ARTICLE

doi:10.1038/nature23874

# Commensal bacteria make GPCR ligands that mimic human signalling molecules

Louis J. Cohen[1,2], Daria Esterhazy[3], Seong-Hwan Kim[1], Christophe Lemetre[1], Rhiannon R. Aguilar[1], Emma A. Gordon[1], Amanda J. Pickard[4], Justin R. Cross[4], Ana B. Emiliano[5], Sun M. Han[1], John Chu[1], Xavier Vila-Farres[1], Jeremy Kaplitt[1], Aneta Rogoz[3], Paula Y. Calle[1], Craig Hunter[6], J. Kipchirchir Bitok[1] & Sean F. Brady[1]

**Commensal bacteria are believed to have important roles in human health. The mechanisms by which they affect mammalian physiology remain poorly understood, but bacterial metabolites are likely to be key components of host interactions. Here we use bioinformatics and synthetic biology to mine the human microbiota for *N*-acyl amides that interact with G-protein-coupled receptors (GPCRs). We found that *N*-acyl amide synthase genes are enriched in gastrointestinal bacteria and the lipids that they encode interact with GPCRs that regulate gastrointestinal tract physiology. Mouse and cell-based models demonstrate that commensal GPR119 agonists regulate metabolic hormones and glucose homeostasis as efficiently as human ligands, although future studies are needed to define their potential physiological role in humans. Our results suggest that chemical mimicry of eukaryotic signalling molecules may be common among commensal bacteria and that manipulation of microbiota genes encoding metabolites that elicit host cellular responses represents a possible small-molecule therapeutic modality (microbiome-biosynthetic gene therapy).**

Although the human microbiome is believed to have an important role in human physiology, the mechanisms by which bacteria affect mammalian physiology remain poorly defined[1]. Bacteria rely heavily on small molecules to interact with their environment[2]. Although it is likely that the human microbiota similarly relies on small molecules to interact with its human host, the identity and functions of microbiota-encoded effector molecules remain mostly unknown. The study of small molecules produced by the human microbiota and the identification of the host receptors that they interact with should help to define the relationship between bacteria and human physiology and provide a resource for the discovery of small-molecule therapeutics.

We recently reported on the discovery of commendamide, a human microbiota-encoded, GPCR-active, long-chain *N*-acyl amide that suggests a structural convergence between human signalling molecules and microbiota-encoded metabolites[3]. *N*-acyl amides, such as endocannabinoids, are able to regulate diverse cellular functions owing, in part, to their ability to interact with GPCRs. GPCRs are the largest family of membrane receptors in eukaryotes and are likely to be key mediators of host–microbial interactions in the human microbiome. The importance of GPCRs to human physiology is reflected by the fact that they are the most common targets of therapeutically approved small-molecule drugs. The GPCRs with which human *N*-acyl amides interact are implicated in diseases, including diabetes, obesity, cancer and inflammatory bowel disease, among others[4,5]. With numerous possible combinations of amine head groups and acyl tails, long-chain *N*-acyl amides represent a potentially large and functionally diverse class of microbiota-encoded GPCR-active signalling molecules.

Here, we combined bioinformatic analysis of human microbiome sequencing data with targeted gene synthesis, heterologous expression and high-throughput GPCR activity screening to identify GPCR-active *N*-acyl amides encoded by human microbiota. The bacterial effectors that we have identified provide mechanistic insights into potential functions of the human microbiome and suggest that these GPCR-active small molecules and their associated microbial biosynthetic genes have the potential to regulate human physiology.

## Isolation of commensal *N*-acyl amides

To identify *N*-acyl synthase (NAS) genes within human microbial genomes, the Human Microbiome Project (HMP) sequence data was searched with BLASTN using 689 NAS genes associated with the *N*-acyl synthase protein family PFAM13444 (ref. 3). The 143 unique human microbial (hm) *N*-acyl synthase genes (hm-*NAS* genes), which we identified, fall into four major clades (clades A–D, Fig. 1a) that are divided into a number of distinct sub-clades (Fig. 1a). Forty-four phylogenetically diverse hm-*NAS* genes were selected for synthesis and heterologous expression. This set included all hm-*NAS* genes from clades sparsely populated with hm-*NAS* sequences and representative examples from clades heavily populated with hm-*NAS* sequences (Fig. 1a).

Liquid chromatography–mass spectrometry (LC–MS) analysis of ethyl acetate extracts derived from *Escherichia coli* cultures transformed with each construct revealed clone-specific peaks in 31 cultures. hm-*NAS* gene functions could be clustered into six groups based on the retention time and mass of the heterologously produced metabolites (Extended Data Fig. 1 and Supplementary Table 1). Molecule isolation and structural elucidation studies were carried out on one representative culture from each group (Supplementary Information). This analysis identified six *N*-acyl amide families that differ by amine head group and fatty acid tail (Fig. 1b, families 1–6): (1) *N*-acyl glycine; (2) *N*-acyloxyacyl lysine; (3) *N*-acyloxyacyl glutamine; (4) *N*-acyl lysine/ornithine; (5) *N*-acyl alanine; (6) *N*-acyl serinol. Each family was isolated as a collection of metabolites with different acyl substituents. The most common analogue within each family is shown in Fig. 1b. Long-chain *N*-acyl ornithines, lysines and glutamines have been reported as natural products produced by soil bacteria and some human pathogens[6–8].

[1]Laboratory of Genetically Encoded Small Molecules, Rockefeller University, New York, New York 10065, USA. [2]Division of Gastroenterology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [3]Laboratory of Mucosal Immunology, Rockefeller University, New York, New York 10065, USA. [4]Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [5]Laboratory of Molecular Genetics, Rockefeller University, New York, New York 10065, USA. [6]Comparative Biosciences Center, Rockefeller University, New York, New York 10065, USA.

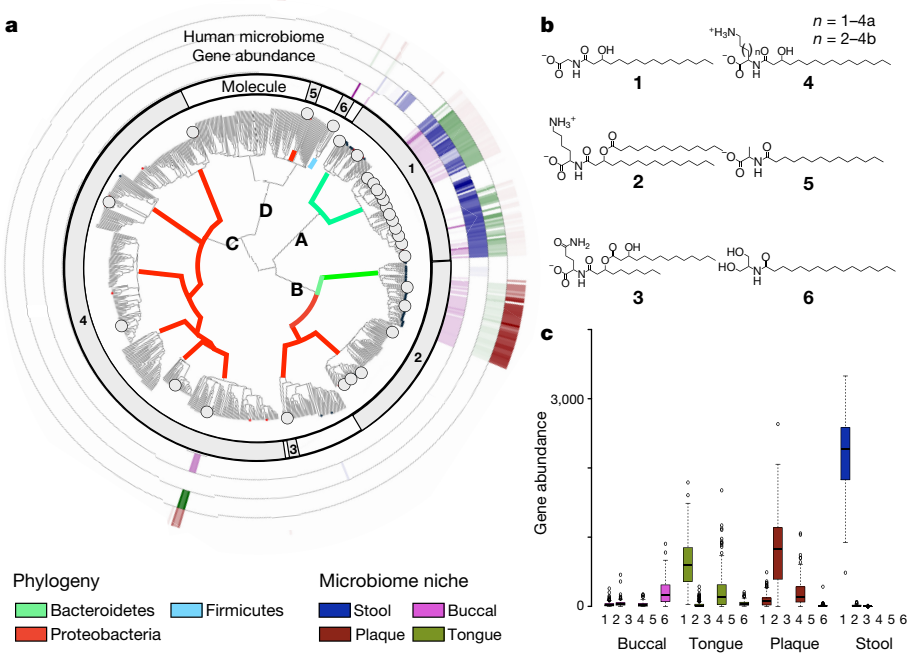



**Figure 1 | hm-*NAS* genes in gastrointestinal microbiota. a**, Phylogenetic tree of *N*-acyl genes from PFAM13444. hm-*NAS* genes have a dot at the branch tip. Black dots were not synthesized, red dots were synthesized, but no molecule was detected and large grey dots mark genes that produced *N*-acyl amides. Branches are coloured by bacterial phylogeny. **b**, The major heterologously produced metabolite from each *N*-acyl family (1–6) is shown. **c**, hm-*NAS* gene distribution and abundance (reads per kilobase of gene per million reads (RPKM)) based on molecule family (1–6). Body site and molecule designations in **a** are based on the analysis shown in **c** and **b**. Box plots are median and 1st to 3rd quartile. HMP patient samples analysed include $n = 133$ tongue, 127 plaque, 148 stool, 122 buccal samples.

Functional differences in NAS enzymes follow the pattern of the NAS phylogenetic tree, with hm-*NAS* genes from the same clade or sub-clade mostly encoding the same metabolite family (Fig. 1a). With the exception of one NAS that is predicted to use lysine and ornithine as substrates, hm-NAS proteins appear to be selective for a single amine-containing substrate. The most common acyl chains incorporated by hm-NAS proteins are 14–18 carbons in length. These can be modified by β-hydroxylation or a single unsaturation. Three hm-NAS enzymes contain two domains. The second domain is either an aminotransferase that is predicted to catalyse the formation of serinol from glycerol (Fig. 1b, family 6 and Extended Data Fig. 2) or an additional acyltransferase that is predicted to catalyse the transfer of a second acyl group (Fig. 1b, families 2 and 3). To explore NAS gene synteny, we looked for gene occurrence patterns around NAS genes in the human microbiome. The only repeating pattern that we saw was that some NAS genes appear adjacent to genes predicted to encode acyltransferases. This is reminiscent of the two domain NAS proteins that we found produce di-acyl metabolites (families 2 and 3). There were rare instances where NAS proteins potentially occur in gene clusters, but none of these were used in this study.

To look for native *N*-acyl amide production by commensal bacteria, organic extracts from cultures of species containing the hm-*NAS* genes that we examined were screened by LC–MS. On the basis of retention time and mass, we detected the production of the expected *N*-acyl amides by commensal species predicted to produce *N*-acyl glycines, *N*-acyloxyacyl lysines, *N*-acyl lysine/ornithines and *N*-acyl serinols. The only case where we did not detect the expected *N*-acyl amide was for *N*-acyloxyacyl glutamines (Extended Data Fig. 1).

## hm-*NAS* genes are enriched in gastrointestinal bacteria

A BLASTN search of NAS genes against human microbial reference genomes and metagenomic sequencing data from the HMP showed that NAS genes were enriched in gastrointestinal bacteria relative to bacteria found at other body sites (Fisher's exact test $P < 0.05$, gastrointestinal versus non-gastrointestinal sites; Fig. 1 and Supplementary Table 2). Within gastrointestinal sites that were frequently sampled in the context of the HMP (for example, stool, buccal mucosa, supragingival plaque and tongue) hm-*NAS* gene families show distinct distribution patterns (Fig. 1c, two-way ANOVA $P < 2 \times 10^{-16}$). Despite very large person-to-person variation in microbiota species composition, most of the *N*-acyl amide synthase gene families that we studied can be found in over 90% of patient samples. *N*-acyoxyacyl glutamine (12%) and *N*-acyl alanine (not detected) synthase genes are the only exceptions. Taken together, these data suggest that *NAS* genes are highly prevalent in the human microbiome and unique sites within the gastrointestinal tract are probably exposed to different sets of *N*-acyl amide structures.

