Moving beyond microbiome-wide associations to causal microbe identification

Neeraj K. Surana1,2 & Dennis L. Kasper1

Microbiome-wide association studies have established that numerous diseases are associated with changes in the microbiota1,2. These studies typically generate a long list of commensals implicated as biomarkers of disease, with no clear relevance to disease pathogenesis3,4. If the field is to move beyond correlations and begin to address causation, an effective system is needed for refining this catalogue of differentially abundant microbes and to allow subsequent mechanistic studies3,4. Here we demonstrate that triangulation of microbe–phenotype relationships is an effective method for reducing the noise inherent in microbiota studies and enabling identification of causal microbes. We found that gnotobiotic mice harbouring different microbial communities exhibited differential survival in a colitis model. Co-housing of these mice generated animals that had hybrid microbiotas and displayed intermediate susceptibility to colitis. Mapping of microbe–phenotype relationships in parental mouse strains and in mice with hybrid microbiotas identified the bacterial family Lachnospiraceae as a correlate for protection from disease. Using directed microbial culture techniques, we discovered Clostridium immunis, a previously unknown bacterial species from this family, that—when administered to colitis-prone mice—protected them against colitis-associated death. To demonstrate the generalizability of our approach, we used it to identify several communal organisms that induce intestinal expression of an antimicrobial peptide. Thus, we have used microbe–phenotype triangulation to move beyond the standard correlative microbiome study and identify causal microbes for two completely distinct phenotypes. Identification of disease-modulating commensals by microbe–phenotype triangulation may be more broadly applicable to human microbiome studies.

The microbiota regulates various facets of host physiology, including immune responses, metabolic functions, and behaviour5. Numerous microbiome-wide association studies (MWAS) have linked the microbiome to a panoply of diseases, offering hope that rational microbiome alteration is a feasible treatment modality for many ailments5. However, MWAS have produced long lists of implicated microbes without clearly elucidating their causal role, correlations have not always held up in subsequent studies, and notable differences have been observed between human and animal studies5,6. Although many correlations may simply reflect biomarkers of disease, causal links between the microbiome and disease susceptibility have been sporadically defined in studies that have initially identified immunomodulatory bacteria, functionally categorized the microbiota on the basis of immune recognition, or used complex bioinformatic heuristics to identify disease-modulating bacteria6,7. Ultimately, a generalizable pathway that refines the catalogue of differentially abundant microbes identified by MWAS to include only those most probably causally related to the phenotype is lacking.

Previously, we generated and characterized gnotobiotic mice colonized with a mouse microbiota (MMb) or a human microbiota (HMb)8,9. We have now explored how these mice differ in susceptibility to colitis, using a chemically induced model of inflammatory bowel disease. We found that—similar to germ-free mice—MMb mice were exquisitely sensitive to dextran sodium sulfate (DSS)-induced colitis, with severe weight loss and 100% mortality (Fig. 1a, b). By contrast, HMb and specific pathogen-free (SPF) mice lost significantly less weight and rarely died (Fig. 1a, b). This dichotomy in survival did not reflect lack of colonic inflammation in HMb or SPF mice: the degree of inflammation did not differ at day 5. However, MMb mice had slightly more severe inflammation than either SPF or HMb mice at day 10 (Fig. 1c).

We compared the faecal microbiota of HMb and MMb mice in an attempt to identify the microbe(s) responsible for modulating disease severity. Given the different underlying microbial sources for these mice, we were not surprised to identify more than 150 bacterial taxa that were differentially abundant between the two groups (Extended Data Fig. 1a). Comparison of the faecal microbiota of MMb and SPF mice demonstrated approximately 100 differentially abundant bacterial taxa (Extended Data Fig. 1b). In both cases, the sheer number of implicated microbes makes prioritizing and pursuing potential leads extremely challenging. Focusing on some of the most differentially abundant organisms detected in MWAS has identified some of the few microbes that have been causally related to a phenotype4. Using this approach in HMb and MMb mice, we identified 26 taxa present in one group and absent in the other. We chose four of these organisms, orally administered them to HMb or MMb mice that lacked them, and challenged the mice with DSS. No organism significantly augmented colitis severity (Extended Data Fig. 2). This result highlights the challenges of relying solely on organisms that can be dichotomized as present or absent.

