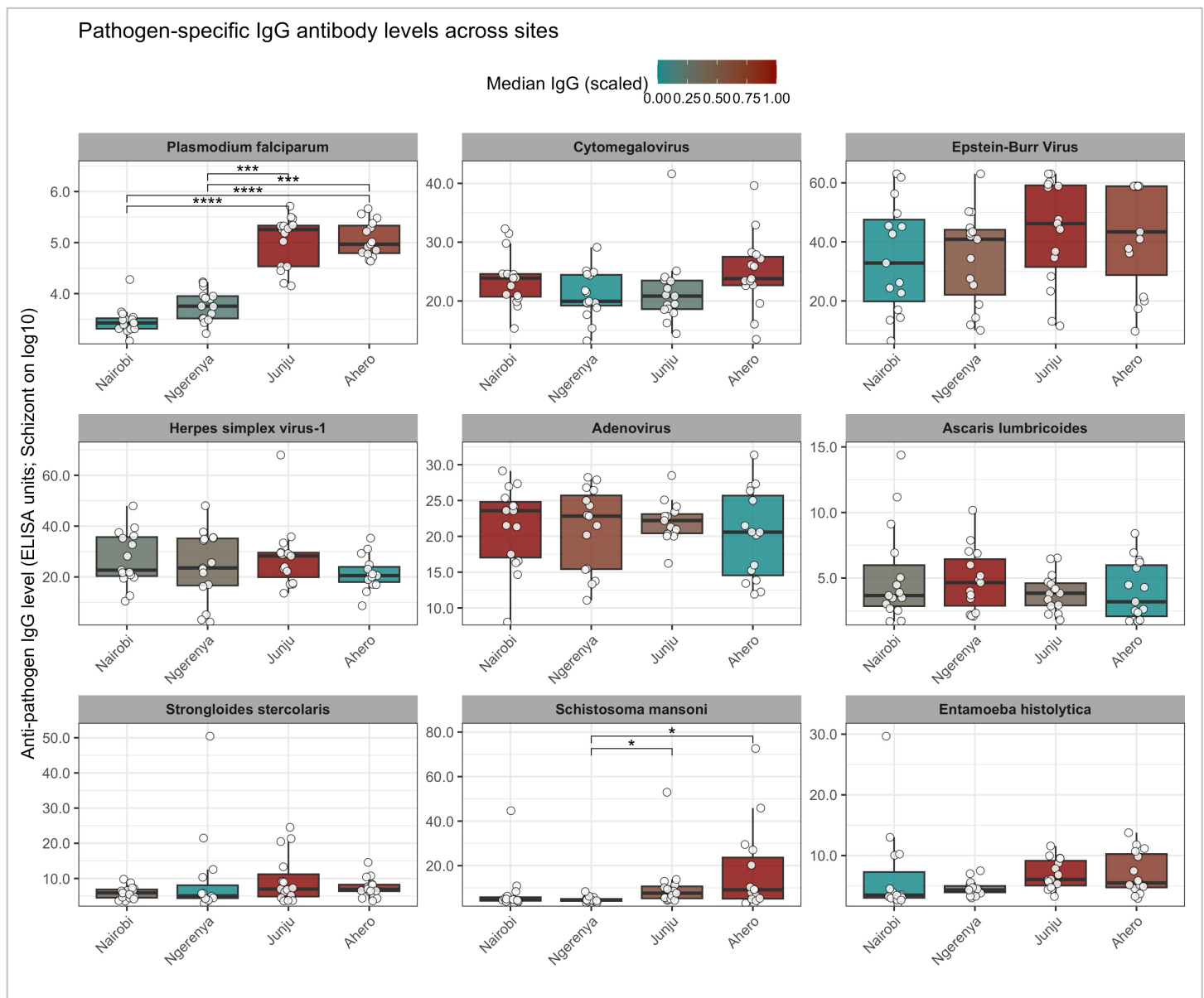


1 Long-term malaria exposure in pathogen-endemic environments is associated with distinct immune states



2

3 **Supplementary Figure 1. Pathogen-specific IgG antibody levels across the four study regions.** Boxplots show total IgG

4 responses to nine pathogens in participants from four Kenyan regions ($n = 15$ biologically independent samples per site).

5 The y-axis represents antibody levels in raw standard ELISA units (except *P. falciparum*, plotted on a log10 scale). Boxes

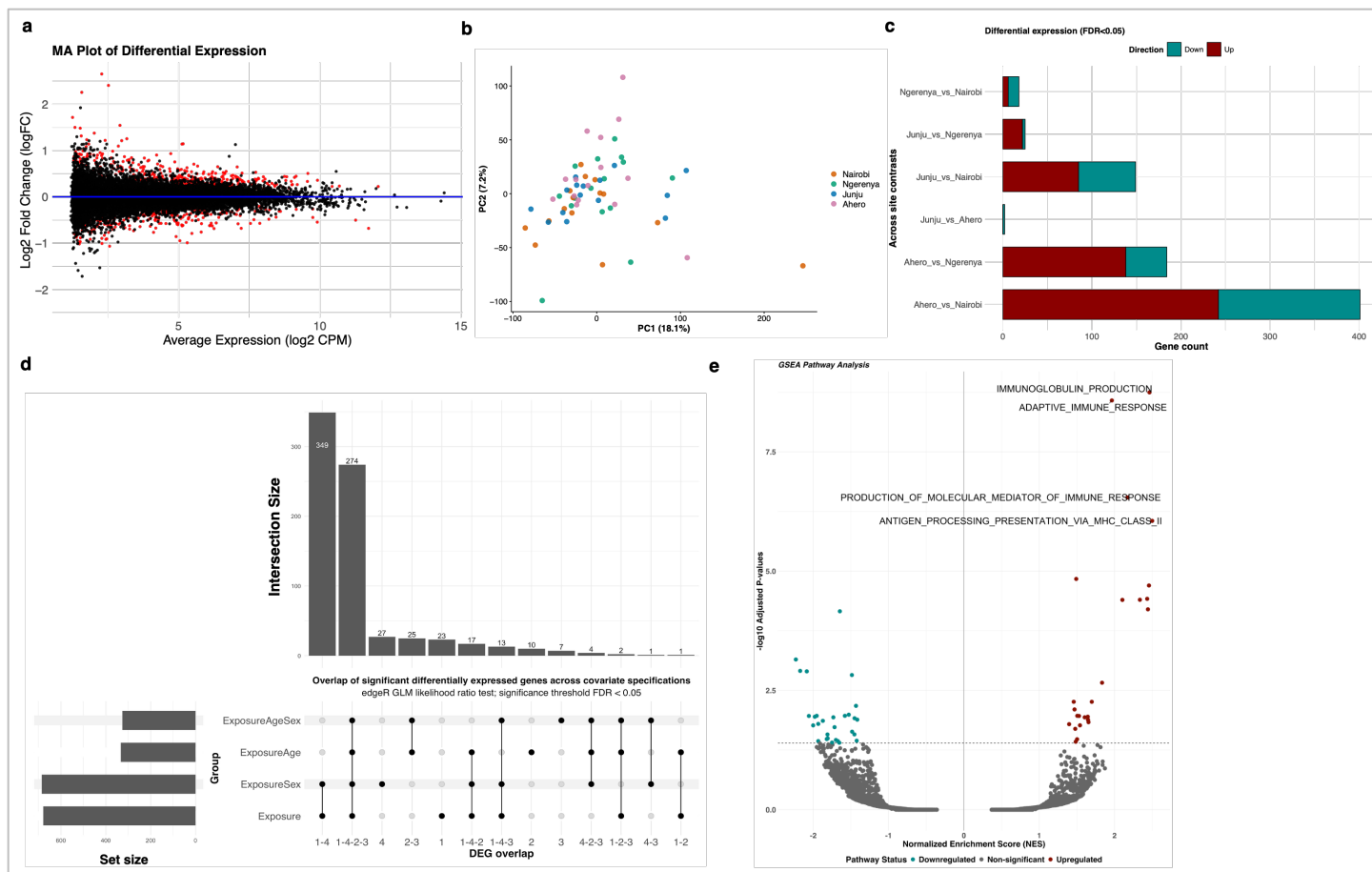
6 represent the interquartile range with median lines; whiskers denote $1.5 \times \text{IQR}$; outliers are plotted individually. The colour

