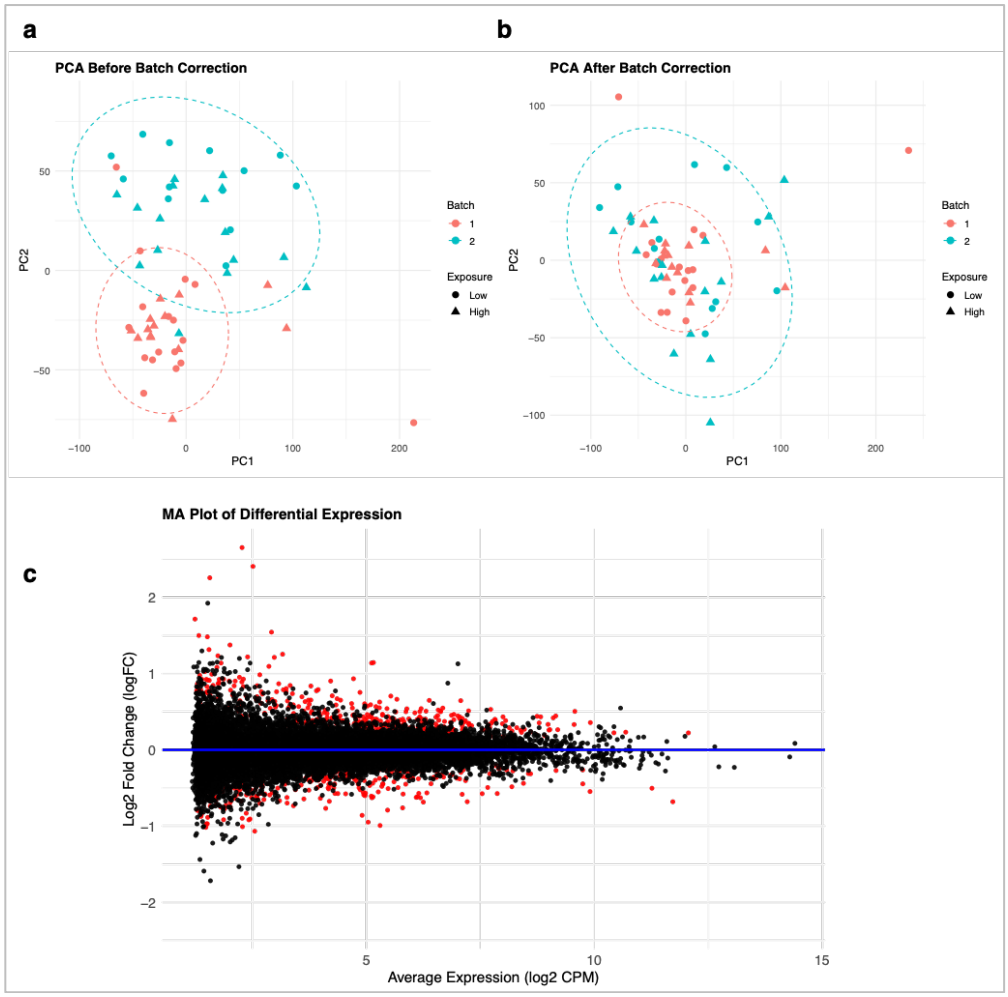
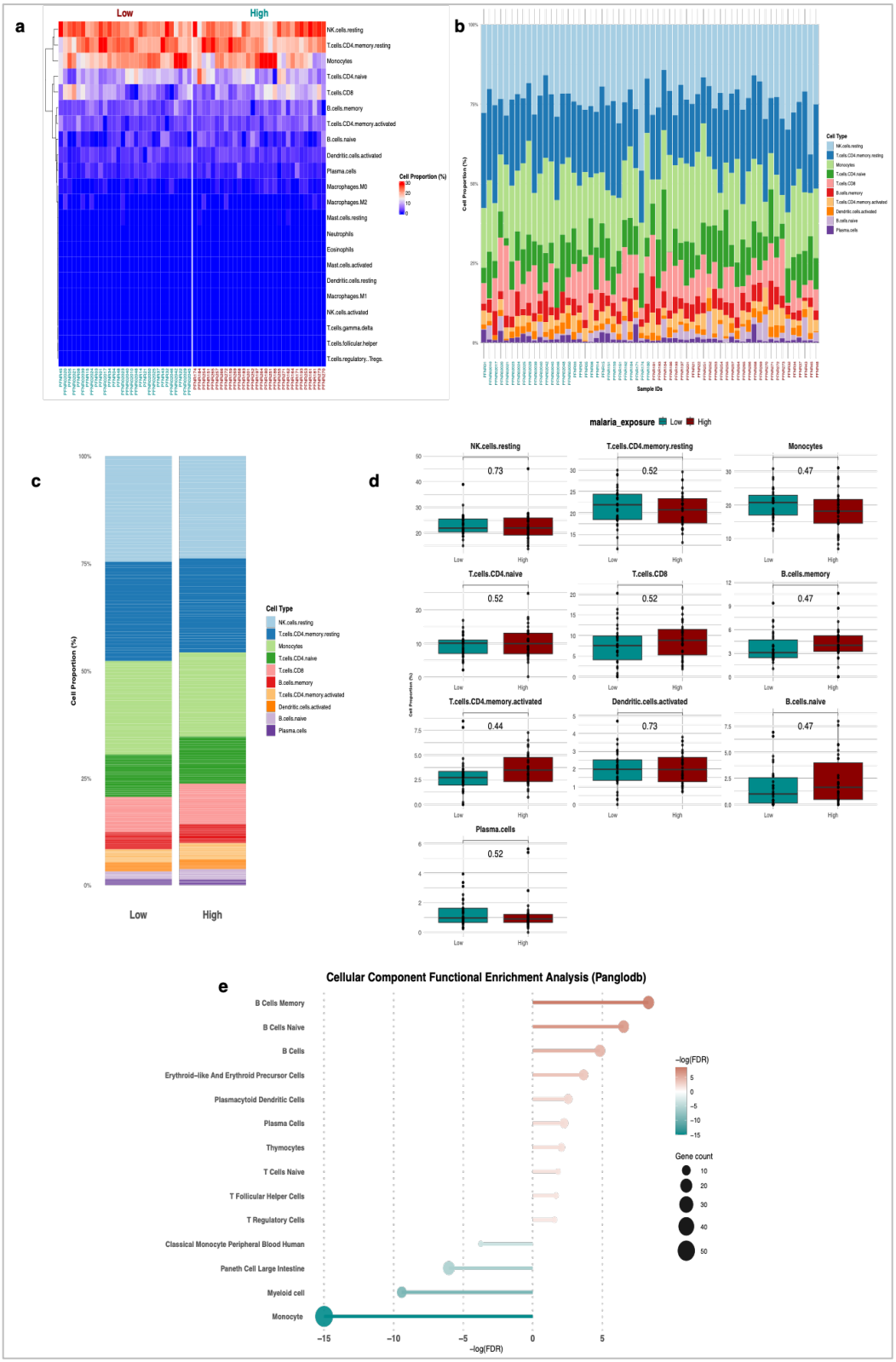
Lifelong malaria exposure drives B cell activation and post-transcriptional regulation while suppressing inflammatory transcriptional programs