When we searched existing metatranscriptome sequence data from stool and supragingival plaque microbiomes to look for evidence of

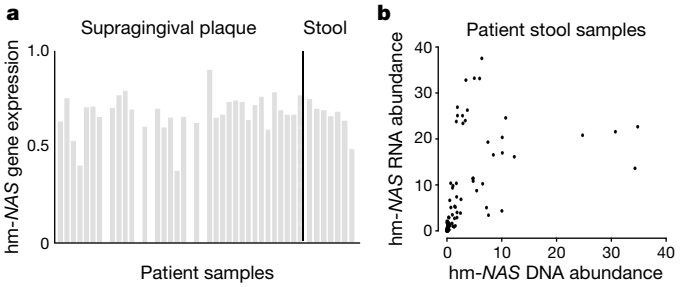



**Figure 2 | *N*-acyl synthase gene expression *in vivo*. a**, hm-*NAS* expression within bacterial genomes. Gene expression analysis for an *N*-acyl glycine hm-*NAS* gene in a stool metatranscriptome dataset and an *N*-acyloxyacyl lysine hm-*NAS* gene in a supragingival plaque metatranscriptome dataset. Gene expression is normalized to the expression of all genes from a bacterial genome containing the hm-*NAS* gene that was heterologously expressed—*Bacteroides dorei* in stool, *Capnocytophaga ochracea* in plaque. 1, highly expressed; 0, not expressed). **b**, hm-*NAS* expression relative to DNA. Comparison of hm-*NAS* gene abundance based on RNA- or DNA-derived reads obtained from individual patient stool samples. Abundance is measured in RPKM. $n = 24$ patient stool samples analysed—eight patients with three samples per patient. $n = 38$ patient plaque samples analysed.

hm-*NAS* gene expression in the gastrointestinal tract, we observed site-specific hm-*NAS* gene expression that matches the predicted body site localization patterns for hm-*NAS* genes in metagenomic data. Across patient samples, hm-*NAS* genes are transcribed to varying degrees relative to the average level of transcription for each gene in the bacterial genome (Fig. 2a). In the stool metatranscriptome dataset, both RNA and DNA sequencing datasets were available allowing a more direct sample-to-sample comparison of hm-*NAS* gene expression levels. When metatranscriptome data were normalized using the number of hm-*NAS* gene-specific DNA sequence reads detected in each sample, we observed what appears to be differential expression of hm-*NAS* genes in different patient samples (Fig. 2b). Datasets for which bacterial genes, transcripts and metabolites can be tracked in a single sample will be necessary to explore how hm-*NAS* gene transcription variation affects metabolite production.

## *N*-acyl-amides interact with gastrointestinal GPCRs

The major *N*-acyl amide isolated from each family was assayed for agonist and antagonist activity against 240 human GPCRs (Fig. 3 and Extended Data Fig. 3). The strongest agonist interactions were: activation of GPR119 by *N*-palmitoyl serinol (half maximal effective concentration ($EC_{50}$) = 9 μM), activation of sphingosine-1-phosphate receptor 4 (S1PR4) by *N*-3-hydroxypalmitoyl ornithine ($EC_{50}$ = 32 μM) and activation of G2A by *N*-myristoyl alanine ($EC_{50}$ = 3 μM). Interactions between bacterial *N*-acyl amides and GPCRs were also specific (Fig. 3a, b). In each experiment, no other GPCRs reproducibly showed greater than 35% activation relative to the endogenous ligands. The strongest antagonist activities were observed for *N*-acyloxyacyl glutamine against two prostaglandin receptors, PTGIR and PTGER4 (Fig. 3c; PTGIR, half maximal inhibitory concentration ($IC_{50}$) = 15 μM; PTGER4, $IC_{50}$ = 43 μM). PTGIR was specifically antagonized by *N*-acyloxyacyl glutamine, whereas PTGER4 was antagonized by *N*-acyloxyacyl glutamine as well as other *N*-acyl amides (Fig. 3c). Alternative GPCR screening methods could identify interactions in addition to those uncovered here.

On the basis of data from the Human Protein Atlas, GPCRs targeted by human microbial *N*-acyl amides are localized to the gastrointestinal tract and its associated immune cells. In mouse models, this collection of GPCRs have been reported to affect diverse mucosal functions, including metabolism (GPR119), immune cell differentiation (S1PR4, PTGIR and PTGER4), immune cell trafficking (S1PR4, G2A) and tissue repair (PTGIR)[9–14]. It is not possible at this time to look for co-localization of GPCR and hm-*NAS* gene expression in specific gastrointestinal niches, as neither the HMP nor the Human Protein Atlas are sufficiently comprehensive in their survey of human body sites. Nonetheless, 16S and metagenomic deep-sequencing studies link bacteria containing hm-*NAS* genes or hm-*NAS* genes themselves to specific locations in the gastrointestinal tract where the GPCRs of interest are expressed (Extended Data Fig. 4).

## Similarity of bacterial and human GPCR ligands

Human microbiota-encoded *N*-acyl amides bear structural similarity to endogenous GPCR-active ligands (Fig. 4). The clearest overlap in structure and function between bacterial and human GPCR-active ligands is for the endocannabinoid receptor GPR119 (Figs 4, 5). Endogenous GPR119 ligands include oleoylethanolamide (OEA) and the dietary lipid-derivative 2-oleoyl glycerol (2-OG)[15,16]. In our heterologous expression experiment, we isolated both the palmitoyl and oleoyl analogues of *N*-acyl serinol. The latter only differs from 2-OG by the presence of an amide instead of an ester and from OEA by the

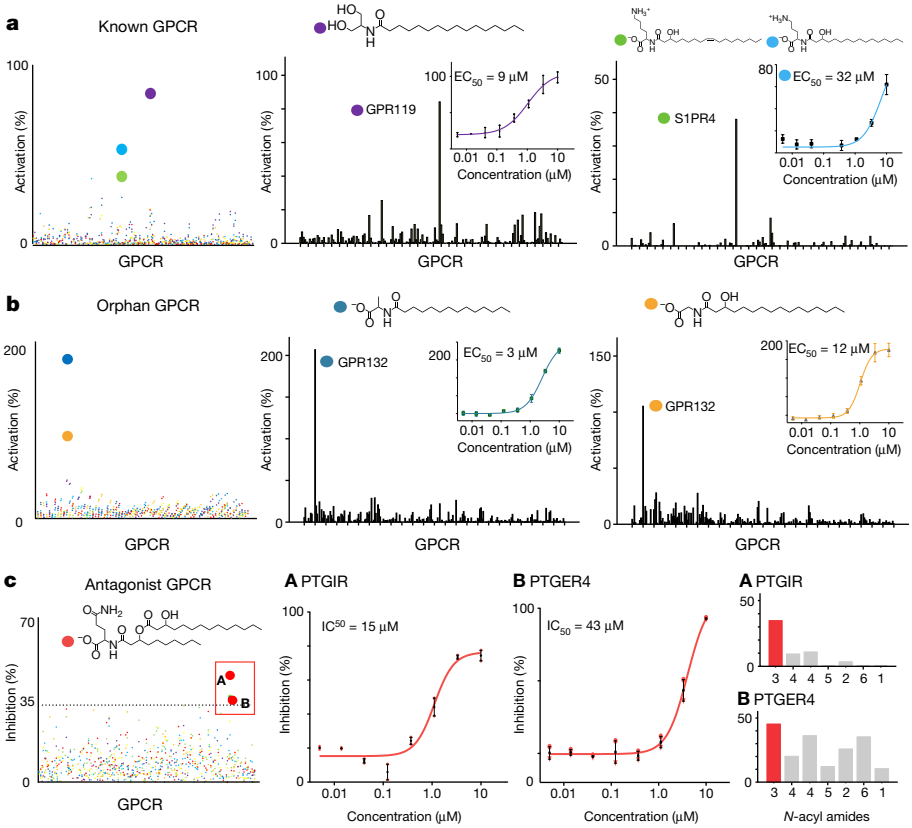



**Figure 3 | GPCR activity screen of human *N*-acyl amides. a**, **b**, Screen of *N*-acyl amides (colour coded) for agonist activity against 168 GPCRs with known ligands (**a**) and 72 orphan GPCRs (**b**). Dot plots display data for all *N*-acyl amides assayed against all GPCRs. Screen was performed once. Bar graphs show the strongest *N*-acyl GPCR agonist interactions compared to all GPCRs. Insets show dose response curves and $EC_{50}$ data (each dose performed in duplicate). **c**, Screen of *N*-acyl amides as antagonists in the presence of endogenous ligands. Screen was performed once. A, PTGIR is specifically inhibited by *N*-acyloxyacyl glutamine. B, PTGER4 is inhibited by structurally diverse human *N*-acyl amides. Error bars are mean ± s.e.m.

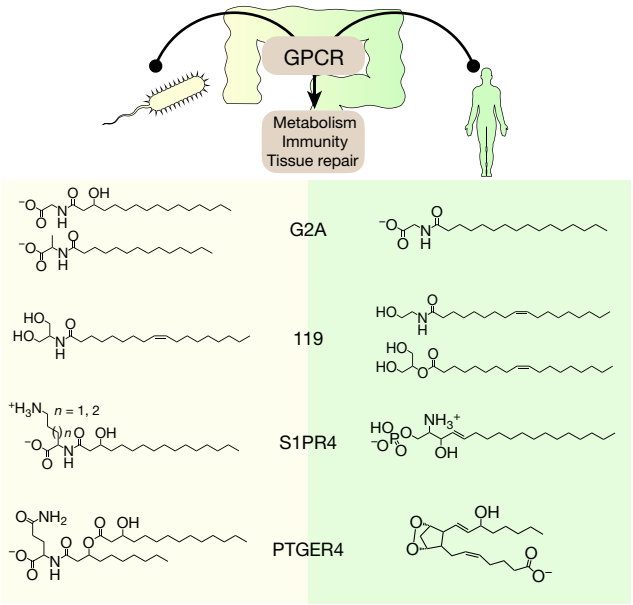



**Figure 4 | Structural mimicry of GPCR ligands.** Comparison of microbiota-encoded and human GPCR ligands suggests structural and functional complementarity.

presence of an additional methanol substituent. *N*-oleoyl serinol is a similarly potent GPR119 agonist as the endogenous ligand OEA ($EC_{50} = 12\,\mu M$ versus $7\,\mu M$), but elicits an almost twofold greater maximum GPR119 activation (Fig. 5a). *N*-palmitoyl derivatives of all 20 natural amino acids were synthesized and none activated GPR119 by more than 37% relative to OEA (Fig. 5b). The generation of a potent and specific long-chain *N*-acyl-based GPR119 ligand therefore necessitates a more complex biosynthesis than the simple *N*-acylation of an amino acid, as is commonly seen for characterized NAS enzymes. In this case, the biosynthesis of *N*-acyl serinols is achieved through the coupling of an NAS domain with an aminotransferase that is predicted to generate serinol from glycerol (Extended Data Fig. 2).

The endogenous agonist for S1PR4, sphingosine-1-phosphate (S1P) and the *N*-3-hydroxypalmitoyl ornithine/lysine family of bacterial agonists share similar head group charges. S1P is a markedly more potent agonist ($EC_{50} = 0.09\,\mu M$ versus $EC_{50} = 32\,\mu M$); however, the bacterial agonists are more specific for S1PR4. The bacterial *N*-3-hydroxypalmitoyl ornithine did not activate S1PR1, 2 or 3 in our GPCR screen, whereas S1P activates all four members of the S1P receptor family that were tested.

No direct comparison could be made between the microbiota-derived and endogenous ligands for PTGIR or PTGER4, because there are no known endogenous antagonists for these receptors. Many human GPCRs remain orphan receptors lacking known endogenous ligands. Ligands for at least some of these receptors will undoubtedly be found among the small molecules produced by the human microbiota. G2A is an orphan receptor and therefore does not have a well-defined endogenous agonist, although it has been reported to respond to lysophosphatidylcholine[17,18]. We found that the bacterial metabolites *N*-3-hydroxypalmitoyl glycine (commendamide) and *N*-palmitoyl alanine both activate G2A. Mammals produce *N*-palmitoyl glycine, which differs from commendamide by the absence of the β-hydroxyl and which activates G2A, based on our synthetic *N*-acyl studies[19].

GPR119 is the most extensively studied of the GPCRs activated by bacterial ligands we identified (Fig. 4). Mechanisms that link endogenous GPR119 agonists (OEA, 2-OG) to changes in host phenotype are well-defined as a result of the exploration of GPR119 as a therapeutic target for diabetes and obesity[20–23]. GPR119 agonists are thought not only to primarily affect glucose homeostasis, but also gastric emptying and appetite through both GPR119-dependent hormone release from enteroendocrine (GLP-1, GIP, PYY) and pancreatic β-cells (insulin) as well as GPR119-independent mechanisms, including PPARα modulation[9,16,24–30]. Mouse enteroendocrine GLUTag cells have been used as a model system for measuring the ability of potential GPR119 agonists to induce GLP-1 release. When administered to GLUTag cells at equimolar concentrations, microbiota-encoded *N*-oleoyl serinol or the endogenous ligands OEA and 2-OG induced GLP-1 secretion to the same magnitude (Fig. 5c). To provide an orthogonal measurement of GPR119 activation by *N*-acyl serinols, HEK293 cells were stably transfected with a GPR119 expression construct. Both OEA and *N*-oleoyl serinol increased cellular cAMP concentrations in a GPR119-dependent manner (Extended Data Fig. 5).