7 gradient from dark cyan to dark red reflects the relative median IgG scale shown in the legend. Pairwise comparisons

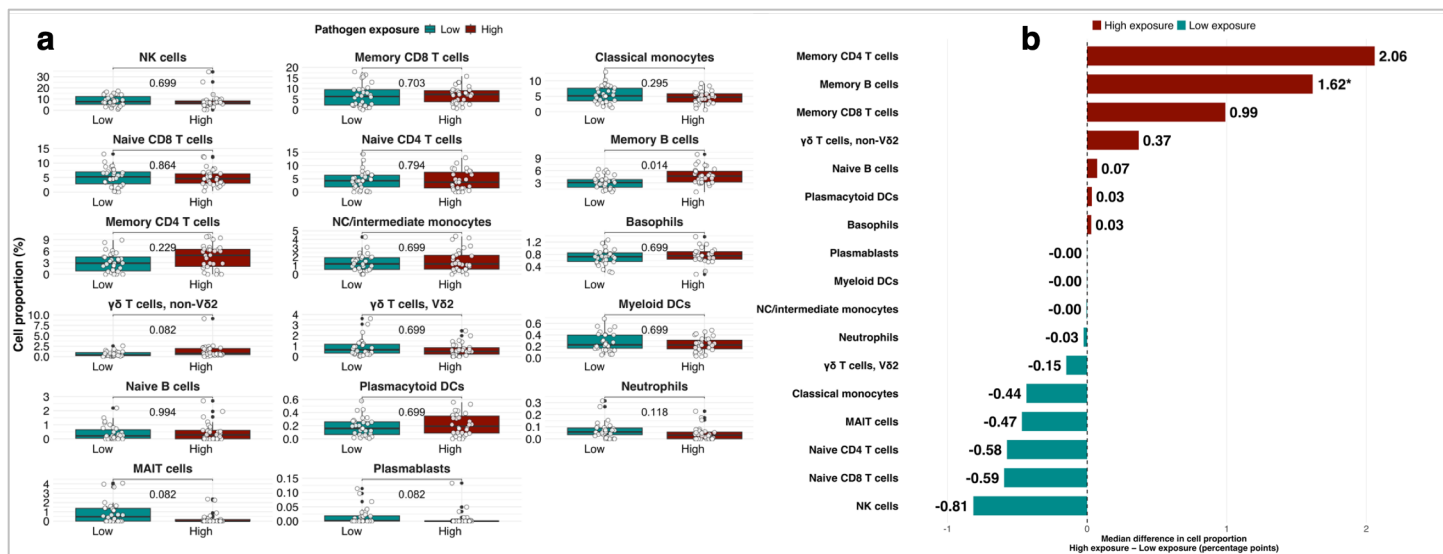
8 between regions were assessed using two-tailed Dunn's tests with Benjamini-Hochberg adjustment, and significance levels

9 are indicated by asterisks ($\text{FDR} < 0.05 = (*)$, $< 0.01 = (**)$, $< 0.001 = (***)$). Only *P. falciparum* and *S. mansoni* showed

