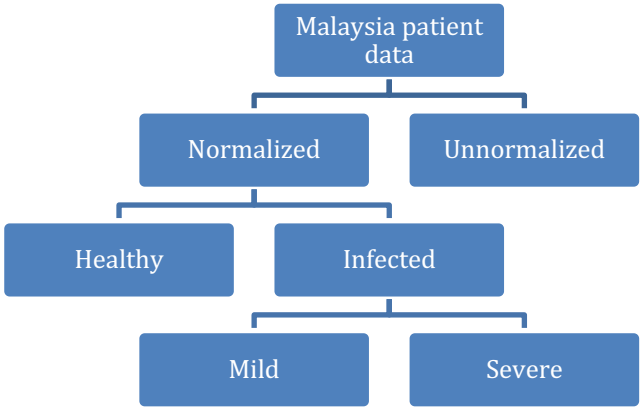
Malaria is a disease affecting hundreds of millions of people every year, in about 100 countries. While *P. falciparum* and *P. vivax* are the more common species of the parasite causing malaria in humans, there has been a rise in severe and fatal cases diagnosed in recent years as *P. knowlesi* zoonotic infections. To better understand the effects of the infection as well as host-pathogen interactions, metabolomics is used to evaluate the biological processes changed in the host in the course of the infection. In this study, we are analyzing patient clinical and metabolomics data generated from plasma samples obtained from cases of human *P. knowlesi* malaria in Malaysia. We aim to identify the metabolites that change when there is an infection, and possibly detect metabolites and their corresponding pathways that relate to disease severity.

Introduction

Every year there are millions of cases of malaria, and according to the World Malaria Report published by the World Health Organization (WHO), there were 212 million new cases of malaria in 2015. The disease is caused by different species of the *Plasmodium* parasite. While there is a large focus on *P. falciparum* and *P. vivax*, which according to the WHO are the greatest threat, three other species still represent a public health concern. The species *P. knowlesi*, often found in macaques, was originally believed to rarely infect humans. However, a study by Singh et al. in 2004 [1] identified this *P. knowlesi* malaria as a significant zoonotic disease in Malaysia, and Cox-Singh et

al. then reported its widespread prevalence in the country [2]. These studies also helped *P. knowlesi* to gain recognition as the 5th human malaria parasite species. The property of being in both macaques, a great model organism, and humans make it ideal for study. While for the most part malaria caused by *P. knowlesi* can be treated without complication, a retrospective study found that 39% of *P. knowlesi* cases were severe, and of these, 27% were fatal [3]. It is therefore of great importance to gain an improved understanding of what distinguishes cases that are severe from those that are not.

Through advancements in biological technologies we can now get a deeper understanding of what is happening within the body. One of these advancements is in the field of metabolomics. Metabolomics uses mass spectrometry to generate the unique chemical fingerprints of metabolites found within a sample. This information helps to identify and characterize ongoing biological processes and understand how they change under different conditions. Surowiec et al. found that metabolite signature profiling could aid in malaria diagnostics and prognostics [4]. A study by our group, Gardinassi et al., previously demonstrated a relationship between parasitemia levels and certain metabolite abundance [5]. Here, by using metabolomics, this study hopes to distinguish what metabolites are involved in *P. knowlesi* infection of humans, and which if any metabolites have a relationship with disease severity. This study will investigate the



importance of normalizing metabolomics data, and then explore the differences between healthy and infected individuals. Finally, this study aims to use metabolomics to find metabolic changes in relation to disease severity (Fig. 1).

Methods

Samples

Samples were collected by Professor Balbir Singh and his group based at the Malaria Research Centre at the Universiti Malaysia Sarawak in Kuching, Sarawak, Malaysia. Patients were recruited from the district hospital in Kapit and surrounding clinics where they often presented with symptoms. As Prof. Balbir Singh's team is based in Kuching, they needed to travel on a regular basis to Kapit to obtain samples.

All patients diagnosed with malaria (detectable parasites by blood smear) were treated according to the standard of care for Malaysia. All infected patients in the study were hospitalized until they were smear-negative in at least two consecutive peripheral blood smears, with public health follow up after release. To provide appropriate negative controls for this study, samples from age and gender-matched individuals were taken from healthy participants from the area surrounding Kapit and within Kapit. Consent for all individuals was obtained according to a local ethical board approved protocols, and a thumbprint used as a signature for participants who were illiterate.

Plasma samples were stripped of identifying patient information and were shipped on dry ice to Emory University. Emory University's Institutional Review Board (IRB) reviewed protocols, local approvals, and consent forms for this study and approved an IRB exemption based on the de-identification of samples and data. Plasma from a total of 143 infected patients and 111 uninfected patients were analyzed in this study.

