

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

Corresponding Author: Mr John Kimotho

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have satisfactorily addressed my previous comments, and the revisions have further improved the manuscript. The new focus on a broader framework of exposure to multiple pathogens, rather than malaria alone, provides a more transparent and balanced interpretation of the findings, although malaria remains the predominant pathogen exposure in the study population.

I have no further comments and recommend the manuscript for publication.

Reviewer #2

(Remarks to the Author)

The authors are proposing an interesting study by comparing individuals living in a high versus low endemic environment. Here the main differentiator, as presented by the authors is the malaria prevalence in these regions.

Comments:

The cellular deconvolution could be tried with other tools such as DOI: 10.1016/j.celrep.2019.01.041 which proved to provide nice correlation with immunophenotyping by flow cytometry (with up to 29 immune cell subsets). This may enable to highlight some differences not observed using the LM22 method (<https://doi.org/10.1038/s41467-023-43181-7>). If the regions are really different from an endemic standpoint, differences in the frequency of T cell subsets could be observed. It was shown that such differences could be observed in infants with seropositivity/negativity to infections such as cytomegalovirus (DOI: 10.1128/JVI.00052-07). Studies have shown that gamma/delta 2+ are enriched following malaria infection, this subset does not appear in the analysis.

Authors should try to correlate the genes, soluble molecules and cellular phenotype in order to reveal potential differences in the groups. As the immune response and memory is a complex network of interactions, these should be analyzed together.

The data generated shows a strong malaria bias, which should be highlighted in the title.

The data presented is a comparative analysis and does not clearly support the theories put forward. The recalibration should be tested in a dedicated longitudinal study. This can be performed using individuals that lived for increasing number of years in endemic environments.

It will be important to discuss further the recalibration concept and include some references such as doi: 10.1111/all.12364 which shows the adaptation to the environment in the context of allergy/asthma.

Is there data showing malaria reactivation in low versus high endemic areas? Can this impact and explain some of the differences observed.

Does the data differ if participants are analyzed based on the top and low 25% IgG for anti-P. falciparum? If not, then we could assume the data is mostly due to malaria infection. Data may also be compared with individuals with very high IgG (above 10x5) and high IgG (10x4 to 10x5) and identify if the same patterns are observed.

Reviewer #3

(Remarks to the Author)

This is a well-designed study and potentially important in understanding how long-term residence in pathogen-endemic environments shapes baseline immune states in healthy adults. The authors combine transcriptomic and cytokine/chemokine profiling to provide evidence that individuals from higher-exposure settings exhibit immune signatures distinct from those of individuals living in lower-exposure settings. Specifically, they show that individuals from higher-exposure settings display enhanced adaptive (B cell compartment) immune signatures coupled with restrained inflammatory transcriptional programs. Overall, the analyses performed to support these conclusions are statistically sound but with a few exceptions. In my opinion, the manuscript is strong and contains notable conceptual and methodological strengths. However, some parts require further clarification.

Major concerns:

1. The authors interpret their findings (difference in gene expression) as consequences of "long-term pathogen exposure." However, some (perhaps the majority) of the observed differential gene expression may be attributable to subclinical infection rather than to long-term immune adaptation. This raises the following question: do the authors have information on Plasmodium infection status at the time of sample collection? Specifically, do they have access to frozen blood samples that could be used for PCR-based detection of subclinical Plasmodium infection, in order to rule out that the differences in gene expression they report reflect differences in infection status between the groups rather than immune adaptation?
2. The differential expression analysis identified 658 differentially expressed genes (DEGs) in the primary model, but only 333 DEGs after adjustment for age. It is not clear from either the Results or the Methods section which gene set was carried forward for the downstream functional enrichment analyses. This should be clearly stated and justified.
3. If age is considered a confounding factor, why was the unadjusted statistical model carried forward as the primary analysis rather than the age-adjusted model? Given that age substantially reduced the number of DEGs, the rationale for prioritizing the unadjusted analysis should be explained more clearly.
4. I previously highlighted the strength of the within-country design in reducing confounding due to ancestry. However, the data would be stronger if the authors could provide information on participant ethnicity/ancestry (self-reported), if available. If ethnicity is correlated with geographic location, population structure could explain part of the observed transcriptional differences independently of long-term residence in pathogen-endemic settings. To address this, the authors could provide: (1) self-reported ethnicity for each group, if available; and (2) if ethnicity/ancestry is confounded with group, a differential expression model corrected for ancestry (genotype principal components), which can be derived using genotypes called directly from the RNA-seq reads.
5. If the RNA-seq data are not adequate for a population-structure analysis based on genetic ancestry, please provide at least a PCA of the expression data with individuals colored by region, to help the reader assess whether location contributes to the overall transcriptional structure. The PCA shown in Figure 1C is based only on pathogen-specific IgG responses and is therefore not sufficient for this purpose, as it is derived from a very limited number of variables.

