

A Deeper Dive into the Syntactical Trends Among High Impact Factor Abstracts in Scientific Literature

Steven Huang^{1*}

Abstract

With the Human Microbiome becoming an increasingly popular topic among the scientific community, there's been a growing need to publish newly obtained data in the form of research articles to contribute to the forefront of this field. However, only the most enticing research articles—often through a well-written abstract—will impact this new field the best. Similarly, the same can be applied to the forefronts of other scientific fields. Here, we assessed the content and syntactical trends between the abstracts of a range of high-impact factor Human Microbiome research articles from *Nature* as a model for what an appealing abstract would entail. Specifically, this study focused on assessing the syntactical variations found within the first third of a typical scientific abstract and offering potential explanations for the different possible use cases of the variants.

Keywords

Abstract, Impact Factor, Altmetric Score, Biomedical Abstract

¹*Department of Molecular and Cellular Biology, University of California, Berkeley*

Introduction

In many cases, the abstract of a hypothesis testing focused journal article under the biomedical sciences aims to not only provide an overview of the content presented within the article but also attract potential readers into reading the article¹. By offering an initial impression to the important portions discussed in the article, the abstract of hypothesis testing journal articles in the biomedical sciences highlights elements of the entire article's hypothesis, objective, experiments, and result findings that either align or derail with the article's original hypothesis.

However, many biomedical science articles that follow a simple abstract structure of an article being published in a journal—whether it is a top journal like *Nature*, or *The New England Journal of Medicine*—don't necessarily correlate with a high individual impact factor on outreach to audiences^{1, 2}. In fact, a biomedical science article's impact factor depends on a myriad of ways that its respective audience is impacted, whether it be through citations, views, or file downloads. Up to this point, any prospective practitioners that enter the field of biomedical sciences may have trouble discerning between writing an appealing abstract which garners the attention of many, or an acceptable abstract that isn't as popular.

In this study, we specifically focused on analyzing a set of high-impact factor abstracts in the Human Microbiome field published in "*Nature*," a scientific journal ranked among the top scientific journals that cover information from all fields between the sciences and technology³. We hypothesized that the organizational pattern of syntactical elements in abstracts of high impact Human Microbiome articles can model successful syntactical elements in high-impacting abstracts from other fields. Because newer topics at the forefront of the sciences may not have many examples of

highly impactful writing, this study models what successful, impactful writing may look like for those fields.

Methods

High Impact Factor Article Search-up

The articles from this study were all obtained directly from the *Nature* journal's online database, accessed by UC resources. To increase our specificity of articles, we then navigated directly to the *Nature* journal website and searched the keywords "Human microbiome AND Diet" and used the filters "Nature," "Research," and "2018-2022" to access a similar batch of articles to get 53 total search results. Among the 53 total search results, there were 15 articles that did not match our keyword search by search engine error, giving us a total of 38 articles to analyze (Extended Data Fig. 1a).

Individual Article Abstract Analysis

Among each of the n=38 articles obtained, we checked for the presence of fifteen components of a well-written abstract as categorical variables based on Bahadoran et al.'s guide on writing a successful biomedical science abstract (Supplementary Table 1). Each analysis followed the same four steps: **1.)** Download the article from "*Nature*" to the annotation editor, GoodNotes. **2.)** Following 15 components of an abstract in order, identify and annotate the presence of those components within the abstract section of the downloaded article. **3.)** Mark a "1" for components that are present and mark a "0" for components that are not present⁴. Use the word counter software, CountWordsWorth, and determine the average number of words per sentence and the total number of words among each abstract (Supplementary Table 2).