Untargeted Metabolomics

With the assistance of the Emory Clinical Biomarkers Laboratory, directed by Drs. Dean P. Jones, Shuzhao Li, and Karan Uppal, plasma samples were run on by mass spectrometry and data was processed for untargeted metabolomics analysis. This work was completed by Ms. ViLinh Tran and Dr. Luiz Gardinassi. The measurement of metabolites within the plasma samples was done using liquid chromatography-mass spectrometry (LC-MS) and high-resolution metabolomics (HRM) workflows. In brief, the samples are analyzed on a High Field QExactive machine in triplicate, and the data generated undergoes peak detecting noise removal and alignment, done using the programs apLCMS and xcms with xMSanalyzer. This creates a data table with mass-to-charge ratio (*m/z*), retention time and intensity for each *m/z* across all samples. This table was then adjusted for batch-effect correction using the program ComBat.

Statistical Analysis

In the present study, ComBat-corrected data from the hilic column (positive mode) was used for all analyses. This results table included the intensity data for each unique metabolite feature in each sample. The ComBat-corrected file had to be cleaned and prepped for analysis. Using the metadata for the experiment, a class labels file had to be generated. This class labels file provides a mapping of all sample barcodes to their designated comparison group for analysis. Preparation of this file also involved removing all non-sample values and information guides from the Combat corrected file and renaming all the samples to have the correct barcode in the same format as the class labels file (Appendix 1).

Next, both the cleaned results table and the class labels file was run through xmsPANDA (R package courtesy of Dr. Uppal and used in [6]), a program designed for performing statistical analyses of HRM data. When comparing the effect of normalization, the values for the percent maximum missing both the group and overall were set to NA as opposed to 0.8 and 0.5 respectively. For all comparisons performed in this study, the false discovery rate (FDR) for the normalized run was 0.2 (Appendix 2).

Next the data was run through *mummichog* v1.0.5 [7]. This was to predict the pathways affected by the metabolites. This software combines pathway analysis and metabolite identification, this allows for a fuller picture, as opposed to finding the significant metabolites independently and then using the smaller dataset to find significant pathways. To perform this analysis, another table had to be created using a specific output from xmsPANDA (limma sorted values). The test statistic used for this analysis was the max fold change.

To annotate the *m/z* features found, xMSannotator was used[8]. It provided annotations from two databases, Human Metabolome Database (HMDB) and LIPID MAPS. Another script was created to find and match all annotations found for a specific *m/z* and all pathways. This took two part due to the fact there could be multiple annotations for a single feature (Appendix 3-4). The resulting table contains the *m/z* in limma significant order with annotations from both databases, including the database ID, the name, the match score, the match category and the adduct and any of the pathways found through mummichog that used that metabolite (Appendix 1).

Results

First the data were run with and without normalization. Without normalization, the differences in the metabolites between uninfected and infected is not readily identifiable.

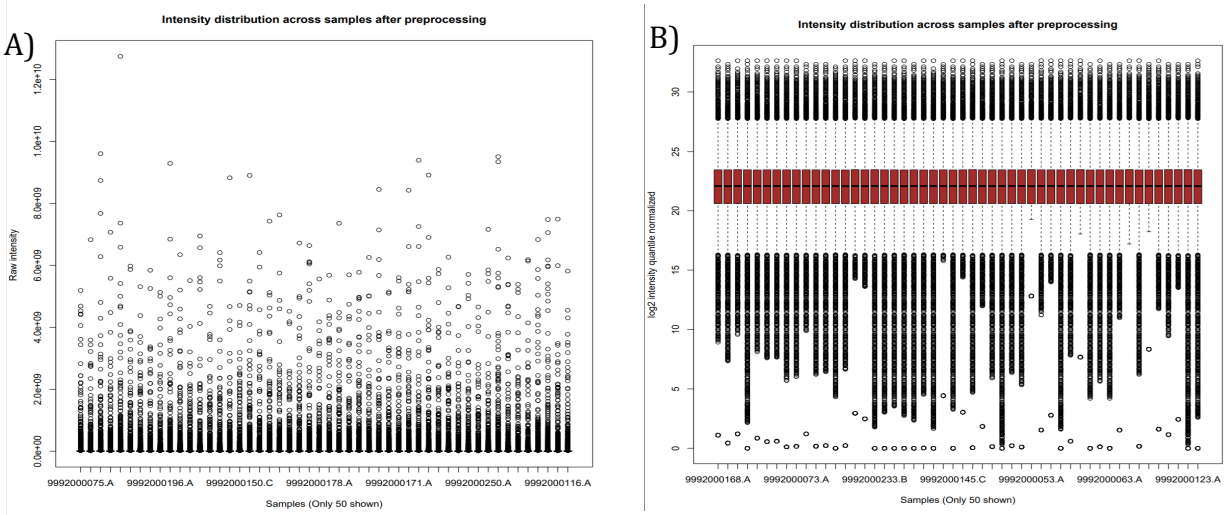


Figure 2**. Intensity distribution across samples after processing**. A) is the intensity distribution of the unnormalized data. The reported intensities are simply the raw intensities recorded B) is the intensity distribution for the normalized data. Th

All we can compare is the raw intensity of the metabolite data across samples, introducing a high degree of noise in the data. With the normalized data, we have an improved ability to compare two different groups (Fig. 2).

