nature portfolio

Peer Review File

DNA methylation signatures underpinning blood neutrophil to lymphocyte ratio during first week of human life

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

In this manuscript, Martino et al. investigated the role of epigenetic regulation during the first week of postnatal life in 720 West African neonates. Whilst the study is of interest and could provide valuable insights into early-life determinants of immune system development and neonatal sepsis risk, there remains several questions/concerns as listed below.

Specific Comments:
Page 4, Line 105: "These associations were also observed in the validation cohort, suggesting that methylation remodeling is a robust process" – How many achieved statistical significance, and at what p-value threshold in validation cohort?

Methods – Sample Collection (pg 16):
- Participation criteria stated in manuscript differs from that at clinicaltrial.gov – Clarify which is the correct one please
- What's the rationale of vaginal birth only since we are not investigating microbiome here, which is well established to be affected by method of birth

Methods – Epigenetics (pg 16): "After sample and probe filtering, the final dataset consisted of 1,267 samples and 747,905 CpG probes." - Can we have a breakdown of how many samples/probes were lost due to the different QC criteria respectively?

Methods – Statistical analysis (pg 19-20):
- Is adjusting for chip (Sentrix ID) alone sufficient, which does not take into account for example, chip position? Did the authors consider doing a more comprehensive adjustment, e.g. the use of control probes, or control probes PC?
- There is no adjustment for cell counts/composition in main model – It is well established that cell composition impacts upon the association analyses
- "SNPs were genome-wide significant at a threshold of P <= 1x10-8" – This does not seem to be agreement with the 'standard' GWAS significance threshold of 5E-8, which is based upon a Bonferroni correction for all the independent common SNPs across the human genome.

Results:
- The analyses in the manuscript centred around genomic regions (e.g. Results – page 4). Why are there no results presented at the individual CpG level?
- GWAS: 100+ genetic associations have been previously reported in GWASes of adult NLRs in European populations – did the authors performed a lookup of these variants in their GWAS, and how do they fare? Conversely, did the authors look up their 13 genomic risk loci claimed here, and how do they perform in the adult NLR in European population?

General comments:
Since the study aims to investigate the role of epigenetic regulation during the first week of postnatal life, why now directly perform an epigenome-wide association study (EWAS) for NLR?

It is understandable that parents of newborns are cautious/careful about excessive blood-taking, but is there any (even small/pilot) subset of individuals with multiple timepoints taken? Right now, it comes across somewhat as a 'patchwork' kind of effort to piece together comparisons across different days from different individuals.

There should be a statement on sample size/power considerations, in particular taking into consideration the somewhat modest sample size of the study for a GWAS.

The authors did not perform any meQTL (SNP-methylation) or eQTL (SNP-expression) analysis, which is a little wasted since the data was already available. This could be used for downstream Mendelian Randomization analyses to better establish causality.

Table 1: What does the ‘Correlation’ column refers to - Strength of eQTL association? How are the ‘target genes’ defined for the analysis – nearest genes, predicted by (?), or?

Figure S4: Pearsons’s correlation between Flow cytometry-derived and epigenetically-inferred NLR appears to be strong. What about correlation between individual cell counts by cytometry and estimated cell counts?

Minor:
Page 4: Undefined first use of abbreviation: “DOL”

Reviewer #2 (Remarks to the Author):

Martino et al. describe a comprehensive study of the DNA methylation profile in the first week of life in a cohort of West African neonates. The works provides a good correlation between the Blood Neutrophil to lymphocyte ratio (NRL) determined by flow cytometry and by DNA methylation profiling. The present study is part of the Expanded Program on Immunization - Human Immunology Project Consortium (EPIC-HIPC). They nicely describe a number of transcription factores related to the NRL. Unfortunately the study of those newborns that develop sepsis or other infection is rather limited.
Some points need to be clarify:
Do the samples from the septic group or the focal infection group belong to the main cohort or the confirmation cohort? . They wrote: "Twelve newborns developed acute localized infections and 21 developed sepsis within the first week of life and were matched to 33 healthy controls based on sex, vaccination status, DOL, and time of blood collection after birth. " Then, the members of the two cohorts were not followed to see infections/sepsis?
The use of a public available umbilical cord blood to study the methylation profile of premature birth newborns is a serious limitation of the work. Authors should take into account that most septic process take place in premature neonates thus, it is more important to know the DNA methylation profiles in these patients.
Line 303: Clarify: "This molecular NLR may be a promising marker for severe disease risk stratification in newborns, as we found strong correlations with flow cytometry measures, but stronger associations with early onset sepsis compared to flow. "
Line 337: ". A second validation cohort comprising 45 newborns was recruited from the main cohort. " Please clarify, if there are 2 validation cohorts? Is this "second" validation cohort part of the main cohort? This must be clear since a validation cohort cannot be included in the main cohort.
More information about sepsis patients must be included .
In general terms clinical information about the patient enrolled is rather limited.