The general inability to move beyond correlations and address causation has been the Achilles heel of microbiome research. In considering how to render microbiome analyses more specific, we noted that typical MWAS often compare microbial populations with little similarity to one another1,8—a feature that our study of MMb and HMb mice exemplifies. Comparison of animals with more similar microbiotas should result in a shorter list of phenotype-associated microbial taxa. We reasoned that co-housing of mice with different microbiotas would generate hybrid-microbiota animals microbially related to their parental strains. If the microbial effect on disease were dominant, mapping of microbe–phenotype relationships in parental mouse strains and in hybrid-microbiota mice would enable us to triangulate disease-modulating organisms.

To determine whether the microbiota’s effect is dominant in susceptibility to DSS-induced colitis, we co-housed HMb and MMb mice for 3 weeks. Both HMb mice co-housed with MMb mice (HMb8MMb-3w) and MMb mice co-housed with HMb mice (MMb8HMb-3w) had intermediate phenotypes (Fig. 2a), a result that suggests bi-directional microbe transfer. To limit the degree of microbial change, we defined the shortest period of co-housing that produced a phenotypic difference, co-housing HMb and MMb mice for 1 or 3 days and then separating the two groups. We challenged mice with DSS on day 14 after initial co-housing to allow...
Published comprehensive analysis of the immunomodulatory capacity of commensal microbes are known to be required for induction of Reg3γ, few specific organisms have been defined. In our recently reported analysis of faecal microbiota of mice and human subjects, Lachnospiraceae was Lachnospiraceae. This family of Gram-positive, anaerobic, abundant bacterial taxa (Fig. 2e and Extended Data Fig. 1a–d). We previously showed that HMb mice have ileal levels of Reg3γ, which are important in maintaining spatial segregation between the epithelial layer and the microbiota16. Although commensal microbes are known to be required for induction of Reg3γ expression17, few specific organisms have been defined. In our recently published comprehensive analysis of the immunomodulatory capacity of taxonomically diverse commensal microbes, none of 28 bacteria assessed for small-intestinal expression of Reg3γ caused significant induction18. We previously showed that HMb mice have ileal levels of Reg3γ expression similar to those in germ-free mice and lower than those in MMb mice12. We now found that Reg3γ expression in HMbHMb-1d mouse was restored to levels comparable to those in MMb mice, with no change observed in MMbHMb-1d mice (Fig. 4a). Using microbial–phenotype triangulation, we identified taxonomic that were differentially abundant in two pairwise comparisons but were unchanged between MMb and MMbHMb-1d mice (Fig. 4b). Thus we refined the list of associated microbes to seven taxa, three of which (Ruminococcus gnavus, Lactobacillus reuteri, and segmented filamentous bacteria (SFB)) exhibit bioinformatic resolution to the species level (Fig. 4b and Extended Data Table 1).
HMb mice treated with R. gnavus or L. reuteri had greater ileal Reg3γ expression than a control (Fig. 4c). The previous demonstration that SFB induces Reg3γ expression in germ-free and SPF mice confirms the more general applicability of results obtained using our specific gnotobiotic mice.8 Our requirement that organisms be unchanged between MMb and MMbHMb-1d mice correctly excluded a relative of L. reuteri and SFB (present in MMb mice) or 0.2% and 0.3% abundance, respectively, and absent in MMbHMb-1d mice. Correctly identifying organisms associated with our phenotype and substantiated these correlations by add-back experiments to fulfill Koch’s postulates. Although Koch’s postulates and their modern revisions traditionally apply only to pathogen association with disease, our findings with a disease-protective commensal (Extended Data Fig. 2). Taken together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.

Our results delineate a bioinformatically straightforward approach that can be used to triangulate specific microbiota members that are likely to influence disease pathogenesis. Combination of microbiota analyses from multiple sets of microbiobically related mice allowed us to trade sensitivity for increased specificity. In both the colitis and Reg3γ analyses, we bioinformatically pinpointed a limited number of taxa identified for Reg3γ induction had no effect in the colitis model. (Extended Data Fig. 2). Together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.

