nature research

Corresponding author(s): Trevor D. Lawley
Nigel Field

Last updated by author(s): Aug 31, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	No software was used for data collection.
Data analysis	Bowtie2 v2.3.0, vegan_2.5-5, phyloseq_1.28.0, microbiome_1.6.0, ggplot2_3.1.1, ggpubr_0.1.8, Maaslin_0.0.5, R v3.6.0, Trimmomatic v0.35, Kraken v1.0, Bracken v1.0, MetaPhlAn v2.6, StrainPhlAn v1.0, CheckM v1.0.12, FastANI v1.1, Roary 1.007001, Prokka v1.13, RAxML v8.0.0, BIONJ, MAFFT v7.2040, FastTree v2.1.10, ABRicate v0.8.13, MLSTcheck v2.1.1630910, Mash v2.1, iTOL v4

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data generated and analysed in this study have been deposited in the European Nucleotide Archive under accession numbers ERP115334 and ERP024601. Raw faecal samples and bacterial isolates are available from the corresponding authors upon request.

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	No sample size calculation was employed as this is an observational study. Participants were recruited on a voluntary basis in hospitals. This study presents the largest gut metagenome dataset of newborn babies (596 individuals, 1202 samples)
Data exclusions	None. All sequencing data that passed the pre-established sequencing quality control (described in Methods) were included for analysis.
Replication	No replication was employed as this is an observational study.
Randomization	Randomization was not employed as this is an observational study. Participants were recruited on a voluntary basis in hospitals.
Blinding	No blinding used as this is an observational 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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study population are 771 mothers (mean 33.82±4.99 years of age) and their term newborn babies (49.4% female, 50.6% male) born in three NHS hospitals in the UK (London and Leicester). Babies were sampled in their neonatal period at 4 to 21 days of age, and further sampled in later infancy at 4 to 12 months of age.
Recruitment	Participants were recruited and consented before or after birth whilst on the labour ward, by their attending midwives. All pregnant women at the labour ward to give birth were eligible to participate, except those under 16 years of age and non-UK residents intending to return abroad immediately after delivery.
Ethics oversight	The study was approved by the NHS London - City and East Research Ethics Committee (REC reference 12/LO/1492).

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