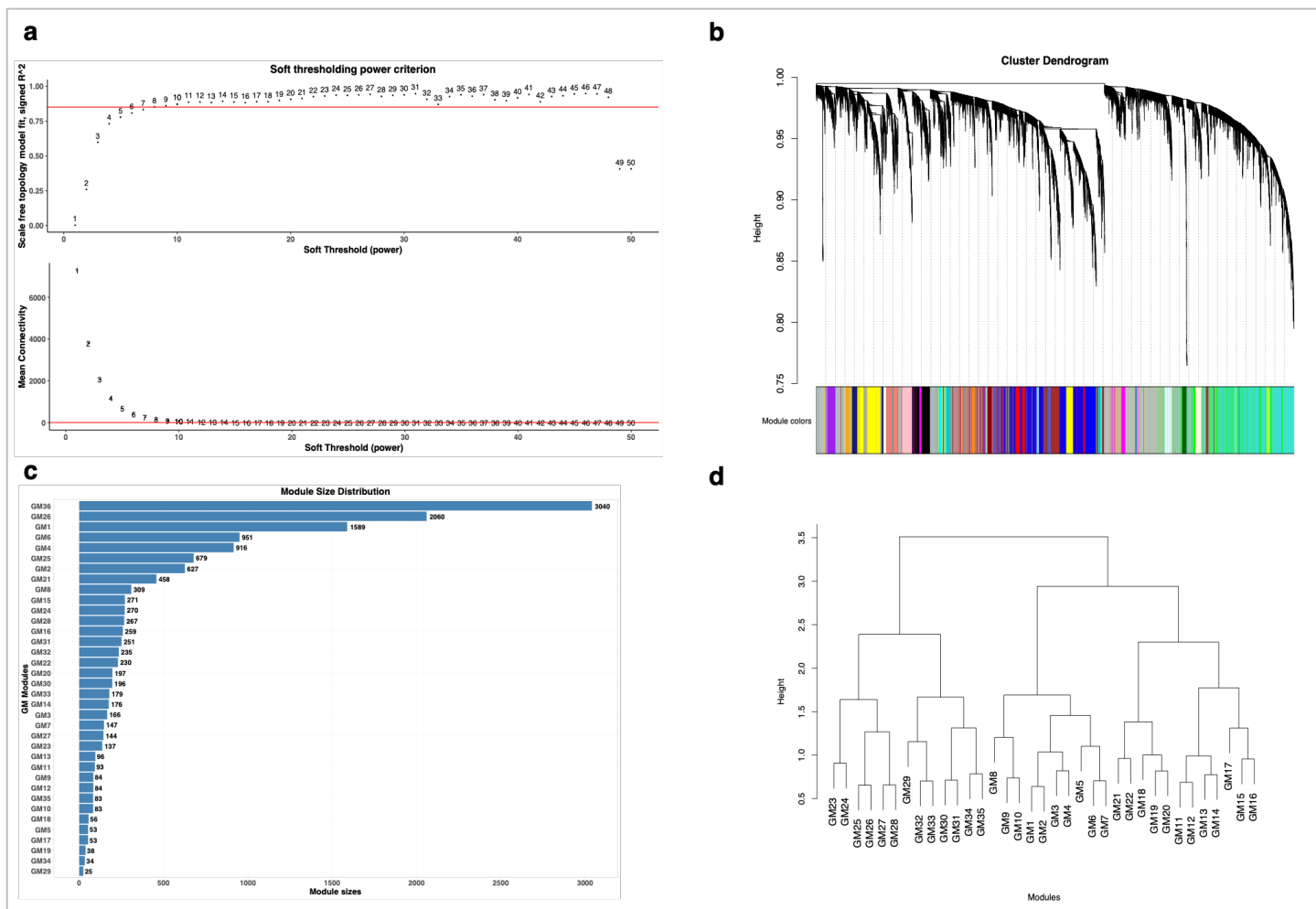
10 statistically significant regional differences.



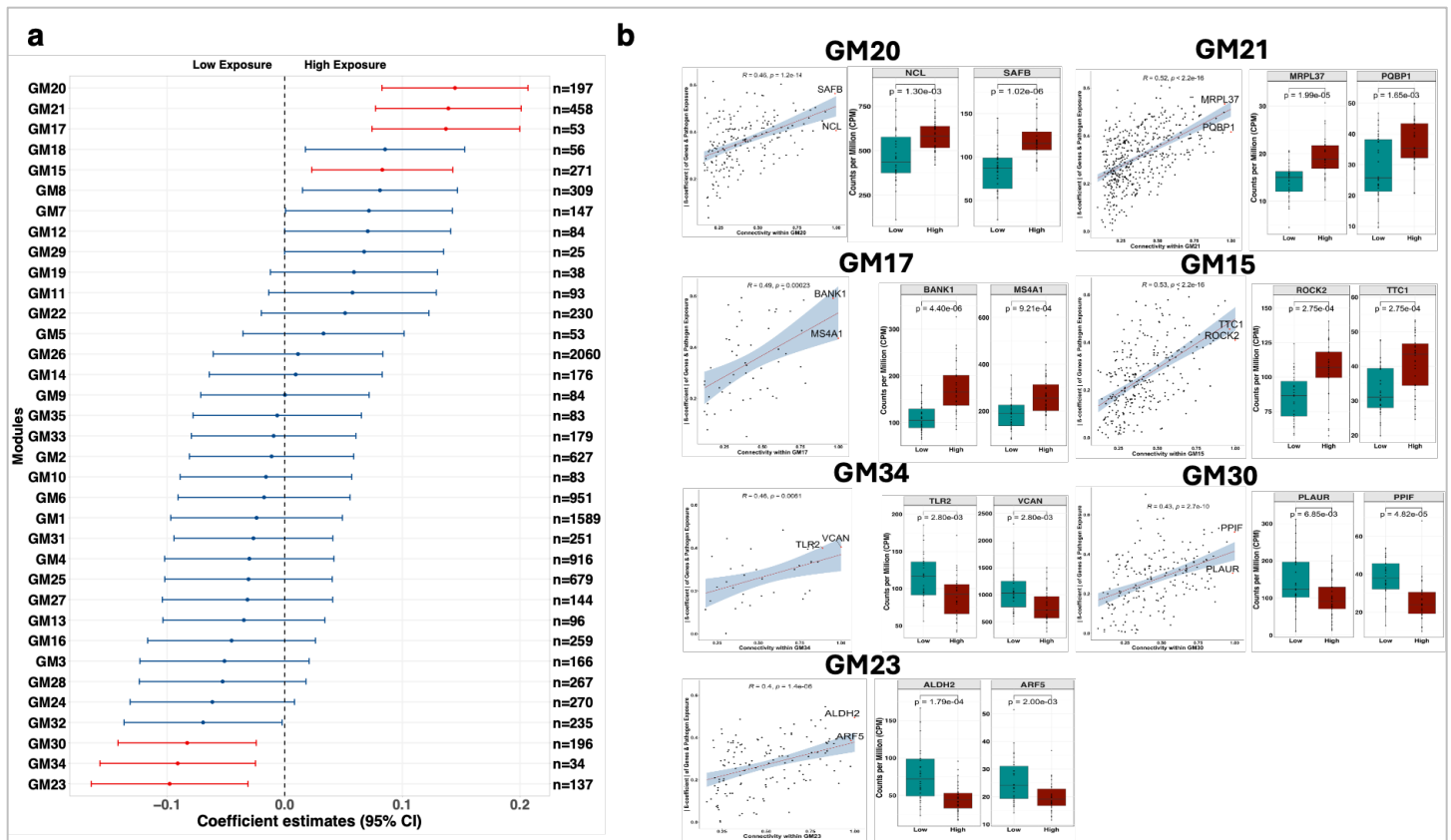
Supplementary Figure 2. Transcriptomic variation and differential gene expression across study regions. (A) MA plot showing each gene as a point, with average expression ($\log_2$ CPM) on the x-axis and $\log_2$ fold change on the y-axis. Genes significantly upregulated or downregulated at $FDR < 0.05$ are highlighted in red. (B) Principal component analysis (PCA) of global PBMC transcriptomes based on normalized expression values from all expressed genes. Samples are coloured according to recruitment site (Nairobi, Ngerenya, Junju, and Ahero). (C) Summary of differential expression across site-to-site contrasts, showing the number of upregulated (dark red) and downregulated (dark cyan) genes per contrast at $FDR < 0.05$. (D) UpSet plot showing the overlap of significantly differentially expressed genes identified using *edgeR* GLM likelihood ratio tests ($FDR < 0.05$) under alternative covariate specifications. Differential expression was assessed using a primary model including pathogen exposure, and additional models adjusting for sex, age, or both age and sex. Horizontal bars indicate the number of genes in each intersection, while connected dots denote the covariate specifications in which genes were detected. Set size bars indicate the total number of significant genes identified in each model. (E) Volcano plot of enriched biological processes from gene set enrichment analysis (GSEA), based on all expressed genes ranked by $\log_2$ FC. Significantly upregulated pathways (positive normalised enrichment score [NES]) are shown in dark red; downregulated pathways (negative NES) in dark cyan; non-significant pathways in grey. The y-axis represents $-\log_{10}$ (FDR), and the horizontal dashed line denotes the significance threshold ($FDR < 0.05$). *Abbreviations:* CPM counts per million; MA, M (log ratio) vs A (mean average) plot; DEG, differentially expressed gene.