HMb mice treated with R. gnavus or L. reuteri had greater ileal Reg3γ expression than a control (Fig. 4c). The previous demonstration that SFB induces Reg3γ expression in germ-free and SPF mice confirms the more general applicability of results obtained using our specific gnotobiotic mice.8 Our requirement that organisms be unchanged between MMb and MMbHMb-1d mice correctly excluded a relative of L. reuteri and SFB (present in MMb mice) or 0.2% and 0.3% abundance, respectively, and absent in MMbHMb-1d mice. Correctly identifying organisms associated with our phenotype and substantiated these correlations by add-back experiments to fulfill Koch’s postulates. Although Koch’s postulates and their modern revisions traditionally apply only to pathogen association with disease, our findings with a disease-protective commensal (Extended Data Fig. 2). Taken together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.

Our results delineate a bioinformatically straightforward approach that can be used to triangulate specific microbiota members that are likely to influence disease pathogenesis. Combination of microbiota analyses from multiple sets of microbiobically related mice allowed us to trade sensitivity for increased specificity. In both the colitis and Reg3γ analyses, we bioinformatically pinpointed a limited number of taxa identified for Reg3γ induction had no effect in the colitis model. (Extended Data Fig. 2). Together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.

Our results delineate a bioinformatically straightforward approach that can be used to triangulate specific microbiota members that are likely to influence disease pathogenesis. Combination of microbiota analyses from multiple sets of microbiobically related mice allowed us to trade sensitivity for increased specificity. In both the colitis and Reg3γ analyses, we bioinformatically pinpointed a limited number of taxa identified for Reg3γ induction had no effect in the colitis model. (Extended Data Fig. 2). Together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.

Our results delineate a bioinformatically straightforward approach that can be used to triangulate specific microbiota members that are likely to influence disease pathogenesis. Combination of microbiota analyses from multiple sets of microbiobically related mice allowed us to trade sensitivity for increased specificity. In both the colitis and Reg3γ analyses, we bioinformatically pinpointed a limited number of taxa identified for Reg3γ induction had no effect in the colitis model. (Extended Data Fig. 2). Together, microbe–phenotype triangulation facilitated identification of taxonomically diverse microbes causally related to two disparate outcomes.
facilitating the discovery of key disease-modulating components of the microbiota.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 February; accepted 7 November 2017.
Published online 6 December 2017.

1. Lynch, S. V. & Pedersen, O. The human intestinal microbiome in health and disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).
2. Surana, N. K. & Kasper, D. L. Deciphering the tête-à-tête between the microbiota and the immune system. *J. Clin. Invest.* **124**, 4197–4203 (2014).
3. Sze, M. A. & Schloss, P. D. Looking for a signal in the noise: revisiting obesity and the microbiome. *MBio* **7**, e01018–16 (2016).
4. Gilbert, J. A. et al. Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* **535**, 94–103 (2016).
5. Gevers, D. et al. The treatment-naive microbiome in new-onset Crohn’s disease. *Cell Host Microbe* **15**, 382–392 (2014).
6. Atarashi, K. et al. Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
7. Blanton, L. V. et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* **351**, aad3311 (2016).
8. Ivanov, I. I. et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–499 (2009).
9. Kau, A. L. et al. Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Cell Transl. Med.* **7**, e2724 (2015).
10. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).
11. Palm, N. W. et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158**, 1000–1010 (2014).
12. Chung, H. et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578–1593 (2012).
13. Baumgart, M. et al. Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn’s disease involving the ileum. *ISME J.* **1**, 403–418 (2007).
14. Frank, D. N. et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
15. Kang, S. et al. Dysbiosis of fecal microbiota in Crohn’s disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* **16**, 2034–2042 (2010).
16. Vaishnava, S. et al. The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255–258 (2011).
17. Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bacterial factor. *Science* **331**, 1126–1130 (2006).
18. Geva-Zatorsky, N. et al. Mining the human gut microbiota for immunomodulatory organisms. *Cell* **168**, 928–943 (2017).
19. Falkow, S. Molecular Koch’s postulates applied to microbial pathogenicity. *Rev. Infec. Dis.* **10** (Suppl. 2), S274–S276 (1988).
20. Fredricks, D. N. & Relman, D. A. Sequence-based identification of microbial pathogens: a reconsideration of Koch’s postulates. *Clin. Microbiol. Rev.* **9**, 18–33 (1996).
21. Koch, R. Untersuchungen über Bakterien: V. Die Atiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus anthracis. *Cohns Beitr. Biol. Pflanz.* **2**, 277–310 (1876).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Couter for technical assistance; S. Edwards, J. Ramos, and T. Sherpa for assistance with gnotobiotic mice; R. Bronson for review of histology; J. McCoy for editorial assistance; and members of the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author Contributions N.K.S. conceived the study, designed and performed experiments, and analysed all data. D.L.K. supervised all aspects of the project. N.K.S. and D.L.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to N.K.S. (neeraj.surana@childrens.harvard.edu) or D.L.K. (dennis.kasper@hms.harvard.edu).