Minor concerns:

6. The color schemes used for the study sites are inconsistent across Figures 1B, 1C, and 1D. Consistent color coding would improve readability and interpretation.
7. The authors should ensure consistent terminology throughout the manuscript when referring to "high exposure," "pathogen exposure," "endemic exposure, as these terms appear to be used interchangeably in some sections.
8. How much RNA was used as input for library preparation? This information is not provided in the Methods section. Were equivalent amounts of RNA used for all samples?
9. The legend of figure 5 does not have panel C.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Dear Editor,

Manuscript Title: Long-term malaria exposure in pathogen-endemic environments is associated with distinct immune states

Reviewer's #1 full report

"The authors have satisfactorily addressed my previous comments, and the revisions have further improved the manuscript. The new focus on a broader framework of exposure to multiple pathogens, rather than malaria alone, provides a more transparent and balanced interpretation of the findings, although malaria remains the predominant pathogen exposure in the study population."

Report acknowledgement: We thank Reviewer #1 for re-evaluating the revised manuscript and for their positive feedback. We are pleased that the revisions satisfactorily addressed all the concerns raised.

Reviewer's #2 full report

"I appreciate The authors are proposing an interesting study by comparing individuals living in a high versus low endemic environment. Here the main differentiator, as presented by the authors is the malaria prevalence in these regions."

Comments:

The cellular deconvolution could be tried with other tools such as DOI: 10.1016/j.celrep.2019.01.041 which proved to provide nice correlation with immunophenotyping by flow cytometry (with up to 29 immune cell subsets). This may enable to highlight some differences not observed using the LM22 method (<https://doi.org/10.1038/s41467-023-43181-7>). If the regions are really different from an endemic standpoint, differences in the frequency of T cell subsets could be observed. It was shown that such differences could be observed in infants with seropositivity/negativity to infections such as cytomegalovirus (DOI: 10.1128/JVI.00052-07). Studies have shown that gamma/delta 2+ are enriched following malaria infection, this subset does not appear in the analysis.

Authors should try to correlate the genes, soluble molecules and cellular phenotype in order to reveal potential differences in the groups. As the immune response and memory is a complex network of interactions, these should be analyzed together.

The data generated shows a strong malaria bias, which should be highlighted in the title.

The data presented is a comparative analysis and does not clearly support the theories put forward. The recalibration should be tested in a dedicated longitudinal study. This can be performed using individuals that lived for increasing number of years in endemic environments.

It will be important to discuss further the recalibration concept and include some references such as doi: 10.1111/all.12364 which shows the adaptation to the environment in the context of allergy/asthma.

Is there data showing malaria reactivation in low versus high endemic areas? Can this impact and explain some of the differences observed.

Does the data differ if participants are analyzed based on the top and low 25% IgG for anti-P. falciparum? If not, then we could assume the data is mostly due to malaria infection. Data may also be compared with individuals with very high IgG (above 10x5) and high IgG (10x4 to 10x5) and identify if the same patterns are observed.”

Response to reviewer #2

Report acknowledgement: We thank the reviewer for the positive assessment of our study and for recognizing the value of comparing individuals from high- and low-endemic settings. We appreciate the constructive comments and suggestions, which have helped strengthen the manuscript. Our point-by-point responses are provided below.