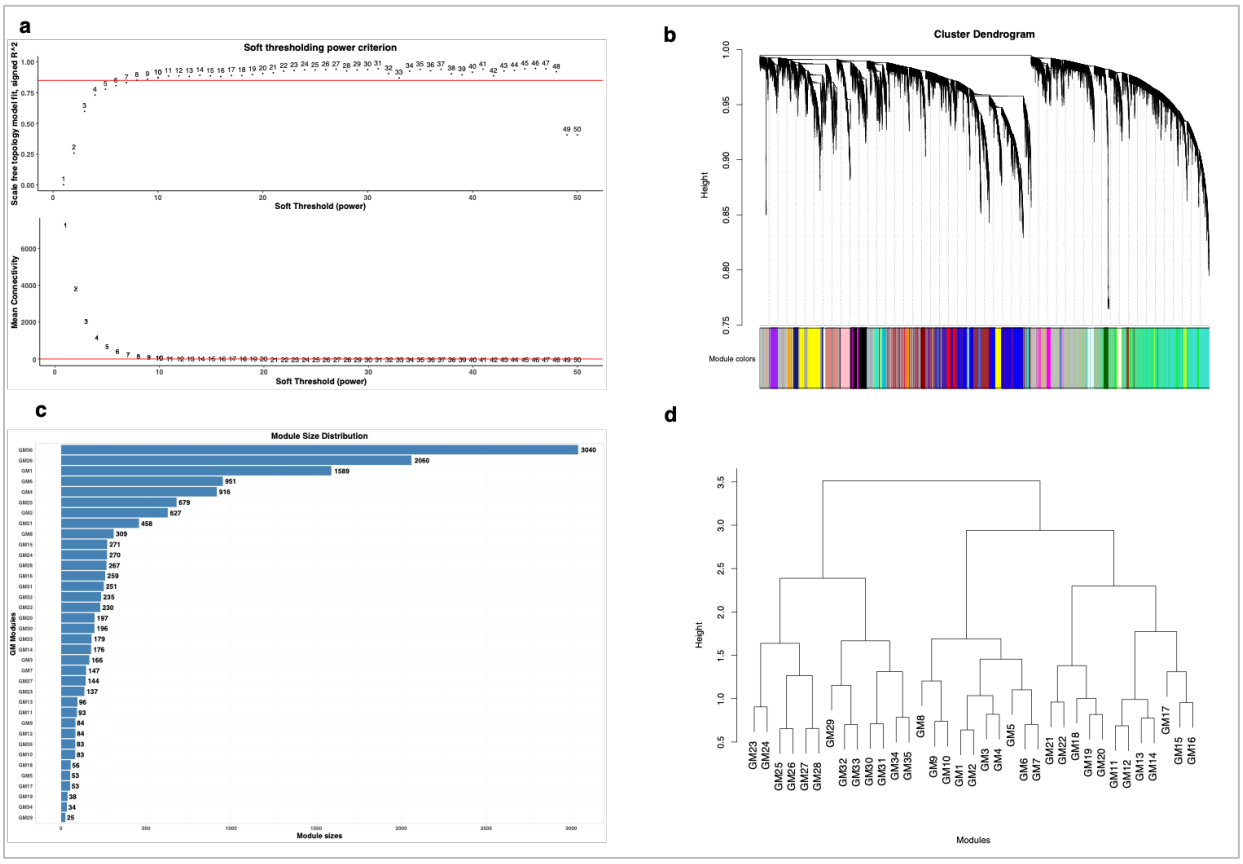


Supplementary Figure 1. Principal component analysis (PCA) of PBMC transcriptomes before and after batch correction, and MA plot of differentially expressed genes (DEGs). (**A–B**) PCA was performed on PBMC gene expression data (CPM) to assess sample clustering by library preparation batch and malaria exposure group. (A) Prior to batch correction, samples cluster by library preparation batch. (B) After batch correction using *removeBatchEffect*() from the *limma* package, batch-associated clustering is reduced. Dashed ellipses represent 95% confidence intervals around each group's centre. (**C**) The MA plot displays each gene as a point, with batch-corrected average expression (log_2 CPM) on the x-axis and log_2 fold change on the y-axis. Genes significantly upregulated (positive fold change) or downregulated (negative fold change) at a false discovery rate (FDR) < 0.05 are highlighted in red. The horizontal blue lines indicate the log_2 fold-change thresholds. Most DEGs occur at moderate expression levels rather than among the most highly expressed genes, suggesting that expression magnitude does not solely drive differential expression. *Abbreviations*: CPM, counts per million; MA, M (log ratio) vs. A (mean average) plot; DEG, differentially expressed gene; PCA, principal component analysis.

