nature portfolio

Corresponding author(s): Mahesh S. Desai

Last updated by author(s): May 20, 2022

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	No custom software was used for data collection.
Data analysis	No custom algorithms or software were used for data analysis. The analysis pipelines in this study employ tools that have been previously described in published literature and are described below. The code used to analyze these data can be found at https://github.com/DII-LIH-Luxembourg. Mass cytometry data analysis: Each organ was analyzed independently following the same analysis pipeline, as previously described by Leonard et al (doi:10.1111/all.14716.). FCS files were normalized with the normalization passport EQ-P13H2303_ver2. FCS files were uploaded into FlowJo™ software for cleaning. Unsupervised analysis was carried out in RStudio (version 1.0.143, R version 3.4.4) using the R package 'flowcore' version 1.44.2. Single cells were clustered using FlowSOM. Statistical analysis was performed using the R package 'limma' version 3.34.9 and 'lme4' version 1.1-18-1. Differential abundance between the conditions was tested by using the R package 'multcomp' version 1.4-8. One-way ANOVA followed by Tukey's post hoc test and adjustment for multiple comparisons using Benjamini-Hochberg method was performed using GraphPad Prism version 9.0.0 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com). 16S rRNA gene sequencing and analysis: Raw sequences were processed using QIIME2 version 2020.6 with DADA2 for sequence quality control and taxonomic assignment was performed using VSEARCH against the SILVA 138 reference database. Further analyses were performed in R version 4.0.2 using the package 'phyloseq' version 1.34.0. PCoA plots were generated using the package 'vegan' version 2.5-7, with clustering significance testing using package 'pairwiseAdonis' version 0.4. Differential enrichment analysis was performed using the package 'DESeq2' version 1.30.1, which implements the Wald test to determine statistical significance (p value adjustment using Benjamini-Hochberg method). Visualizations were generated using 'ggplot2' version 3.3.5 and 'forcats' version 0.5.1. CE-TOFMS-based metabolomics: Analysis of metabolome spectra and concentration calculation was done using the proprietary software "MasterHands" version 2.19.0.2 (Keio University) as previously described (doi:10.1007/s11306-009-0178-y). Heatmaps of top differentially

abundant cecal metabolites were generated based on a linear mixed-effect model using limma in MetaboAnalyst 5.0. Visualization of intersecting sets were performed using 'ComplexUpset' version 1.3.3 and 'ggVennDiagram' version 1.2.0. PCA plots were generated using the 'stats' version 4.0.279, 'factoextra' version 1.0.7, and members of the tidyverse ecosystem ('dplyr' version 1.0.5, 'ggfortify' version 0.4.13, 'ggforce' version 0.3.3, 'ggrepel' version 0.9.1). Correlation matrix between B vitamin concentrations and immune cell frequencies was calculated using 'corrplot' version 0.92.

RNA sequencing and analysis: Fastq files were processed using KneadData, within the bioBakery suite of workflows. Adapters were removed using Trimmomatic and fragments below 50% of the total expected read length were filtered out. Bowtie2 was used to map and remove contaminant reads corresponding to either rRNA databases or the mouse genome. Clean fastq files were concatenated and passed to HUMAnN 3 for metagenome mapping with MetaPhlAn 3. In this pipeline, unaligned reads were translated using DIAMOND for protein identification. Gene family abundances, pathway abundances, and pathway coverage data for all samples were each joined into a single table and re-calculated as counts per million (CPM) and relative abundance using HUMAnN 3. Statistically significant changes in the abundance of transcripts mapping to genes up or downstream of B vitamins were analyzed in GraphPad Prism version 9.0.0 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com) by one-way ANOVA, followed by Tukey's multiple comparison test (adjustment using Benjamini-Hochberg method).

