Natural Product Discovery from the Human Microbiome*

Matthew R. Wilson¹, Li Zha, and Emily P. Balskus²

From the Department of Chemistry and Chemical Biology, Harvard University
Cambridge, MA 02138

Running title: Natural products from the human microbiome

To whom correspondence should be addressed: Prof. Emily P. Balskus, Department of Chemistry and Chemical Biology, Harvard University, Mallinckrodt Building 303N, 12 Oxford Street, Cambridge, Massachusetts 02318, Telephone: (617)-496-9921; FAX: (617)-496-5866; E-mail: balskus@chemistry.harvard.edu.

Keywords: bacteria, microbiome, biosynthesis, secondary metabolism, natural product, host-microbe interplay

Abstract

Human-associated microorganisms have the potential to biosynthesize numerous secondary metabolites that may mediate important host-microbe and microbe-microbe interactions. However, there is currently a limited understanding of microbiome-derived natural products. A variety of complementary discovery approaches have begun to illuminate this microbial “dark matter”, which will in turn allow detailed mechanistic studies of these molecules’ effects on microbiome and host. Herein, we review recent efforts to uncover microbiome-derived natural products and describe the key approaches that were used to identify and characterize these metabolites, discuss potential functional roles of these molecules, and discuss challenges related to this emerging research area.

Context for Secondary Metabolite Discovery

Secondary or specialized metabolites (natural products) represent a vast array of small molecules produced by many organisms, including microorganisms (1). Since the mid-20th century, activity-based screening of cultivatable microbes, mostly Actinomycetes soil bacteria, and subsequent isolation efforts enabled the discovery of numerous bioactive secondary metabolites including various polyketides (PKs)³, nonribosomal peptides (NRPs), and hybrids thereof (2). Advances in DNA sequencing, molecular biology, and cultivation methods have also enabled the more recent discovery of secondary metabolites from unculturable microorganisms (3-6). Both microbial natural products and semisynthetic derivatives have significantly impacted human health, serving as a major source of clinically-used antibiotics, antifungals, anticancer agents, immunosuppressants, and other drugs (7). While natural product discovery has predominantly focused on medicinal applications, recent studies have begun to shed light on the fascinating ecological roles of secondary metabolites in producing organisms and their surrounding environments, including the human body (8, 9).

Secondary Metabolites from the Human Microbiota

Trillions of microorganisms colonize the human body and significantly influence human health and disease. These organisms are collectively known as the human microbiome and possess 100-fold more genes than the human genome (10, 11). A recent analysis of genomic and metagenomic sequencing data from the Human Microbiome Project (HMP) showed that human-associated bacteria encode the biosynthetic machinery to synthesize a vast array of secondary metabolites (12). However, the identities of nearly all of these natural products are unknown, and their biological roles are not well understood. Uncovering the mechanisms by which these
molecules mediate host-microbe and microbe-microbe interactions could reveal factors influencing the microbiome’s effect on the host, a source of narrow-spectrum therapeutics, and inspire new strategies for treating human disease.

In this minireview, we describe several recent efforts to discover natural products from human-associated bacteria in order to highlight strategies for further exploring the biosynthetic capacity of the human microbiome. We also briefly discuss potential biological roles of these metabolites. Our goal is not to provide a comprehensive overview of all small molecules isolated from the human microbiome as this topic has been reviewed recently (13-18). Instead, we showcase how distinct approaches for natural product discovery can be used alone or in combination to unravel the complex metabolic interplay within the human microbiome.

**Sequence-Based Metagenome Mining: Lactocillin**

Advances in DNA sequencing technologies and the accompanying increase in the available (meta)genomic sequencing data have enabled computational strategies for natural product discovery (19, 20). Genome mining approaches take advantage of the fact that the genes responsible for constructing microbial secondary metabolites are typically co-localized in biosynthetic gene clusters (BGCs). Automated programs such as ClusterFinder, a hidden Markov model-based probabilistic algorithm, have been developed to rapidly and accurately identify BGCs from DNA sequencing data (21). Recently, Fischbach and co-workers used ClusterFinder to identify more than 14,000 BGCs from HMP reference genomes (12). These human-associated BGCs are predicted to synthesize a broad range of natural product classes, including saccharides, NRPs, PKs, and ribosomally encoded and posttranslationally modified peptides (RiPPs).