- 1. Reviewer comment:** *“The cellular deconvolution could be tried with other tools such as DOI: 10.1016/j.celrep.2019.01.041 which proved to provide nice correlation with immunophenotyping by flow cytometry (with up to 29 immune cell subsets). This may enable to highlight some differences not observed using the LM22 method (<https://doi.org/10.1038/s41467-023-43181-7>). If the regions are really different from an endemic standpoint, differences in the frequency of T cell subsets could be observed. It was shown that such differences could be observed in infants with seropositivity/negativity to infections such as cytomegalovirus (DOI: 10.1128/JVI.00052-07). Studies have shown that gamma/delta 2+ are enriched following malaria infection, this subset does not appear in the analysis ”.*

Response: We appreciate this valuable suggestion. As recommended, we repeated the cellular deconvolution using the ABsolute Immune Signal (ABIS) framework (Monaco et al., 2019), which estimates a broader range of PBMC immune populations (29 cell types) than the LM22 reference matrix (22 cell types). In our dataset, ABIS resolved 17 immune cell populations, including subsets not represented in LM22 such as $\gamma\delta$ V δ 2 and $\gamma\delta$ non-V δ 2 T cells. Overall, the ABIS analysis closely recapitulated the findings obtained using LM22, demonstrating enrichment of adaptive immune populations in individuals with high pathogen exposure, including significantly increased memory B-cell abundance together with reduced classical monocyte signatures. Although ABIS enabled the assessment of $\gamma\delta$ T-cell subsets, neither $\gamma\delta$ V δ 2 nor $\gamma\delta$ non-V δ 2 T cells differed significantly between exposure groups after multiple-testing correction. This may reflect the fact that our study compared healthy adults with differing histories of long-term pathogen exposure rather than individuals with acute or recent malaria infection. In addition, deconvolution methods infer relative cell proportions from bulk transcriptomic profiles and may have limited sensitivity for detecting subtle differences in low-abundance immune subsets compared with direct immunophenotyping. Thus, the additional immune subsets resolved by ABIS did not alter the principal biological conclusions of the study but provided independent confirmation that our observations were robust to the choice of deconvolution algorithm. We have incorporated these new sensitivity analyses into the revised manuscript in Results section (lines 272–278) and **Supplementary Figure 3A-B**.

- 2. Reviewer comment:** *“ Authors should try to correlate the genes, soluble molecules and cellular phenotype in order to reveal potential differences in the groups. As the immune response and memory is a complex network of interactions, these should be analyzed together.”*

Response: We appreciate this valuable suggestion. We have performed an integrated analysis combining representative features from the major immune layers investigated in this study, including pathogen exposure serology, plasma cytokines, exposure-associated transcriptomic modules, and inferred immune

cell populations. Specifically, we integrated these features using a LASSO logistic regression model to derive an integrated immune state score, allowing us to identify the immune components that most strongly distinguished individuals from high- and low-pathogen-exposure settings. The resulting integrated immune state score robustly separated the two exposure groups and identified RNA processing (GM20 and GM21), B-cell activation (GM17), memory B-cell abundance, pathogen exposure, cytokine responses, and inflammatory signalling (GM30) as the principal contributors to this coordinated immune phenotype. To further address the reviewer's suggestion that these immune components should be analysed together, we have performed Spearman correlation analysis of the features included in the integrated immune state model. This analysis identified coordinated associations among the integrated immune features, with the strongest relationships observed between transcriptional modules and inferred immune cell populations. In particular, the RNA processing and metabolism modules (GM20 and GM21), together with the B-cell activation module (GM17), formed a positively correlated cluster with memory B-cell abundance, whereas the inflammatory signalling module (GM30) clustered with classical monocytes, defining an opposing innate inflammatory programme. Importantly, this integrated analysis provides a systems-level view of the immune landscape, strengthening the biological interpretation of the distinct immune states identified in the study. We have incorporated these analyses into the revised manuscript as new **Figure 6**, together with a corresponding Results section (Lines 401-429), Methods subsection (Lines 879-894), and Discussion describing this integrated immune state.

- 3. Reviewer comment:** *"The data generated shows a strong malaria bias, which should be highlighted in the title."*

Response: We thank the reviewer for this helpful suggestion. We agree that malaria represents the dominant exposure gradient across our study sites and is likely the principal contributor to the immune differences observed, although these occur within broader pathogen-endemic environments. To better reflect the study design and primary findings, we have revised the title to: **"Long-term malaria exposure in pathogen-endemic environments is associated with distinct immune states."**

- 4. Reviewer comment:** *"The data presented is a comparative analysis and does not clearly support the theories put forward. The recalibration should be tested in a dedicated longitudinal study. This can be performed using individuals that lived for increasing number of years in endemic environments."*

