Spring 2022

Functional Profiling of Human Commensal Metabolites in Colorectal Cancer and Pain Sensation

Yiyun Cao
Yale University Graduate School of Arts and Sciences, yiyuncao.lifesciences@gmail.com

Abstract

Functional Profiling of Human Commensal Metabolites in Colorectal Cancer and Pain Sensation

Yiyun Cao

2022

The gastrointestinal tract contains the largest number and greatest diversity of microbes which are referred to collectively as the gut microbiota. Alterations in microbiota dysbiosis are associated with diverse disease states, including colorectal cancer (CRC), inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS). However, the causative mechanisms mediated by the gut microbiota are still to be illustrated. Particularly, the gut commensals encode ~150 times more genes than the human genome and thousands of metabolites that mostly lack functional annotations or structural identification. Therefore, the functional profiling of human commensal metabolites is promising to reveal the molecular mechanisms of the gut microbiota regulating homeostasis or disease outcomes.

Through systematic forward-screening methods on a large collection of human gut commensals from divergent phylogenies, we investigated how the gut microbiota influences physiological outcomes through small molecule metabolites in CRC colorectal cancer (CRC) and pain sensation. 1) We established an electrophoresis-based pipeline to evaluate the genotoxicity of microbial small molecule metabolites causing DNA damage. Using comparative metabolomics and bioactivity-guided natural product-discovery techniques, we discovered a previously undescribed family of genotoxic metabolites—termed the indolimines—produced by the CRC-associated species *Morganella morganii* whose ability to exacerbate CRC independent of inflammation was confirmed in gnotobiotic mice. Through transposon-based random mutagenesis, we identified the gene *aat* (encoding AAT_I protein, aspartate aminotransferase fold type I) responsible for the synthesis of primary amines, precursors of indolimines. The *att- M. morganii* exhibited defects of genotoxicity. This project reveals the existence of a previously unexplored universe of genotoxic small molecules from the human microbiome and implies a broader role for microbiota-derived genotoxins in CRC. In addition, the cancer-promoting mechanism of *M. morganii* is dissected at genetic, molecular and physiological levels. The understanding about genotoxins could instruct the diagnoses or therapies of CRC in the future. 2) We established a calcium influx-based pipeline to investigate the

endogenous agonists of TRPV1 ion channel derived from the gut microbiota. We found that *Klebsiella* species and *Accidaminococcus intestini* enabled to produce heat-stable, small-molecule metabolites to activate TRPV1 channel. Particularly, *A. intestini* was confirmed to sensitize TRPV1+ nociceptive dorsal root ganglion (DRG) neurons, suggesting the effects of commensal metabolites in pain sensation and relevant diseases like IBS. Untargeted global metabolomics and screening of human small molecule metabolite library implied that microbial-specific phospholipids might be a novel class of endogenous ligands of TRPV1. We are still working on bacteria engineering to identify the genes or gene clusters responsible for the synthesis of TRPV1 agonists. This study indicates a potential mechanism by which the gut commensals regulate visceral pain sensation and provides a new direction for IBS diagnoses or therapies. In summary, both projects underline the significance of functional profiling of the gut microbiota in dissecting causative mechanisms in human diseases, especially those with complex etiology such as cancer, pain syndromes and neurological disorders.

Functional Profiling of Human Commensal Metabolites in Colorectal Cancer and Pain Sensation

A Dissertation
Presented to the Faculty of the Graduate School
Of
Yale University
In Candidacy for the Degree of
Doctor of Philosophy

By
Yiyun Cao

Dissertation Director: Noah Palm

May 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

Acknowledgments

I would like to express my deep and sincere gratitude to my adviser, Dr. Noah Palm, for his invaluable guidance and support throughout my years at Yale University. His attitude towards science impressed me and he is always encouraging me to become a pure scientist. First, he is always trying to avoid opportunistic projects and focus on the real significant biological questions. We discussed such choice as ‘planting a tree’ but not just ‘clipping leaves’, which definitely makes my life more difficult because all projects are challenging, risky and time-consuming. Luckily, Noah is very optimistic, patient to my progress and supportive when I feel anxious. I remember that he said ‘graduate school is hard and 99% of your work might fail, but it’s good enough if the 1% would work’. I feel so lucky to stay in such a great lab to explore new areas bravely, and glad to say that more than 1% worked finally. Noah also enjoys science, explores various directions and follows scientific updates. He is open to many novel techniques and new concepts, and generous to share his ideas. As a career model, Noah influences me deeply and leads me to develop such good habits. For example, I really enjoy CYOA (choose your own adventure) meetings in our lab, from which I brought out my two projects: one is my thesis project and the other one further inspires my future research interest.

Besides the mentoring on science, Noah also showed me how to be a great person. He is an active and generous collaborator. For scientists who work in an interdisciplinary area, it’s necessary to build up a good collaborative environment. Noah is humble and respects everyone, from senior professors to naïve fresh trainees. He showed me that it might be quicker to do everything by oneself but it must be better to work together as a team. As his student, I was also supported to attend various conferences and communicate with other scientists, even though I didn’t give formal presentations and some conferences seemed not directly related to my projects. I went to NCI meeting, AAI meeting, Neuron-Immune axis symposium and Gut-Brain axis symposium in the past years. Noah also helped me with writing (fellowships, proposals, manuscript) and presenting. He sent me ‘good job’ after every RIP talk which makes me more confident and comfortable to give public talks. Noah is also very considerate when I had travelling

problems during pandemic. He took care of my ongoing experiments, helped me come back with transition in Dubai, and protected us from COVID. I was also impressed by Noah's work-life balance. Although super busy with grant writing or meetings every day, he still spends enough time with his family. He makes me believe that science could absolutely fit well with personal life.

I also want to thank my thesis committee, Dr. Andrew Goodman, Dr. Jason Crawford and Dr. Akiko Iwasaki, for their insightful comments and constructive suggestions to my projects and career plan. They put forward significant questions during my committee meetings and pointed out what the key problems I should prioritize. Sometimes I feel stress and it's impossible to solve the hard questions but all such tough requirements made my projects better and better. They also gave me a lot of suggestions on postdoc applications and how to become an independent researcher.

I am also extremely grateful to current and former members of the Palm lab for all of their help and support. I would like to thank Haiwei Chen, Tyler Rice, Connor Rosen and Agata Bielecka who taught me a lot when I just joined this lab. It's hard for me to quicky make friends with strangers, so they gave me warmth, kindness, generosity to be a part of this family. Mytien Nguyen, Nicole Sonnert, Anjelica Martin, Deguang Song, Yi Yang, Shana Leopold and everyone else contributed to my project and the lab environment. My great collaborators also supported a lot to my projects. I started the genotoxin project with Dr. Mengzhao Xue from Dr. Seth Herzon's Lab. The night we brainstormed together and the time we did experiments together were exciting. Then, I was lucky to have Dr. Joonseok Oh from Dr. Jason Crawford's Lab, who helped me push the project forward on the chemical side. Joonseok is also a good friend and we always keep our progresses on the same page and communicate in time. In addition, I also got help from other labs or facilities: Dr. Bentley Lim, Dr. Djordje Bajic, Dr. Yiqiao Bao and Amy Ontai from Dr. Andrew Goodman's Lab helped with bacteria engineering; Dr. Sheila Umlauf and Dr. Yulia Surovtseva from Yale Center for Molecular Discovery helped with mutant library pooling; Dr. Manolis Roulis and Dr. Esen Sefik from Dr. Richard Flavell Lab helped with mouse work; Dr. Viktor Feket and Dr. Marco Mastrotto from Dr. Elena Gracheva Lab helped with the TRPV1 project; Dr. Jaymin Patel and Dr. Christina Cho from Dr. Jason Crawford Lab also provided help;

YCGA center helped with sequencing and data analysis. My friends and classmates, Weiwei Han, Yixiang Wang, Jiawei Wang, Tianyang Mao solved my various questions and problems. I appreciate it greatly that Yale is such a great place to work collaboratively and communicate openly and freely.

I want to give my special thanks to my friends who enriched my life beyond science. Hao Xu not only taught me many experiments when I just arrived here, but also encouraged me to stay in academia when I had a difficult time. He also took great care of my cat Knight during my travels, especially during pandemic. Xiruo Li travelled with me together to Florida, San Francisco, Los Angeles during winter breaks and we shared so many interests from novels, animations to TV shows. I also had great time with Runfeng Miao, Peiqi Li, Danyan Li and other friends at Yale. Ziqi Chen, Sihan Xiong, Yuyao Wu, and other friends in China. We miss each other a lot and hope to meet in-person soon.

I would like to thank my family for being supportive during my Ph.D. My parents are always encouraging me to work towards my interests. As the first student admitted by college in my family, I couldn't get professional or experienced advice on my major. However, my parents are proud of me and satisfied with my choices as long as I feel happy. Such an attitude also influences my life and career, making me focus more on my own feelings and dreams. My little sister Yige, my cat Knight and Yige's cat Feifei are amazing sources of happiness. In addition, I also want to particularly thank my models in various areas. Dr. E. John Wherry is a great scientist who I randomly knew from a conference in 2015. He kindly spent time talking with me through Zoom when I felt confused about my career plan. His enthusiasm impressed me and made me decide to pursue PhD studies. He also supported my graduate applications and led me to my current happy life. Jingting Bai is a great actor who I've known since 2016. He has similar age to me and his attitude towards his acting career encouraged me a lot. Influenced by him, I'm more determined to work on science, focus on the questions that I'm excited about, take time to learn and improve myself step by step, and be brave to push my limits. In addition, I'm also more determined to be a kind and gentle person, always help others generously and forgive their

mistakes. Similarly, Jackson Wang in music, Yuzuru Hanyu in figure skating make me believe there are always people doing the correct things.

Finally, I want to give my thanks to the Chinese government. From a relatively poor place, I could get free but high-quality education until high school. Then I was admitted by Tsinghua University, the best college in China and provided various fellowships and international exchange-opportunities that widen my views. After my graduation, China Scholarship Council fellowship was a great support for my graduate life at Yale. In the future, I hope my knowledge could help more students like me.

Chapter 1: INTRODUCTION

Microbiota

Humans and other mammals are constitutively colonized by trillions of microorganisms that are also defined as the microbiota [1]. These microorganisms consisting of bacteria, fungi, virus and parasites colonize almost all mucosal and barrier surfaces including the skin, airways, vagina and the gastrointestinal (GI) tract [2]. As one of the largest barrier systems, GI tract is a complex ecosystem where various interactions (host-to-microbe, microbe-to-host and microbe-to-microbe) happen through intrinsic or extrinsic factors derived from the immune system, metabolic system, neuron system, diet, drugs and the microbiota [3]. Healthy human harbors over 10^{14} of bacterial cells in the gut, with diversity both along the longitudinal axis of the GI tract and in microhabitats, such as the lumen, mucus layers and crypts [4]. In addition, the gut microbiota also exhibits inter-individual diversity because of personalized nutrition, antibiotic utilization and location [5]. Overall, the physiological effects of the gut microbiota on the host are still largely unknown because of such huge diversity and dynamics.

With the development of the next-generation-sequencing (16S ribosomal RNA sequencing, metagenomics and transcriptomics) and anaerobic microbe culturomics, the gut microbiota has been observed to correlate with various human diseases. Locally, gastrointestinal inflammatory diseases, for example, inflammatory bowel disease (IBD), have been well described to be influenced by the gut microbiota [6, 7]. The mechanisms range from metabolite-mediated immune cells activation or differentiation, protein-protein interactions, to changes of microenvironments like hypoxia [8, 9]. Besides local effects, more recent studies have revealed the roles of the gut microbiota in remote organs and systemic diseases, including neurological disorders (Alzheimer's diseases, Parkinson diseases, multiple sclerosis, IBS), cardiovascular diseases, allergy or autoimmune diseases, cancer, respiratory diseases and so on [10]. Although the correlation relationship has been widely reported, the gut microbiota still needs more functional investigations to evaluate the causative relationships at molecular or cellular levels and provide instructive directions for future translational applications.

Microbe-to-Host Interactions in Immunity

The best understanding of microbe-to-host interactions is mediated by the innate immune system through germ line-encoded pattern recognition receptors (PRRs) that detect pathogen-associated molecular patterns (PAMPs) [11]. Lipopolysaccharide (LPS) from various Gram-negative bacteria activates Toll-like receptor 4 (TLR4) expressed by the intestinal epithelial cells and induces production of antimicrobial peptide pro-inflammatory cytokines [12]. Capsular polysaccharide A (PSA) from *Bacteroides fragilis* can interact with TLR2 on dendritic cells (DCs) and then induce regulatory T cells (Tregs) [13, 14]. However, such interactions are very limited because of the conserved repertoire of PRRs and PAMPs. The gut microbiota consists of hundreds of species, encoding about 150 times more genes than the human genome and even more proteins or metabolites that could serve signals for microbe-to-host interactions [1]. Furthermore, the heterogeneity of mucosal cells is much larger than other organs, so the existence of various immune cells, endocrine cells, absorptive cells, stem cells, neurons within the same ecosystem constructs more complex interactive networks.

In the past decades, how the microbiota regulates mucosal adaptive immunity has also been aggressively investigated both in homeostasis and diseases. Secretory immunoglobulin A (sIgA) production makes the gut special compared to other immune organs. As a major neutralizing antibody subset, sIgA was thought to just protect the host from bacteria invasion [15]. However, recent studies have found that the gut microbiota can utilize sIgA to acquire better colonization ability. Both bacteria like *B. fragilis* [16] and fungi *Candida albicans* [17, 18] have been described to gain commensalism in the gut through decreasing the virulence by coated with sIgA. Therefore, sIgA plays a critical role in microbe-host co-evolution. In addition, IgA-seq revealed that the gut microbiota exhibited a variety of IgA-coating and IgA+ population exacerbated colitis than IgA- population [19]. The systemic benefits of intestinal IgA plasma cells were described in multiple sclerosis that these commensal-reactive cells circulated and populated in CNS and then attenuated inflammation [20, 21].

T cells are also regulated by the gut microbiota, especially Th17/Treg balance in colitis [22]. Segmented filamentous bacteria (SFB) was the first commensal bacteria identified in the terminal ileum of C57BL/6 Mice from Taconic Farms but not Jackson Laboratory, leading to different Th17 abundance and inflammatory responses [23]. SFB-derived ATP and SFB-induced serum amyloid A protein (SAA) and reactive oxygen species (ROS) activate DCs to produce IL-1β, IL-6 and IL-23, promoting Th17 differentiation [24]. Other colitis-associated bacteria including *Citrobacter rodentium* and *Escherichia coli* O157 were also found to induce Th17 responses. Such induction could also depend on epithelial adhesion, mediated by intimin in *C. rodentium* for example [24]. The observation of SFB antigen-specific Th17 cells further supported the model that epithelial-adhesive commensal bacteria could be captured by intestinal DCs, then microbial specific antigens are presented to prime adaptive immunity [25, 26].

Besides the regulation mediated by direct adhesion, adaptive immunity is also shaped by microbial secreted metabolites, including short-chain fatty acids (SCFAs), bile acids and amino acid metabolites [27].

The regulatory effects of SCFAs have been widely described by different groups. The most common microbial SCFAs are acetate, propionate and butyrate, produced through fermentation of dietary fibers in the small intestine [28]. Treg cells express SCFA-receptor FFAR and produce anti-inflammatory cytokine IL-10 [29]. SCFAs can also activate inflammasome or inhibit histone deacetylases (HDACs) in the epithelial cells and enhance barrier functions [28, 30]. Beyond Th17/Treg axis, SCFAs also modulate type 2 immunity to decrease allergy or asthma [31], protect against pathogen infections through shifting macrophage metabolism [32], and cross the blood-brain barrier (BBB) to decrease production of inflammatory cytokines from microglia [33].

The major primary bile acids cholic acid (CA) and chenodeoxycholic acid (CDCA) are produced by host in the liver through oxidation of cholesterol and conjugated to taurine or glycine [27]. The gut microbiota converts conjugated bile acids delivered into the GI tract into secondary bile acids, such as deoxycholic acid (DCA) and lithocholic acid (LCA). Farnesoid X receptor (FXR) and Takeda G-protein receptor 5 (TGR5) can recognize bile acids and decrease inflammatory responses in chemical-induced colitis model [34]. DCA and LCA produced by *Clostridium*

scindens could also protect from *Clostridium difficile* infection [35]. While some conjugated bile acids have cytotoxicity on effector T cells, some bile acids such as 3-oxoLCA and isoalloLCA exhibited protective functions. 3-OxoLCA directly binds with retinoid-related orphan receptor-γt (RORγt) to inhibit Th17 differentiation and isoalloLCA produce mitochondrial reactive oxygen species (mitoROS) to increase FOXP3 expression and Treg differentiation [36].

The best understandings about amino acid metabolites in immunity are tryptophan metabolites that are increased in high-protein diets, including indoles, tryptamine, serotonin and kynurenine [37]. *Lactobacillus* species are the best-appreciated commensals that produce various indoles. Indoles such as such as indole-3-pyruvic acid (IPA) or indole-3-aldehyde (I3A) decrease both mucosal and systemic inflammation as the aryl hydrocarbon receptor (AhR) ligands, promoting Treg differentiation and IL-22 production by innate lymphoid cells (ILCs) [38, 39]. Indoles can also protect the host from infections through colonization resistance or direct downregulation of virulent factors and toxins [40]. However, AhR activation in regulating tumors is controversial: it can enhance tumor malignancy and progression in chronic lymphocytic leukemia (CLL) [41], while prevent or halt tumorigenesis in intestinal inflammatory tumor model [42]. Such results might be due to broad expression or different functions of AhR in various cell types or different tumor types.

In summary, microbe-to-host interactions significantly regulate both mucosal and systemic immune responses, either through direct contact or secreted metabolites.

Colorectal Cancer

Colorectal cancer (CRC) is the third commonly diagnosed and the second leading cancer death globally [43]. Generally, the development of the society tends to be followed by a rise in CRC incidence rates because of a western diet, alcohol drinking, smoking and sedentary lifestyle [43]. Although genetic factors such as germline mutations of MLH1 and APC contribute to CRC, most CRC cases are sporadic without any family history or inherited genetic mutations and attributable to various environmental risks [44]. Statistically, the heritability of CRC is 35–40% and only 5% are hereditary cancer syndromes such as Lynch syndrome (or hereditary nonpolyposis colorectal

cancer, HNPCC) or familial adenomatous polyposis (FAP) [43]. The emerging trend of CRC rising incidence at younger ages further highlights the importance of understanding CRC drivers and designing more prevention methods and therapeutics.

In general, the natural history of CRC follows four major stages: initiation, promotion, progression and metastasis [45]. Irreversible genetic damage initiates the CRC rising and the predisposing genetic mutations direct subsequent neoplastic transformation. In the promotion stage, cells exhibit abnormal growth and proliferation, forming neoplasm or polyps. In the progression phase, the abnormal cells further acquire genetic or epigenetic alterations that provide selective growth advantages and metastatic potential. In the metastasis stage, those aggressive tumor cells could spread from the primary organ to other organs through bloodstream or lymphatic vessels. The duration of each phase has wide ranges, and it usually takes decades to complete the four stages into CRC [43].

CRC have three major but not mutually exclusive kinds of genetic and epigenetic aberrations. First, chromosomal instability (CIN) means abnormalities in chromosomal copy number such as aneuploidy and polyploidy and structure that could be caused by mitotic errors. Second, CpG island methylator phenotype (CIMP) means epigenetic hypermethylation at repetitive CG dinucleotides of promoters of tumor suppressor genes. Third, microsatellite instability (MSI) means alterations in the length of short nucleotide tandem repeats that are driven by abnormal DNA mismatch-repair responses [43, 46].

CRC is also an aetiologically heterogenous disease, with distinct subtypes based on tumor anatomical location or global molecular alterations. Basically, CRC develops through three distinctive pathways. Adenoma–carcinoma sequence is a classic pathway for major sporadic CRC, accounting for 85-90% sporadic CRC [47]. Normal cells progress to small adenoma, then large adenoma and finally cancer because of gradual stepwise accumulation of mutations on *Apc*, *Kras* and then *Tp53* [48]. This pathway is predominantly associated with the CIN alterations [43]. Serrated pathway, accounting for 10-15% sporadic CRC, also develops polyp precursors to cancer [47]. Different from adenomatous polyps, normal cells from serrated pathway transit to hyperplastic polyp because of *Braf* mutation, then serrated adenomas and finally cancer because

of CIMP [49]. The arising of adenomas is also claimed associated with the transformation of intestinal stem cells into cancer stem cells at intestinal crypts [50]. Less than 2% sporadic CRC is driven by inflammatory pathway. IBD patients, particularly ulcerative colitis patients, exhibited 2.4-fold higher risk of CRC. Driven by chronic inflammation, normal cells transform into indefinite dysplasia, low-grade dysplasia, then high-grade dysplasia and finally cancer. Dysplasia in this pathway is present in flat mucosa and exhibits multifocal and obscuring lesions, different from discrete adenomas in the other two pathways [43].

The heterogeneity of CRC is traditionally defined based on tumor anatomical sites: proximal colon, distal colon and rectum [43]. Such differences actually correlate with the variety of microbial or host characteristics in colon. From proximal colon to rectum, there is a progressive increase in pH, decrease in oxygen level, increase of microbial loads, diversity, and metabolite abundance, indicating that diverse factors promote tumorigenesis at distinct locations [4].

Microbes and Cancer

Microbes have been observed to promote, diminish or have no effect on various types of cancer. A well-known example of microbe-inducing tumorigenesis is chronic infection induced by *Helicobacter pylori* (*H. pylori*) that may cause stomach cancer [51]. In 2021, the US Department of Health and Human Services (HHS) added it to the 15th Report on Carcinogens. The attempt to apply microbes to cancer treatments dates back to the late 19th century when Coley's toxins was developed as a toxin mixture filtered from killed *Streptococcus pyogenes* and *Serratia marcescens* [51].

Recently, more studies indicated the significance of the gut microbiota in cancer immunotherapy. Particularly, different groups found that melanoma patients exhibited various responses to the checkpoint immunotherapy. The gut microbiota was then highlighted because fecal microbiota transplantation (FMT) from responders could facilitate tumor regression in non-responder patients, due to the increase of intratumoral CD8 T cells [52, 53]. Similarly, the effects of the gut microbiota on systemic anti-tumor response were also observed in mice, through shaping the immune responses in tumor microenvironment [54]. *Lactobacillus acidophilus* can prime myeloid

cells to produce ROS, leading to DNA damage and tumor cell death [55]. *Lactobacillus johnsonii* and *Enterococcus hirae* induced peripheral Th17response [56, 57]. *Bifidobacterium* can facilitate anti-PD-L1 efficacy by expanding intratumoral CD8+ T cells [58]. Therefore, the gut microbiota has been considered to combine with and facilitate chemotherapy, radiotherapy or immunotherapy in personalized cancer therapies.

Microbes could also have tumor-initiation or tumor-promotion effects, mediated by tumor-associated microbial species (microbial cancer markers) that have been recently defined based on multiple metagenomic analysis of human cancer patient cohorts [59, 60]. Generally, the gut microbiota contributes to CRC through three mechanisms [51]: First, bacteria-induced mucosal barrier breach or microbial metabolites can activate inflammatory pathways, shaping the intestine or tumor microenvironment. Second, direct contact from microbes activates dysregulated signaling pathways in the intestinal epithelial cells. Third, secreted microbial products, either small molecules or large proteins, enable to directly cause DNA damage of intestinal epithelial cells, a key driver of genetic or epigenetic alterations and tumorigenesis. These three mechanisms are not mutually exclusive because microbes can function as a whole organism.

Both high-grade (e.g., IBD) and lower-grade (e.g., high-fat diet or obesity) inflammation influenced by the gut microbiota drive a tumor-promotion milieu. Inflammatory factors include reactive oxygen and nitrogen species (ROS and RNS), cytokines (e.g., IL-6 and IL-17), and chemokines (e.g., CXCL12 and CCL20). Immune cells that participate the shape of tumor microenvironment include CD8 T cells, Th17 cells, Th1 cells, Treg cells, macrophages and NKT cells [61]. Recently, the critical roles of *Fusobacterium nucleatum* and enterotoxigenic *Bacteriodes fragilis* (ETBF) in promoting Th17 responses in CRC have been well indicated with multiple mouse models including Apc^{Min} mice, AOM/DSS CRC model and IL-17 knockout mice [62-65]. In addition, some immune-modulating factors also have direct effects on epithelial or tumor cells. For example, GPR43 expression, one SCFA receptor, decreased in cancerous versus healthy tissues in human CRC patients. And restoration of GPR43 expression could increase apoptosis of human colon cancer cells upon SCFA exposure [30]. The inflammatory environment induced by the gut microbiota could in turn support the growth of pathogenic

microbes or opportunistic pathobionts. For example, environmental and nutritional changes caused by inflammation confer a growth advantage to facultative bacteria Enterobacteriaceae, and blooms of adherent-invasive *E. coli* (AIEC), is found in IBD and CRC patients [66].

The best understanding of protein-protein interactions between CRC-associated microbes with the host intestinal epithelial cells is the binding of FadA adhesin from *F. nucleatum* with E-cadherin. This direct contact can activate downstream Wnt/β-catenin signaling pathway, leading to abnormal cell proliferation [67]. In addition, Fap2 from *F. nucleatum* could bind with TIGIT, a significant immune checkpoint, leading to antitumor immune evasion or immunotherapy resistance [68]. Besides the membrane protein interactions, direct contact is also necessary for delivery of bacterial intracellular products into the host epithelial cells through various secretion systems.

Secreted microbial products form a large pool of 'dark matter' because of mostly unknown chemical structures and functions. Bacterial genotoxins are defined because of the DNA-damaging ability, further leading to senescence or apoptosis if DNA is not properly repaired [69]. Commonly known genotoxins include cytolethal distending toxin (CDT) produced by Gram-negative bacteria, typhoid toxin produced by *Salmonella enterica* serovars, and colibactin produced by the phylogenetic group B2 *E. coli*. Large-protein CDT and typhoid toxin promote DNA single (SSB) and double strand break (DSB), harboring an active subunit CdtB that is functionally and structurally homologous to DNAse I [69]. The carcinogenesis induced by CDT was shown that wild type but not *cdtB* mutant *Campylobacter jejuni* strain promoted cancer in the Apc^{Min}/DSS model [70]. Similarly, *Salmonella Typhi* is associated with higher risk of hepatobiliary carcinoma in human [69].

Colibactin is a well-studied small molecule genotoxin. In 2006, genotoxic *E. coli* was firstly discovered to induce megalocytosis in mammalian cells [71]. The potential genotoxic molecule was named 'colibactin' and identified to be encoded by a 54-kb *pks/clb* genomic locus. This biosynthetic gene cluster (BGC) is widely distributed in the B2 phylogenic group that comprises commensals and extraintestinal pathogenic *E. coli* strains [71]. Then the same group found that *pks/clb* genomic locus was also distributed in other phylogenetically related species, including

Klebsiella pneumoniae, *Enterobacter aerogenes*, and *Citrobacter koseri* [72]. The carcinogenesis induced by *clb*+ *E. coli* was revealed *in vitro* by γ-H2AX induction, *tk* and *hprt* gene mutation assay and soft agar transformation assay [73]. From 2012 to 2018, several groups independently indicated that colibactin-producing *E. coli* can promote CRC in mice with different models [74, 75]. In addition, it's observed that *pks/clb* island is widely distributed and significantly enriched in IBD and CRC patients [65]. Recently, the structure of colibactin and the DNA adducts were elucidated, indicating colibactin induced DNA damage through cross-linking DNA double strands [76, 77]. Furthermore, the tumor-promoting effects of colibactin was supported by the results that *clb*+ *E. coli*-induced mutational signature can be detected in human CRC genomes [78]. The colibactin example provides robust proof-of-concept for the importance of microbiota-derived small molecule genotoxins in CRC. However, given the remarkable complexity and diversity of metabolites produced by human gut commensals, we hypothesized that novel microbiota-derived genotoxins might remain undiscovered.

Functional Somatic Syndromes

Functional somatic syndromes (FSS) are defined mostly based on symptoms, such as unpleasant pain, allodynia or hyperalgesia [79]. FSS patients usually have explicit and highly elaborated self-reports, and the symptoms are lack of standard treatment because of complicity and multi-organ syndromes. Some common types of FSS include irritable bowel syndrome (IBS), fibromyalgia/chronic widespread pain (FM/CWP), and multiple chemical sensitivity (MCS) [80]. The most important risk factors include female sex, age, infections and stress. Recently, the gut microbiota has been proposed as one of key contributors to FSS, especially IBS and FM, with the findings that dysbiosis is significantly associated with the pain syndromes [81, 82].

Prevalence of IBS is about 11% globally, characterized by abdominal pain or discomfort, stool irregularities and bloating, but lack of clear structural or biochemical abnormalities [83]. A major part of IBS patients also exhibit other somatic, visceral pain disorders or psychiatric comorbidities, such as migraine, depression and anxiety. IBS have different subtypes: IBS-C (IBS with constipation); IBS-D (IBS with diarrhea), IBS-M (mixed IBS) and IBS-U (unsubtyped IBS) [82].

Similar to other FSS, IBS diagnosis is mainly based on self-report symptoms and exclusion of other diseases, and the therapeutic methods are range from drug treatment to psychotherapy. Unfortunately, there is still no treatment that could cure IBS, but just relieve the symptoms. Therefore, it is critical to establish the link between molecular or cellular pathologies with IBS to develop better diagnostic markers or therapeutic methods.

IBS and Gut Microbiota

The potential mechanisms of IBS have been investigated rapidly recently, with growing evidence reveals that epithelial permeability, immune activation, neuronal mediators and the gut microbiota dysbiosis could regulate multiple intestinal sensorimotor functions, including the hypothalamus–pituitary–adrenal (HPA) axis, the enteric nervous system (ENS) and the extrinsic nervous system (gut-brain axis) [82]. In addition, psychological factors such as depression, anxiety and stress, could also in turn influence bowel functions like motility [82].

Generally, post-infectious IBS and IBS-D have been found with increased epithelial permeability, resulting from abnormal expression of tight junction proteins, such as occludin and zonula occludens protein 1 [84]. ‘Leaky’ gut was observed through confocal laser endomicroscopy of IBS patients, probably caused by bacteria or chemical induced low-grade inflammation [85]. In addition, increased bile acids also enable to accelerate colonic transit, leading to diarrhea and visceral hypersensitivity in IBS-D [86].

The observation that about one-third of IBD patients exhibit IBS-like symptoms in remission indicates the association between inflammatory immune responses with IBS [87]. Particularly, IBS patients have more activated mast cells in the degranulating state, secreting more proteases, histamine and polyunsaturated fatty acids [88]. Interestingly, the immune activation is coupled with responses to microbial pathogens, suggesting the roles of the gut microbiota.

Neuroimmune interactions is an essential part in IBS as the bridge of pathologies and neuronal pain sensation. Histamine and proteases derived from the mast cells, and serotonin derived from the enteroendocrine cells regulate ENS sensitivity in IBS [89]. Some bacterial-derived serine and cysteine proteases may also act on intestinal nociceptors to induce pain sensation to the brain

[82]. In addition, some immune mediators were shown to affect the neuronal structure in the gut. For example, nerve growth factor (NGF) primarily derived from the mast cells can promote neurite growth [90].

As a complex ecosystem, the gut microbiota has been described as a key contributor to the human physiology. Similar in IBS, the change of gut microbiota composition is promising [91]. Fecal microbiota transplantation (FMT) from IBS patients to gnotobiotic mice can induce visceral hypersensitivity, change intestinal transit time and epithelial permeability [92]. At taxonomic level, Firmicutes to Bacteroidetes ratios increased in IBS patients. Two groups of Clostridiales were shown depleted in IBS, and *Ruminococcus torques* was enriched [93]. The mechanisms of microbiota regulating IBS include immune modulation through metabolites like SCFA, proteases and toxins, PAMPs like polysaccharides. PAMPs could also be directly detected by neurons through receptors like TLRs, and DAMPs such as adenosine and ATP produced by epithelial cells after bacterial infection or stimulation are also ligands of neurons [82, 94]. In addition, spore-forming commensals induced enterochromaffin cells to produce 90% serotonin, stimulating the intestinal sensory neurons [95, 96]. Reciprocally, activated neurons could influence epithelial or immune cells through neuropeptides such as CGRP and substance P, potentially further shaping the gut microbiota composition [97, 98].

Nociceptive Neurons

Sensory neurons construct a major system that send the external signals to the brain and then direct us to react properly to various stimuli. Traditional sensations consist of smell, taste, touch, vison and hearing, and different types of neurons are specialized to detect odors, chemicals, force, temperature, light and sound to evoke perceptions or motor responses. Particularly, pain is initiated by the noxious stimuli induced activation of nociceptors that consist of both Aδ fiber axons and C fiber axons [99]. Aδ fiber axons are usually stimulated by mechanical stimuli. These myelinated and large diameter neurons allow a fast action potential to travel at a rate of about 20 meters/second towards the CNS characterized. Nonmyelinated, small diameter, slow-

conducting C fiber axons are usually stimulated by chemical stimuli, and both of them respond to thermal stimuli [100].

The activation of nociceptors largely depends on the membrane sensors at peripheral nerve terminals, such as transient receptor potential (TRP) channels and G protein- coupled receptors (GPCRs) [99]. TRPV1-4 mediate warm-hot range, TRPV1 can also mediate chemical sensation, such as capsaicin and spider toxin, TRPA mediates chemical sensation and TRPM8 mediate cold sensation [99]. GPCRs sense specific ligands such as bradykinin, histamine and cytokines to stimulate sensory neurons [99]. Increasing recent evidence highlight that nociceptor neurons are remarkably diverse, with various molecular expression patterns [101]. Therefore, further classification of nociceptors based on molecular patterns may contribute to the understandings of distinct responses or identification of specialized neural circuits for diverse contexts of nociceptive sensation.

Upon sensing noxious stimuli, action potentials are generated at nociceptor terminals that are transduced to the dorsal root ganglia (DRG) of the spinal cord and relayed to the brain to be processed and perceived as pain [99]. Although it's clear how skin nociceptors respond to stimuli, initiate pain sensation and avoid behaviors, visceral pain sensation like IBS still seems ambiguous.

TRPV1 and Nociception

The studies about ligands or chemical stimuli that enable to elicit pain started with natural products. Well-known examples of natural noxious compounds include capsaicin from hot chilli peppers, cooling agent menthol from mint, isothiocyanates from wasabi and thiosulfinates from garlic [99]. Remarkably, these structurally different irritants induce pain through targeting specific TRP channels expressed by distinct nociceptor subtypes: capsaicin is a ligand of TRPV1; menthol is a ligand of TRPM8; isothiocyanates and thiosulfinates are ligands of TRPA1 [99]. TRP channels exhibit similar overall structures that have cytoplasmic amino and carboxyl termini, and six transmembrane segments with a pore region between TM5 and TM6. However, they are different in ankyrin or TRP domains at the intracellular terminus [102]. Considering the diversity of

vertebrate TRP channels and their significant roles in various biological processes and diseases, the understandings of natural products or endogenous ligands of TRP channels and their respective functions on stimulating nociceptors would shed light on related clinical and therapeutic translations.

Firstly discovered in *Drosophila* eye, TRP channels are essential to stimulate photoreceptor cells after light-activating rhodopsin [103]. In this GPCR-operated TRP channel activation model, the role of phospholipase C (PLC) is revealed in hydrolyzing plasma membrane phosphatidyl-inositol-4,5-bisphosphate (PIP_2), leading to fully activation of TRP channels by second messengers and other downstream metabolites or mediators. Consistently, endogenous lipids are a major class of endogenous ligands of TRP channels through regulating the lipid-TRP interactions [99]. However, there is a broad range of TRP channel ligands based on their special structures that provide multi binding sites for diverse molecules.

The best understandings of TRP channel ligands and respective physiological roles are about TRPV1 and TRPV1-expressing nociceptors. TRPV1 was identified through screening of cDNA expression library of capsaicin-responsive sensory neurons [104]. The cloned receptor TRPV1 (VR1) was confirmed to be activated by capsaicin and noxious temperature, suggesting its function in pain sensation [104]. From an evolutionary angle, mammals are sensitive to capsaicin so the plants can protect from digestion, while birds can tolerate capsaicin and help seed dispersal. Therefore, through comparing rat versus chicken TRPV1, the cytoplasmic loop between TM2 and TM3 was proposed to be the putative vanilloid binding pocket for capsaicin and other vanilloid chemicals, which was also confirmed through mutation studies [105]. More following studies also revealed that other molecules could activate or sensitize TRPV1. Extreme acidosis (pH<6) activates TRPV1 while mild acid conditions (pH 6.5) just increase the TRPV1 sensitivity. The dual functions both contribute to the extracellular pore region between TM5 and TM6, but different sites [106]. Moreover, spider toxins are also TRPV1 agonists through binding with the extracellular pore region between TM5 and TM6 [107]. Similar to other TRP channels, the cytoplasmic termini and intracellular loops of TRPV1 is also regulated by PIP_2 and other lipids. Besides the direct activation on TRPV1 channel, signaling pathways mediated by GPCR, tyrosine

kinase and cytokine receptors also directly activate or sensitize TRPV1, through the regulation of PLC and downstream second messengers and mediators [99].

Functional studies revealed that TRP channel expression defines distinct nociceptor subtypes. Briefly, TRPV1+ nociceptors respond to noxious heat and mechanical stimuli, TRPM+ nociceptors respond to cold temperature, and TRPA1+ nociceptors, a subset of TRPV1+ nociceptors, are important in chemonociception besides heat hypersensitivity [99]. These nociceptors could be further classified based on molecular patterns, but they are not necessarily exclusive. Such classification just simplifies the manipulation of specific nociceptors in functional studies.

Overall TRPV1+ nociceptors account for about 30-50% somatosensory neurons, and the nonmyelinated, slowly conducting C fibers also express essential neuropeptides such as CGRP and substance P that could function on immune cells [108]. These fibers also express various receptors to sense immune modulators such as cytokines, chemokines, lipid mediators and growth factors or microbial compounds such as formyl peptides and toxins, suggesting a critical role in the neuroimmune axis [108, 109].

Proinflammatory cytokines such as TNF and IL-1β contribute to mechanical and thermal pain sensitivity in arthritis models through influencing TRPV1+ nociceptors [110]. Anti-inflammatory cytokine IL-10 could prevent pain sensitivity through decreasing both pro-algesic cytokine release and sodium-channel activation [111]. Release of lipid mediators such as prostaglandins and leukotrienes during inflammation contribute to pain sensitivity, while resolvins and maresins prevent analgesic actions, working on nociceptor nerve terminals [112]. Release of neurotrophic growth factors such as nerve growth factor (NGF) and brain-derived neurotrophic factor (BDNF) during inflammation contribute to nerve terminal sprout and signal transduction both in peripheral and CNS [112]. Some immune modulators like TNF and IL-6 could also increase the TRPV1 expression [109].

Although the neuroimmune axis are investigated more in the immune to neuron direction, increasing evidence have been shown recently that nociceptors can directly sense microbial compounds and shape the immune responses (neuron to immune direction) [109]. This area was

initiated by bacterial infection studies. Pathogen *Staphylococcus aureus* is Gram-positive bacteria leading to pyogenic and painful infections. The fact that acute pain dynamics and phenotypes correlate with bacterial load, but not immune responses suggests that TRPV1+ nociceptors are stimulated earlier [113]. It has been demonstrated that pore-forming toxins (PFT) derived from *S. aureus* such as α-haemolysin (αHL), γ-haemolysin AB and phenol soluble modulin α3 (PSMα3) could directly initiate calcium influx and action potentials through assembly into nociceptor membranes, increasing the pain sensitivity to mechanical and thermal stimuli [113, 114]. In addition, N-formyl peptides derived from *S. aureus* can indirectly sensitize TRPV1 through formyl peptide receptor 1 (FPR1) [113]. Similarly, PFT streptolysin S (SLS) derived from *Streptococcus pyogenes* also activate TRPV1+ nociceptors [115]. Microbial PAMPs LPS can activate TRPV1+ nociceptors through TLR4 expressed on the nerve terminals or TRPA1, and bacterial flagellin can activate a subset of TRPV1+ nociceptors with TLR5 [109]. Fungi cell wall component zymosan from *Candida albicans* also activate TRPV1+ nociceptors, although the receptor is still unclear [116].

Traditionally, pain was thought to contribute to host defense through avoiding behaviors to protect from further injury or damage. However, recent studies highlight that the stimulated nociceptors release neuropeptides to further modulate immune responses. In *S. aureus* and *S. pyogenes* infections, CGRP actually inhibited immune responses like neutrophil recruitment and activation to facilitate the bacterial survival but reduce inflammatory skin damage [113, 115]. While in *C. albicans*, CGRP facilitates anti-fungi immunity through Th17 responses [116]. CGRP [117-119], substance P, vasoactive intestinal peptide (VIP) [120] and pituitary adenylyl cyclase-activating polypeptide (PACAP) predominantly have anti-inflammation effects on macrophages or myeloid cells, although pro-inflammation effects are also observed in some contexts [109]. In addition, neuropeptides can also influence other immune cells including mast cells, dendritic cells, and innate lymphoid cells [121, 122], contributing to various diseases such as allergy [123], asthma, colitis, helminth infections. Taken together, these studies indicate a huge translational potential of manipulating nociceptor sensory neurons in various diseases.

Chapter 2

Human commensal bacteria produce novel genotoxic metabolites and exacerbate colorectal cancer

Overview

Colorectal cancer (CRC) is the third most common malignancy and the second leading cause of cancer death worldwide [43]. Two-thirds of all CRC cases occur in individuals without a family history of CRC or inherited CRC-predisposing genetic mutations [44]. Thus, environmental risk factors that promote the acquisition and accumulation of somatic genetic and epigenetic aberrations are chief contributors to CRC development. The gut microbiota can regulate intestinal tumorigenesis through diverse mechanisms [8, 51, 124] such as short-chain fatty acid-producing clostridia species that induce regulatory T cells as well as temper inflammation-induced tumorigenesis [28] and *Fusobacterium nucleatum* strains that enhance tumor growth by inducing epithelial proliferation through FadA-mediated engagement of E-cadherin and activation of Wnt/β-catenin signaling [67]. Microbial products can also trigger DNA modifications in intestinal epithelial cells [125]. For example, the 20 kDa *Bacteroides fragilis* toxin induces DNA damage through induction of reactive oxygen species (ROS) [126], while cytolethal distending toxin (CDT) from pathogenic Gram-negative bacteria has direct DNase activity [127].

Small molecule metabolites from the microbiome are also known to influence CRC risk by directly causing DNA damage. Select *Escherichia coli* strains produce the reactive small molecule genotoxin colibactin, which directly alkylates and crosslinks DNA, triggering double-stranded DNA breaks (DSBs) and facilitating intestinal tumorigenesis in mouse models [71, 73-75]. The colibactin biosynthetic machinery is encoded by a 54 kb hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) gene cluster referred to as the *pks* or *clb* locus [71], and the mature chemical structure of colibactin responsible for the pathway's DNA interstrand crosslinking activity was recently determined [77, 128]. Human CRC tumors also contain mutational

signatures consistent with colibactin-induced DNA damage, implicating colibactin in human CRC development [78, 129].

The colibactin paradigm illustrates the importance of microbiota metabolite-induced DNA damage in human CRC. However, aside from colibactin, the role of microbiota-derived small molecule genotoxins in CRC initiation or progression remains mostly unexplored. Given the enormous complexity and diversity of metabolites produced by bacteria [130], we hypothesized that diverse taxa from the human gut microbiome may produce previously undiscovered or unappreciated small molecules that cause DNA damage in intestinal epithelial cells and contribute to the development of CRC. Here, we established a pipeline to evaluate the genotoxicity of small molecule metabolites derived from over 100 phylogenetically diverse human gut microbes which were isolated from the microbiomes of inflammatory bowel disease (IBD) patients [19]. We identified a diverse set of microbes that produced genotoxic small molecule metabolites, including the Gram-positive bacteria *Clostridium perfringens* and *Clostridium ramosum*, and the CRC-associated Gram-negative species *Morganella morganii*. However, none of these isolates produced known genotoxins, such as colibactin, or encoded known genotoxin-producing biosynthetic gene clusters. We combined untargeted metabolomics and bioactivity-guided natural product discovery techniques, to isolate, characterize, and synthesize a family of previously undescribed genotoxic metabolites—termed the indolimines—from *M. morganii* under *in vitro* and *in vivo* culture conditions. We also confirmed that genotoxic *M. morganii* exacerbated intestinal tumorigenesis in a gnotobiotic mouse model of CRC. Through transposon-based random mutagenesis, we identified gene *aat* (encoding AAT_I protein, aspartate aminotransferase fold type I) responsible for the synthesis of primary amines, precursors of indolimines. The *att- M. morganii* exhibited defects of genotoxicity. These studies thus uncover an expanded universe of genotoxic gut microbes and metabolites that may critically influence CRC initiation and progression.

Results

Establishing a pipeline to identify genotoxic bacteria from the human gut microbiota

We established a pipeline to screen diverse human gut microbes based on their ability to directly damage DNA. We then applied this pipeline to a gut microbiota culture collection assembled by anaerobic culturomics of stool samples from 11 IBD patients [19], as IBD patients are considered to be at a significantly increased risk of developing CRC [131]. This collection consists of 122 unique bacterial isolates that span 5 phyla, 9 classes, 10 orders, and 18 families, as well as multiple strains that were assigned to the same species (**Fig. 1A**).

To probe for genotoxicity, we evaluated the activity of each isolate in a plasmid DNA damage assay. As microbiome metabolites such as colibactin can be recalcitrant to isolation [77], we focused our primary studies on co-incubation of individual bacterial isolates with linearized pUC19 plasmid DNA (**Fig. 1A**). This assay is based on the principle that the extent and modes of DNA damage can be assessed by electrophoresis under native and denaturing conditions [132, 133]: The intact linearized pUC19 DNA forms a clear band at ~2.7 kb under native conditions, while denaturation with increasing concentrations of NaOH reveals a ~1.3 kb band of single-stranded DNA. The formation of DNA interstrand cross-links (ICLs) prevents unwinding under denaturing conditions, thereby resulting in slower migration of duplexed DNA. Alkylation at many of the sites in DNA bases is known to decrease the stability of the glycoside bonds, resulting in deglycosylation and fragmentation, which is detectable as smaller fragments of higher mobility following electrophoresis [134]. Extensive DNA damage, for example, by DNase-mediated degradation, results in a loss of DNA even under native conditions. Finally, DNA damage caused by restriction enzyme-like molecules produces multiple bands under native conditions and even smaller fragments under denaturing conditions.

We confirmed that plasmid DNA was stable under diverse anaerobic cultivation conditions including incubation in Gifu medium, which supports the growth of all isolates in our collections (**Fig. S1**). To minimize the damage caused by bacterial DNases that are often produced in the stationary phase of bacterial growth, we measured growth curves for all 122 isolates in our collection and established a T_E (time point of exponential phase) and T_S (time point of stationary phase) for each isolate (**Table S1**). The isolates were then clustered into 7 groups that exhibited similar growth dynamics (**Fig. S2**). We selected two culture conditions for the initial screening:

anaerobic co-incubation with DNA to T_S; or anaerobic co-incubation with DNA to T_E which was followed by aerobic co-incubation to T_S to approximate the oxygen stress encountered in an inflammatory gut environment. Finally, we purified the linearized pUC19 DNA from the bacterial cultures via column purification and performed gel electrophoresis under native and gradient denaturing conditions (0 %, 0.2 %, 0.4 % and 1 % NaOH) (**Fig. 1A**, **Fig. S3A-G**).

We used the relative intensity reduction (RIR, %) of DNA after co-incubation with bacteria as a general measure of bacterially-induced DNA damage (**Fig. 1B**, **Table S2**). We found that diverse gut microbes exhibited DNA damaging activities, which suggests that microbiota-mediated genotoxicity may be more widespread than previously appreciated. While previously described microbiota-derived genotoxins discovered in a case-by-case manner are primarily produced by Gram-negative bacteria (e.g., *E. coli*, *B. fragilis and K. oxytoca*) [65, 135], we observed that multiple Gram-positive microbes from the phyla Actinobacteria and Firmicutes also caused significant DNA damage. DNA damaging activity was largely independent of culture conditions, although a few microbes displayed slightly varied genotoxicity in the presence or absence of oxygen stress (**Fig. 1B**, **Table S2**).

From the primary screening, we selected 24 bacterial isolates that exhibited strong DNA damaging activity and 18 phylogenetically-related non-genotoxic isolates for evaluation in a secondary screening (**Fig. 1A**). We re-established precise growth curves for each isolate (**Table S1**) and re-screened all 42 isolates under four distinct culture conditions, including co-incubation of DNA with bacterial supernatants collected from anaerobic cultures at T_S (**Fig. S3H-J**, **Table S2**). The vast majority of isolates that caused DNA damage in our primary screening also exhibited genotoxicity upon secondary screening (**Fig. 1C**). Moreover, for 18 of these isolates, the level of DNA damage induced by clarified bacterial supernatants was comparable to that derived by direct incubation with live bacterial cultures (**Fig. 1C**).

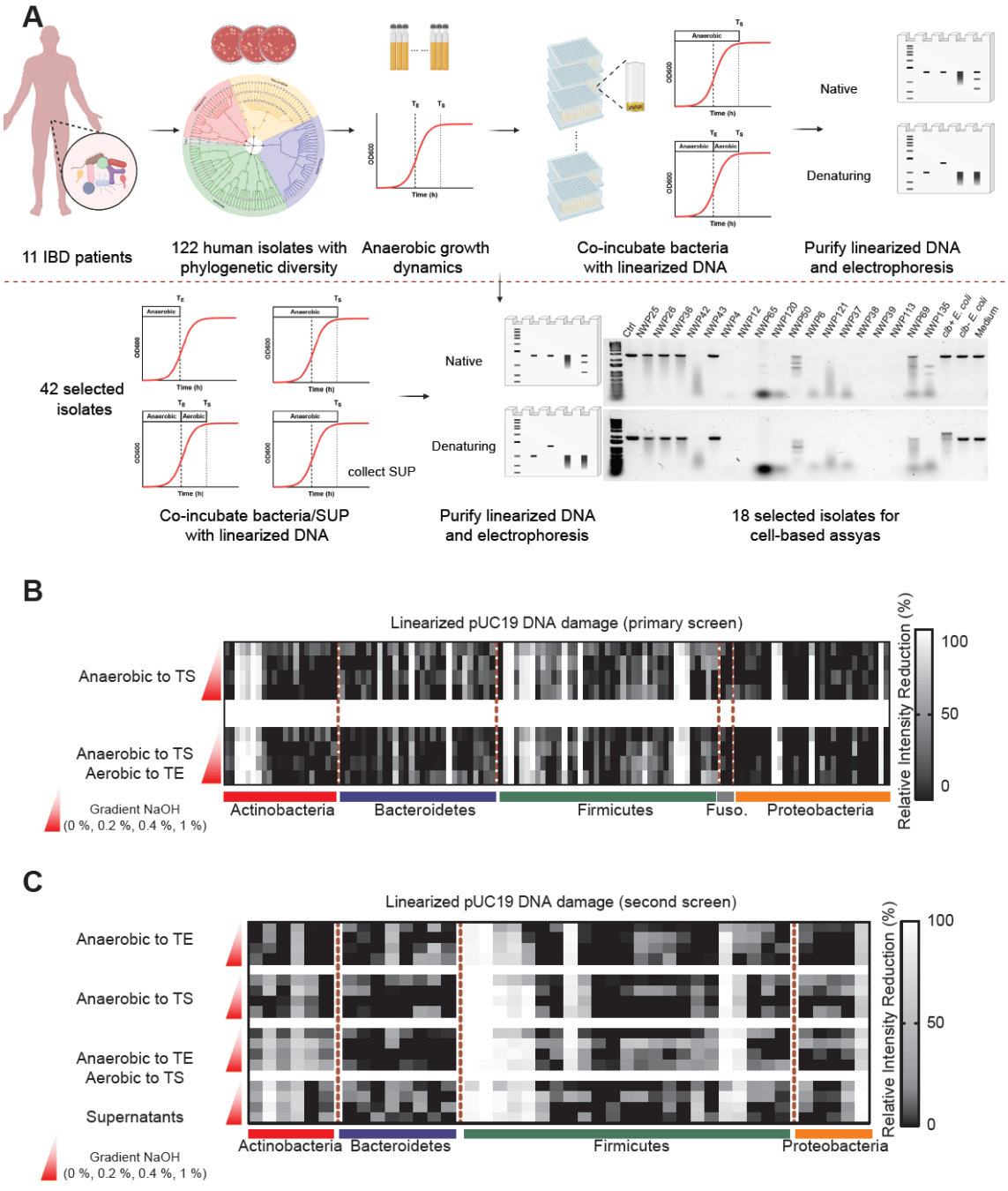


Figure 1. Establishing a pipeline to identify genotoxic bacteria from the human gut microbiota.

(A) Overview of functional screening of gut microbes for direct genotoxicity. 122 phylogenetically diverse bacterial isolates from 11 IBD patients (shaded based on phylum: Red, Actinobacteria; Blue, Bacteroidetes; Orange, Proteobacteria; Gray, Fusobacteria) were evaluated for genotoxicity

via co-incubation with plasmid DNA followed by gel electrophoresis. Bacterial growth curves for all isolates were determined via OD_{600} and individual isolates were co-cultured with linearized plasmid DNA under anaerobic conditions to stationary phase (T_S, light dashed line), or co-cultured under anaerobic conditions to exponential phase (T_E, bold dashed line), and then under aerobic conditions to stationary phase. After co-incubation, DNA damage was assessed via gel electrophoresis of purified plasmid DNA under native or denaturing conditions.

(B) Diverse human gut bacteria exhibit direct DNA damaging activities. Bacterial genotoxicity was determined by calculating the relative intensity reduction (RIR, %) of linearized pUC19 DNA bands after co-incubation with 122 diverse human gut bacteria (as outlined in **A**) as compared to medium only controls. pUC19 DNA was then purified via column purification and treated with or without gradient NaOH (0 %, 0.2 %, 0.4 %, 1 %) before evaluating DNA integrity via gel electrophoresis.