Based on the widespread distribution of RiPPs across multiple body sites and their potential clinical relevance, the Fischbach group searched HMP shotgun metagenomic assemblies for pathways encoding thiopetides, a well-studied group of RiPPs. Of the 13 identified thiopetide BGCs, bcg66 was selected for further study as it resembled the well-characterized BGC that produces the antibiotic thiocillin (22). Comparative metabolomics between a wild-type *Lactobacillus gasseri* strain harboring the bcg66 cluster and an isogenic mutant successfully guided the isolation of lactocillin (Figure 1). As *L. gasseri* is a vaginal isolate, this work represents the first structural characterization of a thiopetide natural product from a member of the human microbiome.

The structural similarity of this human-associated thiopetide to thiocillin suggested that lactocillin might exhibit antibacterial activity against Gram-positive organisms. As expected, purified lactocillin was active against Gram-positive pathogens, including residents of the urogenital tract such as *Staphylococcus aureus*, *Enterococcus faecalis*, and *Corynebacterium aurimucosum*. Interestingly, lactocillin did not show antibacterial activity against other vaginal *Lactobacillus* strains, raising the possibility that lactocillin production may have evolved to prevent the colonization of harmful organisms without effecting nearby beneficial bacteria. Additional colonization experiments with the wild-type *Lactobacillus gasseri* strain and the lactocillin isogenic mutant are needed to test this proposal.

Overall, the lactocillin case study illustrates how sequence-based discovery approaches can rapidly identify BGCs from both genomes and complex metagenomes. The ongoing application of these tools, along with continued advances in DNA sequencing and synthesis as well as improved approaches for heterologous expression of BGCs, will continue to provide access to the chemical output of the human microbiome (23).

**Functional Metagenome Mining: Commendamide**

As detailed in the previous section, sequenced-based approaches can provide structural information about natural products prior to isolation. However, a drawback to this approach is its inability to yield functional information about secondary metabolites that are not linked to cultivatable strains. Functional metagenomics is an alternative strategy that can address this issue. This approach involves extracting DNA from an environmental sample (eDNA) and heterologously expressing this genetic information in an easily cultivable host such as *Escherichia coli* (Figure 2,
The resulting clone libraries are then screened for a particular phenotype (e.g., antibacterial activity) or the presence of specific genes (e.g., biosynthetic genes) and the resulting clones of interest then analyzed for the production of clone-specific metabolites (25). Although functional metagenomic studies have successfully identified bioactive metabolites from soil eDNA (26, 27), this approach was not applied to the human microbiome until relatively recently.

In 2015, Brady and co-workers reported the identification of commendamide, a gut microbial metabolite involved in a host-microbe interaction (Figure 2, 28). They discovered this metabolite by screening *E. coli* clone libraries constructed from human stool metagenomic DNA. A cell-based reporter assay was used to identify clones that modulated nuclear factor-κB (NF-κB), an important transcription factor that responds to a variety of cellular processes. A total of 26 unique bacteria effector genes (Cbegs) were identified from examining a 75,000-member cosmid library. One effector gene family, Cbeg12, was annotated as an N-acetyl transferase. Heterologous expression of Cbeg12 and metabolite profiling revealed the production of N-acetyl-3-hydroxy palmitoyl glycine, a new natural product that was named commendamide. Cbeg12 was identified in several commensal *Bacteroides* genomes, and cultivation of *Bacteroides vulgatus* confirmed production of this metabolite.

Commendamide resembles various long-chain N-acetyl amino acids identified in soil microbes and humans (29). These mammalian and bacterial signaling molecules exhibit a broad range of activities including antibacterial (29), N-type calcium channel activation (30), and agonist activity against diverse G-protein coupled receptors (GPCR) (31). A broad screen for agonist activity against 242 GPCRs revealed that commendamide specifically activates GPCR132/G2A, a critical receptor that regulates cellular proliferation and immunity (32). Disruption or dysregulation of G2A activity is associated with autoimmune syndrome (33) and atherosclerosis (34). Thus, it is possible that the commensal *B. vulgatus* produces commendamide to interact with the host immune system via G2A modulation.

The discovery of commendamide showcases the ability of functional metagenomics to rapidly link a single open reading frame present in stool metagenomic DNA to a biologically active metabolite. In the near future, the development of additional heterologous hosts better suited for expression of DNA from prominent members of the human microbiome will undoubtedly increase the efficiency of this natural product discovery strategy (35).

