nature research

Corresponding author(s): Sarkis Mazmanian, Ashlee Earl, Gregory Donaldson

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	RNAseq data was collected on an Illumina HiSeq2500 Flow cytometry data was collected on a Miltenyi MACSQuant
Data analysis	Reads were trimmed using FASTX toolkit v.0.0.13 Bacterial read mapping was performed with Bowtie2 2.2.9 Read counts per gene were quantified using bedtools 2.25.0 Differential expression analyses were performed with edgeR 3.12.1 Genomic comparisons were made using the MEME suite 4.12.0 (GLAM2 and GLAM2Scan) Structural modeling was done with Phyre 2.0 Graphing and statistical analysis was performed with GraphPad Prism 7 Flow Cytometry data were collected using MACSQuantify 2.0 Flow Cytometry data were analyzed with FlowJo 10.0.8

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 upon request. 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

RNA-Seq and hsRNA-Seq data have been deposited in the NCBI Short Read Archive under a project accession number, PRJNA438372. The B. fragilis NCTC9343 genome used for mapping is available at Genbank GCA_000025985.1. Analysis code is avilable on Github: https://github.com/wenchichou/bugInHost. All other source data are provided with the paper.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes were based on previous published reports or empirically chosen based on pilot studies because expected effect sizes were unknown.
Data exclusions	Data were not excluded.
Replication	In vivo experiments were separately performed on at least two cohorts of mice (born months apart). Replication was successful in each case.
Randomization	Age-matched mice (within 1 week) were randomly sorted into experimental groups.
Blinding	Blinding was used for histological scoring (Fig. 5d). All other experiments were unblinded because the microbiological data collection methods are objective (ie CFU plating).

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

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

Antibodies

Antibodies used	rat anti-mouse CD4 (clone RM4-5) PE-Cy7 (eBioscience ThermoFisher #25-0042-81) (1:200 dilution) rat anti-mouse IFNgamma (clone XMG1.2) FITC (eBioscience ThermoFisher #11-7311-41) (1:200 dilution) rat anti-mouse IL-10 (clone JES5-16E3) PE (eBioscience ThermoFisher #12-7101-81) (1:200 dilution) rat anti-mouse IL-17A (clone eBio17B7) PerCP-Cy5.5 (eBioscience ThermoFisher #45-7177-82) (1:200 dilution) rat anti-mouse FOXP3 (clone FJK-16s) APC (eBioscience ThermoFisher #17-5773-82) (1:200 dilution) We did not keep record of the lot numbers
Validation	All were validated for flow cytometry using mouse samples by the manufacturer. Further validation details: CD4: relative expression verification by manufacturer and 80 citations available on website IFNgamma: cell treatment verification by manufacturer and 140 citations available on website

IL-10: 43 citations available on website
IL-17A: 54 citations available on website
FOXP3: relative expression verification by manufacturer and 270 citations available on website

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NA
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Animals and other organisms

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

Laboratory animals	Swiss Webster and C57BL/6 mice (Mus musculus) were purchased from Taconic Farms and bred at Caltech. For all experiments female 7-8 week-old germfree mice were used with the following exceptions: Fig. 3d-f (half male, half female), DNBS experiments (3 week-old, just-weaned mice were colonized to allow the induction of colitis at 7 weeks of age). For Fig. 4 mice with full microbiome, 7 week-old excluded flora Swiss Webster mice were purchased from Taconic.
Wild animals	No wild animals
Field-collected samples	No field-collected samples

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Mesenteric lymph nodes from gnotobiotic C57BL/6 mice were isolated and processed by dissociating tissues through a 70 µm cell strainer (BD Falcon) to generate single cell suspensions.
Instrument	Miltenyi MACSQuant 10
Software	The MIltenyi software for the MACSQuant was used for acquisition of data. Flowjo 10.0.8 was used to analyze the data.
Cell population abundance	No sorting in this study
Gating strategy	A live/dead stain over SSC gate was used to identify live cells followed by a CD4+ gate to identify T cells.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.