Reviewer #3 (Remarks to the Author):

This study investigated the relationship between blood neutrophil to lymphocyte ratio (NLR) and DNA methylation on days of life 0, 1, 3, and 7. This investigation included samples from 673/720 and 45 newborns in the main and validation cohorts, and analyzed the DNA methylation levels of different immune cells and their transcription factor. Furthermore, this study found that the NLR baseline can predict sepsis/infection outcomes, and the genome-wide association study indicated the SNP variant was related to the NLR level. Cumulatively, these findings elucidate a possible predictor of newborn immune status. However, the analysis of data is not understood enough, especially the clinical translational significance. I have the following issues that need to be addressed.
1. More clarification in the methods and a representative gating strategy in the supplementary

would be helpful for the proper interpretation of the flow cytometry data.
2. The data presented in Figure1 included the samples on DOL 0, 1, 3, and 7, whereas the time points in Figure 3 are DOL 1, 2, 3, 4. It would be more convincing to standardize the time points for sample acquisition.
3. Statistical analysis should be performed on Figure 2C and p-values should be displayed to support the conclusion that "A reduction in neutrophil counts and increase in T-lymphocytes over the first week of life in both cohorts".
4. In Figure4, the sepsis were divided into two types include early and late sepsis on D0, whereas there is no classification of sepsis in D1, D3 and D7. And the definition of sepsis classification should be clarified in method section. Besides, the sample size of validation cohorts is too small to support the result that "epigenetic NLRs were increased in neonates with sepsis in samples collected during the infection period (Figure4B)".
5. The description of "Reduced NLRs at birth predicted sepsis outcomes" is inappropriate and the evidence is insufficient. Investigators should include correlation analysis of epigenetic NLR and sepsis outcomes, such as SOFA score.
6. The figure legend of Figure5C is absence. What does genotypes 1, 2, and 3 refer to ?

Response to Reviewer Comments

We appreciate the thoughtful comments from all reviewers. Our response to these comments will strengthen the manuscript and improve the clarity and impact of our findings. Below we have provided a point by point response. In the revision we have also used the explicit EPIC-HIPC cohort codes to refer to the main (EPIC-002) and validation (EPIC-003) cohorts throughout to enhance clarity ofr the manuscript. Overall, these revisions have further strengthened the manuscript which we hope is now acceptable.

Reviewer #1

Specific Comments

- **Reviewer Comment:** "Page 4, Line 105: “These associations were also observed in the validation cohort, suggesting that methylation remodeling is a robust process” – How many achieved statistical significance, and at what p-value threshold in validation cohort?"
    - **Response:** Thank you for the question regarding the validation cohort analysis. In the valiation cohort 22,676/22,836 tested associations (>99%) were significant at FDR < 0.05. Our approach to validating our ontogeny model was focused on assessing correlation of t-statistics from the model fit on both the discovery and validation cohort. This approach leverages both magnitude and directionality of t-statistic to assess generalizability, which is appropriate given the difference in sample size between the main (1267 samples) and validation cohort (86 samples). The high R-squared value (0.92) indicates a strong positive correlation between the t-statistics, suggesting that the observations from the main cohort are reflected in the validation cohort. Additionally, the Bland-Altman plot confirmed minimal variation (less than 5%) between data sets for the ontogeny associated CpG
        - **Action:** In response to the feedback we have amended p4 to now read: “These associations were also observed in the validation cohort (22,676/22,836, FDR < 0.05), suggesting that methylation remodeling is a robust process (Figure 1C and S1).”

- **Reviewer Comment:** "Methods – Sample Collection (pg 16): - Participation criteria stated in manuscript differs from that at clinicaltrial.gov – Clarify which is the correct one please - What’s the rationale of vaginal birth only since we are not investigating microbiome here, which is well established to be affected by method of birth"
    - **Response:** Thank you for pointing out this difference. Both are correct and state the inclusion criteria is healthy term infants. We do acknowledge the clinical trial registration does not explicitly mention vaginal delivery as inclusion criteria, because more than 90% of births occurred this way at our site in The Gambia, so it is not so much a rationale, rather a norm.
        - **Action:** We have clarified in the text (p17):

“The inclusion criteria were healthy term newborns (gestational age > 36 weeks), born vaginally as is the vast majority (>90%) of births in our study population”

- **Reviewer Comment:** "Methods – Epigenetics (pg 16): “After sample and probe filtering, the final dataset consisted of 1,267 samples and 747,905 CpG probes.” - Can we have a breakdown of how many samples/probes were lost due to the different QC criteria respectively?"
    - **Action:** We have provided this supplementary information as a new Table S7 and referenced in the text: “After sample and probe filtering, the final dataset consisted of 1,267 samples and 747,905 CpG probes (Table S7).”