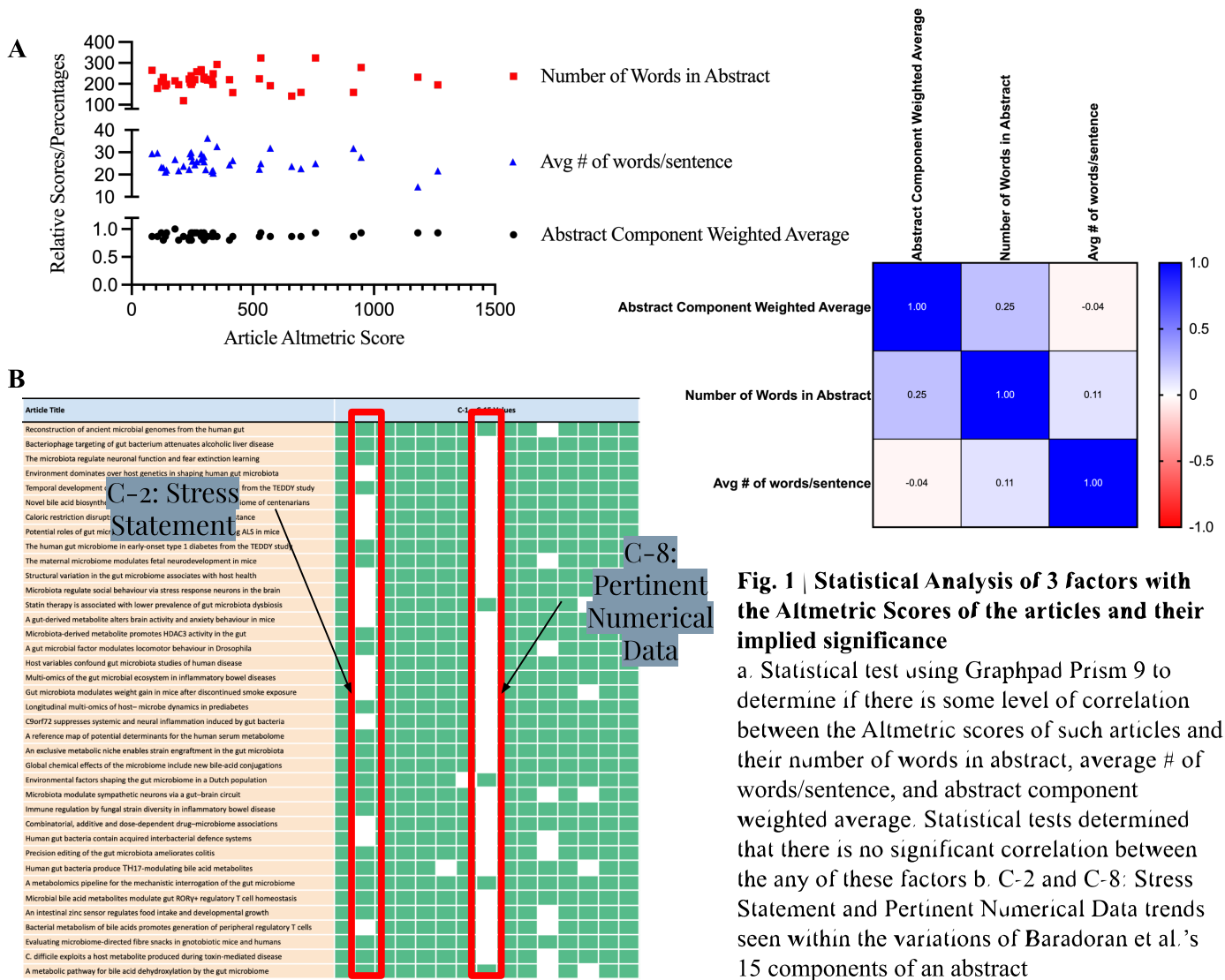


Fig. 1 | Statistical Analysis of 3 factors with the Altmetric Scores of the articles and their implied significance

a. Statistical test using Graphpad Prism 9 to determine if there is some level of correlation between the Altmetric scores of such articles and their number of words in abstract, average # of words/sentence, and abstract component weighted average. Statistical tests determined that there is no significant correlation between the any of these factors b. C-2 and C-8: Stress Statement and Pertinent Numerical Data trends seen within the variations of Baradoran et al.'s 15 components of an abstract

Prevention of Bias from Impact Factor

To prevent bias in marking the presence of components of an article caused by knowing its impact factor, we assigned impact factors to each article using the Altmetric score after all the 38 articles' abstracts were analyzed (Supplementary Table 2). The Altmetric score is a measure of an article's overall impact on its audiences given that it provides a quantitative value that summarizes the reach, promiscuity, potential bias, and source mentions of an article^{4, 5, 6}.

Statistical Analysis and Article Correlation

To obtain statistical data and perform quantitative analysis between the 38 abstracts, we first took the weighted average of the 15 components of a well-written abstract for each article, then performed our statistical analysis and tests on Prism 9, a graphing and biostatistical analysis software, using their correlation test between the Article Altmetric score and four factors: Abstract Component Weighted Average, Abstract Total Word Count, and Abstract Average # of Words/Sentence (Supplementary Table 2).