After performing normalization, we compared the plasma metabolome of healthy and malaria-infected populations globally. Here in the global principal components analysis (PCA), we can see while there is an overlap between uninfected and infected, there is a separate cluster of infected cases that cluster away from the group in the bottom right of the plot (green dots, Fig. 3).

After running a limma analysis comparison the two groups, we found 924 significant features, which can be seen on Figure 4a in a hierarchical clustering analysis. When the PCA is run with just significant features a stronger distinction between the clusters can be seen (Fig. 4b)

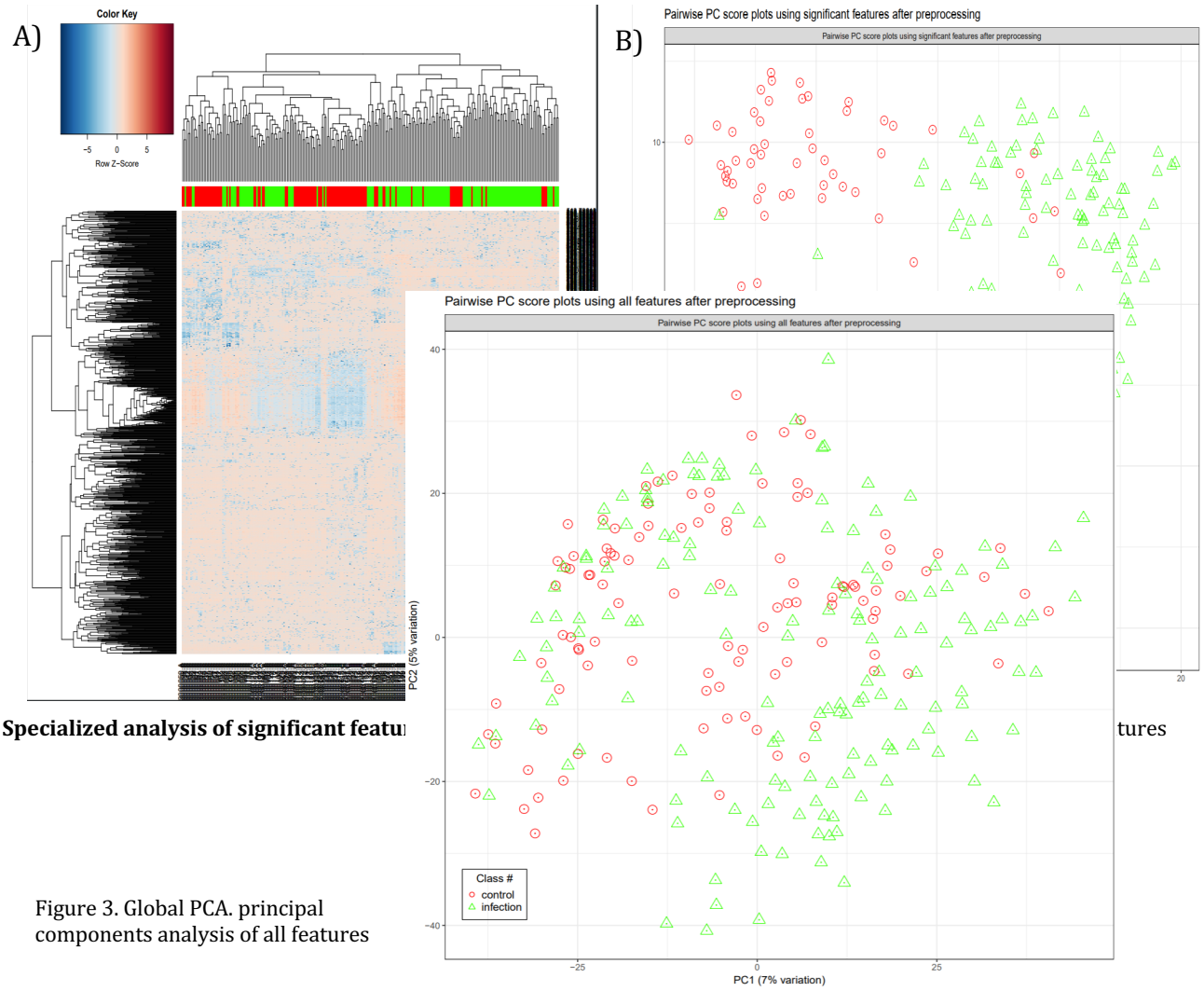


Figure 4. **Specialized analysis of significant featu** tures

Figure 3. Global PCA. principal components analysis of all features

After running pathway analysis and removing pathways with less the 3 metabolites there were 13 significant pathways found to have significantly different metabolic activity in uninfected versus infected individuals (Fig. 5). These pathways related to a range of metabolic functions, including amino acids (e.g. lysine, tryptophan), lipids (e.g. phosphatidylinositol), vitamins (e.g. Vitamin B3), nucleic acids (e.g. purine), and sugars (e.g. hexose).

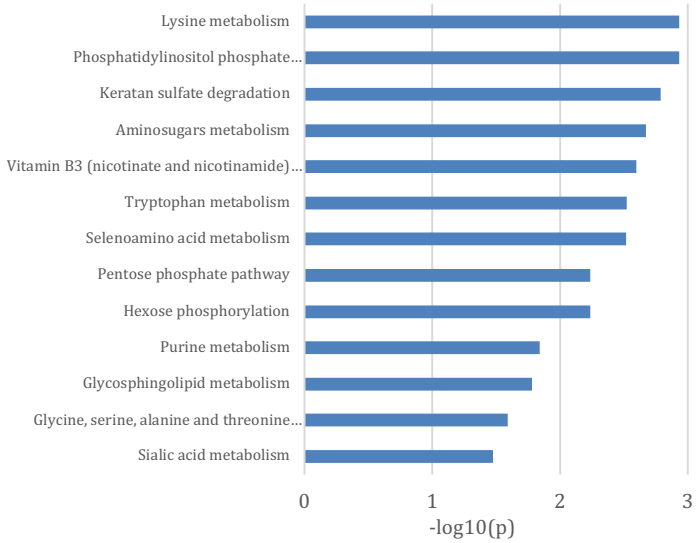


Figure 5. **Significant pathways.** List of most significant pathways with 3 or more metabolites and their corresponding p values

Discussion

Several insights can be gleaned from the analyses in the present study. First, when studying metabolomics, it is important to normalize your data. This is necessary to reduce noise that may be preventing you from seeing any sort of results. This noise exists because biology and human life, in addition to laboratory processes steps, is rarely perfect and there are many situational factors that need to be controlled for.

Both in the global PCA, HCA, and the PCA with significant features there is a clear difference in healthy and infected individuals. Some of the important metabolites and pathways affected are different amino acids, like leucine and arginine, biliverdin, glycerophosphocholine, and different LysoPCs These are similar to those found in the literature[9][5][6]. The Gupta et al. study similarly found significant enrichment of heme