- **Reviewer Comment:** "Methods – Statistical analysis (pg 19-20): - Is adjusting for chip (Sentrix ID) alone sufficient, which does not take into account for example, chip position? Did the authors consider doing a more comprehensive adjustment, e.g. the use of control probes, or control probes PC? - There is no adjustment for cell counts/composition in main model – It is well established that cell composition impacts upon the association analyses - “SNPs were genome-wide significant at a threshold of P <= 1x10-8” – This does not seem to be agreement with the ‘standard’ GWAS significance threshold of 5E-8, which is based upon a Bonferroni correction for all the independent common SNPs across the human genome."
    - **Response:** Thank you for your valuable feedback. Regarding covariate adjustment, our main model included sample plate, sex, Sentrix ID, and Subject ID. We believe this adequately captures technical variation, as evidenced by the high concordance of effect sizes and directions in our replication cohort. While additional adjustments like control probe PCs are valid approaches, we prioritized a parsimonious model with robust replication. Given the large sample size of our main cohort, further covariate inclusion would likely have minimal impact.

      As our primary goal was to capture biologically relevant variation in cell composition, we intentionally did not adjust for cell counts in the main model. Instead, we performed two key analyses:
      Post-hoc analysis: We refit the model for ontogeny-associated CpGs, including cell counts as covariates. This revealed minimal attenuation of associations (53/22,836), supporting the robustness of our findings (reported on page 7).
      EpiDish analysis: We utilized the EpiDish approach to specifically identify differentially methylated cell types, incorporating cell counts as interaction terms. This allowed us to directly investigate cell-type-specific effects (page 7).

      Regarding the GWAS significance threshold, in the revision we now report on the 5E-8 threshold and have revised table 3 accordingly. We have amended the methods to reflect that fact that genome-wide significance is reported at Bonferroni adjusted P < 0.05.
        - **Action:** Revision of Table 3 reporting statistics at 5E-8. Amended text line 256 (p13): We detected 13 associations at genome-wide significance (Table3) and used the model statistics as input to the FUMA GWAS tool to identify genomic risk loci and perform expression quantitative trait loci (eQTL) analysis. The strongest significant genome-wide genetic signals were detected at 13 genomic risk loci (Figure 5 A, B, Table S4-5).
        - **Action:** Revision of methods p20: “P <=5x10-8”

- **Reviewer Comment:** "Results: - The analyses in the manuscript centred around genomic regions (e.g. Results – page 4). Why are there no results presented at the individual CpG level?
    - **Response:** We prioritized the analysis of differentially methylated regions (DMRs) because they represent broader, biologically relevant patterns of coordinated methylation changes, often reflecting underlying regulatory mechanisms. Additionally, DMR analyses increase statistical power compared to analysing individual CpG sites, particularly when the effect size at individual sites is small. However, we recognize the importance of individual CpG information. Therefore, we focused on reporting the most biologically relevant associations in Supplementary Table S1, which were derived by integrating DMRs with RNA-seq data. This table includes information on individual CpG associations within these prioritized DMRs. By taking this approach, we aim to strike a balance between highlighting the broader epigenetic changes of functional relevance and providing detailed information on individual CpG associations for further investigation.

- **Reviewer Comment:** - GWAS: 100+ genetic associations have been previously reported in GWASes of adult NLRs in European populations – did the authors performed a lookup of these variants in their GWAS, and how do they fare? Conversely, did the authors look up their 13 genomic risk loci claimed here, and how do they perform in the adult NLR in European population?"
    - **Response:** We thank the reviewer for highlighting the importance of cross-population comparisons. We attempted a look-up analysis of our findings against previously reported NLR associations in European populations (EFO_0008447, GWAS Catalog). Unfortunately, full summary statistics are not available, limiting our ability to perform a comprehensive comparison. While some overlap with our significant associations was observed (3/151 variants), these did not reach significance in our African newborn cohort in covariate adjusted models. This may be attributable to differences in LD structure between populations. While a more thorough cross-population comparison would be valuable, we emphasize that the primary focus of our study was the epigenetic landscape of newborn immune ontogeny. The GWAS analysis was a secondary exploration, limited in scope to exploring the relative contribution of genes and environment appropriate for the sample size. Our study was not intended to identify robust genetic signals for NLR nor reproduce previous associations, but the reported findings offer novel insights into the heritability of NLR in an understudied population, providing a foundation for future research in larger cohorts.
        - **Action:** On page 15 of the discussion we have included the following text:
            - Given population differences in LD structure, meta-analysis across European and African cohorts for NLR genetic associations will provide futher insights into the genetic acitechture of blood cell traits.