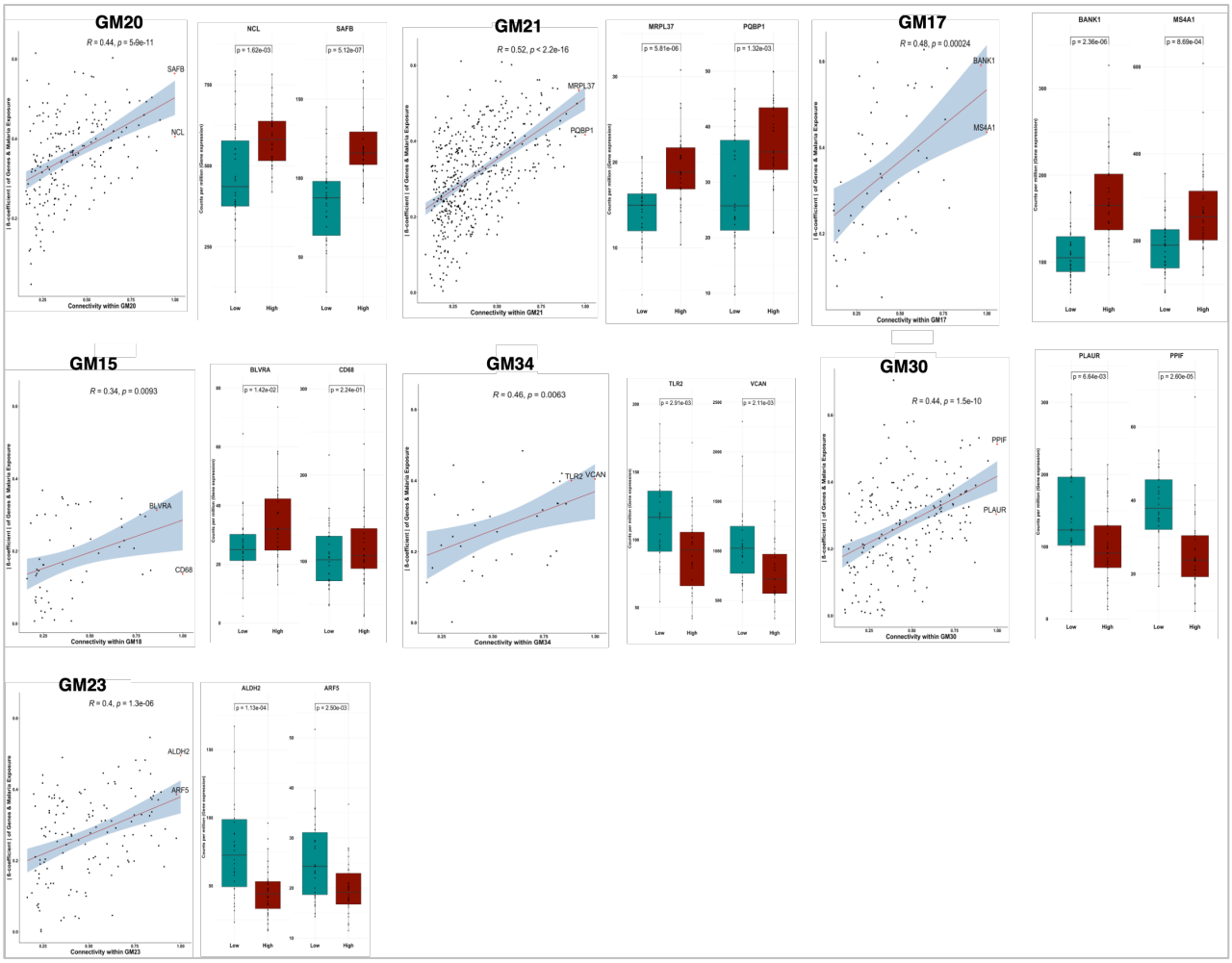


Supplementary Figure 2: Immune cell profiling by digital cytometry and cellular enrichment analysis. (**A**) Heatmap showing the proportions (%) of 22 leukocyte subtypes across malaria exposure groups, estimated from gene expression data (counts per million) using the LM22 [1] gene signature matrix. Cell types are ordered by abundance and clustered hierarchically. The x-axis represents study participant identifiers. (**B**) Bar plots displaying the top 10 most abundant cell

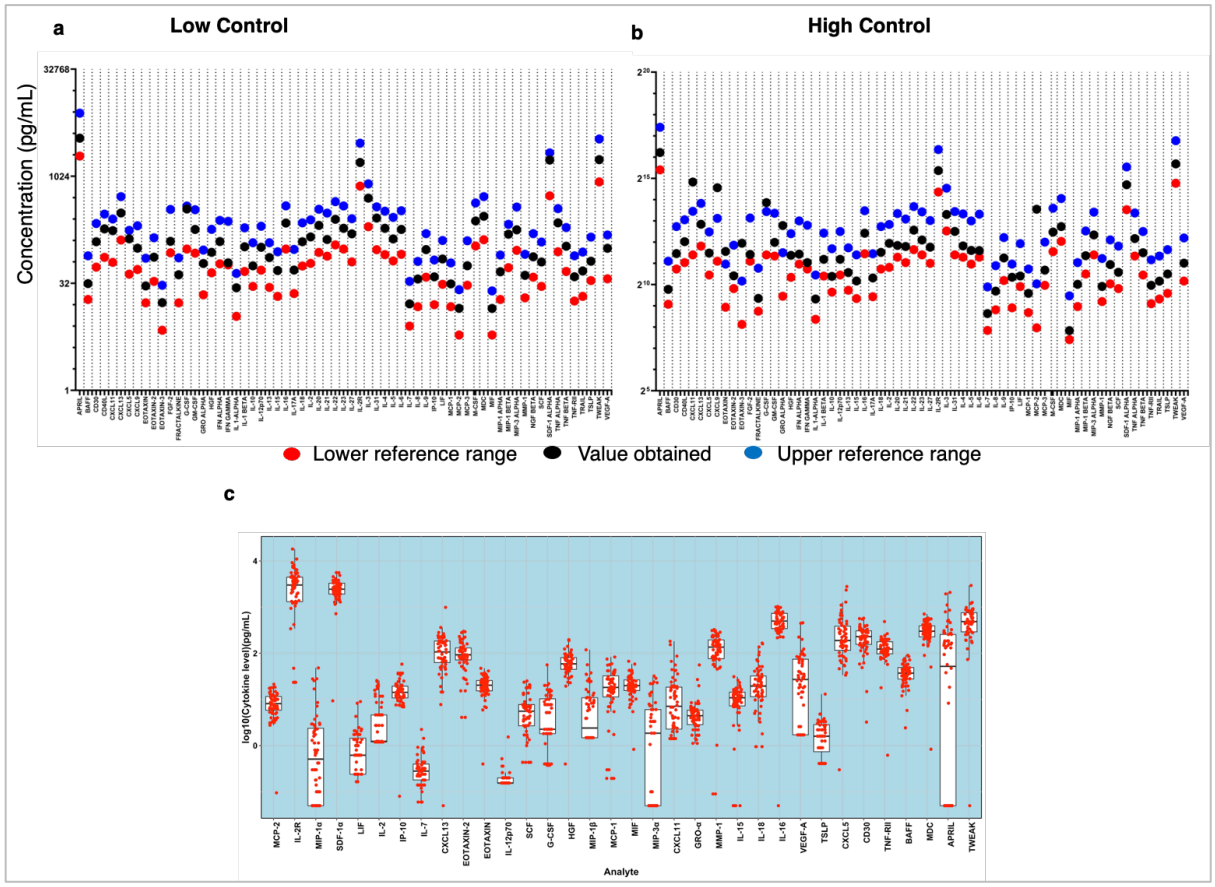
types per individual, grouped by exposure (darkred: high, darkcyan: low). The x-axis represents study participant identifiers. **(C)** Stacked bar plots showing the group median proportions (%) of the top 10 most abundant cell types. (**D**) Groupwise comparisons of the top 10 cell types assessed using two-tailed *t*-tests (parametric) or *Mann-Whitney U* tests (non-parametric), with Benjamini-Hochberg correction (FDR < 0.05) for multiple testing. Box plots indicate median ± IQR, whiskers = 1.5×IQR and outliers as single points. (**E**) Dot plot depicting the results of enrichment analysis performed via the Enrichr platform using the PanglaoDB Augmented 2021 gene set library. The input gene list consisted of differentially expressed genes. The x-axis shows -log (FDR), indicating statistical significance of enrichment, while the y-axis lists enriched cell types. Dot size represents the number of genes overlapping with each cell-type signature (Gene count), and colour intensity reflects the strength of enrichment (-log (FDR)). This analysis highlights key immune cell populations associated with the upregulated/downregulated gene signature.