**Ecological Approach: Lugdunin**

Studying the chemical ecology of members of the human microbiome is another powerful approach for identifying natural products that mediate interactions in this environment (8, 9). Instead of isolating a natural product and then searching for its biological function, an ecology-based approach starts with a relevant biological activity and then identifies the molecule(s) responsible. Several methods exist for connecting phenotypes to metabolites, including traditional bioassay-guided fractionation, targeted gene knockout studies, and metabolomics. Ecology-guided strategies have led to the discovery of many biologically active secondary metabolites from microbes living together with marine sponges, plants, fungi, invertebrate animals, and humans (8, 36).

Investigation of a potential ecological interaction within the human nasal microbiome led to the recent discovery of lugdunin, a secondary metabolite produced by the commensal bacterium *Staphylococcus lugdunensis* (37). The potentially beneficial properties of this organism were first illuminated when Zipperer and co-workers made the key observation that *S. lugdunensis* strongly antagonized the human pathogen *S. aureus* (Figure 3). Transposon mutagenesis traced this observed activity back to a 30-kb gene cluster encoding a nonribosomal peptide synthase (NRPS) and additional biosynthetic enzymes, suggesting that a NRP was likely responsible for the observed antibacterial activity. Overcoming initial cultivation hurdles, Zipperer and co-workers isolated lugdunin, a NRP antibiotic containing several additional features that could not be predicted using current bioinformatic analysis methods (Figure 3).

These isolation efforts, along with chemical synthesis, provided sufficient quantities of lugdunin to further explore its functional role in
the nasal cavity. Minimum inhibitory concentration assays demonstrated that lugdunin exhibits potent antimicrobial activity against Gram-positive bacteria, including vancomycin resistant Enterococcus spp. and S. aureus, with no detectable signs of resistance. This commensal bacterial metabolite also reduced and in some cases completely eradicated S. aureus in both skin and animal infection models. Moreover, epidemiological studies showed that S. aureus colonization rates are significantly lower in humans colonized with S. lugdunensis as compared to S. lugdunensis-negative patients. These studies suggest that lugdunin production by S. lugdunensis may interfere with S. aureus colonization in vivo.

Interspecies competition is a relatively common occurrence between human-associated microbes, particularly in nutrient-poor body sites such as the skin microbiota (38). Analyzing the composition of microbial communities from these sites using new statistical tools could reveal additional pairs of microbes that co-occur or are differentially present, revealing symbiotic, mutualistic, or antagonistic interactions to guide further natural product discovery (39).

**In Vitro Biochemistry: Colibactin**

The full or partial in vitro reconstitution of a cryptic natural product biosynthetic pathway is an unconventional strategy for gaining structural information about a secondary metabolite that can be uniquely enabling in certain situations. This approach involves characterizing the products that arise from combining purified biosynthetic enzymes with potential co-factors and substrates (40, 41). Unlike heterologous expression of a biosynthetic gene cluster, an in vitro approach can facilitate an understanding of the detailed chemical mechanisms involved in synthesizing a natural product (42). This strategy also bypasses potential problems with microbial cultivation and regulating gene cluster expression in vivo. Finally, partial in vitro reconstitution yields structural information that can assist with isolation efforts as exemplified by the ongoing characterization of colibactin, a secondary metabolite made by commensal and pathogenic E. coli.

In 2006, Oswald and co-workers reported that certain E. coli strains cause cell cycle arrest and DNA double strand breaks in mammalian cell lines (43). Transposon mutagenesis linked genotoxicity to a 54-kb biosynthetic gene cluster (the pks island) that encodes both polyketide synthase (PKS) and NRPS enzymes. This biosynthetic machinery was hypothesized to produce a small molecule genotoxin, colibactin. Although colibactin is produced by strains of the most well-studied bacterial species, the active genotoxin has eluded all isolation attempts and its chemical structure is currently unknown.

The first insights into colibactin’s structure originated from comparative bioinformatic analyses of the pks island and in vitro biochemical characterization of the biosynthetic enzymes involved in a self-resistance mechanism (44). Bioinformatics suggested that colibactin was initially made as an inactive, N-acylated precursor (precolibactin) that could be processed by a peptidase in the final stages of biosynthesis, releasing the active natural product. Based on this hypothesis, we biochemically characterized ClbN and ClbB, the NRPS modules predicted to initiate colibactin biosynthesis, and showed that they synthesize and elongate an N-acyl-D-Asn ‘prodrug motif’. Moreover, we demonstrated that a periplasmic peptidase, ClbP, removes this scaffold from precolibactin mimics. The relevance of this reactivity for colibactin biosynthesis was later supported by the successful isolation and structural characterization of the hydrolyzed ‘prodrug motif’ from colibactin-producing E. coli (45) and in vivo studies showing that ΔclbP mutants lack genotoxicity (46).