- **Reviewer Comment:** "General comments: Since the study aims to investigate the role of epigenetic regulation during the first week of postnatal life, why now directly perform an epigenome-wide association study (EWAS) for NLR?"
    - **Response:** We appreciate the reviewer's insightful question regarding the choice of analysis plan. Our study prioritizes characterizing the dynamic epigenetic landscape over the first seven days of life, with the goal of identifying novel epigenetic patterns and pathways that may be relevant to a range of health outcomes, including NLR. By first identifying DMRs associated with DOL, we prioritized specific regions for future EWAS in larger cohorts. Moreover, NLR and DOL were highly correlated so an EWAS of NLR would produce similar results.

- **Reviewer Comment:** "It is understandable that parents of newborns are cautious/careful about excessive blood-taking, but is there any (even small/pilot) subset of individuals with multiple timepoints taken? Right now, it comes across somewhat as a 'patchwork' kind of effort to piece together comparisons across different days from different individuals."
    - **Response:** We understand this perspective. Every infant in this study had 2 small blood samples drawn, which was acceptable to the parents and local IRB in the first week of life. All infants had sampling at DOL0 and again on their DOL 1,3 or 7 in a randomised fashion. We deliberately included a large number of newborns to account for the need to stagger subsequent sampling post baseline to avoid taking more than 2 samples per infant. It needs to be appreciated that these were all healthy infants, who did not have any medical need for a blood test. We emphasize that our design is considerate and unique given logistical challenges; there are very few, if any newborn cohorts with multiple blood draws in this very early window of development and usually they would have had bloods taken for a medical need rather than a research protocol.

- **Reviewer Comment:** "There should be a statement on sample size/power considerations, in particular taking into consideration the somewhat modest sample size of the study for a GWAS."
    - **Response:** A sample size statement was available in the original manuscript methods section: "The sample size was determined as described previously (*19*).". The EPIC-002 main cohort sample size was designed to support multi-omic biomarker analysis and they are quite extensive calculations. For brevity, we refer to the original paper describing various sample size calculations. We acknowledge that for GWAS our sample size is modest, prioritizing detection of moderate to large effect sizes only. In the original manuscript we discussed this as a limitation. Specifically: "Our sample size for GWAS limited identification of the full spectrum of genetic associations. However, our results are meaningful, identifying target molecules for follow-up studies." Also: "Our heritability estimates of 87% from this African cohort is high, explained by our inclusion of non-European LD scores from the UK Biobank to calculate heritability, which can produce noisy estimates in small sample sizes." To address this feedback we have now added the following comments to the Discussion.
        - **Action:** New comment added p15 "Our study focused on an underrepresented African cohort, prioritizing detection of moderate to large

effect sizes. Despite this, our sample size, modest for GWAS, limited identification of the full spectrum of genetic associations."

- **Action:** New comment added to the methods p16: The sample size for EPIC-002 was optimized for multi-omic analysis as described previously (*19*)

- **Reviewer Comment:** "The authors did not perform any meQTL (SNP-methylation) or eQTL (SNP-expression) analysis, which is a little wasted since the data was already available. This could be used for downstream Mendelian Randomization analyses to better establish causality."
  - **Response:** On page 13 we did report eQTL associations of GWAS hits using the FUMA GWAS tool to annotate SNPs to GTEx expression profiles: "The variants on chromosome 22 were whole blood expression QTLs (GTExv8 catalogue) for the CBX6 polycomb transcriptional repressor protein, which is required to balance pluripotency and differentiation in mammals (Table S6) [30]" Table S6 includes the results of this analysis. We appreciate this suggestion to include meQTL analyses. We do acknowledge the value of meQTL analyses that are being pursued as a separate project. We prioritized DMR analysis to capture biologically relevant, time-dependent epigenetic changes in early life. While we acknowledge the value of meQTL analysis for understanding the genetic regulation of methylation, it was not the primary focus of this study, as meQTLs typically represent stable genetic effects and are less relevant to our focus on dynamic, ontogeny-associated changes. However, comprehensive meQTL analysis is an ongoing project utilizing the data generated in this study. In this paper, we chose to prioritize presenting the most novel and biologically relevant findings, namely the characterization of ontogeny-associated DMRs. Integrating meQTL analyses would significantly expand the scope and complexity of the paper, potentially detracting from the primary focus on early-life epigenetic dynamics. Given the lack of meQTL studies in African cohorts we do recognize the value of these studies and are pursuing these as separate research questions.
  -
- **Reviewer Comment:** Table 1: What does the 'Correlation' column refers to - Strength of eQTL association? How are the 'target genes' defined for the analysis – nearest genes, predicted by (?), or?
  - **Response:** We have now clarified this further in the revision.
    - **Actions:** P26. We have added a new footnote to Table 1: *Correlation: Pearson's correlation coefficient between lead CpG in DMR (most significant) and transcript expression for transcripts within 1.5kb from the DMR.*