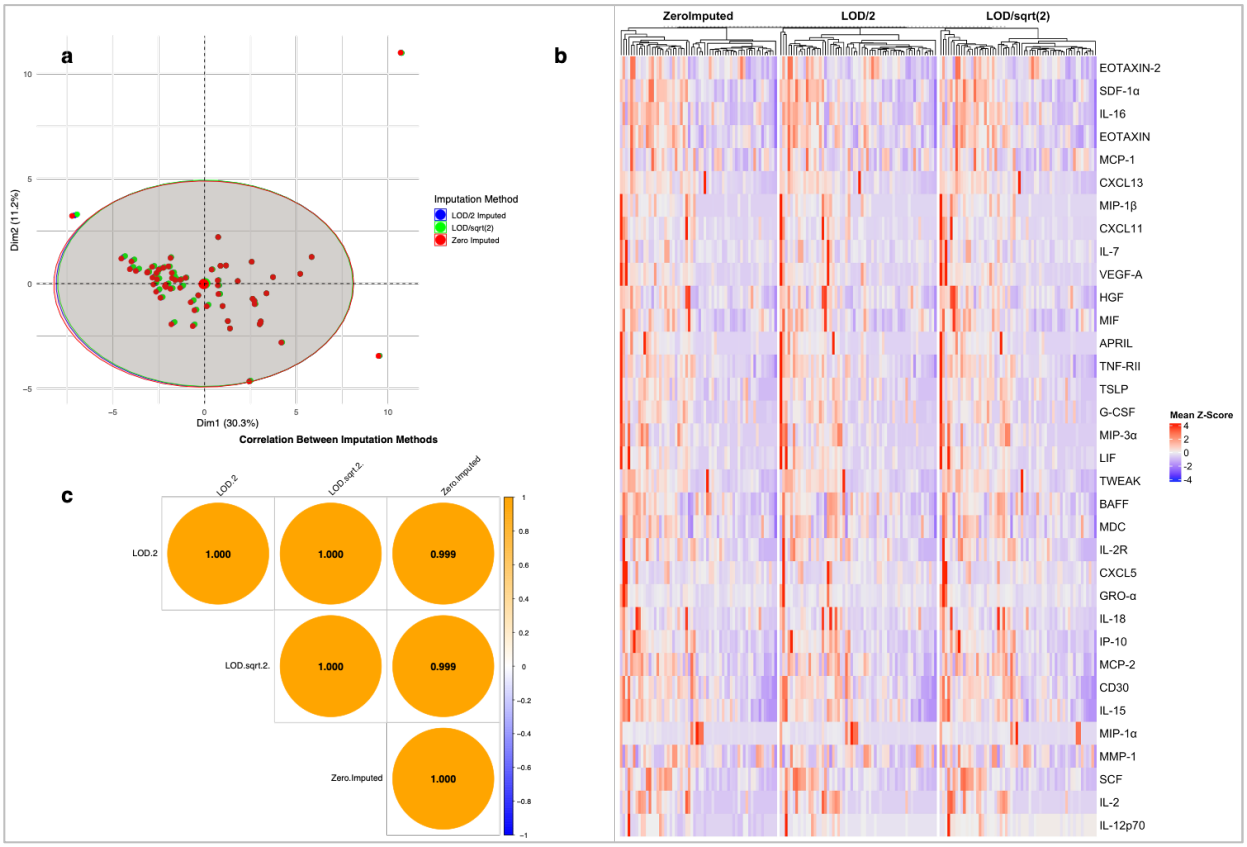
Supplementary Figure 3: Weighted gene co-expression network analysis (WGCNA) construction. (A) A soft thresholding power of 9 was chosen based on scale-free topology fit (R^2 = 0.85, top panel) and minimum mean connectivity (bottom panel) as it resulted in the fewest unassigned genes. Red horizontal lines on both panels indicate selected thresholds. (**B)** Hierarchical clustering dendrogram of genes with module assignments. Here, 36 modules, including unassigned GM36 (grey module), were obtained from 14,576 genes and annotated with colour labels to indicate module membership. Branch height represents gene distance; the colour bar indicates module membership. A signed biweight mid-correlation network was used as it resulted in networks with the fewest unassigned genes. (**C**) Bar plot showing all 36 modules and their sizes. The smallest module had 25 genes (GM29) while the largest assigned module contained 2060 genes (GM26). (**D)** Eigengene dendrogram showing module relationships through hierarchical clustering. *Abbreviations*: GM, gene module.