Reviewer Information Nature thanks J. Faith, M. Lathrop, A. Macpherson and the other anonymous reviewer(s) for their contribution to the peer review of this work.
METHODS

Mice. Germ-free Swiss Webster mice were bred and maintained in vinyl isolators in the animal facility at Harvard University. MMb and HMb mice have been bred and maintained at this facility in separate vinyl isolators since their initial characterization. Experimental manipulation of gnotobiotic mice was performed in sterile cages (Innovive) in which animals received autoclaved food and water. SPF Swiss Webster mice obtained from Taconic Biosciences were fed an autoclaved diet similar to that given to gnotobiotic mice for 1 week before the start of the study and for the duration of all experiments. Mice used in experiments were sex- and age-matched (typically 5–10 weeks old) and drawn randomly from the same litter, when feasible. All procedures were approved by the Harvard Medical Area Standing Committee on Animals and were conducted in accordance with National Institutes of Health guidelines.

DSS colitis. DSS experiments were performed as previously described. In brief, mice were given 4% DSS (molecular mass 36,000–50,000 Da; MP Biomedicals) ad libitum in their drinking water for 7 days, with the DSS solution changed every 2–3 days. From day 7 until the end of the experiment, the mice were given autoclaved water with no DSS. Animals were weighed every 1–2 days, and any mouse that appeared moribund was euthanized. A pathologist blinded to treatment groups conducted a histological assessment of colons. The histological score represented the combined scores for inflammation and ulceration; both elements were scored 0–4, with 0 being normal.

Co-housing experiments. For co-housing of MMb and HMb mice before DSS experiments, we placed two mice per group together in a cage (total, four mice) for the indicated period. For a co-housing period of 1 day or 3 days, the MMb and HMb mice were separated at the relevant time point, placed into a new sterile cage until day 14 (to allow physiological changes to occur), and then challenged with DSS. For co-housing of MMb and HMb mice for faecal microbiota analysis, we placed one mouse per group in a cage (total, two mice) for 1 day, after which the mice were separated and individually housed until day 14. Faecal pellets were collected on days 0 and 14 and were frozen at −80 °C until further processed.

16S rDNA sequencing and analysis. Faecal samples were added to a tube containing 400 μl of zirconia/silica beads (0.1 mm in diameter; Biospec), 20% sodium dodecyl sulfate, 500 μl of buffer PBI (Qiagen), and 250 μl of phenol: chloroform:isoamyl alcohol (25:24:1; pH 7.9; Ambion) and then homogenized by bead beating for 2.5 min. After centrifugation (10,000 g for 5 min), the DNA in the aqueous phase was purified with a QIAquick PCR purification kit (Qiagen) according to the manufacturer’s instructions except that the PE wash step was performed twice. DNA was eluted with 50 μl of EB buffer (Qiagen).

Detailed protocols used for 16S rDNA amplification and sequencing have been described. In brief, the V4 region of the 16S rDNA gene was PCR-amplified (35 cycles, primers 515F and 806R) in triplicate; 1 μl of purified DNA and 5′ Hot Master Mix (Five Prime) were used. Amplicons were quantified by Quant-It (Invitrogen), pooled in equimolar concentration, and size-selected (375–425 bp) on the Pippen Prep (Sage Sciences) to reduce non-specific amplification products. Sequencing was performed on a MiSeq sequencer (Illumina; 2 × 250-bp paired-end reads with V2 chemistry).

Microbial diversity was analysed with Quantitative Insights Into Microbial Ecology (QIIME versions 1.8 and 1.9)24. The closed-reference operational taxonomic unit workflow in QIIME and the Greengenes reference database (May 2013) were used to cluster reads into operational taxonomic units with 97% identity and to assign taxonomy to representative operational taxonomic units24,25. Operational taxonomic unit tables were rarified to a depth of 40,000 (Figs 2 and 4) and 13,000 sequences per sample (Fig. 3). Principal-coordinates beta-diversity visualizations were created with Emperor as packaged in QIIME26. The linear discriminant analysis effect size algorithm in Galaxy (http://huttenhower.sph.harvard.edu/galaxy/) was used for additional statistical analyses27.