The characterization of this self-resistance mechanism led to the hypothesis that ΔclbP mutant strains would accumulate inactive precolibactins that might be more readily isolated. Using this approach, several groups have identified candidate precolibactins from E. coli ΔclbP mutant strains, including metabolites that contain a spirocyclopropane ring and thiazole heterocycle(s) (47-53). These structural features are found in other DNA alkylating natural products, suggesting that colibactin may directly interact with DNA (54, 55).

Although biosynthetic proposals for the isolated candidate precolibactins involve many of the pks enzymes, several genes that are essential for genotoxicity were not invoked (56). The metabolites isolated to date are therefore not likely
Natural products from the human microbiome

precursors to the active genotoxin and that additional pks enzymes play critical roles in biosynthesis. Using in vitro biochemical characterization, Piel and co-workers showed that several of these ‘missing’ enzymes generate the rare PKS extender unit aminomalonate (57). Recently, we demonstrated that this building block can be accepted by multiple PKS modules from the pks island in vitro (53). As the aminomalonate forming- and utilizing-genes are required for genotoxicity, aminomalonate is likely incorporated into colibactin. Recently, Qian and co-workers isolated and structurally characterized an aminomalonate-containing candidate precolibactin supporting this hypothesis (52). Because two essential biosynthetic enzymes remain unaccounted for, future studies will be needed to uncover the structure(s) of the true precolibactin. Since chemical synthesis allows access to many known candidate precolibactins (58), further biochemical experiments with synthetic substrates could provide the information needed to decipher the complete precolibactin structure (59). More broadly, strategically combining bioinformatics and/or in vitro reconstitution of cryptic biosynthetic pathways with chemical synthesis represents an exciting new strategy for accessing microbiome-derived natural products (60).

Outlook and Future Challenges

Our understanding of the structures and functions of natural products made by the human microbiome is in its infancy. The case studies highlighted here showcase the utility of multiple approaches for identifying and characterizing microbiota-derived metabolites (13). Integrating additional emerging strategies for natural product discovery into such efforts, such as synthetic refactoring of cryptic BGCs (61) and elicitation of cryptic gene cluster expression by small molecules (62), should enable the discovery of additional bioactive secondary metabolites from the human microbiome.

While the studies discussed here illustrate the tremendous potential of harnessing secondary metabolites from the human microbiome, they also highlight many current challenges. Given the overwhelming amount of sequencing data, how do we prioritize BGCs from the human microbiome for discovery? As in the case of lactocillin, one can focus on BGCs that produce secondary metabolites with structural similarity to known natural products possessing well-defined activities. Resistance gene-guided genome mining may also reveal potential activities of human-associated bacterial BGCs prior to isolation (63, 64). Moreover, one can envision searching metagenomes for BGCs encoding the biosynthetic machinery needed to construct specific ‘chemical warhead’ motifs using either sequence- or function-based approaches (65). Finally, analysis of the distribution of BGCs across human subjects, including various patient populations, may help to identify the pathways most relevant to human biology.

Another major challenge is establishing whether human-associated microbes actually produce particular secondary metabolites in and on the human body. In all of the studies mentioned in here, culturable strains were used to isolate and characterize natural products outside of a host context. However, it is unclear if these molecules are produced in human microbiomes. Sensitive analytical techniques, such as MALDI-imaging mass spectrometry, may help to address this issue (66). Notably, such tools have confirmed the presence of numerous secondary metabolites on the human skin (67), in lung sputum (68), and tissues of various organs (69). Developing better methods for noninvasive sampling of difficult to reach body sites, including the respiratory and gastrointestinal tracts, should prove valuable in identifying microbial natural products present in these environments.

Identifying bioactive microbiome-derived natural products and understanding their interactions with the host and other microbes has implications for human health. For instance, the biosynthetic enzymes that produce harmful secondary metabolites could be targeted for inhibition (70). Human-associated microbial metabolites could represent a source of new drugs (12). The bacterial producers of these molecules could also be potential diagnostic biomarkers or probiotic therapies (37). Finally, exploring the activities of human microbiome-derived metabolites may provide insights into important targets within the host that can be modulated for therapeutic benefit (28).