(C) Relative intensity reduction (RIR, %) of linearized pUC19 DNA bands in a secondary screening of 42 putative genotoxic and non-genotoxic isolates selected based on primary screening results (**B**). Linearized pUC19 DNA was co-incubated under indicated conditions, isolated via column purification after co-incubation and treated with or without NaOH (0 %, 0.2 %, 0.4 %, 1 %) before evaluating DNA integrity via gel electrophoresis.

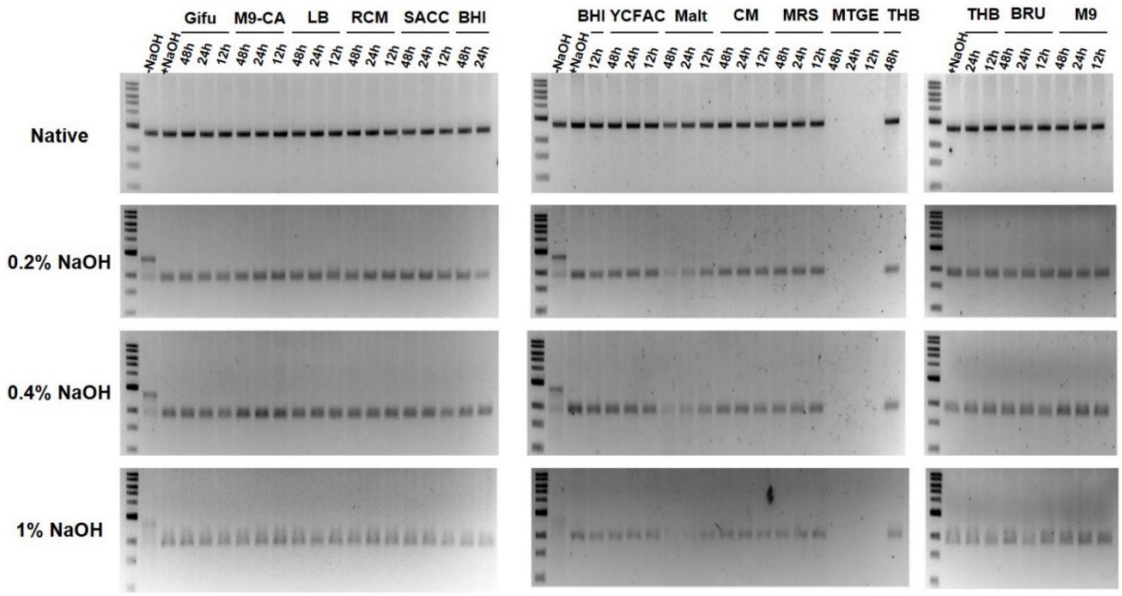
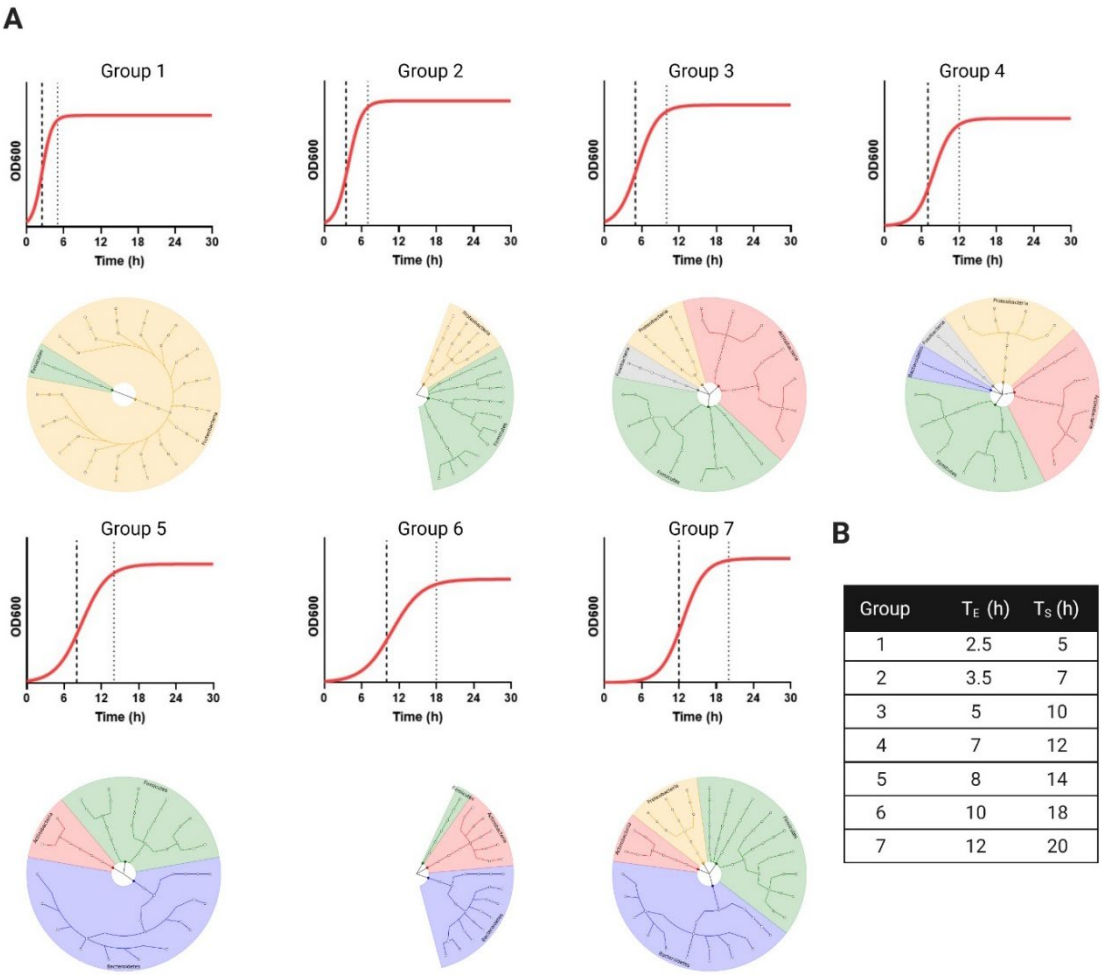


Figure S1. Assessment of plasmid DNA stability in diverse bacterial media.

Linearized pUC19 DNA was incubated in 14 different media for 48 h, 24 h, or 12 h under anaerobic conditions. DNA integrity was evaluated by gel electrophoresis under native or gradient denaturing conditions.

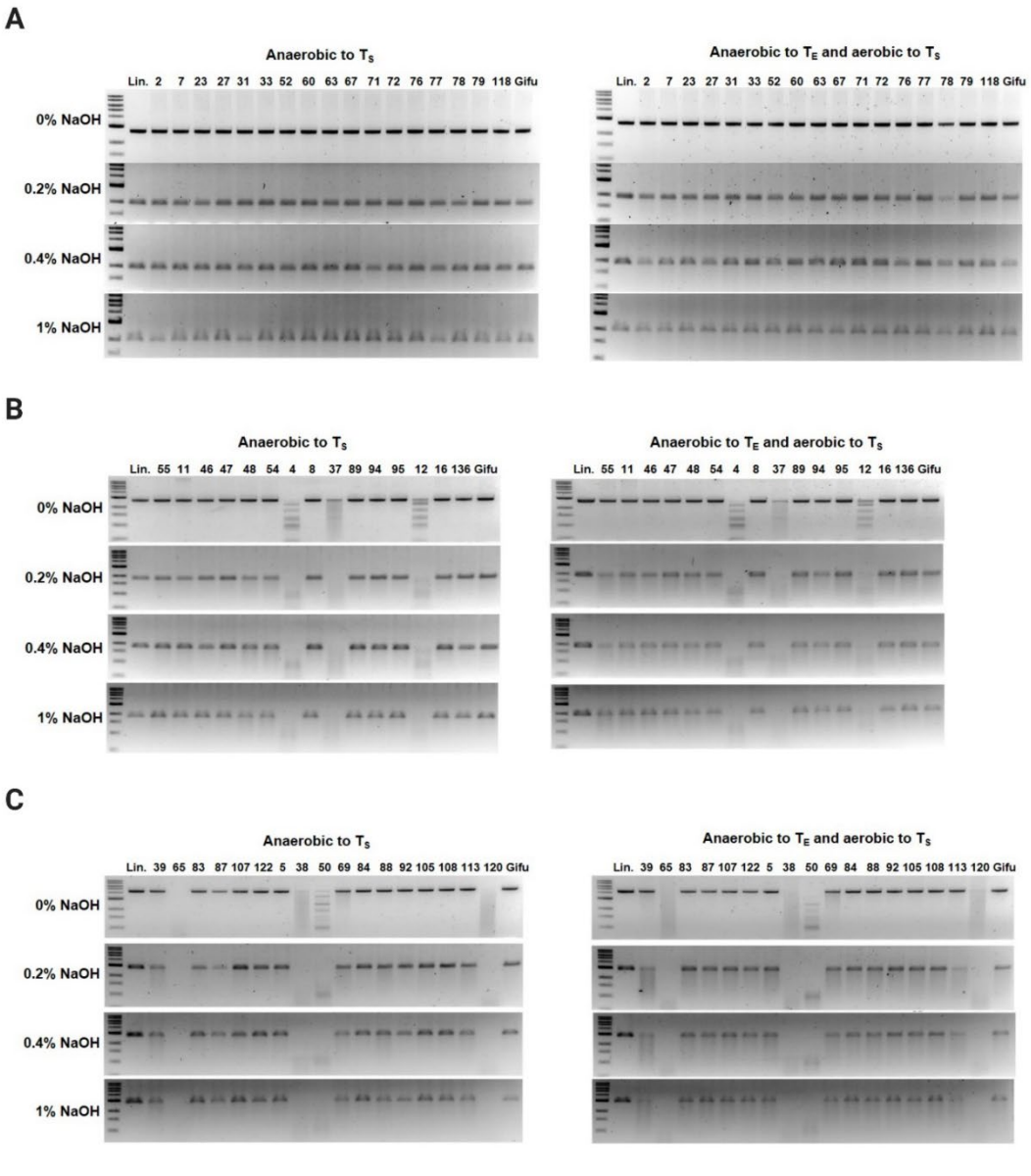


Group	T_E (h)	T_S (h)
1	2.5	5
2	3.5	7
3	5	10
4	7	12
5	8	14
6	10	18
7	12	20

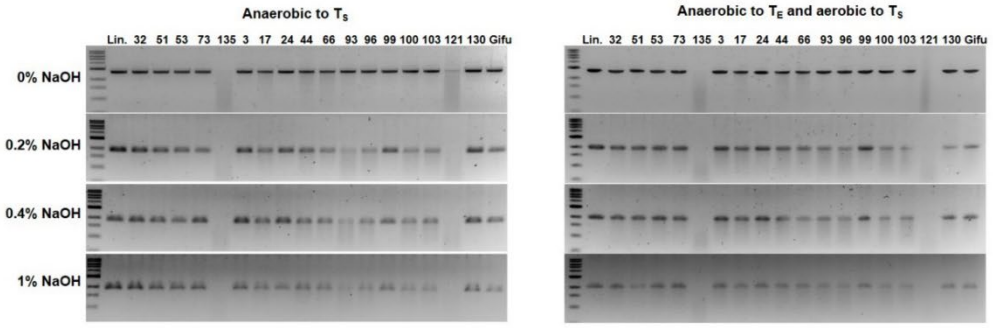
Figure S2. Clustering of human gut isolates into seven groups with similar growth dynamics.

(A) Representative growth curves (top graph) and respective phylogenetic compositions (shaded phylogenetic trees) for seven groups of bacterial strains with similar growth dynamics. Bold dashed line, T_E (time-point of exponential phase of bacterial growth); Light dashed line, T_S (time-point of stationary phase of bacterial growth). Phylogenetic trees are shaded based on phylogeny (phyla). Red, Actinobacteria; Blue, Bacteroidetes; Orange, Proteobacteria; Gray, Fusobacteria.

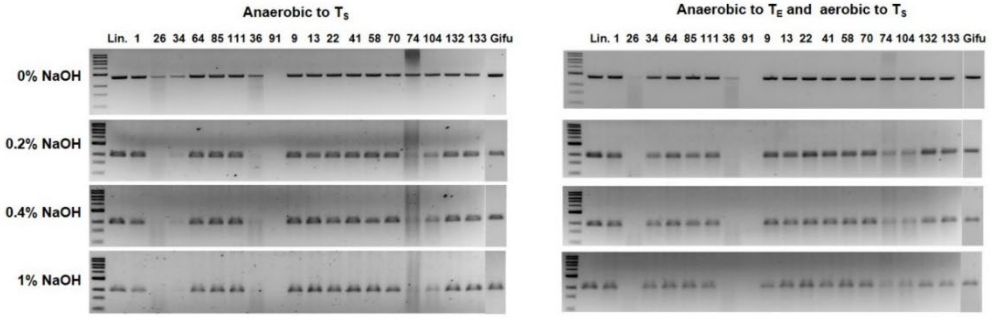
(B) Table summarizing average T_E and T_S for each group.


A
Anaerobic to T_S
Lin. 2 7 23 27 31 33 52 60 63 67 71 72 76 77 78 79 118 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 2 7 23 27 31 33 52 60 63 67 71 72 76 77 78 79 118 Gifu
B
Anaerobic to T_S
Lin. 55 11 46 47 48 54 4 8 37 89 94 95 12 16 136 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 55 11 46 47 48 54 4 8 37 89 94 95 12 16 136 Gifu
C
Anaerobic to T_S
Lin. 39 65 83 87 107 122 5 38 50 69 84 88 92 105 108 113 120 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 39 65 83 87 107 122 5 38 50 69 84 88 92 105 108 113 120 Gifu

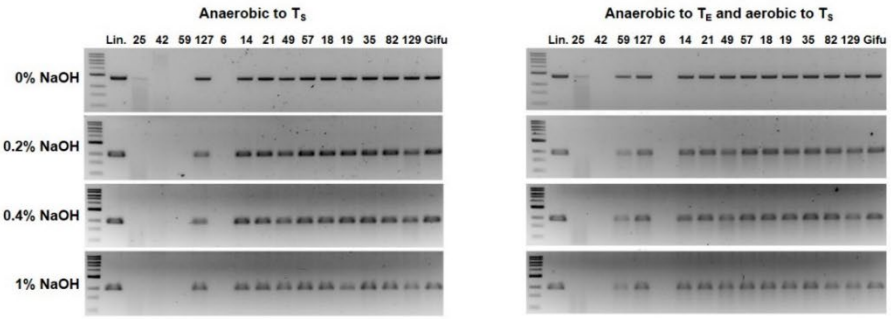
D


Anaerobic to T_S
Lin. 32 51 53 73 135 3 17 24 44 66 93 96 99 100 103 121 130 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 32 51 53 73 135 3 17 24 44 66 93 96 99 100 103 121 130 Gifu

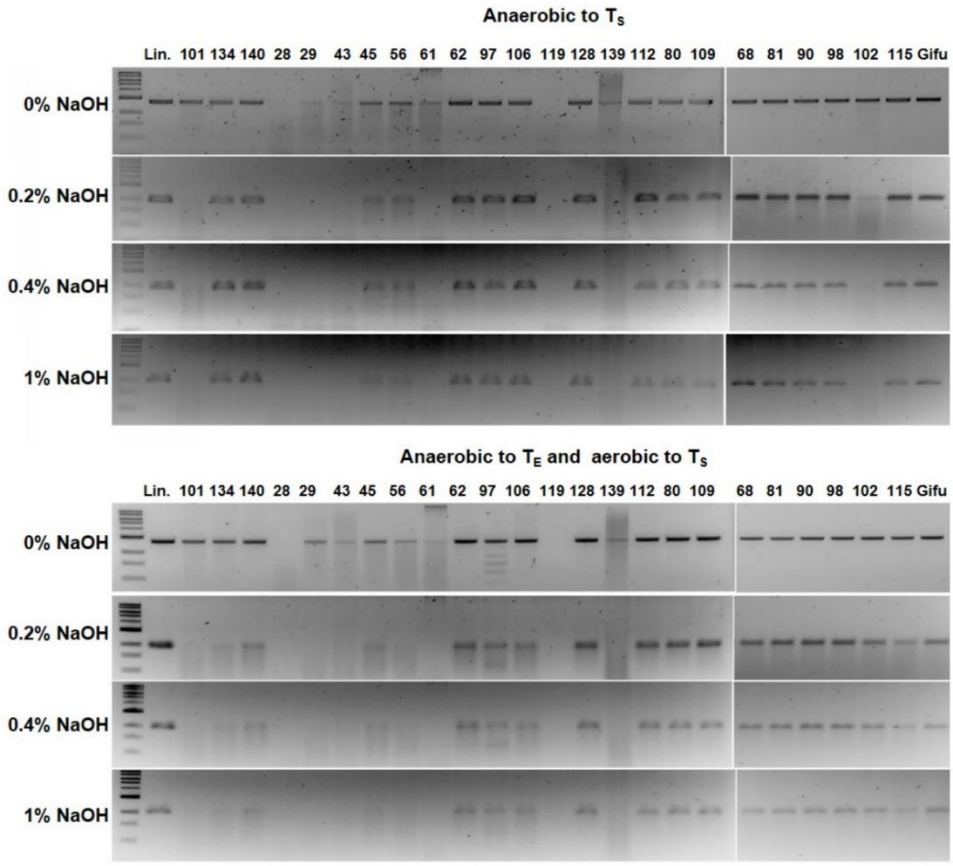
E


Anaerobic to T_S
Lin. 1 26 34 64 85 111 36 91 9 13 22 41 58 70 74 104 132 133 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 1 26 34 64 85 111 36 91 9 13 22 41 58 70 74 104 132 133 Gifu

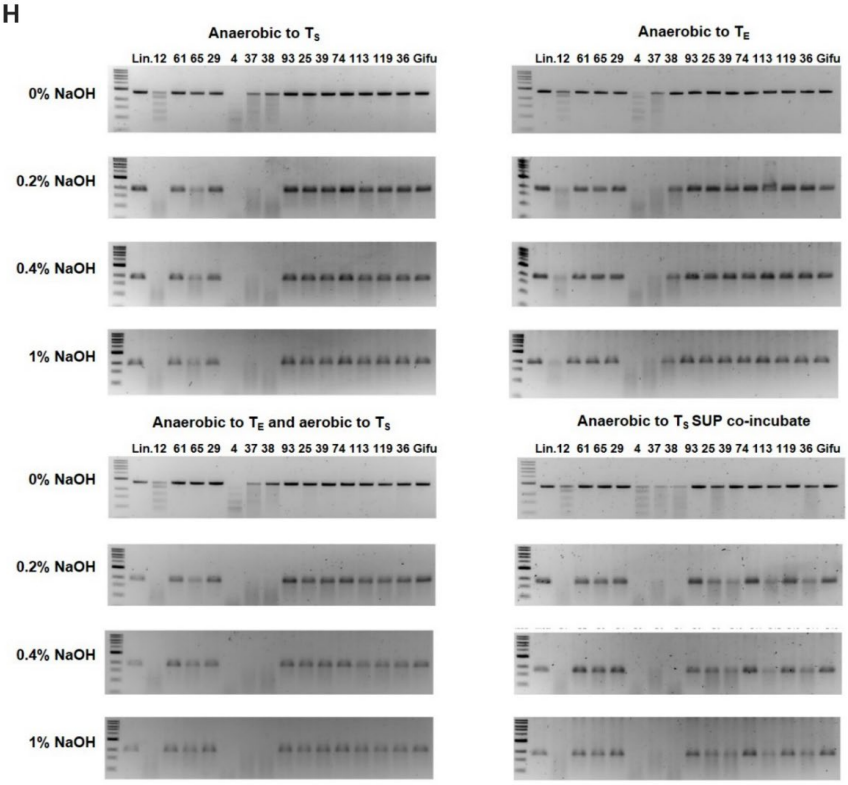
F


Anaerobic to T_S
Lin. 25 42 59 127 6 14 21 49 57 18 19 35 82 129 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 25 42 59 127 6 14 21 49 57 18 19 35 82 129 Gifu

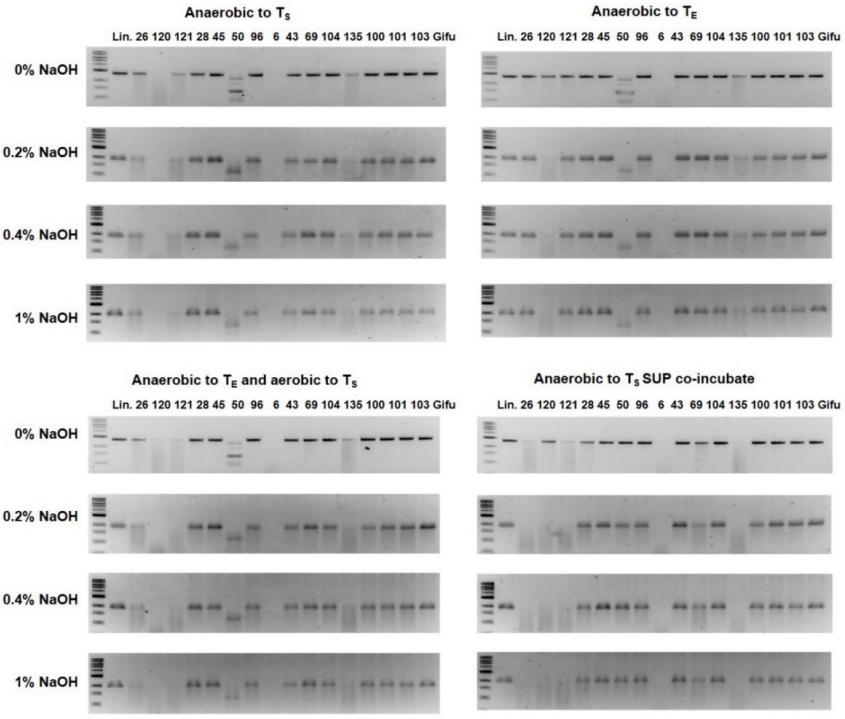
G


Anaerobic to T_S
Lin. 101 134 140 28 29 43 45 56 61 62 97 106 119 128 139 112 80 109 68 81 90 98 102 115 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E and aerobic to T_S
Lin. 101 134 140 28 29 43 45 56 61 62 97 106 119 128 139 112 80 109 68 81 90 98 102 115 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH

H


Anaerobic to T_S
Lin.12 61 65 29 4 37 38 93 25 39 74 113 119 36 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E
Lin.12 61 65 29 4 37 38 93 25 39 74 113 119 36 Gifu
Anaerobic to T_E and aerobic to T_S
Lin.12 61 65 29 4 37 38 93 25 39 74 113 119 36 Gifu
Anaerobic to T_S SUP co-incubate
Lin.12 61 65 29 4 37 38 93 25 39 74 113 119 36 Gifu

I


Anaerobic to T_S
Lin. 26 120 121 28 45 50 96 6 43 69 104 135 100 101 103 Gifu
0% NaOH
0.2% NaOH
0.4% NaOH
1% NaOH
Anaerobic to T_E
Lin. 26 120 121 28 45 50 96 6 43 69 104 135 100 101 103 Gifu
Anaerobic to T_E and aerobic to T_S
Lin. 26 120 121 28 45 50 96 6 43 69 104 135 100 101 103 Gifu
Anaerobic to T_S SUP co-incubate
Lin. 26 120 121 28 45 50 96 6 43 69 104 135 100 101 103 Gifu

J

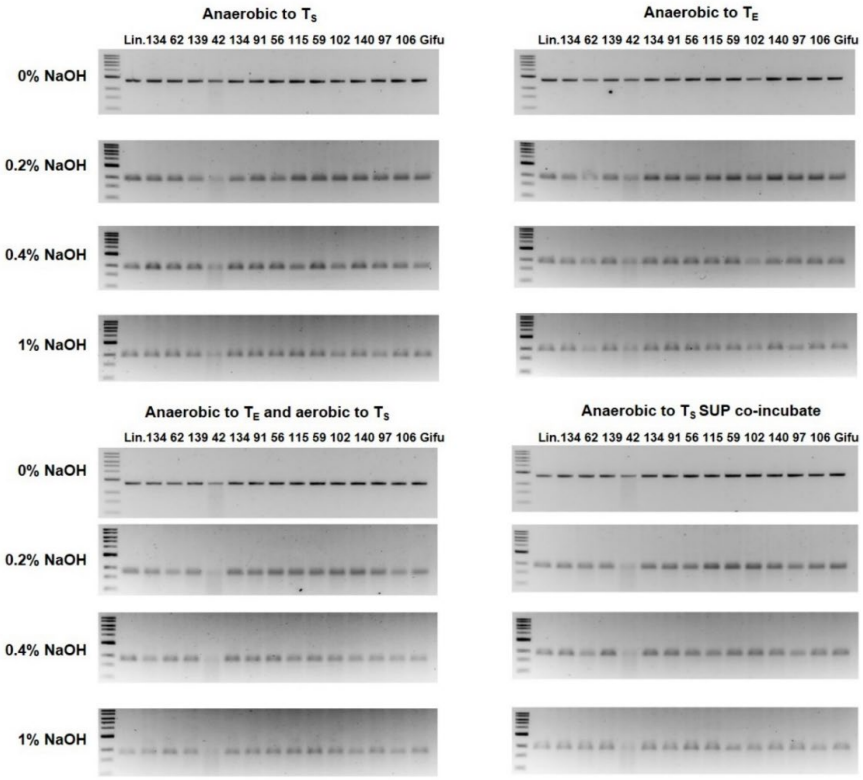


Figure S3. Native and denaturing DNA gel electrophoresis images from primary and secondary screening of DNA damage induced by diverse human gut microbes.

(A-G) Primary screening data. Linearized pUC19 DNA was co-incubated with 122 isolates under two culture conditions: co-incubation to T_S under anaerobic conditions or anaerobic co-incubation to T_E followed by aerobic co-incubation to T_S. pUC19 DNA was isolated via column purification after co-incubation and treated with or without NaOH (0 %, 0.2 %, 0.4 %, 1 %) before evaluating DNA integrity via gel electrophoresis.

(H-J) Secondary screening. Linearized pUC19 DNA was co-incubated with selected isolates (based on primary screening data) anaerobically to T_S or T_E, co-incubated anaerobically to T_E and then aerobically to T_S, or co-incubated with bacterial supernatants from isolates cultured anaerobically to T_S for 4 h. pUC19 DNA was isolated via column purification after co-incubation and treated with or without NaOH (0 %, 0.2 %, 0.4 %, 1 %) before evaluating DNA integrity via gel electrophoresis.

Small molecule metabolites produced by human gut microbes induced DNA damage

To determine whether the putative genotoxic bacterial isolates we identified via electrophoresis-based screening (**Fig.1**, **Fig. S4A**) produce genotoxic small molecules that cause DNA damage in human cells, we separated their supernatants (SUP) into small- (<3 kDa SUP) and large- (>3 kDa SUP) molecular weight fractions and evaluated the relative genotoxicity of these fractions in HeLa cells. Similar to cisplatin, a DNA-crosslinking chemical commonly used for cancer therapies but also inducing drug-resistance, small molecules from most isolates enhanced the expression of γ-H2AX, a marker of DNA double-strand breaks (DSBs) [136], in HeLa cells (**Fig. 2A**). However, these genotoxic isolates triggered different cellular responses. Particularly, all four isolates of *Clostridium perfringens*, one phylogenetically related isolate *Clostridium ramosum*, and the two isolates of *Morganella morganii* induced DSBs through small molecule metabolites, without affecting cell viability (**Fig. 2A**, **Fig. S4**). *Clostridium* species are well known to produce clostridial toxins inducing DNA damage (**Fig. S4B**) and cell death [137], indicated with cell size and granularity (**Fig. S4C**), or Annexin V and 7-AAD (**Fig. S4D-F**). However, it's unclear how the small molecule metabolites derived from *Clostridium* species contribute to DNA damage and their pathogenic virulence. In contrast, *Bifidobacterium adolescentis* and *Bifidobacterium dentium* induced cell death with small molecule metabolites, like cisplatin. Considering just one isolate showed γ-H2AX induction from *Bifidobacterium breve*, *Streptococcus mitis*, *Lactobacillus salivarius*, *Pediococcus acidilactici* and *Enterococcus asini*, we focused our subsequent studies on isolates from *C. perfringens*, *C. ramosum,* and *M. morganii*. Consistent with their ability to induce γ-H2AX (**Fig. 2B**), we found that small molecule metabolites from *C. perfringens*, *C. ramosum,* and *M. morganii* also induced cell cycle arrest in HeLa cells (**Fig. 2C**), further implicating these taxa as potential genotoxin producers.

Next, to enrich for genotoxic small molecules, we performed ethyl-acetate extractions using supernatants and found that ethyl-acetate extracts from *C. perfringens*, *C. ramosum*, *M. morganii*, and *clb*+ *E. coli* cultures nicked circular pUC19 plasmid DNA, while extracts from *clb*- *E. coli* or Gifu medium alone had negligible impacts on DNA integrity (**Fig. 2D**). Similarly, ethyl-acetate extracts from genotoxic species induced γ-H2AX expression (**Fig. 2E-F, Fig. S4G**) and tailing in

an alkaline comet unwinding assay that measures DNA DSBs at the single-cell level (**Fig. 2G-H**) in HeLa cells. These data suggested the potential of genotoxin isolation and identification from bacterial supernatants.

Recent meta-analyses have identified cross-cohort microbial signatures associated with CRC. Consistent with the findings that *Clostridiaceae*, *Erysipelotrichaceae*, and *M. morganii* are prevalent members of the CRC-associated microbiome [59, 60], we found that *C. perfringens*, *C. ramosum,* and *M. morganii* also have higher relative abundance in IBD patients (UC and CD patients) than healthy controls (noIBD) in the publicly available Human Microbiome Project database (**Fig. 2I-K**) [5]. Particularly, *M. morganii* exhibited significant higher prevalence and abundance than *C. perfringens* or *C. ramosum* (**Fig. 2I-K**). Therefore, we focused the subsequent studies on *M. morganii*.

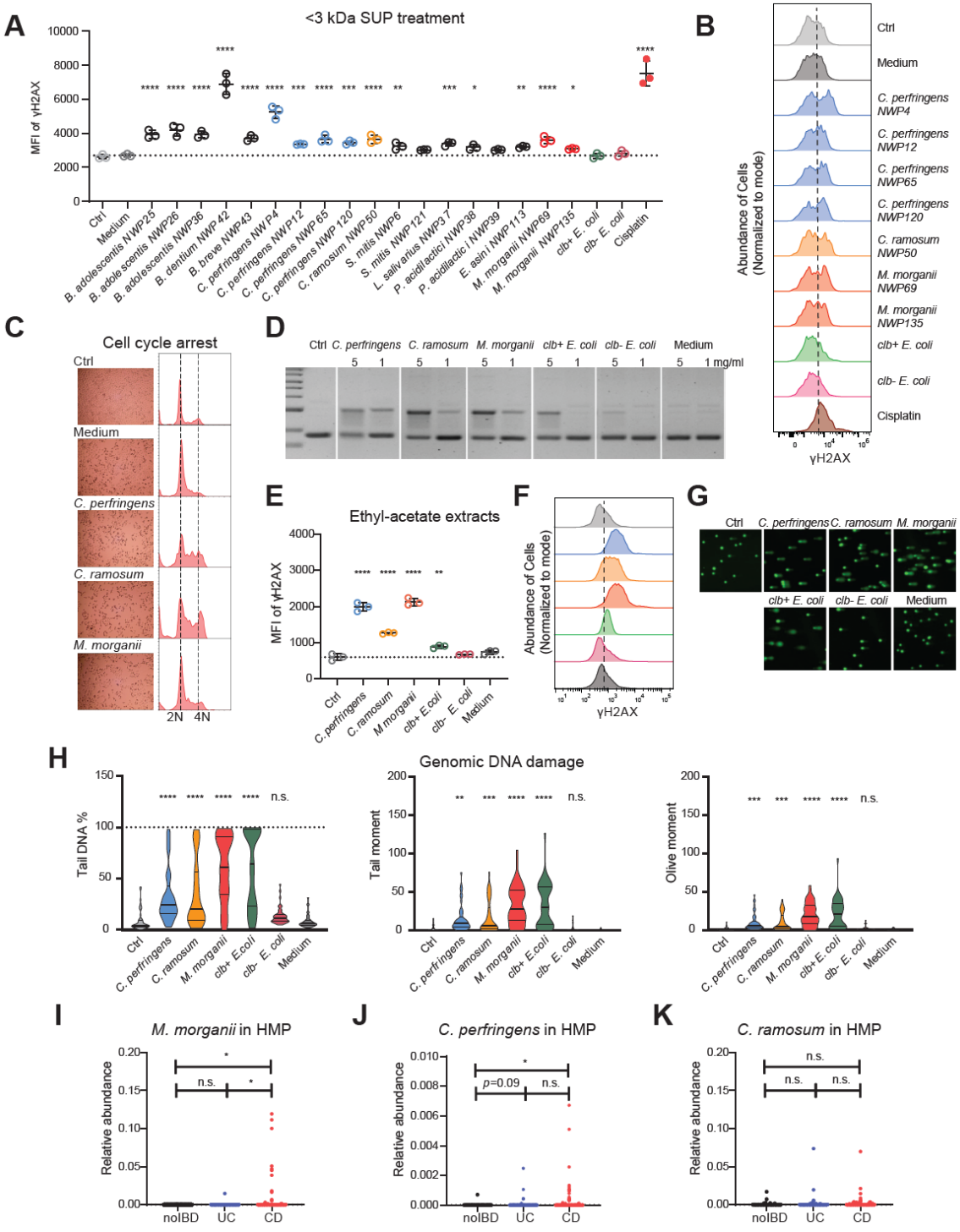


Figure 2. Small molecule metabolites produced by human gut microbes induced DNA damage.

(A) MFI (geometric mean fluorescence intensity) of γ-H2AX staining of HeLa cells treated with 40 % (v/v) PBS (Ctrl) or <3 kDa SUP (small-molecule bacterial supernatants) for 5-6 h.

(B) Representative histograms of γ-H2AX staining of HeLa cells treated with <3 kDa SUP from *C. perfringens, C. ramosum* or *M. morganii* isolates.

(C) HeLa cells were treated with 40 % (*v/v*) PBS or <3 kDa SUP from medium, *C. perfringens, C. ramosum* or *M. morganii* isolates for 24 h. Cell cycle arrest was evaluated by propidium iodide (PI) staining based on flow cytometry.

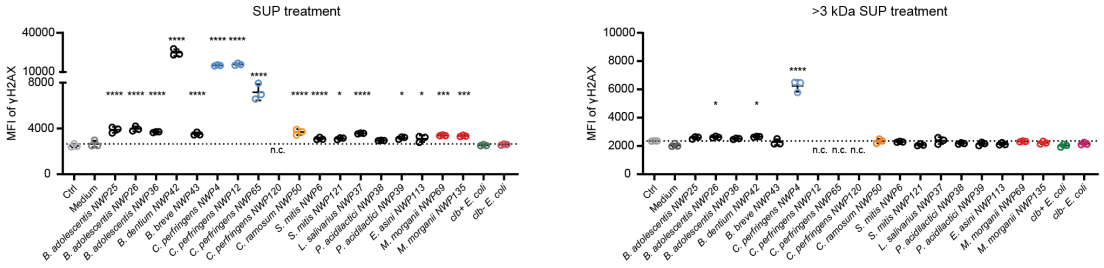
(D-H) Assessment of DNA damage induced by ethyl-acetate extracts of *C. perfringens, C. ramosum* or *M. morganii* supernatants. Evaluation of nicking of circular pUC19 DNA (top band = nicked DNA) after co-incubation for 5-6 h. Ctrl, control pUC19 DNA in TE buffer. (D); γ-H2AX staining (E and F) and genomic DNA comets (G-H) in HeLa cells after treatment with 5 mg/ml extracts or without (Ctrl) for 5-6 h. n = 49 for comet assay analysis.

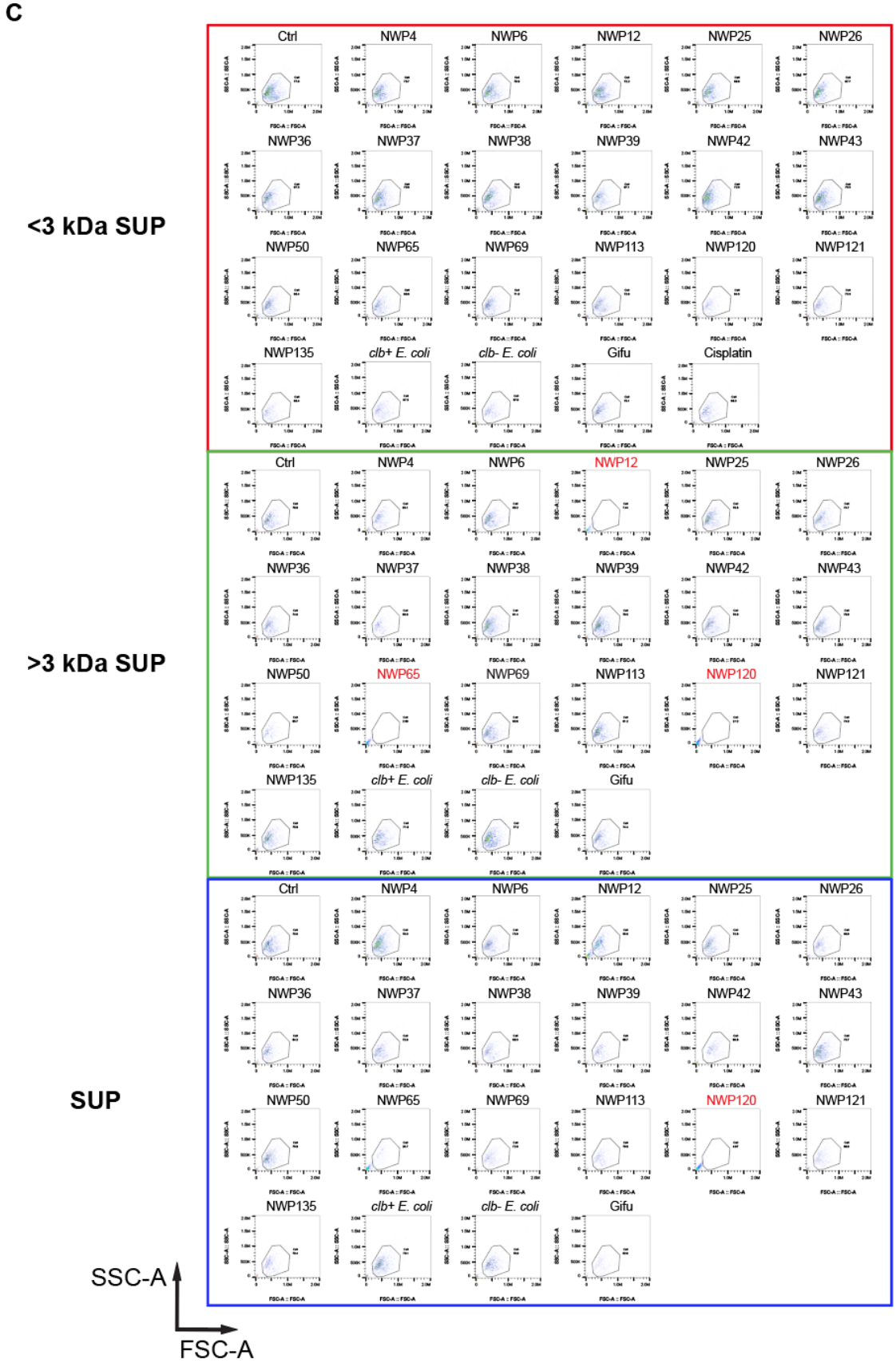
(I-K) Relative abundance of *C. perfringens, C. ramosum* or *M. morganii* in the Human Microbiome Project database. noIBD, healthy controls; UC, Ulcerative Colitis patients; CD, Crohn's Disease patients. n.s., not significant; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$, one-way ANOVA.

A

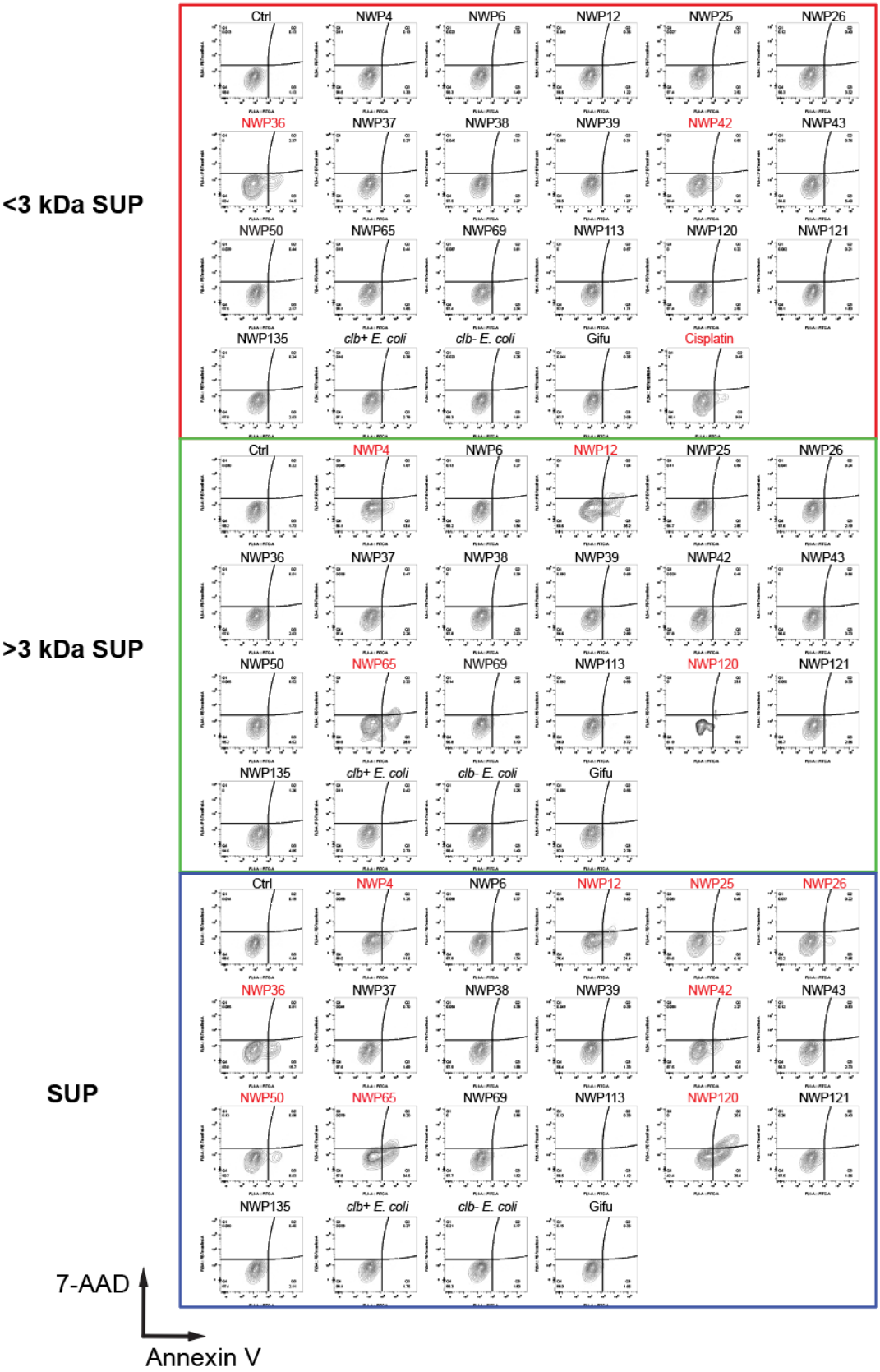
Phylum	Class	Order	Family	Genus	Species	No. of Strains
Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	*Bifidobacterium adolescentis*	3; NWP25, NWP26, NWP36
Actinobacteria	Actinomycetia	Actinobacteria_c	Bifidobacteriaceae	Bifidobacterium	*Bifidobacterium dentium*	1; NWP42
Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	*Bifidobacterium breve*	1; NWP43
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium	*Clostridium perfringens*	4; NWP4, NWP12, NWP65, NWP120
Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Clostridium_g6	*Clostridium ramosum*	1; NWP50
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	*Streptococcus mitis*	2; NWP6, NWP121
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	*Lactobacillus salivarius*	1; NWP37
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Pediococcus	*Pediococcus acidilactici*	2; NWP38, NWP39
Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	*Enterococcus asini*	1; NWP113
Proteobacteria	Gammaproteobacteria	Enterobacterales	Morganellaceae	Morganella	*Morganella morganii*	2; NWP69, NWP135

B




C
<3 kDa SUP
>3 kDa SUP
SUP
Ctrl
NWP4
NWP6
NWP12
NWP25
NWP26
NWP36
NWP37
NWP38
NWP39
NWP42
NWP43
NWP50
NWP65
NWP69
NWP113
NWP120
NWP121
NWP135
clb+ E. coli
clb- E. coli
Gifu
Cisplatin
SSC-A
FSC-A

D


<3 kDa SUP
>3 kDa SUP
SUP
Ctrl
NWP4
NWP6
NWP12
NWP25
NWP26
NWP36
NWP37
NWP38
NWP39
NWP42
NWP43
NWP50
NWP65
NWP69
NWP113
NWP120
NWP121
NWP135
clb+ E. coli
clb- E. coli
Gifu
Cisplatin
7-AAD
Annexin V

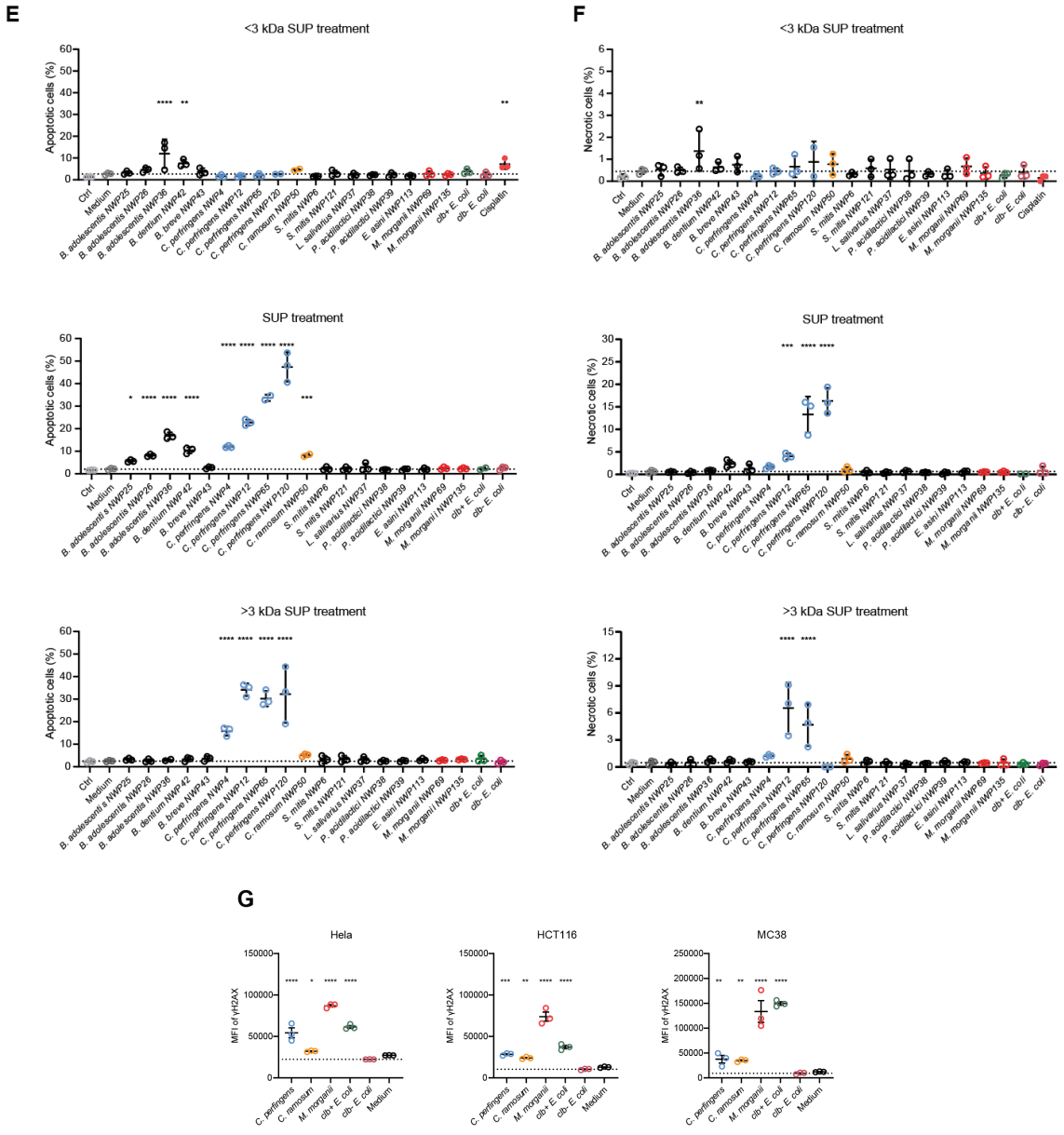


Figure S4. Selection of 18 putative genotoxic isolates and representative cell death flow cytometry plots.

(A) 18 potential genotoxic isolates were selected based on two rounds of *in vitro* screening.

(B) MFI of γ-H2AX staining of HeLa cells treated with 40 % (v/v) unfractionated bacterial supernatants (SUP), or >3 kDa SUP (large-molecules) for 5-6 h.

(C-D) Flow cytometry plots depicting cell size and granularity (FSC/SSC) (C) or cell death, apoptosis and necrosis (D) of HeLa cells after treatment with 40 % (v/v) <3 kDa SUP, unfractionated SUP or >3 kDa SUP for 5-6 h.

(E-F) Percent of apoptotic cells (%) (E) or necrotic cells (%) (F) of HeLa cells treated with 40 % (v/v) <3 kDa SUP, unfractionated SUP or >3 kDa SUP for 5-6 h.

(G) MFI of γ-H2AX staining of HeLa, HCT116 or MC38 cells treated with ethyl-acetate extracts of *C. perfringens, C. ramosum*, *M.* morganii, *clb+ E. coli, clb- E. coli* supernatants or medium for 5-6 h.

M. morganii produces a previously uncharacterized genotoxin that is distinct from colibactin

The biosynthetic machinery involved in the production of microbial metabolites, including previously characterized small molecule genotoxins, is often encoded by biosynthetic gene clusters (BGCs) [138]. For example, colibactin production is encoded by a multimodular PKS-NRPS pathway in *E. coli*, and tilimycin and tilivalline are encoded by an NRPS pathway in *Klebsiella oxytoca* [135, 139]. However, BGC analyses of genotoxic *C. perfringens*, C*. ramosum* and *M. morganii* using antiSMASH [140] failed to detect any known genotoxin-encoding BGCs (**Fig. S5A**, **Table S3**). While *M. morganii* harbors one NRPS/PKS gene cluster, this BGC is entirely distinct from the *clb* genomic island. In fact, *M. morganii* lacks key genes involved in colibactin synthesis, such as *clbI* and *clbP* (**Fig. S5B**) [141, 142]. This is consistent with recent analyses of 69 publicly available *Morganella* genomes, which suggests that *clb* genes are absent in *Morganella* spp. [143]. The genotoxicity caused by *M. morganii* is also distinct from that caused by colibactin—while *clb+ E. coli* caused DNA crosslinking, DNA exposed to *M. morganii*-derived metabolites displayed a smearing pattern under both native and denaturing conditions (**Fig. S5C**). Finally, as previously reported, colibactin-induced γ-H2AX may require cell-to-cell contact between bacterial and host cells as both separation via a filter abrogates activity and supernatants from *clb+ E. coli* failed to induce dramatic increases in γ-H2AX in cell lines, likely due to documented colibactin instability [71, 144, 145]. By contrast, *M. morganii* supernatants and small molecule metabolites elicited significant increases in γ-H2AX (**Fig. S5D-G**). Together, these data suggest that *M. morganii* produces a previously uncharacterized genotoxin(s) that is distinct from colibactin and is readily diffusible.

