

Article

# Individualized Transcriptional Resolution of Complicated Malaria in a Colombian Study

Mónica L. Rojas-Peña <sup>1,†</sup>, Meixue Duan <sup>1,†</sup> , Dalia Arafat <sup>1</sup>, Lina Rengifo <sup>2</sup>, Socrates Herrera <sup>2</sup> , Myriam Arévalo-Herrera <sup>2,‡</sup>  and Greg Gibson <sup>1,\*,‡</sup>

<sup>1</sup> Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA 30332, USA; monicarojasp@gmail.com (M.L.R.-P.); mduan31@gatech.edu (M.D.); darafat.gt@gmail.com (D.A.)

<sup>2</sup> CAUCASECO, Cali, Colombia; lrengifo@inmuno.org (L.R.); sherrera@inmuno.org (S.H.); marevalo@inmuno.org (M.A.-H.)

\* Correspondence: greg.gibson@biology.gatech.edu; Tel.: +1-404-385-2343

† These authors contributed equally as first authors.

‡ These authors contributed equally as senior authors.

Received: 14 August 2018; Accepted: 10 September 2018; Published: 14 September 2018



**Abstract:** To evaluate whether recovery from complicated malaria follows a common trajectory in terms of immunological mechanism or, rather, is highly individualized for each patient, we performed longitudinal gene expression profiling of whole blood. RNA sequencing (RNAseq) was performed on blood samples obtained from eight patients on four consecutive days between hospital admission and discharge. Six patients were infected with *Plasmodium falciparum*, and two with *Plasmodium vivax*; one patient was a pregnant woman infected with *P. falciparum*, who was hospitalized for several weeks. The characterization of blood transcript modules (BTM) and blood informative transcripts (BIT) revealed that patients' responses showed little commonality, being dominated by the balance of gene activity relating to lymphocyte function, inflammation, and interferon responses specific to each patient. Only weak correlations with specific complicated malaria symptoms such as jaundice, thrombocytopenia, or anemia were observed. The differential expression of individual genes, including transcripts derived from the human leukocyte antigen (HLA) complex, generally reflected differences in the underlying immune processes. Although the results of this pilot study do not point to any single process that might provide a target for complicated malaria treatment or prevention or personalized medical strategies, larger patient series and more extensive blood sampling may allow the classification of patients according to their type of response in order to develop novel therapeutic approaches.

**Keywords:** transcriptomics; RNAseq; gene expression; longitudinal profiling; *Plasmodium*

## 1. Introduction

Malaria is the most prevalent parasitic disease worldwide with over 200 million clinical cases and about 445,000 deaths reported globally every year [1]. While most clinical cases are characterized by classical signs and symptoms such as fever, chills, headache, and malaise that rapidly disappear upon early diagnosis and prompt treatment, failure to recover quickly can lead to a more severe and complicated disease that usually requires hospitalized management [2–5]. In these cases, multiple organs can suffer irreversible damage, leading to multi-systemic failure and eventually to death [6]. In Colombia, a low-malaria-transmission country of Latin America, the most common complications include vascular obstruction, inflammatory processes, and damage to as well as cell death in various organs. Symptoms include jaundice, severe thrombocytopenia, liver and kidney injury, and acute anemia, with cerebral infection affecting a minority of patients [7–10]. It is not precisely known what

mechanisms are involved in the progression to severe disease or recovery from infection, although it has been shown that multiple arms of the immune system are engaged [11–13]. In *Plasmodium vivax*, a significant modification of peripheral blood gene expression is first observed as parasites begin to appear in the blood stream 12 to 14 days after initial infection [14]. It involves both myeloid and lymphoid cell types, as well as inflammatory and interferon responses. No clear pattern of Th1, Th2, or Th17 engagement has emerged, and, although high parasitemia is the major factor influencing the extent of transcriptional divergence [15], these results have not yet produced a clear picture of the chain of immune responses that control the infection. Complement activation is critical for the removal of *Plasmodium*-infected red blood cells [16,17], but very little is known about how variation among individuals, and indeed within individuals across infections, results in responses that range from asymptomatic infection to severe and complicated disease leading to mortality.

Recently, RNA sequencing (RNAseq) has emerged as the preferred tool for gene expression profiling, since it provides an unbiased view of the abundance of both genes and exons [18,19]. Compared with microarray analysis, technical artefacts seem to be less prevalent, but its application to complex tissues still suffers from the problem that cell type-specific differential gene expression can be imputed only indirectly. Studies of semi-purified immune cell subtypes certainly increase resolution but are more expensive, necessarily biased towards cells of a priori interest, and there is always the concern that purification modifies gene expression. Consequently, bioinformatics approaches are evolving that allow the deconvolution of the abundance of cell types as well as the inference of the gene activity pathways that are most strongly affected in a disease or treatment [20]. In the near future, single-cell RNAseq [21] will allow the comprehensive comparison of gene expression variation among all cell types of a mixture, potentially obviating the need for bioinformatic deconvolution.

By applying a machine learning approach to over 30,000 peripheral blood profiles, Li et al. [22] identified about 300 “blood transcript modules” (BTM) enriched for gene activity, known targets of key regulatory factors, or specific biochemical pathways in dozens of immune cell types. The derivation of the first principal component (PC1) of each BTM provides a summary of the major gene ontologies that characterize individual samples. Similarly, Chaussabel and colleagues [23] found 28 modules that are conserved across datasets but characteristic of 11 different immune diseases. We reduced these to nine super-modules, or “blood informative transcript” (BIT) axes [24], after recognizing that there is a very high correlation between the PC1 scores in multiple cohorts of healthy people. Subsequent analysis of dozens of datasets, including the results presented below, confirmed that BTM cluster into no more than a dozen groups, each of which covaries with one of the BITs. Furthermore, we show here that, by utilizing single-cell RNAseq profiles, we can now also characterize T-lymphocyte activity as helper, cytotoxic, or natural killer (NK) cell-related, from whole RNA profiles.

The objective of this study was to pilot the use of longitudinal RNAseq profiling during treatment and recovery from complicated malaria to characterize the immunological pathways engaged in the resolution of the infection. Whole blood samples were obtained upon hospitalization and three more times over the next four days. Most patients were discharged on days 4 or 5, but one pregnant woman (CM02) remained sick and hospitalized for 49 days; she was followed for almost a month and suffered a malaria relapse 15 days after first hospital discharge. Specifically, in this study, we examined: (i) whether there is a distinct signature of complicated malaria upon hospitalization of a patient, (ii) whether there is a common pattern of recovery characterized by the engagement of particular BTM or BIT, and (iii) whether individualized profiles associate with particular symptoms of complicated malaria.