## hm-*NAS* expression alters blood glucose in mice

The functional overlap between endogenous and bacterial metabolites suggested that bacteria expressing microbiota-encoded GPR119 ligands might elicit host phenotypes that mimic those induced by eukaryotic ligands. Endogenous and synthetic GPR119 ligands have been associated with changes in glucose homeostasis that are relevant to the aetiology and treatment of diabetes and obesity, including a study where mice were orally administered bacteria that had been engineered to produce a eukaryotic enzyme that increases endogenous GPR119 ligand (OEA) precursors[9,16,24–27,31]. The metabolic effect of the endogenous GPR119 ligands is believed to occur at the intestinal mucosa, because the delivery of OEA intravenously fails to lower blood glucose in mice during an oral glucose-tolerance test (OGTT)[26]. Consequently, we sought to determine whether mice colonized with bacteria that had been engineered to produce *N*-acyl serinols would exhibit predictable host phenotypes. Gnotobiotic mice were colonized with *E. coli* engineered to express the *N*-acyl serinol synthase gene in an isopropyl-β-D-thiogalactoside (IPTG)-dependent manner. Control mice were colonized with *E. coli* expressing an empty vector. On the basis of the number of colony-forming units detected in faecal pellets, both cohorts of mice were colonized to the same extent (Extended Data Fig. 7). After one week of exposure to IPTG, both cohorts were fasted overnight and subjected to an OGTT. At 30 min after the challenge, we observed a statistically significant decrease in blood glucose levels for the group colonized with *E. coli* expressing the *N*-acyl serinol synthase gene (Fig. 5d). Mass spectrometry analysis of metabolites present in caecal samples revealed the presence of *N*-acyl serinols in the treatment cohort, but not in the control cohort (Extended Data Fig. 6). After two weeks of withdrawing IPTG from the drinking water, we no longer observed a difference in blood glucose between the two cohorts after an OGTT[32] (Fig. 5e).

To further explore the metabolic phenotype induced by *N*-acyl serinols we repeated the OGTT experiment in an antibiotic-treated mouse model. In this study, we compared mice colonized with *E. coli* expressing an active *N*-acyl serinol synthase to mice colonized with *E. coli* expressing an NAS point mutant (p.Glu91Ala; Extended Data Fig. 8) that no longer produced *N*-acyl serinols. In this model, the glucose lowering effect of colonization with *N*-acyl serinol-producing *E. coli* remained significant (Fig. 5f). In the antibiotic-treated mice, we measured GLP-1 and insulin concentrations after glucose gavage. Both hormones were significantly increased in the treatment group compared to the control group (Fig. 5g, h). In all mouse models, the observed correlation between hm-*NAS* gene induction and increased glucose tolerance is similar in magnitude to several studies with small-molecule GPR119 agonists including glyburide, an FDA-approved therapeutic for diabetes[24,25].

## Discussion

Our characterization of human microbial *N*-acyl amides, together with other investigations of the human microbiota, suggests that host–microbial interactions may rely heavily on simple metabolites built from the same common lipids, sugars and peptides that define many human signalling systems (for example, neurotransmitters,

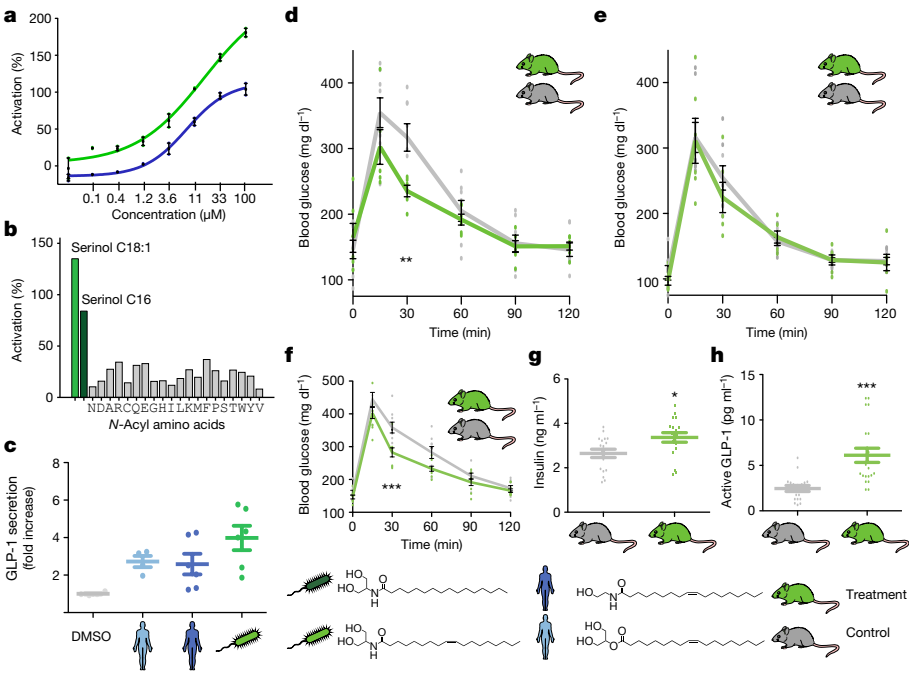



**Figure 5 | *N*-acyl serinols affect GLP-1 secretion *in vitro* and glucose homeostasis *in vivo*.** **a**, β-Arrestin GPR119 activation assay using microbiota (green) and human (blue) ligands (each dose was performed in duplicate). **b**, β-Arrestin assay comparing microbiota ligands and 20 synthesized *N*-palmitoyl amino acids (screen was performed once). **c**, Release of GLP-1 by GLUTag cells (ANOVA, $P < 0.05$, data combined from two independent experiments, $n = 4$ for DMSO and 2-OG, $n = 6$ for OEA and *N*-oleoyl serinol). **d**, OGTT in gnotobiotic mice. Treatment mice ($n = 6$ mice, data combined from two independent experiments) were colonized with *E. coli* producing *N*-acyl serinols and control mice ($n = 8$ mice, data combined from two independent experiments) were colonized with *E. coli* containing an empty vector (two-way ANOVA, Bonferroni post hoc test). **e**, OGTT after withholding IPTG to stop *N*-acyl gene expression (no difference, two-way ANOVA, *n* is the same as in **d**). **f**, OGTT in an antibiotic-treated mouse cohort ($n = 9$ mice in both groups, data combined from two independent experiments, two-way ANOVA, Bonferroni post hoc test). **g**, **h**, Insulin (**g**; $n = 6$ mice in both groups, one experiment, technical triplicates) and GLP-1 (**h**; $n = 9$ control mice, $n = 10$ treatment mice, data combined from two independent experiments, technical replicates) measured at 15 min after glucose gavage in the antibiotic-treated cohort (unpaired, two-tailed *t*-test). Data are mean ± s.e.m. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

bioactive lipids and glycans). This is not surprising as the genomes of the bacterial taxa common to the human gastrointestinal tract (for example, Bacteroidetes, Firmicutes and Proteobacteria) are often lacking in gene clusters that encode the production of complex secondary metabolites (for example, polyketides, non-ribosomal peptides and terpenes). It appears that biosynthesis of endogenous mammalian signalling molecules as well as those produced by the human microbiota may rely on the modest manipulation of primary metabolites. As a result, the structural conservation between metabolites used in host–microbial interactions and endogenous mammalian signalling metabolites may be a common phenomenon in the human microbiome. Evolutionarily, the convergence of bacterial and human signalling systems through structurally related GPCR ligands is not unreasonable as GPCRs are thought to have developed in eukaryotes to allow structurally simple signalling molecules to regulate increasingly complex cellular interactions[33–35]. The structural similarities between microbiota-encoded *N*-acyl amides and endogenous GPCR-active lipids may be indicative of a broader structural and functional overlap among bacterial and human bioactive lipids, including other GPCR-active *N*-acyl amides, eiconasoids (prostaglandins and leukotrienes) and sphingolipids. Sphingolipid-based signalling molecules may also be common in the human microbiome, because prevalent bacterial species are known to synthesize membrane sphingolipids[36].

The GPCRs with which bacterial *N*-acyl amides were found to interact are all part of the same 'lipid-like' GPCR gene family. The potential importance of this GPCR family for the regulation of host–microbial interactions is suggested by their localization in areas of gastrointestinal track, which are enriched in bacteria that are predicted to synthesize GPCR ligands (Extended Data Fig. 4). Lipid-like GPCRs have been shown to have roles in disease models that are correlated with changes in microbial ecology, including colitis (S1PR4, PTGIR and PTGER4), obesity (GPR119), diabetes (GPR119), autoimmunity (G2A) and atherosclerosis (G2A, PTGIR)[9,10,13,14]. The fact that the expression of an NAS gene in a gastrointestinal-colonizing bacterium is sufficient to alter host physiology suggests that the interaction between lipid-like GPCRs and their *N*-acyl amide ligands could be relevant to human physiology and warrants further study. Using LC–MS analysis, we observed most of the microbiota-encoded *N*-acyl amides reported here in human stool samples (Extended Data Fig. 9). Further studies will be needed to better define the distribution and concentration of these metabolites throughout the gastrointestinal tract, especially at the mucosa, where the physiological activity of these metabolites probably occurs. Notably, *Gemella* spp. predicted to encode *N*-acyl serinols are tightly associated with the small intestinal mucosa, supporting this site as a potentially important location for *N*-acyl-amide-mediated interactions[37]. Because the mouse model system used here relies on induced expression of NAS genes, it will also be important to understand how these genes are natively regulated.

Current strategies for treating diseases associated with the microbiome, such as inflammatory bowel disease or diabetes, are not believed to address the dysfunction of the host–microbial interactions that are likely to be part of the disease pathogenesis. Bacteria engineered to deliver bioactive small molecules produced by the human microbiota have the potential to help address diseases of the microbiome by modulating the native distribution and abundance of these metabolites. Regulation of GPCRs by microbiota-derived *N*-acyl amides is a particularly noteworthy therapeutic strategy for the treatment of human diseases, because GPCRs have been extensively validated as therapeutic targets. As our mechanistic understanding of how human microbiota-encoded small molecules effect changes in host physiology grows, the potential for using 'microbiome-biosynthetic gene therapy' to treat human disease by complementing small molecule deficiencies

in native host–microbial interactions with microbiota-derived biosynthetic genes should increase accordingly. The use of functional metagenomics to identify microbiota-encoded effectors combined with bioinformatics and synthetic biology to expand effector molecule families provides a generalizable platform to help to define the role of microbiota-encoded small molecules in host–microbial interactions.

**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 17 October 2016; accepted 1 August 2017.**
**Published online 30 August 2017.**