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

The raw fastq files for this study from 16S rRNA gene sequencing and RNA sequencing have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB51707 (https://www.ebi.ac.uk/ena/browser/view/PRJEB51707). The mass cytometry datasets for colonic lamina propria, lung, and spleen have been uploaded to the FlowRepository database under accession number FR-FCM-Z56W (https://flowrepository.org/). Raw spectral data from CE-TOFMS are available under Project ID PR001381 on Metabolomics Workbench105 (https://www.metabolomicsworkbench.org), which is the NIH Common Fund's National Metabolomics Data Repository (NMDR), supported by NIH grant U2C-DK119886. The data can be accessed directly via the Project DOI: 10.21228/M86T35.

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	Under the original animal protocol, a minimum sample size of 5 mice per group was determined to be sufficient to have at least 90% power to detect statistically significant ($p<0.05$) changes in mucus layer thickness. As this readout was not assessed in the present publication, we further verified the sample size requirements by performing a second power calculation based on the total SCFA concentration in cecal contents, according to a previous study in specific-pathogen-free mice fed a fiber-rich 1 (FR1) or fiber-free (FF) diet (doi: 10.1080/19490976.2021.1966263).
Data exclusions	In 16S gene sequencing analyses, taxa not observed more than once on average across all samples were removed and the data was rarefied to the minimum library size. For RNA sequencing data, host RNA, ribosomal RNA and viral RNA were excluded from analyses. Unmapped reads were also ignored in downstream analyses after confirming that there were no bacterial hits with significant sequence similarity using NCBI BLAST. The cecal and serum metabolomes of mouse ID 569 (FF-fed group) was excluded from subsequent analyses because it fell outside the 0.975 confidence interval for the group centroid (PC1 vs PC2), as calculated when the suspected outlier was removed. We report the metabolite concentrations for this sample in Supplementary Table and confirm that the overall conclusions made in this study are unaffected by its exclusion. Only metabolites which were present in at least half of one experimental group were included in downstream analyses.
Replication	We conducted the experiment with 5 distinct diet groups under specific-pathogen free conditions and 2 diet groups under gnotobiotic (14SM) and germ-free conditions. We note low internal variation in these groups and that the study was adequately powered based on the statistical significance of key readouts. Consequently, it was determined that replication of the study and use of additional animals was not warranted.
Randomization	Mice were housed in individual cages with up to 5 mice per cage. Each cage was assigned at random to one of the 5 dietary groups: fiber-rich 1, fiber-rich 2, inulin-supplemented, fiber-supplemented, or fiber-free.
Blinding	Blinding was not determined to be necessary for these animal experiments because none of the readouts were subject to subconscious researcher biases (e.g. disease symptom or histological scoring).

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
[ ]	[x] Antibodies
[x]	[ ] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[ ]	[x] Animals and other organisms
[x]	[ ] Human research participants
[x]	[ ] Clinical data
[x]	[ ] Dual use research of concern

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[x]	[ ] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used