## 2. Materials and Methods

### 2.1. Study Design and Subject Recruitment

Regardless of age, sex, or ethnic group, eight malaria patients admitted to a third-level hospital with confirmed *Plasmodium falciparum* ( $n = 6$ ) or *P. vivax* ( $n = 2$ ) were enrolled. The three study

recruitment sites are located on the Pacific coast region of Colombia (Choco, Valle del Cauca, and Nariño Departments) and were sampled over the course of complicated malaria episodes in 2015. Clinical complications attributable to malaria infection were recorded, and complicated malaria was defined on the basis of existing guidelines of the World Health Organization (WHO) and the Colombian Ministry of Health [25,26], requiring clinical severity and one or more clinical conditions such as anaemia (Hb < 7 g/dL in adults and children), renal dysfunction (serum creatinine > 1.5 mg/dL), severe thrombocytopenia ( $\leq 20,000$  platelets/ $\mu\text{L}$ ), and hyperparasitaemia (>50,000 parasites/ $\mu\text{L}$ ). The patients were medically monitored daily until discharge upon recovery from their clinical complications. All patients provided written informed consent for participation before enrolment, and the study protocol was approved by the Institutional Review Boards (IRB) for the Malaria Vaccine and Drug Development Center (CECIV). In addition, the Georgia Institute of Technology IRB provided subsequent approval for genomic analyses.

## 2.2. Blood Sample Collection

All study participants were bled by arm venipuncture at four time points during hospitalization, with the exception of the pregnant woman (CM02), who had two additional bleedings during a second hospitalization, because of a parasite relapse a few weeks after her first hospital discharge. Blood (~1 mL) was collected into Tempus tubes (Thermo-Fisher Scientific, Waltham, MA, USA), which preserve whole blood RNA indefinitely at 4 °C [27]. mRNA was extracted using Tempus Blood RNA tube isolation kits, following the protocol provided by the manufacturer, and the RNA quality was determined using the Agilent Bioanalyzer RNA Integrity score (RIN) (Agilent, Santa Clara, CA, USA) [28].

## 2.3. Whole Transcriptome Sequencing (RNAseq) and Preprocessing of the Data

Library preparation for RNAseq was performed using the Illumina TruSeq Stranded mRNA Sample Low-Throughput (LT) preparation protocol [29]. Short read sequencing was performed in rapid run mode with 10 samples per lane on an Illumina HiSeq 2100 (Illumina, San Diego, CA, USA) at Georgia Tech, generating 100 bp single-end libraries with an average of ~30 million single-end reads per sample.

The Tophat2 package [30] was used to align the reads to the human genome hg19/GRCh37, followed by UCSC reference annotation and HTSeq [31] to estimate the transcript abundance. The expression of known transcripts was calculated as counts per million (cpm) and filtered by cpm equal or greater than 1 in at least three of the samples; a total of 13,958 genes passed this criterion and were scaled on the log<sub>2</sub> scale by trimmed mean of the m-values (TMM) normalization in edgeR [32].

## 2.4. Analysis of RNAseq Data

### 2.4.1. Principal Variance Component Analysis

We performed principal variance component analysis (PVCA) to assess the overall contributions of patient, time-point, gender, and *Plasmodium* species to the transcriptional variance using the PVCA package in R v3.3.1. The percentile value for the minimum amount of the variance explained by the PC was set to 75%, and the average contributions of each covariate were computed as the average contribution to each PC weighted by the variance explained by the PC. PC1 mainly captures inter-individual variation, whereas PC2 and PC3 derive from a mixture of sources, with a strong influence of the parasite species.

### 2.4.2. Modular Analysis

Blood transcript modules [22] and BIT [24] analyses were performed by computing the PC1 of all the genes in each module or the top genes most closely correlated with each BIT axis. Standard analysis of variance (ANOVA) was then used to compare these scores among treatment classes, and

two-way hierarchical clustering was computed using Ward's method in JMP Genomics version 8 (SAS Institute, Cary, NC, USA). The Module and Axis scores for each individual at each time point are listed in Supplementary Table S1.

In order to refine the contributions of NK cells, T-helper cells, and T-cytotoxic cells, we identified the top 10 genes differentially expressed between these three cell types in the single-cell RNAseq dataset reported in reference [21]. The first principal component for each of these three sets is, as expected, highly correlated with Axis T, but the residuals of the regression of each PC1 on Axis T provide a measure of inflation or repression of the gene activity in the three cell types.

#### 2.4.3. Differentially Expressed Gene Calling and Gene Ontology Analysis

One-way ANOVA analysis between each patient and the remaining patients was conducted to detect patient-specific differentially expressed genes, using the following arbitrary  $p$ -value thresholds to identify approximately 50 to 150 genes per individual:  $p < 10^{-4}$ : CM09, CM14, CM44, and CM52, detecting 35, 48, 101, and 53 genes, respectively;  $p < 10^{-5}$ : CM41 detecting 55 genes;  $p < 10^{-6}$ : CM04, CM05, detecting 163 and 86 genes, respectively;  $p < 10^{-8}$ : CM02, detecting 242 genes (the larger sample size for this individual increased the statistical power, resulting in more detections).

After removing duplicated differentially expressed genes (DEGs), the TMM expression levels of the 781 genes listed in Supplementary Table S2 were subjected to two-way hierarchical clustering using Ward's method, and eight differentially expressed gene modules were observed. Functional annotation of these modules by gene ontology and other categories was assessed using the ToppFun tool of the ToppGene suite online [33]. A full list of results is provided in Supplementary Table S3.

### 3. Results

#### 3.1. The Individual Is the Major Source of Variation as Patients Recover from Infection

The characteristics of the study participants are shown in Table 1. The age range was 3 to 69 years (mean 34 years), with four individuals of each sex. Five patients were from the mid-Pacific lowland region of Chocó (CH), one from the Valle del Cauca (VC) region, and two from the southwest Pacific coast region of Nariño. Clinical presentations included respiratory distress, jaundice, and oral intolerance to drug administration, while blood laboratory tests revealed severe thrombocytopenia and/or anemia, as well as renal and hepatic dysfunction, in several instances; as expected, parasitemia was lower in the two *P. vivax* malaria cases than in four of the six *P. falciparum* cases [10].