In summary, the human microbiome represents an exciting new frontier for natural
Natural products from the human microbiome

product research, but simultaneously presents substantial challenge to researchers. Efficiently identifying and characterizing microbiome-derived secondary metabolites that mediate important interactions will require collaboration between chemists, engineers, and biologists, additional tools and approaches suitable for exploring the chemistry of this complex microbial habitat.

Acknowledgements: We thank Nitzan Koppel and Vayu Maini Rekdal for helpful discussions regarding this manuscript.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: All authors contributed to the conception and design of the paper. MRW and LZ wrote the manuscript. All authors participated in revising the manuscript and approved the final version of the minireview before submission.

References
1. Berdy, J. (2005) Bioactive microbial metabolites. J Antibiot (Tokyo) 58, 1-26
2. Katz, L., and Baltz, R. H. (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43, 155-176
3. Lewis, K., Epstein, S., D’Onofrio, A., and Ling, L. L. (2010) Uncultured microorganisms as a source of secondary metabolites. J Antibiot (Tokyo) 63, 468-476
4. Wilson, M. C., and Piel, J. (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem Biol 20, 636-647
5. Sidebottom, A. M., and Carlson, E. E. (2015) A reinvigorated era of bacterial secondary metabolite discovery. Curr Opin Chem Biol 24, 104-111
6. Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., Goulding, D., and Lawley, T. D. (2016) Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. Nature 533, 543-546
7. Newman, D. J., and Cragg, G. M. (2016) Natural Products as Sources of New Drugs from 1981 to 2014. J Nat Prod 79, 629-661
8. Crawford, J. M., and Clardy, J. (2011) Bacterial symbionts and natural products. Chem Commun (Camb) 47, 7559-7566
9. Cantley, A. M., and Clardy, J. (2015) Animals in a bacterial world: opportunities for chemical ecology. Nat Prod Rep 32, 888-892
10. Weinstock, G. M. (2012) Genomic approaches to studying the human microbiota. Nature 489, 250-256
11. Human Microbiome Project, C. (2012) A framework for human microbiome research. Nature 486, 215-221
12. Donia, M. S., Cimermanic, P., Schulze, C. J., Wieland Brown, L. C., Martin, J., Mitreva, M., Clardy, J., Linington, R. G., and Fischbach, M. A. (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158, 1402-1414
13. Yang, J. Y., Karr, J. R., Watrous, J. D., and Dorrestein, P. C. (2011) Integrating ‘-omics’ and natural product discovery platforms to investigate metabolic exchange in microorganisms. Curr Opin Chem Biol 15, 79-87
14. Dorrestein, P. C., Mazmanian, S. K., and Knight, R. (2014) Finding the missing links among metabolites, microbes, and the host. Immunity 40, 824-832
15. Joice, R., Yasuda, K., Shaﬁquat, A., Morgan, X. C., and Huttenhower, C. (2014) Determining microbial products and identifying molecular targets in the human microbiome. Cell Metab 20, 731-741
16. Sharon, G., Garg, N., Debelius, J., Knight, R., Dorrestein, P. C., and Mazmanian, S. K. (2014) Specialized metabolites from the microbiome in health and disease. Cell Metab 20, 719-730
17. Donia, M. S., and Fischbach, M. A. (2015) Small molecules from the human microbiota. Science 349, 1254766
18. Koppel, N., and Balskus, E. P. (2016) Exploring and Understanding the Biochemical Diversity of the
Human Microbiota. Cell Chem Biol 23, 18-30
19. Medema, M. H., and Fischbach, M. A. (2015) Computational approaches to natural product discovery. Nat Chem Biol 11, 639-648
20. Ziemert, N., Alanjary, M., and Weber, T. (2016) The evolution of genome mining in microbes - a review. Nat Prod Rep 33, 988-1005
21. Cimermanic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158, 412-421