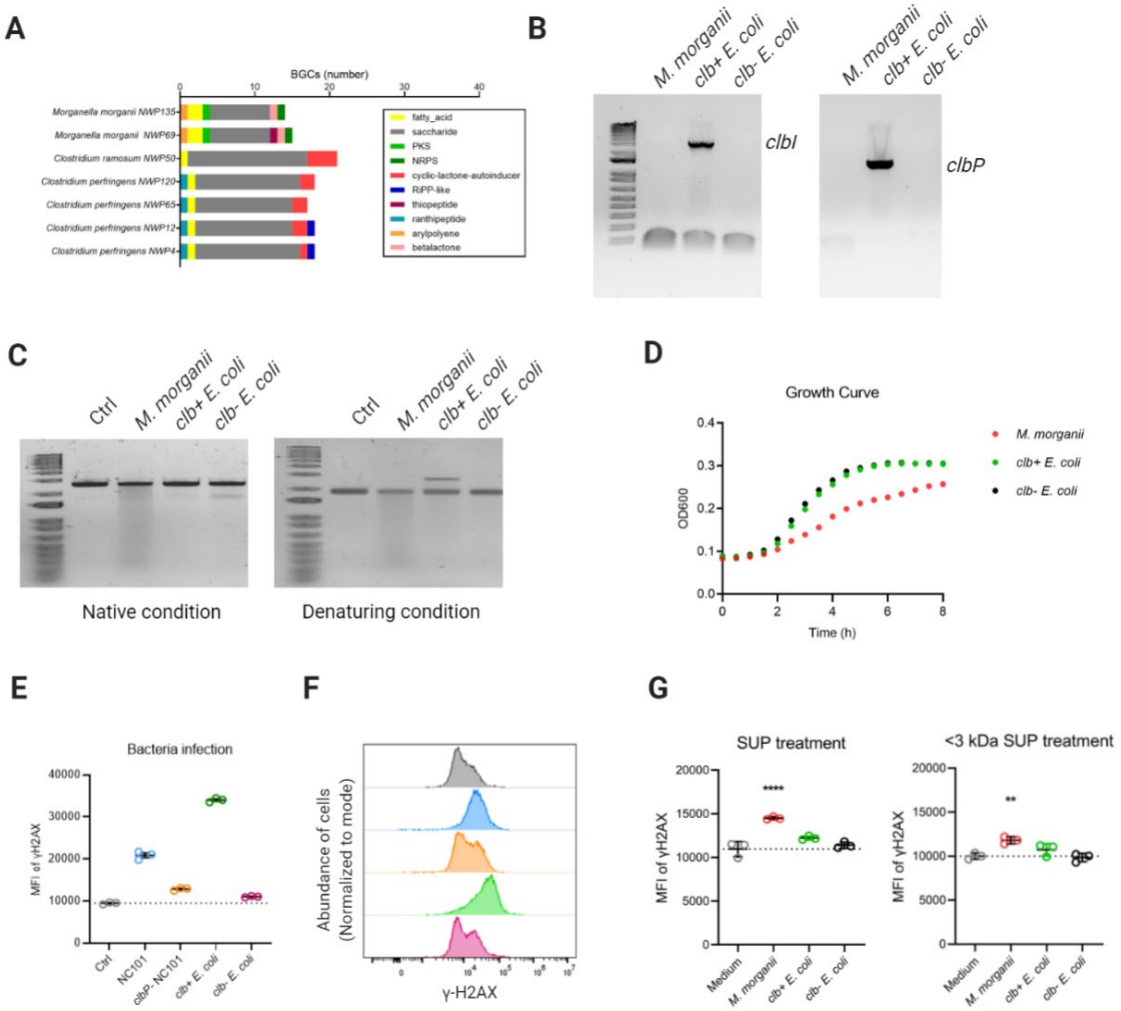


Figure S5. Genotoxic human gut commensals, including *M. morganii*, lack known genotoxin-encoding BGCs and induce DNA damage via colibactin-independent mechanisms.

(A) Biosynthetic gene clusters (BGCs) encoded by genotoxic isolates of *C. perfringens, C. ramosum*, and *M. morganii*. BGCs were identified using antiSMASH as described in the materials and methods. (**Table S3**)

(B) PCR-based detection of the colibactin synthesis-related genes, *clbI* and *clbP*, in *M. morganii*, *clb+ E. coli*, and *clb- E. coli*.

(C) DNA damage induced by *M. morganii*, *clb+ E. coli* and *clb- E. coli* based on linearized pUC19 DNA electrophoresis under native or denaturing conditions.

(D) Growth dynamics of *M. morganii*, *clb+ E. coli* and *clb- E. coli*.

(E-F) MFI and representative histograms of γ-H2AX after infecting HeLa cells with colibactin-producing strains, NC101 and *clb+ E. coli*, or non-colibactin-producing strains, *clbP-* NC101 and *clb- E. coli*, for 4 h.

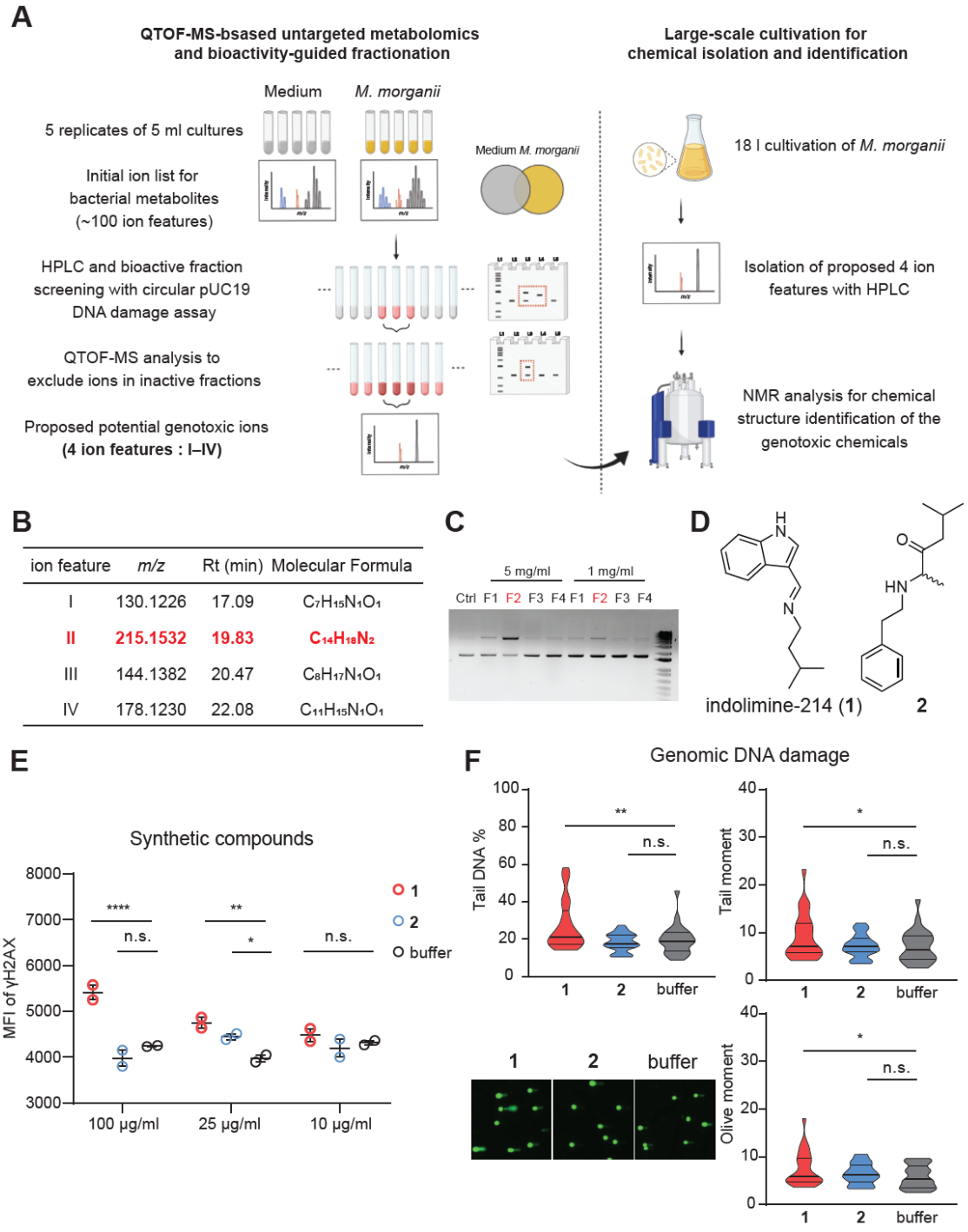
(G) MFI of γ-H2AX in HeLa cells treated with 40 % (v/v) SUP or <3 kDa SUP from *M. morganii*, *clb+ E. coli* and *clb- E. coli* for 5-6 h. * $p<0.05$; ** $p<0.01$; **** $p<0.0001$, one-way ANOVA.

Isolation and identification of a previously undescribed genotoxic metabolite derived from M. morganii

To identify specific genotoxin(s) produced by *M. morganii*, we employed a combination of ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS)-based untargeted metabolomics and bioactivity-guided fractionation using small-scale cultures, followed by large-scale cultivation and isolation for unambiguous structure elucidation and genotoxicity analyses (**Fig. 3A**). We generated an initial candidate ion list of the most abundant *M. morganii*-derived metabolites relative to Gifu medium control (~100 ion features, **Table S4**) via comparative metabolomics. We then performed two rounds of activity-guided fractionation using preparative high-performance liquid chromatography (HPLC) and circular pUC19 plasmid-based genotoxicity assays, then profiled the resulting fractions and subfractions using UPLC-QTOF-MS-based metabolomics (**Fig. S6A-B**). To identify potential genotoxins, we excluded ions present in inactive fractions from the aforementioned initial ion list and ultimately identified 4 ion features (I–IV) as potential genotoxic hits (**Fig. 3B**, **Table S4**). To enable structural elucidation and genotoxic activity assessment for these compounds, we performed a large-scale cultivation (18 liters) and ethyl acetate extraction of *M. morganii* supernatant based on previously observed retention times that imply relatively low polarity of the compounds of interest. This crude extract was subjected to two rounds of HPLC to generate four semi-pure fractions enriched in the four target ion features (F1-F4 enriched in I–IV, respectively). One of these fractions (F2) exhibited dose-dependent genotoxicity in circular pUC19 plasmid-based genotoxicity assays (**Fig. 3C**). Based on UPLC-QTOF-MS analyses, F2 was primarily comprised of two previously undescribed metabolites with *m/z* values of 215.1543 (compound **1**, target ion feature II) and 234.1852 (compound **2**) at a ratio of 4:6. This fraction was recalcitrant to further purification by preparative HPLC using diverse combinations of stationary and mobile phases. Thus, the chemical structures of the two components were characterized as a mixture using one- and two-dimensional nuclear magnetic resonance (NMR) spectroscopy analyses (**Fig. 3D**, **Fig. S7**). Compound **1** was featured as an imine chemical containing a carbon–nitrogen double bond. To better describe the chemical structure of this novel metabolite, we termed compound **1** as

indolimine-214, conjugation of indole-3-aldehyde and the leucine-derived metabolite isoamylamine. Compound **2** was a *M. morganii*-derived metabolite that was absent from the initial ion list, but enriched during extraction and fractionation. To confirm which compound exerted the observed genotoxicity, we synthesized both indolimine-214 (**1**) and the phenethylamine-derived compound **2**.

Synthetic indolimine-214 (**1**) was unstable due to the reversible nature of its imine bridge. Therefore, we fractionated fresh synthetic material and assessed the purity of each fraction using ^{1}H NMR to secure pure compounds for genotoxicity analysis (**Fig. S6C**). Nonetheless, neither synthetic indolimine-214 (**1**) nor the phenethylamine-derived metabolite **2** induced DNA damage in a circular pUC19 plasmid-based genotoxicity assay even at high concentrations (1 mg/ml). However, the mixture of indolimine-214 (**1**) and compound **2** elicited dose-dependent DNA damage at the experimentally observed isolation ratio of 4:6 (F2, **Fig. S6D**), suggesting that compound **2** may facilitate indolimine-214-induced DNA damage in cell-free assays. We next assessed potential genotoxicity of the synthetic compounds individually in HeLa cells and found that indolimine-214 (**1**) alone, but not compound **2**, triggered increased γ-H2AX in a dose-dependent manner (**Fig. 3E**) and induced tailing in an alkaline comet unwinding assay (**Fig. 3F**). Furthermore, the genotoxicity of synthetic indolimine-214 (**1**) correlated directly with its purity (i.e., in relation to its hydrolytic degradation products, **Fig. S6E**). Thus, we defined metabolite indolimine-214 (**1**) as a previously undescribed genotoxic small molecule metabolite produced by *M. morganii*.



ion feature	*m/z*	Rt (min)	Molecular Formula
I	130.1226	17.09	$C_7H_{15}N_1O_1$
II	**215.1532**	**19.83**	**$C_{14}H_{18}N_2$**
III	144.1382	20.47	$C_8H_{17}N_1O_1$
IV	178.1230	22.08	$C_{11}H_{15}N_1O_1$

Figure 3. Isolation and identification of a previously undescribed genotoxic metabolite derived from *M. morganii*.

(A) Overview of isolation and identification of genotoxins derived from *M. morganii*.

(B) Proposed 4 candidate ion features initially detected from *M. morganii* cultures. Rt, retention time.

(C) Evaluation of nicking of circular pUC19 DNA (top band = nicked DNA) after co-incubation overnight with F1–F4 fractions enriched with ion features I–IV, respectively. Ctrl, control pUC19 DNA in TE buffer.

(D) Chemical structures of compounds indolimine-214 (**1**) and **2**.

(E) MFI of γ-H2AX staining for HeLa cells treated with synthetic compounds at indicated concentrations for 5 h. n.s., not significant; * $p<0.05$; ** $p<0.01$; *** $p<0.001$, two-way ANOVA.

(F) Genomic DNA comets in HeLa cells after treatment with 100 µg/ml synthetic compounds for 5-6 h. n = 25 for comet assay analysis, n.s., not significant; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$, one-way ANOVA.

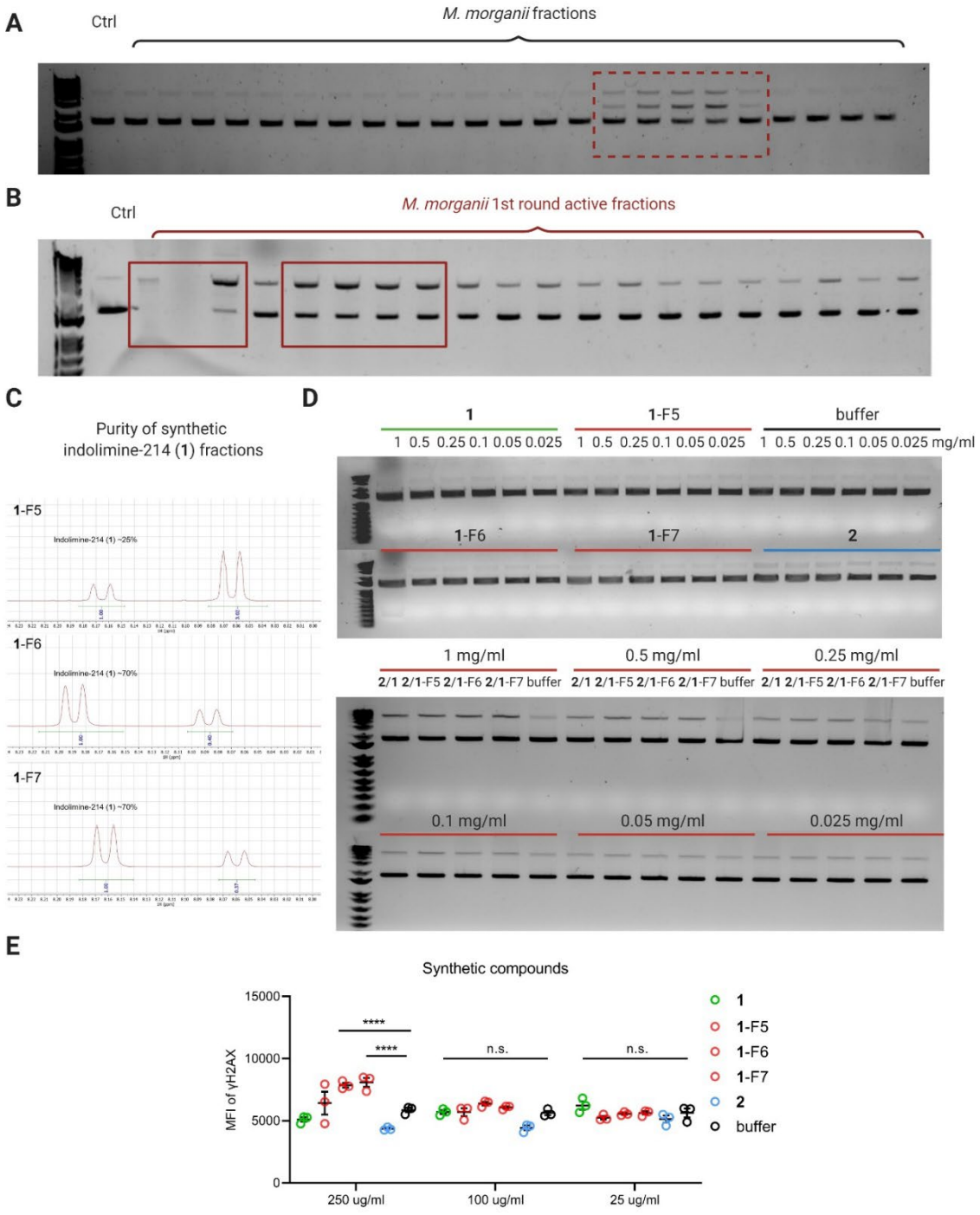


Figure S6. Isolation and assessment of bioactive genotoxic metabolites derived from *M. morganii*.

(A-B) Gel electrophoresis of two rounds of bioactive fraction screening. Evaluation of nicking of circular pUC19 DNA (top band = nicked DNA) after overnight co-incubation. Active fractions were labeled with dash or solid red squares.

(C) Purity assessment of synthetic indolimine-214 (**1**) fractions via integration values of H-8′ in **1** (**Fig. S7**).

(D) Assessment of DNA damage induced by individual (top) or mixed (bottom) synthetic compounds. Evaluation of nicking of circular pUC19 DNA (top band = nicked DNA) after co-incubation overnight.

(E) MFI of γ-H2AX staining for HeLa cells treated with synthetic compounds at indicated concentrations for 5 h. **** p<0.0001, two-way ANOVA.

	Indolimine-214 (**1**)			Compound **2**	
Position	δ_H	δ_C	Position	δ_H	δ_C
1	3.49	59.2	1	3.00, 3.12	45.7
2	1.48	40.2	2	2.90	31.6
3	1.68	25.2	3		136.8
4, 5	0.91	22.3	4, 8	7.22	128.9
1′	8.42	155.2	5, 7	7.31	129.0
2′		114.2	6	7.24	127.1
3′	7.70	130.6	1′	1.39	13.8
4′		136.9	2′	4.19	60.8
5′	7.38	111.5	3′		206.5
6′	7.12	122.3	4′	2.44, 2.52	46.8
7′	7.05	120.1	5′	2.04	23.9
8′	8.15	121.4	6′, 7′	0.87	22.5
9′		124.8	N*H*	8.92	
N*H*	11.42				

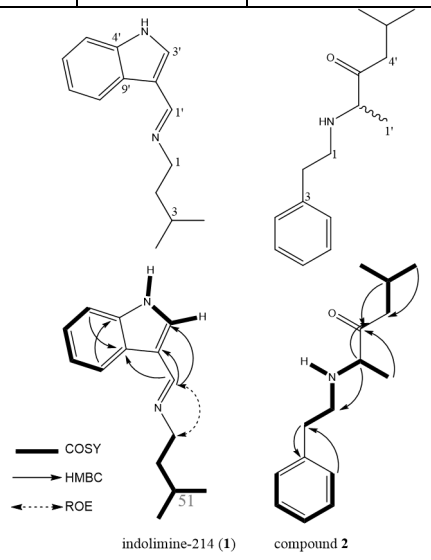


Figure S7. ^{1}H and ^{13}C NMR spectroscopic data of natural and synthetic compounds, indolimine-214 (1) and compound 2 (600 MHz, DMSO-d_6).

M. morganii produces multiple genotoxic indolimines in vivo and exacerbates CRC in gnotobiotic mice

M. morganii is enriched in fecal samples from IBD- and CRC-patients [59] and in CRC tumors [146]. As such, we chose to examine the potential impact of *M. morganii* on CRC *in vivo* as compared to a phylogenetically-related non-genotoxic control. The Enterobacterales strain *E. coli* NC101 exhibits genotoxicity *in vitro* [147] and exacerbates CRC in AOM/Il10$^{-/-}$ gnotobiotic mice via the production of colibactin [74]. However, an isogenic *ΔclbP E. coli* mutant (*clbP*- NC101) with a disrupted *clbP* gene failed to induce γ-H2AX in HeLa cells (**Fig. S5E-F**) or CRC in AOM/Il10$^{-/-}$ gnotobiotic mice [74]. UPLC-QTOF-MS-based quantification of indolimine-214 (**1**) with reference to its synthetic standard confirmed that *M. morganii* produces high levels of this metabolite *in vitro* (~40 μg/ml, **Fig. 4A**), which are comparable to the concentrations of synthetic compound sufficient to induce genotoxicity in HeLa cells (**Fig. 3E-F**). By contrast, indolimine-214 (**1**) was undetectable in supernatants from colibactin-producing *E. coli* strains, *clb*+ *E. coli* and NC101, and isogenic non-genotoxic controls, *clb*- *E. coli* and *clbP*- NC101(**Fig. 4A**). Correspondingly, *M. morganii* supernatants and small molecule metabolites induced γ-H2AX in HeLa cells, while *clb*- *E. coli* strains failed to induce detectable increases in γ-H2AX (**Fig. 4B**, **Fig. S5G**).

To assess genotoxic indolimine production *in vivo*, we colonized germ-free mice with either *M. morganii* or *clbP*- NC101. Cecal contents from mice colonized with *M. morganii* contained high levels of genotoxic indolimine-214 (**1**), but this compound was undetectable in mice colonized with *clbP*- NC101 (**Fig. 4C**). Moreover, in the process of quantifying cecal indolimine-214 (**1**), we observed the production of two additional previously undescribed indolimines. As indolimine-214 (**1**) consists of a conjugation of indole-3-aldehyde and primary amine isoamylamine, we speculated that *M. morganii* may also produce related indole conjugates. Indeed, we found that cecal contents from mice colonized with *M. morganii,* but not mice colonized with *E. coli,* contained two additional indolimines: conjugates of indole-3-aldehyde with isobutylamine (indolimine-200, compound **3**, *m/z* 201.1386) or phenethylamine (indolimine-248, **4**, *m/z* 249.1386) (**Fig.4D-E**, **Fig. S8**). These additional indolimines were also detected in *in vitro M.*

morganii bacterial cultures, but not *E. coli* cultures, as confirmed by synthetic standards (**Fig. 4F**). Finally, synthetic indolimine-200 (**3**) and indolimine-248 (**4**) also triggered increased γ-H2AX in a dose-dependent manner (**Fig. 4G**) Thus, *M. morganii* produced multiple genotoxic indolimines both *in vitro* and *in vivo*.

We performed bulk RNA-seq for colonic epithelial cells isolated from mice treated with or without indolimine-214 (**1**) to evaluate the genotoxicity *in vivo* (**Fig. 4H**). Gene ontology enrichment analysis of upregulated genes in epithelial cells after indolimine-214 (**1**) treatment revealed that double-strand repair via break-induced replication exhibited the highest fold enrichment after statistical over-representation test, consistent with the DSB results observed *in vitro* (**Fig. 4I**). Genomic instabilities during DNA double-strand break-induced repair (BIR) pathway are a prominent source of mutation clusters in cancer [148]. In addition, the enrichment of phosphatidylinositol-mediated signaling and cellular response to DNA damage stimulus suggested the potential involvement of downstream singling pathways related to tumorigenesis, for example, PI3K pathway (**Fig. 4I**) [149, 150]. Representative genes induced by indolimine-214 (**1**) in colonic epithelial cells included *Mcm3* and *Mus81* regulating DSB repair; *Pik3cd*, *Plcb1* and *Plcb2* regulating PI3K pathway; and *Cd74*, *Alkbh2*, *Parp3*, *Mnat1*, *Brat1* and *Nabp1* regulating DNA damage cellular response (**Fig. 4J**). Together, the alterations of genes or gene sets after indolimine-214 (**1**) treatment indicated the potential effects of genotoxin-producing microbes on tumor promotion.

To elucidate the effects of indolimine-producing *M. morganii* on CRC, we measured the induction of colorectal tumors in gnotobiotic mice colonized with *M. morganii* or its phylogenetically-related non-genotoxic control strain *E. coli* (*clbP*- NC101) after repeated administration of azoxymethane (AOM) and dextran sulfate sodium (DSS) (**Fig. 4K**). We observed that *M. morganii* colonized mice developed more adenomatous polyps and invasive adenocarcinomas (**Fig. 4L**), and exhibited increased tumor numbers and overall tumor burden as compared to mice colonized with the *clbP*- NC101 (**Fig. 4M**). Overall, these data suggest that genotoxin-producing *M. morganii* can exacerbate CRC.

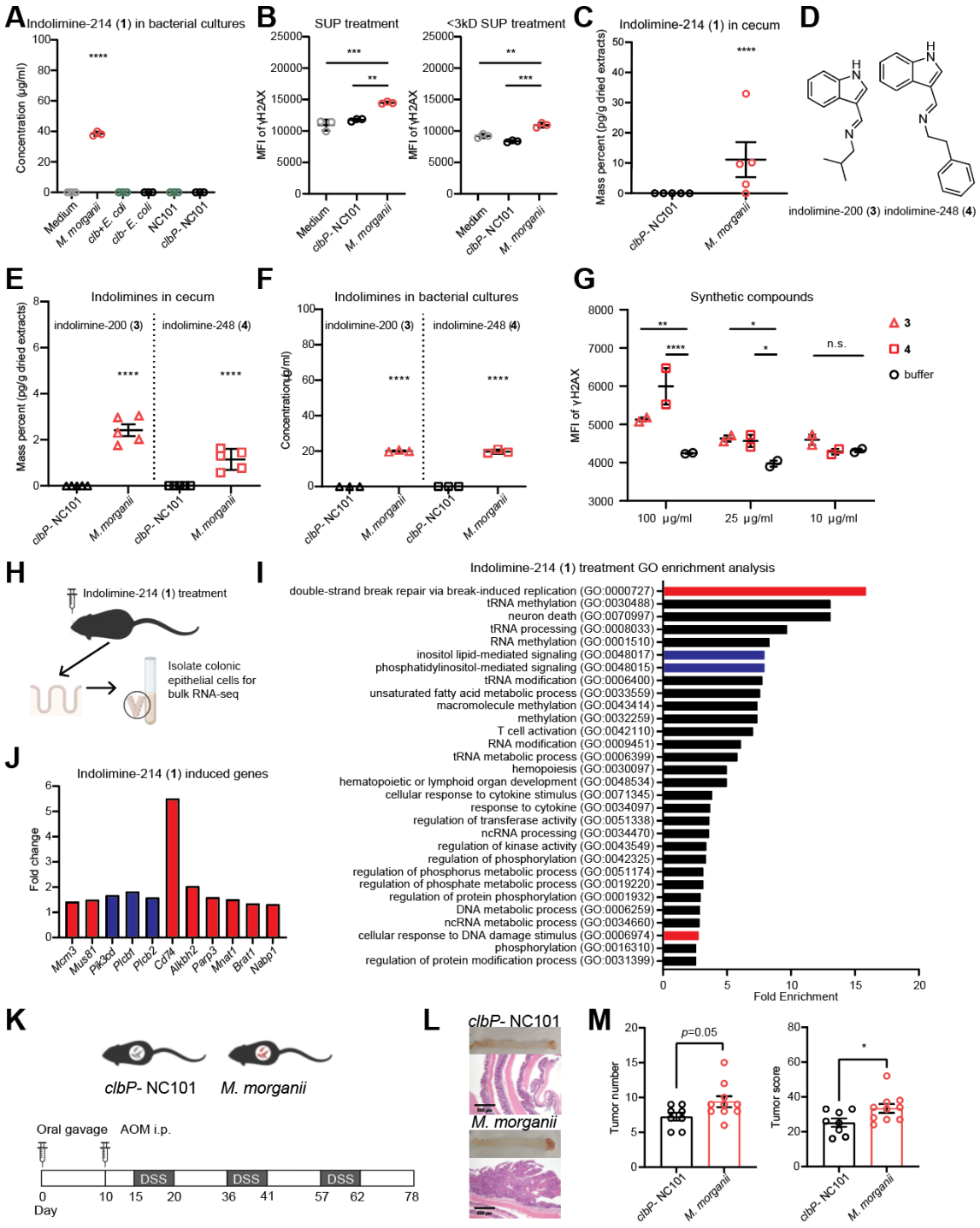

A
Indolimine-214 (1) in bacterial cultures
Concentration (μg/ml)
Medium
M. morganii
clbA-E. coli
clbB-E. coli
NC101
clbP- NC101
B
SUP treatment
<3kD SUP treatment
MFI of γH2AX
C
Indolimine-214 (1) in cecum
Mass percent (pg/g dried extracts)
D
indolimine-200 (3) indolimine-248 (4)
E
Indolimines in cecum
indolimine-200 (3)
indolimine-248 (4)
F
Indolimines in bacterial cultures
G
Synthetic compounds
n.s.
3
4
buffer
100 μg/ml
25 μg/ml
10 μg/ml
H
Indolimine-214 (1) treatment
Isolate colonic epithelial cells for bulk RNA-seq
I
Indolimine-214 (1) treatment GO enrichment analysis
double-strand break repair via break-induced replication (GO:0000727)
tRNA methylation (GO:0030488)
neuron death (GO:0070997)
tRNA processing (GO:0008033)
RNA methylation (GO:0001510)
inositol lipid-mediated signaling (GO:0048017)
phosphatidylinositol-mediated signaling (GO:0048015)
tRNA modification (GO:0006400)
unsaturated fatty acid metabolic process (GO:0033559)
macromolecule methylation (GO:0043414)
methylation (GO:0032259)
T cell activation (GO:0042110)
RNA modification (GO:0009451)
tRNA metabolic process (GO:0006399)
hemopoiesis (GO:0030097)
hematopoietic or lymphoid organ development (GO:0048534)
cellular response to cytokine stimulus (GO:0071345)
response to cytokine (GO:0034097)
regulation of transferase activity (GO:0051338)
ncRNA processing (GO:0034470)
regulation of kinase activity (GO:0043549)
regulation of phosphorylation (GO:0042325)
regulation of phosphorus metabolic process (GO:0051174)
regulation of phosphate metabolic process (GO:0019220)
regulation of protein phosphorylation (GO:0001932)
DNA metabolic process (GO:0006259)
ncRNA metabolic process (GO:0034660)
cellular response to DNA damage stimulus (GO:0006974)
phosphorylation (GO:0016310)
regulation of protein modification process (GO:0031399)
Fold Enrichment
J
Indolimine-214 (1) induced genes
Fold change
Mcm3
Mus81
Pik3cd
Plcb1
Plcb2
Cdf4
Akbh2
Parp3
Mnat1
Brat1
Nabp1
K
clbP- NC101
M. morganii
Oral gavage
AOM i.p.
DSS
Day
L
clbP- NC101
M. morganii
M
p=0.05
Tumor number
Tumor score

Figure 4. *M. morganii* produces multiple genotoxic indolimines *in vivo* and exacerbates CRC in gnotobiotic mice.

(A) UPLC-QTOF-MS quantification of indolimine-214 (**1**) in medium or bacterial supernatants of *M. morganii*, *clb+ E. coli*, *clb- E. coli*, NC101 *E. coli* or *clbP-* NC101 *E. coli*. **** $p<0.0001$, one-way ANOVA.

(B) MFI of γ-H2AX in HeLa cells treated with 40 % (*v/v*) SUP or <3 kDa SUP from medium, *M. morganii,* or *clbP-* NC101 *E. coli* for 5-6 h. ** $p<0.01$; *** $p<0.001$, one-way ANOVA.

(C) QTOF-MS quantification of indolimine-214 (**1**) in cecal contents of gnotobiotic mice colonized by *M. morganii,* or *clbP-* NC101 *E. coli*. **** $p<0.0001$, Student's *t*-test.

(D) Chemical structures of compounds indolimine-200 (**3**) and indolimine-248 (**4**).

(E) QTOF-MS quantification of indolimine-200 (**3**) and indolimine-248 (**4**) in cecal contents of gnotobiotic mice colonized by *M. morganii*, or *clbP-* NC101 *E. coli*. **** $p<0.0001$, Student's *t*-test.

(F) QTOF-MS quantification of indolimine-200 (**3**) and indolimine-248 (**4**) in bacterial supernatants of *M. morganii* or *clbP-* NC101 *E. coli*. **** $p<0.0001$, Student's *t*-test.

(G) MFI of γ-H2AX staining for HeLa cells treated with synthetic compounds at indicated concentrations for 5 h. n.s., not significant; * $p<0.05$; ** $p<0.01$; **** $p<0.0001$, two-way ANOVA.

(H) Schematic of experimental design for genotoxicity evaluation of indolimine-214 (**1**) *in vivo*.

(I) Gene ontology enrichment analysis for upregulated genes of colonic epithelial cells isolated from mice treated with or without indolimine-214 (**1**). Top 30 GOs with higher fold enrichment score were shown.

(J) Fold changes of representative upregulated genes from related GOs in (I).

(K) Schematic of experimental design for CRC induction in age-matched gnotobiotic mice colonized with *M. morganii* or *clbP-* NC101 *E. coli*.

(L-M) Representative colon tissue and histology images (L) and tumor burden (tumor number and tumor score) (M) in gnotobiotic mice colonized with *M. morganii* or *clbP-* NC101 *E. coli*. Each dot represents one mouse (n = 8-10 per group), n.s., not significant; * $p<0.05$, Student's *t*-test.

	Indolimine-200 (**3**)			Indolimine-248 (**4**)	
Position	δ_H	δ_C	Position	δ_H	δ_C
1	3.58	58.2	1	3.80	60.2
2	2.11	28.1	2	2.98	36.3
3	1.04	17.7	3		138.5
1′	8.73	160.4	4, 8	7.22	128.0
2′		107.2	5, 7	7.22	127.3
3′	8.46	Not Detected	6	7.22	127.8
4′		137.6	1′	8.30	157.4
5′	7.61	112.6	2′		112.0
6′	7.42	124.4	3′	7.72	131.4
7′	7.41	123.3	4′		136.8
8′	8.13	118.7	5′	8.03	119.2
9′		123.1	6′	7.16	120.2
			7′	7.42	110.9
			8′	8.03	119.2
			9′		124.3

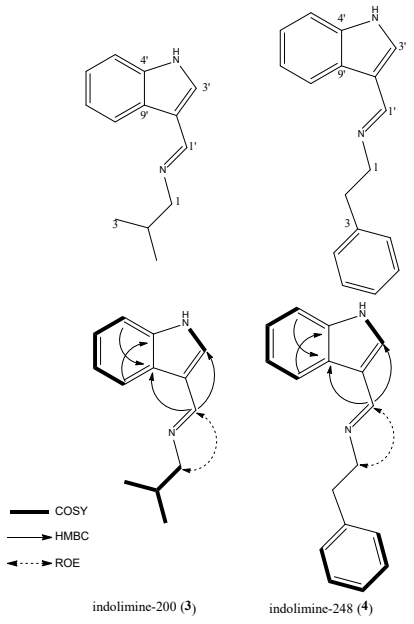


Figure S8. ^{1}H and ^{13}C NMR spectroscopic data of natural and synthetic compounds, indolimine-200 (3) and 248 (4) (600 MHz, methanol-d_4).

Indolimine synthesis required a pyridoxal-dependent decarboxylase in M. morganii

To identify the genes responsible for the indolimine synthesis in *M. morganii*, we focused on the potential active imine group. Imine formation is a reversible process that starts with the nucleophilic addition of a primary amine to the carbonyl group of an aldehyde or ketone. Therefore, we hypothesized that indolimine-214 (**1**), indolimine-200 (**3**) and indolimine-248 (**4**) could use indole-3-aldehyde (IAld) and respective primary amines (isoamylamine, isobutylamine and phenethylamine) as precursors (**Fig. 5A**). Synthesis of primary amines is a one-step reaction catalyzed by microbial decarboxylases from common amino acids [151]. Based on the whole-genome sequence of *M. morganii* NWP135, we found that this isolate contained 18 decarboxylases but just three proteins had clear orthologs greater than 60% protein sequence similarity of the valine decarboxylase in *S. viridifaciens* (**Fig. 5B**, **Table S5**) [151]. To identify the gene mediating primary amine synthesis in *M. morganii*, *E. coli* BL21(DE3) was separately transformed with the plasmid pET28-Peg harboring the codon optimized DNA sequences of potential proteins (Peg1085, Peg1320 or Peg3098). Then we induced the expression of proteins with IPTG and detected indolimine-214 (**1**), indolimine-200 (**3**) or indolimine-248 (**4**) in bacterial supernatants after feeding with IAld and leucine, valine or phenylalanine respectively. Peg1085 was found sufficient to synthesize indolimines using IAld and respective amino acids as precursors (**Fig. 5C**). Although Peg1085 was functionally predicted as glutamate decarboxylase in our WGS analysis based on RAST (**Fig. 5B**, **Table S5**) [152], it was annotated as aspartate aminotransferase (AAT) superfamily (fold type I) of pyridoxal phosphate (PLP)-dependent enzymes in NCBI. Therefore, the gene encoding AAT_I involved indolimine synthesis in *M. morganii* (**Fig. 5D**).

To evaluate if AAT_I is necessary for indolimine synthesis, we constructed a random mutagenesis library of *M. morganii* NWP135 with Ez-Tn5 transposon. After Ez-Tn5 electrotransformation and single-colony picking, we performed combinatory pooling and NGS sequencing with customized primers as described in Sudoku [153, 154]. From ~16, 000 colonies, we identified one mutant strain (*aat- M. morganii*) that transposon was inserted 7 bp after the start codon (**Fig. S9**), generating a disrupted *aat* gene with larger size (**Fig. 5F**). As expected, the *aat-*

M. morganii mutant lost the ability to induce DNA damage both in cell-based γ-H2AX induction assay (**Fig. 5G**) and cell-free plasmid DNA electrophoresis assay (**Fig. 5H**) compared to the wild-type strain *aat*+ *M. morganii*. Furthermore, indolimines were completely absent in the supernatants of *att*- *M. morganii* mutant (**Fig. 5I**). Taken together, these data indicated that *aat* gene encoding AAT_I was both sufficient and necessary for indolimine synthesis and regulated the genotoxicity of *M. morganii*.

Discussion

Aside from a small number of case studies [155], the taxonomic distribution and repertoire of small molecule genotoxins produced by the microbiota remain mostly unexplored. Here, we undertook a systematic evaluation of the carcinogenic potential of a diverse selection of human gut microbes based on the reasoning that somatic mutations resulting from DNA damage are critical mediators of tumor initiation and progression [156]. Although diverse taxa from the human gut microbiota exhibited genotoxicity, we just focused on *M. morganii* because of the higher prevalence and abundance in IBD patients who usually have higher risk to develop CRC. We identified and characterized a previously undiscovered family of the genotoxic *M. morganii*-derived small molecules termed the indolimines and determined that colonization with genotoxin-producing *M. morganii* exacerbated CRC in gnotobiotic mice. We also dissected the synthesis of indolimines and found that *aat* gene encoding aspartate aminotransferase fold type I (AAT_I) was necessary and sufficient to mediate the one-step reaction converting amino acids into primary amines, precursors of indolimines. The *att*- *M. morganii* constructed with transposon-based random mutagenesis exhibited defects of genotoxicity.

By revealing the existence of a previously uncharted universe of microbiota-derived genotoxins and defining the indolimines as a novel family of bioactive microbiota-derived small molecules, these studies imply an expanded role for genotoxic metabolites in CRC. More broadly, novel genotoxins, including the indolimines, may also impact diverse aspects beyond tumor initiation. For example, bacterial production of genotoxins can enhance competitive fitness for species that thrive in inflammatory microenvironments and thereby shape microbiota composition [157, 158],

which may select for the evolution or maintenance of DNA-damaging compounds by individual commensal species. Notably, somatic mutations can be detected in colonic epithelial cells even in early life, which suggests persistent mutagenesis throughout the lifespan of an individual [159]. Furthermore, while CRC patients display increased carriage of *clb+ E. coli*, *clb+* taxa (including *E. coli* relatives, such as *Klebsiella* species) are also found in healthy individuals [72]. Altogether, these observations support a model whereby genotoxic gut microbes contribute to CRC development by persistently inducing DNA damage in host epithelial cells, which synergizes with chronic inflammation in the gut microenvironment, along with additional environmental factors, and eventually facilitates the initiation and progression of CRC.

Finally, these studies underscore the power of function-based assessments of the microbiome to provide new insights into the diverse impacts of indigenous microbes on host biology and disease susceptibility. Recent illumination of the ‘tumor microbiome’ [160] beyond the gut highlights an additional potential role for microbial genotoxins in tumor initiation and progression. Furthermore, the broad distribution of genotoxicity across diverse gut species suggests that resident microbes from other mucosal tissues may also produce previously undiscovered genotoxins. Thus, in addition to revealing an expanded diversity and significance of microbiota-derived genotoxins in CRC, these studies provide a roadmap for future identification and characterization of novel microbiota-derived genotoxins across diverse tissues and disease states.

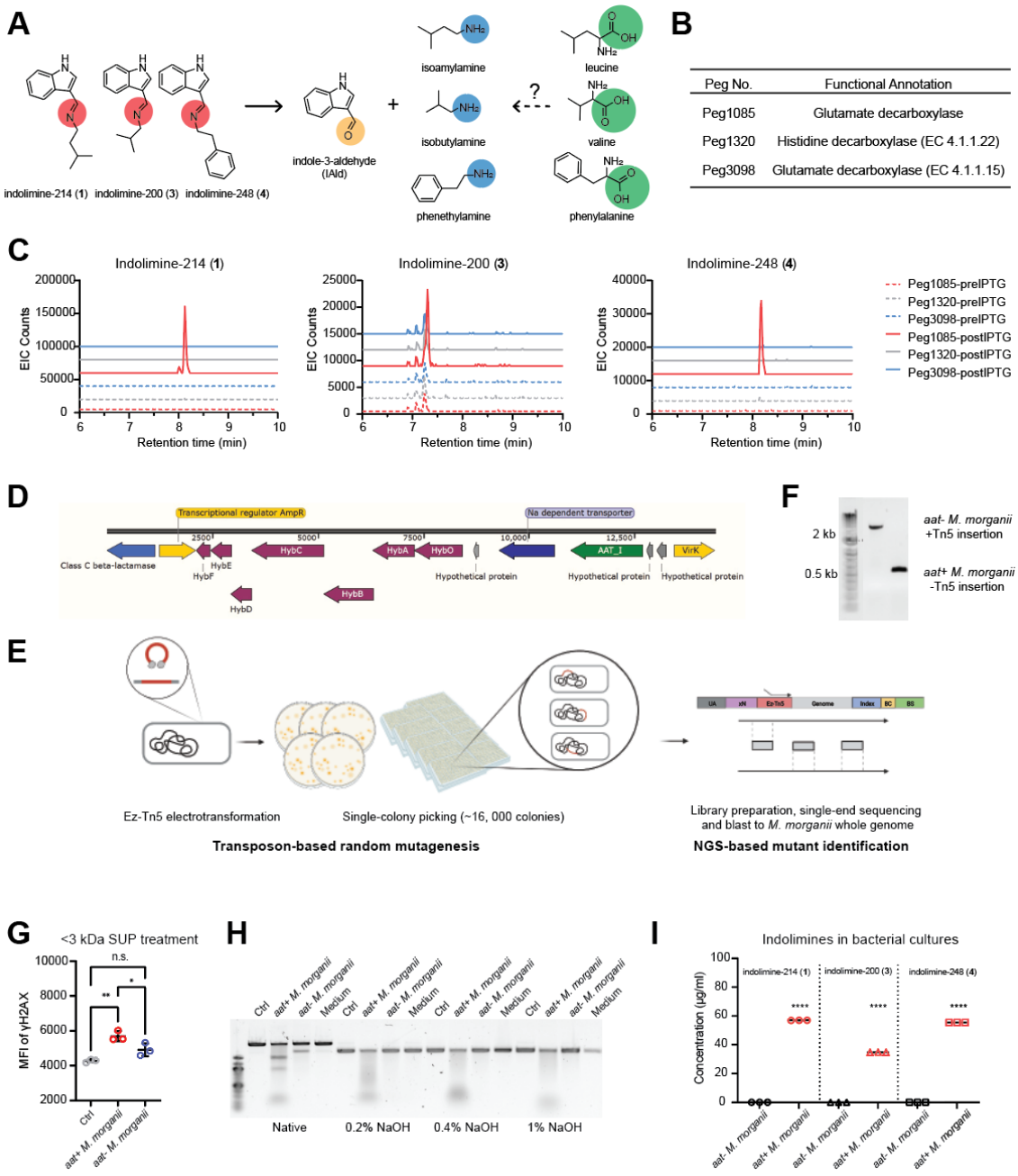


Figure 5. Indolimine synthesis required a pyridoxal-dependent decarboxylase in *M. morganii*.

(A) Hypothesized indolimine synthesis from IAld and primary amines.

(B) Potential target proteins in *M. morganii* NWP135 with orthologs of valine decarboxylase.

(C) QTOF-MS identification of indolimine-214 (**1**) indolimine-200 (**3**) and indolimine-248 (**4**) in *E. coli* BL21(DE3). *E. coli* cells were separately transformed with the plasmid pET28-Peg harboring the codon optimized DNA sequences of Peg1085, Peg1320 or Peg3098. Indolimines were detected after IPTG induction and feeding with IAld and leucine, valine and phenylalanine respectively. preIPTG, bacterial supernatants before IPTG induction; post-IPTG, bacterial supernatants after IPTG induction and precursor feeding.

(D) Schematic gene fragment containing the gene encoding AAT_I in *M. morganii* NWP135.

(**E**) Schematic pipeline of transposon-based random mutagenesis and NGS-based mutant identification.

(**F**) Gel results of PCR products of *aat* gene in *aat-* or *aat+ M. morganii*.

(**G**) MFI of γ-H2AX in HeLa cells treated with 40 % (*v/v*) <3 kDa SUP from *aat+ M. morganii,* or *aat- M. morganii* for 5-6 h. n.s., not significant; * p<0.05; ** p<0.01, one-way ANOVA.

(H) Gel electrophoresis of cell-free DNA damage assay. linearized pUC19 DNA was co-incubated with medium, *aat+ M. morganii* or *aat- M. morganii* for 7-8h, isolated via column purification and treated with or without NaOH (0 %, 0.2 %, 0.4 %, 1 %) before evaluating DNA integrity via gel electrophoresis.

(I) QTOF-MS quantification of indolimine-214 (**1**), indolimine-200 (**3**) and indolimine-248 (**4**) in bacterial supernatants of *aat+ M. morganii* or *aat- M. morganii*. **** p<0.0001, Student's *t*-test.

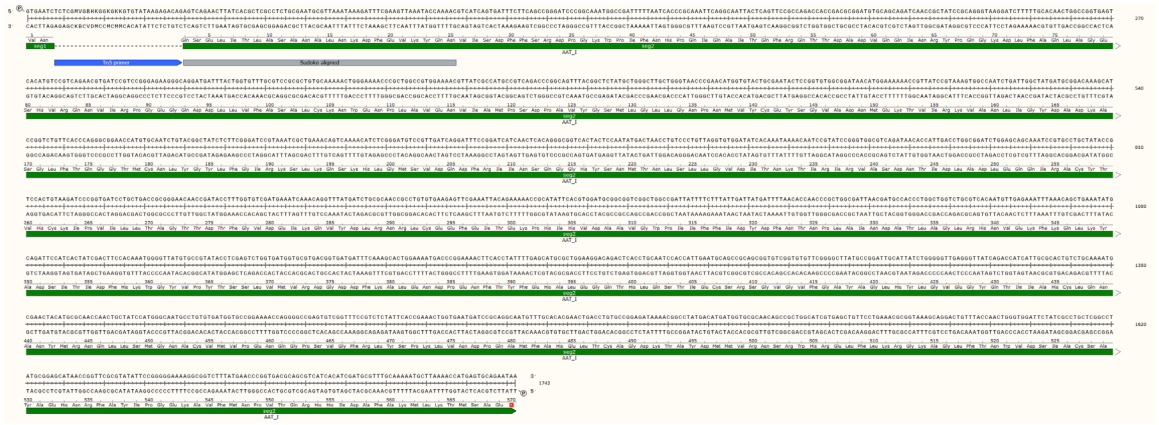


Figure S9. Alignment of Sudoku amplicon with *aat* gene.

The 100 bp Sudoku amplicon contained 32 bp Ez-Tn5 transposon sequence and 68 bp genomic sequence after the insertion site. The insertion site was 7 bp after the start codon of *aat* gene.

Future Directions

Indolimines produced by M. morganii induced tumorigenesis independent of inflammation

With the *aat- M. morganii*, we found that the tumorigenesis promoted by genotoxins was independent of inflammation, through comparing the proinflammatory ability between indolimine-producing or non-producing *M. morganii*, with colibactin-producing or non-producing NC101 as controls. Acute DSS model showed that there were no differences among bacteria-colonizing mice, compared to germ-free mice in weight loss, colon length, bacterial load and fecal lipocalin 2 (**Fig. 6A-E**), indicating genotoxins (both colibactin and indolimines) didn't change bacterial abundance and inflammatory level *in vivo*.

To confirm the tumorigenesis effects of genotoxin-producing microbes in a more complex background, we plan to repeat the AOM/DSS CRC model with mice also colonized with a non-genotoxic community (Geno-) (**Fig. 6F-G**). The Geno- community consisted of 7 species expanding different phyla, selected based on our primary cell-free DNA damage assay (**Fig. 1B**). We chose this relatively simple community instead of complex natural human microbiota because of the findings that diverse gut microbes exhibited DNA damaging activities and microbiota-mediated genotoxicity may be more widespread than previously appreciated (**Fig. 1**). We also performed neonatal colonization of *clbP*+ NC101 *E. coli*, *clbP*- NC101 *E. coli*, *aat*+ *M. morganii* or *aat- M. morganii* in SPF *Apc*^{min} mice to investigate the tumorigenesis of genotoxin-producing microbes in a complex community background.

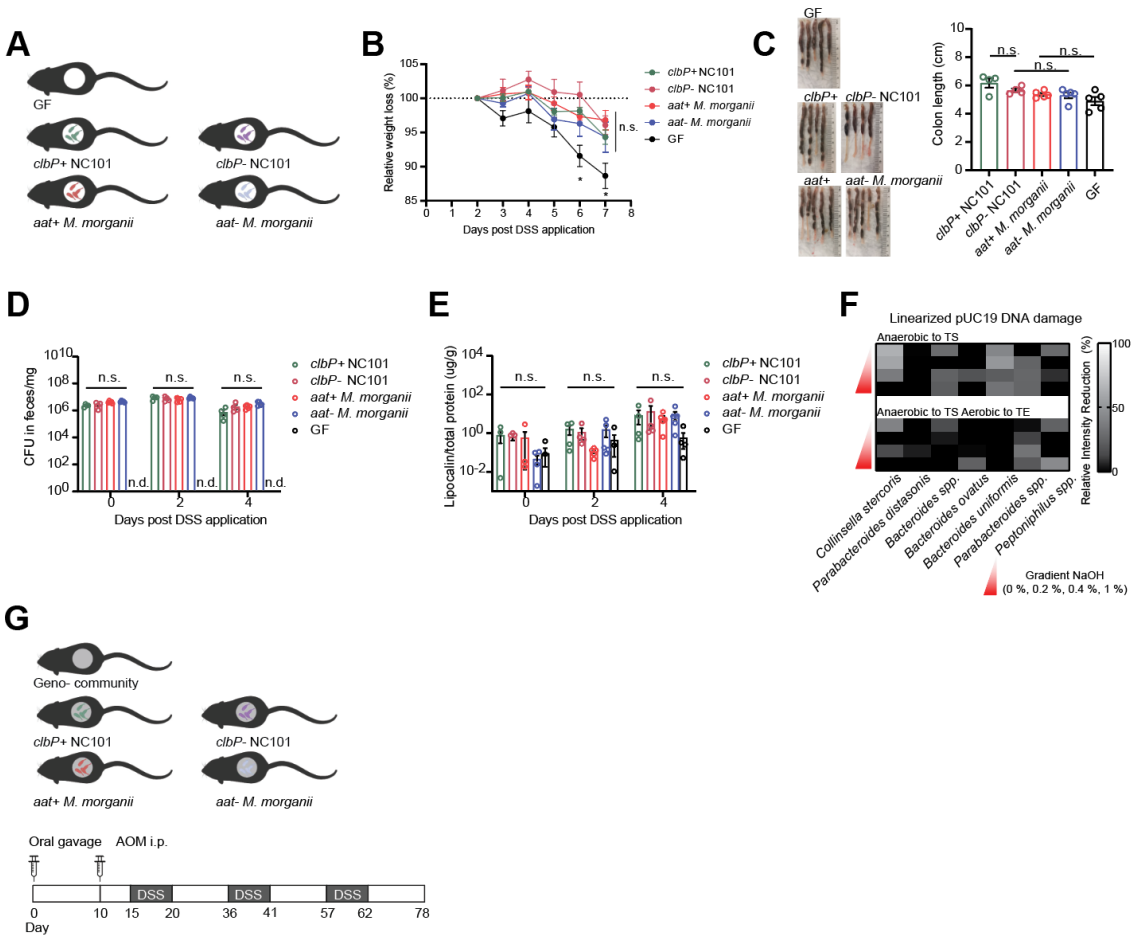


Figure 6. Indolimines produced by *M. morganii* induced tumorigenesis independent of inflammation.

(A) Schematic of experimental design for acute DSS model in age-matched germ-free (GF) mice, mice colonized with *clbP*+ NC101 *E. coli*, *clbP*- NC101 *E. coli*, *aat*+ *M. morganii* or *aat*- *M. morganii*.

(B) Relative weight loss (%) of GF or bacteria-colonizing mice in acute DSS model. n.s., not significant; * $p<0.05$, two-way ANOVA.

(C) Colon lengths on day 7 of GF or bacteria-colonizing mice in acute DSS model. n.s., not significant.

(D) CFU (colony-forming unit) on day 7 of GF or bacteria-colonizing mice in acute DSS model. n.d., no detection; n.s., not significant.