**Table 1.** Patient demographic, clinical, and laboratory information.

Subject	Site	Sex	Parasite Spp/Counts	Clinical Criteria §	Laboratory Criteria	Age	Hospital Days
CM09	Valle del Cauca	M	<i>Pf.</i> /25.1	OI, RD	ST	69	5
CM14	Chocó	M	<i>Pf.</i> /19.6	OI	Hep, ST, MH	37	5
CM41	Chocó	F	<i>Pf.</i> /142	OI, Jau	ST	49	4
CM44	Chocó	F	<i>P.v.</i> /5.7	OI, Jau	MH, RI, ST	25	5
CM52	Chocó	M	<i>P.v.</i> /5.7	RD	None	3	4
CM05	Chocó	F	<i>Pf.</i> /11.6	RD	ST	38	4
CM02 *	Chocó	F	<i>Pf.</i> /5.2, 4.3	RD	SA	28	11, 18
CM04	Nariño	M	<i>Pf.</i> /20.8	OI	RI, Hep	22	4

\* 25-week pregnant, two visits. **Parasite spp:** *Pf.* = *Plasmodium falciparum*; *P.v.* = *Plasmodium vivax*. **Counts:** Parasites  $\times 1000/\mu\text{L}$ . § Clinical criteria OI = oral intolerance, Jau = jaundice, RD = respiratory distress. Laboratory criteria ST = severe thrombocytopenia, SA = severe anemia, MH = macroscopic hemoglobinuria, RI = renal injury, Hep = hepatic failure.

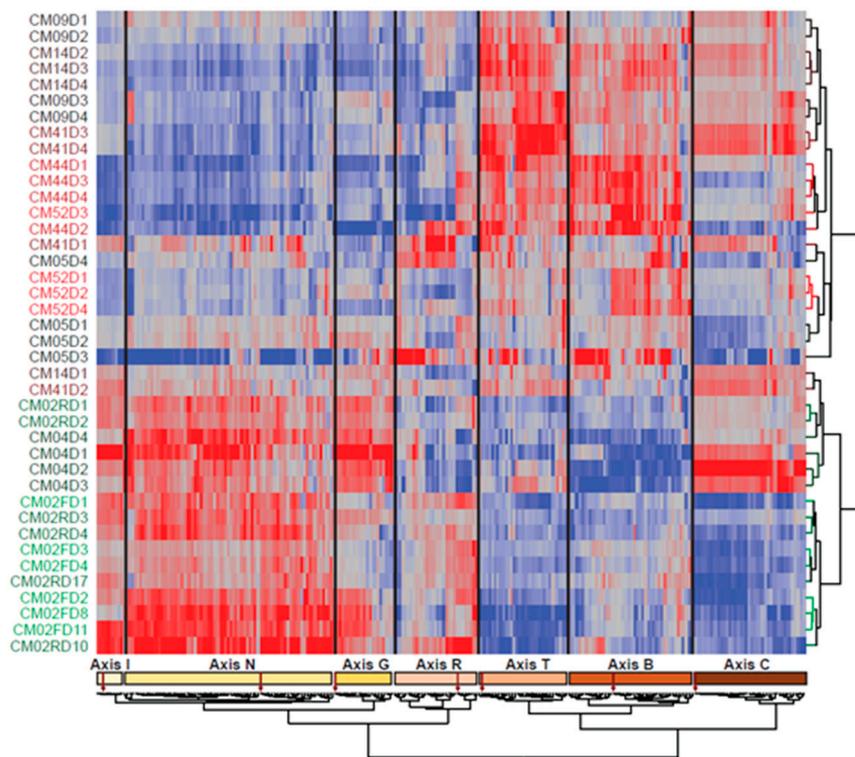
The PVCA of the full dataset consisting of 40 RNAseq samples from the eight complicated malaria cases revealed that 54% of the variation was recorded among individuals, 12% between the two *Plasmodium* species, 10% between males and females, and <1% between days. The contributions of each source of variation to individual PC are listed in Supplementary Table S4. The first principal

component mainly captures inter-individual variation, whereas PC2 and PC3 derive from a mixture of sources and include strong influences of the parasite species. Although there was a considerable variation across days within individuals, this analysis clearly showed that overall there was not a common profile of resolution whereby patients moved from an acute profile at hospitalization to a healthier profile at discharge. It is unclear whether the parasite and sex components were significant, given that just two patients were infected with *P. vivax*, and that, in large samples of healthy individuals, sex accounts for only a few percent of blood gene expression variation (e.g., [24,34]). In order to contrast complicated with uncomplicated malaria, we added another 24 samples to the dataset, consisting of 12 Colombian individuals previously subjected to experimental malaria infection, who were sampled before and after infection [14]. All but one of the complicated malaria samples clustered with these experimentally infected samples (data not shown), all of which were immediately resolved following standard malaria treatment, and none were associated with severe symptoms. Consequently, there is no evidence that the complicated cases are characterized by a distinct profile of infection that is qualitatively different from uncomplicated malaria.

### 3.2. Modular Analysis Resolves the Major Immune Activities Engaged in Each Individual

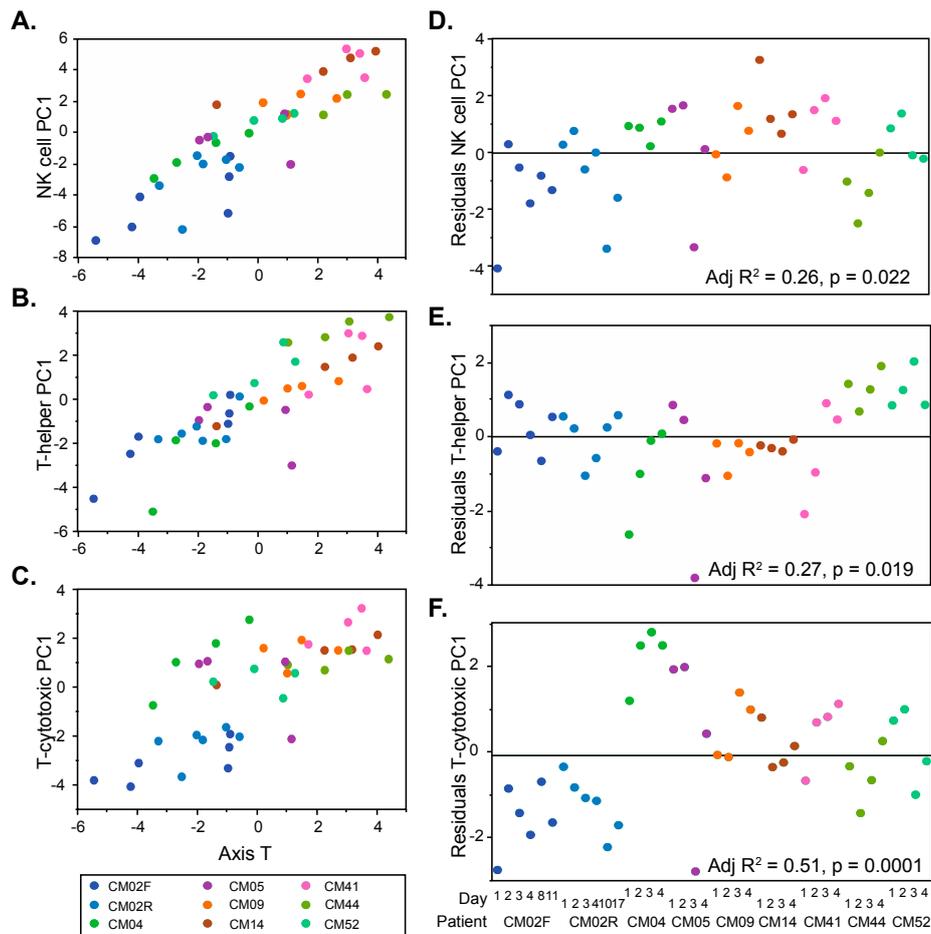
To gain insight into the immunological pathways that vary within and among individuals, we next performed BTM and BIT analysis by generating the PC1 for each set of BTM module genes or the 10 characteristic BIT axes (see Materials and Methods for details). An initial nomenclature considered Axis 1 through Axis 9 [24]; however, for a better interpretation of results, we recently [35] renamed the Axes T, B, N, R, I, G, and C for T cell, B cell, Neutrophil, Reticulocyte, Interferon response, General, and Cell cycle, corresponding to Axes 1, 3, 5, 2, 7, 4, and a newly discovered Axis 10, respectively. The former Axes 6, 8, and 9 were excluded from the analysis because we do not know their biological function. Axis C is enriched for genes involved in mitosis and cell cycle regulation, and the PC1 score for this Axis is very similar to that of multiple BTM annotated to cell cycle functions.