metabolites as we find hemolysis related metabolites, like biliverdin, indicating blood cell lysis and possibly a malaria specific response. Like the Uppal et al. study, we found changes in levels of glycerophosphocholines which relate to more general infection and immune pathways, but could be used to target a more general response. One thing of note in the global analysis is that not all the patients infected clustered out. What made these patients different, could it be disease severity? The next part of this study hopes to look at this.

References

1. Singh, B., et al., *A large focus of naturally acquired Plasmodium knowlesi infections in human beings.* Lancet, 2004. **363**(9414): p. 1017-24.
2. Cox-Singh, J., et al., *Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening.* Clin Infect Dis, 2008. **46**(2): p. 165-71.
3. William, T., et al., *Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia.* Emerg Infect Dis, 2011. **17**(7): p. 1248-55.
4. Surowiec, I., et al., *Metabolic Signature Profiling as a Diagnostic and Prognostic Tool in Pediatric Plasmodium falciparum Malaria.* Open Forum Infect Dis, 2015. **2**(2): p. ofv062.
5. Gardinassi, L.G., et al., *Metabolome-wide association study of peripheral parasitemia in Plasmodium vivax malaria.* Int J Med Microbiol, 2017.
6. Uppal, K., et al., *Plasma metabolomics reveals membrane lipids, aspartate/asparagine and nucleotide metabolism pathway differences associated with chloroquine resistance in Plasmodium vivax malaria.* PLoS One, 2017. **12**(8): p. e0182819.
7. Li, S., et al., *Predicting network activity from high throughput metabolomics.* PLoS Comput Biol, 2013. **9**(7): p. e1003123.
8. Uppal, K., et al., *xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data.* BMC Bioinformatics, 2013. **14**: p. 15.
9. Gupta, S., et al., *Extensive alterations of blood metabolites in pediatric cerebral malaria.* PLoS One, 2017. **12**(4): p. e0175686.