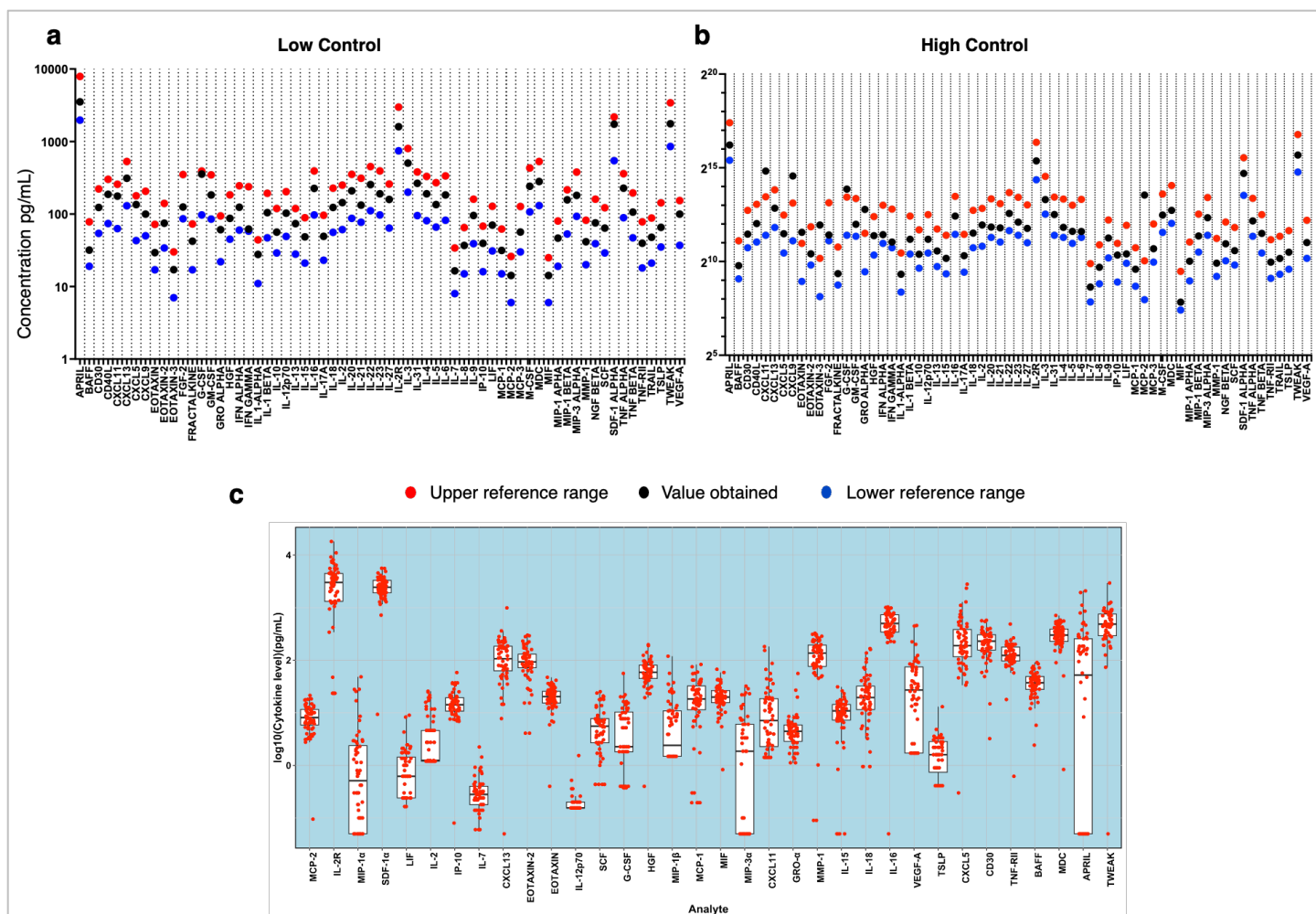
Supplementary Figure 3. PBMC-specific cellular deconvolution using ABIS. (A) Estimated proportions of 17 immune cell populations inferred from PBMC RNA-seq using the ABIS deconvolution framework¹ in individuals from low- and high-pathogen-exposure settings ($n = 30$ biologically independent samples per exposure group). Boxes represent the median and interquartile range (IQR), whiskers extend to $1.5 \times$ IQR, and points represent individual participants. Group comparisons were performed using two-sided t-tests for parametric data or two-sided Wilcoxon rank-sum tests for non-parametric data, with Benjamini-Hochberg adjustment for multiple comparisons. **(B)** Median differences in estimated cell proportions (high minus low pathogen exposure; percentage points; $n = 30$ biologically independent samples per exposure group), ranked by effect size to facilitate comparison of the direction and magnitude of immune cell differences between exposure groups. Positive values indicate higher estimated proportions in the high-exposure group, whereas negative values indicate lower estimated proportions. Bars marked with * indicate statistically significant differences after Benjamini-Hochberg correction (adjusted $P < 0.05$).