Two-way hierarchical clustering of the 253 BTM and 7 BIT scores in Figure 1 with individual samples as rows revealed a very clear structure, whereby the individual samples all tended to cluster with one another, and the BTM fell into six clusters indicated along the bottom, each of which included one BIT axis as indicated. The leftmost cluster actually included both Axis I and Axis N (though they are in distinct sub-clusters, as shown by the lower dendrogram), indicating that inflammation and interferon response were tightly coordinated in this dataset. The Reticulocyte Axis is right next to BTM173 (erythrocyte differentiation) and BTM171 (heme biosynthesis) and immediately adjacent to the two platelet activation modules (BTM32\_0 and BTM32\_1). All five patients who reported severe thrombocytopenia had at least two time points with a standard deviation more than one unit below the mean for these two modules (although CM52 had the two lowest scores). Intriguingly, on the right of the figure, all of the cell cycle modules and Axis C quite cleanly distinguished four patients with consistently low mitotic activity (CM02, CM05, CM44, and CM52) from the remainder. Three of these patients experienced respiratory distress, as did CM09, whose cell cycle expression level was intermediate. See Supplementary Table S5 for the identities of the BTM in each cluster.



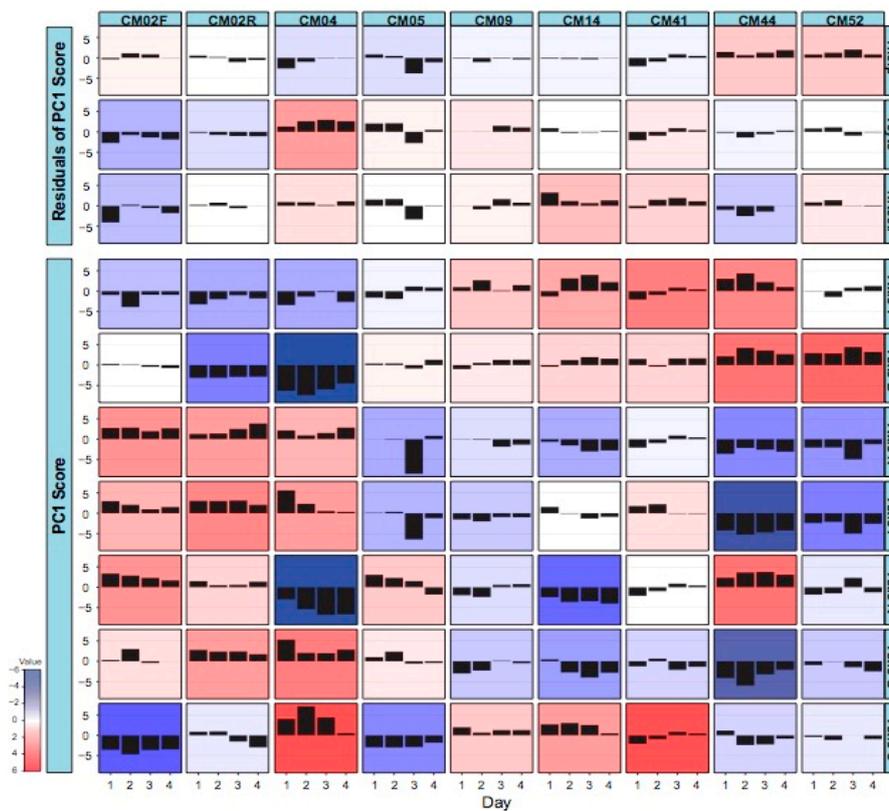
**Figure 1.** Two-way hierarchical clustering of blood transcript modules (BTM) and blood informative transcripts (BIT) by individual and time. Each row represents one patient on the day indicated by the last number of the label, and each column is the principal component (PC1) score for a BTM, with co-clustered BIT PC1 highlighted. The red-to-blue scale indicates positive to negative scores. The figure highlights the tendency of the samples to cluster by individual and provides an overview of the axes of variation shared by particular patients.

The B and T cell activity Axes clustered with 78 BTM annotated to various aspects of lymphoid signaling and activity. We attempted to further resolve the immune cell contributions by taking advantage of recently reported single-cell RNAseq results from whole blood [21], which document dozens of genes significantly elevated in NK cells, cytotoxic T cells, or helper T cells (see Materials and Methods). Figure 2A–C show that PC1 for each of these gene sets is highly correlated with Axis T, explaining the lack of resolution of the sub-types in Figure 1. However, the residuals of the regressions significantly differed among patients, as shown in Figure 2D–F. Our interpretation is that the individuals with positive residuals had a higher abundance of the genes representative of that cell type, likely implying the elevated abundance of the cell type itself. By this reasoning, CM04, CM14, and CM41 had consistently elevated NK cell activity, CM44 and CM52 had elevated helper T cell activity, and CM04 had very high cytotoxic T cell activity, whereas CM02 showed a remarkably low activity of that component.