22. Wieland Brown, L. C., Acker, M. G., Clardy, J., Walsh, C. T., and Fischbach, M. A. (2009) Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. Proc Natl Acad Sci U S A 106, 2549-2553
23. Milsh滕en, A., Schneider, J. S., and Brady, S. F. (2014) Mining the metabiome: identifying novel natural products from microbial communities. Chem Biol 21, 1211-1223
24. Brady, S. F., Simmons, L., Kim, J. H., and Schmidt, E. W. (2009) Metagenomic approaches to natural products from free-living and symbiotic organisms. Nat Prod Rep 26, 1488-1503
25. Charlop-Powers, Z., Milsh滕en, A., and Brady, S. F. (2014) Metagenomic small molecule discovery methods. Curr Opin Microbiol 19, 70-75
26. Wang, G. Y., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., Meurer, G., Saxena, G., Andersen, R. J., and Davies, J. (2000) Novel natural products from soil DNA libraries in a streptomyces host. Org Lett 2, 2401-2404
27. Brady, S. F., and Clardy, J. (2005) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. Angew Chem Int Ed Engl 44, 7063-7065
28. Cohen, L. J., Kang, H. S., Chu, J., Huang, Y. H., Gordon, E. A., Reddy, B. V., Termei, M. A., Craig, J. W., and Brady, S. F. (2015) Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist. Proc Natl Acad Sci U S A 112, E4825-4834
29. Brady, S. F., Wagenaar, M. M., Singh, M. P., Janso, J. E., and Clardy, J. (2000) The cytosporones, new octaketide antibiotics isolated from an endophytic fungus. Org Lett 2, 4043-4046
30. Rimmerman, N., Bradshaw, H. B., Hughes, H. V., Chen, J. S., Hu, S. S., McHugh, D., Vefring, E., Janso, J. E., and Clardy, J. (2000) The cytosporones, new octaketide antibiotics isolated from an endophytic fungus. Org Lett 2, 4043-4046
31. Fu, J., Gaetani, S., Oveisi, F., Lo Verme, J., Serrano, A., Rodriguez De Fonseca, F., Rosengarth, A., Luecke, H., Di Giacomo, B., Tarzia, G., and Piomelli, D. (2003) Oleylethanolamide regulates feeding and body weight through activation of the nuclear receptor PPAR-alpha. Nature 425, 90-93
32. Weng, Z., Fluckiger, A. C., Nisitani, S., Wahl, M. I., Le, L. Q., Hunter, C. A., Fernal, A. A., Le Beau, M. M., and Witte, O. N. (1998) A DNA damage and stress inducible G protein-coupled receptor blocks cells in G2/M. Proc Natl Acad Sci U S A 95, 12334-12339
33. Le, L. Q., Kabarowski, J. H., Weng, Z., Satterthwaite, A. B., Harvill, E. T., Jensen, E. R., Miller, J. F., and Witte, O. N. (2001) Mice lacking the orphan G protein-coupled receptor G2A develop a late-onset autoimmune syndrome. Immunity 14, 561-571
34. Tlaskalova-Hogenova, H., Stepankova, R., Kozakova, H., Hudecovic, T., Vannucci, L., Tuckova, L., Rossmann, P., Hrncir, T., Kverka, M., Zakostelska, Z., Klimesova, K., Pribylova, J., Bartova, J., Sanchez, D., Fundova, P., Borovska, D., Sutkova, D., Zidek, Z., Schwarzer, M., Drastich, P., and Funda, D. P. (2011) The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. Cell Mol Immunol 8, 110-120
35. Iqbal, H. A., Low-Beinart, L., Obijialu, J. U., and Brady, S. F. (2016) Natural Product Discovery through Improved Functional Metagenomics in Streptomyces. J Am Chem Soc 138, 9341-9344
36. Piel, J. (2009) Metabolites from symbiotic bacteria. Nat Prod Rep 26, 338-362
37. Zipperer, A., Konnerth, M. C., Laux, C., Berscheid, A., Janek, D., Weidenmaier, C., Burian, M., Schilling, N. A., Slavetinsky, C., Marschal, M., Willmann, M., Kalbacher, H., Schitteck, B., Brotz-Oesterhelt, H., Grond, S., Peschel, A., and Krismer, B. (2016) Human commensals producing a novel antibiotic impair pathogen colonization. Nature 535, 511-516
38. Grice, E. A., and Segre, J. A. (2011) The skin microbiome. Nat Rev Microbiol 9, 244-253
39. Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E.
S., Mitzenmacher, M., and Sabeti, P. C. (2011) Detecting novel associations in large data sets. *Science* **334**, 1518-1524