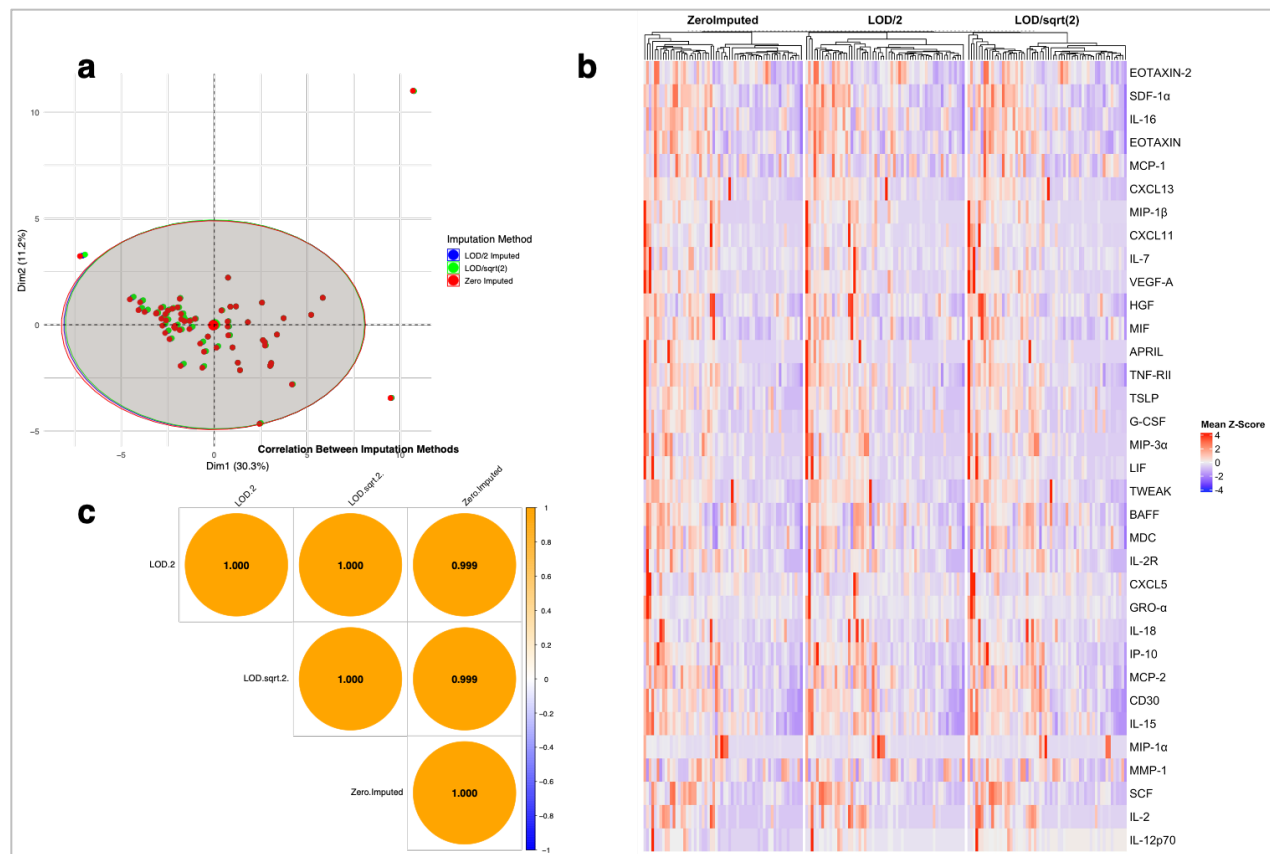
Supplementary Figure 4. Weighted gene co-expression network analysis (WGCNA) construction. (A) Selection of the soft-thresholding power used for network construction. A power of 9 was selected because it achieved a scale-free topology fit of $R^2 = 0.85$ while maintaining low mean connectivity. Red horizontal lines indicate the selected thresholds. **(B)** Hierarchical clustering dendrogram of genes with module assignments. A signed weighted gene co-expression network was constructed using biweight midcorrelation (bicor), with a soft-thresholding power of $\beta = 9$. A total of 35 assigned modules plus the unassigned grey module (GM36) were obtained from 14,576 genes. Branch height represents gene dissimilarity, and the colour bar indicates module membership. **(C)** Bar plot showing the sizes of all 35 assigned modules and the unassigned grey module. The smallest assigned module contained 25 genes (GM29), while the largest assigned module contained 2,060 genes (GM26). **(D)** Eigengene dendrogram showing relationships among modules based on hierarchical clustering. *Abbreviations:* GM, gene module.