**Figure 2.** Lymphocyte activity inference in each individual. Each plot (A–C) shows how a PC1 based on natural killer (NK) cell, T-helper cell, or T-cytotoxic cell specific transcripts correlates with Axis T and varies by individual, indicated by different colors. Graphs (D–F) plot the residuals of the linear regressions by individual, highlighting patient-specific lymphocyte activity. *p*-values and adjusted  $R^2$  estimates were calculated through an analysis of variance (ANOVA).

The time course of BIT Axis scores for each patient is summarized in the matrix of histograms presented in Figure 3. The results strongly imply that each patient confronted the malaria infection with a quite different immune response. In short, the following patterns emerged. Table 1 lists the clinical presentation features for each patient, but there was no clear relationship between the clinical data and the gene expression data, except for a few cases mentioned below.



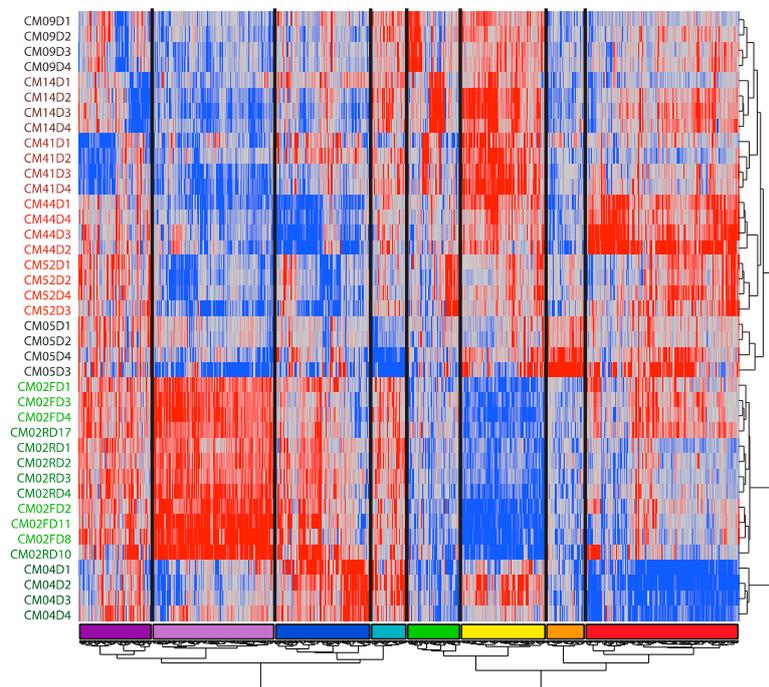
**Figure 3.** Standardized change in BIT by individual and time. Each box shows the difference between the sample PC1 score and the sample mean for the indicated individual from Day 1 to Day 4 of sampling. Red implies strong elevation, and blue strong reduction in the expression of genes in the Axis.

- CM02, who presented with severe anemia clinical manifestations and relatively low parasitemia, had generally low cytotoxic T-lymphocyte and natural killer cell activity, which may explain her failure to resolve the infection. She also had particularly low mitotic activity during the initial infection and low B cell function during the relapse. All this was compensated by elevated inflammatory and interferon responses.
- CM04, who experienced renal and liver failure, also had a notable inflammatory response and low B cell activity, but this was offset by high cytotoxic T cell and mitotic activity. His reticulocyte numbers were relatively low, consistent with the absence of anemia, and his interferon response settled down upon hospitalization.
- CM05, with severe thrombocytopenia, was notable for a dramatic drop in inflammation and interferon signaling specifically on day 3 and for a relatively high cytotoxic T cell function on days 1 and 2.
- CM09, with severe thrombocytopenia and high parasitemia, had a rather stable profile without any dominant characteristics, perhaps correlating with his less complicated disease or with his age (69), which suggests that he may have built up more tolerance to malaria than the younger patients.
- CM14, with hepatic failure and severe thrombocytopenia, was notable for very high NK cell signatures and relatively low reticulocytes.
- CM41, who experienced severe thrombocytopenia and jaundice, presumably due to liver dysfunction, had relatively high NK cell and cytotoxic T cell activity, as well as elevated cell cycle gene expression, similar to CM14.
- CM44 and CM52, both infected with *P. vivax*, had similar transcriptome profiles. CM52 was a 3-year-old boy at the time of this episode of malaria and was mainly hospitalized for respiratory distress. CM44 presented a more extreme condition (severe thrombocytopenia, jaundice, and

renal injury) with respect to her combination of low inflammation (perhaps related to a relatively low parasitemia) and particularly high B cell and helper T cell activity.

### 3.3. Patient-Specific Gene Expression Reinforces the Conclusions of the Modular Analysis

Given that each patient showed a relatively unique modular pattern during hospitalization, we next asked whether their gene expression patterns followed patient-specific trajectories. To do so, we identified patient-specific DEGs using one-way ANOVA (see Materials and Methods). After removing duplicated DEGs, we obtained 781 DEGs in total and performed two-way hierarchical clustering, observing eight DEG modules (Figure 4). Gene ontology functional enrichment analysis of the DEGs within each module revealed striking concordances with the modular analysis (Supplementary Table S3). It is also notable that with this mode of patient-specific gene expression analysis, the four samples for each patient always clustered together.



**Figure 4.** Two-way hierarchical clustering of differentially expressed gene (DEG) by individual and time. Each row represents one patient on the day indicated by the last number of the label, and each column is the standardized expression of a DEG in ANOVA contrasting each individual with the remaining sample. The figure highlights clusters of genes that tend to be patient-specific, but shows (like the BTM analysis) that these are to some extent shared by pairs of patients. The colored bar at base indicates clusters discussed in the text.