1. Koppel, N. & Balskus, E. P. Exploring and understanding the biochemical diversity of the human microbiota. *Cell Chem. Biol.* **23,** 18–30 (2016).
2. Meinwald, J. & Eisner, T. Chemical ecology in retrospect and prospect. *Proc. Natl Acad. Sci. USA* **105,** 4539–4540 (2008).
3. Cohen, L. J. *et al.* Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. *Proc. Natl Acad. Sci. USA* **112,** E4825–E4834 (2015).
4. Cani, P. D. *et al.* Endocannabinoids — at the crossroads between the gut microbiota and host metabolism. *Nat. Rev. Endocrinol.* **12,** 133–143 (2016).
5. Pacher, P. & Kunos, G. Modulating the endocannabinoid system in human health and disease—successes and failures. *FEBS J.* **280,** 1918–1943 (2013).
6. Moore, E. K. *et al.* Lysine and novel hydroxylysine lipids in soil bacteria: amino acid membrane lipid response to temperature and pH in *Pseudopedobacter saltans*. *Front. Microbiol.* **6,** 637 (2015).
7. Geiger, O., González-Silva, N., López-Lara, I. M. & Sohlenkamp, C. Amino acid-containing membrane lipids in bacteria. *Prog. Lipid Res.* **49,** 46–60 (2010).
8. Zhang, X., Ferguson-Miller, S. M. & Reid, G. E. Characterization of ornithine and glutamine lipids extracted from cell membranes of *Rhodobacter sphaeroides*. *J. Am. Soc. Mass Spectrom.* **20,** 198–212 (2009).
9. Flock, G., Holland, D., Seino, Y. & Drucker, D. J. GPR119 regulates murine glucose homeostasis through incretin receptor-dependent and independent mechanisms. *Endocrinology* **152,** 374–383 (2011).
10. Schulze, T. *et al.* Sphingosine-1-phospate receptor 4 ($S1P_4$) deficiency profoundly affects dendritic cell function and TH17-cell differentiation in a murine model. *FASEB J.* **25,** 4024–4036 (2011).
11. Le, L. Q. *et al.* Mice lacking the orphan G protein-coupled receptor G2A develop a late-onset autoimmune syndrome. *Immunity* **14,** 561–571 (2001).
12. Konya, V., Marsche, G., Schuligoi, R. & Heinemann, A. E-type prostanoid receptor 4 (EP4) in disease and therapy. *Pharmacol. Ther.* **138,** 485–502 (2013).
13. Kabashima, K. *et al.* The prostaglandin receptor EP4 suppresses colitis, mucosal damage and CD4 cell activation in the gut. *J. Clin. Invest.* **109,** 883–893 (2002).
14. Manieri, N. A. *et al.* Mucosally transplanted mesenchymal stem cells stimulate intestinal healing by promoting angiogenesis. *J. Clin. Invest.* **125,** 3606–3618 (2015).
15. Hansen, K. B. *et al.* 2-Oleoyl glycerol is a GPR119 agonist and signals GLP-1 release in humans. *J. Clin. Endocrinol. Metab.* **96,** E1409–E1417 (2011).
16. Overton, H. A. *et al.* Deorphanization of a G protein-coupled receptor for oleoylethanolamide and its use in the discovery of small-molecule hypophagic agents. *Cell Metab.* **3,** 167–175 (2006).
17. Khan, S. Y. *et al.* Lysophosphatidylcholines activate G2A inducing $G_{\alpha i-1}$-/$G_{\alpha q/11}$-$Ca^{2+}$ flux, $G_{\beta\gamma}$-Hck activation and clathrin/β-arrestin-1/GRK6 recruitment in PMNs. *Biochem. J.* **432,** 35–45 (2010).
18. Kabarowski, J. H. G2A and LPC: regulatory functions in immunity. *Prostaglandins Other Lipid Mediat.* **89,** 73–81 (2009).
19. Rimmerman, N. *et al. N*-palmitoyl glycine, a novel endogenous lipid that acts as a modulator of calcium influx and nitric oxide production in sensory neurons. *Mol. Pharmacol.* **74,** 213–224 (2008).
20. Ritter, K., Buning, C., Halland, N., Pöverlein, C. & Schwink, L. G protein-coupled receptor 119 (GPR119) agonists for the treatment of diabetes: recent progress and prevailing challenges. *J. Med. Chem.* **59,** 3579–3592 (2016).
21. Nunez, D. J. *et al.* Gut hormone pharmacology of a novel GPR119 agonist (GSK1292263), metformin, and sitagliptin in type 2 diabetes mellitus: results from two randomized studies. *PLoS ONE* **9,** e92494 (2014).
22. Ha, T. Y. *et al.* Novel GPR119 agonist HD0471042 attenuated type 2 diabetes mellitus. *Arch. Pharm. Res.* **37,** 671–678 (2014).
23. Katz, L. B. *et al.* Effects of JNJ-38431055, a novel GPR119 receptor agonist, in randomized, double-blind, placebo-controlled studies in subjects with type 2 diabetes. *Diabetes Obes. Metab.* **14,** 709–716 (2012).
24. Chu, Z. L. *et al.* A role for β-cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucose-dependent insulin release. *Endocrinology* **148,** 2601–2609 (2007).
25. Chu, Z. L. *et al.* A role for intestinal endocrine cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucagon-like peptide-1 and glucose-dependent insulinotropic peptide release. *Endocrinology* **149,** 2038–2047 (2008).
26. Lauffer, L. M., Iakoubov, R. & Brubaker, P. L. GPR119 is essential for oleoylethanolamide-induced glucagon-like peptide-1 secretion from the intestinal enteroendocrine L-cell. *Diabetes* **58,** 1058–1066 (2009).
27. Serrano, A. *et al.* Oleoylethanolamide: effects on hypothalamic transmitters and gut peptides regulating food intake. *Neuropharmacology* **60,** 593–601 (2011).
28. Fu, J. *et al.* Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR-α. *Nature* **425,** 90–93 (2003).
29. Lan, H. *et al.* GPR119 is required for physiological regulation of glucagon-like peptide-1 secretion but not for metabolic homeostasis. *J. Endocrinol.* **201,** 219–230 (2009).
30. Lauffer, L., Iakoubov, R. & Brubaker, P. L. GPR119: "double-dipping" for better glycemic control. *Endocrinology* **149,** 2035–2037 (2008).
31. Chen, Z. *et al.* Incorporation of therapeutically modified bacteria into gut microbiota inhibits obesity. *J. Clin. Invest.* **124,** 3391–3406 (2014).
32. Mimee, M. Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. *Cell Syst.* **1,** 62–71 (2015).
33. Vaudry, H. Molecular evolution of GPCRs: What we know and what the future holds. *J. Mol. Endocrinol.* **52,** E1–E2 (2014).
34. Lovejoy, D. A., Chang, B. S., Lovejoy, N. R. & del Castillo, J. Molecular evolution of GPCRs: CRH/CRH receptors. *J. Mol. Endocrinol.* **52,** T43–T60 (2014).
35. Hla, T. Genomic insights into mediator lipidomics. *Prostaglandins Other Lipid Mediat.* **77,** 197–209 (2005).
36. Wieland Brown, L. C. *et al.* Production of α-galactosylceramide by a prominent member of the human gut microbiota. *PLoS Biol.* **11,** e1001610 (2013).
37. Ou, G. *et al.* Proximal small intestinal microbiota and identification of rod-shaped bacteria associated with childhood celiac disease. *Am. J. Gastroenterol.* **104,** 3058–3067 (2009).

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

**Acknowledgements** We thank the High-Throughput and Spectroscopy Resource Center, Center for Clinical and Translational Science and the Comparative Bioscience Center at Rockefeller University for the use of their facilities; members of the Mangelsdorf laboratory at UT Southwestern and D. Drucker at Mount Sinai Hospital, Toronto for the use of the GLUTag cell line; A. Milshteyn, A. Estrela and J. Craig for their critical review of the manuscript. This work was supported in part by a grant from the Robertson Foundation, the Center for Basic and Translational Research on Disorders of the Digestive System, the Leona M. and Harry B. Helmsley Charitable Trust, Rainin Foundation, U01 GM110714-1A1 (S.F.B.), GM122559-01 (S.F.B.), the Crohn's and Colitis Foundation Career Development Award (L.J.C.) and NIDDK K08 DK109287-01 (L.J.C.).

**Author Contributions** L.J.C. and S.F.B. designed research; L.J.C. assisted with all experiments; S.-H.K. assisted with molecule characterization; E.A.G., P.Y.C., J.K.B. and R.R.A. assisted with gene cloning; D.E., A.B.E., S.M.H., C.H. and A.R. assisted with mouse experiments; J.C., X.V.-F., J.K. assisted with molecule synthesis; A.J.P. and J.R.C. assisted with metabolite analysis in human and mouse samples; L.J.C. and C.L. analysed data; L.J.C. and S.F.B. wrote the paper.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. 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 S.F.B. (sbrady@rockefeller.edu).

**Reviewer Information** *Nature* thanks D. J. Drucker and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Bioinformatics analysis of human *N*-acyl synthase genes.** Protein sequences for members of the PFAM family 13444 acetyltransferase (GNAT) domain (http://pfam.xfam.org/family/PF13444) ($n = 689$) were downloaded and corresponding gene sequences identified based on the European Bioinformatics Institute (EBI) number. A multiple sequence alignment was performed using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), generating a phylogenetic tree in Newick format with the '–guidetree-out' option. The 689 PFAM sequences were queried against the Human Microbiome Project (HMP) clustered gene index datasets and reference genome datasets with BLASTN (http://hmpdacc.org/HMGC/). The PFAM13444 sequences that aligned to a HMP gene (expectation ($E$) $< 1 \times 10^{-40}$ and >70% identity) were identified and comprise the human microbial *N*-acyl synthase (hm-*NAS*) gene dataset (143 hm-*NAS* genes). Reference genomes for 111 out of 143 hm-*NAS* genes were identified (Supplementary Table 2).

To determine the abundance of hm-*NAS* genes within microbiomes at specific human body sites, hm-*NAS* genes were queried against HMP whole-metagenome shotgun-sequencing data on a per-body site basis (http://hmpdacc.org/HMASM/). Each hm-*NAS* gene was BLASTN searched against the non-redundant gene sets from the following body sites: buccal mucosa, anterior nares, mid vagina, posterior fornix, vaginal introitus, retroauricular crease (combined left and right), stool, supragingival plaque and tongue dorsum. These body sites were chosen, because they contained sequence data from the largest number of unique patients[38]. hm-*NAS* genes and highly similar genes in the HMP non-redundant gene set ($E < 1 \times 10^{-40}$) were aligned to shotgun-sequencing reads from each patient sample taken from different sites in the human microbiome. Aligned reads were normalized to hm-*NAS* gene length and sequencing depth of each dataset. The normalized count of the reads aligned to each hm-*NAS* gene or its highly similar gene from the HMP non-redundant gene set were scaled (0–1) and colour coded per body site, and added as concentric rings around the phylogenetic tree (Fig. 1a). To determine the variability and distribution of hm-*NAS* genes that correspond to specific *N*-acyl amide families 1–6 (Fig. 1) in the human microbiome, normalized read counts for each hm-*NAS* gene from each *N*-acyl amide family were plotted separately per body site as RPKM (Fig. 1c). The tree in Fig. 1 was plotted using graphlan (https://huttenhower.sph.harvard.edu/graphlan).

**Analysis of metatranscriptome datasets.** Two RNA-seq datasets were identified with multiple patient samples taken from separate sites in the human microbiome[39,40]. One RNA-seq dataset was part of the HMP (http://hmpdacc.org/RSEQ/) and generated from supragingival samples taken from twin pairs with and without dental caries. The second RNA-seq dataset was generated from stool samples and compared different RNA extraction methods. We used only samples labelled 'whole', which functioned as controls for the original study[39]. Alignment of all hm-*NAS* genes to each dataset only identified hm-*NAS* genes from *N*-acyl amide family 1 and 2 in each of the RNA-seq datasets (1 in stool, 2 in supragingival plaque). To explore whether hm-*NAS* gene expression might vary between patient samples, we performed two different analyses. In the first analysis we identified reference genomes containing hm-*NAS* genes identical to those we used in heterologous expression experiments for molecule families 1 and 2 (*Bacteroides dorei* for compound **1**, *Capnocytophaga ochracea* for compound **2**). RNA-seq reads were aligned to all of the genes from each reference genome. For each genome the average per gene read density normalized for gene length was compared to the read density seen for the hm-*NAS* gene. The percentile of the normalized expression of each hm-*NAS* gene was then plotted (0 for not expressed, 1 for the most expressed) and compared between patient samples for each RNA-seq dataset (Fig. 2a). In the second analysis, the direct correlation between DNA and RNA abundance was determined for the stool metatranscriptome dataset for which DNA reads were also available[39]. RNA-seq and shotgun-sequenced DNA reads were aligned to the 15 hm-*NAS* genes from *N*-acyl amide family 1 that encodes *N*-acyl glycines (Supplementary Table 1). The reads were normalized (RPKM) and each hm-*NAS* gene from each patient sample was plotted as a single point with DNA and RNA read counts on the *x* and *y* axis, respectively (Fig. 2b).

**Heterologous expression of PFAM13444 genes in *E. coli*.** The 44 hm-*NAS* genes that we examined by heterologous expression were codon optimized, appended with NcoI and NdeI sites at the N and C terminus, respectively, and synthesized by Gen9. Genes obtained from Gen9 were digested with NdeI and NcoI and ligated into the corresponding restriction sites into a pET28c vector (Novagen). For heterologous expression purposes the resulting constructs were transformed into *E. coli* EC100 containing the T7 polymerase gene integrated into its genome (*E. coli* EC100:DE3). *E. coli* EC100:DE3 hm-*NAS*-containing strains were inoculated into 10 ml of Luria–Bertani (LB) broth supplemented with kanamycin (50 μg ml$^{-1}$) and grown overnight (37 °C with shaking 200 r.p.m.). The next day, 1 ml of overnight culture was used to inoculate 50 ml of LB broth supplemented with kanamycin (50 μg ml$^{-1}$) and isopropyl β-D-1-thiogalactopyranoside (IPTG) (25 μM). Cultures were incubated at 30 °C for four days with shaking (200 r.p.m.). Each culture broth was extracted with an equal volume of ethyl acetate and the resulting extracts were dried *in vacuo*. Crude extracts were resuspended in 50 μl of methanol and analysed by reversed phase HPLC–MS (XBridge$^{Tm}$ $C_{18}$ 4.6 mm × 150 mm) using a binary solvent system (A/B solvent of water/acetonitrile with 0.1% formic acid:10% B isocratic for 5 min, gradient 10% to 100% B over 25 min). Clone-specific metabolites encoded by each hm-*NAS* gene were identified by comparing experimental extracts with extracts prepared from cultures of *E. coli* EC100:DE3 transformed with an empty pET28c vector.