Each of these tests was weighted at a critical value of 0.05 when p-values were analyzed for hypothesis testing. The null and alternative hypotheses for correlation tests are: H_0 : There is no correlation between the Altmetric score and the 3 variables. H_a : There is some sort of correlation between the Altmetric score and at least 1 of the 3 variables.

Findings and Discussion

Article Syntactical Components Show No Significant Difference but Have Variation Trends

Initially, we hypothesized that because all our articles should be already highly impactful, their syntactical trends should be relatively similar, and therefore, their Altmetric scores are large and impactful enough to the point where there should be no correlation between variations in Altmetric score and syntactical trends. We then performed a robustness test to determine if our articles we tested for whether our 38 samples are statistically identical or not through a correlation test.

In our test, we correlated the Altmetric scores of our 38 samples and 3 variables: Abstract Component Weighted Average ($r = 0.246$, $P\text{-value} \approx 0.1892$), Abstract Total Word Count ($r = 0.113$, $P\text{-value} \approx 0.8499$), and Abstract Average # of Words/Sentence ($r = -0.043$, $P\text{-value} \approx 0.2341$). Results on Prism 9 have shown that there is no significant correlation between the Altmetric score and any of the 3 variables (Fig. 1a.). Additional analysis of our Altmetric scores has also shown that our minimum and maximum scores are between the 96th and 99th percentile in terms of impact⁴. In short, our robustness test has shown that there is no correlation between variations in Altmetric score and syntactical trends of our abstracts, highlighting the importance of this study proceeding with a relatively pure sample of articles that are similar enough in terms of syntactical elements and impact.

Through this test, we're able to direct our focus towards analyzing the smaller variations that have been displayed inside our data, specifically Bahadoran et al.'s 15 components of an abstract. Specifically, we observed that there is the variation found in the number of abstracts that contain stress statements (C-2). Roughly $\frac{1}{3}$ of the abstracts present does not contain a stress statement. Additionally, we also observed that almost all abstracts are missing one of Bahadoran et al.'s components, Pertinent Numerical Data (C-8, Fig. 1b.).

Variation Within the First Third of the Abstract

Within the first third of the abstract, we see one major variation

in its general structure around the presence of the stress statement, a phrase consisting of conjunctions–statements like “but,” “however,” “therefore,” etc. that focus on identifying the major gap in knowledge in a field. Generally, the stress statement in a scientific hypothesis testing article follows the explanation of background information and precedes a statement about the article’s hypotheses or questions to be addressed⁷. However, our results have shown that 14 of the 38 articles are missing a stress statement and instead just have a long piece of background information followed by the article’s hypothesis or questions.

In a deeper analysis, we determined that based on our initial observations there are a total of 4 unique variations in the syntactical structure of the first third of the abstract. These 4 variations were split into 2 groups, one group consisting of abstracts that contain the stress statement and another consisting of abstracts without the stress statement. Both groups contained one variation with a longer background and another with a shorter background. Further, we noticed that the group of abstracts without a stress statement contained a modified version of the stress statement’s emphasis on the gap in knowledge in the field. Specifically, we see the longer versions of the abstracts with Needham et al. or Funabashi et al. and shorter versions of the abstracts with Zeevi et al. and Gacesa et al.’s abstracts (Fig. 2a, Fig. 2b, Fig. 2c, Fig. 2d).