- **Reviewer Comment:** Figure S4: Pearsons's correlation between Flow cytometry-derived and epigenetically-inferred NLR appears to be strong. What about correlation between individual cell counts by cytometry and estimated cell counts?
  - **Response:** The associations between flow cytometry-derived cell counts and epigenetically-inferred cell counts were indeed very strong and statistically

significant for all individual cell populations contributing to the NLR calculation (i.e., neutrophils, lymphocytes, monocytes, eosinophils, and basophils).

- **Action:** We have amended p11 text to make this explicit: "We examined the correlation between the epigenetically inferred NLR and flow cytometry derived NLR and found strong significant positive correlations ($R = 0.72$, $P < 2.2^{-16}$, Figure S4). Importantly, this strong correlation was also observed for each of the individual cell populations from which NLR was derived (data not shown)."
  We believe that the existing Figure S4, which demonstrates the strong correlation between the overall NLR values, is sufficient to support the validity of our epigenetic inference method. Given the high concordance already shown, presenting additional figures for each individual cell type contributing to NLR would be redundant and unnecessarily increase the length of the supplementary materials.

- **Reviewer Comment** Minor: Page 4: Undefined first use of abbreviation: "DOL"
  - **Response:** We apologies for the confusion as the abbreviation likely was removed during revisions of the initial draft.
    - **Action:** In the abstract and the first instance on p4 we have included 'days of life (DOL)'

Reviewer #2

Specific Comments

- **Reviewer Comment:** Do the samples from the septic group or the focal infection group belong to the main cohort or the confirmation cohort? They wrote: "Twelve newborns developed acute localized infections and 21 developed sepsis within the first week of life and were matched to 33 healthy controls based on sex, vaccination status, DOL, and time of blood collection after birth. " Then, the members of the two cohorts were not followed to see infections/sepsis?
  - **Response:** We agree the clarity of manuscript could be improved. Sepsis outcomes were recorded for the larger main cohort (EPIC-002) only, and the matched controls were from the same cohort. EPIC-002 was sufficiently large enough to observe neonatal sepsis events, however the (EPIC-003) validation cohort was purely intended to validate signatures observed in the main cohort and was not large enough to observe neonatal infections.
    - **Action:** In the revision we have clarified the text on P11:

"In the EPIC-002 main cohort twelve newborns developed acute localized infections and 21 developed sepsis within the first week of life and were matched to 33 healthy controls from the same cohort based on sex, vaccination status, DOL, and time of blood collection after birth."

- **Reviewer Comment:** The use of a public available umbilical cord blood to study the methylation profile of premature birth newborns is a serious limitation of the work. Authors should take into account that most septic process take place in premature neonates thus, it

is more important to know the DNA methylation profiles in these patients.

- **Response:** Thank you for this insightful comment. We agree that mechanisms that predispose premature newborns to sepsis is a research priority. Of note, the EPIC-002 main cohort was not designed as a sepsis study, it is a study of neonatal immunity in which sepsis was observed in a relatively small number of healthy term newborns. The observation that reduced NLR at birth was observed in term newborns who developed sepsis logically raised the question of whether NLR was a general marker of immune maturity. We used publicly available data on preterms to explore this, as there were no pre-term births in EPIC-002, which supported the concept. We do not believe this exploratory analysis constitutes a limitation of our work, rather we view it as enhancing the findings regarding NLR, ontogeny and sepsis risk within the constraints of this study of healthy newborns.

- **Reviewer Comment:** Line 303: Clarify: "This molecular NLR may be a promising marker for severe disease risk stratification in newborns, as we found strong correlations with flow cytometry measures, but stronger associations with early onset sepsis compared to flow."
  - **Response:** Thank you, the text is intended to convey the meaning that NLR derived from methylation data are fundamentally different from NLR derived by flow, and the latter was a stronger risk factor for sepsis in regression models than flow.
    - **Action:** In the revision we have amended the text as follows (p15).

"This molecular NLR may be a promising marker for severe disease risk stratification in newborns, as we found that it correlated strongly with flow cytometry measures of NLR, but exhibited stronger associations with early onset sepsis than NLR derived by flow."