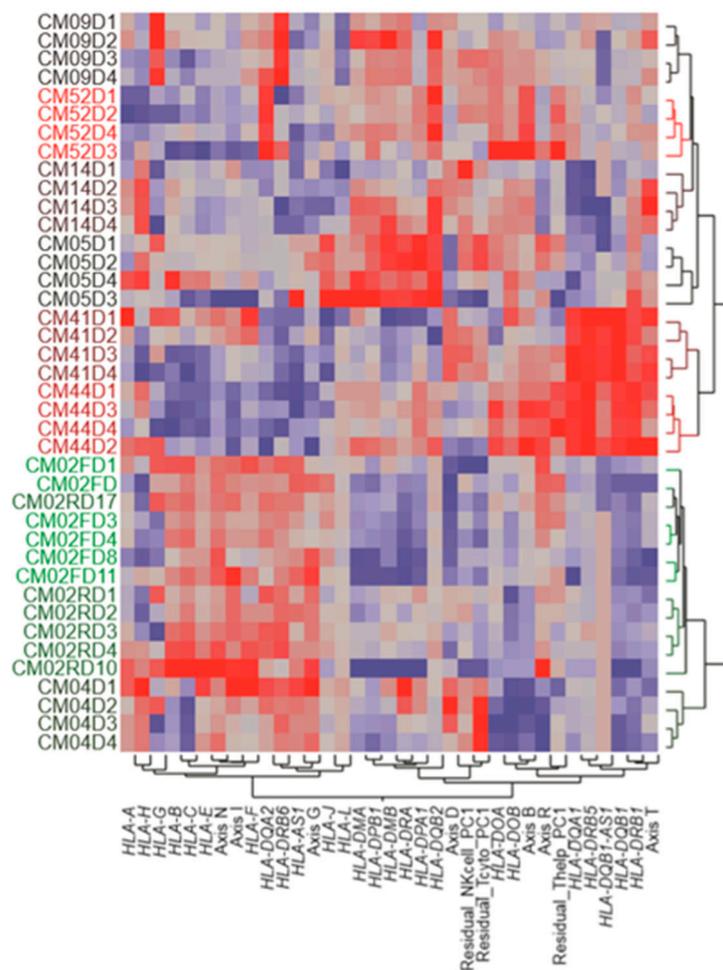
The expression of the red cluster genes was high in CM44 and CM02 during the first three days, and in CM05 on days 3 and 4, but low in CM04. Gene ontology analysis revealed enrichment for erythropoiesis and B cell differentiation markers, which was concordant with the modular analysis Axis R and Axis B variation patterns reported above. Similarly, gene expression in the yellow module was high in CM41, CM14, CM44, and CM09 but low in CM02, which, as gene ontology analysis suggested, may indicate gamma-delta T cell deficiency and mitochondrial dysfunction. Once again, this agreed with the Axis T of the modular analysis. Additionally, gene expression of the dark blue module, which is related to type I interferon signaling, was upregulated in CM02, CM04, and CM41 on days 1 and 2 but downregulated in CM44, matching the Axis I pattern of the modular analysis. Furthermore, gene expression of the light purple module was high in CM02, relatively high in CM04,

but low in other patients. Gene ontology analysis implies that this module is related to neutrophils and inflammation, in line with the Axis N pattern of the modular analysis.

Even though the BIT and BTM patterns were validated by the modular analysis, a small number of more detailed patient-specific results were also revealed by the DEG analysis. For example, gene expression of the orange module was high in CM05 but low in other patients. Ten BTM related to myeloid dendritic cell activation were also upregulated in this patient.

### 3.4. Variable Expression of HLA Complex Genes among Patients

Since the human leukocyte antigen (HLA) complex is known to play an important role in the processes of protection against intracellular pathogens, including the immune response during malaria infection, we also examined the collective regulation of 25 HLA genes and their relationship to the BIT Axes (Figure 5). Overall, three clusters of co-regulated HLA expression were observed. The HLA class I genes, *HLA-A* to *HLA-H*, were all relatively highly expressed in CM02 and CM04, the two patients with the strongest inflammatory response. In addition, three individuals displayed an unusually high expression of single class I genes, namely, *HLA-G* in CM09, and *HLA-H* in CM14. Two class II genes were also notably uniquely elevated in these two individuals, i.e., *HLA-DRB6* and *HLA-DQA2*, respectively. It is not known whether these specific patterns may have influenced the complicated course of the disease.



**Figure 5.** Two-way hierarchical clustering of HLA complex gene expression by individual and time. Each row represents one patient on the day indicated by the last number of the label, and each column is the standardized expression of the indicated HLA gene or lymphocyte activity estimate. The figure highlights three clusters of co-expression, namely, HLA class I to the left and two HLA class II clusters in the middle and right of the grid.

Regarding the expression of HLA class II genes, these showed relatively low levels in the two patients with high neutrophil and interferon activity but separated into two sub-groups. One included *HLA-DMA*, *HLA-DMB*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, and *HLA-DQB2*, the expression of which was particularly high in CM05, CM09, and CM52, all of whom were among the high inflammatory response individuals. This suggests a trade-off between inflammation and HLA-class I and class II gene expression. The second sub-group included *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*, and *HLA-DRB5*, whose expression was notably elevated in CM41 and CM44, two individuals with relatively high T cell activity in general but otherwise contrasting profiles. Larger sample sizes will be required to ascertain whether and how HLA class II gene expression is associated with complicated malaria.

#### 4. Discussion

Given that malaria is a blood stream infection, recovery from which requires elimination of parasite-infected red blood cells, our expectation was that longitudinal transcriptome profiling would reveal common regulatory signatures of engagement of one or more arms of the immune system. At least, elevated reticulocyte development and reduced engagement of innate and interferon-mediated responses, as patients move from acute clinical manifestations to healthy condition and release from hospital, would be expected upon recovery. However, while there were signs of such responses in a majority of the patients, the timing and strength of the signals revealed by Axes R, N, and I were notably variable among patients. No other strong signatures, for example of engagement of B cells, a subset of T cells, or perhaps a pathway such as antigen processing, were observed across the dataset. Rather, the overwhelming pattern was one of highly individualized profiles of immune-related recovery. The dominant variance component in the dataset was between-individual effects, and each person displayed a personal trajectory of recovery.

Such individualization is reminiscent of our findings in healthy individuals [36]. In that study, we found, by contrasting the RNAseq profiles of a dozen of individuals at three six-month intervals, that all subjects consistently clustered together, whether considering the major PC of the entire transcriptome or the Axis scores. This gave rise to the notion of an “omic personality”, a consistent profile of gene expression over time that reflects the baseline healthy status of an individual [36]. This baseline expression is quite likely to have a major influence on the recovery profiles.

Similarly, microarray-based profiling of lupus erythematosus patients [37] discovered at least eight dominant trajectories of gene expression over time. Clustering transcripts on the basis of their correlation with various clinical traits of relevance to pathology, including the systemic lupus erythematosus disease activity index (SLEDAI), showed that some patients had a predominantly B cell-related pattern, while, in others, reticulocytes, NK cells, or T cells seemed to produce the strongest correlations. Despite the well-documented interferon response seen in lupus, a plasmablast signature present in most patients, and neutrophilia associating with nephritis, it was the personalized nature of the immune gene expression monitoring that dominated the analysis.

These observations raise the question of whether baseline peripheral blood profiles might be predictive of the course of a disease or provide personalized information regarding the best course of the curative treatment. The lupus study was ambiguous on this point, and, despite trends such as association of low cell cycle activity with respiratory distress, our pilot study is too small to make any definitive statements. What is clear is that the different pairs of patients with high B cell (CM44 and CM52), NK cell (CM14 and CM41), or cytotoxic T cell (CM04 and CM05) activity did not share complicated malaria attributes to the exclusion of other patients, so a close matching of immune profile to clinical presentation is unlikely.