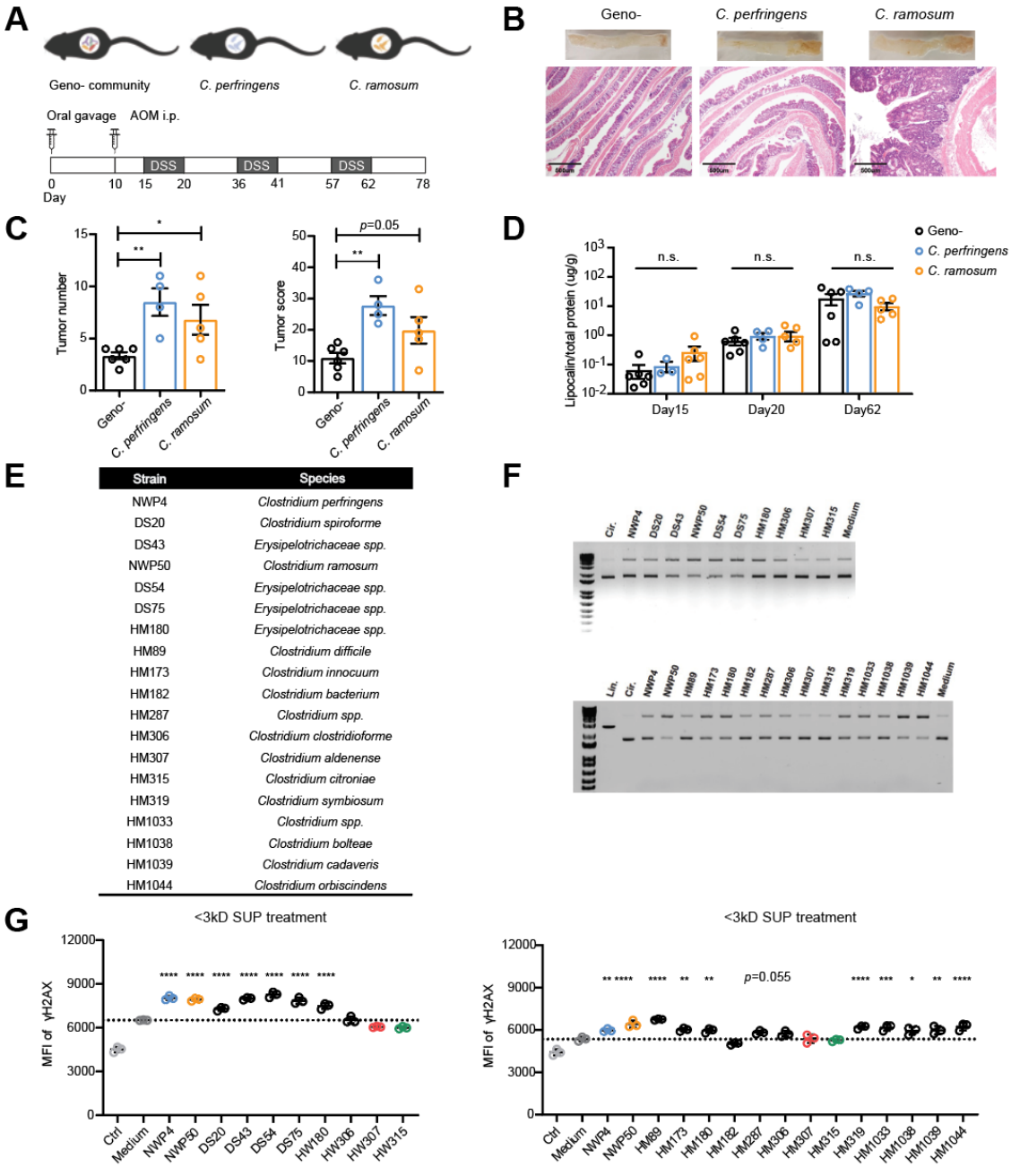
(E) Fecal lipocalin 2 levels on day 7 of GF or bacteria-colonizing mice in acute DSS model. n.s., not significant.

(F) Relative intensity reduction (RIR, %) of linearized pUC19 DNA bands after co-incubation with 7 non-genotoxic species from Fig. 1B.

(G) Schematic of experimental design for CRC induction in age-matched gnotobiotic mice colonized with *clbP*+ NC101 *E. coli*, *clbP*- NC101 *E. coli*, *aat*+ *M. morganii* or *aat*- *M. morganii* in a non-genotoxic (Geno-) community background.

Clostridium species exacerbates CRC in gnotobiotic mice

In our previous screening, *C. perfringens* and its phylogenetically closely related species, *C. ramosum* also showed genotoxicity (**Fig. 1**, **Fig. 2**). Similarly, *C. perfringens* and *C. ramosum* exacerbated tumorigenesis in AOM/DSS CRC model independent of inflammation, compared to the Geno- community (**Fig. 7A-D**). *Clostridium* species were previously found to be CRC-associated microbes, and most studies were related to the production of clostridial toxins. However, it's unclear how the small molecule metabolites derived from *Clostridium* species contribute to DNA damage and tumorigenesis. Besides multiple strains of *C. perfringens* and *C. ramosum*, we also evaluated the genotoxicity of other *Clostridium* species both in cell-free and cell-based assays. Among the 19 *Clostridium* species, most exhibited similar levels of genotoxicity both in cell-free and cell-based DNA damage assays (**Fig. 7 E-G**), except two species from healthy humans (*C. aldenense* and *C. citroniae*). We already confirmed that there were no indolimines in bacterial supernatants of *C. perfringens* and *C. ramosum*. Therefore, the identification of small-molecule genotoxins shared by those phylogenetically *Clostridium* species would expand our understandings about both microbial genotoxins and pathological effects of *Clostridium* species.



Strain	Species
NWP4	*Clostridium perfringens*
DS20	*Clostridium spiroforme*
DS43	*Erysipelotrichaceae spp.*
NWP50	*Clostridium ramosum*
DS54	*Erysipelotrichaceae spp.*
DS75	*Erysipelotrichaceae spp.*
HM180	*Erysipelotrichaceae spp.*
HM89	*Clostridium difficile*
HM173	*Clostridium innocuum*
HM182	*Clostridium bacterium*
HM287	*Clostridium spp.*
HM306	*Clostridium clostridioforme*
HM307	*Clostridium aldenense*
HM315	*Clostridium citroniae*
HM319	*Clostridium symbiosum*
HM1033	*Clostridium spp.*
HM1038	*Clostridium bolteae*
HM1039	*Clostridium cadaveris*
HM1044	*Clostridium orbiscindens*

Figure 7. *Clostridium* species exacerbates CRC in gnotobiotic mice.

(A) Schematic of experimental design for CRC induction in age-matched gnotobiotic mice colonized with Geno- community, *C. perfringens* or *C. ramosum*.

(B-D) Representative colon tissue and histology images (B), tumor number and tumor score (C), and fecal lipocalin 2 levels (D) in gnotobiotic mice colonized with Geno- community, *C. perfringens* or *C. ramosum*. Each dot represents one mouse (n = 4-6 per group), n.s., not significant; * $p<0.05$; ** $p<0.01$, one-way ANOVA.

(E) List of 19 *Clostridium* species.

(F) Assessment of DNA damage induced by <3 kDa supernatants of 19 *Clostridium* species. Evaluation of nicking of circular pUC19 DNA (top band = nicked DNA) after co-incubation for 5-6 h. Lin., linearized pUC19 DNA; Cir., circular pUC19 DNA.

(**G**) MFI of γ-H2AX in HeLa cells treated with 40 % (*v/v*) <3 kDa SUP from 19 *Clostridium* species for 5-6 h. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$, one-way ANOVA.

Other functions of genotoxins

Although most studies about genotoxins are related to the host-microbe interactions, it's also interesting to investigate the functions of genotoxins in shaping microbial community or mediating colonization resistance. Instead of attacking the host, the primary goal of bacterial genotoxins is likely to enhance competitive fitness within the context of a diverse microbiota or better survival/persistence within the host by shaping intestinal niches [69]. Indeed, recent studies have found that colibactin could target bacteria carrying prophages, inducing lytic development via the bacterial SOS response. In addition, colibactin-producing *E. coli* could also compete against pathogen *Vibrio cholerae*. It would be also promising to investigate if indolimines produced by *M. morganii* similarly mediate inter-species competition.

M. morganii random mutagenesis library

In this project, we constructed a random mutagenesis library of *M. morganii* with Ez-Tn5 transposon, consisting of ~16, 000 colonies (4 times to ~4, 000 genes in genome of *M. morganii*). Comparing homologous recombination, transposon insertion is more powerful to generate relatively stable knockout mutants, without spontaneous loss of insertion fragments when the engineered microbe doesn't harbor matched transposase. Therefore, this library opens doors to more studies related to *M. morganii* both *in vitro* and *in vivo*, including virulent factors, colonization factors, metabolic networks and inter-species colonization competition. In addition, the optimized protocol in this project could be also expanded to other species, with commercially available Ez-Tn5 kit and customized primers for sequencing.

Materials and Methods

Bacteria strains and plasmids

NWP and DS strains were isolated from 11 IBD patients (NWP) or healthy controls via anaerobic culturomics as previously described [19] and identified via 16S rRNA gene sequencing (V4 region). HMP strains were obtained from BEI Resources, NIAID, and NIH as part of the Human Microbiome Project. The *E. coli* K-12 BW25113 strains with and without the *clb* island (*clb*+ *E. coli* or *clb*- *E. coli*) were generated by the Crawford laboratory as previously described [161]. The *E. coli* NC101 wild-type and *ΔclbP* mutant strains were generated by the Jobin laboratory as previously described [74]. 25 µg/ml of chloramphenicol or 50 µg/ml kanamycin were added when appropriate.

NWP strains were cultured in Gifu Anaerobic Broth (GAM Broth) at 37 °C under anaerobic conditions in a Coy Laboratory Products Inc. chamber (10 % CO_2, 4 % H_2, 86 % N_2). Growth curves were measured for 48 h starting at OD_{600} = 0.01 using a BioTek PowerWave HT Microplate Spectrophotometer.

The 2686 bp plasmid pUC19 was purchased from New England Biolabs and linearized with the endonuclease EcoRI (New England Biolabs, 5 U/mg DNA). Linearized plasmid DNA was purified using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs) and eluted with 10 mM Tris–1 mM EDTA pH 7.5 buffer.

Media

Gifu Anaerobic Medium (GAM Broth) was purchased from Himedia Laboratories (M1801); Luria Broth Base (Miller's LB Broth Base)™ was purchased from Invitrogen (12795027); Chopped Meat Medium (CM, AS-811), Brucella Broth (BRU, AS-105), Yeast Casitone Fatty Acids Broth with Carbohydrates (YCFAC, AS-680), and MTGE Anaerobic Enrichment Broth (MTGE, AS-778) were purchased from Anaerobe Systems; BD Difco™ Lactobacilli MRS Broth (MRS, 288130), Reinforced Clostridial Medium (RCM, 218081), Malt Extract Broth (211320), and Bacto™ Brain Heart Infusion (BHI, 237500) were purchased from BD Biosciences; Todd Hewitt Broth (THB, DST47500) was purchased from DOT Scientific, Inc.; M9 minimal medium was prepared with 5x

M9 salts (30 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 5 g NH_4Cl for 1l stock), 0.4 % glucose, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$; M9-CA was prepared with M9 minimal medium supplemented with 0.2 % Bacto™ Casamino Acids (BD Biosciences, 223050); Standard amino acid complete (SACC) medium was prepared as published [162].

In vitro linear DNA gel electrophoresis assay

All bacterial strains were classified into seven groups with similar growth dynamics based on their individual growth curves. Screening experiments were designed based on approximate T_E and T_S within these seven groups. Two-day cultures of each isolate were diluted to OD_{600} = 0.01 in 250 µl liquid media and co-incubated with 1 µg linearized pUC19 plasmid DNA in 96-well deep well plates. After co-incubation, bacterial cultures were centrifuged to remove bacterial cells and supernatants were collected for linear DNA extraction using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs). Purified linear DNA samples were eluted in 20 µl DEPC-treated water.

For native gel electrophoresis, 1 µl of DNA was mixed with 6x purple gel loading dye without SDS (New England Biolabs). For denaturing conditions, 1 µl of DNA was treated with 0.2 %, 0.4 %, or 1 % NaOH denaturing buffers on ice for 30 min, then mixed with 6x purple gel loading dye. Gel electrophoresis was performed using 1 % agarose TBE gels for 1-2 h at 90 V. Gels were stained with 5000x SYBR Gold (Thermo Fisher Scientific) for 2 h at room temperature before UV visualization. DNA band intensity (~2.7 kb under native condition and ~1.3 kb under denaturing conditions) was quantified with ImageJ and the relative intensity reduction (RIR, %) resulting from DNA damage was calculated by comparing DNA band intensity after co-incubation with bacterial supernatants versus medium only controls.

Ethyl-acetate extraction of bacterial metabolites

Bacterial cultures were centrifuged to remove bacterial cells and other large particles. The resulting clarified bacterial supernatants were extracted three times using two equivalent volumes of ethyl acetate, and the organic layers were collected and dried via rotary evaporator. The dried

samples were placed under vacuum for another 6 h before weighing the dried extracts. Dried crude extracts were reconstituted with dimethyl sulfoxide as a stock solution at 500 mg/ml.

In vitro circular DNA damage assessment via gel electrophoresis

Circular pUC19 plasmid DNA (100 ng) was co-incubated with bacterial supernatants or ethyl-acetate extracts at 37 °C under aerobic conditions for 6 h. Native DNA gel electrophoresis was performed using 1 % agarose TBE gels containing ethidium bromide for 1-2 h at 90 V.

Cell culture

HeLa cells were cultured in high glucose Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich, D6429) supplemented with 10 % FBS (Sigma-Aldrich, 12306C) and 1 % Penicillin-Streptomycin (Thermo Fisher Scientific, 15140122).

To assess DNA damage caused by bacterial supernatants, anaerobic overnight bacterial cultures were collected and filtered using Amicon Ultra-0.5 Centrifugal Filter Unit, 3KDa (EMD Millipore, UFC500324). Hela cells were treated with unfractionated SUP (bacterial supernatants), <3 kDa SUP (small-molecule metabolites) or >3 kDa SUP (large-molecules) diluted 40 % (*v/v*) in culture medium supplemented with 10 mM HEPES (Thermo Fisher Scientific, 155630080) for 4-6 h.

To assess DNA damage using chemical extracts of bacterial metabolites, Hela cells were treated with 5 mg/ml ethyl-acetate extracts diluted in culture medium supplemented with 10 mM HEPES for 4-6 h.

To assess DNA damage caused by bacterial infection with colibactin-producing *E. coli*, *E. coli* cultures were incubated overnight, collected, washed with PBS and resuspended in culture medium supplemented with 10mM HEPES. Hela cells (~70 % confluent) were infected for 4 h at a multiplicity of infection of 100.

To assess DNA damage using pure compounds, Hela cells were treated with 100 µg/ml, 25 µg/ml, or 10 µg/ml indolimine-214 (**1**), compound **2**, or mixed **1**+**2** compounds (at a 4: 6 ratio) and diluted in culture medium supplemented with 10 mM HEPES for 4-6 h.

After infection or co-incubation with bacterial supernatants, HeLa cells were washed with PBS and harvested using 0.25 % Trypsin-EDTA (Sigma-Aldrich, T4049) for assessment via flow cytometry or comet assay.

Flow cytometry

For intracellular γ-H2AX staining, post-treated HeLa cells were washed with PBS once and fixed with Ebioscience™ IC Fixation Buffer (Thermo Fisher Scientific, 00822249) in the dark for 20 min at room temperature. Intracellular γ-H2AX staining (1:100 AF647 anti-phospho-H2AX; BioLegend, 613408) was conducted in Ebioscience™ Permeabilization Buffer (Thermo Scientific, 00833356) for 1 h after 20 min permeabilization at room temperature. 10,000-20,000 events per sample were collected using a CytoFLEX flow cytometer (Beckman Coulter, Inc.) and MFI (geometric mean fluorescence intensity) was analyzed with FlowJo_v10.7.1.

For cell cycle analysis, post-treated HeLa cells were washed once with PBS and fixed with 90 % ice-cold ethanol in the dark for 30 min on ice. After washing, cells were suspended in PBS containing 50 µg/ml propidium iodide and 100 µg/ml RNaseA. Cells were incubated for 30 min at 37 °C before being collected using a CytoFLEX flow cytometer (Beckman Coulter, Inc.).

Comet assay

After treatment with ethyl-acetate extracts or pure compounds, HeLa cells were washed with PBS and collected via trypsinization. DNA damage was assessed via alkaline single cell gel electrophoresis using the Trevigen Comet Assay kit (Trevigen, Inc.) according to the manufacturer's instructions. Briefly, cells were embedded in Comet LMAgarose and loaded onto CometSlide™. Slides were placed flat at 4 °C in the dark for 10min and then immersed overnight in Lysis Solution at 4 °C. After overnight incubation, slides were immersed in Alkaline Unwinding Solution for 20 min at room temperature and then subjected to electrophoresis under 15 V for 70 min at 4°C. After gently washing slides with water and 70 % ethanol, DNA was stained with Invitrogen™ SYBR™ Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific, S11494) for 30 min in the dark. Comet images were acquired using a Leica DMRB fluorescence microscope. Tail

DNA % (Tail DNA content as a percentage of Comet DNA content), Tail moment (Tail length multiplied by Tail DNA %), and Olive moment (DNA % and the distance between the intensity-weighted centroids of head and tail) were analyzed with OpenComet v1.3.1., and 30-50 comets were recorded for each sample.

Chemical isolation, identification, synthesis, and quantification

Ultraviolet/visible (UV/Vis) spectra were obtained on an Agilent 1260 Infinity system equipped with a photo diode array (PDA) detector (Agilent Technologies, Inc.). The nuclear magnetic resonance (NMR) spectroscopy data were generated at 25 °C on an Agilent 600 MHz NMR spectrometer (DD2) equipped with an inverse cold probe (3 mm), using standard NMR pulse libraries. High-performance liquid chromatography mass spectrometry (HPLC-MS) analysis was performed on an Agilent 1260 Infinity system using a Phenomenex Luna C_{18}(2) column (100 Å, 5 μm, 4.6 × 150 mm, Phenomenex) or a Phenomenex phenyl-hexyl column (100 Å, 5 μm, 4.6 × 150 mm) using the PDA detector coupled with a single quadrupole electrospray ionization mass spectrometry instrument (ESI-MS, Agilent Technologies, Inc. 6120). Purification of metabolites addressed in the current study was implemented using an Agilent Prepstar HPLC system with a preparative Agilent Polaris C_{18}-A 5 μm column (21 × 250 mm), a Phenomenex Luna C_{18}(2) column (100 Å, 10 μm, 10 × 250 mm), or an Agilent phenyl-hexyl column (100 Å, 10 μm, 10 × 250 mm). High-resolution ESI-MS (HR-ESI-MS) data were recorded on an Agilent iFunnel 6550 quadrupole time-of-flight (QTOF) MS instrument fitted with an electrospray ionization (ESI) source (positive mode) linked to an Agilent 1290 Infinity HPLC system with the aforementioned analytical columns. Experimental electronic circular dichroism spectra were obtained on a Chirascan CD spectrometer (Applied Photophysics, Inc.).

Initial untargeted metabolomics at a small scale was performed to identify *M. morganii*-derived metabolites against the rich Gifu medium components. 100 µL of each supernatant from *M. morganii* overnight cultures or Gifu medium alone (5 × 5 ml) were subjected to UPLC-QTOF-MS without extraction and the resultant raw data were processed using Mass Profiler Professional (Agilent Technologies, Inc.) or XCMS online (*53*). Metabolomics analyses led to the identification

of ~100 ion features mainly present in *M. morganii* relative to the rich Gifu medium control (initial ion list, **Table S4**). The bacterial cultivations were then pooled and concentrated, and the residue was fractionated into ~30 fractions utilizing preparative HPLC (Fr.1 to Fr.30; 5→50 % MeCN in water with 0.01 % trifluoroacetic acid (TFA) for 30 min, 8 ml/min, 1 min elution collection window). These fractions (100 µL) were analyzed utilizing UPLC-QTOF-MS and concentrated for cell-free electrophoresis assays with circular pUC19 plasmid DNA and for investigation of bioactive entities in the fractions. Fr.20–24 exhibited genotoxic activity which led these active fractions to be combined. Ions present in inactive fractions were excluded from the initial ion list, which initially suggested ~20 ion features for potential genotoxic small molecules in the active fractions. The combined active fractions (Fr.20–24) were again subjected to preparative HPLC for further fractionation (Fr.20–24–1 to Fr.20–24–30; 5→30 % MeCN in water with 0.01 % TFA for 30 min, 4 ml/min, 1 min elution collection window). The sub-fractions were assessed for their genotoxicity as described above and analyzed with UPLC-QTOF-MS. This second round of HPLC fraction screening revealed that Fr.20–24–1 to Fr.20–24–3 and Fr.20–24–5 to Fr.20–24–8 were active in the assay and the exclusion process of ions in the inactive fraction was applied to ultimately identify 4 potential genotoxic small molecules (I–IV, **Fig. 3**).

Having identified those potential hits, a large cultivation of *M. morganii* (18 liters) was performed to characterize their structures and assess genotoxicity. The supernatant was extracted twice with ethyl acetate (20 liters each) and the extracted residue was chromatographed over a preparative HPLC system (Fr.1 to Fr.60; 5→50 % MeCN in water with 0.01 % trifluoroacetic acid (TFA) for 1 h, 8 ml/min, 1 min elution collection window). Single quadrupole MS analyses showed that Fr.19 to Fr.23 possessed the 4 ion features, leading them to be pooled together (Fr.19–23). The combined fraction was further separated into 60 fractions (Fr.19–23–1 to Fr.19–23–60; 5→30 % MeCN in water with 0.01 % TFA for 1 h, 4 ml/min, 1 min elution collection window), and the sub-fractions were analyzed. We combined Fr.19–23–13 to Fr.19–23–15 into F1, Fr.19–23–17 to Fr.19–23–20 into F2, Fr.19–23–22 to Fr.19–23–23 into F3, and Fr.19–23–26 to Fr.19–23–30 into F4, based upon each pooled fraction (i.e., F1–F4) being enriched with the targeted ion features I–IV, respectively. These 4 semi-pure fractions were evaluated for their genotoxicity utilizing the

aforementioned genotoxicity assay. Chromatographic analysis of genotoxic F2 demonstrated that II (indolimine-214 (**1**), one of the targeted ion features) and another bacterial metabolite (compound **2**) were present at a ratio of 4: 6.

The structures of the two compounds were characterized as a mixture utilizing one- and two-dimensional NMR analyses, including ^{1}H NMR, correlation spectroscopy (COSY), heteronuclear single quantum correlation (HSQC), and heteronuclear multiple bond correlation (HMBC) experiments.

Indolimine-214 (**1**) was synthesized by the addition of indole-3-carbaldehyde (1 g) and isopentylamine (1 g) to methanol (10 ml) at room temperature. The reaction was screened by single quadrupole MS analysis and after 3 h, the solvent was evaporated and the residue was purified utilizing preparative HPLC application. Upon purification, NMR analysis of indolimine-214 (**1**) showed at least ~20% degradation into the two starting materials, due to the reversible reaction mechanism in water for formation of the target compound. Compound **2** was synthesized by dissolving phenethylamine (500 mg) and 5-methylhexane-2,3-dione (500 mg) in a mixture of water and methanol (1:1, 5 ml) followed by addition of sodium cyanoborohydride (1 g). The reaction mixture was warmed to 50 °C and incubated overnight. The crude mixture was concentrated and subjected to preparative HPLC to obtain compound **2**. The aforementioned semi-fraction F2 was purified using analytical HPLC application: phenyl-hexyl column (Phenomenex, 100 Å, 250 × 4.6 mm, 5 µm), 18→19% MeCN in water with 0.1% formic acid for 30 min, 0.7 mL/min, 35 injections. Pure compound **2** was then garnered and its electronic circular dichroism was measured to establish that the compound is a racemic mixture. Other indolimines (**3** and **4**) were detected in bacterial and cecum samples and their structures were validated upon synthesis employing a similar manner for indolimine-214 (**1**).

All NMR structure raw data are shown in **Appendices Figure S10-41**.

The UPLC-QTOF-MS-based quantification of indolimines was performed by a standard curve using synthetic standards followed by integration of ion counts. For quantification in bacterial production, synthetic standards were added to a Gifu medium and extracted in an identical manner to bacterial supernatants to account for extraction efficiencies in the quantification

workflow. To quantify the metabolites in cecum contents, collected materials were dried and resuspended with methanol and water (1:1, 2 mL). The supernatants were concentrated and resuspended with methanol and water (1:1, 100 µL) and analyzed utilizing QTOF-MS. The synthetic standards were prepared and analyzed in the identical mixture of solvents.

Mice studies

Azoxymethane (AOM, Millipore Sigma, A5486) was dissolved in PBS at 2 mg/ml. Dextran sulfate sodium (DSS) was purchased from TdB Labs AB (batch no. DB001-42).

Age- and sex-matched 5-8 week old germ-free wild-type C57BL/6 mice were used for all studies. All mice were bred and maintained at the Yale University School of Medicine and all treatments were approved by the Yale Animal Care and Use Committee (IACUC) (IACUC protocol number: 11513, the Yale Animal Welfare Assurance number: D16-00146).

Germ-free mice were colonized via oral gavage with individual bacterial cultures or mixed bacterial communities (~10^8 CFU in 100 ul) grown under anaerobic conditions and stored in a sealed glass vial until immediately prior to oral gavage. For acute DSS mode, after 1-2 weeks, mice were treated with 2% DSS in the drinking water for 7 days. Fecal samples were collected on day 0, 2, 4 and 7. Mice were weighted every day. On day 7, mice were sacrificed, colon tissues were collected, measured for length and then fixed in Bouin's buffer for 24 h. Fixed tissues were transferred into 70% EtOH and sent to Histowiz for H&E staining and pathology consultation.

For AOM/DSS CRC model, after 10 days, wild-type mice received intraperitoneal injections of AOM (10 mg/kg). 5 days after the first AOM injection, mice were treated with 2 % DSS in the drinking water for 5 days, followed by 16 days of untreated, sterile water. This cycle was repeated twice with 1.5 % DSS. On day 78, fecal samples were collected and mice were sacrificed, colon tissues were collected, and tumor number and overall tumor load were recorded. Macroscopically-visible tumors were enumerated for each mouse and tumor load was calculated as a sum of tumor scores (graded by size of all tumors) per mouse. Tumor sizes were graded from 1 to 5 as follows: Grade 1, very small but detectable tumor; Grade 2, tumor covering up to one-eighth of colonic circumference; Grade 3, tumor covering up to one-fourth of the colonic

circumference; Grade 4, tumor covering up to half of the colonic circumference; and Grade 5, tumor covering more than half of the colonic circumference.

Colon tissues were swiss-rolled (lumen side out) and fixed in 4 % PFA (J.T.Baker®™, S89807) in PBS overnight. Paraffin embedding and H&E staining was performed by Yale Pathology Tissue Services (YPTS).

Neonatal colonization was performed after birth. Briefly, pups were orally instilled with ~1×10^8 colony forming units (CFU) of bacteria 5-6 times within 1 week. Genotyping was performed at weaning to identify *Apc*^{Min} or wild-type mice. Mice were aged until 14 weeks for tumor enumeration. Colons and small intestines were isolated, and tumors were counted macroscopically. Blinded assessment of H&E-stained tumor tissues was performed to identify aberrant crypt foci (ACF), adenomas (ADE), and adenocarcinomas (ACA).

Bulk RNA-seq of colonic epithelial cells

Mice were orally gavaged with indolimine-214 (**1**) or DMSO every day for 5 days (indolimine-214 was dissolved in DMSO at 100mg/ml, gavage 50ul). On day 5, colon tissues were dissected and opened longitudinally, washed with ice-cold PBS, cut into small pieces and digested in 20 ml DPBS+2mM EDTA at 100 rpm 37°C for 15min in the shaker. Then add BSA to stop isolation, vigorously shake the tube by hand for 30s and filter the cells with 70 μm membrane. Cells were centrifuged and collected, then RNA was extracted with RNeasy Mini Kit (Qiagen, 74106) and DNase on-column digestion following the protocol. Colon RNAseq libraries were prepared following the protocols of iScript cDNA Synthesis Kit (Bio-rad, 1708890) and Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus (Illumina, 20040529). Quality of libraries was checked and HiSeq paired-end, 100 bp sequencing was perfromed by Yale Center for Genome Analysis staff.

Sequencing data were trimmed, aligned, and gene counts quantified using Partek Flow (v6.0) [163]. Upregulated genes in indolimine-214 treatment group were determined with a threshold of 0.4 on log2 (fold change) and *p* value cutoff of 0.05. Then the gene list was analyzed for GO

enrichment using statistical over-representation test in Panther v14 available at http://geneontology.org.

ELISA

Fecal pellets were weighed, dissolved in 1 mL PBS per 100 mg fecal material, and disrupted via beadbeating (MP Biomedicals, Lysing Matrix D, 6913) for 10 s in a Biospec BeadBeater. Fecal water samples were collected after spinning at 10,000g for 5 min.

Fecal lipocalin-2 ELISAs were performed using the Mouse Lipocalin-2/NGAL DuoSet ELISA kit per the manufacturer's instructions (R&D Systems, #DY-1857). Briefly, plates were coated for 2 h at 37 °C with 50 µL of Capture Antibody diluted in PBS. After blocking for 1 h at room temperature with 100 µL of 1 % BSA/PBS, 50 µL fecal water samples or standards were incubated for 2 h at room temperature. 50 µL of Detection Antibody diluted in 0.1 % BSA/PBS was incubated for 1 h at room temperature followed by Streptavidin-HRP diluted 1:200 in 0.1 % BSA/PBS for 30 min at room temperature.

Absorbance was measured at 450 nm after development using the Pierce TMB Substrate Kit (Thermo Fisher Scientific, 34021) and 2M H_2SO_4.

Total protein levels were measured using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225) according to the manufacturer's instructions.

Biosynthetic gene cluster (BGC) analysis

Raw whole-genome sequence reads for all bacterial genomes have been deposited in the NCBI Sequence Read Archive (SRA) database in FASTQ format. Genome assemblies were performed as described in our previous study (*54*). Briefly, all Illumina paired-end reads were filtered and trimmed using Trimmomatic v.0.38 (*55*) with the following parameters: ILLUMINACLIP: NexteraPE-PE.fa:2:30:12:1:true LEADING:3 TRAILING:3 MAXINFO:40:0.994 MINLEN:36. The four output files after trimming included two (forward and reverse) FASTQ files with paired reads and two FASTQ files with unpaired reads. All four files from each strain were assembled into contigs using SPAdes 3.13.0 [164] with the default parameters for paired-end libraries.

Assembled contigs were input into the antiSMASH portal (https://antismash.secondarymetabolites.org/#!/start) for biosynthetic gene cluster (BGC) exploration with default settings (Detection strictness: loose). Results are summarized in **Table S3**.

Gene identification and IPTG-inducing indolimine synthesis

The 3 potential proteins were found after NCBI blast with the AA sequence of valine decarboxylase in *Streptomyces viridifaciens* and contained >60% similarity (**Table S5**). The codon-optimized DNA sequences (IDT) of Peg1085, Peg1320 and Peg3098 were synthesized by IDT. Add BamHI and HindIII restriction enzyme sites on both ends, synthesize gBlocks and construct over-expression system with pET-28 plasmids. Then the plasmids were separately electrotransformed into *E. coli* BL21(DE3) (NEB, C2527I).

Valine decarboxylase:

MSTSSASSGPDLPFGPEDTPWQKAFSRLRAVDGVPRVTAPSSDPREVYMDIPEIPFSKVQIPPD
GMDEQQYAEAESLFRRYVDAQTRNFAGYQVTSDLDYQHLSHYLNRHLNNVGDPYESSSYTLNS
KVLERAVLDYFASLWNAKWPHDASDPETYWGYVLTMGSSEGNLYGLWNARDYLSGKLLRRQH
REAGGDKASVVYTQALRHEGQSPHAYEPVAFFSQDTHYSLTKAVRVLGIDTFHSIGSSRYPDEN
PLGPGTPWPTEVPSVDGAIDVDKLASLVRFFASKGYPILVSLNYGSTFKGAYDDVPAVAQAVRDI
CTEYGLDRRRVYHDRSKDSDFDERSGFWIHIDAALGAGYAPYLQMARDAGMVEEAPPVFDFRL
PEVHSLTMSGHKWMGTPWACGVYMTRTGLQMTPPKSSEYIGAADTTFAGSRNGFSSLLLWDY
LSRHSYDDLVRLAADCDRLAGYAHDRLLTLQDKLGMDLWVARSPQSLTVRFRQPCADIVRKYSL
SCETVYEDNEQRTYVHLYAVPHLTRELVDELVRDLRQPGAFTNAGALEGEAWAGVIDALGRPDP
DGTYAGALSAPASGPRSEDGGGS

Peg1085 Glutamate decarboxylase:

MNQSELITLASANALNKDFEVKYQNVISDFFSRDPGKWPIFNHPQIQAITQFRQTTDADVQQINRY
PQGKDLFAQLAGESHVRQNVIRPGEGQDDLLVFASALCKNWENPLAVENVIAMPSDPAVYGSM
LGLLGNPNMVYCEYSGVADNMEKTVIRKVANLIGYDADKASGLFTQGGTMCNLYGYLFGIRKSL
KQSKHLGMSVDQDFRIINSQGGHYSNMTNLSLLGVDITNKTIRIRVASDNTIDLADLEQQIRACYT
VHCKIPVILLTAGTTDTFGVDEIKQVYDLRNRLCEEFEITEKPHIHVDAAVGWPIIFFIDYDFNTNPL
AINDATLAGLRHNVEKFKQLKYADSITIDFHKWGYVPYTSSLVMVRDGDDFKALENDPENFTYFE
HALEGQTHLQSTIECSRSGVGVFGAYAGLHYLGVEGYQTIIAHCLQNANYMRNQLLSMGNACVM
VPENQGPSVGFRLYSPKLVNDPQAMFAHELTCAGDKTAYDMMVRNSRWHRELFLKRGKAGLF
TNWVDSIACSAYAEHNRFAYIPGEKAVFMNPVTQRHHIDAFAKMLKTMSAE

ATGAATCAATCGGAATTGATTACACTGGCCTCCGCGAATGCTCTGAACAAAGACTTTGAAGT
AAAATATCAAAACGTGATTTCCGACTTCTTTTCCCGTGACCCTGGAAAATGGCCCATTTTCAA
TCACCCCCAAATCCAAGCAATTACACAGTTCCGTCAGACCACAGATGCGGATGTACAGCAGA
TCAATCGCTACCCTCAGGGTAAGGACTTGTTCGCGCAACTGGCAGGCGAATCGCACGTCCG
TCAGAACGTTATCCGTCCAGGTGAAGGTCAGGACGATTTACTTGTGTTCGCATCAGCCTTAT
GCAAAAATTGGGAGAACCCCTTAGCCGTAGAGAACGTCATTGCTATGCCAAGCGACCCAGC
GGTCTATGGCTCCATGTTGGGTTTATTAGGAAATCCGAACATGGTATACTGTGAATATTCTGG

CGTAGCAGACAATATGGAGAAAACCGTGATTCGTAAGGTGGCAAATTTAATTGGATACGATG
CTGACAAAGCTAGCGGCTTGTTCACCCAGGGGGGAACTATGTGTAATTTGTATGGATACTTA
TTTGGTATTCGCAAATCGCTTAAACAAAGTAAGCATTTAGGCATGTCAGTCGACCAGGATTTT
CGTATTATCAATTCGCAGGGGGGTCACTACAGCAACATGACTAATCTTTCTCTTCTTGGGGT
AGACATCACGAACAAAACTATCCGCATCCGTGTGGCCAGCGATAACACTATTGATCTTGCGG
ACCTTGAGCAACAAATTCGTGCCTGTTACACTGTTCATTGCAAAATTCCTGTCATCTTGCTTA
CCGCTGGGACAACCGATACGTTTGGTGTTGACGAAATCAAACAAGTCTACGATCTGCGCAAT
CGTCTTTGCGAGGAGTTCGAGATCACGGAGAAACCGCATATCCATGTAGACGCTGCGGTTG
GGTGGCCAATCATTTTCTTTATCGACTATGATTTCAATACGAATCCCCTTGCTATTAATGATG
CAACCTTGGCAGGGTTGCGCCATAATGTAGAGAAATTCAAGCAATTAAAATATGCGGATTCC
ATTACAATCGACTTCCACAAATGGGGCTACGTCCCATACACGAGCTCCCTGGTTATGGTGCG
CGACGGAGACGATTTCAAGGCGTTAGAAAATGATCCTGAAAATTTTACATATTTTGAACATGC
GCTTGAAGGGCAAACACATTTACAATCAACAATTGAATGTAGCCGCTCGGGCGTAGGGGTG
TTCGGAGCGTACGCGGGTCTTCACTACCTTGGTGTAGAGGGCTATCAGACTATCATTGCCCA
CTGTCTGCAAAACGCGAATTATATGCGTAACCAGCTGTTATCAATGGGGAACGCCTGCGTCA
TGGTCCCTGAGAACCAAGGTCCCTCCGTCGGTTTCCGTTTATACTCACCAAAACTGGTCAAC
GACCCGCAAGCTATGTTTGCTCACGAATTAACGTGTGCTGGGGATAAGACCGCTTACGACAT
GATGGTGCGTAATTCACGTTGGCACCGCGAGTTATTCCTGAAACGCGGAAAGGCCGGTTTG
TTCACTAACTGGGTGGACTCTATCGCTTGTAGTGCCTATGCTGAGCACAACCGCTTTGCCTA
TATTCCGGGAGAAAAGGCAGTCTTCATGAACCCTGTTACCCAACGCCATCACATCGACGCGT
TCGCTAAGATGCTGAAGACGATGTCAGCCGAGTAA

Peg1320 Histidine decarboxylase (EC 4.1.1.22):

MTLSINDQNKLDAFWAYCVKNQYFNIGYPESADFDYTNLERFLRFSINNCGDWGEYCNYLLNSF
DFEKEVMEYFADLFKIPFEQSWGYVTNGGTEGNMFGCYLGREIFPDGTLYYSKDTHYSVAKIVKL
LRIKSQVVESLPNGEIDYDDLMKKIADDKEAHPIIFANIGTTVRGAIDDIAEIQKRLKAAGIKREDYYL
HADAALSGMILPFVDDAQPFTFADGIDSIGVSGHKMIGSPIPCGIVVAKKENVDRISVEIDYISAHD
KTITGSRNGHTPLMLWEAIRSHSTEEWKRRITRSLDMAQYAVDRMQKAGINAWRNKNSITVVFP
CPSERVWREHCLATSGDVAHLITTAHHLDTAQIDKLIDDVIADFNLHAA

ATGACCTTGTCCATTAATGACCAAAACAAACTTGATGCTTTCTGGGCGTACTGTGTTAAAAAT
CAATACTTTAATATCGGCTATCCAGAATCGGCGGACTTCGATTACACTAATTTAGAACGTTTT
TTGCGCTTTTCGATCAACAATTGTGGTGACTGGGGTGAATACTGTAATTACTTATTGAATAGC
TTTGACTTCGAAAAGGAAGTCATGGAGTATTTTGCAGACTTATTCAAGATCCCCTTCGAGCAG
TCTTGGGGCTATGTCACTAACGGCGGGACAGAAGGCAATATGTTTGGATGTTATCTGGGCC
GCGAAATTTTCCCAGATGGTACACTTTATTACAGTAAGGATACGCATTACTCCGTGGCAAAAA
TTGTAAAACTTTTACGTATCAAAAGCCAAGTAGTAGAGTCCTTACCAAACGGGGAAATTGATT
ATGACGATCTTATGAAGAAGATTGCTGATGACAAAGAGGCACATCCAATCATTTTCGCAAATA
TCGGAACAACAGTTCGCGGGGCCATCGATGATATTGCCGAAATTCAGAAACGCCTTAAGGC
GGCGGGAATCAAACGTGAAGATTACTATTTGCACGCAGACGCCGCTTTGTCCGGGATGATTT
TGCCTTTTGTAGACGATGCCCAGCCATTCACCTTCGCAGATGGAATTGACTCAATTGGAGTA
AGTGGCCACAAAATGATCGGATCACCTATTCCTTGTGGGATTGTCGTCGCGAAAAAGGAGAA
TGTTGACCGTATTTCCGTGGAGATCGACTACATTTCGGCGCATGATAAAACGATTACTGGAT
CTCGCAACGGACACACACCCCTTATGTTGTGGGAGGCTATCCGCTCGCATTCGACCGAGGA
GTGGAAGCGTCGCATCACACGCTCACTGGACATGGCTCAGTATGCCGTTGACCGTATGCAG
AAGGCGGGCATCAACGCCTGGCGCAACAAAAACTCCATTACAGTGGTCTTCCCTTGTCCCT
CTGAGCGTGTATGGCGCGAGCACTGTCTTGCGACTAGCGGTGATGTTGCCCACCTTATTAC
AACCGCGCACCACTTGGACACGGCGCAAATTGACAAGCTGATCGATGACGTTATTGCCGAC
TTTAACTTACATGCCGCTTAA

Peg3098 Glutamate decarboxylase (EC 4.1.1.15):

MSLHAVKGKNSSEFIDIYASTDSDAKLPKYKMPDDSSDPRIIYSVVRDELLLDGNSRQNLATFCTT
WVEDEVKQLMTDSVDKNMIDKDEYPQTAEIESRCVHIIADLWNSPQAQETIGCSTTGSSEAAML
GGLAMKWRWRKNREKQGKETGKPNLVTGPVQVCWEKFARYFDVELRQIPLEGDALGMQPSDL

RKYCDENTIGVVATLGVTFTGIYEPVAELAKELDAIQRDTGLDIPLHVDGASGGFIAPFIQPELVWD
FRIERVKSINSSGHKYGLSPLGVGWVVWRSKEDLPEELVFNVDYLGGNMPTFALNFSRPGGQIIA
QYYNFLRLGRAGYTKIQQACADTAQWLADELNKLGIFDLVYDGRGALPAVAYKLKPGVTQFNLY
DLSDRIRTRGWLIASYPLPADREKTVVQRIMIRHGVSRDLAALLLDDIKRAIDHFRQNPVVNSTAK
ATFHHG

ATGTCTCTTCATGCCGTCAAAGGCAAAAACTCGAGTGAGTTTATTGACATCTATGCCTCGAC
GGATAGCGACGCTAAGTTACCAAAATATAAAATGCCTGATGACTCATCTGACCCACGCATCA
TTTATTCAGTCGTCCGCGATGAGCTTCTTCTGGACGGCAATTCCCGTCAAAACTTAGCCACT
TTTTGCACCACATGGGTAGAGGACGAGGTTAAACAGTTAATGACAGACTCAGTTGACAAAAA
CATGATTGATAAGGATGAGTACCCCCAAACGGCCGAGATTGAAAGCCGCTGTGTTCACATCA
TCGCTGACTTGTGGAATTCACCCCAAGCACAAGAGACCATCGGGTGCTCTACCACTGGTTCT
TCCGAAGCTGCGATGTTAGGGGGGTTAGCTATGAAATGGCGCTGGCGTAAGAACCGTGAAA
AGCAAGGTAAGGAGACAGGAAAACCGAATTTAGTAACCGGACCCGTACAAGTTTGCTGGGA
GAAATTCGCACGTTACTTCGATGTTGAATTACGTCAAATTCCCCTTGAGGGTGACGCCTTGG
GTATGCAACCAAGCGACCTTCGCAAGTATTGCGATGAGAATACTATTGGGGTAGTGGCGAC
GCTGGGAGTTACGTTTACCGGGATCTACGAACCCGTAGCAGAATTGGCTAAGGAGTTGGAC
GCTATCCAGCGCGACACCGGACTGGACATTCCATTGCATGTGGATGGCGCCTCGGGGGGG
TTCATCGCACCATTTATTCAACCTGAATTAGTATGGGACTTTCGTATTGAACGCGTCAAATCA
ATCAACTCCTCTGGTCACAAGTACGGTTTGAGTCCATTGGGGGTTGGCTGGGTAGTTTGGC
GCTCCAAAGAAGATCTTCCCGAGGAATTGGTGTTCAATGTCGATTATTTGGGGGGCAATATG
CCTACTTTCGCCTTAAATTTTAGCCGTCCAGGTGGCCAAATTATCGCTCAATATTATAATTTC
CTTCGTTTGGGGCGTGCGGGGTATACTAAGATCCAACAAGCGTGCGCCGACACAGCTCAGT
GGCTGGCGGACGAGCTTAATAAATTAGGGATCTTCGATTTAGTTTACGATGGTCGTGGGGCA
CTTCCTGCTGTAGCCTATAAACTTAAACCGGGTGTTACCCAGTTCAATCTGTATGATTTAAGT
GACCGTATTCGCACTCGCGGATGGTTAATTGCCAGCTACCCATTGCCGGCTGATCGTGAAA
AAACAGTAGTGCAGCGTATCATGATTCGCCATGGTGTTAGCCGCGACCTTGCAGCTTTGTTG
TTGGATGACATTAAACGTGCTATTGACCATTTTCGCCAAAACCCCGTAGTGAATTCTACTGCT
AAAGCGACGTTCCACCACGGTTAA

Primers for pET-28 system:

Peg1085	tacgtGGATCCATGAATCAATCG
	tcgatAAGCTTTTACTCGGCT
Peg1320	tacgtGGATCCATGACCTTG
	tcgatAAGCTTTTAAGCGGCA
Peg3098	tacgtGGATCCATGTCTCTTCA
	tcgatAAGCTTTTAACCGTGGT

To induce production of indolimines from transformed *E. coli*, single colonies were picked and grown overnight at 37 °C in LB with carbenicillin. Dilute 1:100 in 2 ml TB medium with Carb and grow 3-4 h at 37 °C. Prepare 1 ml TB+Carb+1 mM IPTG+1 mM IAld+1 mM amino acids (leucine, valine or phenylalanine) and prewarm to 37 °C before use. After 3-4 h remove 1 ml at 37 °C and collect supernatants as pre-IPTG samples. Add prewarmed 1 ml TB+Carb+1 mM IPTG+1 mM

IAld+1 mM amino acids and incubate at room temperature overnight for slow induction. Then collect supernatants as post-IPTG samples and bring pre-IPTG and post-IPTG samples for QTOF-MS based identification of indolimines.

Random mutagenesis library construction

The random mutagenesis library of *M. morganii* NWP135 was constructed with EZ-Tn5™ <R6Kγori/KAN-2>Tnp Transposome™ Kit (Lucigen, TSM08KR) following the manufacture's protocol. Briefly, *M. morganii* NWP135 was inoculated into medium from single colony and cultured overnight. Electroporation materials including transposome, electroporation buffer (0.5 M sucrose) and 2mm Gene Pulser/MicroPulser Cuvettes (Bio-rad, 1652086) were put on ice in advance. Overnight growth culture was diluted into OD600=0.01 and incubated into exponential stage in Gifu (3~4h). 15 ml bacteria was collected through centrifuging at 6,500 g for 10 min at 4 °C, washed 3 times with ice-cold 0.5 M sucrose and resuspend in 100 µl 0.5 M sucrose (x300 concentrated). 1 µl Transposome (~100 ng transposon DNA) was added into bacteria and incubated on ice for 30 min-1 h. Then the mixture was transferred into 2mm cuvettes and electroporated at 2500 V, 25 µF, 200 Ω. The mixture was immediately transferred into 900 µl pre-warmed SOC medium and recover bacteria for 3 generation (~2 h). The recovered bacteria was plated on ~25 LB agar plates with kanamycin and incubated in overnight (do not incubate for longer time to avoid false-positive colonies) at 37 °C.

~150 ml fresh *M. morganii* cultures were used to make ~250 agar plates. Then we used automatic colony picker (Molecular Devices QPix 420) to pick ~16, 000 single colonies into 384-well plates containing LB (Kan+) medium.

NGS-based mutant identification

Overnight-growth bacterial cultures were mixed with 40% glycerol/LB at 1:1 to make frozen stocks. Fresh bacterial cultures were performed combinatory pooling with Biomek NX^{P} Automated Liquid Handler (Beckman Coulter) and NGS-based sequencing as Sudoku [153, 154]. Briefly, 40 384-well plates containing fresh bacterial cultures were aligned into 5 rows and 8 columns. Then

~16, 000 single colonies were volume-equally pooled into 5 plate row libraries (PR1-PR5), 8 plate column libraries (PC1-PC8), 16 row libraries (R1-R16) and 24 column libraries (C1-C24), 54 libraries in total with H_2O as a blank control library. Genomic DNA from pooled libraries were extracted with DNeasy UltraClean Microbial Kit (Qiagen, 12224-250). 200 ng DNA from every library was used to construct amplicon sequencing libraries through a two-step hemi-nested PCR reaction with customized primers [165]. The first reaction amplified a portion of the genome adjacent to the Tn5 transposon with forward Tnseq1.1 and reverse Tnseqarb1 and Tnseqarb2 primers. The second reaction added Illumina universal adaptor and xN (random sequences to avoid saturation of any color channel) with 4 universal forward primers, and Illumina index sequence, diverse barcode sequence (unique barcode for every library) and flow-cell-binding sequence with reverse primers. All PCR reactions were performed with OneTaq DNA polymerase (NEB, M0480L) following the protocol of Sudoku. 5 µg pooled PCR products were purified at 0.5-1 kb size by gel extraction with Monarch DNA Gel Extraction Kit (NEB, T1020S). After DNA quantification with Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854), the library was diluted into 2-4 nM for denaturation and 10 pM for loading into Miseq Reagent kit v3 (Illumina, MS-102-3001). Sequencing was performed in single-index, 100-bp single-end read mode with Miseq machine (Illumina).

Primers:

PCR 1 Primers

Tnseq1.1	ACCTACAACAAAGCTCTCATCAACC
Tnseqarb1	GAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNGATAT
Tnseqarb2	GAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNACGC

PCR 2 Primers

Forward ID	Primer sequence
Tnseq-4N'	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGGTTGAGATGTGTATAAGAGACAG
Tnseq-5N'	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGGGTTGAGATGTGTATAAGAGACAG
Tnseq-6N'	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNGGGTTGAGATGTGTATAAGAGACAG
Tnseq-7N'	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNGGGTTGAGATGTGTATAAGAGACAG
Reverse ID	Primer sequence

	CAAGCAGAAGACGGCATACGAGATAACGATGGGTGACTGGAGTTCA
BC-Adaptor_01	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAAGCGCAAGTGACTGGAGTTCA
BC-Adaptor_02	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAATTGCCGGTGACTGGAGTTCAG
BC-Adaptor_03	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACAGCTCAGTGACTGGAGTTCAG
BC-Adaptor_04	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACCTTGGAGTGACTGGAGTTCAG
BC-Adaptor_05	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACGAAGGTGTGACTGGAGTTCA
BC-Adaptor_06	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACGAGATGGTGACTGGAGTTCA
BC-Adaptor_07	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACGTCAACGTGACTGGAGTTCAG
BC-Adaptor_08	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACGTTCCTGTGACTGGAGTTCAG
BC-Adaptor_09	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACTCGAGTGTGACTGGAGTTCAG
BC-Adaptor_10	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATACTGACTGGTGACTGGAGTTCAG
BC-Adaptor_11	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGACCACTGTGACTGGAGTTCAG
BC-Adaptor_12	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGACTGTCGTGACTGGAGTTCAG
BC-Adaptor_13	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGAGACACGTGACTGGAGTTCA
BC-Adaptor_14	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGCAGAACGTGACTGGAGTTCA
BC-Adaptor_15	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGCATCGTGTGACTGGAGTTCAG
BC-Adaptor_16	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGGTGTTCGTGACTGGAGTTCAG
BC-Adaptor_17	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGTCTCAGGTGACTGGAGTTCAG
BC-Adaptor_18	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATAGTGGACAGTGACTGGAGTTCA
BC-Adaptor_19	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATATATCGGCGTGACTGGAGTTCAG
BC-Adaptor_20	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATATCGCCTTGTGACTGGAGTTCAG
BC-Adaptor_21	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATATGCGGTTGTGACTGGAGTTCAG
BC-Adaptor_22	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATATGCTTCCGTGACTGGAGTTCAG
BC-Adaptor_23	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATATGGAAGCGTGACTGGAGTTCA
BC-Adaptor_24	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCAACCTTCGTGACTGGAGTTCAG
BC-Adaptor_25	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCAACTGCAGTGACTGGAGTTCAG
BC-Adaptor_26	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCACACTCTGTGACTGGAGTTCAG
BC-Adaptor_27	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCACTTCACGTGACTGGAGTTCAG
BC-Adaptor_28	ACGTGTGCTCTT

	CAAGCAGAAGACGGCATACGAGATCAGAGACAGTGACTGGAGTTCA
BC-Adaptor_29	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCAGTAGAGGTGACTGGAGTTCA
BC-Adaptor_30	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCAGTCTGAGTGACTGGAGTTCAG
BC-Adaptor_31	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCATCGTAGGTGACTGGAGTTCAG
BC-Adaptor_32	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCATGCATGGTGACTGGAGTTCAG
BC-Adaptor_33	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCCGGTTAAGTGACTGGAGTTCAG
BC-Adaptor_34	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCCTACGTTGTGACTGGAGTTCAG
BC-Adaptor_35	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCCTTAAGGGTGACTGGAGTTCAG
BC-Adaptor_36	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCGAATACCGTGACTGGAGTTCAG
BC-Adaptor_37	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCGATCCTTGTGACTGGAGTTCAG
BC-Adaptor_38	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCGCCATAAGTGACTGGAGTTCAG
BC-Adaptor_39	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCGTACCAAGTGACTGGAGTTCAG
BC-Adaptor_40	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTAGAGGAGTGACTGGAGTTCA
BC-Adaptor_41	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTAGGTACGTGACTGGAGTTCAG
BC-Adaptor_42	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTCTGACTGTGACTGGAGTTCAG
BC-Adaptor_43	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTGATGTGGTGACTGGAGTTCAG
BC-Adaptor_44	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATCTTCAGCTGTGACTGGAGTTCAG
BC-Adaptor_45	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGAACACGAGTGACTGGAGTTCA
BC-Adaptor_46	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGAAGCAACGTGACTGGAGTTCA
BC-Adaptor_47	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGAAGGTTGGTGACTGGAGTTCA
BC-Adaptor_48	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGACTCAGTGTGACTGGAGTTCAG
BC-Adaptor_49	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGAGAAGTCGTGACTGGAGTTCA
BC-Adaptor_50	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGATGAGCAGTGACTGGAGTTCA
BC-Adaptor_51	GACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGATGTCGTGTGACTGGAGTTCAG
BC-Adaptor_52	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGCAACCTAGTGACTGGAGTTCAG
BC-Adaptor_53	ACGTGTGCTCTT
	CAAGCAGAAGACGGCATACGAGATGCAATTCGGTGACTGGAGTTCAG
BC-Adaptor_54	ACGTGTGCTCTT

Barcode assignment:

Pool	Barcode	Reverse Complement	Pool	Barcode	Reverse Complement
R1	AACGATGG	CCATCGTT	C12	CACTTCAC	GTGAAGTG
R2	AAGCGCAA	TTGCGCTT	C13	CAGAGACA	TGTCTCTG
R3	AATTGCCG	CGGCAATT	C14	CAGTAGAG	CTCTACTG
R4	ACAGCTCA	TGAGCTGT	C15	CAGTCTGA	TCAGACTG
R5	ACCTTGGA	TCCAAGGT	C16	CATCGTAG	CTACGATG
R6	ACGAAGGT	ACCTTCGT	C17	CATGCATG	CATGCATG
R7	ACGAGATG	CATCTCGT	C18	CCGGTTAA	TTAACCGG
R8	ACGTCAAC	GTTGACGT	C19	CCTACGTT	AACGTAGG
R9	ACGTTCCT	AGGAACGT	C20	CCTTAAGG	CCTTAAGG
R10	ACTCGAGT	ACTCGAGT	C21	CGAATACC	GGTATTCG
R11	ACTGACTG	CAGTCAGT	C22	CGATCCTT	AAGGATCG
R12	AGACCACT	AGTGGTCT	C23	CGCCATAA	TTATGGCG
R13	AGACTGTC	GACAGTCT	C24	CGTACCAA	TTGGTACG
R14	AGAGACAC	GTGTCTCT	PR1	CTAGAGGA	TCCTCTAG
R15	AGCAGAAC	GTTCTGCT	PR2	CTAGGTAC	GTACCTAG
R16	AGCATCGT	ACGATGCT	PR3	CTCTGACT	AGTCAGAG
C1	AGGTGTTC	GAACACCT	PR4	CTGATGTG	CACATCAG
C2	AGTCTCAG	CTGAGACT	PR5	CTTCAGCT	AGCTGAAG
C3	AGTGGACA	TGTCCACT	PC1	GAACACGA	TCGTGTTC
C4	ATATCGGC	GCCGATAT	PC2	GAAGCAAC	GTTGCTTC
C5	ATCGCCTT	AAGGCGAT	PC3	GAAGGTTG	CAACCTTC
C6	ATGCGGTT	AACCGCAT	PC4	GACTCAGT	ACTGAGTC
C7	ATGCTTCC	GGAAGCAT	PC5	GAGAAGTC	GACTTCTC
C8	ATGGAAGC	GCTTCCAT	PC6	GATGAGCA	TGCTCATC
C9	CAACCTTC	GAAGGTTG	PC7	GATGTCGT	ACGACATC
C10	CAACTGCA	TGCAGTTG	PC8	GCAACCTA	TAGGTTGC
C11	CACACTCT	AGAGTGTG	H2O	GCAATTCG	CGAATTGC

Sequencing results from Miseq were used to identify a mutant *M. morganii* strain with transposon-inserted *aat* gene. Briefly, the raw sequences were firstly trimmed to discard Ez-Tn5 transposon sequences with cutadapt

$cutadapt -j 10 –g GGGTTGAGATGTGTATAAGAGACAG -o *.fastq.gz *.fastq.gz

Then the trimmed sequences were blasted to *aat* gene of *M. morganii* NWP135 with Geneious Prime. One out of ~16, 000 colonies was found to be blasted with *aat* gene, indicating that the transposon was inserted into this gene. This strain was located in [R3, C9, PR1, PC4], Plate4 C9.