- **Reviewer Comment:** Line 337: ". A second validation cohort comprising 45 newborns was recruited from the main cohort. " Please clarify, if there are 2 validation cohorts? Is this "second" validation cohort part of the main cohort? This must be clear since a validation cohort cannot be included in the main cohort.
  - **Response:** We apologise for the confusion. There were two cohorts and in the revision we have used the original EPIC-HIPC cohort codes to refer to the larger main cohort (EPIC-002), and validation cohort (EPIC-003). Both were recruited from the same study site at different times.
    - **Action**To clarify we have now included the cohort numbers EPIC-002 and EPIC-003 throughout the manuscript
    - **Action:** We have amended the text on P16 to:

A second validation cohort (EPIC-003) comprising 45 newborns was recruited from the same study site for validation purposes, following the same protocol.

- **Reviewer Comment:** In general terms clinical information about the patients enrolled is rather limited.
  - **Response:** We appreciate this comment and agree that clinical information is indeed rather limited for our participants. However, our study by design focused on healthy term newborns that did not have any known medical complications prior to enrolment, i.e. during pregnancy or birth because our primary study endpoint was to establish the trajectory of immune ontogeny in healthy newborns over the first week of life. Given our study's focus on healthy newborns over the first week of

life we thus did not attempt to alter routine clinical care to obtain additional information (e.g. healthy infants would only come to the attention of the study team if a clinical condition was to occur, but otherwise were managed at home through the existing public health care services as per standard practice). However, we included as a secondary endpoint clinically diagnosed neonatal sepsis as the incidence was high enough in this otherwise low-risk cohort to conduct meaningful analysis.

- **Reviewer Comment:** More information about sepsis patients must be included.
    - **Response:** For those clinically diagnosed with sepsis, the clinical information we present in the manuscript was all that is collected during medical care in our study set up. In The Gambia, clinically diagnosed neonatal sepsis is a frequent condition and in the absence of microbiological confirmation, consists of a syndromic approach to diagnosis. While blood cultures (and possibly CSF) will occasionally be obtained (based on the clinical judgement of the responsible physician), very few of these reveal a pathogen in The Gambia (or in high-income settings). Unfortunately, CRP is not available in The Gambia.

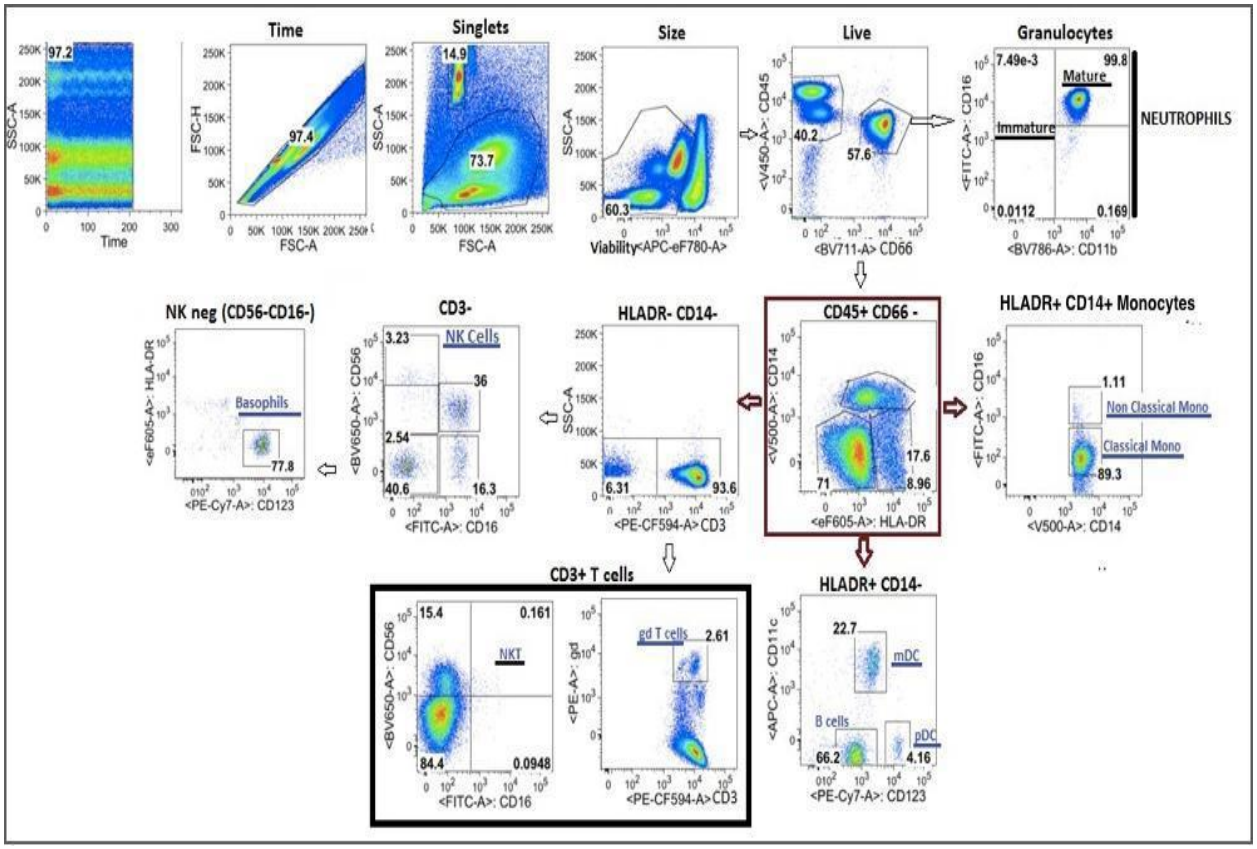
Editorial Note: Gating strategies below reprinted from Montante, S. et al. Breastfeeding and Neonatal Age Influence Neutrophil-Driven Ontogeny of Blood Cell Populations in the First Week of Human Life. *J Immunol Res* **2024**, 1117796, doi:10.1155/2024/1117796 (2024).