***N*-acyl amide isolation and structure determination.** For each group of clones that, based on LC–MS analysis, were predicted to produce a different *N*-acyl amide family, we chose one representative clone for use in molecule isolation studies. Each representative clone was grown in 1.5 l of LB broth in a 2.7-l Fernach flask (30 °C, 200 r.p.m.). After four days, cultures were extracted twice with an equal volume of ethyl acetate. Dried ethyl acetate extracts were partitioned by reversed-phase flash chromatography (Teledyne Isco, $C_{18}$ RediSep RF Gold 15 g), using the following mobile phase conditions: water:acetonitrile with 0.1% formic acid, 10% acetonitrile isocratic for 5 min, gradient to 100% acetonitrile over 20 min (30 ml per minute). Fractions containing clone-specific metabolites were pooled and semi-preparative reversed-phase HPLC was used to separate individual *N*-acyl amide molecules (Supplementary Information). The structures of compounds **2**–**6** were determined using a combination of HRMS, $^{1}$H, $^{13}$C and 2D NMR data (Supplementary Information). Compound **1** was described in our previous study[3].

**hm-*NAS* gene-containing bacterial species culture broth analysis.** *Capnocytophaga ochracea* F0287 (compound **2**), *Klebsiella pneumoniae* WGLW1-5 (compound **3**), *Neisseria flavescens* SK114 (compounds **4a** and **4b**), and *Gemella haemolysans* M341 (compound **6**) were obtained from the Biodefense and Emerging Infections Research Resources Repository (BEI Resources) HMP catalogue. Compound **1** was previously identified in culture broth extracts from cultures of *Bacteroides vulgatus*[3]. Each chosen bacteria strain contained an hm-*NAS* gene related to one which was heterologously expressed to produce compound **2**, **3**, **4a**, **4b** or **6**. Strains were inoculated under sterile conditions into 2 l of LYBHI medium (brain–heart infusion medium supplemented with 0.5% yeast extract (Difco), 5 mg l$^{-1}$ hemin (Sigma-Aldrich), 1 mg ml$^{-1}$ cellobiose (Sigma-Aldrich), 1 mg ml$^{-1}$ maltose (Sigma-Aldrich), 0.5 mg ml$^{-1}$ cysteine (Sigma-Aldrich)) and grown anaerobically (*C. ochracea*) or aerobically (*N. flavescens*, *G. haemolysans* and *K. pneumoniae*) for seven days. Culture broths were extracted with an equal volume of ethyl acetate. To look for the presence of *N*-acyl amides, these extracts were examined by HPLC–MS as was described for the original heterologous expression experiments. With the exception of family 3, the *N*-acyl metabolite that was heterologously expressed could be identified in the culture broth extracts from the bacteria that contained that hm-*NAS* gene (Extended Data Fig. 1).

**GPCR screen of *N*-acyl amide small molecules.** For each of the six *N*-acyl amide families (1–6), the analogue produced at the highest level in our heterologous expression experiments was assayed for GCPR activity. In the case of family 4, the major lysine analogue (*N*-3-hydroxyoleoyl lysine) was also screened. For family 1, the major glycine analogue (*N*-3-hydroxypalmitoyl glycine) was previously screened[3]. Using β-arrestin cell-based assays at 10 μM ligand concentration, agonist and antagonist activity was assessed by DiscoveRx Corporation against 168 GPCRs with known ligands as well as 72 orphan GPCRs. The most potent interactions between *N*-acyl amides and GPCRs were validated by repeating the assay in duplicate and generating dose–response curves. Synthetic *N*-acyl amides were assayed in the same way. HEK293 cells expressing GPR119 were exposed to equimolar concentrations of *N*-palmitoyl serinol or OEA. cAMP concentration was used as an orthogonal assay to measure GPR119 activation. cAMP was measured in HEK293 cells engineered to express a cAMP-sensitive ion channel that permits cAMP measurement in live cells by monitoring calcium efflux (ACTOne cells, Codex Biosolutions)[41]. Calcium efflux was measured using the ACTone membrane potential kit (Codex Biosolutions CB-80500-201). For our analysis, ACTOne HEK293 cells transfected with GPR119 (ACTOne-GCPR119) were compared with cells not transfected with GPR119. Each cell line was exposed to equimolar concentrations of OEA, *N*-oleoyl serinol or DCEA (5-(*N*-ethylcarboxamido)adenosine), which stimulates GPCR ADORA2B in the parental HEK293 cell line. For a reference to the quality control of each GPCR reporter cell line from DiscoveRx, please refer to https://www.discoverx.com/targets/cell-based-assay-list?class=gpcr&pcat=stable%20cell%20lines&readout=arrestin or Codex Biosolutions http://codexbiosolutions.com/actone_cell_lines.php.

**Synthesis of proteinogenic amino-acid-containing *N*-acyl-palmitoyl analogues.** Wang resins with preloaded amino acids were purchased from Matrix Innovation. Coupling reagents (PyBOP and Cl$^{-}$HOBt) were purchased from P3 BioSystems. Palmitoyl chloride and all other reagents were purchased from Sigma-Aldrich. Dimethylformamide (DMF) was added to preloaded Wang resins (~80 mg) and incubated for 30 min. Removal of *N*-Fmoc from swollen resins was accomplished by two rounds of piperidine treatment (20% solution in DMF (v/v), 3 ml) for 3 and

10 min, followed by several washes with DMF. Palmitoyl chloride (1 equivalent) in DMF was then added and the resin suspension was shaken for 2 h at room temperature. The *N*-acylated amino acid product was cleaved from the resins by treatment with trifluoroacetic acid (TFA) supplemented with 2.5% (v/v) water and 2.5% (v/v) triisopropylsilane (TIPS). After evaporation of TFA, the crude product was purified by automated reversed-phase flash chromatography (Teledyne Isco, $C_{18}$ RediSep RF Gold 15 g), binary solvent system: water and acetonitrile supplemented with 0.1% acetic acid. All final products were verified by mass spectrometry (Supplementary Table 3).

***In vitro* study of GLP-1 release from GLUTag cells.** OEA and 2-OG were purchased from Cayman Chemical Company and resuspended in DMSO to a concentration of 10 mM. *N*-oleoyl serinol was isolated and purified in the same manner as *N*-palmitoyl serinol described above and its identity was confirmed by $^1$H NMR and HRMS. *N*-oleoyl serinol was resuspended at a concentration of 10 mM in DMSO. GLUTag cells were obtained from the Mangelsdorf Laboratory (University of Texas Southwestern) with permission from D. Drucker (Mount Sinai Hospital Toronto). Cells were not tested for mycoplasma or validated as they came from the group who generated the cell line. GLUTag cells were grown in DMEM, low glucose, GlutaMAX (ThermoFisher) supplemented with 10% FBS and 1% penicillin–streptomycin. After the cells reached 80% confluency, they were collected and plated 1:1 into 24-well culture plates in fresh culture medium at 50,000 cells per well. After overnight growth in culture plates, cells were washed twice with Krebs buffer supplemented with 20 μl per ml DPP4 inhibitor (Millipore). GLUTag cells were incubated for 30 min in supplemented Krebs buffer and compounds were added at 1 μM and 100 μM. Cells were incubated with compounds for 2 h. Medium was then collected, centrifuged at 500*g* (4 °C) for 5 min and the cell-free supernatant was analysed for GLP-1 levels using the Active GLP-1 V2 kit (Mesoscale Discovery). Data from two independent experiments were analysed together for Fig. 5c ($n = 6$ wells for OEA, *N*-oleoyl serinol and $n = 4$ wells for 2-OG and DMSO).

**Colonization of germ-free and antibiotic-treated mice with *N*-acyl serinol-producing *E. coli*.** All experimental procedures were approved by the Animal Care and Use Committee of The Rockefeller University. Germ-free C57BL/6 mice were maintained in sterile isolators with autoclaved food and water in the Gnotobiotic Facility of the Mucida Laboratory at The Rockefeller University. Wild-type C57BL/6 mice were purchased from Jackson Laboratory. Eight-week-old mice were used for all experiments. For colonization studies, 5 ml of an overnight culture (LB with 50 μg ml$^{-1}$ kanamycin) of *E. coli* transformed with pET28c:hm-*NAS* *N*-acyl serinol synthase (treatment group) or *E. coli* transformed with the empty pET28c vector (control group) was centrifuged at 500*g* for 2 min, the supernatant was decanted and the cells were resuspended in 2 ml of sterile PBS. Germ-free mice were gavaged with 100 μl of bacterial culture immediately upon removal from sterile isolators. Antibiotic-treated mice were given water supplemented with 500 μg ml$^{-1}$ ampicillin for one week followed by gavage with *E. coli* clones. Suppression of endogenous microbiota was confirmed before gavage by culturing stool on non-selective LB agar and demonstrating no colony formation. After colonization, mice were housed in specific-pathogen-free conditions and fed autoclaved water and food. Water was supplemented with 35 μg ml$^{-1}$ kanamycin and 25 mM IPTG[32]. Faecal pellets from mice were analysed each week for three weeks to confirm colonization by the appropriate bacteria and to check for contamination by plating on LB agar with and without 50 μg ml$^{-1}$ kanamycin (Extended Data Fig. 7). A single fresh mouse faecal pellet was weighed and resuspended in PBS at a fixed concentration (400 μl per 40 mg). Subsequently, 1 μl of this suspension was diluted 1:100 in PBS and plated in duplicate. The average number of colonies per plate (two plates) was recorded. Plasmids were isolated and restriction mapped from these colonies to confirm the presence of the correct hm-*NAS* gene insert or lack thereof. Ethyl acetate extracts from broth cultures were also examined as previously discussed to confirm the production of *N*-acyl serinols by bacteria in the treatment group. In the first experiment of germ-free mice, seven mice were studied—3 (one male, two females) in the treatment group and four (two males, two females) in the control group. The independent replicate experiment of germ-free mice also consisted of seven mice (all female, three in the treatment group and four in the control group). Mice were all individually caged and at the end of each week, food consumption and weight were measured. In the antibiotic-treated mouse cohort 18 mice were studied (two independent experiments of nine male mice in the treatment group and nine male mice in the control group). Mice were of the same size and gender distribution in each replicate. In the antibiotic-treated study, the control mouse group was treated with *E. coli* transformed with pET28c:hm-*NAS*, which contains an *N*-acyl serinol synthase gene containing a single point mutation that rendered the enzyme non-functional (E91A) (see below). The animal experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size. Data for all mice that completed the experiments were analysed.

**Generation of an active-site pET28c:hm-NAS *N*-acyl serinol synthase mutant.** Conserved active site residues in bacterial *NAS* genes were identified in previous biochemical and X-ray crystallography studies[42]. To create a catalytically inactive *N*-acyl serinol synthase, we changed a key glutamic acid residue (Glu91) to alanine. The point mutation was created by PCR using a pET28c:hm-*NAS* *N*-acyl serinol synthase vector as template and the following primers forward: GTTCTGTGCGATACGTCTCC; and reverse: GCCTTTCACAGGCAGATATTC. The position of the point mutation is underlined in the forward primer. The resulting PCR reaction was digested with DpnI to remove any remaining methylated vector. The PCR product was then phosphorylated, column purified and blunt end ligated (End-It, Epicentre). The vector was transformed into EC100:DE3 cells and the point mutation was confirmed by Sanger sequencing. When transformed into *E. coli*, the resultant E91A mutant construct did not confer the production of any detectable *N*-acyl serinols. Cultures were grown under the same conditions as the original *N*-acyl serinol synthase producing clone (Extended Data Fig. 8).