To confirm the transposon insertion in *aat* gene of the selected mutant strain, bacterial pellets of mutant and wild-type strains were collected and genomic DNA were extracted. Then the *aat* gene was amplified with primers: transposon-inserted *aat* gene was expected to be ~2.5 kb while wide-type *aat* gene was expected to be ~500 bp.

Primers for *aat* mutant identification

TGACATTACCTCAGCGTAATTTAC

CAGCATAGAGCCGTAAACTGC

Statistical analysis

Statistical analyses were performed with GraphPad Prism 9 (GraphPad Software). Differences between groups were calculated by Student's unpaired t-test, one- or two-way ANOVA with default settings. Significant differences are labeled as n.s., no significance or * p<0.05; ** p<0.01; *** p<0.001; **** p<0.0001.

Figure design

Phylogenetic trees were generated and annotated with GraPhlAn. Schematic figures were created using BioRender.

Tables

Information of tables are shown in **Appendices**.

Chapter 3

The gut microbiota is a source of endogenous ligands in pain sensation

Overview

Functional somatic syndromes (FSS) are prevalent and common in various medical or health-care settings. They're characterized by ongoing unpleasant pain, reduced threshold for pain, or visceral hyperalgesia [79]. Based on the differences of regional syndromes or chemical responses, FSS could be further defined with other terms, such as irritable bowel syndrome (IBS), fibromyalgia/chronic widespread pain (FM/CWP), and multiple chemical sensitivity (MCS) [80]. It has been a diagnostic challenge for many years that FSS lack standard objective clinical signs and just rely on subjective reports. Furthermore, each FSS is not entirely independent of each other, indicated by a substantial overlap of symptoms. For example, patients with IBS may also have joint and muscle pain characterized in FM besides recurrent abdominal pain [82]. Considering the huge burden on society, the mechanisms or clear diagnostic markers of FSS are significant to be investigated.

Recent studies have shown an increasing number of evidence that gut microbiota dysbiosis correlates with FSS, particularly, IBS and FM [81, 82]. IBS patients were found to harbor higher abundance of *Enterobacteriaceae* and *Veillonella*, while lower abundance of *Bifidobacterium* [93]. Fibromyalgia patients were observed in alterations of butyrate- and propionate-metabolizing species, including decreased *Faecalibacterium prausnitzii* and *Bacteriodes uniformis*, and increased *Intestinimonas butyriciproducens*, *Flavonifractor plautii*, *Butyricicoccus desmolans* and *Eisenbergiella tayi* [81]. However, the causation relationship between gut microbiota dysbiosis and pain sensation is still unclear.

Nociceptor sensory neurons are known to mediate pain sensation. These specialized neurons densely innervate peripheral tissues, both mucosal barrier and visceral organs including skin, GI tract, cornea, joints, heart and lung [108]. In the GI tract, sensory neurons located in submucosal and myenteric plexus extend processes into the lamina propria of the mucosal layer where they

can sense nutrients or other chemical signals secreted by epithelial cells, for example, enteroendocrine cells [166]. But it's still unclear if and how these neurons could sense luminal signals derived from the gut microbiota. Nociceptors express diverse molecular receptors or sensors, such as transient receptor potential (TRP) channels and G protein- coupled receptors (GPCRs). Then they send the information of stimuli from peripheral nerve terminals into nodose ganglia (NG), dorsal root ganglia (DRG), or the brain [99]. As one major subset sensing noxious heat or chemical stimuli, TRPV1+ nociceptors have critical role in pain sensation. In addition, TRPV1 expression is increased in IBS patients and correlates with average pain score [167, 168]. TRPV1 channel has a broad-spectrum of ligands, including capsaicin, protons, ATP, toxins from spider venom, and endogenous lipids [99]. Furthermore, pathogens can indirectly sensitize TRPV1 channel through GPCR signaling or pore-forming toxins, regulating skin immunity in infection of *Staphylococcus aureus*, *Streptococcus pyogenes* and *Candida albicans*, and gut immunity in *Salmonella* infection [113, 115, 116, 169, 170]. Therefore, we hypothesized that TRPV1+ nociceptors could also respond to the gut commensal and modulate visceral pain sensation through TRPV1-microbial metabolites interactions. Here we established a pipeline to functionally screen the TRPV1-activating metabolites derived from over 100 phylogenetically diverse human gut microbes. We observed that diverse microbes exhibited various TRPV1-activiting ability. Remarkably, two species, *Klebsiella pneumoniae* and *Acidaminococcus intestini* produced heat-stable, small-molecule metabolites as potential TRPV1 endogenous ligands. In addition, the facts imply the relationship of these two species with visceral pain sensation that *K. pneumoniae* infection causes chest pain [171] and the *Veillonellaceae* order which *A. intestini* belongs to positively correlates with IBS pain score. We also found that live *A. intestini* activated TRPV1+ nociceptors isolated from DRG. Although we haven't identified the microbial metabolites through traditional chemical ways, lysophospholipids might be interesting targets, supported by the result that phospholipids from a human metabolite library also activated TRPV1 in our screening assay. With several promising future directions including metabolite identification, visceral pain mouse model and optimization of TRP-activation assay, we hope to dissect the

mechanisms by which gut commensals regulate visceral pain sensation and find new diagnostic markers or therapeutic methods.

Results

Establishing a GCaMP6s-based assay to identify TRPV1-activating human gut microbes

TRPV1 is a relatively non-selective ion channel with permeability to cations, mainly sodium and calcium [102]. Calcium influx is commonly used to evaluate the activation of neuronal cells. Therefore, we chose GCaMP, a widely used indicator for calcium mobilization to indicate TRPV1 activation. Basically, GCaMP consists of circularly permuted green fluorescent protein (cpGFP), calcium-binding protein calmodulin (CaM), and CaM-interacting M13 peptide. cpGFP has fused N- and C-termini in the middle of the protein, and the new terminus fused to CaM-M13 complex [172]. After TRPV1 is activated, mobilized calcium can bind with CaM-M13 complex, resulting in structural change of cpGFP and generate green fluorescence signal [172]. Among various versions of GCaMP, GCaMP6s exhibited the best sensitivity and strongest GFP signal. Therefore, we established the *in vitro* screening assay through transiently transfecting HEK293T cells with TRPV1 and GCaMP6s. After individually treating the cells with supernatants of 122 diverse human gut microbes, the activation of TRPV1 could be indicated by the fold-change of relative fluorescence units (RFU) compared to negative control cells which were treated with PBS (**Fig. 1A**).

We found that supernatants of diverse human gut microbes activated TRPV1 channel with a various pattern (**Fig. 1B**). Although Gifu medium could also activate TRPV1 channel in this assay, several isolates exhibited significantly stronger ability to increase the calcium mobilization. Based on the fold-change of RFU to Gifu medium, all isolates could be simply classified into low-activation (weaker than Gifu medium), middle-activation (1-2-fold of Gifu medium level) and high-activation hits (>2-fold Gifu medium level) (**Fig. 1C-D**). Five high-activation hits were from the same species *Acidaminococcus intestini* (**Fig. 1D**). *A. intestini* is Gram-negative, anaerobic bacterium from *Veillonellaceae* order. Although the abundance of *A. intestini* in IBS patients has been unknown, it's reported that the amount of stool *Veillonellaceae* positively correlated with IBS

clinical severity. In addition, a large percent of middle-activation hits belonged to *Enterobacteriaceae*, for example, *Klebsiella pneumoniae* (**Fig. 1D**). *K. pneumoniae* is Gram-negative, facultative anaerobic bacterium, causing destructive changes to lungs even though found in the normal flora of the mouth, skin, and intestines. Infection symptoms might include fever, chills, cough, and chest pain. Furthermore, supernatants of *K. pneumonia* and *A. intestini* could activate both mouse and human TRPV1 in the screening assay (**Fig. 1E-F**). Therefore, we focus on these two species for the subsequent studies.

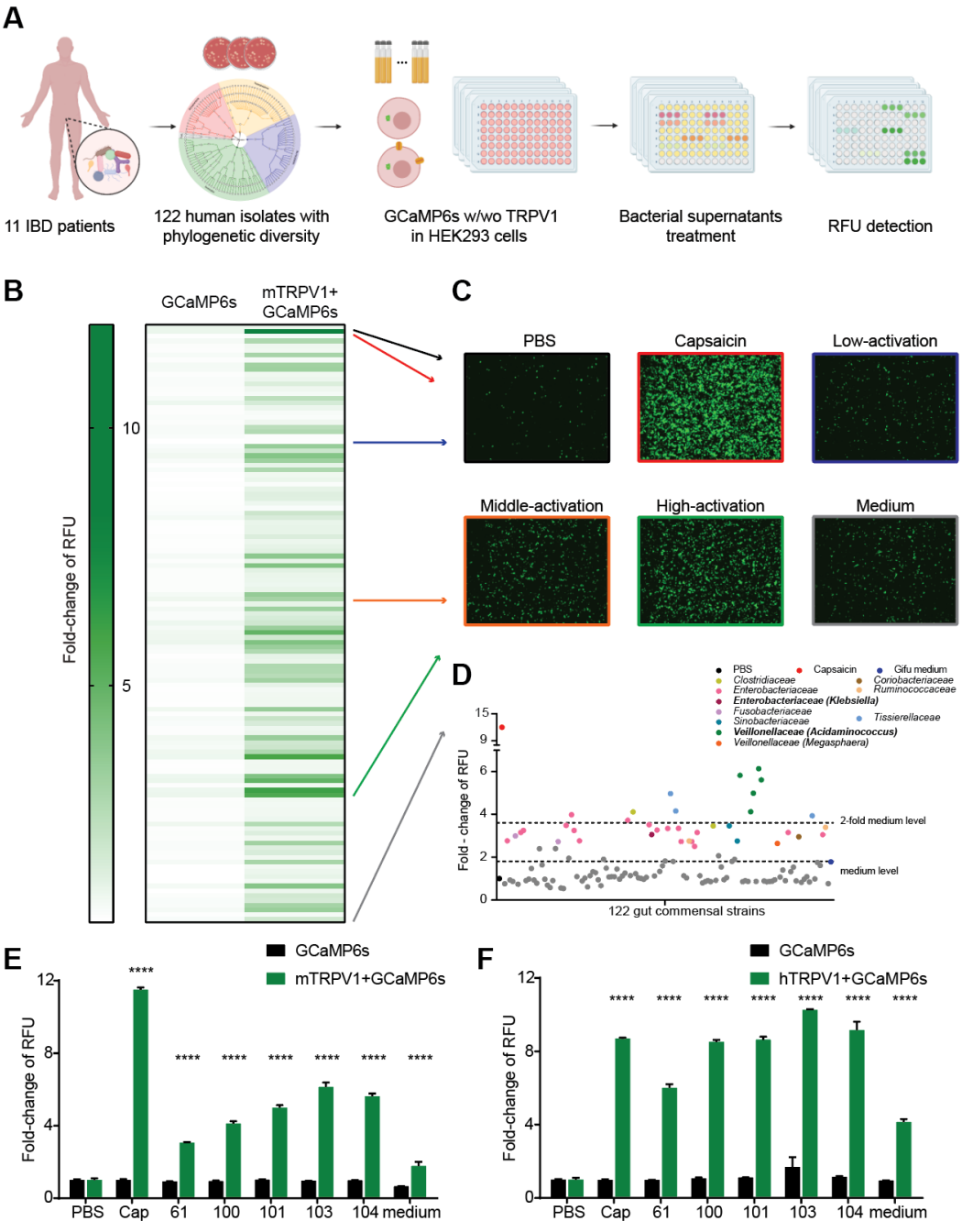


Figure 1. Establishing a GCaMP6s-based assay to identify TRPV1-activating human gut microbes.

(A) Overview of functional screening of gut microbes for TRPV1 activation. 122 phylogenetically diverse bacterial isolates from 11 IBD patients (shaded based on phylum: Red, Actinobacteria;

Blue, Bacteroidetes; Orange, Proteobacteria; Gray, Fusobacteria) were cultured anaerobically. GCaMP6s with or without TRPV1 plasmids were transiently transfected into HEK293T cells. After treating cells with bacterial supernatants, GFP relative fluorescence units were detected with plate reader.

(B) Diverse human gut bacteria exhibit TRPV1-activating activities. Fold-change of RFU was normalized to negative control cells treated with PBS. Every isolate was compared between GCaMP6s with or without mTRPV1 (mouse TRPV1) to avoid potential false-positive results.

(C) Representative images of HEK293T cells treated with PBS, 1 μM capsaicin, bacterial supernatants of low-activation, middle-activation, high-activation isolates or Gifu medium.

(D) Dot-plot of fold-change of RFU of the 122 human isolates. Different colors indicated distinct families.

(E-F) Bacterial supernatants of *K. pneumonia* (NWP61) and *A. intestini* (NWP100, 101, 103, 104) significantly activated mTRPV1 (mouse TRPV1, E) and hTRPV1 (human TRPV1, F) in the screening assay. Cap, 1 μM capsaicin; medium, Gifu medium. **** $p<0.0001$, two-way ANOVA.

The potential microbial TRPV1 agonists were relatively stable small molecules and prevalent in Klebsiella genus and Veillonellaceae family

Since the bacterial supernatants were sufficient to activate TRPV1, we then performed several treatments to the supernatants to understand more characteristics of potential TRPV1-activating microbial metabolites. In our primary screening, all bacterial supernatants were heated at 98 °C for 15 min, indicating that potential TRPV1 agonists were heat stable. To identify if the functional metabolites were large or small molecules, we separated the bacterial supernatants (SUP) into small- (<3 kDa SUP) and large- (>3 kDa SUP) molecular weight fractions and found that just small molecule fractions kept the TRPV1-activating ability (**Fig. 2A-B**). In addition, proteinase K treatment didn't change the activation significantly (**Fig. 2C**). Therefore, the potential microbial TRPV1 agonists were small molecules.

Butanol (BuOH) is a good solvent to extract small molecules in chemistry. We found that the BuOH extracts of *K. pneumoniae* (NWP61) and *A. intestini* (NWP103) still significantly activate TRPV1 compared to medium extracts (**Fig. 2D**), shedding light on future chemical isolation and identification.

Capsazepine (CPZ) is a well-known antagonist of TRPV1 channel, competing the binding site of capsaicin [99]. CPZ significantly blocked the TRPV1 activation mediated by capsaicin in our screening assay (**Fig. 2E**). Interestingly, we found that TRPV1 activation mediated by *A. intestini* SUP was also blocked by CPZ, further indicating the potential direct TRPV1-microbial metabolite interactions (**Fig. 2E**). Furthermore, *A. intestini* SUP also activated TRPV5 but not other TRP channels, including TRPV2, TRPV3, TRPV4, TRPA1, and the TRPV5 activation was not antagonized by CPZ (**Fig. 2F**). Taken together, these results implied a specific interaction between TRPV1 and unknown microbial metabolite(s) derived from *A. intestini*.

We then cultured more phylogenetically related species in the Human Microbiome Project to evaluate the prevalence of potential TRPV1-activating metabolites. All 17 strains closely related to *K. pneumoniae* NWP61 from the *Klebsiella* genus significantly activated TRPV1, although there was a variety of this ability (**Fig. 3A-B**). Similarly, we evaluated more strains closely related to *A. intestini* (NWP96, 100, 101, 103 and 104) (**Fig. 3C-D**). 1 strain of *Acidaminococcus*

fermentans (AF) and 2 strains of *Acidaminococcus sp.* (HM81, 853), from the same genus of *A. intestini*, also activated TRPV1, either cultured in Gifu or RCM medium. However, 7 strains of *Veillonella* species (HM49, 64, DS72, HM562, 778, 895, 1157) didn't activate TRPV1 either cultured in Gifu or RCM medium, and 1 *Veillonella sp.* strain (HM850) just activated TRPV1 after cultured in RCM medium. Such defects may be due to the bacterial growth or metabolite concentration that was indicated by the OD_{600} (**Fig. 3E-F**). Interestingly, 1 *Megasphaera spp.* strain (NWP111) cultured in Gifu medium didn't activate TRPV1, even though the OD_{600} indicated that it grew well (**Fig. 3C-E**). In summary, these results suggested that potential microbial TRPV1-activating metabolites were prevalent in *Klebsiella* and *Veillonellaceae* species. However, culturing conditions might be a critical factor influencing the production or concentration of these metabolites. Therefore, the conditions to perform metabolomics or chemical identification should be carefully considered.

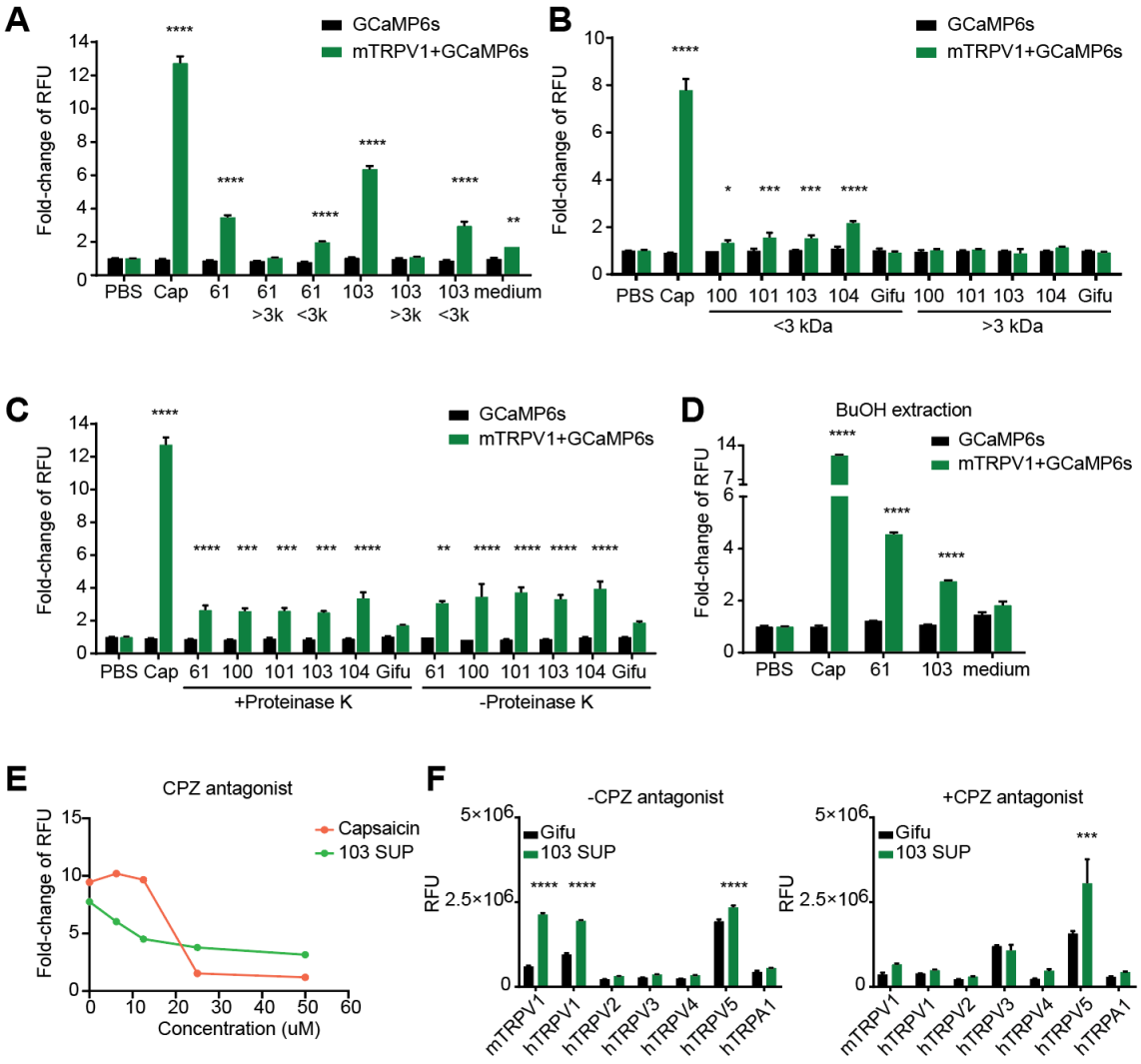


Figure 2. The potential microbial TRPV1 agonists were relatively stable small molecules.

(A-B) Small-molecule fractions of supernatants of *K. pneumonia* (NWP61) and *A. intestini* (NWP100, 101, 103, 104) significantly activated mTRPV1 (mouse TRPV1, E). Cap, 1 μM capsaicin; medium, Gifu medium. * p<0.05; ** p<0.01; *** p<0.001; **** p<0.0001, two-way ANOVA.

(C) Proteinase K treatment didn't affect TRPV1 activation mediated by supernatants of *K. pneumonia* (NWP61) and *A. intestini* (NWP100, 101, 103, 104). Cap, 1 μM capsaicin. ** p<0.01; *** p<0.001; **** p<0.0001, two-way ANOVA.

(D) BuOH extracts of supernatants of *K. pneumonia* (NWP61) and *A. intestini* (NWP103) activated TRPV1. Cap, 1 μM capsaicin. **** $p<0.0001$, two-way ANOVA.

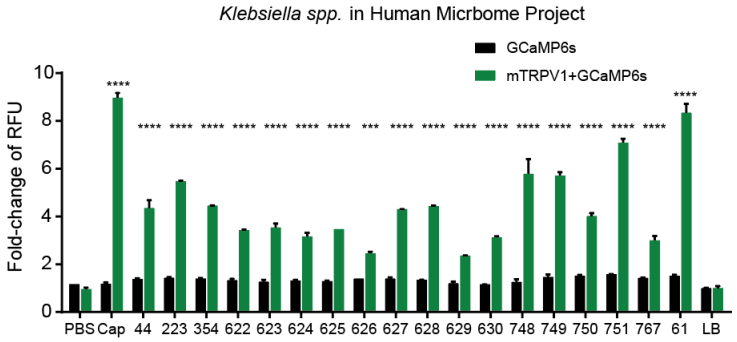
(E) Capsazepine (CPZ) blocked TRPV1 activation mediated by supernatants of *A. intestini* (NWP103). 1 μM capsaicin was positive control.

(F) Capsazepine (CPZ) blocked mTRPV1 (mouse TRPV1), hTRPV1 (human TRPV1) but not hTRPV5 (human TRPV5) activation mediated by supernatants of *A. intestini* (NWP103). *** $p<0.001$; **** $p<0.0001$, two-way ANOVA.

A

Strain	Species
HM44	*Klebsiella sp.*
HM223	*Klebsiella sp.*
HM354	*Klebsiella sp.*
HM622	*Klebsiella oxytoca*
HM623	*Klebsiella oxytoca*
HM624	*Klebsiella oxytoca*
HM625	*Klebsiella oxytoca*
HM626	*Klebsiella oxytoca*
HM627	*Klebsiella oxytoca*
HM628	*Klebsiella oxytoca*
HM629	*Klebsiella oxytoca*
HM630	*Klebsiella oxytoca*
HM748	*Klebsiella pneumoniae*
HM749	*Klebsiella pneumoniae*
HM750	*Klebsiella pneumoniae*
HM751	*Klebsiella pneumoniae*
HM767	*Klebsiella sp.*
NWP61	*Klebsiella pneumoniae*

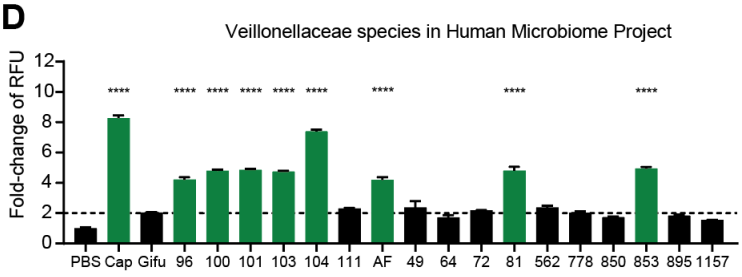
B

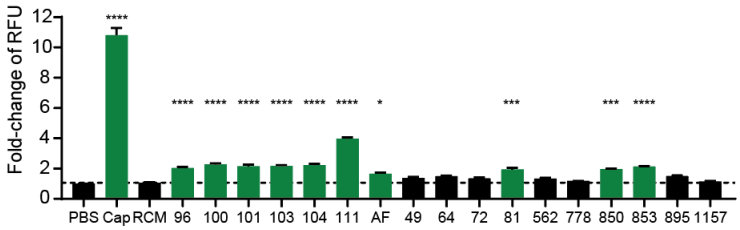


C

Strain	Species
NWP96	*Acidaminococcus intestini*
NWP100	*Acidaminococcus intestini*
NWP101	*Acidaminococcus intestini*
NWP103	*Acidaminococcus intestini*
NWP104	*Acidaminococcus intestini*
NWP111	*Megasphaera spp.*
AF	*Acidaminococcus fermentans*
HM49	*Veillonella sp.*
HM64	*Veillonella sp.*
DS72	*Veillonella parvula*
HM81	*Acidaminococcus sp.*
HM562	*Veillonella sp.*
HM778	*Veillonella sp.*
HM850	*Veillonella sp.*
HM853	*Acidaminococcus sp.*
HM895	*Veillonella sp.*
HM1157	*Veillonella montpellierensis*

D





E

Gifu	96	100	101	103	104	111	AF	49
OD600	0.694	0.68	0.67	0.67	4.24	2.93	1.008	0.11
64	72	81	562	778	850	853	895	1157
0.048	0.106	0.9	0.224	0	0.026	0.8	0.048	0

F

RCM	96	100	101	103	104	111	AF	49
OD600	0.564	0.546	0.536	0.536	0.538	5.11	0.676	0.35
64	72	81	562	778	850	853	895	1157
0.308	0.216	0.616	0.406	0	0.45	0.64	0.374	0

Figure 3. Diverse species from *Klebsiella* genus and *Veillonellaceae* family could activate TRPV1.

(A-B) LB-supernatants of diverse *Klebsiella* species activated TRPV1. **** $p<0.0001$, two-way ANOVA.

(C-D) Gifu- or RCM-supernatants of diverse *Veillonellaceae* species activated TRPV1. * $p<0.05$; *** $p<0.001$; **** $p<0.0001$, one-way ANOVA.

(E-F) OD600 of *Veillonellaceae* species grown in Gifu or RCM.

Lysophospholipids might be TRPV1 endogenous ligands sensitizing nociceptors

Based on our previous screening results, *A. intestini* was a strong TRPV1-activating hit. We then determined if *A. intestini* could sensitize TRPV1+ nociceptive neurons through calcium imaging. Live *A. intestini* was observed to induce robust calcium influx in isolated mouse DRG neurons in a dose-dependent manner within minutes of application (**Fig. 4A**). Interestingly, bacteria-responsive DRG neurons also significantly responded to capsaicin, an indicator of TRPV1+ nociceptors (**Fig. 4B**). Such specificity to TRPV1+ nociceptors was also observed in other pain-inducing pathogens, for example, *Streptococcus pyogenes*, with similar concentration of bacterial cells (4-5x10^9 cfu/ml) [115]. Therefore, *A. intestini* might be a microbial source of TRPV1 endogenous ligands to sensitize nociceptors and regulate visceral pain sensation.

We performed traditional chemical extraction, fractionation and GCaMP6s-based evaluation to identify the potential TRPV1-activating metabolites derived from *A. intestini*. Unfortunately, we couldn't get consistent results after many rounds of such work. To avoid the potential loss of metabolites during this pipeline, we then decided to perform untargeted global metabolomics with raw bacterial supernatants instead of extracts. Since bacterial supernatants contain hugely diverse and complex metabolites from small molecules to large proteins, we finalized the potential ion list after filtering ions generated from different culturing conditions (**Fig. 4C**). Firstly, bacterial supernatants of NWP103 cultured in Gifu medium for 12 h or 24 h could both significantly activate TRPV1 (**Fig. 4D**). 28 ions were present in both conditions but not Gifu and RCM medium. Secondly, bacterial supernatants of NWP103 cultured in RCM medium for 12 h or 24 h also activated TRPV1 (**Fig. 4E**), so 10 of the 28 ions were selected. Finally, bacterial supernatants of *Megasphaera spp.* strain (NWP111) cultured in Gifu or RCM medium for 12 h didn't activate TRPV1 (**Fig. 4D-E**), so 7 of the 10 ions were final potential target chemical hits. Among these 7 ions, 4 lysophospholipid metabolites had higher abundance: 2 isoforms of 14:0 lyso PE (*m/z* 426) and 2 isoforms of 16:1 lyso PE (*m/z* 452) (**Fig. 4F**).

Phospholipids are known to involve various signaling pathways, including activation of TRP channels through modulating their structures on plasma membrane [99, 173]. However, it's still unclear if and how microbial phospholipids regulate TRP channels. Furthermore, 14:0 lyso PE

and 16:1 lyso PE identified here were unique metabolites of *A. intestini*, suggesting a broader role of microbial metabolites in host physiology. However, there are hurdles to handle lipids and we haven't got consistently repeatable data for the evaluation of these metabolites in TRPV1 activation (data not shown).

We also screened a human metabolite library to identify more potential endogenous ligands of TRPV1. Interestingly, besides piperine, a ligand of TRPV1 similar to capsaicin, 16:0 lyso PC in this library also significantly activated TRPV1 (**Fig. 5A-C**). Furthermore, lysophopholipids have been reported to activate TRP channels, including TRPV2, TRPV4, and TRPM8 [173, 174]. All above results suggested that lysophospholipids might be a class of endogenous ligands of TRPV1, partially derived from the gut microbiota.

With hurdles to directly determine the metabolites from bacterial supernatants, we also attempted to find the gene responsible for the production of TRPV1 agonists. We firstly constructed a gain-of-function (GOF) library in which the genomic DNA of *A. intestini* was sheared into ~3 kb fragments and expressed in *E. coli* (**Fig. 5D**) [175]. Unfortunately, we didn't get any promising clones that could significantly activated TRPV1 in the screening assay (data not shown). This might be due to the limitation that *E. coli* is not a proper expression strain because *A. intestini* belongs to Firmicutes, a completely different phylum to *E. coli*. Or the genes responsible for the production of potential metabolites are not located within 3 kb in the genome. Therefore, we still have a long way to clearly understand and dissect the mechanisms by which the gut microbiota activated TRPV1.

We also evaluated the TRPV1 activation of complex microbial mixtures from humans. Even fecal water samples from healthy donors exhibited intra-individual difference in activating TRPV1 (**Fig. 5E**). Therefore, it is promising to evaluate with human samples from patients, like IBS or FSS patients and compare the activation pattern with healthy controls.

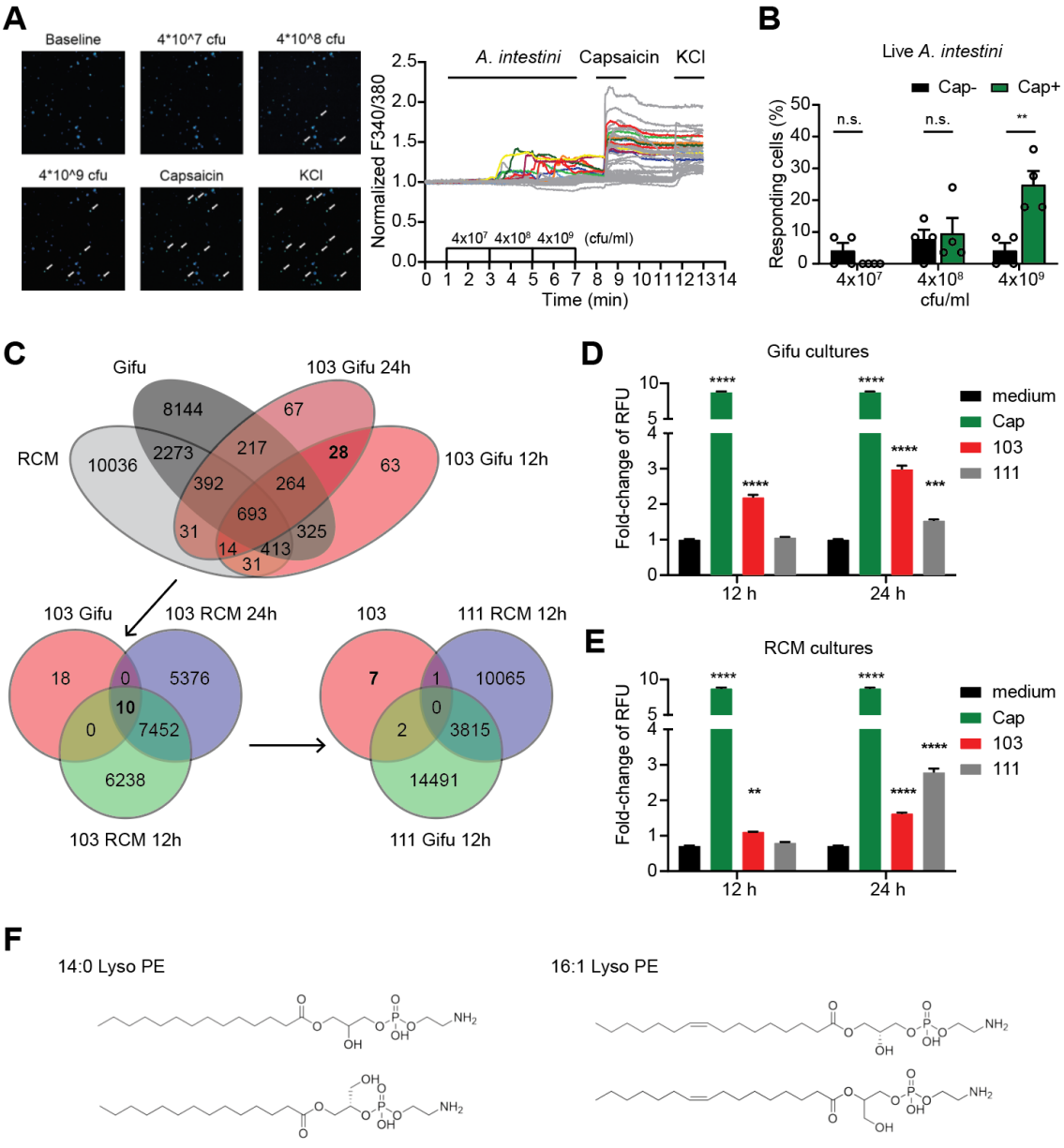


Figure 4. Lysophospholipids derived from *A. intestini* might be TRPV1 endogenous ligands sensitizing TRPV1+ DRG nociceptors.

(A) Representative Fura 2-AM images and calcium traces of isolated mouse DRG neurons. Cells were stimulated with live *A. intestini* ($4x10^7$, $4x10^8$, $4x10^9$ cfu/ml), capsaicin (1 μM), and KCl (135 mM). White arrows indicated cells responding to the stimuli.

(B) Percent of capsaicin non-responsive (Cap-) and capsaicin responsive (Cap+) neurons that responded to *A. intestini* ($4x10^7$, $4x10^8$, $4x10^9$ cfu/ml). ** $p<0.01$, one-way ANOVA.

(C) Venn diagrams of step-by-step untargeted global metabolomics to finalize the potential ion list unique in bacterial supernatants of *A. intestini*.

(D-E) The activation of TRPV1 after treatment with 12- or 24-h Gifu- or RCM-supernatants of *A. intestini* (NWP103) and *Megasphaera spp.* (NWP111. ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$, one-way ANOVA.

(F) Chemical structures of the 4 abundant lysophospholipids derived from *A. intestini*.

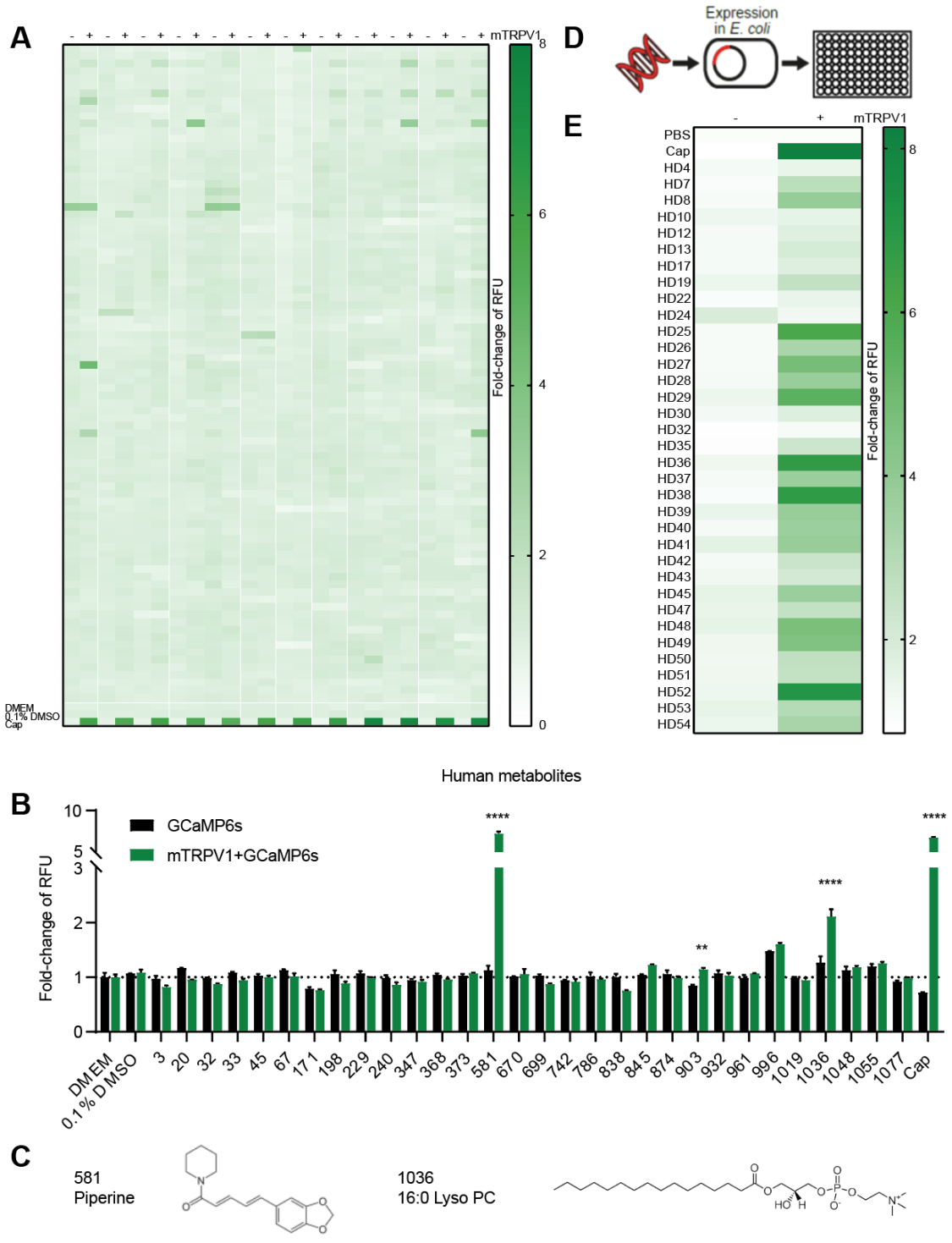


Figure 5. Human metabolites screening for TRPV1 activation.

(A) Fold-change of RFU of 1100 human metabolites in the screening assay. Every metabolite was diluted into 1mM to treat cells. Cap, 1 μM capsaicin; -, GCaMP6s; +, mTRPV1+GCaMP6s.

(B) The activation of TRPV1 after treatment with metabolites selected from **A**. ** $p<0.01$; **** $p<0.0001$, two-way ANOVA.

(C) Chemical structures of piperine and 16:0 lyso PC.

(D) Overview of *A. intestini* GOF library.

(E) Fold-change of RFU of human fecal water samples. Human feces were resuspended in PBS at 100 mg/ml, then heated fecal water samples were used to treat cells. Cap, 1 μM capsaicin; HD, healthy donors.

Discussion

Although the correlations between the gut microbiota dysbiosis and IBS or other FSS diseases have been observed in many studies, the causative effects of the gut microbiota in regulating pain sensation are still unclear. Here, we undertook a systematic evaluation of TRPV1-activation mediated by a diverse selection of human gut microbes based on the reasoning that TRPV1+ nociceptors are a major subset of neurons responding to noxious stimuli and induing pain sensation. We found that several species from the human gut microbiota significantly activated TRPV1 with their bacterial supernatants. Particularly, Gram-negative bacteria *Klebsiella pneumoniae* and *Acidaminococcus intestini* produced heat-stable, small-molecule metabolites as potential TRPV1 endogenous ligands. In addition, their phylogenetically related species, such as *Klebsiella oxytoca*, *Acidaminococcus fermentans* and other unidentified species from *Klebsiella* and *Acidaminococcus* or *Veillonella* genera could also activate TRPV1, although dependent on culturing conditions, suggesting the prevalence of potential TRPV1-activating metabolites in these gut microbes. We also confirmed that live *A. intestini* could activated isolated mouse DRG sensory neurons, with a preference to TRPV1+ (capsaicin-responsive) nociceptors. We have tried hardly to identify the metabolites through traditional chemical isolation or untargeted global metabolomic, and found that lysophospholipids might be a novel interesting class of chemicals as endogenous ligands of TRPV1. There are few studies about the unique phospholipids derived from the gut microbiota, even though host phospholipids are well-known to influence TRP channel activation, which is also supported by our screening results from the human metabolite library. In the future, it still needs more experiments to dissect if and how microbial lipids sensitize nociceptors and regulate pain sensation.

Besides the effects on pain sensation, the existence of unique microbial lipid metabolites also implies an expanded physiological role in the human diseases. More broadly, lipids have been revealed to involve various signaling pathways, including TRP channels, GPCRs and pattern recognition receptors (PRRs), suggesting their significances in homeostasis, olfaction, immunity, cognition and metabolic systems [173]. Indeed, the gut microbiota plays critical roles in many

aspects, thus the mechanisms of microbial lipid metabolites regulating host responses need to be evaluated and illustrated.

Although we focused on TRPV1 in this project, it's striking that bacterial metabolites also activated other TRP channels, for example, TRPV5 (**Fig. 2**). Therefore, mapping a broader swath of microbiota metabolomes against all TRP channels would reveal more promising interactions. A high-throughput technology interrogating all TRP channels will be useful to screen thousands of metabolomes from diverse sources, ranging from individual bacterial strains cultured *in vitro* to complex microbial mixtures growing *in vivo* in mice or humans. Indeed, we observed diverse TRPV1 activation of human fecal samples (**Fig. 5E**), suggesting inter-individual differences of human microbiota in TRP channel activation.

Finally, it's essential to identify the chemical structures of microbial TRPV1-activating metabolites and confirm their functions with more neuronal experiments or mice behavior models related to pain sensation, including mechanical pain, neuropathic pain or other pain models. To exclude other microbial impacts and instruct future precise therapies, it's also critical to identify the genes responsible for the production of TRPV1-activating metabolites. Unfortunately, GOF library construction didn't give us useful information, so we decided to try loss-of-function (LOF) library construction. Although it's challenging to do commensal engineering, especially natural anaerobic strain like *A. intestini*, we could start with *Klebsiella* and combine bioinformatic analysis later, based on our results that the TRPV1-activation ability was shared by phylogenetically related species. *K. pneumoniae* was reported engineerable with transposon. Therefore, we plan to apply the random mutagenesis method optimized in *Morganella morganii* to screen out a mutant strain with defects in TRPV1-activation. The mutant strain would be useful for gene identification, gene prevalence analysis based on whole-genome-sequence and even chemical identification through comparative metabolomics. It will be also possible to do strict *in vivo* physiological evaluation.

Future Directions

LOF library construction of Klebsiella pneumonia

Klebsiella pneumoniae often normally lives in the gut but can cause serious infections if move to other parts through bloodstream, especially in immunocompromised patients. It is a common cause of infections like bacteremia and the symptoms include fever, cough and chest pain. Some *K. pneumoniae* strains have acquired resistance to a broad spectrum of antibiotics, becoming a huge burden of the society. Here we found that *K. pneumonia* and its phylogenetically related species activated TRPV1, suggesting an interaction between host nociceptors and intestinal microbial metabolites. Investigation of how *K. pneumoniae* activate TRPV1 at molecular or cellular levels could provide new therapeutic directions.

In Chapter 2, we optimized a transposon-based whole-genome random mutagenesis method based on Sudoku. Similar to *K. pneumoniae*, *Morganella morganii* is also a Gram-negative, multiple antibiotic-resistance species. Therefore, we plan to apply this method to construct a LOF library of *K. pneumoniae*, either using Ez-Tn5 transposon kit (if our strain is kanamycin-susceptible) or optimizing Himar I Mariner transposon (need to re-design primers for transposon insertion sequencing) [176]. Then our GCaMP6s-based TRPV1-activation assay could screen all mutant clones and identify the genes responsible for the production of TRPV1-activating metabolites.

If the genes could be identified in this way, bioinformatic analysis of more *Klebsiella* species, even *Acidaminococcus* or *Veillonella* species would be useful to evaluate the gene prevalence in other TRPV1-activaitng microbes. In addition, metagenomic analysis of human gut microbiota dataset could provide more information about the gene prevalence cross patient cohorts. Besides gene identification, chemical identification also become easier through comparative metabolomics with mutant and wild-type strains. Furthermore, either mutant strains or pure chemicals could be utilized to confirm the physiological roles in nociceptor sensitization or pain sensation *in vivo* with proper mouse models.

Mouse models for pain sensation

Various mouse models for pain sensation have been designed for different research purposes. Von-filament inducing mechanical pain [177] and heat-plate inducing thermal pain [113] are two

commonly used behavior models. In addition, some other neuropathic models are also widely used for studies of neuron-immune interactions [178]. For visceral pain model like IBS, colon distention is a good indicator for the gut responses. And stressed models are useful for those pain sensation related to cognitive or stress symptoms [179]. IBS is also reported to correlation with inflammation in some cases, therefore, bacteria-inducing infection model like Salmonella infection is also useful [180].

Considering this project mainly focuses on intestinal host-microbe interactions, we prefer to start with models revealing the local responses, like colon distention model [181]. A challenge of this model is that complex surgeries in this model are hard to be done with gnotobiotic mice or mice colonized with anaerobic microbes. Therefore, *K. pneumoniae* is easier than *A. intestini* to be evaluated *in vivo* [182, 183]. Infections with mutant or wild-type *K. pneumoniae* strains could also confidently reveal the mechanisms mediated by TRPV1-activating metabolites. Furthermore, pure chemicals could also overcome many technical hurdles.

A high-throughput screening assay integrated all TRP channels

Previous works in our lab on Presto-Tango [184]and Presto-Salsa suggest a new direction to more broadly investigate the microbial metabolome-TRP channel interactions. Similar to the design of Presto-Salsa, we can also develop a high-throughput screening assay containing all TRP channels (TRPV, TRPM, TRPA, TRPC) distinguished by unique nucleic acid barcodes that can be read out by next-generation sequencing. Different from GCaMP6s assay dependent on transit calcium influx, we need to design a new downstream transcription-based signaling pathway to indicate calcium mobilization and TRP activation. TEV-tTA system that is used in Presto-Salsa or UAS-GAL4 system that has been reported by different groups are good candidates to be optimized [185, 186]. In the future, we hope such high-throughput system could reveal more host-microbe interactions mediated by metabolites from various sources (from bacterial cultures to patient samples) with higher efficiency.

Materials and Methods

Bacteria strains and culturing

NWP, DS and HMP strains were from the same source as those in Chapter 2. *Acidaminococcus fermentans* was from ATCC (25085).

All strains were cultured in Gifu Anaerobic Broth (GAM Broth; Himedia Laboratories, M1801) or RCM (Reinforced Clostridial Medium; BD, DF1808-17-3) at 37 °C under anaerobic conditions in a Coy Laboratory Products Inc. chamber (10 % CO_2, 4 % H_2, 86 % N_2) or aerobic conditions.

pZE21 strain was gift from Dr. Andrew Goodman. *E.coli* S17 λpir carrying pSAM_Kp2.1 (donor strain) and PIR1 *E. coli* carrying pSAM_Kp2.1 (plasmid amplification strain) were gift from Dr. Eric Pamer.

Plasmids

pGP-CMV-GCaMP6s was from Addgene (40753); mTRPV1 (mouse TRPV1) was a gift from Dr. Elena Gracheva; all human TRP channel plasmids were from GenScript; UAS-GAL4 system developed by Dr. Liqun Luo contained pQUAST-p65AD::CaM (Addgene, 64715), pQUAST-MKII::GAL4DBDo (Addgene, 64723), pUAS-GG (Addgene, 24342), pUAS-luc2 (Addgene, 24343); UAS-GAL4 system developed by Dr. Alice Ting contained CMV-GFP-CaM-uTEVp (Addgene, 163028), CD4-MKII-f-hLOV1-TEVcs(ENLYFQ/M)-Gal4 (Addgene, 163027), UAS-mCherry (Addgene, 135457).

Unique barcode plasmids to pair each TRP channel were constructed through inserting barcodes after GFP on the pUAS-GG plasmid. Briefly, vector plasmids were digested with HF-MluI and HF-SalI and purified. Oligos with barcodes (100 μM) were phosphorylated and annealed with T4 PNK (NEB) in a 10 μl system:

1 μl forward oligo (CGCGTataactgNNNNNNNNNNatcatG)

1 μl reverse oligo (TCGACatgatNNNNNNNNNNcagttatA)

1 μl 10X T4 Ligation Buffer

6.5 μl H_2O

0.5 μl T4 PNK

Incubate at 37 °C for 30 min. Then 95 °C for 5 min and ramp down to 25 °C at 5 °C/min.