Reviewer #3

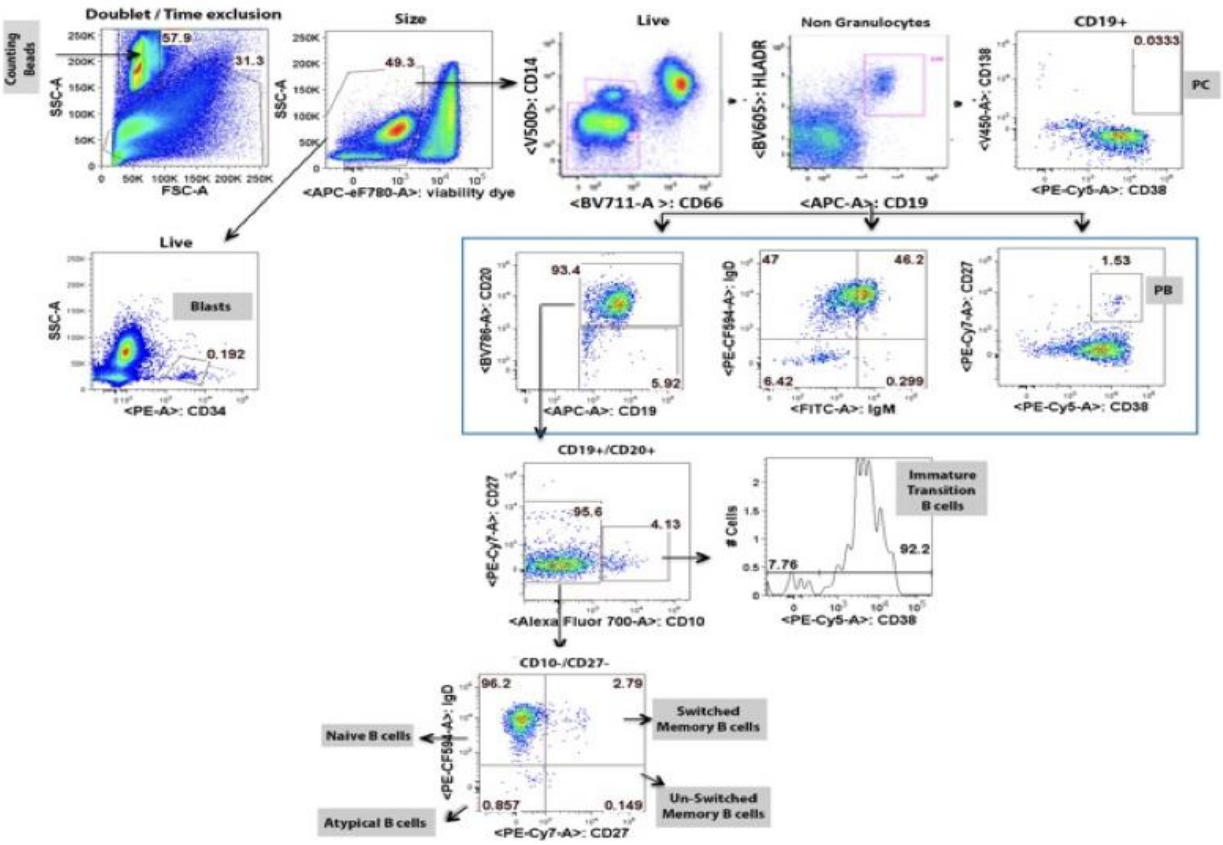
Specific Comments

- **Reviewer Comment:** More clarification in the methods and a representative gating strategy in the supplementary would be helpful for the proper interpretation of the flow cytometry data.
    - **Response:** As we have a separate paper describing flow cytometry for this cohort currently under review, we have provided the gating strategy reported in that paper here for the reviewer to appraise. We will be referencing this paper in the final stages of review in this epigenetic manuscript to refer readers to further details on the flow gating strategies. Two panels and gating strategies were used in this study. One targeting major innate immune cells and a second B-cell specific panel.