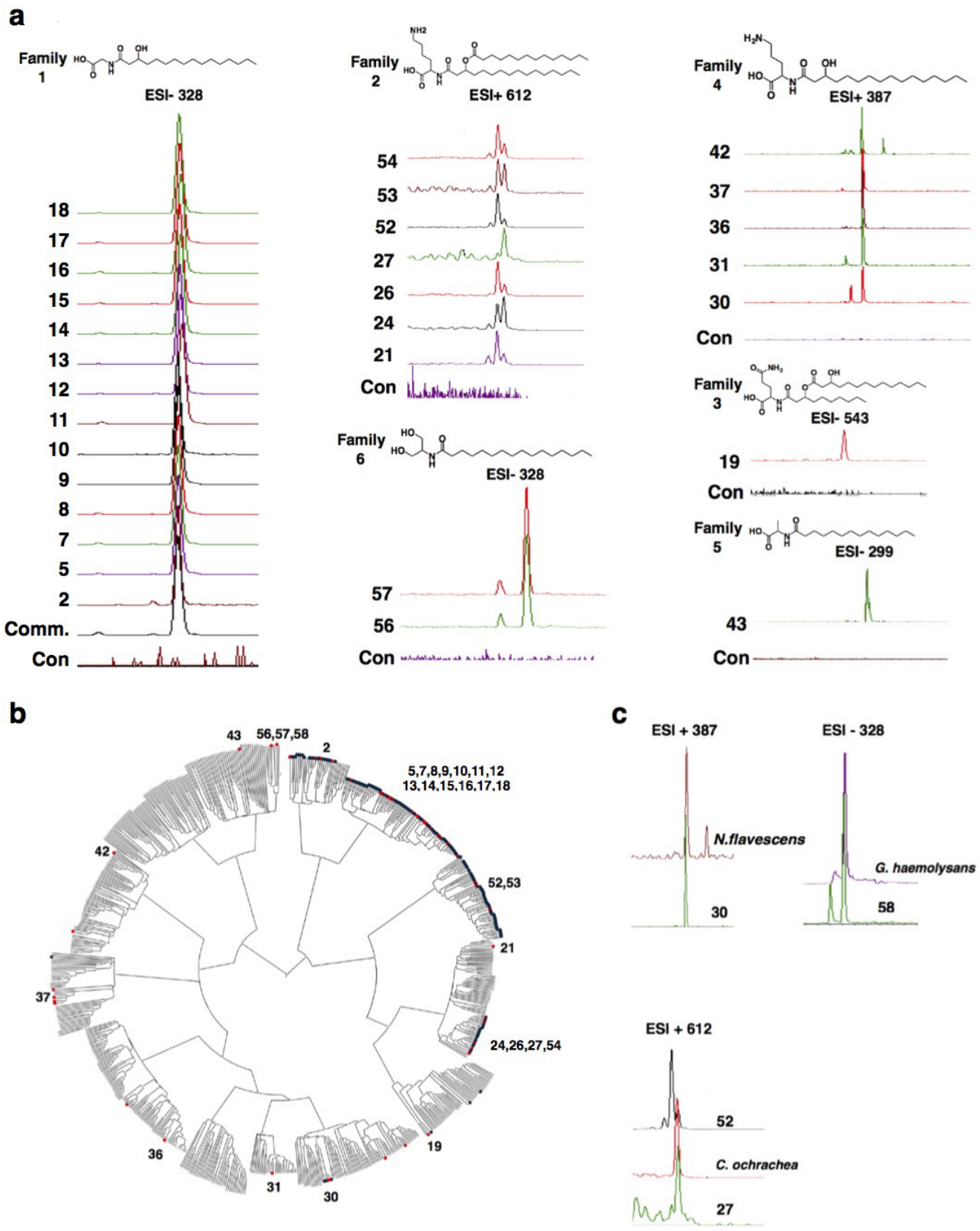
**Oral glucose tolerance test.** One week after colonization, mice were fasted overnight (16 h) and then administered a 2 g kg$^{-1}$ OGTT (40% glucose solution). Blood glucose was measured by tail bleed (Breeze 2 Bayer) at time 0, before the glucose gavage, then 15, 30, 60, 90 and 120 min after gavage. After one week, the IPTG was removed from the drinking water and mice were allowed to equilibrate for an additional two weeks[32]. Three weeks after colonization the OGTT was repeated. Blood glucose levels at each time point during the OGTT tests were compared between groups using a Student's *t*-test with a significance threshold of $P < 0.05$.

**Insulin and GLP-1 measurement.** Mice were given an OGTT as previously described. At 15 min, blood was collected by submandibular bleed and immediately mixed with 10 μl of 0.5 M EDTA and 5 μl of DPPIV inhibitor (Millipore, DPP4-010) per 500 μl of blood. Treated blood was spun at 2,000*g* for 15 min at 4 °C. Plasma was collected and immediately placed at −80 °C. Insulin was measured using the Crystal Chem Ultra Sensitive Mouse ELISA kit and active GLP-1 was measured using the Mesoscale Discovery Active GLP-1 V2 kit. Samples were analysed for insulin in triplicate and GLP-1 in duplicate. Insulin was measured from mice in one experiment ($n = 6$ mice in each group). GLP-1 was measured from mice in two independent experiments ($n = 9$ mice in the control, $n = 10$ mice in the treatment group). All mice were male and of the same size distribution.

***N*-acyl serinol metabolite measurement in mouse caecal samples and human stool.** After withholding IPTG for two weeks, mice from the first experimental set were re-exposed to IPTG in the drinking water for one week to induce hm-*NAS* gene expression and *N*-acyl serinol production. Mice were euthanized and caecal samples taken. Fresh caecal stool from two control mice and two treated mice was resuspended in 5 ml of sterile PBS and extracted 1:1 with ethyl acetate. Crude extracts were dried *in vacuo* and resuspended in methanol normalized by crude extract weight. Each extract was then analysed by reversed-phase liquid chromatography coupled to a 6550 Q-TOF mass spectrometer (Agilent Technologies). Peak identities were confirmed by accurate mass, and also by comparison of chromatographic retention time and MS/MS spectra to those of the purified *N*-palmitoyl serinol standard. For both mice in the treatment group, *N*-palmitoyl serinol could be detected in the caecal samples, whereas *N*-palmitoyl serinol was detected in neither of the mice in the control group (Extended Data Fig. 6). Stool samples were collected from human subjects before bone marrow transplant as part of an observational clinical study conducted at Memorial Sloan Kettering Cancer Center (MSKCC) in collaboration with J.C. All samples were collected with informed consent under protocol numbers 09-067, 09-141 and 06-107 at MSKCC and stored as part of a biospecimen repository. All sample processing was done at MSKCC. Patient age, gender and transplant indication are provided in the table in Extended Data Fig. 9. Fresh stool samples were processed in the same way as the mouse caecal samples described above.

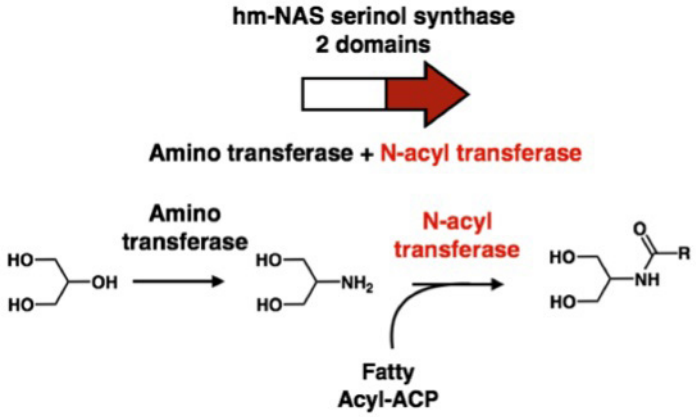
**Data availability.** All figure data are available as Source Data. Gene accession numbers for all cloned genes are provided in Supplementary Table 1. Publically available DNA and RNA datasets that were analysed in this study are referenced accordingly and references contain links to datasets that are available for downloading.


38. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
39. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl Acad. Sci. USA* **111,** E2329–2338 (2014).
40. Peterson, S. N. *et al.* Functional expression of dental plaque microbiota. *Front. Cell. Infect. Microbiol.* **4,** 108 (2014).
41. Tang, Y., Li, X., Han, X., Lu, J. & Diwu, Z. Functional analysis of endogenous β-adrenergic receptor through fluorimetric monitoring of cyclic nucleotide-gated ion channel. *Anal. Biochem.* **360,** 303–305 (2007).
42. Van Wagoner, R. M. & Clardy, J. FeeM, an *N*-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding. *Structure* **14,** 1425–1435 (2006).
43. Chen, Y. *et al.* Dysbiosis of small intestinal microbiota in liver cirrhosis and its association with etiology. *Sci. Rep.* **6,** 34055 (2016).
44. Cheng, J. *et al.* Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease. *BMC Gastroenterol.* **13,** 113 (2013).

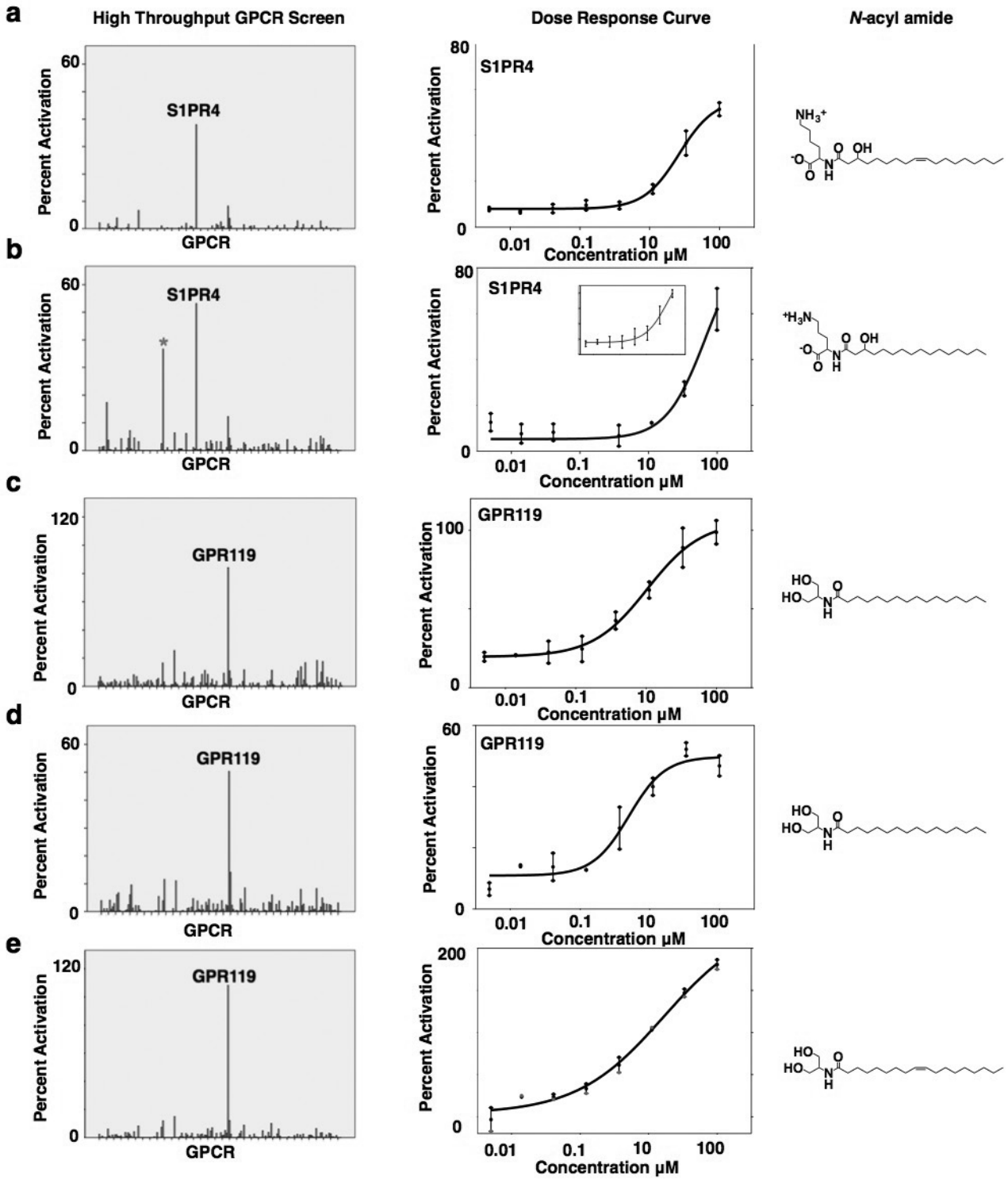



**Extended Data Figure 1 |** See next page for caption.

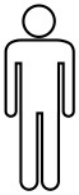
**Extended Data Figure 1 | Analysis of hm-NAS clone families. a**, LC–MS analysis of crude extracts prepared from *E. coli* transformed with each hm-*NAS* gene expression construct (number 1–58, see Supplementary Table 1 for details about each clone number) compared to negative control extracts derived from *E. coli* containing an empty vector (Con). On the basis of metabolite retention time and observed mass hm-*NAS* genes could be grouped into 6 *N*-acyl amide families (1–6). The mass of the major metabolite (pictured) from each *N*-acyl amide family is shown in either the $ESI^+$ or $ESI^-$ mass spectrometry detection mode for each hm-NAS extract, including the control extract. Functional differences in NAS enzymes follow the pattern of the NAS phylogenetic tree, with hm-*NAS* genes from the same clade or sub-clade mostly encoding the same metabolite family. Commendamide (Comm.) was previously isolated and is part of family 1 (ref. 3). **b**, Phylogenetic tree of PFAM13444, showing the location of each hm-*NAS* gene that we synthesized and examined by heterologous expression. **c**, Crude ethyl acetate extracts were prepared from cultures of bacterial species that contain the same or a highly related (>80% nucleotide identity) hm-*NAS* gene that was expressed by heterologous expression. The only exception was for the *N*-acyl alanine family, for which a representative cultured commensal bacterial species was not available. *N*-acyl glycines were previously analysed in the same manner[3]. The extracted-ion chromatogram for the hm-*NAS* gene family is shown for the *E. coli* clone compared to the crude extract from the commensal species. For family 6, hm-NAS clone 58 is pictured in **c** and not in **a** owing to formatting constraints.