50 ng digested pUAS-GG and 1 µl oligo duplex (1:200 dilution) were ligated with 1 µl Quick Ligase (NEB) in a 10 µl system for 10 min at room temperature. Take 2-5 µl ligation system to transform into competent *E. coli*. Pick single colonies, extract plasmids and sequence the barcodes after GFP. 52 pUAS-GG plasmids with unique barcodes were selected as follows:

Well	No.	Barcode	Well	No.	Barcode
A01	GG01	AGGGGAGAAA	E07	GG27	TCTCACCTTG
A02	GG02	AGAGGCTGGA	E09	GG28	GCCTCGTAGC
A03	GG03	GGGGAGGGGA	E10	GG29	GGACAAGCAA
A04	GG04	TTACAATGGA	E12	GG30	TGCTGCAGGG
A05	GG05	GAGGGGGGGG	F01	GG31	TTACCCCGCC
A08	GG06	CGTCCACTTT	F02	GG32	TCTCGTCTGA
A11	GG07	CTATGAGAAG	F03	GG33	ATCTTCTCTT
B01	GG08	CATGTGACAC	F07	GG34	GGATTGTCAC
B02	GG09	TCGAACTCCG	F09	GG35	CGCAGGAGAC
B04	GG10	TGGTGTGTAA	F10	GG36	ATCCTCGGAG
B05	GG11	TATGAATCTG	F11	GG37	CCTGAGTGGC
B06	GG12	CTCCCTTTGT	G01	GG38	CTTGCATGCC
B08	GG13	ACCTCGATCT	G02	GG39	TTACTGTCCT
B10	GG14	GGGATAAGAG	G03	GG40	CACCGGGGTA
B11	GG15	GGGGGTCTAA	G04	GG41	TGTATTTTAT
C02	GG16	CCCCGAGAAC	G05	GG42	GGAACATTAA
C03	GG17	GCCGGCGCGA	G07	GG43	TCCGCTCCTA
C04	GG18	GAAGATGGGA	G08	GG44	GGGAGACGGG
C10	GG19	AATACAACCA	G09	GG45	ATAGACGTGG
D01	GG20	TGGTGTATCA	G10	GG46	CCTCCCAAAT
D02	GG21	CCGACTCCCC	G12	GG47	GTGTATAATC
D03	GG22	CTGATATCGT	H03	GG48	ACCACACTAG
D04	GG23	CAAGGTGTCT	H05	GG49	CTCTATGGAG
D05	GG24	GTCCTGTCCC	H06	GG50	AAAGAATTCG
E01	GG25	ATCCCCAAAG	H09	GG51	GCAAAATGCA
E06	GG26	CAGTGAGCCG	H10	GG52	TTTAATGACT

Cell culturing, PEI transfection and GCaMP6s-based screening

HEK293T cells were cultured in high glucose Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich, D6429) supplemented with 10 % FBS (Sigma-Aldrich, 12306C) and 1 % Penicillin-Streptomycin (Thermo Fisher Scientific, 15140122). Cells were seeded 100 µl/well into 96-well plates and cultured overnight for screening preparation.

150 ng GCaMP6s with or without 50 ng TRP plasmids were mixed with 0.6 µl PEI in 20 µl DMEM. The mixture was incubated for 30 min at room temperature, added into cells and incubated overnight for GCaMP6s-based screening.

Bacterial supernatants were collected after high-speed centrifugation. Amicon Ultra-0.5 Centrifugal Filter Unit, 3KDa (EMD Millipore, UFC500324) were used for small- and large-molecule separation. Heat treatment was performed at 95 °C for 15 min. Proteinase K treatment was performed through adding 1 µl proteinase K (Qiagen into 500 µl bacterial supernatants and incubated at 37 °C for 30 min and stopped at 95 °C for 15 min. Discard the cell culture medium and add 50 µl bacterial supernatants per well. After stimulation at 37 °C for 30 min, discard liquids and read GFP signal with plate reader SpectraMax i3x (Ex/Em = 472 nm/512nm). Fold-change of RFU was calculated through normalizing the values of samples to PBS-treating cells.

For antagonist blocking experiments, cells were firstly incubated with capsazepine (CPZ) at 37 °C for 30 min before treatment with samples.

Butanol extraction of bacterial metabolites

Bacterial cultures were centrifuged to remove bacterial cells and other large particles. The resulting clarified bacterial supernatants were extracted with butanol (BuOH). Dried crude extracts were reconstituted with 1% DMSO/PBS at 20 mg/ml.

Dorsal root ganglia (DRG) calcium imaging

Animals (2 mice for 10 coverslips) were euthanized followed by decapitation. Exploit skin and cut the spinal cord out. Use needles to fix spinal cord under dissection-scope. DRG were dissected into 500 µl ice-cold PBS (HBSS with Ca^{2+}) along the spinal cord. Finish dissection within 30 min then chop tissues with scissors in 1.5 ml tube. Add 500ul pre-warmed collagenase P (2 mg/ml in HBSS. flick tubes after 7 min and then incubate another 8 min at 37 °C). Remove collagenase P and add 500 µl pre-warmed 0.25% trypsin (flick tubes after 5 min and then incubate another 5 min at 37 °C). Remove trypsin as much as possible and tissue was suspended in 500 µl DMEM complete media supplemented with 10% FBS and 1% penicillin/streptomycin and mechanically

dissociated using a plastic tipped pipette (1000 µl and then 200 µl, 15 times each). Spin tissues at 100 xg, 3min at room temperature. Remove supernatant and resuspend with 200 µl DMEM and mechanically dissociated 15 times. Add 20 µl suspension to each poly-D-lysine/laminin coverslips. Culture at 37 °C for 30-45 min, add another 500 µl DMEM slowly and culture 1-2 h. For calcium imaging, prepare 10 µM Fura 2-AM and 0.02% Pluronic F-127 in Ringer's buffer as dye. Transfer coverslips into new 12-well plate and wash cells with Ringer's buffer 3 times and load 500 µl dye to each well. Culture them at room temperature in dark for 1 h. Then the cells are washed with 500 µl Ringer's buffer 3 times and keep them in Ringer's buffer for imaging.
Images were obtained using Axio-Observer.Z1 inverted microscope (Zeiss) equipped with an Orca-Flash4.0 camera (Hamamatsu) using the MetaFluor software (Molecular Devices). The Ca signal was reflected by the ratio of 380/340 based on Fura 2-AM.
Cells were stimulated with live bacteria resuspended in Ringer's buffer, washed, stimulated with 1 µM capsaicin, washed, and stimulated with high potassium buffer.
Ringer's buffer:

#	Component	mM	MWg/mol	mg/1000mL	
1	NaCl	140	58.44	8181.6	Or 100ml 10x stock
2	KCl	5	74.5	372.5	Or 100ml 10x stock
3	Glucose	10	180.16	1801.6	
4	CaCl2	2	110.98	221.9	Or 2ml 1M stock
5	MgCl2.6H2O	2	203.3	406.6	Or 2ml 1M stock
6	HEPES	10	238.3	2383.0	Or 10ml 1M stock
	pH to 7.4				

High potassium buffer:

#	Component	mM	MWg/mol	mg/250mL	
1	NaCl	10	58.44	146.1	
2	KCl	135	74.5	2514.4	
3	Glucose	10	180.16	450.4	
4	CaCl2	2	110.98	55.5	Or 0.5ml 1M stock
5	MgCl2.6H2O	2	203.3	101.6	Or 0.5ml 1M stock
6	HEPES	10	238.3	595.8	Or 2.5ml 1M stock
	pH to 7.4				

Gain-of-function (GOF) library construction of *A. intestini*

GOF library construction was followed as described in Goodman Lab. Briefly, A. intestini genomic DNA was extracted from overnight cultures with DNeasy UltraClean Microbial Kit (Qiagen, 12224-

250). DNA was sheared to 2-8 kb by focused ultrasonication (Covaris E220 with miniTUBE red). Size exclusion was performed with chromaspin TE-1000 columns through spinning 5 min at 700 xg, keep flow-through. Or gel purification for the size of interest. Sheared DNA should be end-repaired and added phosphate groups:

1 µg sheared DNA

2 µl NEB2.1 buffer

0.2 µl T4 DNA polymerase

0.2 µl 10mM dNTP

x µl H2O 20 µl in total;

Incubate 10 min at room temperature and inactivate for 10 min at 75 °C, then add:

5 µl T4 PNK buffer 5ul

2 µl T4 PNK 2ul

23 µl H2O 50 µl in total;

Incubate 30 min at 37 °C and inactivate for 20 min at 65 °C; purify DNA with PCR purification. Fragments were cloned into PCR-linearized expression vector pZE21 by blunt-ended ligation (Epicentre FastLinkTM kit). Ligation products were gel purified for 5-10 kb and transformed into *E. coli* 10G Elite competent cells (Lucigen) by electroporation. Overnight grown colonies were picked and arrayed in 96-well format into liquid LB medium supplemented with kanamycin using a colony picking robot (Molecular Devices QPix 420). Hits were screened with GCaMP6s-based assay and validated through sequencing.

Loss-of-function (LOF) library construction of *K. pneumoniae*

LOF library construction was followed as described in Pamer Lab. Briefly, *E.coli* S17 λpir carrying pSAM_Kp2.1 (donor strain) and *K. pneumoniae* (recipient strain) were inoculated into LB with antibiotics and cultured overnight at 37 °C. Then inoculate donor strain culture to 20 ml LB (+ Kan + Strep) from OD600=0.01 and culture for ~5 h; inoculate recipient strain culture to 10 ml LB from OD600=0.01 and culture for ~2 h. When donor OD600=0.2-0.4, recipient OD600=0.4-0.6, wash 16 ml donor and 8 ml recipient twice with PBS, mix them at D:R=2:1. Spin and resuspend in ~400

μl PBS, spot 4 drops (100 μl x 4) on LB plate with filters (1 drop/filter), incubate at 37 °C overnight. Mating spots were suspended in 2 ml PBS (total 4 x 2 = 8ml), wash once with PBS and resuspend in 4 ml PBS and plate 100 μl per M9 (*E. coli* can't grow) plate (with streptomycin for selection), incubate at 37 °C overnight (no more than 20h). Overnight grown colonies were picked and arrayed in 96-well format into liquid LB medium supplemented with streptomycin using a colony picking robot (Molecular Devices QPix 420). Hits were screened with GCaMP6s-based assay and validated through NGS-bsed transposon sequencing similar to the protocol in Chapter 2.

Appendices

Table S1. Phylogenies and growth dynamics of 122 human gut isolates.

Based on Qiime 1.6, GreenGenes database:

Grow in Gifu medium

Nr	Phylum	Family	Genus	Species	48h OD600	TE(h)	TS(h)
1	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	5.4	11	14
2	Proteobacteria	Enterobacteriaceae	Other	Other	2.33	2.5	4
3	Firmicutes	Streptococcaceae	Streptococcus	luteciae	1.98	4.5	10
4	Firmicutes	Clostridiaceae	Clostridium	perfringens	3.91	4	6
5	Fusobacteria	Fusobacteriaceae	Fusobacterium	spp.	2.33	5.5	8
6	Firmicutes	Streptococcaceae	Streptococcus	spp.	1.28	10	16
7	Proteobacteria	Enterobacteriaceae	UC	UC	2.27	2.5	4
8	Proteobacteria	Enterobacteriaceae	UC	UC	2.56	3	6
9	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.94	9	12
11	Firmicutes	Veillonellaceae	Megasphaera	spp.	0.21	5	7
12	Firmicutes	Clostridiaceae	Clostridium	perfringens	4.15	3	5
13	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.71	8.5	12
14	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	1.41	10.5	16
16	Proteobacteria	Enterobacteriaceae			2.78	3.5	5
17	Proteobacteria	Enterobacteriaceae	Morganella	spp.	2.15	6	10
18	Bacteroidetes	Bacteroidaceae	Bacteroides	spp.	3.81	9.5	15
19	Bacteroidetes	Bacteroidaceae	Bacteroides	ovatus	4.31	12.5	15
21	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	2.96	12	16
22	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.88	9	12
23	Proteobacteria	Enterobacteriaceae	UC	UC	2.28	2.5	4
24	Fusobacteria	Fusobacteriaceae	Fusobacterium	spp.	2.63	7.5	10
25	Actinobacteria	Actinobacteria	Bifidobacterium	adolescentis	3.41	6	18
26	Actinobacteria	Actinobacteria	Bifidobacterium	adolescentis	3.29	8	14
27	Proteobacteria	Enterobacteriaceae	UC	UC	2.32	2.5	4
28	Proteobacteria	Enterobacteriaceae	Morganella	spp.	2.04	13	24
29	Proteobacteria	Enterobacteriaceae	UC	UC	6.51	5	24
31	Proteobacteria	Enterobacteriaceae	UC	UC	2.37	2.5	4
32	Bacteroidetes	Bacteroidaceae	Bacteroides	spp.	4.38	7.5	11
33	Proteobacteria	Enterobacteriaceae	UC	UC	2.29	2.5	4
34	Bacteroidetes	Bacteroidaceae	Bacteroides	spp.	4.16	10	14
35	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	5.39	11.5	15
36	Actinobacteria	Actinobacteria	Bifidobacterium	adolescentis	3.81	8.5	13
37	Firmicutes	Lactobacillaceae	Lactobacillus	UC	1.5	4.5	6
38	Firmicutes	Lactobacillaceae	Pediococcus	UC	0.88	5	8
39	Firmicutes	Lactobacillaceae	Pediococcus	UC	0.94	6	9
41	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.75	8.5	12
42	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	spp.	4.69	13	18

43	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	3.62	8	24
44	Proteobacteria	Enterobacteriaceae	Morganella	spp.	1.62	6	10
45	Firmicutes	Streptococcaceae	Streptococcus	spp.	2.33	6	24
46	Firmicutes	Erysipelotrichales	Erysipelotrichaceae	spp.	1.14	5.5	7
47	Firmicutes	Erysipelotrichales	Erysipelotrichaceae	spp.	1.34	5.5	7
48	Firmicutes	Clostridiaceae	Clostridium	perfringens	8.44	4	7
49	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	0.79	12	16
50	Firmicutes	Erysipelotrichales	Erysipelotrichaceae	spp.	0.9	6	8
51	Actinobacteria	Coriobacteriaceae	Collinsella	aerofaciens	1.33	8	11
52	Proteobacteria	Enterobacteriaceae	UC	UC	2.06	2.5	4
53	Actinobacteria	Coriobacteriaceae	Collinsella	aerofaciens	1.63	8.5	11
54	Firmicutes	Clostridiaceae	Clostridium	perfringens	7.54	4	7
55	Firmicutes	Streptococcaceae	Streptococcus	spp.	0.87	5	7.5
56	Bacteroidetes	Bacteroidaceae	Bacteroides	ovatus	4.14	12.5	24
57	Bacteroidetes	Bacteroidaceae	Bacteroides	spp.	4.59	12	16
58	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	5.12	8	12
59	Bacteroidetes	Bacteroidaceae	Bacteroides	UC	4.12	13.5	17
60	Proteobacteria	Enterobacteriaceae	UC	UC	2.14	2.5	4
61	Proteobacteria	Enterobacteriaceae	UC	UC	2.31	5	24
62	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.13	7	24
63	Proteobacteria	Enterobacteriaceae	UC	UC	2.1	2.5	4
64	Bacteroidetes	Porphyromonadaceae	Parabacteroides	distasonis	3.34	8	14
65	Firmicutes	Clostridiaceae	Clostridium	perfringens	9.21	6	9
66	Proteobacteria	Enterobacteriaceae	Morganella	spp.	1.58	6.5	10
67	Proteobacteria	Enterobacteriaceae	UC	UC	2.08	2.5	4
68	Firmicutes	[Tissierellaceae]	Peptoniphilus	spp.	0.98	10	20
69	Proteobacteria	Enterobacteriaceae	Morganella	spp.	1.59	5	8
70	Firmicutes	[Tissierellaceae]	Peptoniphilus	spp.	0.96	8	12
71	Proteobacteria	Enterobacteriaceae	UC	UC	1.96	2.5	4
72	Proteobacteria	Enterobacteriaceae	UC	UC	2.01	2.5	4
73	Actinobacteria	Coriobacteriaceae	Collinsella	aerofaciens	2.07	8	11
74	Firmicutes	Lachnospiraceae	[Ruminococcus]	gnavus	2.53	5	12
76	Firmicutes	Lachnospiraceae	[Ruminococcus]	gnavus	2.1	2.5	4
77	Proteobacteria	Enterobacteriaceae	UC	UC	1.93	2.5	4
78	Proteobacteria	Enterobacteriaceae	UC	UC	2.07	2.5	4
79	Proteobacteria	Enterobacteriaceae	UC	UC	1.93	2.5	4
80	Bacteroidetes	Bacteroidaceae	Bacteroides	spp.	3.69	15.5	21
81	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	2.07	11.5	20
82	Bacteroidetes	Bacteroidaceae	Bacteroides	ovatus	4.3	11.5	15
83	Actinobacteria	Coriobacteriaceae	Collinsella	stercoris	1.58	6.5	9
84	Actinobacteria	Coriobacteriaceae	Collinsella	stercoris	1.71	6	8
85	Firmicutes	Clostridiales (order)	UC	UC	0.99	9	14
87	Actinobacteria	Coriobacteriaceae	Collinsella	stercoris	1.68	6.5	9

88	Firmicutes	Lactobacillaceae	Lactobacillus	reuteri	1.34	5	8
89	Firmicutes	Lactobacillaceae	Lactobacillus	reuteri	1.42	3.5	6
90	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	1.55	11.5	20
91	Bacteroidetes	Bacteroidaceae	Bacteroides	UC	3.7	9.5	13
92	Proteobacteria	Sinobacteriaceae	Nevskia	spp.	0.15	5	8
93	Firmicutes	Lactobacillaceae	Lactobacillus	reuteri	1.2	4	10
94	Firmicutes	Lactobacillaceae	Lactobacillus	reuteri	1.27	4	6
95	Proteobacteria	Sinobacteriaceae	Nevskia	spp.	0.08	4	6
96	Firmicutes	Veillonellaceae	Acidaminococcus	spp.	0.53	7	10
97	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	1.6	14	24
98	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	1.8	11.5	20
99	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	0.81	7	10
100	Firmicutes	Veillonellaceae	Acidaminococcus	spp.	0.52	6.5	10
101	Firmicutes	Veillonellaceae	Acidaminococcus	spp.	0.45	16.5	30
102	Bacteroidetes	Bacteroidaceae	Bacteroides	UC	1.66	12	20
103	Firmicutes	Veillonellaceae	Acidaminococcus	spp.	0.4	6	10
104	Firmicutes	Veillonellaceae	Acidaminococcus	spp.	0.45	7	12
105	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	0.98	5	8
106	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	0.84	14.5	24
107	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	0.77	6	9
108	Actinobacteria	Bifidobacteriaceae	Bifidobacterium	UC	1	5	8
109	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	2	13	21
111	Firmicutes	Veillonellaceae	Megasphaera	spp.	2.22	9	14
112	Firmicutes	Lachnospiraceae	[Ruminococcus]	gnavus	1.45	12	22
113	Firmicutes	Lactobacillaceae	Pediococcus	UC	0.75	5	8
115	Bacteroidetes	Porphyromonadaceae	Parabacteroides	distasonis	2.09	12.5	20
118	Proteobacteria	Enterobacteriaceae	UC	UC	1.84	2.5	4
119	Firmicutes	Ruminococcaceae	Faecalibacterium	prausnitzii	0.64	7	24
120	Firmicutes	Clostridiaceae	Clostridium	perfringens	7.08	5.5	8
121	Firmicutes	Streptococcaceae	Streptococcus	spp.	1.53	6	10
122	Actinobacteria	Coriobacteriaceae	Eggerthela	lenta	0.13	6.5	9
127	Bacteroidetes	Bacteroidaceae	Bacteroides	uniformis	1.77	10	17
128	Firmicutes	Erysipelotrichaceae	Allobaculum	spp.	0.6	15	24
129	Bacteroidetes	Porphyromonadaceae	Parabacteroides	spp.	2.38	10.5	15
130	Actinobacteria	Coriobacteriaceae	Collinsella	stercoris	1.73	7.5	10
132	Firmicutes	[Tissierellaceae]	Peptoniphilus	spp.	0.79	8	12
133	Bacteroidetes	Bacteroidaceae	Bacteroides	fragilis	4.29	8	12
134	Firmicutes	Ruminococcaceae	Oscillospira	spp.	0.73	9	30
135	Proteobacteria	Enterobacteriaceae	Morganella	spp.	1.21	6	11
136	Proteobacteria	Enterobacteriaceae	UC	UC	2.88	3	5
139	Firmicutes	Ruminococcaceae	Oscillospira	spp.	4.2	8	24
140	Firmicutes	Ruminococcaceae	Oscillospira	spp.	0.85	17.5	30

Table S2. Quantification of relative intensity reduction (RIR, %) of linearized pUC19 plasmid DNA bands for *in vitro* gel electrophoresis-based screening.

1st screening:

Nr.	TS				TE to TS				Average (%)	
	native	0.2% NaOH	0.4% NaOH	1% NaOH	native	0.2% NaOH	0.4% NaOH	1% NaOH		
14	17.023	1.355	6.824	0.000	10.653	16.588	28.439	6.222	10.888	
21	0.000	0.000	10.992	0.000	7.099	8.992	16.475	2.785	5.793	
25	87.602	100.000	100.000	100.000	75.891	100.000	93.814	97.402	94.339	
26	69.855	87.610	92.567	88.564	93.581	100.000	100.000	100.000	91.522	
36	53.684	80.611	78.228	75.852	80.199	92.730	100.000	82.472	80.472	
42	93.552	100.000	100.000	100.000	100.000	100.000	100.000	100.000	99.194	
43	71.850	95.999	95.635	85.479	57.065	83.074	55.591	91.817	79.564	
49	18.491	20.658	31.908	8.113	1.116	32.507	35.844	21.800	21.305	
51	14.256	0.000	0.000	0.000	0.000	0.000	7.580	39.098	7.617	Actinobacteria
53	19.534	0.000	15.153	0.000	0.000	0.000	0.000	0.000	4.336	
73	11.324	4.545	0.000	0.000	0.000	0.000	1.300	0.000	2.146	
83	33.948	15.330	0.000	0.000	29.773	0.000	0.734	0.000	9.973	
84	9.518	0.000	0.000	0.000	7.116	0.000	0.000	0.000	2.079	
87	63.230	57.740	36.227	0.000	39.725	0.000	13.862	0.000	26.348	
99	1.216	0.000	0.000	4.707	0.000	0.000	7.793	3.475	2.149	
105	12.978	0.000	0.000	0.000	10.974	0.000	0.000	0.000	2.994	
106	28.614	11.726	0.000	0.000	0.000	11.500	0.000	41.437	11.660	
107	27.423	0.000	0.000	0.000	31.767	0.000	0.000	0.000	7.399	
108	13.523	0.000	0.000	0.000	10.367	0.000	0.000	0.000	2.986	
122	21.093	0.000	0.000	0.000	41.751	0.000	21.168	0.000	10.502	
130	0.000	0.000	0.000	0.000	0.000	29.291	3.630	0.000	4.115	
1	27.216	0.000	29.320	43.904	2.175	0.000	0.000	0.000	12.827	
9	33.474	0.000	2.353	0.000	0.000	0.000	0.000	17.901	6.716	
13	37.881	0.000	0.000	0.000	7.556	0.000	10.401	0.000	6.980	
18	0.000	0.000	19.851	0.000	0.000	0.000	7.951	0.000	3.475	
19	0.000	0.000	6.770	34.048	3.570	3.708	12.223	7.745	8.508	
22	33.938	0.000	22.560	7.291	0.000	0.000	0.000	0.000	7.974	Bacteroidetes
32	0.192	0.000	0.000	0.000	0.000	0.000	5.066	0.000	0.657	
34	79.939	78.634	89.640	90.542	34.490	35.189	25.723	12.814	55.871	
35	0.000	0.000	0.170	0.000	0.000	0.000	0.000	0.000	0.021	
41	31.909	0.000	0.000	0.000	11.890	0.000	0.000	0.000	5.475	
56	45.872	71.578	47.556	38.192	54.698	79.112	52.674	81.665	58.918	
57	0.000	0.000	9.576	0.000	0.000	0.736	10.183	5.062	3.195	
58	26.892	0.000	22.660	0.000	4.020	0.000	0.000	0.000	6.696	
59	93.801	100.000	100.000	100.000	38.818	67.930	56.066	46.511	75.391	
62	13.974	33.781	0.000	0.000	0.000	0.000	0.000	27.148	9.363	

64	28.797	0.000	23.924	21.853	14.543	4.081	0.000	0.000	11.650
80	45.970	45.247	3.132	16.739	0.000	0.000	0.000	5.692	14.598
81	25.752	5.683	6.757	0.000	27.950	0.000	0.000	30.492	12.079
82	0.000	0.000	17.642	0.000	0.000	0.000	1.060	23.693	5.300
90	24.500	0.000	18.814	0.000	13.254	0.000	0.000	22.490	9.882
91	100.000	100.000	100.000	100.000	95.713	100.000	100.000	98.758	99.309
97	28.659	36.426	17.699	0.000	1.398	15.574	5.274	50.124	19.394
98	20.098	0.000	23.730	0.000	5.088	0.000	0.000	20.932	8.731
102	17.722	75.668	80.748	63.445	1.350	0.000	0.000	37.590	34.565
109	41.700	50.674	21.624	32.192	0.000	0.000	0.000	0.000	18.274
115	11.039	0.000	6.350	4.660	13.899	22.968	17.538	58.441	16.862
127	3.580	37.691	30.179	0.000	8.734	26.569	11.188	0.000	14.743
129	0.000	24.925	26.532	25.642	2.216	11.735	32.925	20.154	18.016
133	37.270	0.000	18.747	0.000	3.345	0.000	10.351	22.436	11.519
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	95.568	95.670	95.095	95.111	95.287	92.597	84.742	81.216	91.911
6	100.000	100.000	100.000	100.000	100.000	100.000	100.000	100.000	100.000
11	30.335	52.068	14.520	35.246	18.661	16.727	0.000	0.000	20.945
12	63.589	90.324	90.271	92.264	55.527	85.718	79.714	77.378	79.348
37	63.665	93.643	93.605	92.614	67.965	92.108	87.754	86.980	84.792
38	81.260	100.000	100.000	100.000	92.151	100.000	100.000	77.366	93.847
39	42.621	20.447	12.759	0.000	19.224	36.817	52.423	46.936	28.903
45	52.179	74.444	42.815	36.076	28.768	45.804	16.857	70.720	45.958
46	32.744	37.730	39.175	30.748	13.863	35.092	0.000	0.000	23.669
47	28.561	22.444	7.224	35.855	13.471	3.470	0.000	0.000	13.878
48	21.403	55.428	27.049	49.429	13.049	25.680	11.451	19.387	27.859
50	88.295	100.000	100.000	100.000	100.000	100.000	100.000	79.101	95.924
54	22.034	46.306	20.570	45.862	0.000	26.102	0.000	7.582	21.057
55	20.413	38.036	12.322	6.020	18.248	42.339	18.497	0.000	19.485
65	100.000	100.000	100.000	100.000	91.871	100.000	100.000	100.000	98.984
68	27.786	0.000	7.974	0.000	23.990	0.000	0.000	44.119	12.984
70	40.147	0.000	3.858	8.371	3.702	0.000	0.000	0.000	7.010
74	25.710	12.471	5.266	27.500	18.018	28.713	33.844	34.328	23.231
76	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
85	36.340	0.000	13.295	9.088	12.049	32.101	12.751	0.000	14.453
88	12.160	0.000	0.000	0.000	5.569	0.000	0.000	0.000	2.216
89	11.386	19.421	0.000	17.674	0.000	0.000	0.000	0.000	6.060
93	0.000	52.675	55.412	62.112	0.000	0.000	36.727	28.744	29.459
94	19.609	7.051	5.010	25.599	11.074	29.751	17.195	8.174	15.433
96	1.815	21.836	22.492	39.363	0.000	22.704	46.936	46.857	25.250
100	0.000	19.110	19.297	41.303	0.000	12.348	52.902	36.400	22.670
101	43.894	85.043	77.088	58.558	16.082	78.825	65.376	94.526	64.924
103	0.000	11.486	3.280	41.889	0.000	47.300	33.702	32.058	21.214

Firmicutes

104	38.476	23.550	33.582	51.558	14.151	23.740	39.517	25.125	31.212	
111	32.862	0.000	15.031	6.118	15.570	7.751	15.824	2.970	12.016	
112	41.046	12.001	16.520	9.575	0.000	0.000	0.000	23.730	12.859	
113	3.685	0.000	0.000	0.000	10.645	46.251	41.137	35.322	17.130	
119	94.972	96.017	99.347	91.261	100.000	97.620	100.000	100.000	97.402	
120	87.965	100.000	100.000	100.000	81.872	95.440	100.000	100.000	95.660	
121	65.946	92.055	91.328	100.000	87.061	92.167	95.008	88.787	89.044	
128	23.848	23.865	0.000	0.000	0.000	0.000	0.000	0.000	5.964	
132	33.825	0.000	0.000	0.000	3.922	0.000	0.723	0.000	4.809	
134	47.957	52.430	0.000	0.000	13.169	50.766	15.668	76.806	32.099	
139	38.368	60.025	62.606	56.959	10.984	43.330	38.206	64.173	46.831	
140	38.370	50.395	0.000	0.000	0.000	10.514	0.000	51.893	18.897	
5	25.668	0.000	0.000	0.000	13.964	0.000	13.321	0.000	6.619	Fuso
24	0.000	0.000	0.000	21.136	0.000	0.000	0.000	0.000	2.642	
2	0.000	0.000	0.000	25.938	0.000	0.000	0.000	0.000	3.242	
7	0.000	0.000	0.000	5.384	0.000	0.000	0.000	0.505	0.736	
8	10.772	16.294	0.000	24.898	0.000	0.000	0.000	0.000	6.495	
16	6.845	14.459	0.000	14.996	0.000	0.000	0.000	5.599	5.237	
17	18.901	2.369	0.000	9.613	0.000	0.000	15.803	6.959	6.706	
23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
27	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
28	95.223	100.000	100.000	100.000	100.000	100.000	100.000	100.000	99.403	
29	81.293	98.287	100.000	100.000	44.824	100.000	77.173	100.000	87.697	
31	0.000	0.000	4.689	26.789	0.000	0.000	0.000	0.000	3.935	
33	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
44	2.403	0.000	0.000	0.000	0.000	0.000	20.615	10.941	4.245	
52	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	Proteobacteria
61	65.722	99.015	95.979	100.000	71.320	97.753	100.000	100.000	91.223	
63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.537	0.192	
66	11.501	12.057	10.078	18.442	0.000	0.000	51.009	17.301	15.048	
67	0.000	0.000	0.000	0.000	0.000	0.000	0.000	11.551	1.444	
69	24.095	0.000	22.807	0.000	30.514	0.000	22.374	18.608	14.800	
71	1.490	0.000	28.189	0.000	0.000	0.000	0.000	0.000	3.710	
72	0.000	0.000	2.098	0.000	0.000	0.000	0.000	0.000	0.262	
77	0.000	0.000	18.467	33.197	0.000	0.000	0.000	0.000	6.458	
78	0.000	13.490	0.000	0.000	25.263	64.090	1.053	34.684	17.322	
79	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
92	15.224	0.000	11.128	0.136	12.272	0.000	0.000	0.000	4.845	
95	7.187	16.580	0.000	9.879	0.000	0.000	0.000	0.000	4.206	
118	0.000	0.000	0.000	4.944	0.000	0.000	0.000	0.000	0.618	
135	96.669	100.000	100.000	100.000	95.827	93.940	94.570	92.828	96.729	
136	21.435	10.961	24.204	28.559	0.000	0.000	0.000	0.000	10.645	

2nd screening:

Nr.	TE				TS				
	native	0.2%	0.4%	1%	native	0.2%	0.4%	1%	
25	0.000	0.000	4.740	10.425	17.275	3.919	0.000	15.858	
26	20.219	19.404	9.652	0.000	44.150	54.654	0.000	0.000	
36	0.000	0.000	0.000	0.000	7.752	0.000	1.333	3.526	Act.
42	37.912	47.347	49.253	48.251	39.505	66.947	24.666	50.992	
43	0.057	0.000	0.000	0.000	29.521	39.842	8.284	10.996	
106	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
34	16.432	0.000	1.438	0.000	13.322	23.782	0.000	0.000	
56	2.555	5.072	0.000	0.000	6.750	18.610	0.000	2.614	
59	0.000	0.000	0.000	6.440	0.000	0.000	0.000	0.000	
62	45.419	52.386	36.647	57.461	11.739	0.612	0.000	0.000	Bac.
91	10.415	0.000	0.000	0.000	16.565	0.000	0.000	0.000	
97	0.000	0.000	5.318	36.900	8.561	0.000	0.000	33.650	
102	41.154	0.000	55.659	34.628	23.303	0.000	0.000	26.848	
115	1.208	0.000	8.218	15.318	0.000	0.000	0.000	0.000	
4	92.151	92.143	93.377	92.822	100.000	100.000	100.000	100.000	
6	100.000	100.000	100.000	95.072	100.000	100.000	100.000	100.000	
12	37.165	62.435	67.022	75.402	69.082	100.000	90.845	100.000	
37	50.561	76.681	81.275	75.954	57.132	86.073	83.083	85.625	
38	0.000	3.794	35.658	36.040	41.651	100.000	95.381	100.000	
39	0.000	0.000	0.000	0.000	3.098	9.263	0.000	16.689	
45	14.102	0.000	0.000	0.000	1.388	0.000	0.000	0.000	
50	93.085	95.847	93.928	90.567	89.988	88.876	88.160	75.373	
65	2.564	2.600	0.000	0.000	33.365	60.324	45.928	57.239	
74	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
93	0.000	0.000	0.000	0.000	1.201	0.000	0.000	0.000	
96	0.000	1.368	0.000	0.000	0.000	6.500	0.000	0.000	Fir.
100	3.457	28.960	25.929	13.486	1.190	25.021	0.000	0.000	
101	1.095	20.415	15.509	5.027	0.000	24.268	0.000	0.000	
103	0.000	17.347	10.691	29.172	0.451	17.973	0.000	0.000	
104	0.748	6.315	1.225	0.000	0.961	3.933	0.000	0.000	
113	0.000	0.000	0.000	0.000	6.410	19.816	13.935	18.432	
119	0.000	0.000	0.000	0.772	0.000	4.477	0.000	7.559	
120	43.685	92.551	85.044	78.356	95.655	100.000	100.000	94.754	
121	30.924	31.940	16.718	0.000	69.577	68.312	55.805	70.215	
134	19.249	28.021	15.338	10.218	0.133	0.000	0.000	0.000	
139	17.404	8.050	0.000	1.711	17.331	31.737	0.000	13.463	
140	0.000	0.000	10.697	0.000	9.426	0.000	0.000	3.608	
28	9.796	11.385	2.332	0.000	26.630	13.989	0.000	0.000	
29	4.341	0.000	0.000	0.000	21.859	3.652	0.000	21.275	
61	4.876	0.000	0.000	0.000	10.199	14.718	0.000	16.125	Pro.
69	2.690	0.000	0.000	0.000	13.123	38.477	0.000	0.000	
135	65.479	67.403	62.220	58.027	72.900	80.455	61.095	57.603	

Nr.	TE to TS				TS supernatant				Average (%)
	native	0.2%	0.4%	1%	native	0.2%	0.4%	1%	
25	4.097	26.585	7.909	27.031	23.737	36.022	15.599	37.635	17.972
26	56.588	67.157	57.556	66.951	86.094	97.004	89.354	95.242	59.562
36	5.679	26.900	18.416	19.384	28.331	53.245	35.713	39.341	19.968
42	56.778	70.234	62.229	66.087	56.524	69.992	70.983	63.957	58.241
43	9.165	60.343	39.923	55.724	0.000	0.000	0.000	0.000	21.150
106	8.552	20.621	11.246	35.639	17.027	9.697	5.531	30.183	11.541
34	12.685	0.000	0.000	9.772	19.574	17.467	0.000	0.000	8.050
56	1.463	0.000	0.000	0.903	5.097	20.526	8.086	0.000	5.337
59	0.000	0.000	0.000	0.000	0.000	0.000	1.598	20.588	1.849
62	38.413	29.860	0.000	0.000	24.656	29.876	29.742	3.745	14.054
91	9.955	0.000	0.000	6.007	18.024	12.234	0.000	5.640	5.702
97	0.000	0.000	0.000	3.262	2.157	28.113	36.136	0.000	9.323
102	1.590	0.000	0.000	0.000	0.000	0.000	0.000	8.781	5.044
115	2.593	0.000	0.000	0.000	11.026	0.000	22.685	0.000	3.025
4	100.000	99.757	99.644	100.000	70.496	96.586	96.013	93.642	96.345
6	100.000	100.000	100.000	95.992	100.000	100.000	100.000	100.000	99.666
12	62.612	96.197	92.887	95.874	53.845	96.109	95.399	86.523	86.614
37	64.336	88.362	95.179	83.664	77.113	83.528	73.595	82.523	80.018
38	21.857	92.233	91.635	85.966	79.570	96.484	89.283	94.373	82.369
39	0.000	20.011	0.000	6.325	0.000	56.415	39.202	48.664	16.639
45	7.549	19.069	17.820	0.000	9.448	11.533	0.000	0.000	5.567
50	95.907	96.508	89.329	84.345	1.325	40.975	0.000	0.000	62.565
65	4.710	60.731	27.057	28.217	0.000	27.474	0.000	22.523	30.631
74	0.000	10.472	0.000	3.290	0.000	0.000	0.000	0.000	1.147
93	0.000	0.000	0.000	1.417	0.000	0.000	9.409	8.562	1.716
96	0.000	56.259	26.813	22.340	0.000	28.722	0.000	0.000	11.719
100	0.000	58.990	11.703	9.397	0.000	31.366	0.000	0.000	11.472
101	0.000	36.274	13.409	13.275	0.000	5.406	0.000	0.000	7.719
103	0.000	24.502	0.000	13.199	9.947	13.601	18.163	3.545	8.449
104	0.000	48.370	8.633	25.106	0.000	7.991	0.000	0.000	7.916
113	2.995	32.231	30.690	20.287	14.658	56.702	65.139	61.986	28.607
119	19.561	37.222	19.247	27.580	0.000	0.000	0.000	11.701	10.612
120	92.650	100.000	98.013	100.000	29.231	88.198	78.793	75.686	87.748
121	87.733	100.000	86.164	88.883	81.376	93.349	82.165	70.548	79.510
134	29.252	0.000	8.729	13.534	24.968	26.913	0.000	5.308	9.070
139	20.465	0.000	0.000	17.119	28.011	27.200	0.000	0.000	12.944
140	0.000	0.000	0.000	29.025	9.570	2.438	6.381	1.914	5.197
28	13.195	37.857	0.000	0.000	39.686	25.737	15.602	0.000	14.391
29	0.000	18.375	5.825	0.000	0.000	4.508	0.000	0.000	6.291
61	1.537	22.237	0.000	0.000	0.000	5.924	0.000	6.131	6.406
69	8.886	35.718	10.049	0.000	32.430	60.524	21.227	26.664	20.592
135	72.068	86.162	64.557	70.445	96.768	100.000	100.000	96.896	79.913

Table S3. Biosynthetic gene cluster (BGC) results for *C. perfringens*, *C. ramosum*, and *M. morganii*. BGCs were identified using antiSMASH.

Species	*Clostridium perfringens*	**Strain**	NWP4			
	Type	**From**	**To**	**Most similar**		**Similarity**
9.1	RiPP-like	36,934	49,138			
20.1	saccharide	16,876	33,253			
23.1	saccharide	1,818	27,875			
24.1	ranthipeptide	1	13,871			
27.1	saccharide	1	25,662			
28.1	fatty_acid	1	23,457	dactylocycline A	Polyketide	5%
30.1	saccharide	3,749	23,327			
31.1	saccharide	1	23,109	S-layer glycan	Saccharide	9%
38.1	saccharide	672	21,325	S-layer glycan	Saccharide	26%
46.1	saccharide	1	18,828			
48.1	saccharide	1	18,325			
70.1	saccharide	1	13,287			
77.1	saccharide	1	11,947			
91.1	saccharide	1	9,868			
95.1	saccharide	1	9,566			
99.1	saccharide	1	9,115			
143.1	saccharide	1	6,006			
158.1	cyclic-lactone-autoinducer	1	4,847			
Species	*Clostridium perfringens*	**Strain**	NWP12			
	Type	**From**	**To**	**Most similar**		**Similarity**
7.1	saccharide	20,378	49,887	S-layer glycan	Saccharide	26%
9.1	saccharide	7,904	49,474			
25.1	fatty_acid	4,517	28,532	dactylocycline A	Polyketide	5%
26.1	saccharide	8,546	29,726			
27.1	ranthipeptide	1	16,778			
32.1	saccharide	1	26,006			
33.1	saccharide	1	24,546	S-layer glycan	Saccharide	9%
46.1	RiPP-like	5,647	17,851			
47.1	saccharide	1	14,210			
56.1	saccharide	1	18,371			
66.1	saccharide	1	16,163			
76.1	saccharide	1	13,975			
79.1	cyclic-lactone-autoinducer	729	13,484			
85.1	saccharide	1	12,350			
103.1	saccharide	1	10,216			
119.1	saccharide	1	8,184			
120.1	cyclic-lactone-autoinducer	1	8,099			
140.1	saccharide	1	6,852			
Species	*Clostridium perfringens*	**Strain**	NWP65			
	Type	**From**	**To**	**Most similar**		**Similarity**

1.1	saccharide	103,415	136,214			
5.1	cyclic-lactone-autoinducer	64,912	78,569			
8.1	cyclic-lactone-autoinducer, saccharide	11,550	43,861			
15.1	ranthipeptide	7,349	28,932			
25.1	saccharide	5,987	41,549			
26.1	saccharide	1	19,636			
31.1	saccharide	1	22,632			
34.1	saccharide	754	30,263	S-layer glycan	Saccharide	26%
36.1	saccharide	13,396	28,336			
50.1	saccharide	1,914	21,151	emulsam	Saccharide	9%
52.1	fatty_acid	1	19,762	dactylocycline A	Polyketide	5%
58.1	saccharide	1	18,371			
59.1	saccharide	1	18,021			
63.1	saccharide	1	17,578			
77.1	saccharide	1	12,236			
82.1	saccharide	1	10,531			
Species	*Clostridium perfringens*	**Strain**	NWP120			
	Type	**From**	**To**	**Most similar**		**Similarity**
8.1	saccharide	20,588	53,553			
11.1	cyclic-lactone-autoinducer	9,853	30,416			
21.1	saccharide	1	37,559			
24.1	ranthipeptide	5,140	26,723			
33.1	saccharide	1	23,102			
35.1	saccharide	13,341	28,235			
39.1	saccharide	7,917	26,025			
40.1	fatty_acid	1	22,360	dactylocycline A	Polyketide	5%
42.1	saccharide	754	24,554	S-layer glycan	Saccharide	26%
45.1	saccharide	1,864	22,757	emulsam	Saccharide	9%
55.1	saccharide	1,614	19,493			
62.1	saccharide	1	17,324			
63.1	saccharide	1	17,312			
64.1	saccharide, cyclic-lactone-autoinducer	1	17,223			
68.1	saccharide	1	16,312			
93.1	saccharide	1	12,139			
103.1	saccharide	1	10,116			
Species	*Clostridium ramosum*	**Strain**	NWP50			
	Type	**From**	**To**	**Most similar**		**Similarity**
1.1	saccharide	1	17,936			
3.1	saccharide	12,765	36,537	locillomycin	NRP + Polyketide	14%
12.1	cyclic-lactone-autoinducer	1	16,235			
20.1	saccharide	1	24,674	exopolysaccharide	Saccharide	13%
22.1	fatty_acid	79	24,285	isoindolinomycin	Polyketide	5%
26.1	cyclic-lactone-autoinducer	12,271	23,737			
29.1	cyclic-lactone-autoinducer	1	19,165			

46.1	cyclic-lactone-autoinducer	1	13,628			
72.1	saccharide	1	11,834			
77.1	saccharide	1	10,971	polysaccharide B	Saccharide	6%
83.1	saccharide	1	10,425			
98.1	saccharide	1	9,532			
117.1	saccharide	1	8,143			
137.1	saccharide	1	6,933			
146.1	saccharide	1	6,687			
147.1	saccharide	1	6,566			
152.1	saccharide	1	6,512			
162.1	saccharide	1	6,180			
181.1	saccharide	1	5,532			
222.1	saccharide	1	4,407			
259.1	saccharide	1	3,809			
Species	*Morganella morganii*	**Strain**	NWP69			
	Type	**From**	**To**	**Most similar**		**Similarity**
2.1	fatty_acid, saccharide	36,705	82,581			
2.2	saccharide	201,760	222,999			
2.3	saccharide	317,572	338,585	polysaccharide B	Saccharide	6%
4.1	saccharide, betalactone	11,743	52,449			
8.1	saccharide	87,232	124,940	arginomycin	Other	6%
9.1	fatty_acid	41,314	62,267			
10.1	thiopeptide	23,628	49,757	O-antigen	Saccharide	14%
17.1	saccharide	1	39,605	O&K-antigen	Saccharide	6%
19.1	NRPS, T1PKS	40,234	81,433			
20.1	saccharide	13,551	52,631			
24.1	saccharide	23,254	52,948	O&K-antigen	Saccharide	3%
30.1	arylpolyene	8,727	40,875	APE Ec	Other	84%
Species	*Morganella morganii*	**Strain**	NWP135			
	Type	**From**	**To**	**Most similar**		**Similarity**
2.1	fatty_acid, saccharide	36,705	82,581			
2.2	saccharide	201,760	222,999			
2.3	saccharide	317,572	338,585	polysaccharide B	Saccharide	6%
3.1	saccharide	326,546	364,254	arginomycin	Other	6%
4.1	NRPS, T1PKS	40,234	95,563	yersiniabactin	NRP + Polyketide	2%
5.1	saccharide	78,904	127,369	O&K-antigen	Saccharide	6%
10.1	fatty_acid	41,314	62,267			
11.1	saccharide	13,551	52,631			
16.1	saccharide, betalactone	11,743	52,449			
20.1	saccharide	23,254	52,948	O&K-antigen	Saccharide	3%
25.1	arylpolyene	1	32,138	APE Ec	Other	84%

Table S4. Initial ion list for bacterial metabolites generated by comparative metabolomics with XCMS and MPP.

The finalized four ion features (I–IV) are highlighted.

m/z	Rt (min)	max intensity	*m/z*	Rt (min)	max intensity	*m/z*	Rt (min)	max intensity
147.012	4.71	4.19E+05	325.076	3.42	3.50E+04	165.004	5.54	3.02E+03
536.576	35.02	6.58E+03	106.073	7.86	1.13E+06	782.633	5.64	2.05E+03
236.058	3.86	2.49E+06	262.036	34.23	1.90E+04	395.067	13.3	5.62E+03
118.061	3.14	3.88E+06	178.123	22.08	8.34E+03	247.021	27.27	1.20E+03
536.576	35.87	7.92E+03	51.0224	7.81	2.03E+04	187.562	2.77	9.31E+04
536.576	35.55	8.14E+03	384.157	4.16	6.10E+05	439.144	13.23	1.94E+04
536.576	34.24	5.06E+03	80.0575	7.86	1.06E+04	335.229	3.77	4.70E+04
122.08	7.7	1.02E+06	107.077	7.86	5.61E+04	407.229	7.79	9.05E+04
279.098	7.33	1.25E+06	398.264	7.9	6.73E+03	215.103	11.54	2.83E+04
206.049	7.81	4.60E+03	79.054	7.82	8.73E+04	301.056	2.79	7.75E+03
343.132	35.95	9.37E+03	353.068	2.99	8.26E+04	809.617	6.37	9.25E+02
260.096	4.49	9.45E+05	201.196	16.39	2.67E+03	1256.91	7.11	1.03E+03
136.997	4.1	4.05E+06	181.061	6.3	1.18E+05	1215.95	7.11	7.62E+03
188.063	2.77	2.42E+04	77.0385	7.86	1.81E+05	295.078	3.09	1.64E+05
187.076	2.79	2.03E+04	195.063	8.14	1.48E+05	299.087	15.17	1.19E+05
264.033	34.22	1.92E+03	1237.44	7.11	5.33E+03	1226.94	7.12	6.80E+03
1238.44	7.12	3.63E+03	78.0418	7.81	2.26E+04	389.05	13.3	7.77E+03
216.923	5.75	1.50E+03	115.029	7.82	7.48E+03	242.153	6.74	5.70E+05
248.103	4.29	3.14E+05	248.052	6.14	1.17E+05	168.015	4.12	9.83E+03
340.166	4.71	2.15E+05	442.171	10.61	6.23E+05	661.323	29.77	1.84E+03
321.063	13.3	2.65E+04	118.065	7.86	4.04E+05	1238.94	7.12	2.49E+03
479.125	13.3	1.57E+04	473.107	13.3	1.90E+04	95.0492	7.85	1.90E+03
1226.44	7.1	5.55E+03	104.058	7.81	1.19E+04	810.639	7.1	2.92E+03
263.039	34.23	3.56E+03	271.15	7.63	2.85E+05	131.069	20.19	9.58E+03
796.968	6.38	3.18E+03	316.099	4.48	1.86E+05	1009.11	5.46	4.45E+03
130.123	17.09	3.27E+04	206.082	17.82	5.66E+03	52.0259	7.82	2.01E+03
138.062	4.72	1.26E+05	162.092	20.69	9.52E+04	1216.44	6.4	2.21E+03
796.635	6.41	2.68E+03	585.174	13.23	8.30E+03	1234.43	7.13	2.20E+03
122.096	7.86	1.27E+06	103.054	7.82	7.45E+04	541.094	13.29	3.28E+03
215.153	19.83	1.42E+04	362.089	13.3	1.86E+04	1237.94	7.11	5.88E+03
360.149	15.81	7.62E+04	215.023	13.3	8.76E+03	424.06	13.3	3.11E+03
331.031	5.75	2.08E+03	120.065	7.82	5.78E+04	144.081	20.69	1.54E+05
123.1	7.89	1.08E+05	105.045	7.82	1.18E+04	294.939	5.73	1.46E+03
105.07	7.86	6.68E+06	119.068	7.81	3.52E+04	144.138	20.47	7.77E+03
121.014	7.62	5.41E+03	175.034	11.8	2.07E+04			

Table S5. Decarboxylases of *M. morganii* NWP135.

contig_id	feature_id	function
NODE_1_length_425693_cov_17.074003	peg.1085	Glutamate decarboxylase
NODE_1_length_425693_cov_17.074003	peg.1320	Histidine decarboxylase (EC 4.1.1.22)
NODE_20_length_63183_cov_18.125495	peg.1389	Phosphatidylserine decarboxylase (EC 4.1.1.65)
NODE_20_length_63183_cov_18.125495	peg.1405	Diaminopimelate decarboxylase and/or diaminopimelate epimerase leader peptide
NODE_21_length_56420_cov_19.612783	peg.1442	L-threonine 3-O-phosphate decarboxylase (EC 4.1.1.81)
NODE_26_length_38472_cov_21.694815	peg.1704	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)
NODE_2_length_393565_cov_15.440495	peg.2014	Benzoylformate decarboxylase (EC 4.1.1.7)
NODE_2_length_393565_cov_15.440495	peg.2226	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23)
NODE_11_length_110829_cov_16.355085	peg.234	Ornithine decarboxylase (EC 4.1.1.17)
NODE_4_length_278258_cov_19.281330	peg.2851	Diaminopimelate decarboxylase (EC 4.1.1.20)
NODE_5_length_257198_cov_21.050687	peg.3098	Glutamate decarboxylase (EC 4.1.1.15)
NODE_5_length_257198_cov_21.050687	peg.3145	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36) / Phosphopantothenoylcysteine synthetase (EC 6.3.2.5)
NODE_5_length_257198_cov_21.050687	peg.3258	Possible carboxymuconolactone decarboxylase family protein (EC 4.1.1.44)
NODE_13_length_101122_cov_16.218392	peg.392	Pyruvate decarboxylase (EC 4.1.1.1); Alpha-keto-acid decarboxylase (EC 4.1.1.-)
NODE_9_length_137453_cov_20.980047	peg.3926	Ornithine decarboxylase (EC 4.1.1.17)
NODE_9_length_137453_cov_20.980047	peg.3936	4-carboxymuconolactone decarboxylase domain/alkylhydroperoxidase AhpD family core domain protein
NODE_14_length_93994_cov_21.112738	peg.512	Lysine decarboxylase family
NODE_16_length_90343_cov_15.968143	peg.725	Biosynthetic arginine decarboxylase (EC 4.1.1.19)

NMR structure raw data Figure S10-41.