Response: We thank the reviewer for this thoughtful comment. A prospective longitudinal study would provide the strongest evidence for the temporal development of the immune states described here and would be valuable for directly examining immune adaptation over time. However, such studies are inherently challenging in human populations because they require prolonged follow-up while accounting for repeated pathogen exposures, maintaining participant retention over many years, and ensuring that samples are collected during periods free of acute illness. Although our study is cross-sectional, it was specifically designed to minimise the influence of recent infection and capture the effects of long-term cumulative exposure. Participants were required to have resided in their respective study areas for at least 10 years, be clinically healthy at enrolment, have no history of acute or severe infection during the preceding six months, and be malaria-negative at sampling. These stringent inclusion criteria increase confidence that the observed immune differences reflect long-term residence in distinct transmission settings rather than transient responses to recent infection. Nevertheless, we agree that our study cannot directly demonstrate the temporal process by which these immune states develop. Rather, our findings identify immune states associated with long-term malaria exposure within pathogen-endemic environments in clinically healthy individuals with well-defined, prolonged residence in distinct

transmission settings. We also highlight the need for carefully designed longitudinal studies as an important future direction to define the kinetics and mechanisms through which these immune states emerge and evolve over time. We have also carefully revised the Discussion to ensure that our interpretation remains fully consistent with the results presented.

5. **Reviewer comment:** “ *It will be important to discuss further the recalibration concept and include some references such as doi: 10.1111/all.12364 which shows the adaptation to the environment in the context of allergy/asthma.*”

Response: We thank the reviewer for this suggestion. The Discussion has been expanded to place our findings within the broader context of environmental immune adaptation. In addition to the suggested reference, we have included complementary literature to provide a broader conceptual framework for understanding how prolonged environmental exposures can remodel human immune phenotypes. These additions have been incorporated into the Discussion (Lines 475-479).

6. **Reviewer comment:** “ *Is there data showing malaria reactivation in low versus high endemic areas? Can this impact and explain some of the differences observed.*”

Response: We thank the reviewer for raising this point. We interpret “malaria reactivation” as referring either to recurrent malaria exposure or possible relapse from non-*falciparum* species. Relapse from non-*falciparum* species such as *P. ovale* or *P. vivax* is biologically possible, however, *P. falciparum* is the predominant malaria species in the studied regions, and relapsing species account for only a small proportion of infections. Importantly, our study was designed to compare clinically healthy adults with prolonged residence in contrasting transmission settings rather than individuals with active malaria. Participants had no symptoms or signs of acute malaria, no history of acute or severe infection during the preceding six months, and a complete blood count was performed to exclude haematological abnormalities. These criteria make ongoing clinical malaria or recent severe infection unlikely explanations for the observed immune differences. Nevertheless, we acknowledge that low-density asymptomatic parasitaemia cannot be completely excluded in high-transmission settings and may contribute to baseline immune variation. Such asymptomatic infections are themselves characteristic of long-term malaria exposure and acquired clinical immunity in endemic populations, making it difficult to disentangle their contribution from the cumulative effects of chronic exposure. Overall, we interpret the observed immune states as being associated with cumulative long-term pathogen exposure rather than acute clinical malaria or recent symptomatic infection.

7. **Reviewer comment:** “ *Does the data differ if participants are analyzed based on the top and low 25% IgG for anti-P. falciparum? If not, then we could assume the data is mostly due to malaria infection. Data may also be compared with individuals with very high IgG (above 10x5) and high IgG (10x4 to 10x5) and identify if the same patterns are observed.*”

Response: We thank the reviewer for this suggestion. We had considered using anti-*Plasmodium falciparum* schizont IgG titres as an alternative classifier for exposure. However, in our study, anti-*Plasmodium falciparum* schizont IgG titres closely mirrored the predefined exposure groups, with minimal overlap between participants from the high- and low-transmission settings; only four individuals in the high-transmission group had titres that overlapped with those observed in the low-transmission group (Fig. 1E). Consequently, stratifying participants according to the highest and lowest 25% of IgG titres, or according to the proposed IgG thresholds, would select the same individuals (differing primarily in sample size rather than participant composition) represented by the current exposure groups and would

therefore not constitute an independent analysis. This is further supported by our site-level differential expression analyses, which showed that comparisons between sites within the same exposure tier identified very few differentially expressed genes (Junju vs Ahero, 2 genes; Ngerenya vs Nairobi, 18 genes), whereas comparisons spanning the greatest exposure differences showed the largest transcriptional divergence (up to 401 genes). Together, these findings indicate that the observed immune signatures track the underlying exposure gradient rather than geographical boundaries. Given the close concordance between schizont IgG titres and the predefined exposure groups, we would expect IgG-based stratification to yield the same biological conclusions. We have clarified this rationale in the revised manuscript.