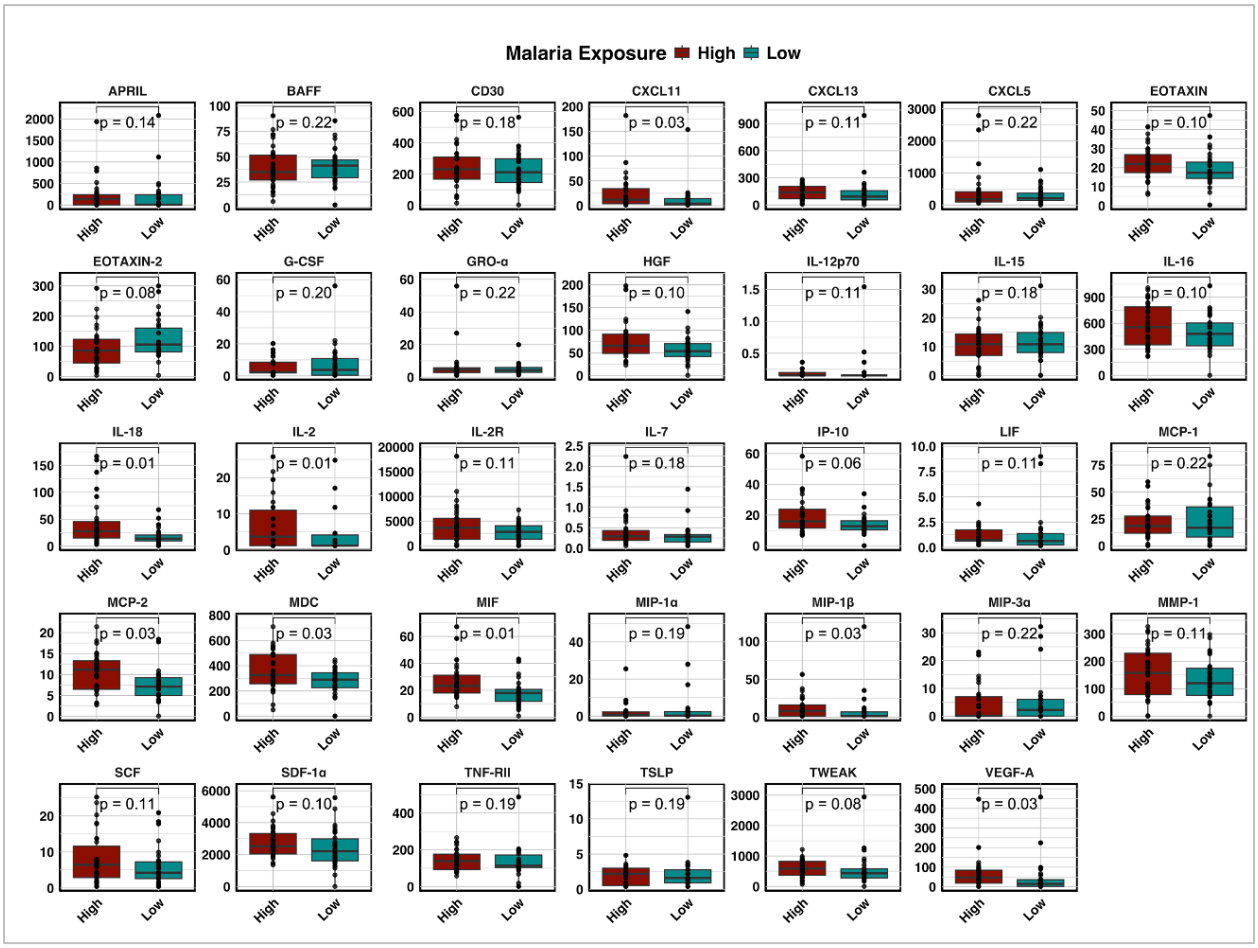
Supplementary Figure 4: Module-trait relationships and hub gene analysis. Scatter plot of absolute effect sizes of gene-malaria exposure beta coefficient (y-axis) versus intramodular connectivity (x-axis) of modules linked to malaria exposure (GM20, GM21, GM17, GM15, GM34, GM30, GM23). The regression line (red coloured with 95% confidence interval bands) highlights correlation strength (*R*) and its significance. Hub genes, labelled red dots, were defined by having the highest connectivity and the strongest gene-malaria exposure effects. Box plots comparing hub gene expression between high and low exposure groups, assessed by two-tailed *t*-tests (parametric) or *Mann-Whitney U* tests (non-parametric) with BH FDR correction ($q < 0.05$) for multiple testing. All modules demonstrated significant positive correlations. Box plots indicate median ± IQR, whiskers = 1.5×IQR and outliers as single points. All identified hub genes in the modules showed significant differences between the two malaria-exposed groups (apart from GM15: CD68). Data shown as median ± interquartile range (IQR). *Abbreviations*: GM, Gene module; | β-coefficient |, absolute beta-coefficient of genes and malaria exposure; R, correlation coefficient.



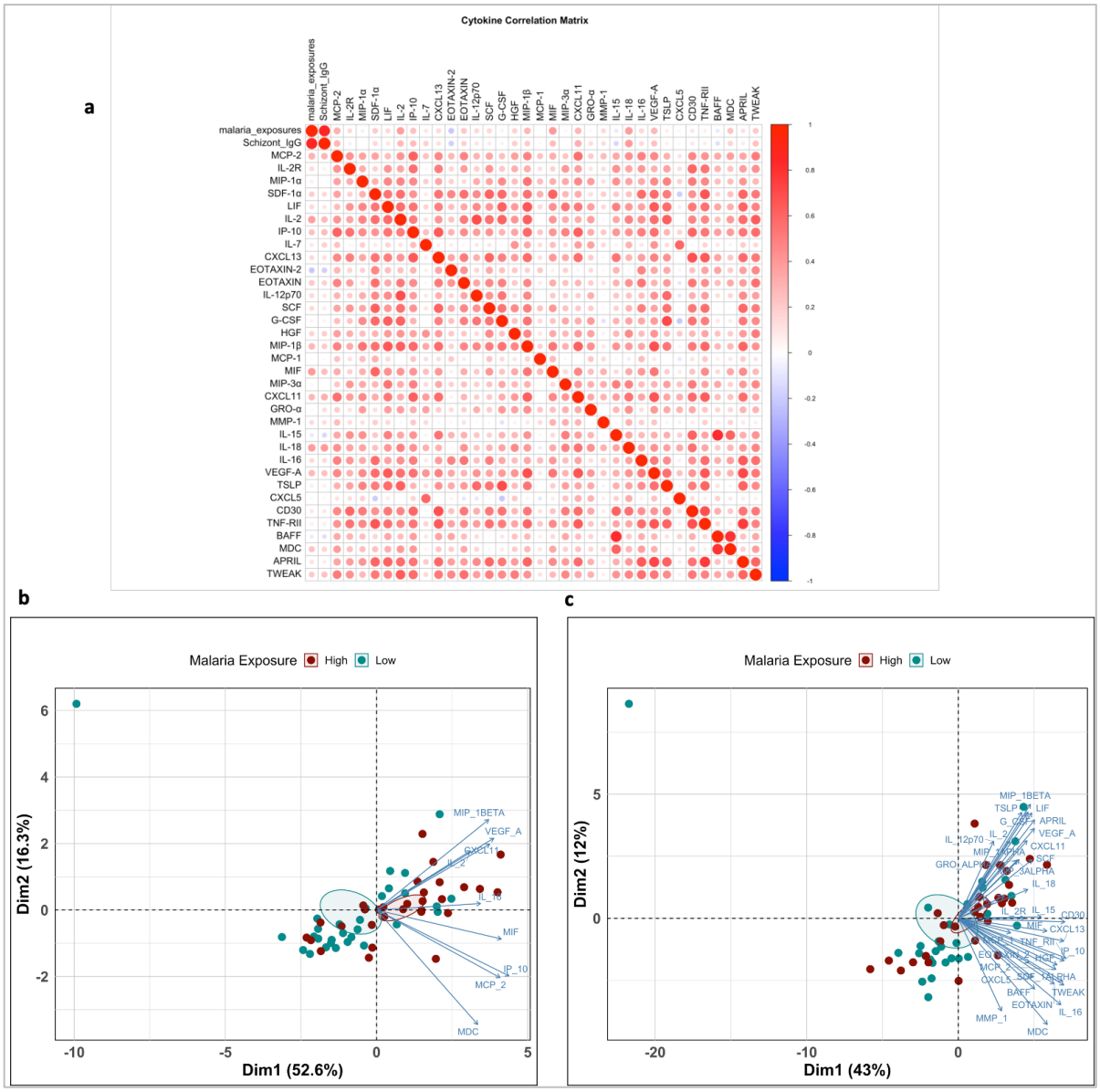
Supplementary Figure 5: Controls performance and soluble markers distribution. Performance evaluation of low (**A**) and high (**B**) controls compared to the reference ranges provided by the manufacturer in the 65 analytes quantified. The x-axis shows the 65 analytes, y-axis shows soluble markers concentration (pg/mL, log-scaled axes). Red dots show the manufacturer's upper analyte reference limit, blue dots represent lower limits, while the black dots represent our obtained concentrations. Some analytes in the high control panel (CXCL9, CXCL11, Eotaxin, Eotaxin-3, G-CSF, GRO alpha, and MCP-2) exceeded upper reference values due to the sigmoid curve plateau effect (expected for high-abundance analytes) [2]. Notably, no plasma concentrations in our study surpassed these limits for these analytes. Control plots were generated using GraphPad Prism (v.10.0.3). (**C**) Log_{10} transformed scale of overview of the 34 soluble markers after filtering out lowly expressed analytes and imputing missing values. Box plots indicate median ± IQR, whiskers = 1.5×IQR and outliers as single points. This visualisation highlights the variability in analyte abundance and provides a foundation for interpreting analyte-related findings in the current study. *Abbreviations*: pg/mL, picograms per millilitre.