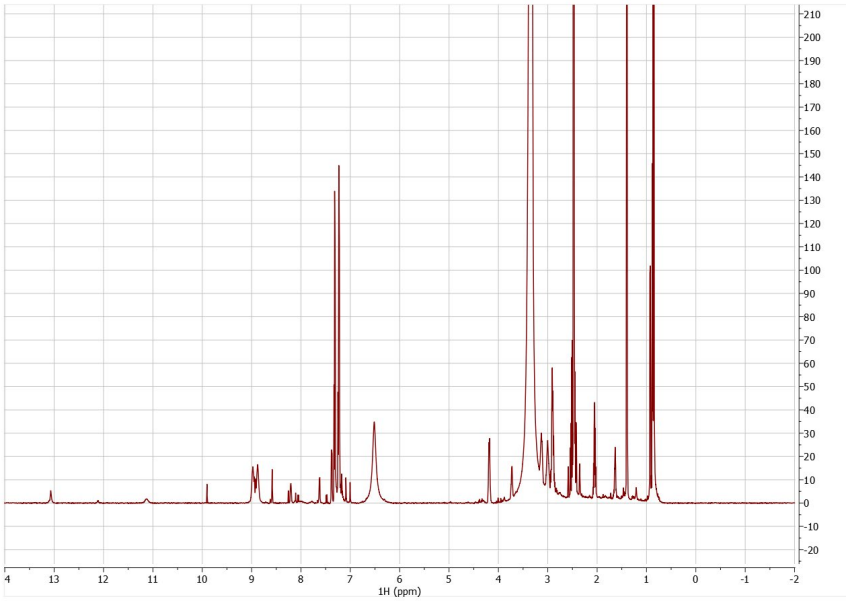


Figure S10. ^{1}H NMR spectrum of the mixture of natural indolimine-214 (1) and compound 2 in DMSO-d_6

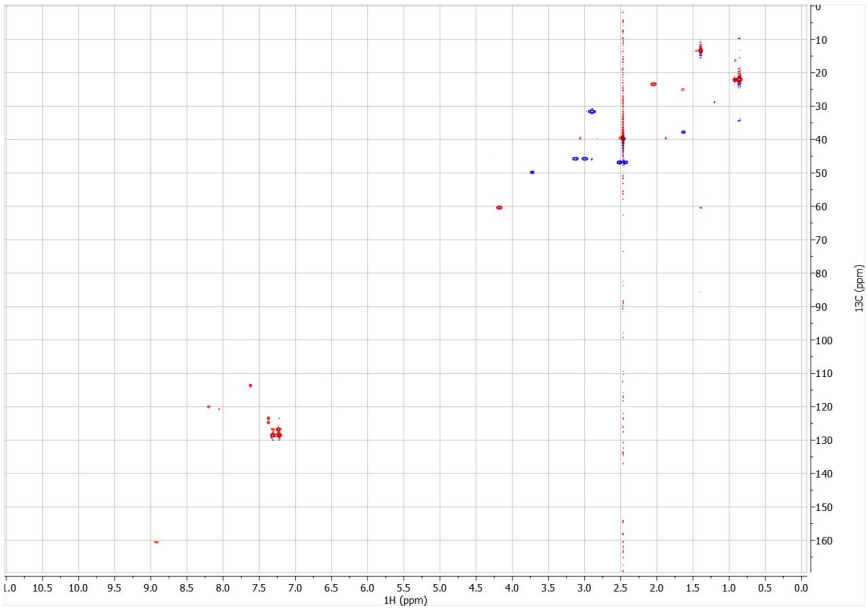


Figure S11. COSY NMR spectrum of the mixture of natural indolimine-214 (1) and compound 2 in DMSO-d_6

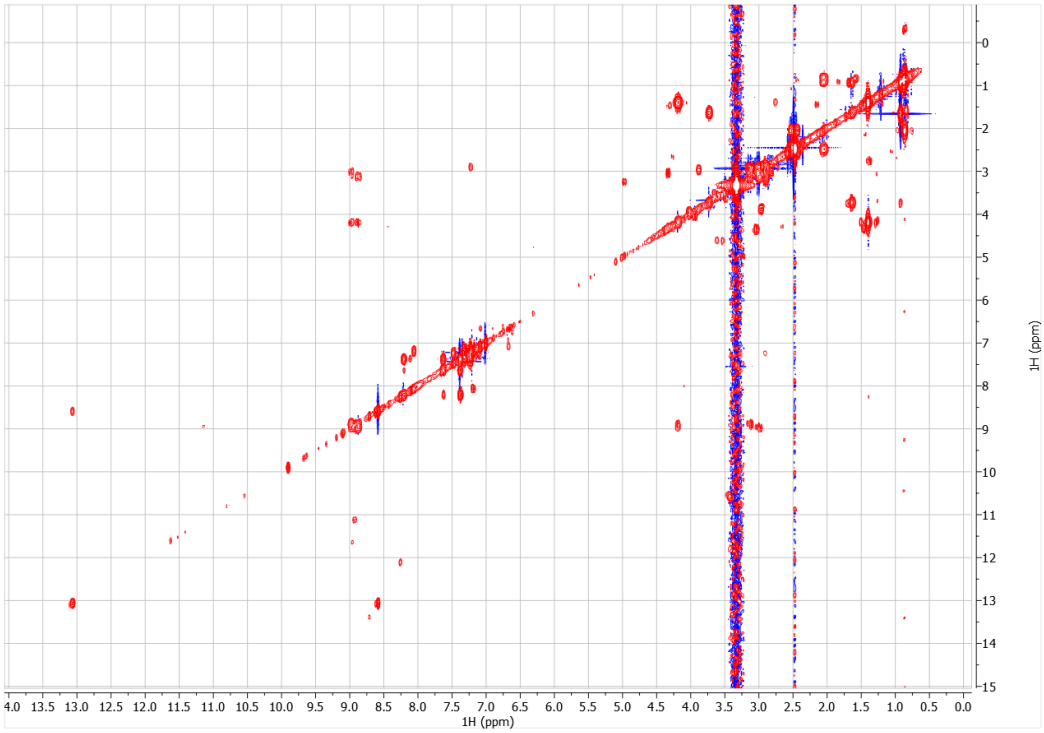


Figure S12. HSQC NMR spectrum of the mixture of natural indolimine-214 (1) and compound 2 in DMSO-d_6

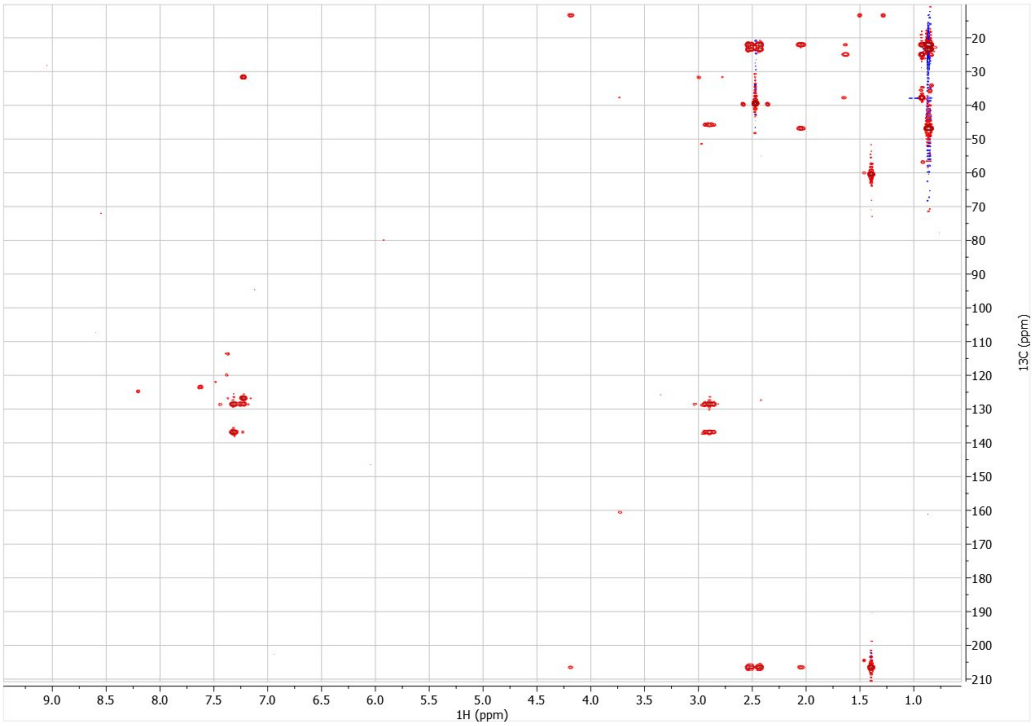


Figure S13. HMBC NMR spectrum of the mixture of natural indolimine-214 (1) and compound 2 in DMSO-d_6

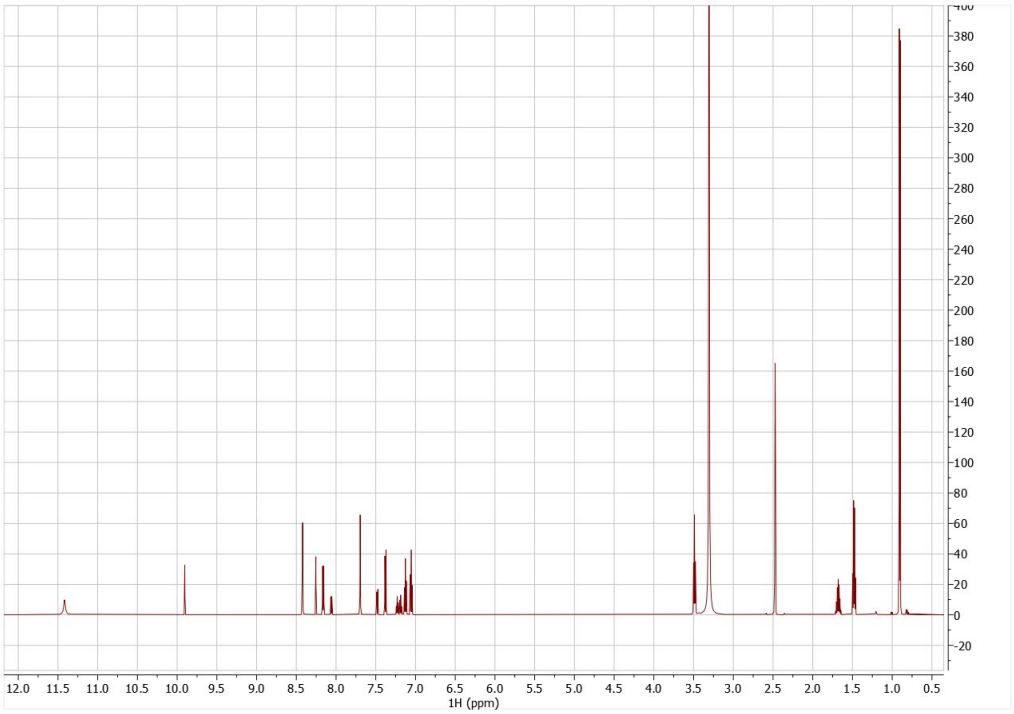


Figure S14. ^{1}H NMR spectrum of synthetic indolimine-214 (1) in DMSO-d_6

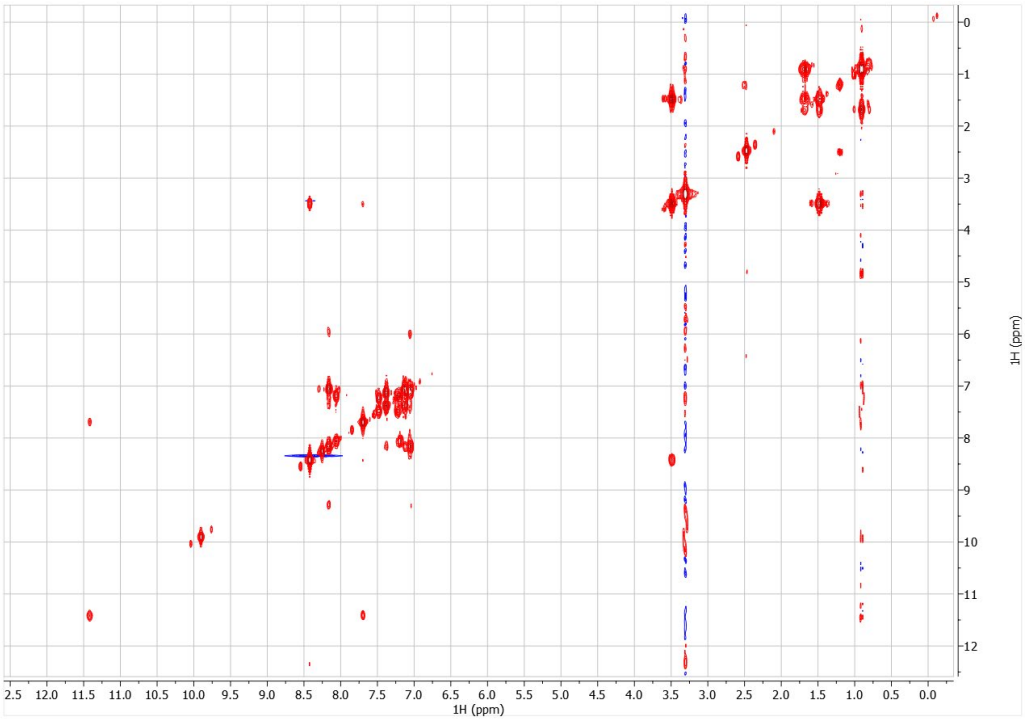


Figure S15. COSY NMR spectrum of synthetic indolimine-214 (1) in DMSO-d_6

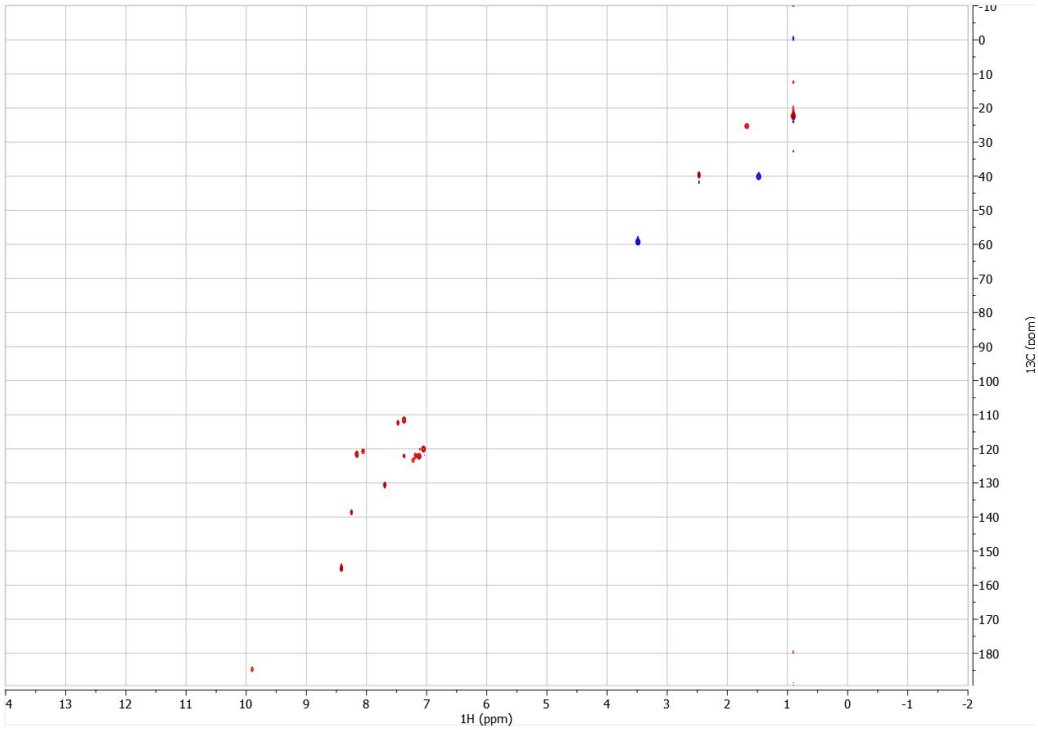


Figure S16. HSQC NMR spectrum of synthetic indolimine-214 (1) in DMSO-d_6

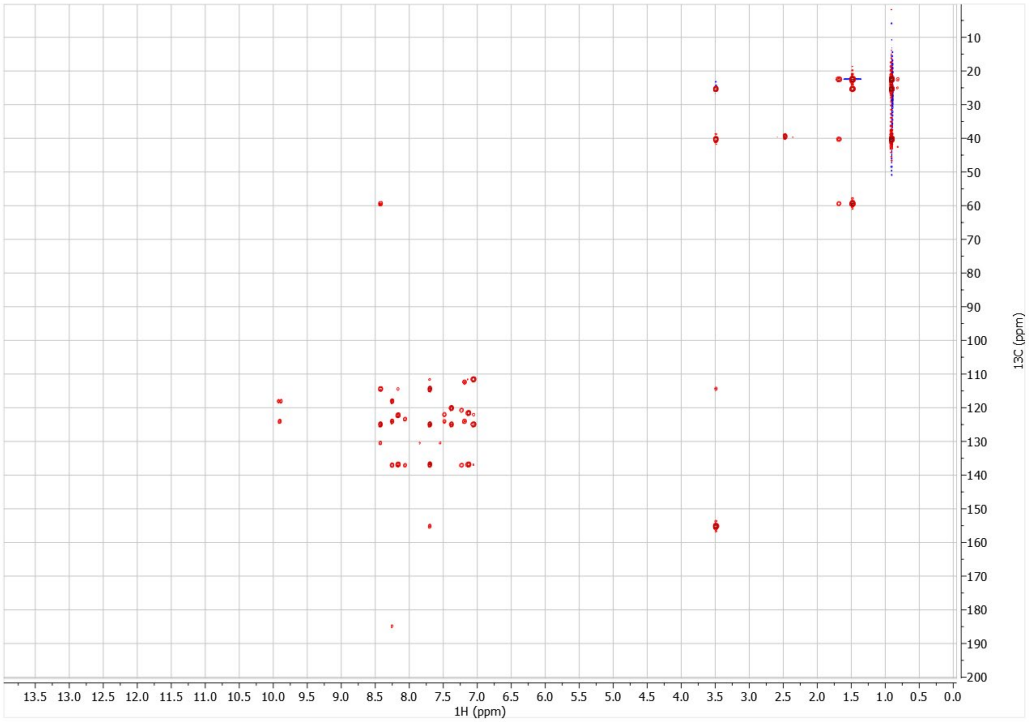


Figure S17. HMBC NMR spectrum of synthetic indolimine-214 (1) in DMSO-d_6

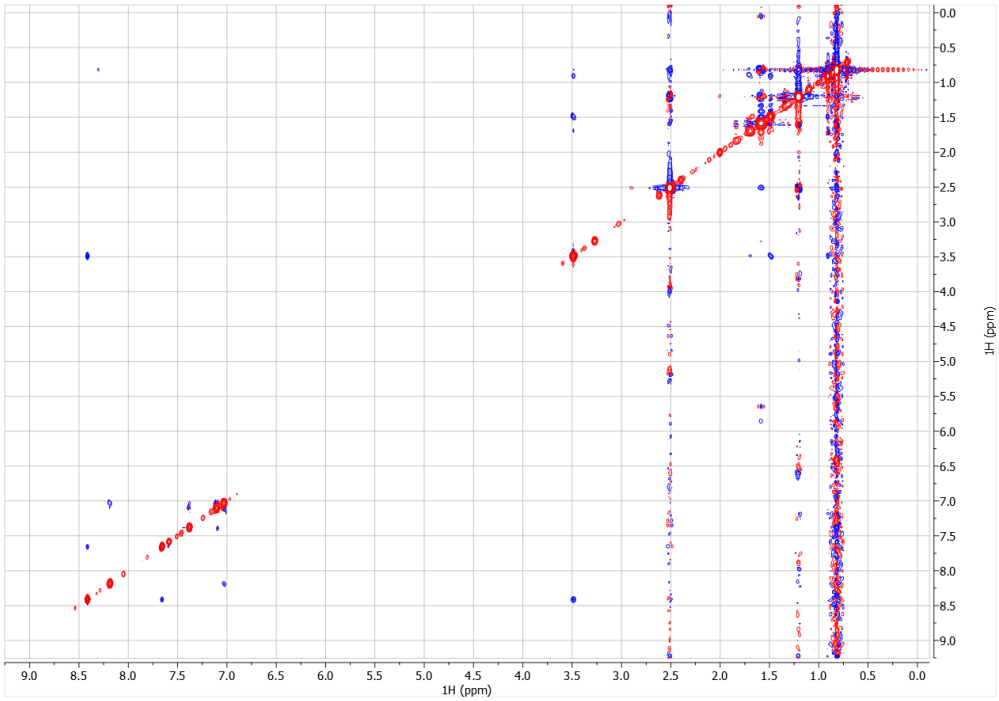


Figure S18. ROESY NMR spectrum of synthetic indolimine-214 (1) in DMSO-d_6 (mixing time: 300 ms)