40. Sattely, E. S., Fischbach, M. A., and Walsh, C. T. (2008) Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Nat Prod Rep* **25**, 757-793

41. Lowry, B., Walsh, C. T., and Khosla, C. (2015) In Vitro Reconstitution of Metabolic Pathways: Insights into Nature's Chemical Logic. *Synlett* **10**, 1008-1025

42. Kuo, J., Lynch, S. R., Liu, C. W., Xiao, X., and Khosla, C. (2016) Partial In Vitro Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis. *ACS Chem Biol* **11**, 2636-2641

43. Nougayrede, J. P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006) Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848-851

44. Brotherton, C. A., and Balskus, E. P. (2013) A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J Am Chem Soc* **135**, 3359-3362

45. Bian, X., Fu, J., Plaza, A., Herrmann, J., Pistorius, D., Zhang, Y., and Muller, R. (2013) In vivo evidence for a prodrug activation mechanism during colibactin maturation. *Chembiochem* **14**, 1194-1197

46. Dubois, D., Baron, O., Cougnoux, A., Delmas, J., Pradel, N., Boury, M., Bouchon, B., Bringer, M. A., Nougayrede, J. P., Oswald, E., and Bonnet, R. (2011) ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J Biol Chem* **286**, 35562-35570

47. Vizcaino, M. I., Engel, P., Trautman, E., and Crawford, J. M. (2014) Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *J Am Chem Soc* **136**, 9244-9247

48. Bian, X. Y., Plaza, A., Zhang, Y. M., and Muller, R. (2015) Two more pieces of the colibactin genotoxin puzzle from Escherichia coli show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety. *Chem Sci* **6**, 3154-3160

49. Brotherton, C. A., Wilson, M., Byrd, G., and Balskus, E. P. (2015) Isolation of a metabolite from the pks island provides insights into colibactin biosynthesis and activity. *Org Lett* **17**, 1545-1548

50. Li, Z. R., Li, Y., Lai, J. Y., Tang, J., Wang, B., Lu, L., Zhu, G., Wu, X., Xu, Y., and Qian, P. Y. (2015) Critical Intermediates Reveals New Biosynthetic Events in the Enigmatic Colibactin Pathway. *Chembiochem* **16**, 1715-1719

51. Vizcaino, M. I., and Crawford, J. M. (2015) The colibactin warhead crosslinks DNA. *Nat Chem* **7**, 411-417

52. Li, Z. R., Li, J., Gu, J. P., Lai, J. Y., Duggan, B. M., Zhang, W. P., Li, Z. L., Li, Y. X., Tong, R. B., Xu, Y., Lin, D. H., Moore, B. S., and Qian, P. Y. (2016) Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat Chem Biol* **12**, 773-775

53. Zha, L., Wilson, M. R., Brotherton, C. A., and Balskus, E. P. (2016) Characterization of Polyketide Synthase Machinery from the pks Island Facilitates Isolation of a Candidate Precolibactin. *ACS Chem Biol* **11**, 1287-1295

54. Wolkenberg, S. E., and Boger, D. L. (2002) Mechanisms of in situ activation for DNA-targeting antitumor agents. *Chem Rev* **102**, 2477-2495

55. Tanasova, M., and Sturla, S. J. (2012) Chemistry and biology of acylfulvenes: sesquiterpene-derived antitumor agents. *Chem Rev* **112**, 3578-3610

56. Balskus, E. P. (2015) Colibactin: understanding an elusive gut bacterial genotoxin. *Nat Prod Rep* **32**, 1534-1540

57. Brachmann, A. O., Garce, C., Wu, V., Martin, P., Ueoka, R., Oswald, E., and Piel, J. (2015) Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. *Chem Commun (Camb)* **51**, 13138-13141

58. Healy, A. R., Vizcaino, M. I., Crawford, J. M., and Herzon, S. B. (2016) Convergent and Modular Synthesis of Candidate Precolibactins. Structural Revision of Precolibactin Precursor. *J Am Chem Soc* **138**, 5426-5432

59. Franke, J., and Hertweck, C. (2016) Biomimetic Thioesters as Probes for Enzymatic Assembly Lines: Synthesis, Applications, and Challenges. *Cell Chem Biol*, in press

60. Chu, J., Vila-Farrés, X., Inoyama, D., Termei, M., Cohen, L. J., Gordon, E. A., Reddy, B., Charlop-Powers, Z., Zebrsoki, H. A., Gallardo-Macias, R., Jaskowski, M., Satish, S., Park, S., Perlin, D. S., Freundlich, J. S., Brady, S. F. (2016) Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat Chem Biol* **12**, 1004-1006

61. Smanski, M. J., Zhou, H., Claesen, J., Shen, B., Fischbach, M. A., and Voigt, C. A. (2016) Synthetic
Seyedsayamdost, M. R. (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci U S A* **111**, 7266-7271

Tang, X., Li, J., Millan-Aguinaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M., and Moore, B. S. (2015) Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem Biol* **10**, 2841-2849