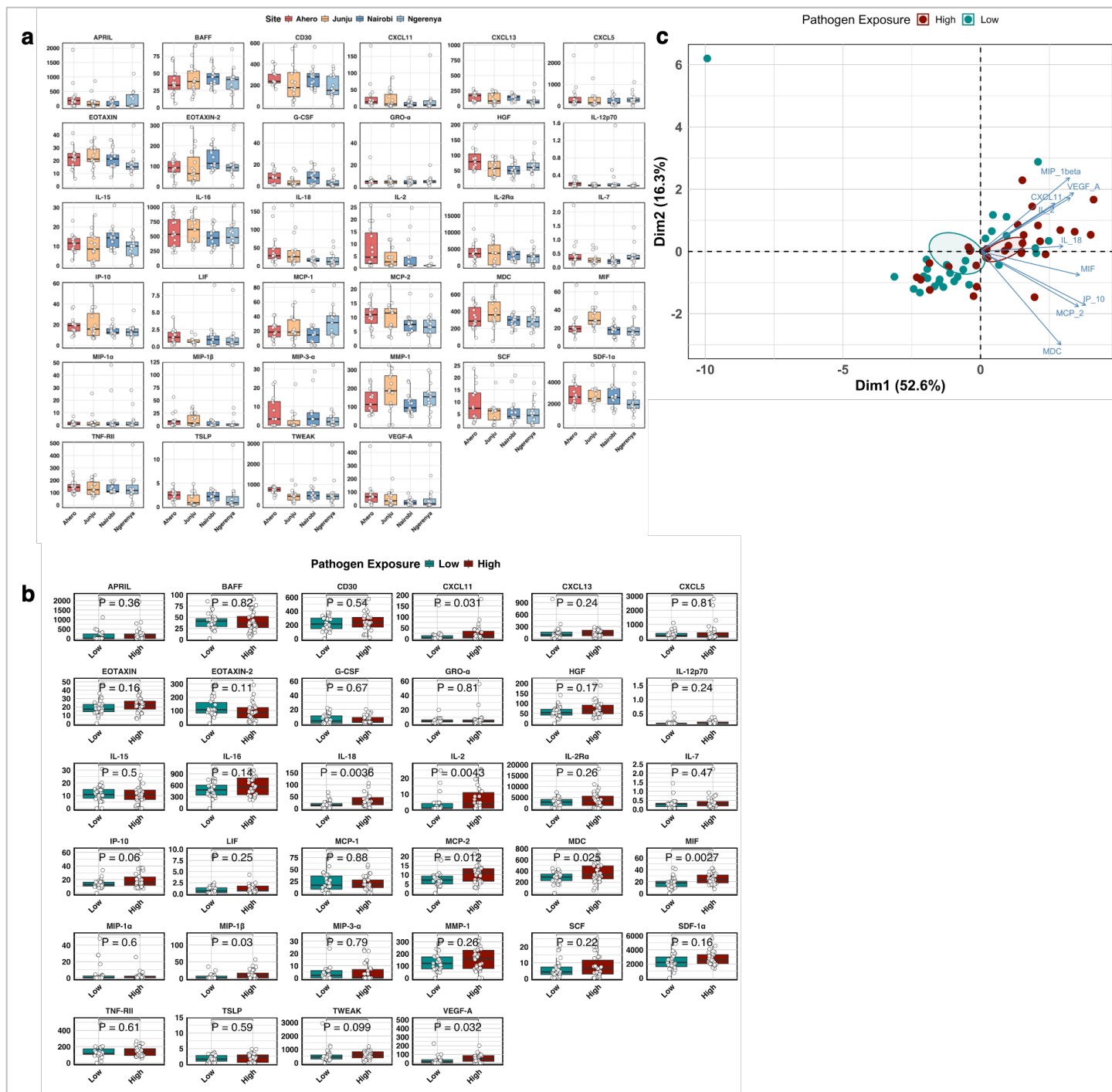
Supplementary Figure 5. Covariate-adjusted module-trait associations and hub gene characteristics in pathogen exposure. (A) Forest plot showing beta coefficients for the association between pathogen exposure and 35 gene module eigengenes, estimated using linear regression adjusted for age and sex (low exposure as the reference group; $n = 60$ biologically independent samples). Points represent the estimated beta coefficients for high versus low pathogen exposure, and error bars represent 95% confidence intervals (CIs). The right y-axis indicates the number of genes within each module. P values were adjusted for multiple testing across modules using the Benjamini-Hochberg method, with significant associations defined as $FDR < 0.05$. (B) Scatter plots of absolute gene-pathogen exposure beta coefficients (y-axis) versus intramodular connectivity (x-axis) for modules associated with pathogen exposure. Associations between intramodular connectivity and absolute gene-pathogen exposure effect sizes were assessed using two-sided Pearson correlation tests. The red line represents the fitted linear regression, and the shaded band represents the 95% confidence interval around the fitted regression line. The correlation coefficient (R) and P value are shown. Hub genes, shown as red points, were defined as genes with the highest intramodular connectivity and strongest gene-pathogen exposure effects. Box plots compare hub gene expression between high- and low-exposure groups ($n = 30$ biologically independent samples per exposure group) using two-sided t -tests for parametric data or two-sided Wilcoxon rank-sum tests for non-parametric data, with Benjamini-Hochberg FDR correction for multiple testing ($q < 0.05$). Box plots indicate the median and IQR, whiskers extend to $1.5 \times IQR$, and outliers are shown as individual points.



Supplementary Figure 6. Control performance and soluble marker distributions. (A, B) Performance evaluation of low (A) and high (B) controls for the 65 quantified analytes compared with the reference ranges provided by the manufacturer. The x-axis shows the 65 analytes, and the y-axis shows soluble marker concentrations (pg/mL; log scale). Red dots indicate the manufacturer's upper reference limits, blue dots indicate the lower reference limits, and black dots represent the measured control concentrations. Some analytes in the high-control panel (CXCL9, CXCL11, Eotaxin, Eotaxin-3, G-CSF, GRO alpha, and MCP-2) exceeded the upper reference limits because of the sigmoid curve plateau effect expected for high-abundance analytes². No plasma concentrations in the study exceeded these limits for these analytes. Control plots were generated using GraphPad Prism (v.10.0.3). **(C)** Distribution of the 34 soluble markers retained after filtering lowly expressed analytes and imputing missing values ($n = 60$ biologically independent samples). Concentrations are shown on a log10-transformed scale. Box plots indicate the median and interquartile range (IQR), whiskers extend to $1.5 \times \text{IQR}$, and outliers are shown as individual points. *Abbreviations:* pg/mL, picograms per millilitre.



Supplementary Figure 7. 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 2³. 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 8. Groupwise analysis of soluble immune markers across pathogen exposure sites. (A) Boxplots showing concentrations (pg/mL) of selected soluble immune mediators across the four study sites ($n = 15$ biologically independent samples per site), illustrating site-level variation in marker abundance. **(B)** The same mediators grouped by pathogen exposure category ($n = 30$ biologically independent samples per exposure group). Statistical comparisons were performed using Wilcoxon rank-sum tests for non-parametric data or two-tailed t-tests for parametric data, with significance assessed using FDR-adjusted q values ($q < 0.05$). Significant differences were observed for IL-18, MIF, IL-2, MDC, MIP-1 β , CXCL11, MCP-2, and VEGF-A. In both panels, boxes represent the interquartile range (IQR), centre lines indicate the median, whiskers extend to $1.5 \times$ IQR, and individual points represent participants. **(C)** Principal component analysis (PCA) of soluble marker profiles based on nine analytes ($n = 60$ biologically independent samples). Points are coloured by pathogen exposure status, with ellipses indicating 95% confidence intervals; axes indicate the percentage of variance explained by PC1 and PC2. Loadings indicate each marker's contribution to the principal components.

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

Variable	Region of residence				Overall N = 60	RIs
	Ahero N = 15	Junju N = 15	Nairobi N = 15	Ngerenya N = 15		
Age (Years)	24 (23, 34)	32 (30, 40)	21 (19, 23)	27 (22, 31)	26 (22, 31)	
Sex						
<i>Female</i>	7 (47%)	3 (20%)	9 (60%)	0 (0%)	19 (32%)	
<i>Male</i>	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
Basophils x10 ³ /μL	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)	0.02-0.10
Monocyte/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)	
Neutrophils/Lymphocyte ratio	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)	
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 (g/dL)	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
Neutrophil-to-lymphocyte ratio (NLR) and monocyte-to-lymphocyte ratio were calculated from absolute cell counts obtained from the complete blood count. <i>Abbreviations:</i> WBC- White blood cells; RBC-Red blood cells; HGB-Haemoglobin; MCV-Mean Cell Volume; MCH-Mean Cell Haemoglobin; MCHC; Mean Cell Haemoglobin Concentration; RDW-Red Cell Distribution 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 estimated regression coefficients (Estimate), standard errors (Std. Error), *t*-values, and *P* values [*Pr*(>|*t*|)] for each predictor variable. Statistical significance of individual regression coefficients was assessed using two-sided *t*-tests. High malaria exposure was significantly associated with increased anti-*P. falciparum* IgG levels, with individuals from high-exposure regions showing a 4.6-unit increase in log₂-IgG (equivalent to a 24.6-fold increase after back-transformation) compared with those from low-exposure regions (*P* = 7.53 × 10⁻¹⁹). Sex and age were not significantly associated with antibody levels.