Reviewer's #3 full report

"This is a well-designed study and potentially important in understanding how long-term residence in pathogen-endemic environments shapes baseline immune states in healthy adults. The authors combine transcriptomic and cytokine/chemokine profiling to provide evidence that individuals from higher-exposure settings exhibit immune signatures distinct from those of individuals living in lower-exposure settings. Specifically, they show that individuals from higher-exposure settings display enhanced adaptive (B cell compartment) immune signatures coupled with restrained inflammatory transcriptional programs. Overall, the analyses performed to support these conclusions are statistically sound but with a few exceptions. In my opinion, the manuscript is strong and contains notable conceptual and methodological strengths. However, some parts require further clarification.

Major concerns:

- 1. The authors interpret their findings (difference in gene expression) as consequences of "long-term pathogen exposure." However, some (perhaps the majority) of the observed differential gene expression may be attributable to subclinical infection rather than to long-term immune adaptation. This raises the following question: do the authors have information on Plasmodium infection status at the time of sample collection? Specifically, do they have access to frozen blood samples that could be used for PCR-based detection of subclinical Plasmodium infection, in order to rule out that the differences in gene expression they report reflect differences in infection status between the groups rather than immune adaptation?*
- 2. The differential expression analysis identified 658 differentially expressed genes (DEGs) in the primary model, but only 333 DEGs after adjustment for age. It is not clear from either the Results or the Methods section which gene set was carried forward for the downstream functional enrichment analyses. This should be clearly stated and justified.*
- 3. If age is considered a confounding factor, why was the unadjusted statistical model carried forward as the primary analysis rather than the age-adjusted model? Given that age substantially reduced the number of DEGs, the rationale for prioritizing the unadjusted analysis should be explained more clearly.*
- 4. I previously highlighted the strength of the within-country design in reducing confounding due to ancestry. However, the data would be stronger if the authors could provide information on participant ethnicity/ancestry (self-reported), if available. If ethnicity is correlated with geographic location, population structure could explain part of the observed transcriptional differences independently of long-term residence in pathogen-endemic settings. To address this, the authors could provide: (1) self-reported ethnicity for each group, if available; and (2) if ethnicity/ancestry is confounded with group, a differential expression model corrected for ancestry (genotype principal components), which can be derived using genotypes called directly from the RNA-seq reads.*

5. If the RNA-seq data are not adequate for a population-structure analysis based on genetic ancestry, please provide at least a PCA of the expression data with individuals colored by region, to help the reader assess whether location contributes to the overall transcriptional structure. The PCA shown in Figure 1C is based only on pathogen-specific IgG responses and is therefore not sufficient for this purpose, as it is derived from a very limited number of variables.

Minor concerns:

6. The color schemes used for the study sites are inconsistent across Figures 1B, 1C, and 1D. Consistent color coding would improve readability and interpretation.

7. The authors should ensure consistent terminology throughout the manuscript when referring to “high exposure,” “pathogen exposure,” “endemic exposure, as these terms appear to be used interchangeably in some sections.

8. How much RNA was used as input for library preparation? This information is not provided in the Methods section. Were equivalent amounts of RNA used for all samples?

9. The legend of figure 5 does not have panel C.”

Response to reviewer #3

Report acknowledgement: We thank the reviewer for the positive assessment of our study and for recognizing its potential contribution to understanding how long-term residence in pathogen-endemic environments shapes baseline immune states. We appreciate the constructive comments and suggestions, which have helped strengthen the manuscript. Our point-by-point responses are provided below.

1. **Reviewer comment:** “The authors interpret their findings (difference in gene expression) as consequences of “long-term pathogen exposure.” However, some (perhaps the majority) of the observed differential gene expression may be attributable to subclinical infection rather than to long-term immune adaptation. This raises the following question: do the authors have information on *Plasmodium* infection status at the time of sample collection? Specifically, do they have access to frozen blood samples that could be used for PCR-based detection of subclinical *Plasmodium* infection, in order to rule out that the differences in gene expression they report reflect differences in infection status between the groups rather than immune adaptation?”