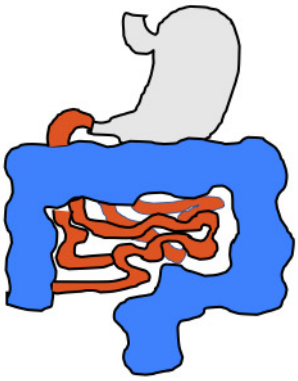

**Extended Data Figure 2 | Proposed biosynthesis of *N*-acyl serinol.** Proposed two-step biosynthesis of *N*-acyl serinol using the two domains encoded by the hm-*NAS N*-acyl serinol synthase gene.



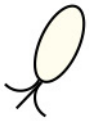

**Extended Data Figure 3 | Validation of hits from the high-throughput GPCR screen.** When structural analogues were independently screened in the GPCR panel (for example, *N*-oleoyl and palmitoyl serinol or *N*-3-hydroxypalmitoyl lysine and ornithine), they yielded the same GPCR profile and when *N*-acyl serinol was re-assayed across all GPCRs in the panel, it also yielded the same GPCR activity profile. **a**, **b**, *N*-3-hydroxypalmitoyl lysine and ornithine both interact with S1PR4. *Did not repeat. **b**, inset technical repeat of *N*-3-hydroxypalmitoyl ornithine dose–response curve. **c**, **d**, Technical repeats of *N*-palmitoyl serinol. **c**–**e**, *N*-palmitoyl and oleoyl serinol both interact with GPR119. Screening data were performed once, dose–response curves were performed in duplicate. Data are mean ± s.e.m.



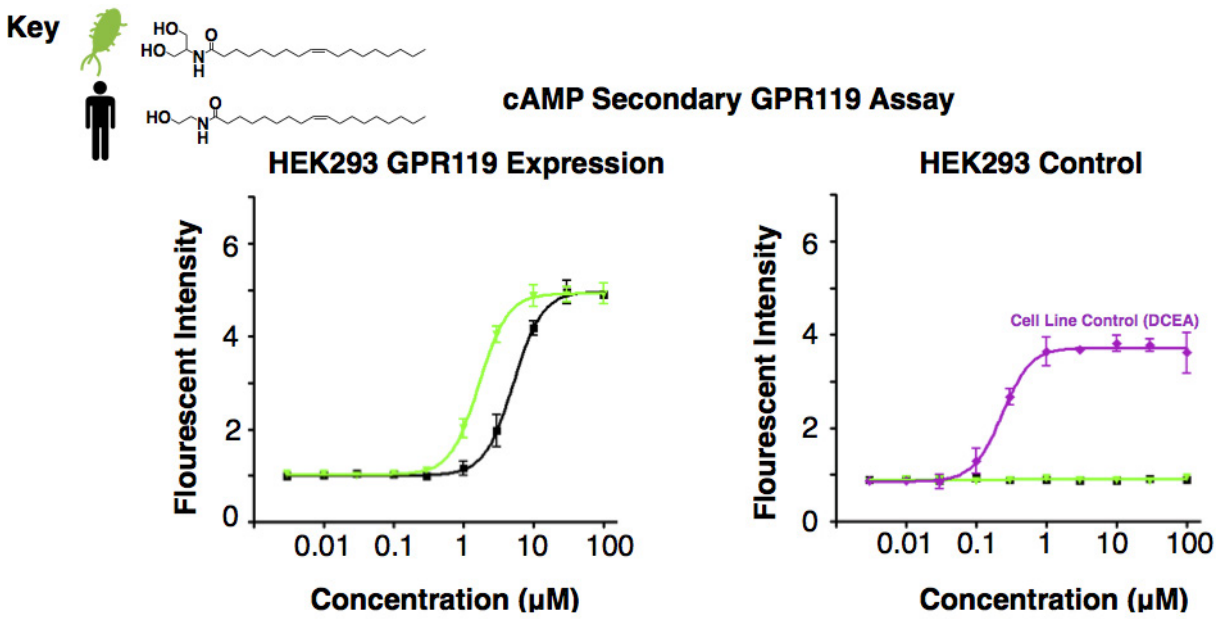
| GPCR | Gastrointestinal Site |
|---|---|
| GPR119 | Duodenum<br>Ileum |
| S1PR4 | Ileum<br>Duodenum<br>Colon |
| GPR132 | Ileum<br>Duodenum<br>Colon |
| PTGIR | Ileum<br>Duodenum<br>Colon |
| PTGER4 | |



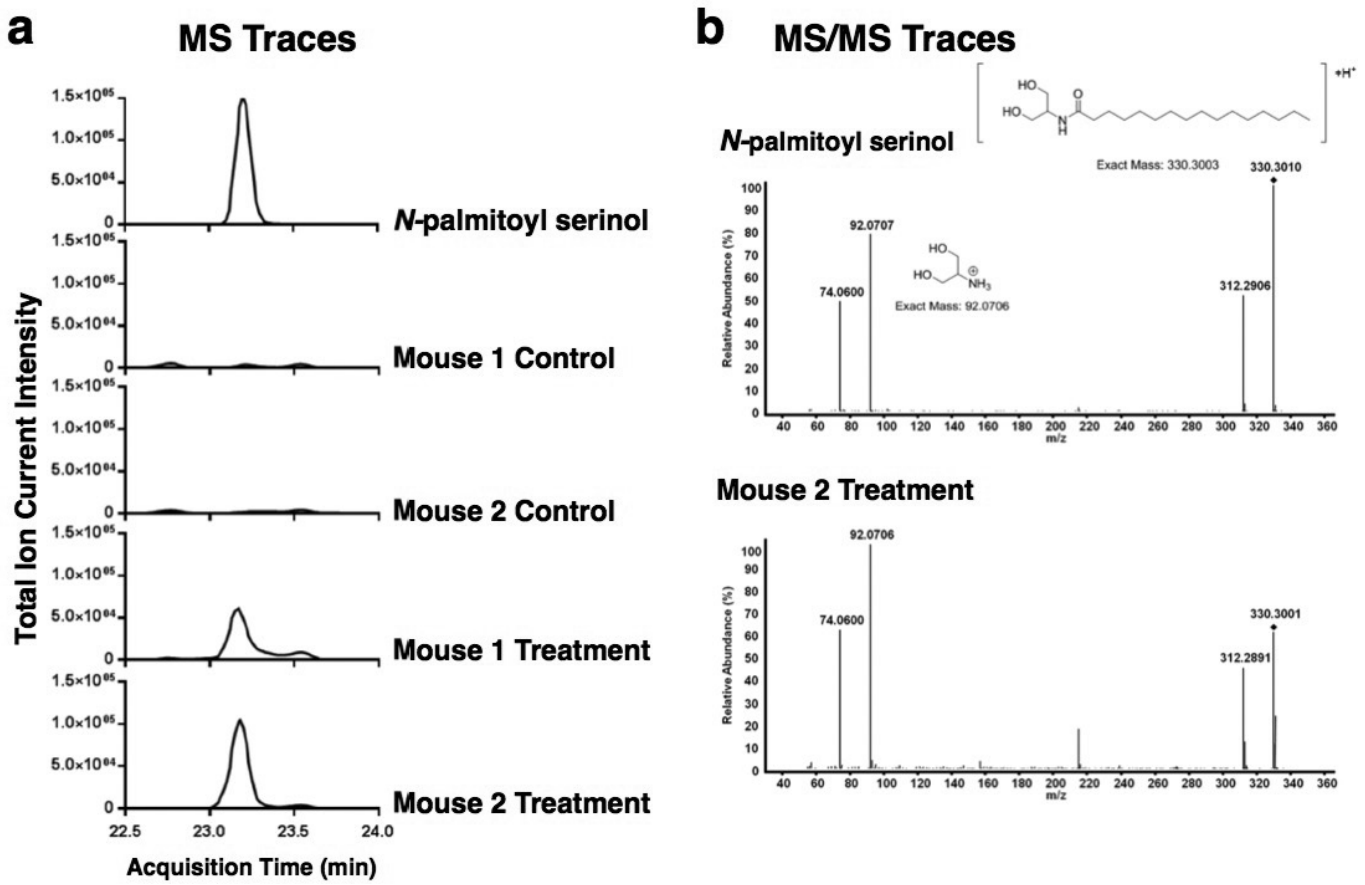
| Gastrointestinal Site | *N*-acyl amide *Genus* |
|---|---|
| Oral<br>Duodenum<br>Ileum | *N*-acyl serinols<br>*Gemella* |
| Oral<br>Duodenum<br>Ileum | *N*-acyl ornithines<br>*Neisseria* |
| Ileum<br>Colon | *N*-acyl glycines<br>*Bacteroides* |
| Colon | *N*-acyloxyacyl glutamine<br>*Klebsiella* |



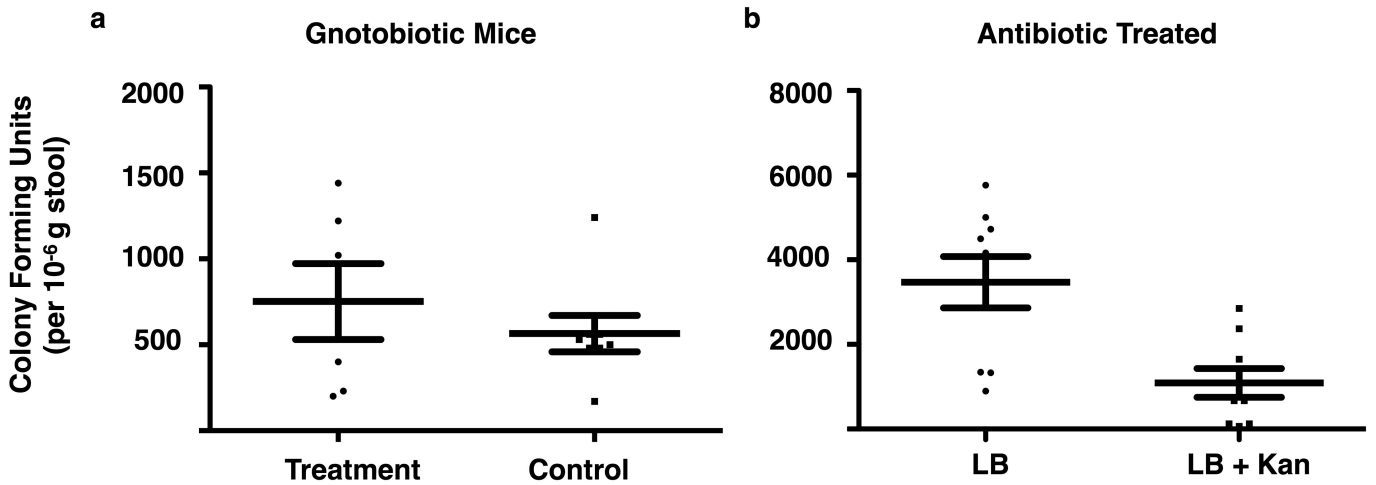
**Extended Data Figure 4 | Combined analysis of protein and transcript expression of GPCRs in the gastrointestinal tract.** Links between GPCR, *N*-acyl amide, bacterial genus and the site where these co-occur in the gastrointestinal tract (coloured) are shown. On the basis of protein expression data (Human Protein Atlas) GPR119 is most highly expressed in the pancreas and duodenum, S1PR4 in the spleen and lymph node, G2A in the lymph node and appendix, PTGIR in the lung and appendix and PTGER4 in the bone marrow and small intestine. From gene expression data in the colon (GTEx dataset, $n = 88$ patient samples from small intestine, 345 patient samples from colon) GPR132, PTGER4 and PTGIR are all expressed alongside the *N*-acyl synthase genes that are known to encode metabolites that target these GPCRs (Fig. 1). In the gastrointestinal tract, GPR119 and S1PR4 are most highly expressed in the small intestine where 16S studies have identified bacteria from the genera *Gemella* and *Neisseria*. All known reference genomes (NCBI) from these genera contain *N*-acyl synthase genes that are highly similar (BLASTN, $E = 2 \times 10^{-132}$) to those we found to encode GPR119 or S1PR4 ligands[37,43,44].