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

Supplementary Table 3. Baseline characteristics of the study participants (based on levels of pathogen exposure)

	Pathogen Exposure				
Variable	High N = 30 ¹	Low N = 30 ¹	P value ²	Overall N = 60 ¹	RI
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)	1.000	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)	0.02-0.10
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)	0.000921	0.92 (0.77, 1.36)	
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)	0.000920	8.86 (8.20, 9.70)	6.0-10.0

¹Data are presented as median (IQR) or frequency (%). ² Group comparisons were performed using two-sided Wilcoxon rank-sum tests (exact where applicable) for continuous variables and Pearson's chi-squared test for categorical variables. Neutrophil-to-lymphocyte ratio (NLR) and monocyte-to-lymphocyte ratio (MLR) were calculated from absolute cell counts obtained from the complete blood count. *Abbreviations:* WBC, white blood cells; RBC, red blood cells; HGB, haemoglobin; MCV, mean cell volume; MCH, mean cell haemoglobin; MCHC, mean cell haemoglobin concentration; RDW, red cell distribution width; MPV, mean platelet volume; CBC, complete blood count; RIs, reference intervals

Supplementary Table 4. Cross-database validation of biological processes upregulated in the high pathogen exposure group. Biological processes were consistently enriched across differential expression analyses using STRING, WebGestalt, Enrichr (GO Biological Process), Literature Lab, ORA (MSigDB C5), and GSEA (MSigDB C5). FDR corresponds to the lowest adjusted value for the lead pathway term within each theme.

Cross-Database Validation of Biological Processes identified by Differential Gene Expression analysis			
Common Biological pathway	Biological pathways	FDR	Databases
Pathways upregulated in high pathogen exposure group			
Regulation of the immune system process	(i)Positive regulation of immune response/Positive regulation of biological process/Positive regulation of cellular process	0.006	STRING
	(ii)Regulation of immune system process/regulation of immune response	0.000	WebGestalt
	(iii) Immunomodulation	0.018	Literature Lab
	(iv)Positive regulation of immune response/Regulation of cell activation	0.026	GSEA (MSigDB C5)
Regulation of B cell activation	(i) Regulation of B cell activation	0.029	STRING
	(ii) Regulation of B cell activation	0.000	WebGestalt
	(iii)Memory B Cells	0.029	Literature Lab
	(iv) Regulation of cell activation	0.041	GSEA (MSigDB C5)
B cell signalling pathways	(i) Positive regulation of signalling	0.042	STRING
	(ii) B cell receptor signalling pathway	ns	WebGestalt
	(iii) B Cell Receptor Signalling Pathway	0.016	Enrichr (GO)
	(iv) B Cell proliferation	0.033	ORA MSigDB (C5)
	(v)B-Cell Signalling/B Lymphocyte Surface	0.012	Literature Lab
	(vi)B-Cell mediated immunity/B Cell Receptor Signalling Pathway	0.026	GSEA (MSigDB C5)
mRNA Processing	(i)Regulation of mRNA metabolic process/post-transcriptional regulation of gene expression	0.032	STRING
	(ii) mRNA metabolic process/mRNA processing	ns	WebGestalt
	(iii) mRNA Processing/ RNA splicing	ns	Enrichr (GO)
	(iv) Regulation of mRNA metabolic process/mRNA processing	0.031	ORA MSigDB (C5)
	(v) RNA Processing, Post-Transcriptional	0.021	Literature Lab
Positive regulation of protein localization	(i) Positive regulation of protein localization/Cellular localization	0.036	STRING
	(ii) positive regulation of protein localization	ns	WebGestalt
	(iii) Regulation of protein localization	0.033	ORA MSigDB (C5)
Abbreviations: GO- Gene ontology; ORA - Over-representation analysis; MSigDB - Molecular Signatures Database; GSEA -Gene-set enrichment analysis; WebGestalt- WEB-based Gene Set Analysis Toolkit; FDR- False Discovery Rate; ns- Not Significant			

Supplementary Table 5. Cross-database validation of biological processes downregulated in the high pathogen exposure group. Biological processes were consistently enriched across differential expression analyses using STRING, WebGestalt, Enrichr (GO Biological Process), Literature Lab, ORA (MSigDB C5), and GSEA (MSigDB C5). FDR corresponds to the lowest adjusted value for the lead pathway term within each theme.

Cross-Database Validation of Biological Processes identified by Differential Gene Expression analysis			
Common Biological pathway	Biological pathways	FDR	Databases
Pathways downregulated in high pathogen exposure group			
Leukocyte chemotaxis	(i) Leukocyte chemotaxis/Leukocyte migration/Chemotaxis/Neutrophil chemotaxis	0.008	STRING
	(ii) Neutrophil migration/granulocyte migration	0.000	WebGestalt
	(iii) Granulocyte Chemotaxis/Neutrophil Chemotaxis	0.026	Enrichr (GO)
	(iv) Neutrophil migration/Granulocyte migration/Myeloid leukocyte migration/Leukocyte chemotaxis/Cell chemotaxis	0.000	ORA MSigDB (C5)
	(v) Neutrophil Activation/Neutrophil Infiltration	0.004	Literature Lab
	(vi) myeloid leukocyte migration	0.036	GSEA (MSigDB C5)
Inflammatory response	(i) Inflammatory response	0.008	STRING
	(ii) Inflammatory Response/Positive Regulation of Inflammatory Response/Regulation of Inflammatory Response	0.033	Enrichr (GO)
	(iii) Positive regulation of type 1 immune response/Regulation of acute inflammatory response	0.016	ORA MSigDB (C5)
	(iv) Acute Inflammatory Response/Inflammatory Response/Cytokine Inflammatory Response/Inflammasome Signalling	0.002	Literature Lab
	(v) Inflammatory response/acute phase response	0.027	GSEA (MSigDB C5)
Blood vessel development	(i) Regulation of vasculature development/ Regulation of angiogenesis/Blood vessel development	0.008	STRING
	(ii) Blood vessel morphogenesis/blood vessel development/angiogenesis/vasculature development	0.000	WebGestalt
	(iii) Positive Regulation of Angiogenesis/Positive Regulation of Vasculature Development	0.003	Enrichr (GO)
	(iv) Regulation of vasculature development	0.009	ORA MSigDB (C5)
	(v) Vascular Remodelling	0.002	Literature Lab
	(vi) Blood vessel morphogenesis/circulatory system development	0.048	GSEA (MSigDB C5)
Regulation of cell population proliferation	(i) Regulation of cell population proliferation	0.008	STRING
	(ii) Regulation of Cell Population Proliferation	0.001	Enrichr (GO)
Abbreviations: GO- Gene ontology; ORA - Over-representation analysis; MSigDB - Molecular Signatures Database; GSEA -Gene-set enrichment analysis; WebGestalt- WEB-based Gene Set Analysis Toolkit; FDR- False Discovery Rate; ns- Not Significant			