Another possible source of patient-specific clinical presentation is the aberrant expression of one or a handful of genes that lead to a complicated course rather than to a standard resolution. We noted several instances of divergent expression of *HLA* or other genes of interest, but samples from several thousand patients will be required to establish any statistical association with a phenotype. Most of the divergent expression was due to an individual having the highest or lowest values for the relevant

axis of variation, but it is likely that mutations and segregating regulatory variants also contribute for specific genes. The individuals CM09, CM14, and C52, whose overall profiles had relatively normal axis scores, also had divergent *HLA* gene expression levels, and these genes may be involved in mediating aspects of the complicated course of disease.

It is not known what causes the strong patterns of covariance of gene expression. Some of it is due to the differential abundance of contributing lymphoid and myeloid cell types, but immune cell type abundance is known to have low heritability [38,39]. Aspects such as interferon response, cell cycle regulation, and general cellular metabolism are likely affected by multiple immune cell types, and the longitudinal consistency of the corresponding axis scores suggests a substantial trans-acting regulatory component. The interaction of the immune system with the parasite would significantly contribute at multiple stages of the infection cycle, from parasite entry and development in the liver cells to parasite multiplication during blood infection. *Plasmodium* antigens likely induce immune responses, which engage all of the major T cell subsets, possibly modulating the immune response, once the acute clinical phase of infection commences. Although, in malaria, the immune system is primed upon sporozoite injection by a mosquito under the patient's skin, it is only during the parasite development in the blood stream that the clinical manifestations appear. Whether or not this knowledge can be used to understand and eventually manipulate the immune responses in a personalized manner is unclear but will likely require longitudinal studies involving thousands of patients across diverse infectious and complicated disease phenotype classes.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2075-4426/8/3/29/s1>, Supplementary Table S1: BTM Module and BIT Axis scores. Supplementary Table S2: List of Differentially Expressed Genes. Supplementary Table S3: Gene Ontology Analysis. Supplementary Table S4: Principal Variance Components. Supplementary Table S5: BTM in each Cluster of DEG.

**Author Contributions:** M.A.-H., S.H. and G.G. conceived the study. M.L.R.-P. processed the RNA samples for sequencing. M.L.R.-P. and M.D. analyzed the data. S.H., M.A.-H. and L.R. performed the sample collection and managed the epidemiological and clinical data. D.A. managed IRB approvals and assisted with sample processing. M.L.R.P. wrote the first draft of the paper. All authors edited, revised, and approved the manuscript.

**Funding:** This work was supported by grants from the US National Institutes of Health (NIAID/ICEMR U19AI089702), Colciencias (360-2011, 458-2012, 719-2013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgments:** The research was conducted in sites corresponding to *Centro Latino Americano de Investigación en Malaria* (CLAIM), Center for Integrative Genomics, and School of Biological Sciences, Georgia Institute of Technology.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. World Health Organization (WHO). *World Malaria Report*; WHO: Geneva, Switzerland, 2017.
2. Chaparro-Narváez, P.E.; Lopez-Perez, M.; Rengifo, L.M.; Padilla, J.; Herrera, S.; Arévalo-Herrera, M. Clinical and epidemiological aspects of complicated malaria in Colombia, 2007–2013. *Malar. J.* **2016**, *15*, 269. [[CrossRef](#)] [[PubMed](#)]
3. Sypniewska, P.; Duda, J.F.; Locatelli, I.; Althaus, C.R.; Althaus, F.; Genton, B. Clinical and laboratory predictors of death in African children with features of severe malaria: A systematic review and meta-analysis. *BMC Med.* **2017**, *15*, 147. [[CrossRef](#)] [[PubMed](#)]
4. Arévalo-Herrera, M.; Lopez-Perez, M.; Medina, L.; Moreno, A.; Gutierrez, J.B.; Herrera, S. Clinical profile of *Plasmodium falciparum* and *Plasmodium vivax* infections in low and unstable malaria transmission settings of Colombia. *Malar. J.* **2015**, *14*, 154. [[CrossRef](#)] [[PubMed](#)]
5. Arévalo-Herrera, M.; Rengifo, L.; Lopez-Perez, M.; Arce-Plata, M.I.; Garcia, J.; Herrera, S. Complicated malaria in children and adults from three settings of the Colombian Pacific Coast: A prospective study. *PLoS ONE* **2017**, *12*, e0185435. [[CrossRef](#)] [[PubMed](#)]
6. World Health Organization. Severe malaria. *Trop. Med. Int. Health.* **2014**, *19* (Suppl. 1), 7–131. [[CrossRef](#)] [[PubMed](#)]