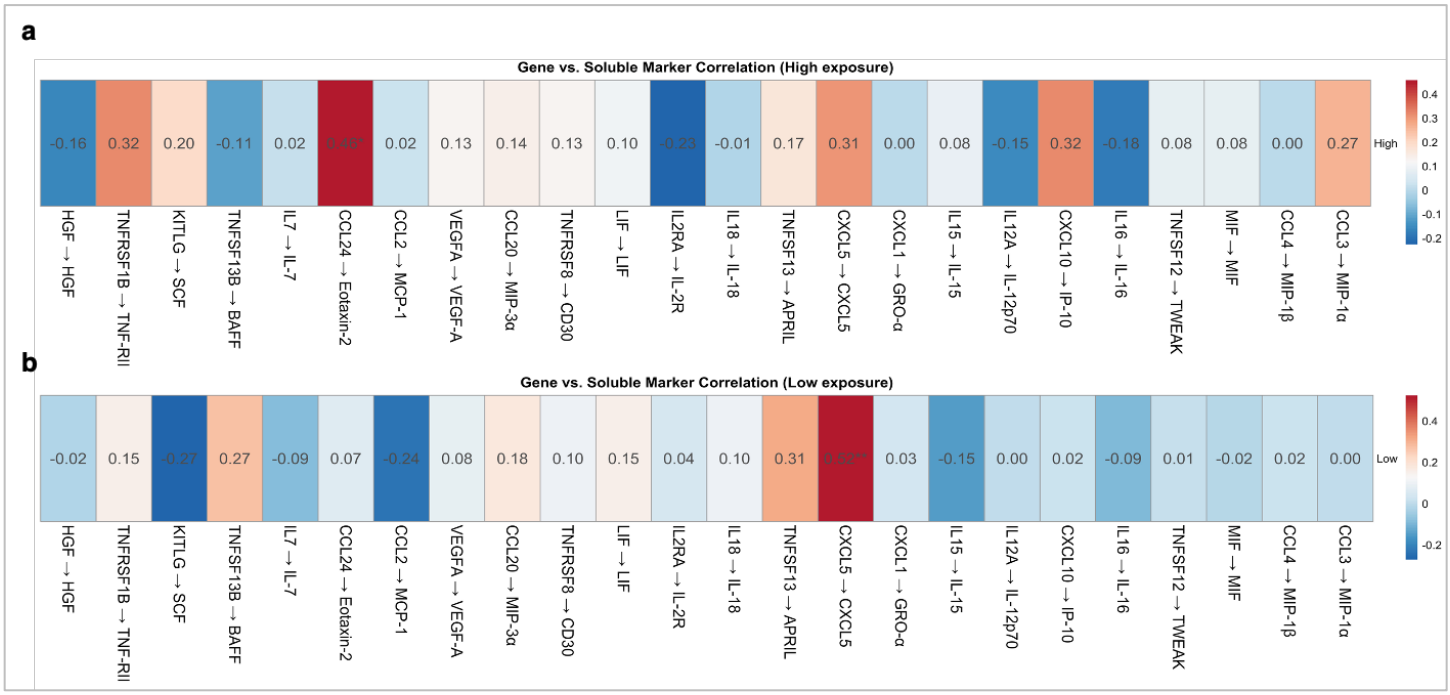
Supplementary Figure 6: Sensitivity analysis of soluble markers data imputation methods for handling missing values and those below the assay's LOD. In LOD/2 imputation, missing values and values below the LOD were imputed with half the LOD value provided by the manufacturer. In LOD/√2, missing values and values below the LOD were imputed with the LOD value divided by the square root of 2 [3]. Lastly, in zero imputation, missing values were replaced with 0, with the assumption that these values were undetectable, while analytes below the LOD but not missing were left unchanged. The analysis showed high similarity among the three imputation methods as demonstrated using (**A**) a PCA plot, (**B**) a heatmap, and (**C**) a correlation plot. LOD/2 imputed values were used in the current study. *Abbreviations*: LOD, limit of detection.



Supplementary Figure 7: Groupwise analysis of soluble markers. Boxplots show concentrations of selected soluble immune markers (pg/mL) comparing individuals with high and low malaria exposure. Statistical comparisons were performed using *Mann-Whitney U* tests (for non-parametric data) or two-tailed *t*-tests (for parametric data), with significance determined using FDR-adjusted *q*-values ($q < 0.05$). Significant differences were observed for IL-18, MIF, IL-2, MDC, MIP-1β, CXCL11, MCP-2, and VEGF-A. Box plots indicate median ± IQR, whiskers = 1.5×IQR and outliers as single points.



Supplementary Figure 8: (**A**) Correlation matrix depicting associations between soluble marker levels, malaria exposure groups, and anti-*P. falciparum* IgG levels. Correlations were calculated using Spearman's or Pearson's method based on data distribution. Colour intensity reflects the strength and direction of the correlations. (**B-C**) Principal component analysis (PCA) plots based on soluble markers expression data: 9 soluble markers in (**B**) and 34 soluble markers in (**C**). Samples are grouped by malaria exposure levels. Axes represent the percentage of variance explained by principal components 1 (x-axis) and 2 (y-axis). Data points are coloured by malaria exposure status, with ellipses indicating 95% confidence intervals for each group. Loadings show the soluble markers' contributions.



Supplementary Figure 9: Gene-protein correlations stratified by malaria exposure. Heatmaps display Spearman correlation coefficients (ρ) between gene expression levels and concentrations of their corresponding soluble proteins in individuals with high (**A**) and low (**B**) malaria exposure (gene → protein). Each cell represents a gene-protein pair, with colour indicating correlation strength and direction (red = positive, blue = negative). Significant correlations were observed for *CCL24* and *CXCL5* genes. Correlation coefficients are shown with asterisks denoting statistical significance ($p < 0.05$; *$p < 0.01$; **$p < 0.001$).

Supplementary Table 1: Baseline characteristics of the study participants (based on region of residence).