Bacterial culture. Serial dilutions of faeces from MMb and HMb mice were spread on brain–heart infusion agar supplemented with colistin (100 μg ml−1), gentamicin (6 μg ml−1), and aztreonam (5 μg ml−1), and cultures were incubated for 7 days in an anaerobic chamber (Goy Industries). MMb cx and HMb cx samples were collected from plates that contained approximately 500–1,000 colonies, resuspended in pre-reduced brain–heart infusion agar, and frozen at −80 °C until needed. To recover specific isolates, we cultured HMb faeces on the medium described above for 5–14 days and picked approximately 60 individual colonies. These colonies were inoculated into pre-reduced chopped meat medium with glucose (Anaerobe Systems), and genomic DNA was isolated with a DNeasy Blood and Tissue Kit (Qiagen). We compared the 16S rDNA gene sequence from each of these isolates with GenBank and the Greengenes reference database to identify bacterial taxonomy.

Probiotic administration to mice. MMb, HMb, and germ-free mice received orally administered Paraprevotella clara, Bacteroides uniformis, SFB, L. reuteri (BEI HM-102), R. gnavus (ATCC 29149), MMb cx, HMb cx, C. innocuum, or C.-immunis (100–150 μl; approximately 1010–1010 colony-forming units). P. clara, B. uniformis, C. innocuum, and C. immunis were isolated from the faeces of HMb mice; SFB was previously obtained from Y. Umesaki (Yakult) and propagated in SFB-monocolonized mice at Harvard Medical School. Seven days later, the mice were challenged with DSS. In some experiments, faecal samples were collected before and after probiotic administration for microbiota analysis. For Reg3γ experiments, HMb mice received orally administered (150–200 μl; approximately 1010–1010 colony-forming units) Parabacteroides distasonis (ATCC 8503; control bacteria), R. gnavus (ATCC 29149), L. reuteri (BEI HM-102), Allobaculum stercorarium (DSM 13633), Maritubuculum intestinale (DSM 28989), or L. vaginalis (DSM 5837). Mice were euthanized 7 days later.

Characterization of C. immunis. Our Lachnospiraceae isolate stains Gram-negative, is resistant to colistin differential disks (Anaerobe Systems), and is sensitive to vancomycin and kanamycin differential disks (Anaerobe Systems). Taken together, these findings are consistent with the isolate being a Gram-positive organism. The closest match for the 16S rDNA gene sequence of this isolate is C. symbiosum, with which it shares 98% identity. MALDI–TOF analysis by VITEK MS (BioMerieux) revealed that the closest match for this isolate is C. cladostridiforme, and MALDI–TOF analysis with a MALDI Biotyper (Bruker) was unable to provide a species- or genus-level identification. Biochemical testing demonstrated that our isolate—unlike C. symbiosum—w is resistant to 20% bile and does not produce acid from mannanose. Moreover, our isolate—unlike C. cladostridiforme—produces abundant amounts of butyrate from peptone–yeast extract–filter solution–glucose broth. Given these genetic, proteomic, and biochemical differences from its closest relatives, we propose that this isolate represents the type strain of C. immunis, a novel species within the family Lachnospiraceae.

We purified genomic DNA, using a Genomic Tip G/100 (Qiagen), and sequenced it on a PacBio RS II at the Yale Center for Genome Analysis, following the manufacturer’s instructions for library preparation. With use of the sequence data from one single-molecule real-time (SMRT) cell, the hierarchical genome-assembly process (HGAP) was able to assemble the genome into seven contigs, with an N50 contig length of 3.4 Mbp and a total genome size of 3.4 Mbp. We annotated the genome with the RAST server (http://rast.nmpdr.org), which indicated that there are 5,905 coding sequences and 75 RNAs. Tetracycline resistance genes were the only antibiotic resistance genes identified by ResFinder28. No virulence genes were identified by VirulenceFinder29, and PathogenFinder predicted that C. immunis is not a human pathogen (probability of being a human pathogen = 0.25)31.