Yeh, H. H., Ahuja, M., Chiang, Y. M., Moore, S., Yoon, O., Hajovsky, H., Bok, J. W., Keller, N. P., Wang, C. C., and Oakley, B. R. (2016) Resistance Gene-Guided Genome Mining: Serial Promoter Exchanges in Aspergillus nidulans Reveal the Biosynthetic Pathway for Fellutamide B, a Proteasome Inhibitor. *ACS Chem Biol* **11**, 2275-2284

Owen, J. G., Charlop-Powers, Z., Smith, A. G., Ternei, M. A., Calle, P. Y., Reddy, B. V., Montiel, D., and Brady, S. F. (2015) Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci U S A* **112**, 4221-4226

Biteen, J. S., Blainey, P. C., Cardon, Z. G., Chun, M., Church, G. M., Dorrestein, P. C., Fraser, S. E., Gilbert, J. A., Jansson, J. K., Knight, R., Miller, J. F., Ozcan, A., Prather, K. A., Quake, S. R., Ruby, E. G., Silver, P. A., Taha, S., van den Engh, G., Weiss, P. S., Wong, G. C., Wright, A. T., and Young, T. D. (2016) Tools for the Microbiome: Nano and Beyond. *ACS Nano* **10**, 6-37

Bouslimani, A., Porto, C., Rath, C. M., Wang, M., Guo, Y., Gonzalez, A., Berg-Lyon, D., Ackermann, G., Moeller Christensen, G. J., Nakatsuiji, T., Zhang, L., Borkowski, A. W., Meehan, M. J., Dorrestein, K., Gallo, R. L., Bandeira, N., Knight, R., Alexandrov, T., and Dorrestein, P. C. (2015) Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci U S A* **112**, E2120-2129

Quinn, R. A., Phelan, V. V., Whiteson, K. L., Garg, N., Bailey, B. A., Lim, Y. W., Conrad, D. J., Dorrestein, P. C., and Rohwer, F. L. (2016) Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *ISME J* **10**, 1483-1498

Ly, A., Buck, A., Balluff, B., Sun, N., Gorzolka, K., Feuchtinger, A., Janssen, K. P., Kuppen, P. J., van de Velde, C. J., Weirich, G., Erlmeier, F., Langer, R., Aubele, M., Zitzelsberger, H., McDonnell, L., Aichler, M., and Walch, A. (2016) High-mass-resolution MALDI mass spectrometry imaging of metabolites from formalin-fixed paraffin-embedded tissue. *Nat Protoc* **11**, 1428-1443

Martin Del Campo, J. S., Vogelaar, N., Tolani, K., Kizjakina, K., Harich, K., and Sobrado, P. (2016) Inhibition of the Flavin-Dependent Monoxygenase Siderophore A (SidA) Blocks Siderophore Biosynthesis and Aspergillus fumigatus Growth. *ACS Chem Biol*, Article ASAP, doi:10.1021/acschembio.6b00666
FOOTNOTES

*Studies of colibactin biosynthesis in the Balskus group have been supported, in whole or in part, by the Damon Runyon-Rachleff Innovation Award, the Smith Family Award for Excellence in Biomedical Research, the Searle Scholars Program, and the Packard Fellowship for Science and Engineering. This work was also supported by National Institutes of Health (1R01CA208834-01).

1Supported by an American Cancer Society – New England Division Postdoctoral Fellowship, (PF-16-122-01-CDD).

3The abbreviations used are: PK, polyketide; NRP, nonribosomal peptide; HMP, Human Microbiome Project; BGC, biosynthetic gene cluster; RiPP, ribosomally encoded and posttranslationally modified peptides; NF-κB, nuclear factor-κB; GPCR, G-protein coupled receptor; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; Asn, asparagine.
FIGURES

**Figure 1**: Overview of the sequenced-based metagenomic analysis method used to discover lactocillin.
Figure 2: Functional metagenomic strategy used to find commendamide.

made by human gut Bacteroides vulgatus
modulates human GPCR/G2A activity (EC₅₀ = 11.8 µM)
Figure 3: Schematic describing the chemical ecology approach that enabled the isolation and characterization of lugdunin.
**Figure 4**: The bioinformatics and *in vitro* biochemistry workflow that uncovered structural features of colibactin. *A*) bioinformatic analysis of the *pkS* gene cluster guided the *in vitro* biochemical characterization of the ‘prodrug’ resistance mechanism and aminomalonate-forming and utilizing enzymes, *B*) an aminomalonate-containing candidate precolibactin isolated from an *E. coli* *pkS* Δ*clbp* mutant.