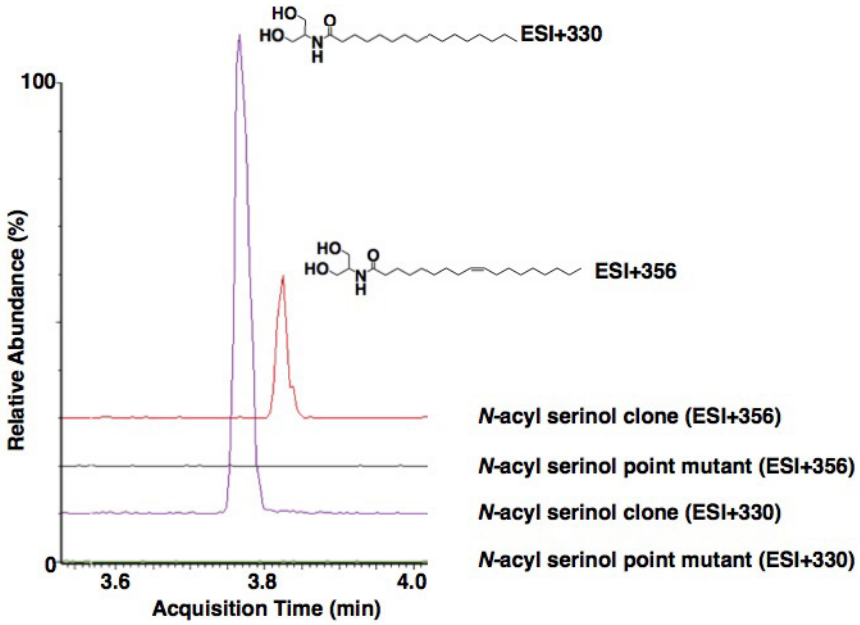

**Extended Data Figure 5 | Secondary assay of GPR119.** ACTOne HEK293 cells (control) and ACTOne HEK293 cells transfected with GPR119 were exposed to equimolar concentrations of the endogenous GPR119 ligand OEA or the bacterial ligand *N*-oleoyl serinol. Relative fluorescent intensity was recorded for each ligand concentration compared to the background signal. All data points were performed in quadruplicate and data are mean ± s.d. An increase in cAMP concentration was observed in HEK293 cells expressing GPR119, but not in control HEK293 cells. The DCEA (5-(*N*-ethylcarboxamido)adenosine) control is presented to confirm cAMP response of the parental cell line. The $EC_{50}$ for *N*-oleoyl serinol (bacterial) was 1.6 μM and for oleoylethanolamide was 5.1 μM, which are consistent with data from the β-arrestin assay (Fig. 5a).



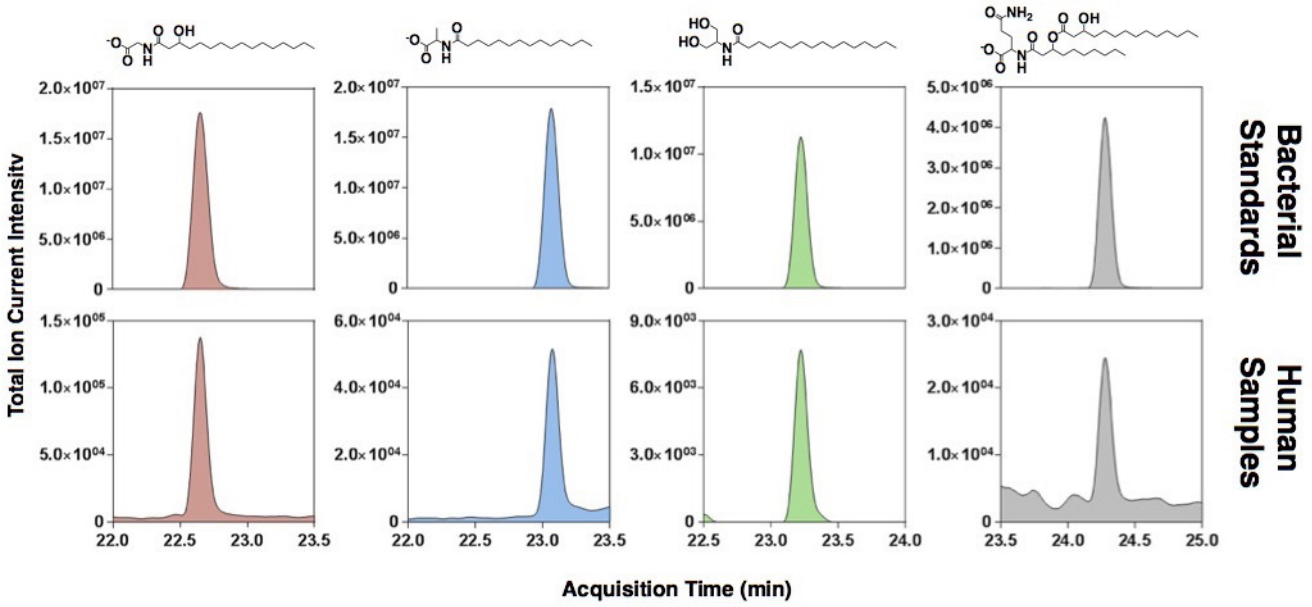

**Extended Data Figure 6 | Identification of *N*-acyl serinol biosynthesis *in vivo*.** **a**, LC–MS analysis of crude caecal extracts. Extracted-ion chromatograms for palmitoyl serinol ($[M + H]^+$ *m/z*: 330.3003) are shown. A peak with the same exact mass and chromatographic retention time as the *N*-palmitoyl serinol standard was present in treatment mice but not control mice. Treatment mice were colonized with *E. coli* containing the *N*-acyl serinol synthase gene. Control mice were colonized with *E. coli* containing the empty pET28c vector. **b**, Identification of *N*-palmitoyl serinol by MS/MS fragmentation of the *m/z* 330.3003 ion. In the MS2 spectrum the diamond indicates the *N*-palmitoyl serinol parent ion and the product ion at *m/z*: 92.0706 shows the presence of the serinol head group.




**Extended Data Figure 7 | Bacterial colonization of mouse model systems.** One week after inoculation with *E. coli*, a single faecal pellet from a colonized mouse was collected, resuspended in 400 μl PBS and plated at a 1:100 dilution onto LB agar plates with or without kanamycin 50 μg $ml^{-1}$. **a**, The number of colony-forming units per $10^{-6}$ g of faeces observed on LB agar plates with kanamycin was similar for the treatment group (*E. coli* with hm-*NAS* gene, $n = 6$ mouse stool samples) and the control group (*E. coli* with empty vector, $n = 8$ mouse stool samples). **b**, In the antibiotic-treated mouse cohort, other colonizing bacteria are present. Stool samples produced threefold more colony-forming units on unselected LB agar plates compared to LB agar plates with kanamycin. Data are mean ± s.e.m. In both cases, when random colonies were picked from the LB with kanamycin plates, all colonies were found to contain the cloning vector. indicating these were in fact *E. coli* colonizing bacteria.




**Extended Data Figure 8 | *N*-acyl serinol synthase point mutant.** LC–MS analysis of crude extracts prepared from cultures of *E. coli* expressing either the *N*-acyl serinol synthase gene or the *N*-acyl serinol synthase gene with an active site point mutation (E91A). *N*-acyl serinol metabolites (for example, *N*-palmitoyl serinol and *N*-oleoyl serinol) are absent from the culture broth with the point mutant ($ESI^+$ mode). This mutant was created to address the possibility that the observed mouse phenotype might be due to overproduction of any protein by *E. coli* and not specifically from *N*-acyl serinol production.




**Patient age and transplant indication**

| Age | Gender | Disease | status | IRB | Treatment |
|---|---|---|---|---|---|
| 69 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 68 | F | Multiple Myeloma | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 41 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 60 | F | Multiple Myeloma | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 32 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 37 | M | Multiple Myeloma | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 25 | M | Hodgkin's Disease | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 66 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 59 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 23 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 53 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 46 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 64 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 38 | M | Non-Hodgkin's Lymphoma | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 60 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 51 | F | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 66 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 38 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |
| 64 | F | Leukemia | consented 06-107 | 06-107 | Bone Marrow Transplant per Standard of Care |
| 45 | M | Leukemia | consented 09-141 | 09-141 | Bone Marrow Transplant per Standard of Care |

**Extended Data Figure 9 | Detection of N-acyl amides in human faecal samples.** High-resolution reversed-phase LC–MS analysis of human faecal extract pooled from 128 samples representing 21 individuals. Extracted-ion chromatograms for individual *N*-acyl amides are shown within a 2 p.p.m. tolerance of the exact mass (M + H). Compounds observed to be present in the human faecal extract were confirmed by alignment to authentic standards (top) and by spiked addition of the pure compound (data not shown). No zwitterionic *N*-acyl amides (*N*-acyl or *N*-acyloxyacyl ornithine/lysines) were detected. Patient demographics are provided in the table. Human stool samples were collected with informed consent at MSKCC under institutional review board numbers 09-167, 09-141, and 06-107.

nature research

Corresponding author(s): Sean Brady

☐ Initial submission ☒ Revised version ☐ Final submission

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but 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. — Data reported in figure 5 and in methods pg 20-21

2. Data exclusions

   Describe any data exclusions. — No data was excluded

3. Replication

   Describe whether the experimental findings were reliably reproduced. — Replicate experiments were reliably reproduced and summarized in the manuscript

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups. — NA - statement made Pg 21

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis. — NA - statement made Pg 21

   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 in 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 clear description of 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. | NA |

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## 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. | NA |

9. Antibodies

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

10. Eukaryotic cell lines

| | |
|---|---|
| a. State the source of each eukaryotic cell line used. | Pg - 20 - only cell line used obtained through the originator |
| b. Describe the method of cell line authentication used. | Pg - 20 - only cell line used obtained through the originator |
| c. Report whether the cell lines were tested for mycoplasma contamination. | No mycoplasma test - cell lines obtained from originator |
| d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | NA |

## 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. | Pg 20 |

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. | Specimens were collected between 2010 and 2014 as part of a biospecimen repository under informed consent. Protocol was approved by Memorial Sloan Kettering IRB protocol #09-067, 09-141 and 06-107. Patients underwent a standard of care bone marrow transplant and stool samples collected before, during and after transplant. This is an observational study in a single cohort and as such there are no comparative statistics. Basic demographic information is provided in a table in the correction statement. |

CORRIGENDUM

doi:10.1038/nature25997

# Corrigendum: Commensal bacteria make GPCR ligands that mimic human signalling molecules

Louis J. Cohen, Daria Esterhazy, Seong-Hwan Kim, Christophe Lemetre, Rhiannon R. Aguilar, Emma A. Gordon, Amanda J. Pickard, Justin R. Cross, Ana B. Emiliano, Sun M. Han, John Chu, Xavier Vila-Farres, Jeremy Kaplitt, Aneta Rogoz, Paula Y. Calle, Craig Hunter, J. Kipchirchir Bitok & Sean F. Brady

*Nature* **549,** 48–53 (2017); doi:10.1038/nature23874

In this Article, the description in the Methods of the human stool samples used for the analysis presented in Extended Data Fig. 9 was incomplete. All samples were collected with informed consent under protocol numbers 09-067, 09-141 and 06-107 at Memorial Sloan Kettering Cancer Center (MSKCC) and stored as part of a biospecimen repository. All sample processing was done at MSKCC. All patients underwent a bone marrow transplant which was standard of care and part of an observational study (not a clinical trial, as originally stated in the Methods). Patient age, gender and transplant indication are provided in the table now added to Extended Data Fig. 9. The Reporting Summary has been updated (the original uncorrected Reporting Summary is provided as the Supplementary Information to this Corrigendum). The original Letter has been corrected online.

**Supplementary Information** is available in the online version of this Corrigendum.