Reg3γ expression analysis. qPCR for Reg3γ was performed as previously described. The distal 1.5 cm of small intestine was collected from the indicated mice, frozen immediately in liquid nitrogen, and stored at −80 °C until needed. Tissues were homogenized in Trizol (Invitrogen), and RNA was purified according to the manufacturer’s instructions, with a subsequent additional cleaning step (RNeasy Mini Kit; Qiagen). CDNA was prepared with the SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific), and qPCR was performed on a LightCycler (Roche) with iQ SYBR Green Supermix (Biorad).

Statistical analysis. Sample-size estimates for each experiment were based on previous laboratory experience. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. Prism 6 (GraphPad Software) was used for all statistical analyses.

Data availability. The 16S rDNA sequences from the microbiota analyses have been deposited at the European Nucleotide Archive under accession number PRJEB23029. The genome sequence for C. immunis has been deposited at DDBJ/European Nucleotide Archive/GenBank under accession number DDBJ/PRJEB23029. The version described in this paper is PDB:G0000000000, Data availability: Data for all figures have been provided with the paper. Any other data that support the findings of this study are available from the corresponding authors upon reasonable request.

22. Wirtz, S., Neufert, C., Weigmann, B. & Neurath, M. F. Chemically induced mouse models of intestinal inflammation. Nat. Protocols 2, 541–546 (2007).
23. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the illumina HiSeq and MiSeq platforms. ISME J. 6, 1621–1624 (2012).
24. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336 (2010).
25. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for microbial life. Nucleic Acids Res. 41, 65–70 (2013).
27. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

28. Kaneuchi, C., Watanabe, K., Terada, A., Benno, Y. & Mitsuoka, T. Taxonomic study of *Bacteroides clostridiiformis* subsp. *clostridiiformis* (Burri and Ankersmit) Holdeman and Moore and of related organisms: proposal of *Clostridium clostridiiformis* (Burri and Ankersmit) comb. nov. and *Clostridium symbiosum* (Stevens) comb. nov. *Int. J. Syst. Evol. Microbiol.* **26**, 195–204 (1976).

29. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* **67**, 2640–2644 (2012).

30. Joensen, K. G. et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J. Clin. Microbiol.* **52**, 1501–1510 (2014).

31. Cosentino, S., Voldby Larsen, M., Møller Aarestrup, F. & Lund, O. PathogenFinder—distinguishing friend from foe using bacterial whole genome sequence data. *PLoS ONE* **8**, e77302 (2013).
Extended Data Figure 1 | Individual MWAS reveal a large number of differentially abundant taxa. Linear discriminant analysis effect size was used to identify differentially abundant taxa in the faecal microbiota of various mice. Taxa coloured red and green were more abundant in that particular group of mice. Taxa coloured yellow did not statistically differ in abundance between groups. Each ring of the cladogram represents a different taxonomic level, starting with kingdom in the centre and ending with genus in the outer ring. a, Comparison of HMB and MMb. b, Comparison of MMb and SPF. c, Comparison of MMb and MMbHMb-1d. d, Comparison of HMB and HMBHMb-1d. The family Lachnospiraceae is indicated by the symbols c4 (a), a6 (b), a1 (c), and a9 (d).
Extended Data Figure 2 | Several taxa that are differentially present in HMb and MMb mice do not augment colitis severity. **a**, Survival of MMb mice (n = 2 mice) and MMb mice orally receiving *P. clara* (n = 4 mice) or *B. uniformis* (n = 4 mice) and subjected to DSS-induced colitis. **b**, Survival of HMb mice (n = 2 mice) and HMb mice orally receiving *L. reuteri* (n = 4), *R. gnatus* (n = 4 mice), or SFB (n = 4 mice) and subjected to DSS-induced colitis.
Extended Data Figure 3 | Culture of MMb faeces on semi-selective medium does not enrich for Lachnospiraceae. The relative abundance of bacterial families present in MMb faeces before (left) and after (right) culture is shown.
Extended Data Figure 4 | M Mb mice given M Mb cx and M Mb mice given H Mb cx have distinct microbiotas. Weighted principal components analysis of the faecal microbiota of M Mb mice before and after gavage with M Mb cx or H Mb cx is shown. The arrow indicates an M Mb mouse that received H Mb cx but died after being challenged with DSS.
Extended Data Figure 5 | The HMbc bacterial consortium is sufficient to protect mice from colitis-associated death. The survival of germ-free mice orally receiving HMbc (n = 10 mice) and subjected to DSS-induced colitis is shown.
Extended Data Figure 6 | Several taxa that are present in MMb mice and absent in HMb mice do not induce Reg3γ expression. qPCR analysis of ileal Reg3γ expression in HMb mice receiving no organisms (n = 4 mice) and in HMb mice receiving orally administered *A. stercoricanis* (n = 4 mice), *M. intestinale* (n = 4 mice), or *L. vaginalis* (n = 4 mice). Reg3γ expression was normalized to germ-free mice (n = 3 mice). Individual (dots) and mean (bars) values are shown.
Extended Data Table 1 | List of bacterial taxa associated with Reg3-γ induction

