nature portfolio

Corresponding author(s): David Martino

Last updated by author(s): Jul 24, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Not applicable.
Data analysis	DNA methylation analysis was performed using the following Bioconductor packages: EpiDISH (v2.10.0) for cell-type deconvolution, Limma (v3.50.3) for differential methylation analysis, mixOmics (v6.18.1) for within subject variance estimation, DMRcate (v2.8.5) for identification of differentially methylated regions, MissMethyl (v1.28.0) for imputation of missing methylation values, mCSEA (v1.14) for methylation enrichment analysis, and methylclock (v1.0.1) for estimating epigenetic age acceleration. The tidyverse suite of packages (v2.0.0) was employed for data manipulation and visualization. Genetic association analyses were conducted using FUMA GWAS (v1.5.4) for functional mapping and annotation of GWAS results, PLINK2 (v2.00) for data management and statistical analysis, and LDSC (v1.0.1) for estimating genetic heritability. Preprocessing of Illumina methylation array data was performed using Minfi (v1.40.0). Mediation analyses were performed using the mediation package (v4.5.0), and linear mixed effects models were fitted using lmerTest (v3.1-3).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Provide your data availability statement here.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Subject sex was recorded in clinical notes and determined by genetic inference from whole blood samples. In cases where recorded sex differed from genetically inferred sex, subjects were excluded from analysis (Table S7). Regression modelling was adjusted for sex such that findings are generalizable to both sexes.
Reporting on race, ethnicity, or other socially relevant groupings	We report on participant ancestry which was determined by genetic inference in this study. Supplementary figure 5 shows the results of the genetic ancestry analysis relative to the 1kG human reference populations.
Population characteristics	Unselected healthy term newborns were enrolled in this study
Recruitment	The EPIC-002 main cohort of 720 term newborns was enrolled at the Medical Research Council Unit, The Gambia. The sample size for EPIC-002 was optimized for multi-omic analysis as described previously (19). The protocol was approved by the local Ethics Committees and by the Institutional Review Board at Boston Children's Hospital (IRB-P00024239) and is registered at clinicaltrials. gov: https://clinicaltrials.gov/ct2/show/NCT03246230. Informed consent was obtained from the parents or guardians of participants. 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), 5-min Apgar scores > 8, and birth weights > 2.5 kg. Subjects were screened for HIV-I, HIV-II, and hepatitis B and excluded if positive. Venous blood samples (2 mL) were obtained from all neonates within the first 24 h of life (DOL0, Visit 1). Participants were randomized to a second blood draw (visit 2) at either DOL1, DOL3, or DOL7.
Ethics oversight	The protocol was approved by the local Ethics Committees and by the Institutional Review Board at Boston Children's Hospital (IRB-P00024239) and is registered at clinicaltrials. gov: https://clinicaltrials.gov/ct2/show/NCT03246230

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The cohort design included careful consideration of statistical power and sample size. Power calculations in the setting of traditional OMIC designs are well-documented (28) and we performed power calculations for OMIC platforms or flow cytometry based on preliminary data (unpublished data). Although we intended our tests of hypothesis to borrow strength across data types, little has been published on statistical power for integrative analyses or cross-platform comparisons, so this characteristic of our design was not taken into account when determining sample size. Power calculations were carried out using group sizes ranging between n = 30 or n = 60 per group, two-sided significance criterion (alpha) of 0.05, and a target 80% power to detect. Since preliminary data for flow cytometry were not available, we performed a worst-case power analysis using a real-world published data on the most variable cell subpopulation where a significant difference was still identified comparing healthy, immunosuppressed, and transplant-tolerant individuals (29). This provided a lower bound of the minimum sample size required to detect a significant difference in cell sub-populations that could act as potential biomarkers. We found n = 30 would result in 80% power to detect a 2.3-fold difference in the relative abundance of a cell population, while n = 60 would result in 80% power to detect 1.8-fold difference, assuming a balanced design. For this epigenetic paper, the sample n=1,276 and the participant n=673
Data exclusions	Poor quality methylation profiles were excluded during QC and detailed in Table S7. Samples that were mismatched between clinically reported sex and genetically inferred sex were excluded (TableS7)

Replication	We recruited a validation cohort and replicated associations in that cohort
Randomization	Computer-generated block randomization sequence. In The Gambia, the code allow block randomization of 720 study participants into 12 groups, using random block sizes of 24 and 48. Each study participant was assigned sequential identification numbers (IDs) starting from 001 to 720 for The Gambia. A check digit character was augmented to each of these numbers to make up the participant IDs which will be generated using the Damm algorithm (https://en.wikipedia.org/wiki/Damm_algorithm; date accessed 08 November 2016). Here the check digit numbers 0–9 were converted to letters A-H and J-K. The IDs for the Gambian cohort were prefixed with the letter "G" which stands for Gambia. The randomization code used the web application developed by the Statistics and Bioinformatics department at MRCG to generate the IDs including the check digit and prefix as detailed above (the app is available on https://stats.mrc.gm/chkdgt-id/).
Blinding	Due to the nature of randomized groups and difficulty of blinding to randomized vaccine groups, staff in direct contact with participants were unblinded, while all laboratory staff at the sites were blinded to participants assigned group throughout the study

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	The study is registered on clinicaltrials.gov with registration number NCT03246230.
Study protocol	The protocol is available on clinicaltrials.gov with registration number NCT03246230.
Data collection	Data collection occurred in The Gambia at the Medical Research Council Unit, The Gambia. Recruitment occurred during 2017-2022
Outcomes	Molecular signature correlating with anti-hepatitis B vaccine antibody response

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*