7. Tobon-Castano, A.; Giraldo-Castro, C.; Blair, S. Prognostic value of clinical and parasitological signs for severe malaria in patients from Colombia. *Biomedica* **2012**, *32* (Suppl. 1), 79–94. [[CrossRef](#)] [[PubMed](#)]
8. Rahimi, B.A.; Thakkinstian, A.; White, N.J.; Sirivichayakul, C.; Dondorp, A.M.; Chokejindachai, W. Severe vivax malaria: A systematic review and meta-analysis of clinical studies since 1900. *Malar. J.* **2014**, *13*, 481. [[CrossRef](#)] [[PubMed](#)]
9. Quispe, A.M.; Pozo, E.; Guerrero, E.; Durand, S.; Baldeviano, G.C.; Edgel, K.A.; Graf, P.C.; Lescano, A.G. *Plasmodium vivax* hospitalizations in a monoendemic malaria region: Severe vivax malaria? *Am. J. Trop. Med. Hyg.* **2014**, *91*, 11–17. [[CrossRef](#)] [[PubMed](#)]
10. Gonzalez, L.M.; Guzman, M.; Carmona, J.; Lopera, T.; Blair, S. Clinical and epidemiologic characteristics of 291 hospitalized patients for malaria in Medellin (Colombia). *Acta Med. Colomb.* **2000**, *25*, 163–170.
11. Langhorne, J.; Ndungu, F.M.; Sponaas, A.M.; Marsh, K. Immunity to malaria: More questions than answers. *Nat. Immunol.* **2008**, *9*, 725–732. [[CrossRef](#)] [[PubMed](#)]
12. Gazzinelli, R.T.; Kalantari, P.; Fitzgerald, K.A.; Golenbock, D.T. Innate sensing of malaria parasites. *Nat. Rev. Immunol.* **2014**, *14*, 744–757. [[CrossRef](#)] [[PubMed](#)]
13. McCall, M.B.; Sauerwein, R.W. Interferon-gamma-central mediator of protective immune responses against the pre-erythrocytic and blood stage of malaria. *J. Leukoc. Biol.* **2010**, *88*, 1131–1143. [[CrossRef](#)] [[PubMed](#)]
14. Rojas-Peña, M.L.; Vallejo, A.; Herrera, S.; Gibson, G.; Arévalo-Herrera, M. Transcription profiling of malaria-naïve and semi-immune Colombian volunteers in a *Plasmodium vivax* sporozoite challenge. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003978.
15. Idaghdour, Y.; Quinlan, J.; Goulet, J.; Berghout, J.; Gbeha, E.; Bruat, V.; de Malliard, T.; Grenier, J.C.; Gomez, S.; Gros, P.; et al. Evidence for additive and interaction effects of host genotype and infection in malaria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16786–16793. [[CrossRef](#)] [[PubMed](#)]
16. Berg, A.; Otterdal, K.; Patel, S.; Gonca, M.; David, C.; Dalen, I.; Nymo, S.; Nilsson, M.; Nordling, S.; Magnusson, P.U. Complement activation correlates with disease severity and contributes to cytokine responses in *Plasmodium falciparum* malaria. *J. Infect. Dis.* **2015**, *212*, 1835–1840. [[CrossRef](#)] [[PubMed](#)]
17. Biryukov, S.; Angov, E.; Landmesser, M.E.; Spring, M.D.; Ockenhouse, C.F.; Stoute, J.A. Complement and antibody-mediated enhancement of red blood cell invasion and growth of malaria parasites. *EBioMedicine* **2016**, *9*, 207–216. [[CrossRef](#)] [[PubMed](#)]
18. Wang, Z.; Gerstein, M.; Snyder, M. RNAseq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
19. Lowe, R.; Shirley, N.; Bleackley, M.; Dolan, S.; Shafee, T. Transcriptomics technologies. *PLoS Comput. Biol.* **2017**, *13*, e1005457. [[CrossRef](#)] [[PubMed](#)]
20. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)] [[PubMed](#)]
21. Zheng, G.X.Y.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **2017**, *8*, 14049. [[CrossRef](#)] [[PubMed](#)]
22. Li, S.; Roupael, N.; Duraisingham, S.; Romero-Steiner, S.; Presnell, S.; Davis, C.; Schmidt, D.S.; Johnson, S.C.; Milton, A.; Rajam, G.; et al. Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines. *Nat. Immunol.* **2014**, *15*, 195–204. [[CrossRef](#)] [[PubMed](#)]
23. Chaussabel, D.; Quinn, C.; Shen, J.; Patel, P.; Glaser, C.; Baldwin, N.; Stichweh, D.; Blankenship, D.; Li, L.; Munagala, I.; et al. A modular analysis framework for blood genomics studies: Application to systemic lupus erythematosus. *Immunity* **2008**, *29*, 150–164. [[CrossRef](#)] [[PubMed](#)]
24. Preininger, M.; Arafat, D.; Kim, J.; Nath, A.P.; Idaghdour, Y.; Brigham, K.L.; Gibson, G. Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet.* **2013**, *9*, e1003362. [[CrossRef](#)] [[PubMed](#)]
25. WHO. Severe falciparum malaria. World Health Organization, communicable diseases cluster. *Trans. R. Soc. Trop. Med. Hyg.* **2000**, *94*, 1–90.
26. MinSalud, Ministerio de la Protección Social. Colombia. *Guía de Atención Clínica de Malaria Ministerio de la Protección Social, Bogotá*; Ministerio de la Protección Social: Bogota, Colombia, 2010; p. 132.

27. Matheson, L.A.; Duong, T.T.; Rosenberg, A.M.; Yeung, R.S. Assessment of sample collection and storage methods for multicenter immunologic research in children. *J. Immunol. Methods* **2008**, *339*, 82–89. [[CrossRef](#)] [[PubMed](#)]
28. Schroeder, A.; Mueller, O.; Stocker, S.; Salowsky, R.; Leiber, M.; Gassmann, M.; Lightfoot, S.; Menzel, W.; Granzow, M.; Ragg, T. The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **2006**, *7*, 3. [[CrossRef](#)] [[PubMed](#)]
29. Sigurgeirsson, B.; Emanuelsson, O.; Lundeberg, J. Analysis of stranded information using an automated procedure for strand specific RNA sequencing. *BMC Genom.* **2014**, *15*, 631. [[CrossRef](#)] [[PubMed](#)]
30. Kim, D.; Pertea, G.; Trapnell, C.; Pimentel, H.; Kelley, R.; Salzberg, S.L. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **2013**, *14*, R36. [[CrossRef](#)] [[PubMed](#)]
31. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [[CrossRef](#)] [[PubMed](#)]
32. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
33. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **2009**, *37*, W305–W311. [[CrossRef](#)] [[PubMed](#)]
34. Whitney, A.R.; Diehn, M.; Popper, S.J.; Alizadeh, A.A.; Boldrick, J.C.; Relman, D.A.; Brown, P.O. Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1896–1901. [[CrossRef](#)] [[PubMed](#)]
35. Mo, A.; Marigorta, U.M.; Arafat, D.; Chan, L.H.K.; Ponder, L.A.; Jang, S.R.; Prince, R.; Kugathasan, S.; Prahalas, S.; Gibson, G. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. *Genome Med.* **2018**, *10*, 48. [[CrossRef](#)] [[PubMed](#)]
36. Tabassum, R.; Sivadas, A.; Agrawal, V.; Tian, H.; Arafat, D.; Gibson, G. Omic personality: Implications of stable transcript and methylation profiles for personalized medicine. *Genome Med.* **2015**, *7*, 88. [[CrossRef](#)] [[PubMed](#)]
37. Banchereau, R.; Hong, S.; Cantarel, B.; Baldwin, N.; Baisch, J.; Edens, M.; Cepika, A.M.; Acs, P.; Turner, J.; Anguiano, E.; et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell* **2016**, *165*, 1548–1550. [[CrossRef](#)] [[PubMed](#)]
38. Brodin, P.; Jojic, V.; Gao, T.; Bhattacharya, S.; Angel, C.J.; Furman, D.; Shen-Orr, S.; Dekker, C.L.; Swan, G.E.; Butte, A.J.; et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell* **2015**, *160*, 37–47. [[CrossRef](#)] [[PubMed](#)]
39. Carr, E.J.; Dooley, J.; Garcia-Perez, J.E.; Lagou, V.; Lee, J.C.; Wouters, C.; Meyts, I.; Goris, A.; Boeckstaens, G.; Linterman, M.A.; et al. The cellular composition of the human immune system is shaped by age and cohabitation. *Nat. Immunol.* **2016**, *17*, 461–468. [[CrossRef](#)] [[PubMed](#)]