Flow cytometry gating strategy for innate immune panel

- 

Flow cytometry gating strategy of B cells panel



- **Reviewer Comment:** The data presented in Figure1 included the samples on DOL 0, 1, 3, and 7, whereas the time points in Figure 3 are DOL 1, 2, 3, 4. It would be more convincing to standardize the time points for sample acquisition.
    - **Response:** We apologise, this is a labelling caused by how the R software interprets group labels. In figure 3 the group levels (DOL0, DOL1, DOL3, DOL7) have been coerced to numerical values, and we did not pick this up.
        - **Action:** In the revision we have corrected Figure 3 to DOL0,1,3,7 as these are the only time points collected.

- **Reviewer Comment:** Statistical analysis should be performed on Figure 2C and p-values should be displayed to support the conclusion that "A reduction in neutrophil counts and increase in T-lymphocytes over the first week of life in both cohorts".
    - **Response:** Thankyou. Although the trends we are describing are obvious and consistent in both datasets, to address this comment we have included a repeated measures ANOVA test and amended figure 2 and its associated text accordingly.
        - **Action:** To clarify we have amended the text on p7:

    "A significant trend for reduction in neutrophil counts and increase in T-lymphocytes with each DOL was observed in both cohorts (Figure 2C) by repeated measures ANOVA ($P < 0.05$), and confirmed by flow cytometry (Figure S3)."

- **Reviewer Comment:** In Figure4, the sepsis were divided into two types include early and late sepsis on D0, whereas there is no classification of sepsis in D1, D3 and D7. And the definition of

sepsis classification should be clarified in method section. Besides, the sample size of validation cohorts is too small to support the result that "epigenetic NLRs were increased in neonates with sepsis in samples collected during the infection period (Figure4B)".

- **Response:** To address this comment we have now amended the methods to include the definitions of EOS and LOS
  - **Action:** The methods p16 now state:

    In line with international agreement, Sepsis was classified as early (<72 hours) or late onset (>72 hours) depending on time of occurrence.
- **Response:** We agree the sample size is limited for this subset of newborns who developed sepsis.
  - **Action:** Figure 4B has been removed in the revision as well as the accompanying text that the reviewer felt was not supported by the analysis.

- **Reviewer Comment:** The description of "Reduced NLRs at birth predicted sepsis outcomes" is inappropriate and the evidence is insufficient. Investigators should include correlation analysis of epigenetic NLR and sepsis outcomes, such as SOFA score.
  - **Response:** For those clinically diagnosed with sepsis, the clinical information we present in the manuscript was indeed all that is normally collected during medical care in our study areas in resource poor settings. Specifically, potential infectious pathogens are rarely identified, nor are peripheral white blood cell counts (including platelets) followed for every septic patient; neither are mean arterial pressures, PaO2/FiO2 captured, or liver and renal function tests routinely performed and or adjunct inflammatory markers (e.g. CRP) are not measured at all. This precludes generating SOFA scores. Regarding the description of NLR as 'predicting' sepsis, this refers to the regression model with sepsis diagnosis as the outcome variable, NLR was a statistically significant predictor, or risk factor. The confusion around the use of the word 'predicted' warrants revision as suggested by the reviewer.
    - **Action:** In the revision we have amended the text to:

    In linear mixed-effect models adjusted for sex, birth weight z-scores, gestational age, and subject ID, the baseline epigenetic NLR was significantly associated with 'any sepsis' outcome diagnosed between DOL1-7

- **Reviewer Comment:** The figure legend of Figure5C is absence. What does genotypes 1, 2, and 3 refer to?
  - **Response:** In the original manuscript the legend for Figure5C was present, however it did not include a description of the genotype codes, and we apologise for the confusion. These codes 0,1,2 refer to the number of copies of the minor allele carried with 0=homozygous major allele, 1=hemizygous, 2=homozygous minor allele.
    - **Action:** In the revised manuscript we have appended the legend for figure 5C as follows:

    (C) Mean difference in log NLR shown as a function of genotype for two top-ranked SNPs on Chromosome 22 (left panel) and Chromosome 18 (right panel). Median and quartiles are shown with exact P-values using the Wilcox test. Genotype codes indicate the presence of minor allele of the SNP variant; 0=homozygous major allele, 1=hemizygous, 2= homozygous minor allele.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have sufficient addressed my comments in the revised version of the manuscript.

Reviewer #2 (Remarks to the Author):

No further comments

Reviewer #3 (Remarks to the Author):

The authors responsed all concerns raised by the reviewer. I have no more questions.