| P. Bacteroidetes.c. Bacteroidia.o. Bacteroidales.f. g.s. |
|--------------------------------------------------------|
| P. Bacteroidetes.c. Bacteroidia.o. Bacteroidales.f. Rikenellaceae.g.s. |
| P. Firmicutes.c. Bacilli.o. Lactobacillales.f. g.s. |
| P. Firmicutes.c. Bacilli.o. Lactobacillales.f. Lactobacillaceae.g. Lactobacillus.s. reuteri |
| P. Firmicutes.c. Bacilli.o. Lactobacillales.f. Streptococaceae.g. Streptococcus.s. |
| P. Firmicutes.c. Clostridia.o. Clostridiales.f. Clostridaceae.g. CandidatusArthromitus.s. |
| P. Firmicutes.c. Clostridia.o. Clostridiales.f. Lachnospiraceae.g. Ruminococcus.s. gnavus |

p, phylum; c, class; o, order; f, family; g, genus; s, species. Taxonomic levels that lack information (e.g., f.g.s) did not match named taxa present in the Greengenes database.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

### Experimental design

1. **Sample size**
   
   Describe how sample size was determined.

   Sample sizes were based on a combination of the resource equation method, limited availability of gnotobiotic mice, and prior lab experience to ensure statistical and biological significance.

2. **Data exclusions**
   
   Describe any data exclusions.

   No data were excluded.

3. **Replication**
   
   Describe whether the experimental findings were reliably reproduced.

   All attempts at replication were successful. The number of replicates for each experiment is listed in the figure legends.

4. **Randomization**
   
   Describe how samples/organisms/participants were allocated into experimental groups.

   Mice were randomized between groups when feasible. When different gnotobiotic mice were being compared, mice were age- and sex-matched.

5. **Blinding**
   
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Investigators were not blinded to the experimental groups given different housing conditions required. The pathologist was blinded to groups when reviewing histology.

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. **Statistical parameters**
   
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

   | n/a | Confirmed |
   |-----|-----------|
   | ☒   | The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
   | ☒   | A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly. |
   | ☒   | A statement indicating how many times each experiment was replicated |
   | ☒   | The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
   | ☒   | A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
   | ☒   | The test results (e.g. \(p\) values) given as exact values whenever possible and with confidence intervals noted |
   | ☒   | A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
   | ☒   | Clearly defined error bars |

*See the web collection on statistics for biologists for further resources and guidance.*
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Qime 1.8 and 1.9 were used for 16S rDNA analysis, and LEfSe was used to identify bacterial taxa that were statistically different between groups. These packages are freely available online.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Unique materials are available from the authors. Other materials are available from commercial sources as indicated.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

N/A

b. Describe the method of cell line authentication used.

no eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

no eukaryotic cell lines were used

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

no commonly misidentified cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Gnotobiotic and SPF Swiss-Webster mice were used throughout the study. Mice used in experiments were sex- and age-matched (typically 5–10 weeks old) and drawn randomly from the same litter, when feasible. Both sexes were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

No human subjects were used in this manuscript.
ERRATUM

doi:10.1038/nature25471

Erratum: Moving beyond microbiome-wide associations to causal microbe identification

Neeraj K. Surana & Dennis L. Kasper

Nature 552, 244–247 (2017); doi:10.1038/nature25019

In this Letter, errors in Fig. 2c were inadvertently introduced during the production process. The key to the survival graph should state 'HMb' for the green line rather than 'MMb', 'HMb^{Mmb-1d}' for the blue line rather than 'MMb^{HMb-1d}' and 'HMb^{Mmb-3d}' for the red line rather than 'MMb^{Mmb-3d}'. The original figure has been corrected online.