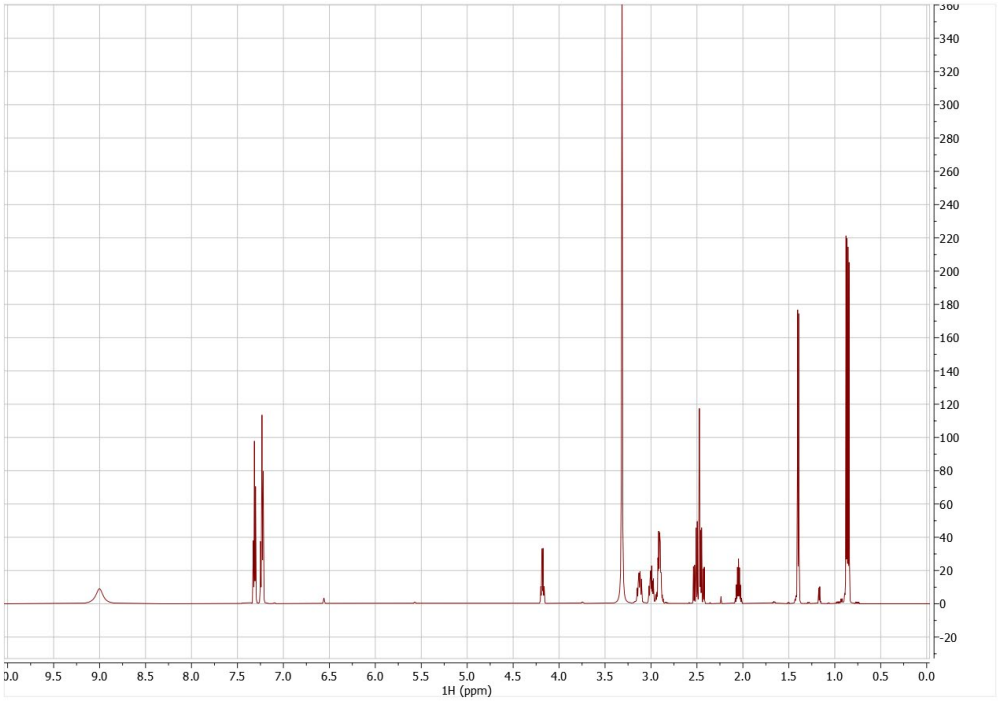


Figure S19. ^{1}H NMR spectrum of synthetic compound 2 in DMSO-d_6

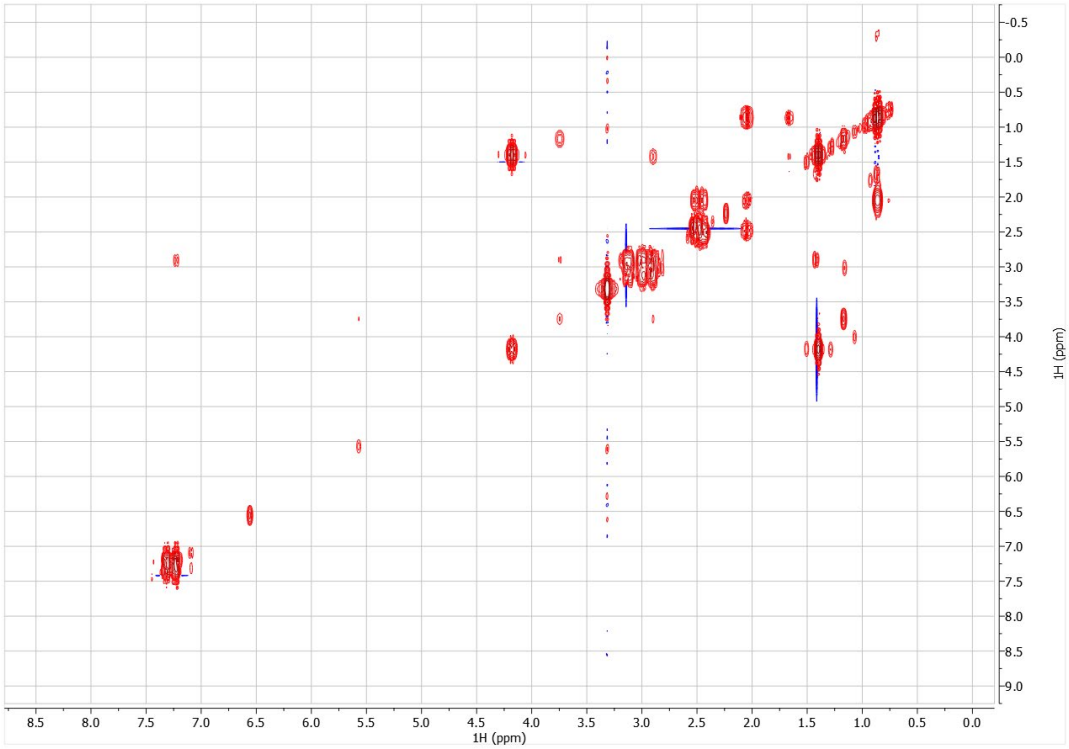


Figure S20. COSY NMR spectrum of synthetic compound 2 in DMSO-d_6

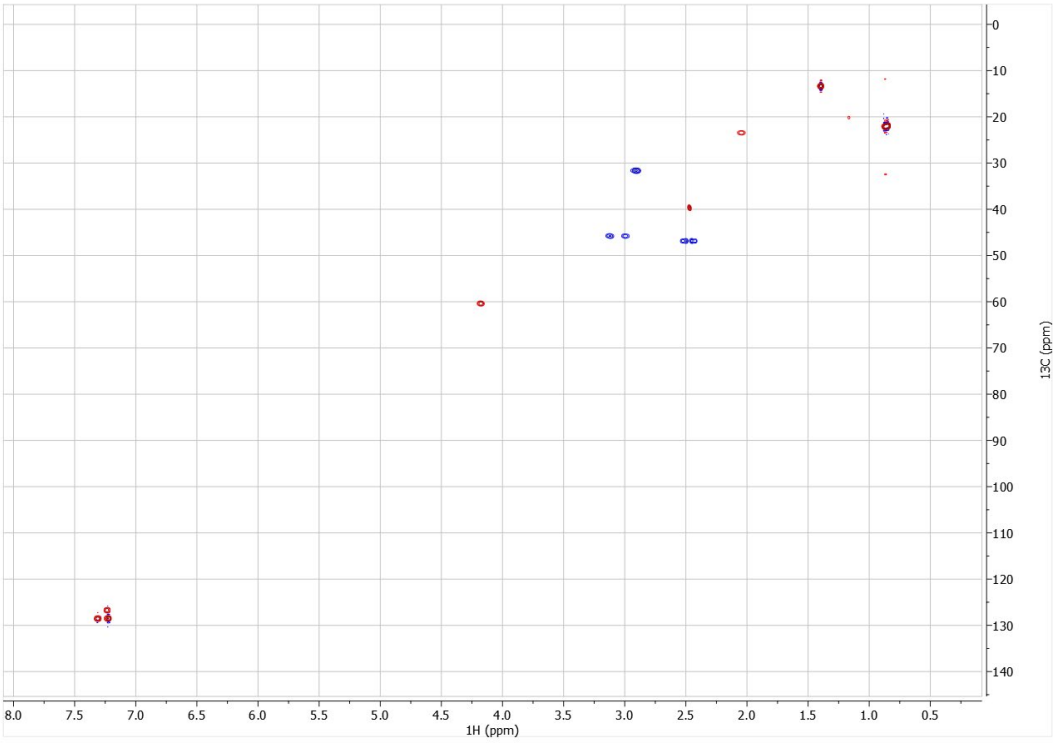


Figure S21. HSQC NMR spectrum of synthetic compound 2 in DMSO-d_6

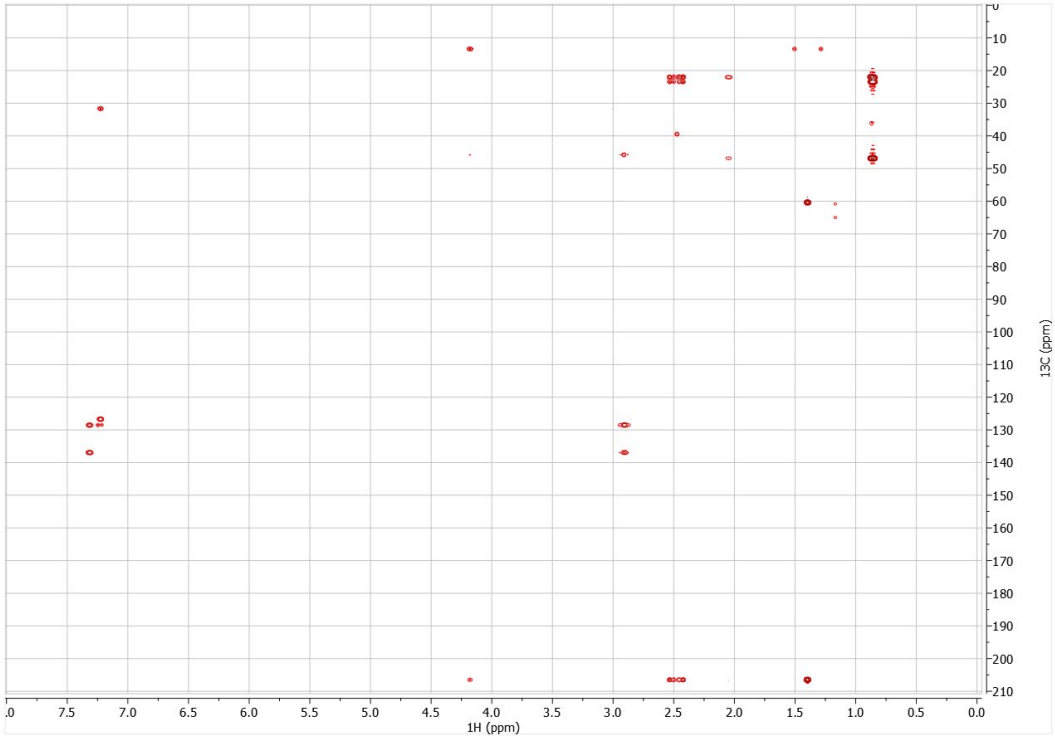


Figure S22. HMBC NMR spectrum of synthetic compound 2 in DMSO-d_6

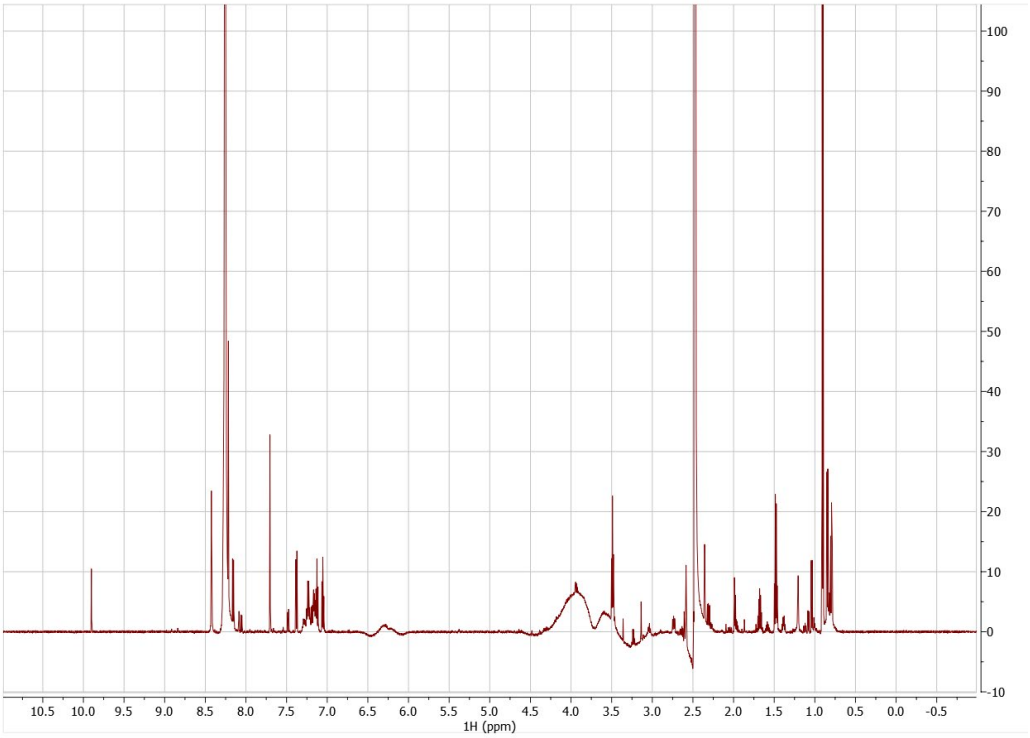


Figure S23. ^{1}H NMR spectrum of natural indolimine-214 (1) in DMSO-d_6

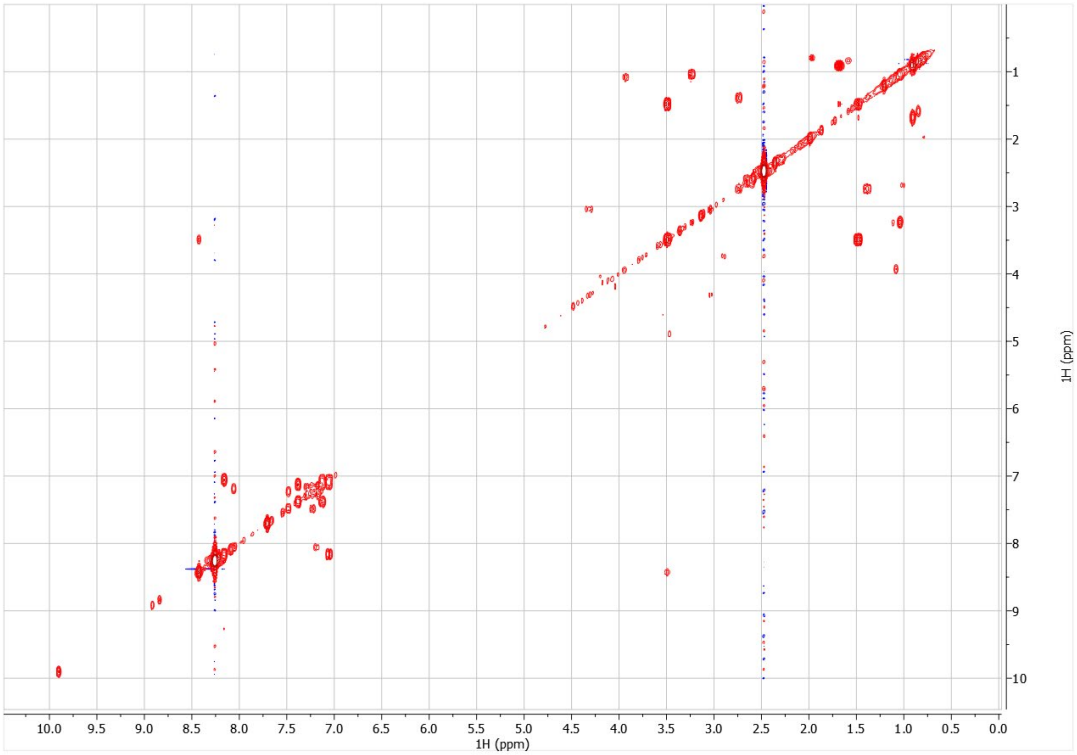


Figure S24. COSY NMR spectrum of natural indolimine-214 (1) in DMSO-d_6

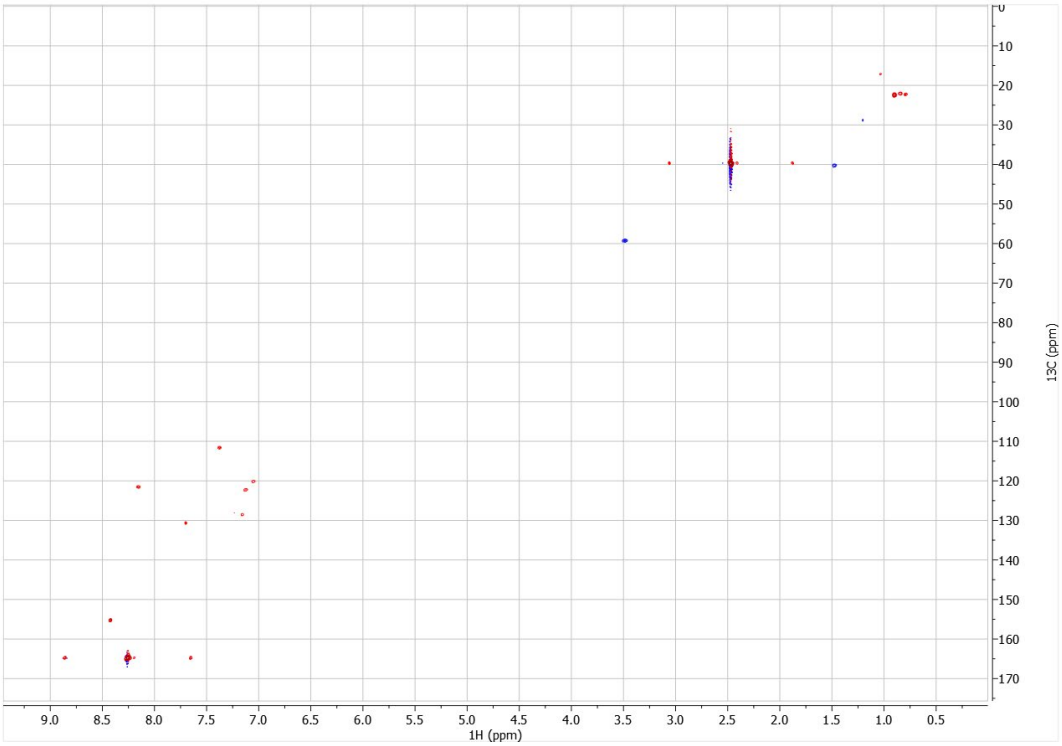


Figure S25. HSQC NMR spectrum of natural indolimine-214 (1) in DMSO-d_6

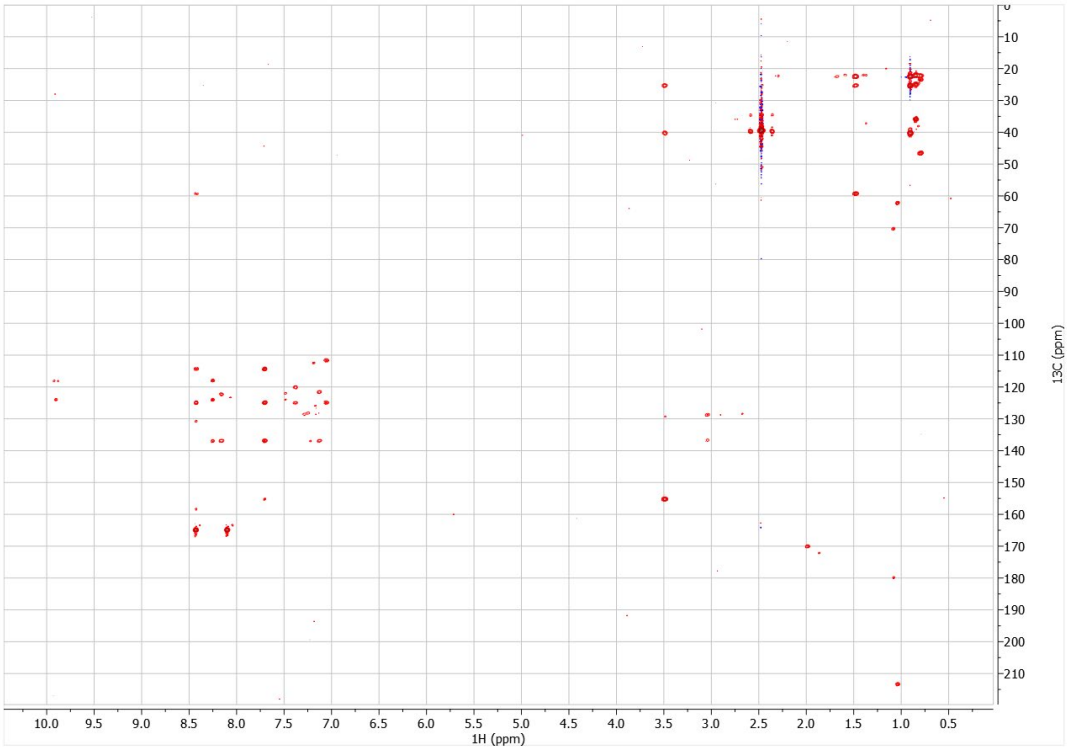


Figure S26. HMBC NMR spectrum of natural indolimine-214 (1) in DMSO-d_6

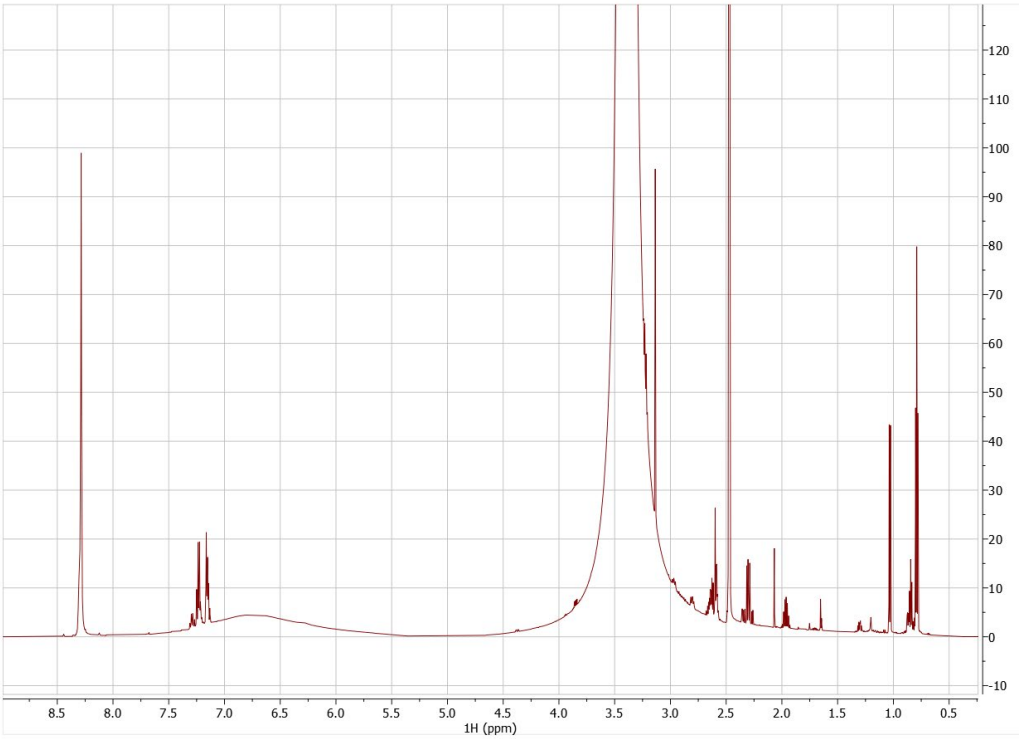


Figure S27. ^{1}H NMR spectrum of natural compound 2 in DMSO-d_6

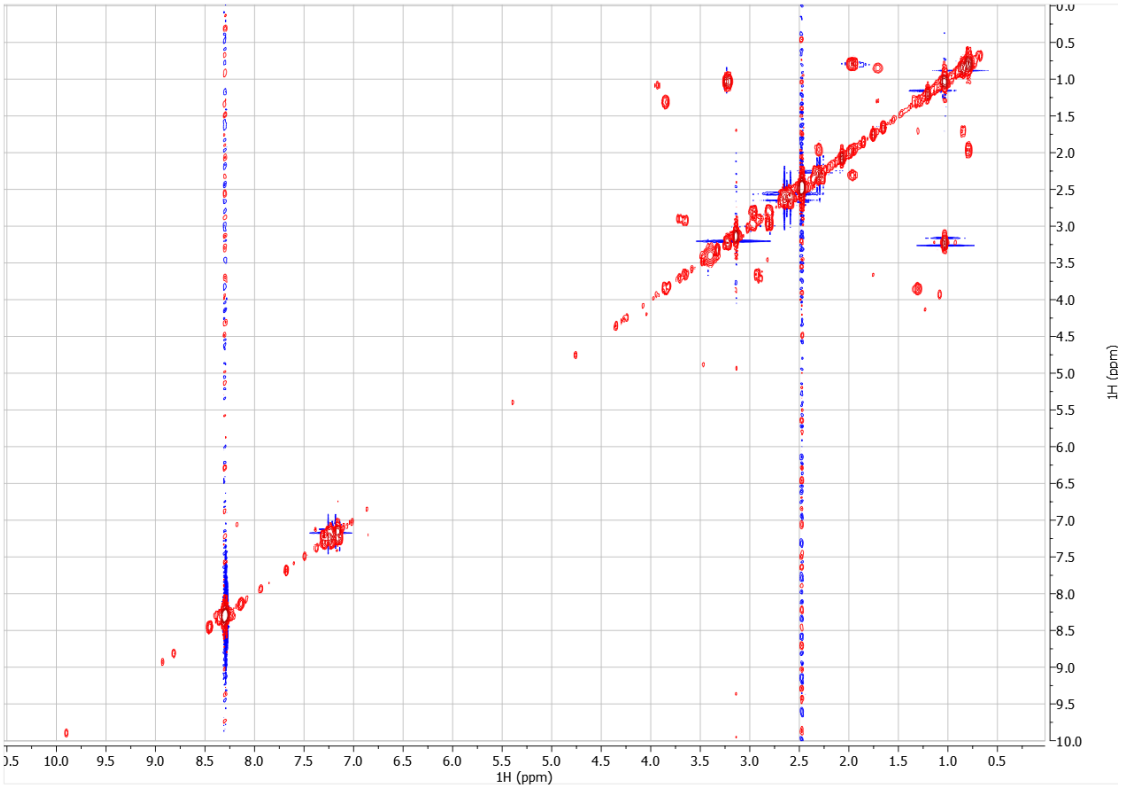


Figure S28. COSY NMR spectrum of natural compound 2 in DMSO-d_6

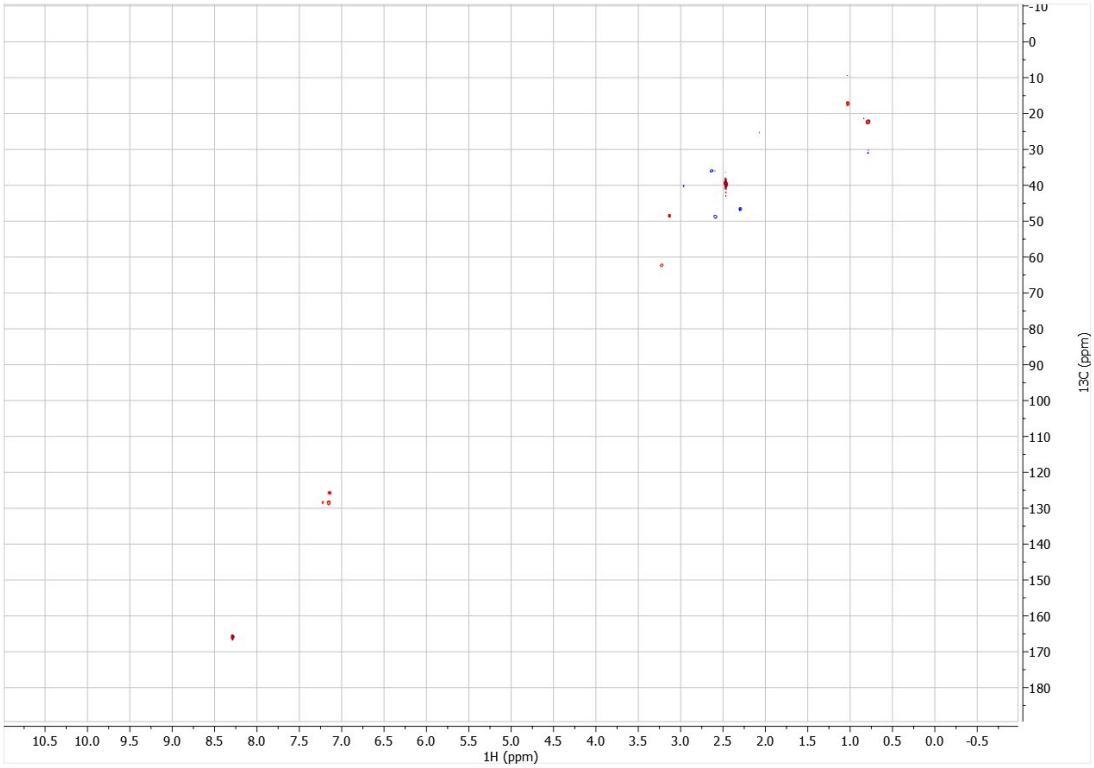


Figure S29. HSQC NMR spectrum of natural compound 2 in DMSO-d_6

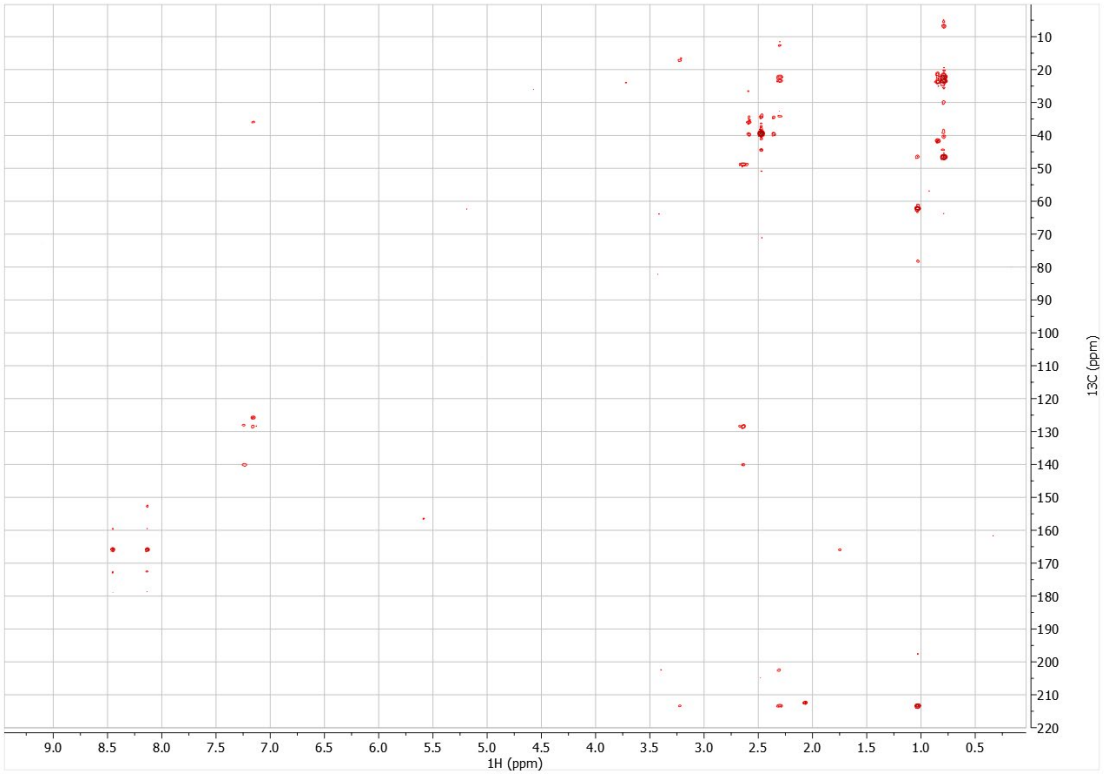


Figure S30. HMBC NMR spectrum of natural compound 2 in DMSO-d_6

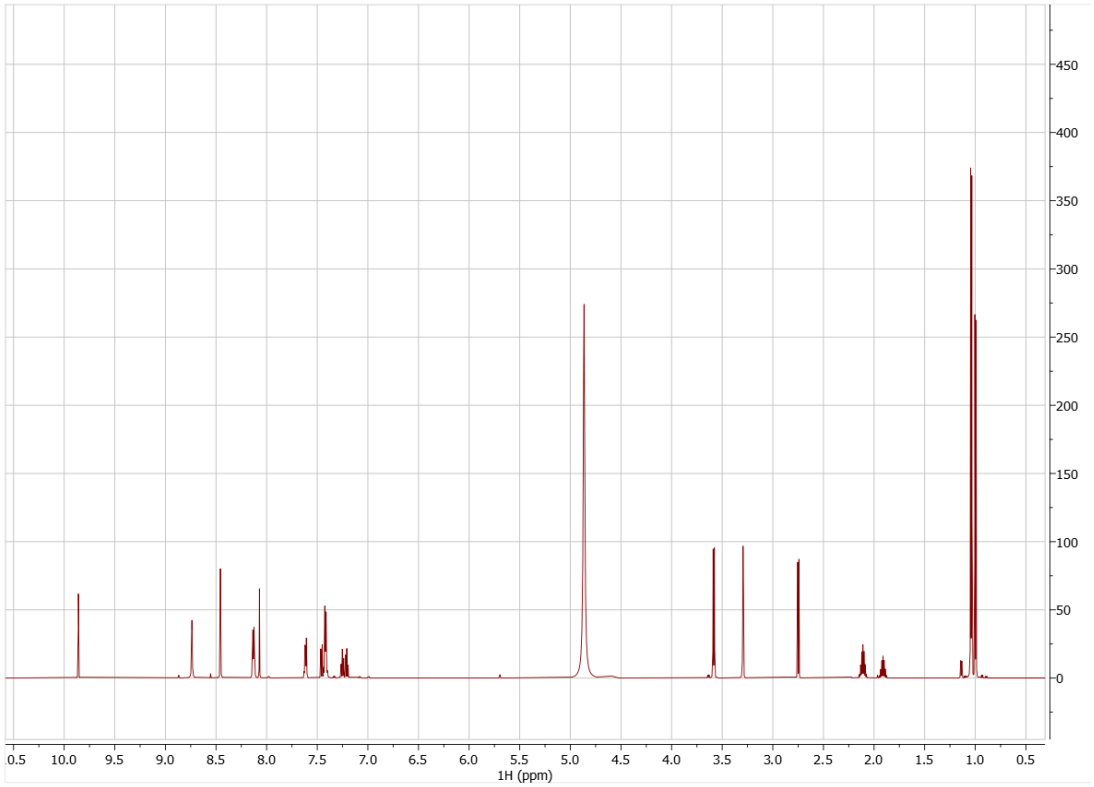


Figure S31. ^{1}H NMR spectrum of synthetic indolimine-200 (3) in methanol-d_4

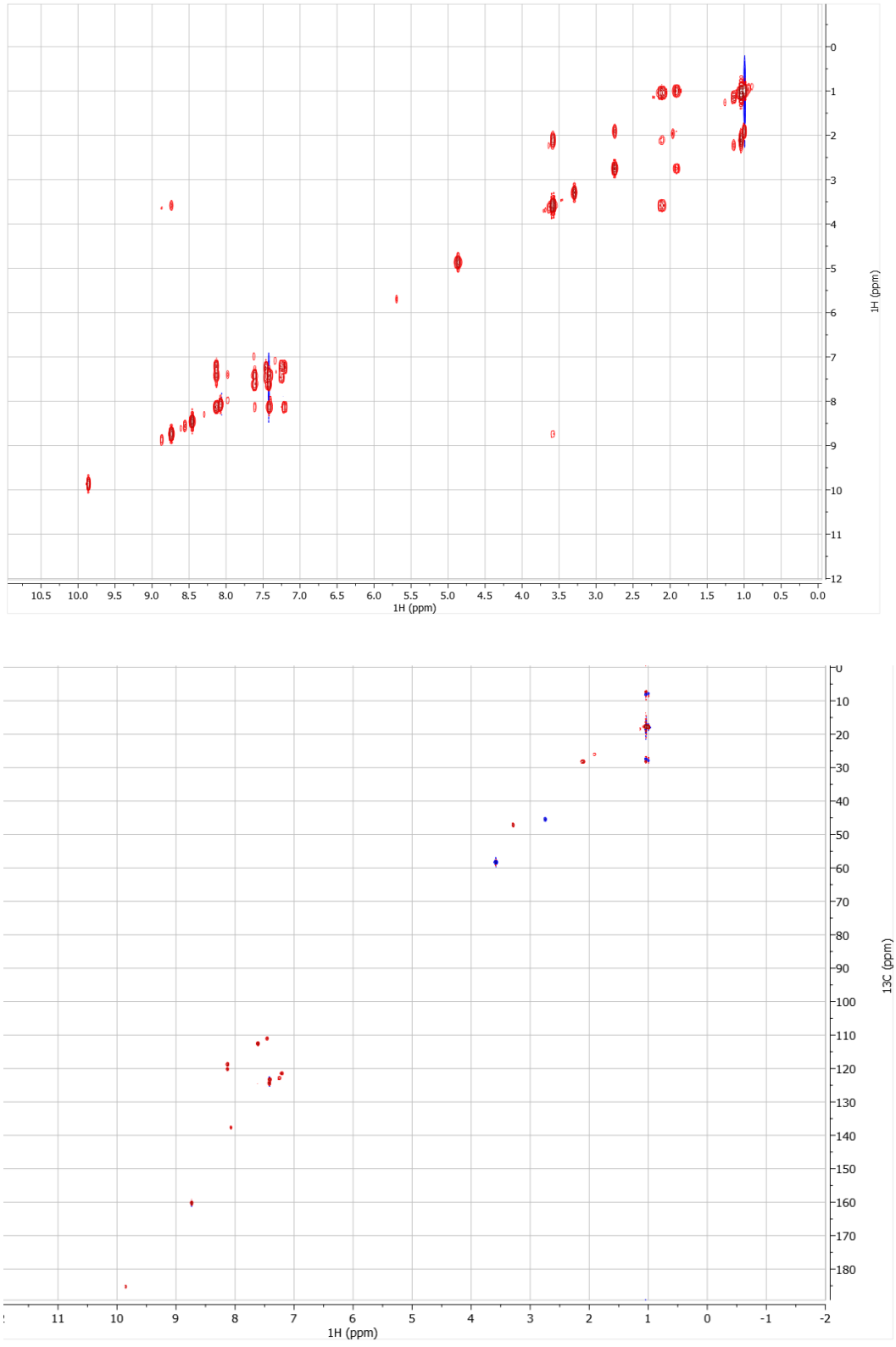


Figure S33. HSQC NMR spectrum of synthetic indolimine-200 (3) in methanol-d_4

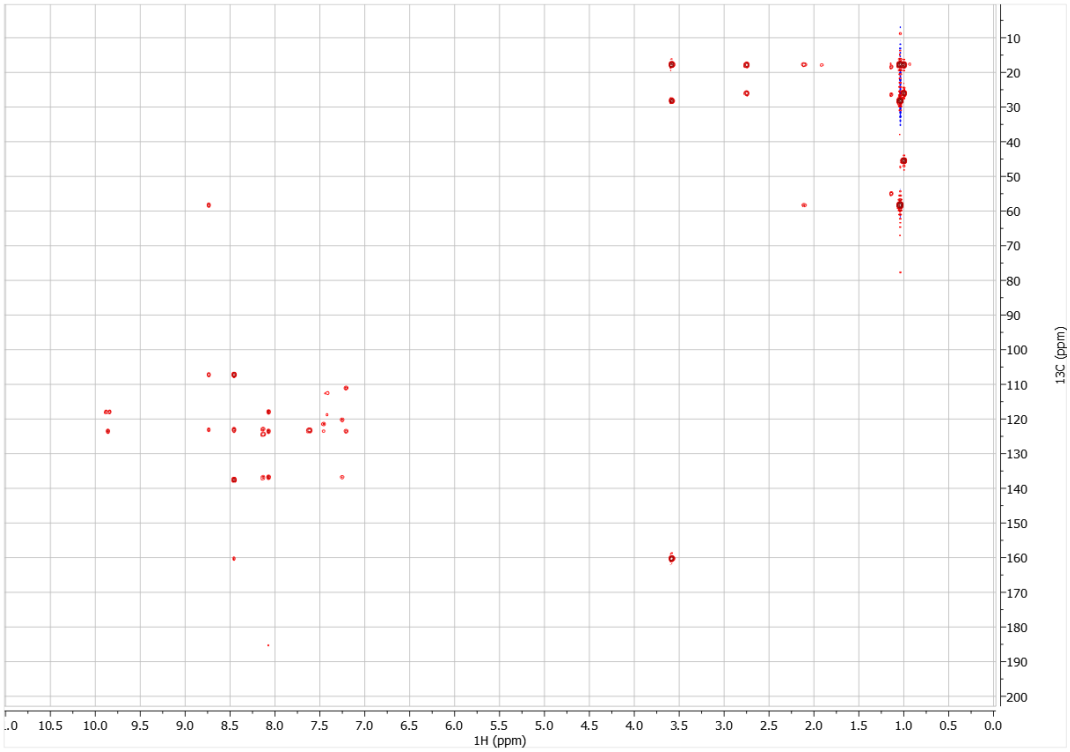


Figure S34. HMBC NMR spectrum of synthetic indolimine-200 (3) in methanol-d_4

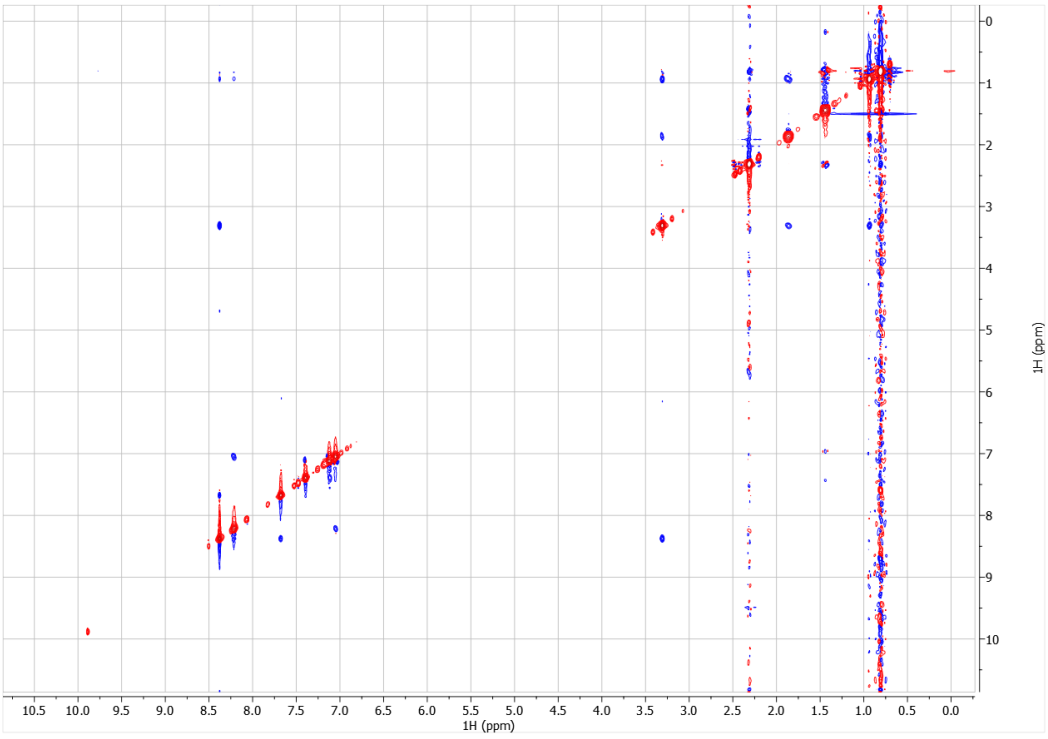


Figure S35. ROESY NMR spectrum of synthetic indolimine-200 (3) in methanol-d_4 (mixing time: 300 ms)

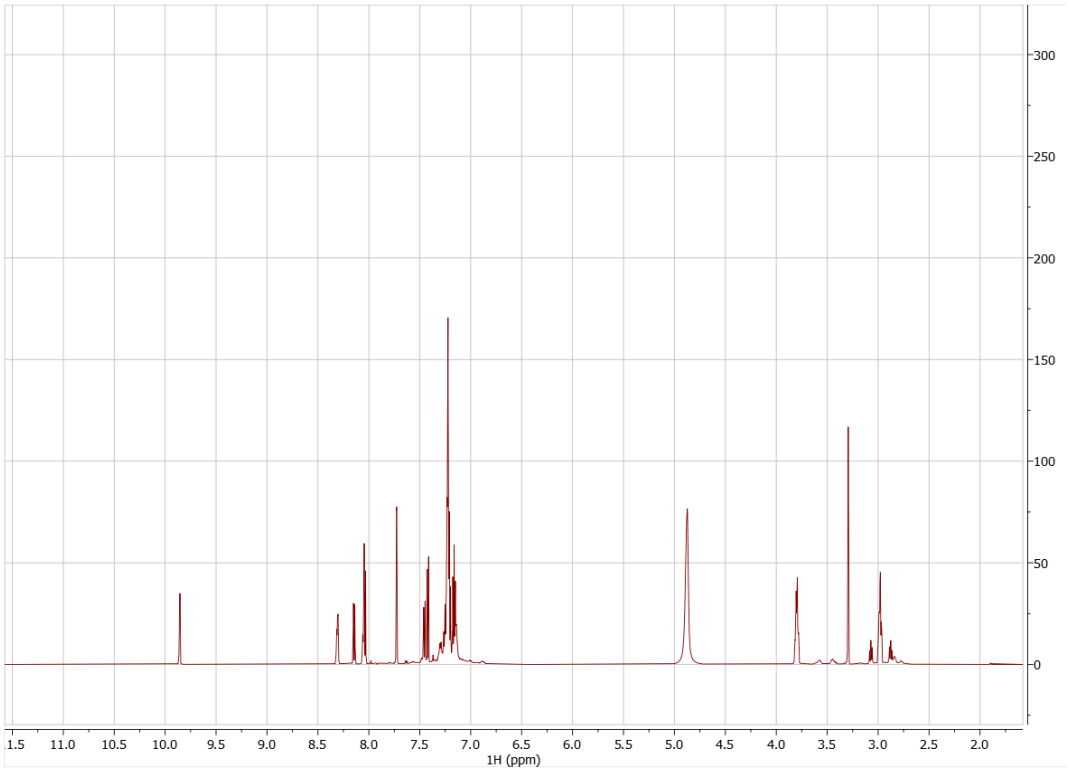


Figure S36. ^{1}H NMR spectrum of synthetic indolimine-248 (4) in methanol-d_4

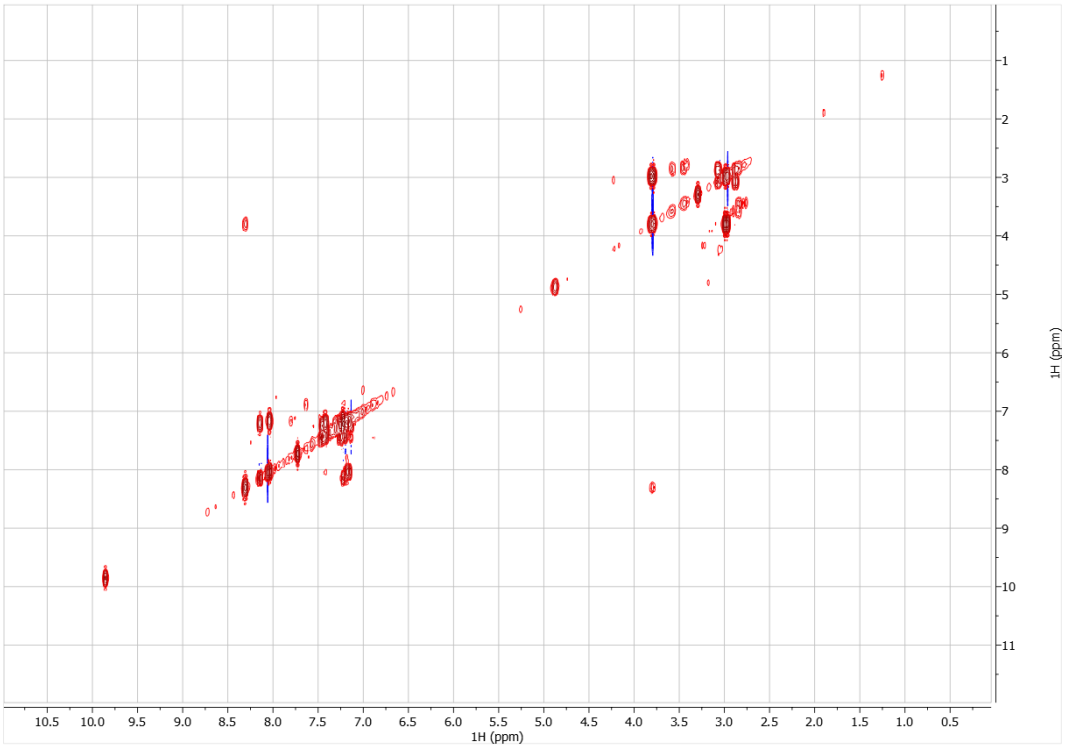


Figure S37. COSY spectrum of synthetic indolimine-248 (4) in methanol-d_4

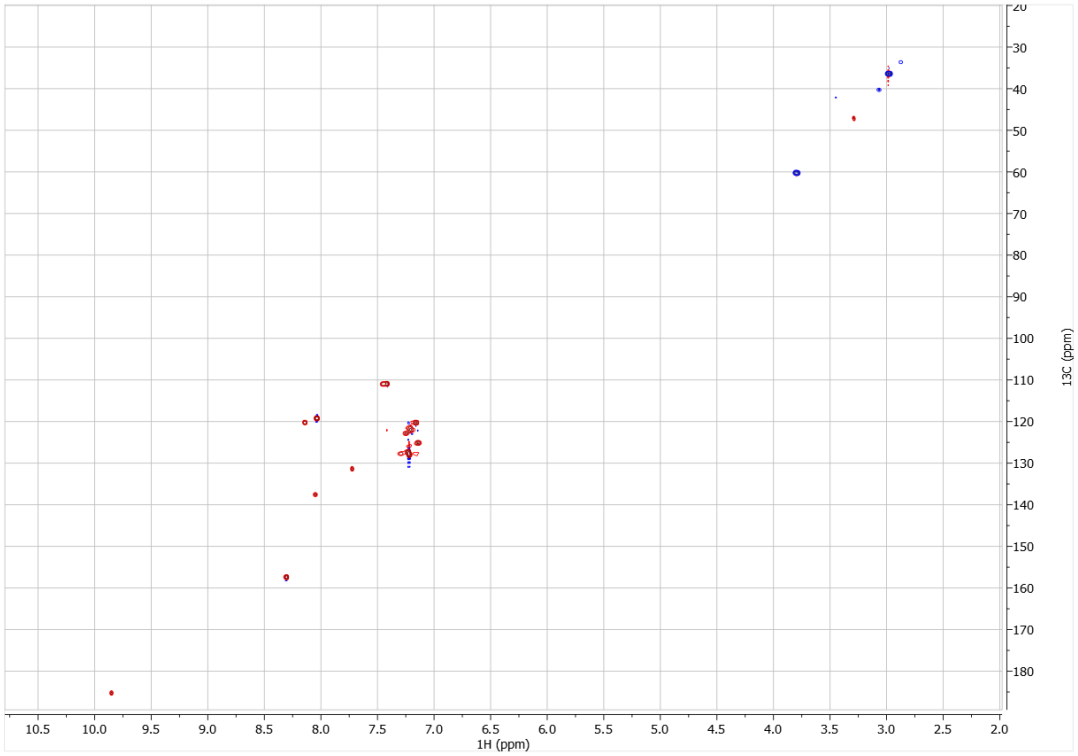


Figure S38. HSQC NMR spectrum of synthetic indolimine-248 (4) in methanol-d_4

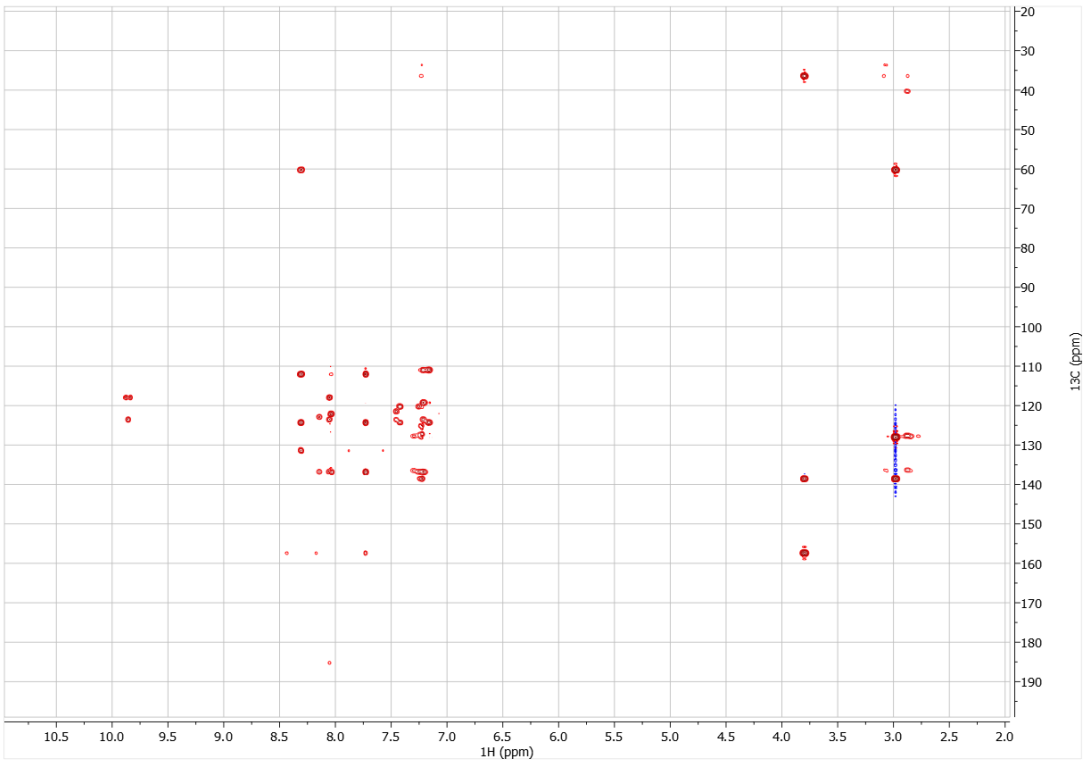


Figure S39. HMBC NMR spectrum of synthetic indolimine-248 (4) in methanol-d_4

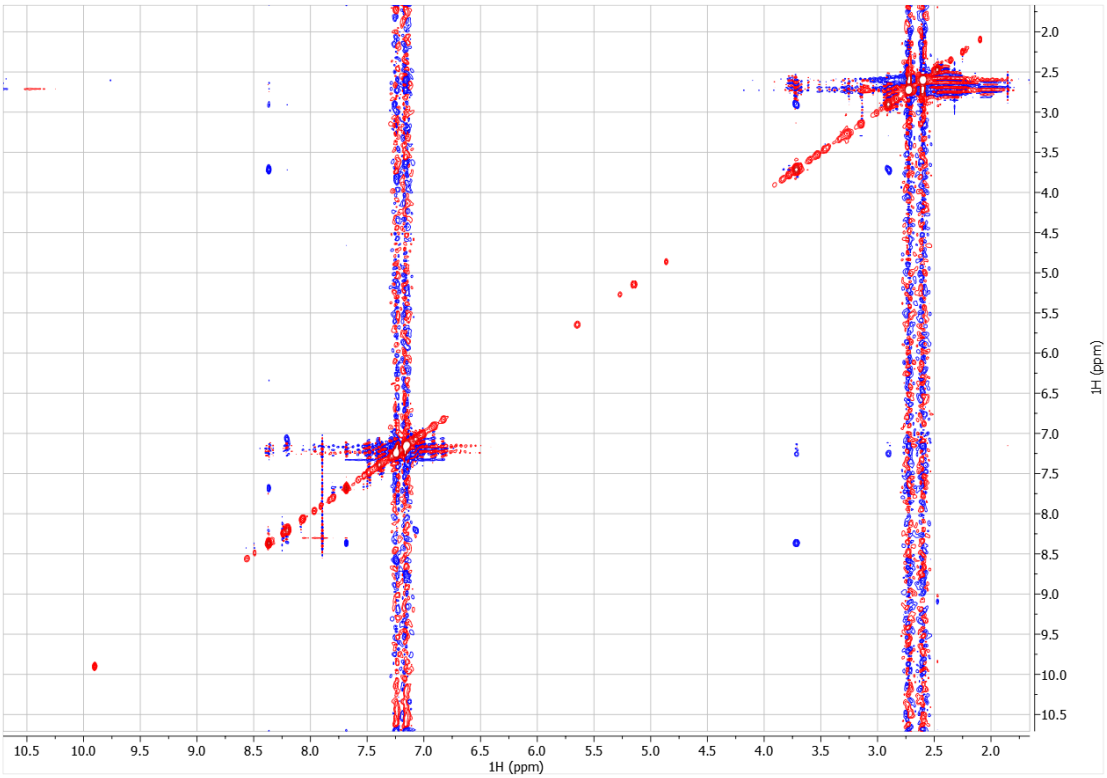


Figure S40. ROESY NMR spectrum of synthetic indolimine-248 (4) in methanol-d_4 (mixing time: 300 ms)

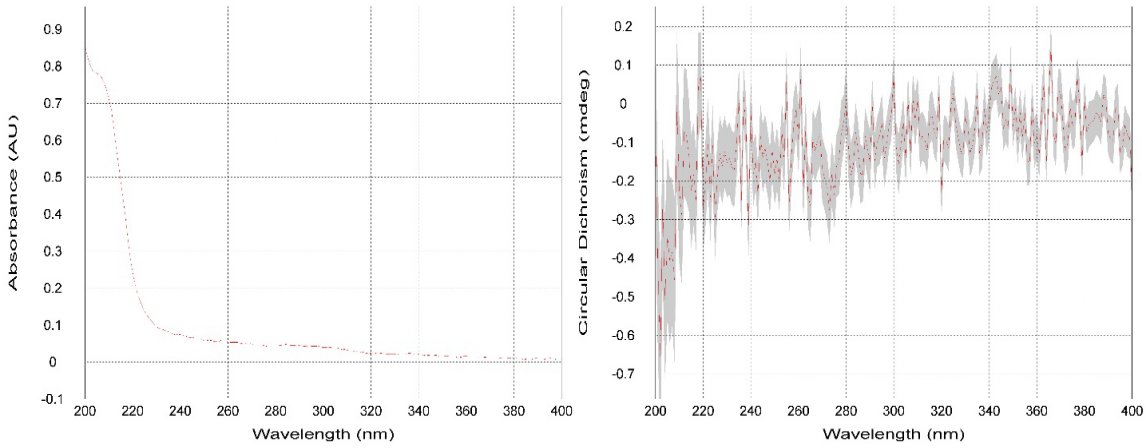


Figure S41. UV (left) and electronic circular dichroism (ECD, right) spectrum of natural compound 2 in methanol

References

1. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. **124**(4): p. 837-48.
2. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-14.
3. Turnbaugh, P.J., et al., *The human microbiome project.* Nature, 2007. **449**(7164): p. 804-10.
4. Donaldson, G.P., S.M. Lee, and S.K. Mazmanian, *Gut biogeography of the bacterial microbiota.* Nat Rev Microbiol, 2016. **14**(1): p. 20-32.
5. Integrative, H.M.P.R.N.C., *The Integrative Human Microbiome Project.* Nature, 2019. **569**(7758): p. 641-648.
6. Lloyd-Price, J., et al., *Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases.* Nature, 2019. **569**(7758): p. 655-662.
7. Huttenhower, C., A.D. Kostic, and R.J. Xavier, *Inflammatory bowel disease as a model for translating the microbiome.* Immunity, 2014. **40**(6): p. 843-54.
8. Brennan, C.A. and W.S. Garrett, *Gut Microbiota, Inflammation, and Colorectal Cancer.* Annu Rev Microbiol, 2016. **70**: p. 395-411.
9. Kamada, N., et al., *Role of the gut microbiota in immunity and inflammatory disease.* Nat Rev Immunol, 2013. **13**(5): p. 321-35.
10. Ruff, W.E., T.M. Greiling, and M.A. Kriegel, *Host-microbiota interactions in immune-mediated diseases.* Nat Rev Microbiol, 2020. **18**(9): p. 521-538.
11. Iwasaki, A. and R. Medzhitov, *Regulation of Adaptive Immunity by the Innate Immune System.* Science, 2010. **327**(5963): p. 291-295.
12. Park, B.S. and J.O. Lee, *Recognition of lipopolysaccharide pattern by TLR4 complexes.* Exp Mol Med, 2013. **45**: p. e66.
13. Mazmanian, S.K., J.L. Round, and D.L. Kasper, *A microbial symbiosis factor prevents intestinal inflammatory disease.* Nature, 2008. **453**(7195): p. 620-5.
14. Round, J.L. and S.K. Mazmanian, *Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota.* Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12204-9.
15. Round, J.L. and N.W. Palm, *Causal effects of the microbiota on immune-mediated diseases.* Sci Immunol, 2018. **3**(20).
16. Donaldson, G.P., et al., *Gut microbiota utilize immunoglobulin A for mucosal colonization.* Science, 2018. **360**(6390): p. 795-800.
17. Ost, K.S., et al., *Adaptive immunity induces mutualism between commensal eukaryotes.* Nature, 2021.
18. Doron, I., et al., *Mycobiota-induced IgA antibodies regulate fungal commensalism in the gut and are dysregulated in Crohn's disease.* Nat Microbiol, 2021. **6**(12): p. 1493-1504.
19. Palm, N.W., et al., *Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease.* Cell, 2014. **158**(5): p. 1000-10.
20. Rojas, O.L., et al., *Recirculating Intestinal IgA-Producing Cells Regulate Neuroinflammation via IL-10.* Cell, 2018.
21. Probstel, A.K., et al., *Gut microbiota-specific IgA(+) B cells traffic to the CNS in active multiple sclerosis.* Sci Immunol, 2020. **5**(53).
22. Honda, K. and D.R. Littman, *The microbiota in adaptive immune homeostasis and disease.* Nature, 2016. **535**(7610): p. 75-84.
23. Ivanov, II, et al., *Induction of intestinal Th17 cells by segmented filamentous bacteria.* Cell, 2009. **139**(3): p. 485-98.
24. Atarashi, K., et al., *Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial Cells.* Cell, 2015. **163**(2): p. 367-80.
25. Goto, Y., et al., *Segmented filamentous bacteria antigens presented by intestinal dendritic cells drive mucosal Th17 cell differentiation.* Immunity, 2014. **40**(4): p. 594-607.
26. Yang, Y., et al., *Focused specificity of intestinal TH17 cells towards commensal bacterial antigens.* Nature, 2014. **510**(7503): p. 152-6.

27. McCarville, J.L., et al., *Microbiota Metabolites in Health and Disease.* Annu Rev Immunol, 2020. **38**: p. 147-170.
28. Koh, A., et al., *From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites.* Cell, 2016. **165**(6): p. 1332-1345.
29. Smith, P.M., et al., *The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis.* Science, 2013. **341**(6145): p. 569-73.
30. Macia, L., et al., *Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome.* Nat Commun, 2015. **6**: p. 6734.
31. Trompette, A., et al., *Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis.* Nat Med, 2014. **20**(2): p. 159-66.
32. Schulthess, J., et al., *The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages.* Immunity, 2019. **50**(2): p. 432-445 e7.
33. Wenzel, T.J., et al., *Short-chain fatty acids (SCFAs) alone or in combination regulate select immune functions of microglia-like cells.* Mol Cell Neurosci, 2020. **105**: p. 103493.
34. Chen, M.L., K. Takeda, and M.S. Sundrud, *Emerging roles of bile acids in mucosal immunity and inflammation.* Mucosal Immunol, 2019. **12**(4): p. 851-861.
35. Buffie, C.G., et al., *Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile.* Nature, 2015. **517**(7533): p. 205-8.
36. Hang, S., et al., *Bile acid metabolites control TH17 and Treg cell differentiation.* Nature, 2019. **576**(7785): p. 143-148.
37. Agus, A., J. Planchais, and H. Sokol, *Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease.* Cell Host Microbe, 2018. **23**(6): p. 716-724.
38. Zelante, T., et al., *Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22.* Immunity, 2013. **39**(2): p. 372-85.
39. Schiering, C., et al., *Feedback control of AHR signalling regulates intestinal immunity.* Nature, 2017. **542**(7640): p. 242-245.
40. Bommarius, B., et al., *A family of indoles regulate virulence and Shiga toxin production in pathogenic E. coli.* PLoS One, 2013. **8**(1): p. e54456.
41. Sadik, A., et al., *IL4I1 Is a Metabolic Immune Checkpoint that Activates the AHR and Promotes Tumor Progression.* Cell, 2020. **182**(5): p. 1252-1270 e34.
42. Metidji, A., et al., *The Environmental Sensor AHR Protects from Inflammatory Damage by Maintaining Intestinal Stem Cell Homeostasis and Barrier Integrity.* Immunity, 2018. **49**(2): p. 353-362 e5.
43. Keum, N. and E. Giovannucci, *Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies.* Nat Rev Gastroenterol Hepatol, 2019. **16**(12): p. 713-732.
44. Jasperson, K.W., et al., *Hereditary and familial colon cancer.* Gastroenterology, 2010. **138**(6): p. 2044-58.
45. Pitot, H.C., *The molecular biology of carcinogenesis.* Cancer, 1993. **72**(3 Suppl): p. 962-70.
46. Grady, W.M. and J.M. Carethers, *Genomic and epigenetic instability in colorectal cancer pathogenesis.* Gastroenterology, 2008. **135**(4): p. 1079-99.
47. Conteduca, V., et al., *Precancerous colorectal lesions (Review).* Int J Oncol, 2013. **43**(4): p. 973-84.
48. Armaghany, T., et al., *Genetic alterations in colorectal cancer.* Gastrointest Cancer Res, 2012. **5**(1): p. 19-27.
49. Kedrin, D. and M.K. Gala, *Genetics of the serrated pathway to colorectal cancer.* Clin Transl Gastroenterol, 2015. **6**: p. e84.
50. Munro, M.J., et al., *Cancer stem cells in colorectal cancer: a review.* J Clin Pathol, 2018. **71**(2): p. 110-116.
51. Garrett, W.S., *Cancer and the microbiota.* Science, 2015. **348**(6230): p. 80-6.
52. Gopalakrishnan, V., et al., *Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients.* Science, 2017.

53. Matson, V., et al., *The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients.* Science, 2018. **359**(6371): p. 104-108.
54. Roy, S. and G. Trinchieri, *Microbiota: a key orchestrator of cancer therapy.* Nat Rev Cancer, 2017. **17**(5): p. 271-285.
55. Iida, N., et al., *Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment.* Science, 2013. **342**(6161): p. 967-70.
56. Viaud, S., et al., *The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide.* Science, 2013. **342**(6161): p. 971-6.
57. Daillere, R., et al., *Enterococcus hirae and Barnesiella intestinihominis Facilitate Cyclophosphamide-Induced Therapeutic Immunomodulatory Effects.* Immunity, 2016. **45**(4): p. 931-943.
58. Sivan, A., et al., *Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy.* Science, 2015. **350**(6264): p. 1084-9.
59. Thomas, A.M., et al., *Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation.* Nature Medicine, 2019. **25**(4): p. 667-+.
60. Wirbel, J., et al., *Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer.* Nature Medicine, 2019. **25**(4): p. 679-+.
61. Wong, S.H. and J. Yu, *Gut microbiota in colorectal cancer: mechanisms of action and clinical applications.* Nat Rev Gastroenterol Hepatol, 2019. **16**(11): p. 690-704.
62. Bullman, S., et al., *Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer.* Science, 2017. **358**(6369): p. 1443-1448.
63. Brennan, C.A., et al., *Fusobacterium nucleatum drives a pro-inflammatory intestinal microenvironment through metabolite receptor-dependent modulation of IL-17 expression.* Gut Microbes, 2021. **13**(1): p. 1987780.
64. Chung, L., et al., *Bacteroides fragilis Toxin Coordinates a Pro-carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells.* Cell Host Microbe, 2018. **23**(2): p. 203-214 e5.
65. Dejea, C.M., et al., *Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria.* Science, 2018. **359**(6375): p. 592-597.
66. Zeng, M.Y., N. Inohara, and G. Nunez, *Mechanisms of inflammation-driven bacterial dysbiosis in the gut.* Mucosal Immunol, 2017. **10**(1): p. 18-26.
67. Rubinstein, M.R., et al., *Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin.* Cell Host Microbe, 2013. **14**(2): p. 195-206.
68. Gur, C., et al., *Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack.* Immunity, 2015. **42**(2): p. 344-55.
69. Martin, O.C.B. and T. Frisan, *Bacterial Genotoxin-Induced DNA Damage and Modulation of the Host Immune Microenvironment.* Toxins (Basel), 2020. **12**(2).
70. He, Z., et al., *Campylobacter jejuni promotes colorectal tumorigenesis through the action of cytolethal distending toxin.* Gut, 2019. **68**(2): p. 289-300.
71. Nougayrede, J.P., et al., *Escherichia coli induces DNA double-strand breaks in eukaryotic cells.* Science, 2006. **313**(5788): p. 848-51.
72. Putze, J., et al., *Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae.* Infect Immun, 2009. **77**(11): p. 4696-703.
73. Cuevas-Ramos, G., et al., *Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells.* Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11537-42.
74. Arthur, J.C., et al., *Intestinal inflammation targets cancer-inducing activity of the microbiota.* Science, 2012. **338**(6103): p. 120-3.
75. Cougnoux, A., et al., *Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype.* Gut, 2014. **63**(12): p. 1932-42.
76. Wilson, M.R., et al., *The human gut bacterial genotoxin colibactin alkylates DNA.* Science, 2019. **363**(6428): p. 709-+.

77. Xue, M., et al., *Structure elucidation of colibactin and its DNA cross-links.* Science, 2019. **365**(6457).
78. Pleguezuelos-Manzano, C., et al., *Mutational signature in colorectal cancer caused by genotoxic pks(+) E. coli.* Nature, 2020.
79. Petersen, M.W., et al., *Irritable bowel, chronic widespread pain, chronic fatigue and related syndromes are prevalent and highly overlapping in the general population: DanFunD.* Sci Rep, 2020. **10**(1): p. 3273.
80. Butler, S., et al., *Chronic widespread pain-the need for a standard definition.* Pain, 2016. **157**(3): p. 541-543.
81. Minerbi, A., et al., *Altered microbiome composition in individuals with fibromyalgia.* Pain, 2019.
82. Enck, P., et al., *Irritable bowel syndrome.* Nat Rev Dis Primers, 2016. **2**: p. 16014.
83. Lovell, R.M. and A.C. Ford, *Global prevalence of and risk factors for irritable bowel syndrome: a meta-analysis.* Clin Gastroenterol Hepatol, 2012. **10**(7): p. 712-721 e4.
84. Bischoff, S.C., et al., *Intestinal permeability--a new target for disease prevention and therapy.* BMC Gastroenterol, 2014. **14**: p. 189.
85. Fritscher-Ravens, A., et al., *Confocal endomicroscopy shows food-associated changes in the intestinal mucosa of patients with irritable bowel syndrome.* Gastroenterology, 2014. **147**(5): p. 1012-20 e4.
86. Valentin, N., et al., *Biomarkers for bile acid diarrhoea in functional bowel disorder with diarrhoea: a systematic review and meta-analysis.* Gut, 2016. **65**(12): p. 1951-1959.
87. Barbara, G., C. Cremon, and V. Stanghellini, *Inflammatory bowel disease and irritable bowel syndrome: similarities and differences.* Curr Opin Gastroenterol, 2014. **30**(4): p. 352-8.
88. Barbara, G., et al., *Mast cell-dependent excitation of visceral-nociceptive sensory neurons in irritable bowel syndrome.* Gastroenterology, 2007. **132**(1): p. 26-37.
89. Buhner, S., et al., *Activation of human enteric neurons by supernatants of colonic biopsy specimens from patients with irritable bowel syndrome.* Gastroenterology, 2009. **137**(4): p. 1425-34.
90. Dothel, G., et al., *Nerve Fiber Outgrowth Is Increased in the Intestinal Mucosa of Patients With Irritable Bowel Syndrome.* Gastroenterology, 2015. **148**(5): p. 1002-U456.
91. Rajilic-Stojanovic, M., et al., *Intestinal microbiota and diet in IBS: causes, consequences, or epiphenomena?* Am J Gastroenterol, 2015. **110**(2): p. 278-87.
92. De Palma, G., et al., *The Adoptive Transfer of Anxiety and Gut Dysfunction From IBS Patients to Axenic Mice Through Microbiota Transplantation.* Gastroenterology, 2014. **146**(5): p. S845-S845.
93. Rajilic-Stojanovic, M., et al., *Global and Deep Molecular Analysis of Microbiota Signatures in Fecal Samples From Patients With Irritable Bowel Syndrome.* Gastroenterology, 2011. **141**(5): p. 1792-1801.
94. Valdez-Morales, E.E., et al., *Sensitization of peripheral sensory nerves by mediators from colonic biopsies of diarrhea-predominant irritable bowel syndrome patients: a role for PAR2.* Am J Gastroenterol, 2013. **108**(10): p. 1634-43.
95. Bellono, N.W., et al., *Enterochromaffin Cells Are Gut Chemosensors that Couple to Sensory Neural Pathways.* Cell, 2017. **170**(1): p. 185-198 e16.
96. Yano, J.M., et al., *Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis.* Cell, 2015. **161**(2): p. 264-76.
97. Lai, N.Y., et al., *Gut-Innervating Nociceptor Neurons Regulate Peyer's Patch Microfold Cells and SFB Levels to Mediate Salmonella Host Defense.* Cell, 2020. **180**(1): p. 33-49 e22.
98. Roza, C. and P.W. Reeh, *Substance P, calcitonin gene related peptide and PGE2 co-released from the mouse colon: a new model to study nociceptive and inflammatory responses in viscera, in vitro.* Pain, 2001. **93**(3): p. 213-9.
99. Julius, D., *TRP channels and pain.* Annu Rev Cell Dev Biol, 2013. **29**: p. 355-84.
100. von Hehn, C.A., R. Baron, and C.J. Woolf, *Deconstructing the neuropathic pain phenotype to reveal neural mechanisms.* Neuron, 2012. **73**(4): p. 638-52.

101. Chiu, I.M., et al., *Transcriptional profiling at whole population and single cell levels reveals somatosensory neuron molecular diversity.* Elife, 2014. **3**.
102. Venkatachalam, K. and C. Montell, *TRP channels.* Annu Rev Biochem, 2007. **76**: p. 387-417.
103. Minke, B. and M. Parnas, *Insights on TRP channels from in vivo studies in Drosophila.* Annu Rev Physiol, 2006. **68**: p. 649-84.
104. Caterina, M.J., et al., *The capsaicin receptor: a heat-activated ion channel in the pain pathway.* Nature, 1997. **389**(6653): p. 816-24.
105. Jordt, S.E. and D. Julius, *Molecular basis for species-specific sensitivity to "hot" chili peppers.* Cell, 2002. **108**(3): p. 421-30.
106. Leffler, A., B. Monter, and M. Koltzenburg, *The role of the capsaicin receptor TRPV1 and acid-sensing ion channels (ASICS) in proton sensitivity of subpopulations of primary nociceptive neurons in rats and mice.* Neuroscience, 2006. **139**(2): p. 699-709.
107. Siemens, J., et al., *Spider toxins activate the capsaicin receptor to produce inflammatory pain.* Nature, 2006. **444**(7116): p. 208-12.
108. Pinho-Ribeiro, F.A., W.A. Verri, and I.M. Chiu, *Nociceptor Sensory Neuron-Immune Interactions in Pain and Inflammation.* Trends in Immunology, 2017. **38**(1): p. 5-19.
109. Baral, P., S. Udit, and I.M. Chiu, *Pain and immunity: implications for host defence.* Nat Rev Immunol, 2019.
110. Cook, A.D., et al., *Immune Cytokines and Their Receptors in Inflammatory Pain.* Trends in Immunology, 2018. **39**(3): p. 240-255.
111. Krukowski, K., et al., *CD8+ T Cells and Endogenous IL-10 Are Required for Resolution of Chemotherapy-Induced Neuropathic Pain.* J Neurosci, 2016. **36**(43): p. 11074-11083.
112. Talbot, S., S.L. Foster, and C.J. Woolf, *Neuroimmunity: Physiology and Pathology.* Annu Rev Immunol, 2016. **34**: p. 421-47.
113. Chiu, I.M., et al., *Bacteria activate sensory neurons that modulate pain and inflammation.* Nature, 2013. **501**(7465): p. 52-7.
114. Blake, K.J., et al., *Staphylococcus aureus produces pain through pore-forming toxins and neuronal TRPV1 that is silenced by QX-314.* Nature Communications, 2018. **9**.
115. Pinho-Ribeiro, F.A., et al., *Blocking Neuronal Signaling to Immune Cells Treats Streptococcal Invasive Infection.* Cell, 2018. **173**(5): p. 1083-+.
116. Kashem, S.W., et al., *Nociceptive Sensory Fibers Drive Interleukin-23 Production from CD301b+ Dermal Dendritic Cells and Drive Protective Cutaneous Immunity.* Immunity, 2015. **43**(3): p. 515-26.
117. Nagashima, H., et al., *Neuropeptide CGRP Limits Group 2 Innate Lymphoid Cell Responses and Constrains Type 2 Inflammation.* Immunity, 2019. **51**(4): p. 682-+.
118. Wallrapp, A., et al., *Calcitonin Gene-Related Peptide Negatively Regulates Alarmin-Driven Type 2 Innate Lymphoid Cell Responses.* Immunity, 2019.
119. Xu, H., et al., *Transcriptional Atlas of Intestinal Immune Cells Reveals that Neuropeptide alpha-CGRP Modulates Group 2 Innate Lymphoid Cell Responses.* Immunity, 2019. **51**(4): p. 696-708 e9.
120. Talbot, J., et al., *Feeding-dependent VIP neuron-ILC3 circuit regulates the intestinal barrier.* Nature, 2020.
121. Cardoso, V., et al., *Neuronal regulation of type 2 innate lymphoid cells via neuromedin U.* Nature, 2017. **549**(7671): p. 277-281.
122. Klose, C.S.N., et al., *The neuropeptide neuromedin U stimulates innate lymphoid cells and type 2 inflammation.* Nature, 2017. **549**(7671): p. 282-286.
123. Perner, C., et al., *Substance P Release by Sensory Neurons Triggers Dendritic Cell Migration and Initiates the Type-2 Immune Response to Allergens.* Immunity, 2020.
124. Tilg, H., et al., *The Intestinal Microbiota in Colorectal Cancer.* Cancer Cell, 2018. **33**(6): p. 954-964.
125. Allen, J. and C.L. Sears, *Impact of the gut microbiome on the genome and epigenome of colon epithelial cells: contributions to colorectal cancer development.* Genome Med, 2019. **11**(1): p. 11.

126. Goodwin, A.C., et al., *Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis.* Proc Natl Acad Sci U S A, 2011. **108**(37): p. 15354-9.
127. Buc, E., et al., *High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer.* PLoS One, 2013. **8**(2): p. e56964.
128. Jiang, Y.D., et al., *Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity.* Journal of the American Chemical Society, 2019. **141**(29): p. 11489-11496.
129. Dziubanska-Kusibab, P.J., et al., *Colibactin DNA-damage signature indicates mutational impact in colorectal cancer.* Nat Med, 2020. **26**(7): p. 1063-1069.
130. da Silva, R.R., P.C. Dorrestein, and R.A. Quinn, *Illuminating the dark matter in metabolomics.* Proc Natl Acad Sci U S A, 2015. **112**(41): p. 12549-50.
131. Adami, H.O., et al., *The continuing uncertainty about cancer risk in inflammatory bowel disease.* Gut, 2016. **65**(6): p. 889-893.
132. Xue, M., et al., *Characterization of Natural Colibactin-Nucleobase Adducts by Tandem Mass Spectrometry and Isotopic Labeling. Support for DNA Alkylation by Cyclopropane Ring Opening.* Biochemistry, 2018. **57**(45): p. 6391-6394.
133. Bossuet-Greif, N., et al., *The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells.* mBio, 2018. **9**(2).
134. Gates, K.S., *An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals.* Chemical Research in Toxicology, 2009. **22**(11): p. 1747-1760.
135. Unterhauser, K., et al., *Klebsiella oxytoca enterotoxins tilimycin and tilivalline have distinct host DNA-damaging and microtubule-stabilizing activities.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(9): p. 3774-3783.
136. Yuan, J., R. Adamski, and J. Chen, *Focus on histone variant H2AX: to be or not to be.* FEBS Lett, 2010. **584**(17): p. 3717-24.
137. Freedman, J.C., A. Shrestha, and B.A. McClane, *Clostridium perfringens Enterotoxin: Action, Genetics, and Translational Applications.* Toxins (Basel), 2016. **8**(3).
138. Shine, E.E. and J.M. Crawford, *Molecules from the Microbiome.* Annu Rev Biochem, 2021. **90**: p. 789-815.
139. Schneditz, G., et al., *Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(36): p. 13181-13186.
140. Blin, K., et al., *antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline.* Nucleic Acids Res, 2019. **47**(W1): p. W81-W87.
141. Trautman, E.P., et al., *Domain-Targeted Metabolomics Delineates the Heterocycle Assembly Steps of Colibactin Biosynthesis.* J Am Chem Soc, 2017. **139**(11): p. 4195-4201.
142. Zha, L., et al., *Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring.* Nature Chemical Biology, 2017. **13**(10): p. 1063-+.
143. Wami, H., et al., *Insights into evolution and coexistence of the colibactin- and yersiniabactin secondary metabolite determinants in enterobacterial populations.* Microb Genom, 2021. **7**(6).
144. Healy, A.R., et al., *Synthesis and reactivity of precolibactin 886.* Nat Chem, 2019. **11**(10): p. 890-898.
145. Shine, E.E., et al., *Model Colibactins Exhibit Human Cell Genotoxicity in the Absence of Host Bacteria.* ACS Chem Biol, 2018. **13**(12): p. 3286-3293.
146. Chen, W., et al., *Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer.* PLoS One, 2012. **7**(6): p. e39743.
147. Tang-Fichaux, M., et al., *The Polyphosphate Kinase of Escherichia coli Is Required for Full Production of the Genotoxin Colibactin.* mSphere, 2020. **5**(6).
148. Sakofsky, C.J., et al., *Break-induced replication is a source of mutation clusters underlying kataegis.* Cell Rep, 2014. **7**(5): p. 1640-1648.

149. Lord, C.J. and A. Ashworth, *The DNA damage response and cancer therapy.* Nature, 2012. **481**(7381): p. 287-94.
150. Hoxhaj, G. and B.D. Manning, *The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism.* Nat Rev Cancer, 2020. **20**(2): p. 74-88.
151. Kim, D.I., et al., *Microbial production of multiple short-chain primary amines via retrobiosynthesis.* Nat Commun, 2021. **12**(1): p. 173.
152. Brettin, T., et al., *RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes.* Sci Rep, 2015. **5**: p. 8365.
153. Baym, M., et al., *Rapid construction of a whole-genome transposon insertion collection for Shewanella oneidensis by Knockout Sudoku.* Nat Commun, 2016. **7**: p. 13270.
154. Anzai, I.A., et al., *Rapid curation of gene disruption collections using Knockout Sudoku.* Nat Protoc, 2017. **12**(10): p. 2110-2137.
155. Dougherty, M.W. and C. Jobin, *Shining a Light on Colibactin Biology.* Toxins (Basel), 2021. **13**(5).
156. Martincorena, I. and P.J. Campbell, *Somatic mutation in cancer and normal cells.* Science, 2015. **349**(6255): p. 1483-9.
157. Tronnet, S., et al., *The Genotoxin Colibactin Shapes Gut Microbiota in Mice.* mSphere, 2020. **5**(4).
158. Silpe, J.E., et al., 2021.
159. Lee-Six, H., et al., *The landscape of somatic mutation in normal colorectal epithelial cells.* Nature, 2019. **574**(7779): p. 532-537.
160. Nejman, D., et al., *The human tumor microbiome is composed of tumor type-specific intracellular bacteria.* Science, 2020. **368**(6494): p. 973-+.
161. Vizcaino, M.I. and J.M. Crawford, *The colibactin warhead crosslinks DNA.* Nat Chem, 2015. **7**(5): p. 411-7.
162. Dodd, D., et al., *A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites.* Nature, 2017. **551**(7682): p. 648-+.
163. Mi, H., et al., *Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0).* Nat Protoc, 2019. **14**(3): p. 703-721.
164. Nurk, S., et al., *Assembling single-cell genomes and mini-metagenomes from chimeric MDA products.* J Comput Biol, 2013. **20**(10): p. 714-37.
165. Lariviere, P.J., et al., *An Essential Regulator of Bacterial Division Links FtsZ to Cell Wall Synthase Activation.* Curr Biol, 2019. **29**(9): p. 1460-1470 e4.
166. Mawe, G.M. and J.M. Hoffman, *Serotonin signalling in the gut--functions, dysfunctions and therapeutic targets.* Nat Rev Gastroenterol Hepatol, 2013. **10**(8): p. 473-86.
167. Akbar, A., et al., *Expression of the TRPV1 receptor differs in quiescent inflammatory bowel disease with or without abdominal pain.* Gut, 2010. **59**(6): p. 767-74.
168. Akbar, A., et al., *Increased capsaicin receptor TRPV1-expressing sensory fibres in irritable bowel syndrome and their correlation with abdominal pain.* Gut, 2008. **57**(7): p. 923-9.
169. Lai, N.Y., et al., *Gut-Innervating Nociceptor Neurons Regulate Peyer's Patch Microfold Cells and SFB Levels to Mediate Salmonella Host Defense.* Cell, 2019.
170. Cohen, J.A., et al., *Cutaneous TRPV1(+) Neurons Trigger Protective Innate Type 17 Anticipatory Immunity.* Cell, 2019. **178**(4): p. 919-932 e14.
171. Galvao, I., et al., *The Metabolic Sensor GPR43 Receptor Plays a Role in the Control of Klebsiella pneumoniae Infection in the Lung.* Front Immunol, 2018. **9**: p. 142.
172. Chen, T.W., et al., *Ultrasensitive fluorescent proteins for imaging neuronal activity.* Nature, 2013. **499**(7458): p. 295-300.
173. Ciardo, M.G. and A. Ferrer-Montiel, *Lipids as central modulators of sensory TRP channels.* Biochim Biophys Acta Biomembr, 2017. **1859**(9 Pt B): p. 1615-1628.
174. Chen, Y., et al., *Epithelia-Sensory Neuron Cross Talk Underlies Cholestatic Itch Induced by Lysophosphatidylcholine.* Gastroenterology, 2021. **161**(1): p. 301-317 e16.
175. Zimmermann, M., et al., *Mapping human microbiome drug metabolism by gut bacteria and their genes.* Nature, 2019.

176. Jung, H.J., et al., *Genome-Wide Screening for Enteric Colonization Factors in Carbapenem-Resistant ST258 Klebsiella pneumoniae.* MBio, 2019. **10**(2).
177. Osteen, J.D., et al., *Selective spider toxins reveal a role for the Nav1.1 channel in mechanical pain.* Nature, 2016. **534**(7608): p. 494-9.
178. Sadler, K.E., et al., *Transient receptor potential canonical 5 mediates inflammatory mechanical and spontaneous pain in mice.* Sci Transl Med, 2021. **13**(595).
179. Vannucchi, M.G. and S. Evangelista, *Experimental Models of Irritable Bowel Syndrome and the Role of the Enteric Neurotransmission.* J Clin Med, 2018. **7**(1).
180. Matheis, F., et al., *Adrenergic Signaling in Muscularis Macrophages Limits Infection-Induced Neuronal Loss.* Cell, 2020. **180**(1): p. 64-78 e16.
181. Christianson, J.A. and G.F. Gebhart, *Assessment of colon sensitivity by luminal distension in mice.* Nat Protoc, 2007. **2**(10): p. 2624-31.
182. Pope, J.L., et al., *Microbial Colonization Coordinates the Pathogenesis of a Klebsiella pneumoniae Infant Isolate.* Sci Rep, 2019. **9**(1): p. 3380.
183. Young, T.M., et al., *Animal Model To Study Klebsiella pneumoniae Gastrointestinal Colonization and Host-to-Host Transmission.* Infect Immun, 2020. **88**(11).
184. Chen, H., et al., *A Forward Chemical Genetic Screen Reveals Gut Microbiota Metabolites That Modulate Host Physiology.* Cell, 2019. **177**(5): p. 1217-1231 e18.
185. Gao, X.J., et al., *A transcriptional reporter of intracellular Ca(2+) in Drosophila.* Nat Neurosci, 2015. **18**(6): p. 917-25.
186. Kim, C.K., et al., *A Molecular Calcium Integrator Reveals a Striatal Cell Type Driving Aversion.* Cell, 2020. **183**(7): p. 2003-2019 e16.