Region of residence						
Variable	**Ahero** N = 15[1]	**Junju** N = 15[1]	**Nairobi** N = 15[1]	**Ngerenya** N = 15[1]	**Overall** N = 60[1]	**RIs**
Age (Years)	24 (23, 34)	32 (30, 40)	21 (19, 23)	27 (22, 31)	26 (22, 31)	
Sex						
Female	7 (47%)	3 (20%)	9 (60%)	0 (0%)	19 (32%)	
Male	8 (53%)	12 (80%)	6 (40%)	15 (100%)	41 (68%)	
Complete blood counts						
WBC x10^3/µL	5.00 (4.70, 5.50)	5.33 (4.35, 5.54)	4.52 (4.02, 6.72)	5.50 (4.10, 5.70)	5.20 (4.23, 5.70)	4.0-11.0
Lymphocytes x10^3/µL	2.38 (2.07, 2.47)	2.53 (1.85, 2.88)	2.01 (1.54, 2.69)	2.12 (1.70, 2.91)	2.23 (1.72, 2.68)	1.50-4.00
Monocytes x10^3/µL	0.37 (0.27, 0.69)	0.43 (0.36, 0.49)	0.44 (0.41, 0.55)	0.16 (0.13, 0.21)	0.37 (0.22, 0.49)	0.20-0.80
Neutrophils x10^3/µL	2.00 (1.72, 2.53)	1.88 (1.52, 2.17)	2.01 (1.76, 3.90)	2.43 (1.69, 3.33)	2.01 (1.70, 2.82)	2.00-7.50
Eosinophils x10^3/µL	0.11 (0.06, 0.23)	0.15 (0.10, 0.21)	0.12 (0.06, 0.19)	0.16 (0.09, 0.34)	0.13 (0.08, 0.23)	0.04-0.40
Monocyte/Lymphocyte ratio	0.06 (0.05, 0.09)	0.02 (0.01, 0.02)	0.02 (0.01, 0.02)	0.03 (0.02, 0.09)	0.02 (0.02, 0.05)	
Neutrophils/Lymphocyte ratio	0.19 (0.15, 0.30)	0.19 (0.13, 0.22)	0.22 (0.18, 0.28)	0.07 (0.06, 0.10)	0.17 (0.11, 0.23)	
Basophils x10^3/µL	0.83 (0.76, 1.21)	0.79 (0.52, 0.99)	1.35 (0.99, 1.71)	1.01 (0.78, 1.43)	0.92 (0.77, 1.36)	0.02-0.10
RBCs x10^6/µL	5.18 (4.70, 5.46)	5.01 (4.70, 6.00)	5.22 (4.94, 5.49)	5.56 (5.10, 6.17)	5.23 (4.81, 5.68)	4.50-6.50
HGB(g/dL)	14.60 (12.90, 15.80)	14.15 (13.70, 14.42)	14.60 (13.90, 15.70)	15.00 (14.10, 15.70)	14.41 (13.75, 15.35)	13.0-18.0
Hematocrit(%)	42.7 (38.1, 46.7)	42.1 (41.4, 44.1)	42.3 (39.7, 44.9)	45.5 (43.1, 48.5)	43.2 (40.7, 46.2)	40.0-54.0
MCV(fL)	83 (80, 90)	85 (74, 88)	81 (79, 85)	85 (73, 88)	83 (79, 87)	76-96
MCH(pg)	28.10 (27.40, 31.10)	28.10 (23.70, 30.00)	28.20 (27.20, 29.60)	28.20 (23.20, 28.90)	28.15 (26.55, 29.50)	27.0-32.0
MCHC(pg)	34.20 (33.80, 34.40)	32.60 (32.30, 33.70)	34.70 (34.30, 35.60)	33.00 (31.70, 33.30)	33.75 (32.90, 34.40)	31.0-35.0
RDW(%)	12.50 (12.30, 13.00)	15.00 (14.20, 16.20)	13.50 (13.20, 14.20)	12.10 (11.70, 13.70)	13.40 (12.50, 14.40)	11.0-16.0
Platelets x10^3/µL	239 (198, 279)	207 (177, 254)	258 (232, 328)	212 (194, 244)	229 (198, 267)	150-400
MPV(fL)	8.10 (7.80, 8.40)	9.10 (8.23, 9.92)	9.60 (9.10, 9.90)	9.00 (8.30, 9.70)	8.86 (8.20, 9.70)	6.0-10.0

Abbreviations: WBC- White blood cells; RBC-Red blood cells; HGB-Haemoglobin; MCV-Mean Cell Volume; MCH-Mean Cell Volume; MCHC; Mean Cell Haemoglobin Concentration; RDW-Red Cell Width; MPV-Mean Platelet Volume; CBC- Complete blood count; RIs-Reference intervals

Supplementary Table 2: Linear regression analysis of factors associated with anti-*P. falciparum* IgG antibody levels. The table presents the estimated effect sizes (Estimate), standard errors (Std. Error), *t*-values, and *p*-values (Pr(>|t|)) for each predictor variable. High malaria exposure is significantly associated with increased anti-*P. falciparum* IgG levels with individuals from high-exposure regions showing a 4.6 increase in $\log_2$-IgG (equivalent to a 24.6-fold increase [back transformed estimate]) compared to those from low-exposure regions ($P = 7.53 \times 10^{-19}$). Sex and age do not exhibit a significant effect on antibody levels.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.22905312	0.653749519	17.17638453	5.52E-24
malaria_exposureHigh	4.622637973	0.349909852	13.21093974	7.53E-19
Age	0.019857151	0.024373404	0.814705694	0.418692068
SexMale	0.430735388	0.350726646	1.228122794	0.224537743

Supplementary Table 3: Baseline characteristics of the study participants (based on levels of malaria exposure)