All antibodies were ordered from Fluidigm. Batch numbers are listed in order by order date (2018, 2019a, 2019b, 2020):
CD45 (clone 30-F11 isotype IgG2b) labelled with 89Y; batch # 0581829, 0161909, 0161909, 0622022
Ly-6G (clone 1A8 isotype Rat IgG2a) labelled with 141Pr; batch # 2751702, 1201823, 1401910, 0132018
CD45R (B220) (clone RA3-6B2 isotype Rat IgG2a) labelled with 144Nd; batch # 2851711, 2851711, 2851711, 2851711
CD69 (clone H1.2F3 isotype Hamster IgG) labelled with 145Nd; batch # 3511606, 3121831, 2211909, 1322039
F4/80 (clone BM8 isotype Rat IgG2a) labelled with 146Nd; batch # 3121701, 0211903, 1901904, 2601904
CD11b (Mac-1) (clone M1/70 isotype Rat IgG2b) labelled with 148Nd; batch # 3171702, 1401907, 1401907, 1401907
CD19 (clone 6D5 isotype Rat IgG2a) labelled with 149Sm; batch # 0761711, 581828, 1901905, 1901905
CD25 (IL-2R) (clone 3C7 isotype Rat IgG2b) labelled with 150Nd; batch # 1171712, 081811, 1981902, 3401917
IgM (clone RMM-1 isotype Rat IgG2a) labelled with 151Eu; batch # 3001411, 1171710, 2681906, 2681906
CD3e (clone 145-2C11 isotype Hamster IgG) labelled with 152Sm; batch # 3181714, 2971806, 2971806, 0132008
CD8a (clone 53-6.7 isotype Rat IgG2a) labelled with 153Eu; batch # 1151403, 0101806, 2531901, 2531901
Foxp3 (clone FJK-16s isotype Rat IgG2a) labelled with 158Gd; batch # 3461707, 1841813, 2421903, 2421903
RORgt (clone B2D isotype IgG1) labelled with 159Tb; batch # 3181709, 3181709, 3181709, 0912006
CD5 (clone 53-7.3 isotype Rat IgG2a) labelled with 160Gd; batch # 2791503, 2791503, 2931525, 2791503
Tbet (clone 4B10 isotype IgG1) labelled with 161Dy; batch # 3391715, 3391715, 3181901, 3181901
Ly-6C (clone HK1.4 isotype Rat IgG2c) labelled with 162Dy; batch # 2341706, 1201817, 2461904, 3431915
CD62L (clone MEL-14 isotype Rat IgG2a) labelled with 164Dy; batch # 2851703, 2851703, 2691906, 2691906
CD326 (EpCAM) (clone G8.8 isotype Rat IgG2a) labelled with 166Er; batch # 0251601, 0251601, 0251601, 0251601
Gata3 (clone TWAJ isotype Rat IgG2b) labelled with 167Er; batch # 3481606, 0391808, 2551902, 0592011
CD206 (MMR) (clone C068C2 isotype Rat IgG2a) labelled with 169Tm; batch # 0711816, 1341903, 2671901, 0282009
CD49b (clone HMa2 isotype Hamster IgG) labelled with 170Er; batch # 2851710, 3511803, 3511803, 3511803
CD44 (clone IM7 isotype Rat IgG2b) labelled with 171Yb; batch # 1931725, 1201828, 2461903, 3181915
CD4 (clone RM4-5 isotype Rat IgG2a) labelled with 172Yb; batch # 1201808, 3391809, 3391809, 3391809
CD117 (ckit) (clone 2B8 isotype Rat IgG2b) labelled with 173Yb; batch # 0331524, 2631811, 2631811, 0452003
I-A/I-E (MHC Class II) (clone M5/114.15.2 isotype Rat IgG2b) labelled with 174Yb; batch # 0791514, 2631807, 2631807, 2631807
CD38 (clone 90 isotype Rat IgG2a) labelled with 175Lu; batch # 0341403, 2041807, 2041807, 2041807
FceR1a (clone 36951 isotype Hamster IgG) labelled with 176Yb; batch # 2381302, 0381905, 0381905, 0381905
CD11c (clone N418 isotype Hamster IgG) labelled with 209Bi; batch # 0081812, 1521804, 1521804, 2007297-27

Validation

Standard, commercially available antibodies and their metal tags were selected in consultation with specialists from Fluidigm and validated on test samples prior to their application in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Experiments were performed with either specific-pathogen-free, gnotobiotic, or germ-free BALB/c mice, purchased from Charles River Laboratories, France. All mice were wildtype, 6-8 weeks old at the start of the study, and female.

Wild animals

Not applicable.

Field-collected samples

Not applicable.

Ethics oversight

Gnotobiotic animal experiments were approved by the Luxembourgish Ministry of Agriculture, Viticulture and Rural Development (LUPA 2019/50). Specific-pathogen-free experiments did not require approval because the feeding was not classified as a manipulation following veterinarian review. All experiments were performed according to the "Règlement grand-ducal du 11 janvier

2013 relatif à la protection des animaux utilisés à des fins scientifiques" based on the "Directive 2010/63/EU" on the protection of animals used for scientific purposes.

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