Supplementary Table 6. WGCNA pathway enrichment analysis: Summary of biological pathways enriched in gene modules identified through WGCNA, stratified by their association with pathogen exposure level. 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 pathogen 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 high-pathogen 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
GM8	309	Regulation of cell motility/regulation of cell migration	ns	STRING
		no enrichment	na	WebGestalt
Modules negatively associated with high-pathogen 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	1.07E-05	WebGestalt
Abbreviations: GM- Gene Module; WebGestalt- WEB-based Gene Set Analysis Toolkit; FDR- False Discovery Rate; ns- Not significant; na- Not applicable				

156
157
158
159

Supplementary Table 7. Correlation analysis of soluble markers with anti-pathogen IgG antibodies. Values represent Spearman correlation coefficients (rho) between plasma analytes and pathogen-specific IgG responses. Red values indicate nominal significance ($p < 0.05$). Positive values denote direct associations, whereas negative values denote inverse associations.

Analyte	Pf	CMV	EBV	HSV1	AdV	Asc	Ent	Sch	Str
IL-15	-0.15	0.40*	0.1	0.23	0.13	0.13	0.04	-0.03	-0.15
IL-18	0.39*	0.26*	0.21	0.15	-0.01	-0.02	0.30*	0.18	0.14
BAFF	-0.12	0.37*	0.17	0.18	0.24	0.04	0.1	-0.08	-0.12
HGF	0.19	0.37*	0.13	-0.03	0	-0.14	0.14	-0.09	-0.1
CXCL13	0.12	0.35*	0.01	-0.05	0.01	-0.03	-0.01	-0.05	-0.09
IL-12p70	0.11	0.35*	0.04	-0.24	0.05	0.04	-0.06	-0.14	-0.07
CXCL11	0.32*	0.09	0.04	0.14	-0.06	0.13	0.18	0.07	-0.01
MCP-2	0.28*	0.13	0.1	-0.05	0.05	0.16	0.17	0.22	0.04
IL-2	0.28*	0.19	0.06	-0.12	0	-0.07	0.2	0.08	-0.01
G-CSF	0.05	0.27*	-0.02	0.15	0.08	0.02	-0.03	0.03	-0.17
CXCL5	0.04	-0.05	0.04	0.07	0.08	-0.09	0.27*	-0.07	0.1
MDC	0.18	0.23	0.2	0.21	0.22	0.05	0.27*	0.11	0.01
IL-7	0.19	0.05	0.14	0.06	-0.04	0.07	0.27*	0	0.19
IP-10	0.27*	0.1	-0.02	-0.1	-0.06	0.1	0.07	-0.08	0.05
MIF	0.26*	-0.14	0.19	0.19	-0.05	-0.06	0.19	0	0.01
EOTAXIN-2	-0.19	0.08	-0.09	-0.14	-0.07	-0.1	-0.26*	-0.11	0
VEGF-A	0.25	0.07	0.11	0.05	0.01	-0.04	0.04	-0.08	0.1
CD30	0.09	0.24	-0.13	0.13	-0.04	0.09	0.06	0.09	-0.18
TWEAK	0.24	0.19	0.02	-0.05	0.09	-0.13	0.09	0.01	0.01
SCF	0.09	0.23	-0.08	-0.24	0.04	-0.11	-0.03	-0.09	-0.1
MIP-1 beta	0.21	0.24	0.11	0.01	0.03	0.03	0.01	-0.02	-0.1
GRO-alpha	0.14	0.13	0.11	0.23	0	-0.01	0.08	-0.15	0.05
MMP-1	0.13	-0.08	0.06	0.09	0	0.23	0	-0.05	0.18
LIF	0.1	0.23	0.01	0.17	0.05	0.08	0	-0.02	0.02
TSLP	0.07	0.19	-0.01	-0.13	-0.08	0	-0.11	-0.03	-0.21
APRIL	0.08	0.19	0.01	-0.01	-0.01	-0.13	-0.03	-0.19	-0.05
EOTAXIN	0.19	0.1	0.04	0.05	0.03	-0.05	0.11	0.15	-0.06
MIP-3 alpha	0	0.18	-0.09	0.1	-0.03	0.13	0.06	-0.11	0.09
IL-2R alpha	0.07	0.08	-0.16	-0.02	0	-0.04	-0.01	-0.01	-0.17
SDF-1 alpha	0.16	0.15	0.04	-0.03	0	-0.02	0	-0.01	0.03
IL-16	0.15	0.02	0.04	-0.12	0.1	0	-0.02	0.04	-0.02
MIP-1alpha	-0.01	0.09	0.07	-0.07	-0.13	0.11	0.02	0.07	-0.1
TNF-RII	0.08	0.09	-0.07	-0.04	-0.07	-0.01	-0.03	-0.12	-0.13
MCP-1	-0.03	-0.02	0.05	0.06	-0.01	-0.06	0.05	-0.06	0.04
Abbreviations: Pf = Plasmodium falciparum; CMV = Cytomegalovirus; EBV = Epstein-Barr virus; HSV1 = Herpes simplex virus 1; AdV = Adenovirus; Asc = Ascaris lumbricoides; Ent = Entamoeba histolytica; Sch = Schistosoma mansoni; Str = Strongyloides stercoralis.									

160

References

- 1 Monaco, G. *et al.* RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell reports* **26**, 1627–1640 (2019).
- 2 Ellington, A. A., Kullo, I. J., Bailey, K. R. & Klee, G. G. Antibody-based protein multiplex platforms: technical and operational challenges. *Clinical chemistry* **56**, 186–193 (2010).
- 3 Verbovšek, T. A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. *RMZ - Materials and Geoenvironment* **58**, 393–404 (2011).