Response: We thank the reviewer for this important comment. The possibility that subclinical *Plasmodium* infection, or asymptomatic parasitaemia, contributed to differences in baseline immune profiles cannot be completely excluded. At enrolment, all participants were clinically healthy, had no signs or symptoms of malaria, no history of acute or severe infection during the preceding six months, and a complete blood count was performed to exclude haematological abnormalities. Thus, the observed immune differences are unlikely to reflect active clinical malaria. We did not perform PCR-based detection of subclinical *Plasmodium* infection, and suitable whole-blood samples were not available for retrospective molecular testing. We therefore cannot completely exclude the possibility that asymptomatic parasitaemia contributed to some of the observed transcriptional differences, particularly among participants from higher-transmission settings. We acknowledge this as a limitation of the study. However, asymptomatic parasitaemia is itself a recognized feature of long-term malaria exposure in endemic populations, reflecting acquired clinical immunity following repeated infections. Consequently, while its contribution cannot be distinguished in the present study, it represents part of the biological continuum of chronic malaria exposure rather than an entirely independent confounding factor. Taken together, because

participants were recruited based on prolonged residence (≥ 10 years) in contrasting transmission settings and were clinically healthy at sampling, we interpret the observed immune signatures as being primarily associated with cumulative long-term malaria exposure alongside other pathogens, rather than acute/recent symptomatic infection.

2. **Reviewer comment:** *“The differential expression analysis identified 658 differentially expressed genes (DEGs) in the primary model, but only 333 DEGs after adjustment for age. It is not clear from either the Results or the Methods section which gene set was carried forward for the downstream functional enrichment analyses. This should be clearly stated and justified?”*

Response: We thank the reviewer for this important point. The downstream functional enrichment analyses were performed using the differentially expressed genes identified in the primary (unadjusted) differential expression analysis (658 differentially expressed genes). Age- and sex-adjusted models were conducted solely as sensitivity analyses to assess the robustness of the observed transcriptional signatures. As demonstrated in the manuscript, adjustment for age primarily reduced the number of detected differentially expressed genes rather than altering the underlying biological signal, with 82% (274/333) of the age-adjusted DEGs overlapping those identified in the primary analysis and only 10 genes uniquely detected following age adjustment. These findings indicate that age adjustment primarily attenuated the exposure-associated signal rather than identifying a distinct transcriptional programme, consistent with age acting as a proxy for cumulative pathogen exposure in this cohort due to prolonged residence in the respective transmission settings. Consequently, adjusting for age is likely to constitute over-adjustment by accounting for variance that is intrinsically related to the exposure of interest. We have now explicitly clarified in both the Results (Lines 188–191) and Methods (Lines 755–757) that all downstream analyses, including functional enrichment analyses, were performed using the DEGs identified from the primary model (658 DEGs), while the covariate-adjusted models were used solely as sensitivity analyses.

3. **Reviewer comment:** *“If age is considered a confounding factor, why was the unadjusted statistical model carried forward as the primary analysis rather than the age-adjusted model? Given that age substantially reduced the number of DEGs, the rationale for prioritizing the unadjusted analysis should be explained more clearly.”*

Response: We thank the reviewer for this important question. As explained above in our response to Comment 2, although age differed between the exposure groups, we did not consider it to be a classical confounder because it is intrinsically linked to cumulative pathogen exposure through prolonged residence in the respective transmission settings. Consequently, adjusting for age is likely to constitute over-adjustment by accounting for variance that is directly related to the exposure of interest. For this reason, the unadjusted model was prespecified as the primary differential expression analysis, whereas the age-adjusted model was used solely as a sensitivity analysis to assess the robustness of the observed transcriptional signatures. As demonstrated in the manuscript, age adjustment attenuated the exposure-associated signal but did not identify a distinct transcriptional programme. We have now clarified this rationale more explicitly in both the Results (Lines 188–191) and Methods (Lines 755–757).

4. **Reviewer comment:** *“I previously highlighted the strength of the within-country design in reducing confounding due to ancestry. However, the data would be stronger if the authors could provide information on participant ethnicity/ancestry (self-reported), if available. If ethnicity is correlated with geographic location, population structure could explain part of the observed transcriptional differences independently of long-term residence in pathogen-endemic settings. To address this,*

the authors could provide: (1) self-reported ethnicity for each group, if available; and (2) if ethnicity/ancestry is confounded with group, a differential expression model corrected for ancestry (genotype principal components), which can be derived using genotypes called directly from the RNA-seq reads.”