Malaria Exposure					
Variable	**High** N = 30[1]	**Low** N = 30[1]	***P*-value**[2]	**Overall** N = 60[1]	**RIs**
Age (Years)	30 (24, 37)	23 (21, 27)	0.002	26 (22, 31)	
Sex			0.781		
Female	10 (33%)	9 (30%)		19 (32%)	
Male	20 (67%)	21 (70%)		41 (68%)	
Complete blood counts					
WBC x10^3/µL	5.07 (4.59, 5.50)	5.32 (4.07, 6.50)	>0.999	5.20 (4.23, 5.70)	4.0-11.0
Lymphocytes x10^3/µL	2.39 (2.00, 2.61)	2.02 (1.67, 2.69)	0.137	2.23 (1.72, 2.68)	1.50-4.00
Monocytes x10^3/µL	0.41 (0.33, 0.56)	0.26 (0.16, 0.44)	0.006	0.37 (0.22, 0.49)	0.20-0.80
Neutrophils x10^3/µL	1.94 (1.68, 2.27)	2.37 (1.76, 3.45)	0.047	2.01 (1.70, 2.82)	2.00-7.50
Eosinophils x10^3/µL	0.12 (0.07, 0.21)	0.14 (0.08, 0.27)	0.894	0.13 (0.08, 0.23)	0.04-0.40
Basophils x10^3/µL	0.04 (0.02, 0.06)	0.02 (0.01, 0.04)	0.103	0.02 (0.02, 0.05)	
Monocyte/Lymphocyte ratio	0.19 (0.15, 0.24)	0.15 (0.07, 0.22)	0.07	0.17 (0.11, 0.23)	
Neutrophils/Lymphocyte ratio	0.81 (0.69, 1.19)	1.28 (0.88, 1.59)	<0.001	0.92 (0.77, 1.36)	0.02-0.10
RBCs x10^6/µL	5.02 (4.70, 5.69)	5.30 (4.99, 5.67)	0.135	5.23 (4.81, 5.68)	4.50-6.50
HGB(g/dL)	14.18 (13.50, 14.98)	14.80 (13.90, 15.70)	0.107	14.41 (13.75, 15.35)	13.0-18.0
Hematocrit(%)	42.4 (40.6, 44.5)	44.0 (41.0, 46.5)	0.237	43.2 (40.7, 46.2)	40.0-54.0
MCV (fL)	83 (79, 88)	82 (79, 86)	0.363	83 (79, 87)	76-96
MCH (pg)	28.10 (26.50, 30.20)	28.20 (26.60, 29.30)	0.779	28.15 (26.55, 29.50)	27.0-32.0
MCHC (pg)	33.70 (32.60, 34.30)	33.85 (33.00, 34.70)	0.391	33.75 (32.90, 34.40)	31.0-35.0
RDW (%)	13.40 (12.50, 15.00)	13.25 (12.10, 14.10)	0.19	13.40 (12.50, 14.40)	11.0-16.0
Platelets x10^3/µL	217 (188, 266)	235 (208, 285)	0.169	229 (198, 267)	150-400
MPV (fL)	8.35 (7.90, 9.10)	9.45 (8.80, 9.90)	<0.001	8.86 (8.20, 9.70)	6.0-10.0

[1]Median (IQR) or Frequency (%); [2]Wilcoxon rank sum test; Pearson's Chi-squared test; Wilcoxon rank sum exact test. *Abbreviations*: WBC- White blood cells; RBC-Red blood cells; HGB-Hemoglobin; MCV-Mean Cell Volume; MCH-Mean Cell Volume; MCHC; Mean Cell Haemoglobin Concentration; RDW-Red Cell Width; MPV-Mean Platelet Volume; CBC- Complete blood count; RIs-Reference intervals.

Supplementary Table 4: WGCNA pathway enrichment analysis: Summary of biological pathways enriched in gene modules identified through WGCNA, stratified by their association with malaria exposure. Under each pathway category, pathway names are listed, as provided by the databases, with additional related pathways from each database shown under the "slash". Modules positively associated with malaria exposure are listed in the upper section, while negatively associated modules are in the lower section. Benjamini-Hochberg adjusted FDR values are provided for the first listed pathway name, which had the smallest FDR, while non-significance (ns) is indicated where a pathway did not reach significance. Pathway enrichment analysis was assessed using STRING and WebGestalt.

Modules	Size	Biological pathways	FDR	Databases
Modules positively associated with malaria exposure				
GM20	197	RNA splicing/mRNA splicing, via spliceosome/mRNA processing/ Regulation of mRNA processing/Gene expression	1.12E-12	STRING
		mRNA processing/RNA splicing/regulation of mRNA processing/regulation of mRNA splicing, via spliceosome	2.28E-13	WebGestalt
GM21	458	Cellular nitrogen compound metabolic process/Cellular metabolic process/RNA metabolic process/Gene expression/Cellular process	6.26E-11	STRING
		RNA processing/RNA splicing/organonitrogen compound biosynthetic process	0.001	WebGestalt
GM17	53	B cell activation/Regulation of B cell activation/Proliferation/Cell surface receptor signalling pathway	7.88E-07	STRING
		B cell receptor signalling pathway/B cell activation/proliferation/regulation of B cell activation	2.15E-08	WebGestalt
GM15	271	no enrichment	na	STRING
		snRNA metabolic process/regulation of DNA repair/ ncRNA processing	ns	WebGestalt
Modules negatively associated with Malaria exposure				
GM23	137	no enrichment	na	STRING
		Protein dephosphorylation/positive regulation of fatty acid biosynthetic process	ns	WebGestalt
GM34	34	no enrichment	na	STRING
		cellular response to cytokine stimulus/response to cytokine	ns	WebGestalt
GM30	196	Cytokine-mediated signalling pathway/Inflammatory response/Neutrophil chemotaxis/Leukocyte chemotaxis	3.46E-10	STRING
		Neutrophil migration/cytokine-mediated signalling pathway/inflammatory response	1.0651E-05	WebGestalt

*Homo sapiens served as the reference background for calculating fold enrichment for pathway enrichment analysis;

Supplementary Table 5: Soluble markers correlation analysis with anti-*P. falciparum* IgG antibodies. Significant correlations were observed for MCP-2, IL-2, 1P-10, MIF, CXCL11 and IL-18. Spearman's test was used to assess the correlations.

Cytokine	Correlation (ρ)	*P*-value
MCP-2	0.282066245	0.029
IL-2R	0.069746126	0.5964
MIP-1α	-0.007969975	0.9518
SDF-1α	0.157129163	0.2305
LIF	0.103593199	0.4309
IL-2	0.2768629	0.0322
IP-10	0.26623873	0.0398
IL-7	0.186976555	0.1526
CXCL13	0.123648596	0.3466
EOTAXIN-2	-0.18956701	0.1469
EOTAXIN	0.185890133	0.155
IL-12p70	0.105881466	0.4207
SCF	0.087264998	0.5073
G-CSF	0.054976701	0.6765
HGF	0.190265795	0.1454
MIP-1β	0.207720509	0.1113
MCP-1	-0.027125805	0.837
MIF	0.260277589	0.0446
MIP-3α	-0.001559327	0.9906
CXCL11	0.315489638	0.0141
GRO-α	0.138575992	0.291
MMP-1	0.130928895	0.3187
IL-15	-0.153948979	0.2402
IL-18	0.39071708	0.002
IL-16	0.146376026	0.2644
VEGF-A	0.245308965	0.0589
TSLP	0.067331927	0.6092
CXCL5	0.039150308	0.7665
CD30	0.090953301	0.4895
TNF-RII	0.079523194	0.5459
BAFF	-0.12093592	0.3573
MDC	0.178718793	0.1719
APRIL	0.079121412	0.5479
TWEAK	0.24193078	0.0626

References

1. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
2. Ellington, A. A., Kullo, I. J., Bailey, K. R. & Klee, G. G. Antibody-based protein multiplex platforms: technical and operational challenges. *Clin. Chem.* **56**, 186–193 (2010).
3. Verbovšek, T. A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. *RMZ - Mater. Geoenvironment* **58**, 393–404 (2011).