Response: We appreciate this important comment. We acknowledge that genetic background can contribute to variation in immune phenotypes and therefore warrants consideration. Participants were recruited from regions representing distinct malaria transmission intensities but within Kenyan populations predominantly comprising Mijikenda communities along the Coast (Junju and Ngerenya), the Luo community in Ahero, and a cosmopolitan population in Nairobi. To assess whether the observed transcriptional differences were more consistent with ethnicity or pathogen exposure more directly, we compared differential gene expression across all study sites (**Supplementary Fig. 2C**). If ethnicity were the principal driver of the observed transcriptional differences, we would expect the greatest divergence between ethnically distinct populations irrespective of exposure history. However, this was not observed. Junju and Ahero, which are predominantly Mijikenda and Luo, respectively, but experience comparable malaria transmission intensities, differed by only two differentially expressed genes. In contrast, Junju and Ngerenya, which are predominantly Mijikenda but differ in pathogen exposure, exhibited greater transcriptional divergence (25 differentially expressed genes). Furthermore, the largest transcriptional differences were consistently observed between sites spanning the greatest differences in pathogen exposure (Ahero vs Nairobi, 401 genes; Ahero vs Ngerenya, 184 genes; Junju vs Nairobi, 149 genes), whereas comparisons between sites within the same exposure category identified very few differentially expressed genes (Junju vs Ahero, 2 genes; Ngerenya vs Nairobi, 18 genes). Together, these findings indicate that the observed transcriptional signatures are more closely associated with the pathogen exposure gradient than with ethnic background, making population structure an unlikely explanation for the principal findings of this study.

- 5. Reviewer comment:** *“If the RNA-seq data are not adequate for a population-structure analysis based on genetic ancestry, please provide at least a PCA of the expression data with individuals colored by region, to help the reader assess whether location contributes to the overall transcriptional structure. The PCA shown in Figure 1C is based only on pathogen-specific IgG responses and is therefore not sufficient for this purpose, as it is derived from a very limited number of variables.”*

Response: We thank the reviewer for this helpful suggestion. Our site-level differential expression analyses indicated that the observed transcriptional differences were more closely associated with the pathogen exposure gradient than with ethnicity or geographic location. To further support this conclusion, as recommended, we have now performed principal component analysis (PCA) of the global RNA-seq transcriptome and included the results as **Supplementary Fig. 2B**. Consistent with the site-level differential expression analyses, participants showed substantial overlap across the four recruitment sites, with no distinct clustering according to geographic location. This was not unexpected, as the transcriptional differences associated with long-term malaria exposure were relatively subtle at the global transcriptome level and therefore were not expected to drive clear separation in the principal components. Taken together, these complementary analyses support the conclusion that the observed transcriptional differences are more closely associated with the pathogen exposure gradient than with recruitment site.

- 6. Reviewer comment:** *“The color schemes used for the study sites are inconsistent across Figures 1B, 1C, and 1D. Consistent color coding would improve readability and interpretation.”*

Response: We thank the reviewer for this helpful observation. We agree that consistent colour coding improves figure readability and interpretation. We have now harmonized the colour scheme used for the four study sites across Figures 1B, 1C, and 1D so that each recruitment site is represented by the same colour.

7. **Reviewer comment:** *“The authors should ensure consistent terminology throughout the manuscript when referring to “high exposure,” “pathogen exposure,” “endemic exposure, as these terms appear to be used interchangeably in some sections.”*

Response: We thank the reviewer for this helpful suggestion. We have carefully reviewed the manuscript and standardized the terminology throughout. We now use "pathogen exposure" as the primary term when referring to the biological exposure of interest, "high-" and "low-pathogen-exposure groups" when describing the analytical groups, and "pathogen-endemic environments/settings" when referring to the geographical context.

8. **Reviewer comment:** *“How much RNA was used as input for library preparation? This information is not provided in the Methods section. Were equivalent amounts of RNA used for all samples?.”*

Response: We thank the reviewer for highlighting that. We have now clarified the RNA input used for library preparation in the Methods section. For library preparation, all samples were normalized to an input of 500 ng total RNA in a final volume of 11 μ L, to ensure equivalent amounts of RNA were used for all samples. This information has now been added to the Methods section (Lines 663-665).

9. **Reviewer comment:** *“The legend of figure 5 does not have panel C?.”*

Response: We thank the reviewer for identifying this omission. The legend for Figure 5 has been corrected to include panel C.