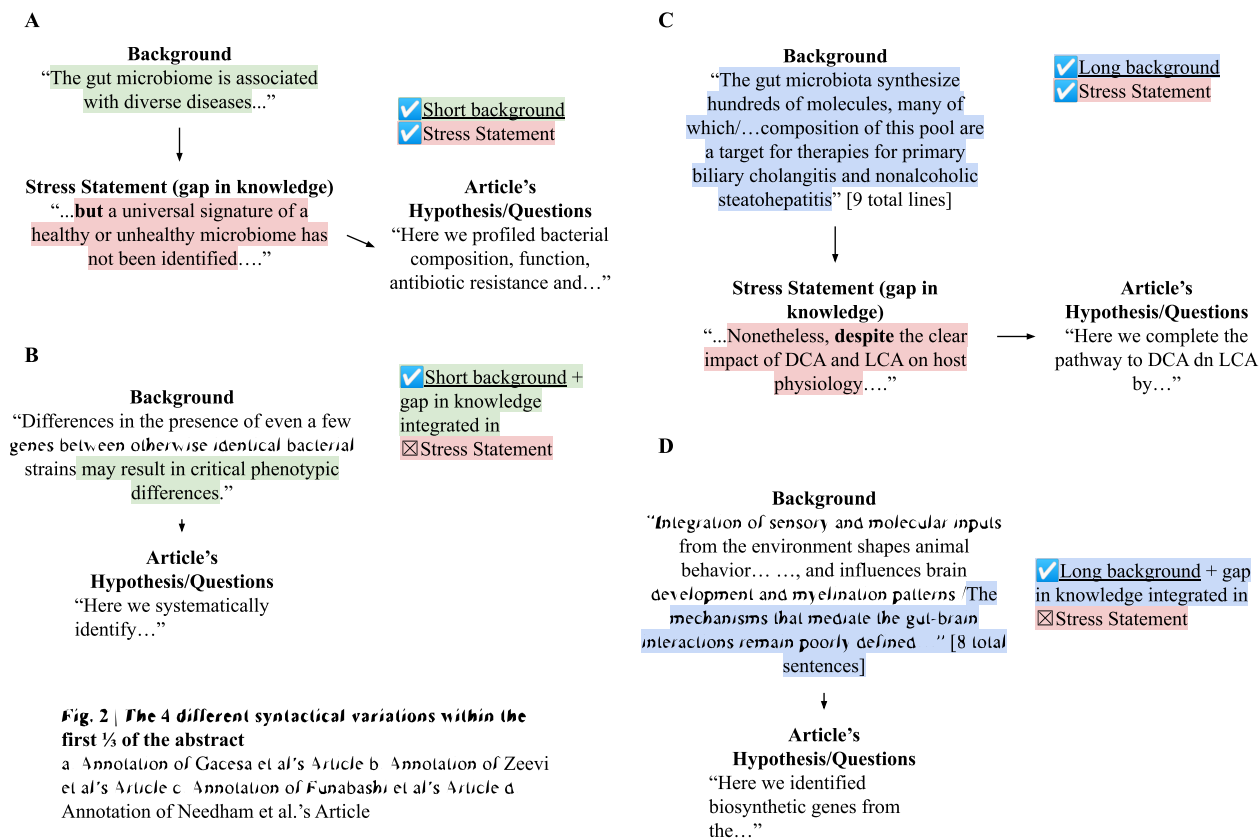


Fig. 2 | The 4 different syntactical variations within the first $\frac{1}{3}$ of the abstract
a Annotation of Gacesa et al’s Article b Annotation of Zeevi et al’s Article c Annotation of Funabashi et al’s Article d Annotation of Needham et al.’s Article

Among these four variants, our results showed each of them following a similar syntactical pattern of starting with a background, then addressing a gap in knowledge—whether it is through a stress statement or integrated in the background—and ending with the hypotheses or questions present (Fig. 2a, Fig. 2b, Fig. 2c, Fig. 2d). Despite the length of the background, it seems that if the author was able to smoothly integrate the gap in understanding into their background, they would do so as displayed through Gascesa et al. and Funabashi et al.’s abstracts (Fig. 2b, Fig. 2d). On the other hand, if the author was not able to integrate the gap in understanding as well, they instead used a simple stress statement as seen with Needham et al. and Zeevi et al.’s abstracts (Fig. 2a, Fig. 2c).

In a broader context, this means we have two successful ways of communicating the gap in knowledge in biomedical scientific abstracts. The first is to incorporate the gap in knowledge in our background smoothly. If we’re unable to do so, we can rely on the second, where we utilize a stress statement to specifically highlight the gap in knowledge.

Variation Among the Numerical Data in the Last Third of the Abstract

Another variation observed among our 38 articles on Bahadoran et al.’s 15 components is that 90% of them were missing the Pertinent Numerical Data category (C-8). This completely contrasts with Bahadoran et al.’s suggestion for including slight numerical values to support conclusions in an abstract¹.

We believe one reason why this may be true for most of our data is that reviewers and other audiences that read an abstract want to understand the significance, final claims, and findings of a study, basing what they read off trust⁷. In this case, under the assumption that the scientific community is built upon the aspect of trust in ethical research, there is no need to present statistical p-values in the abstract when readers can trust the article to present a more complete version of the claims later in the article⁷.

Final Discussion

While the Altmetric score and syntactical components of a biomedical science abstract have been associated together in this study, other factors that affect the Altmetric score remain largely unexplored. Therefore, this study has several limitations. First, other components in a biomedical article like the methods, findings, and discussion section may also have an effect on the articles impact factor as measured by the Altmetric score. Second, because all our successful abstract samples originated from the journal “Nature,” not any other journal, there may be inherent bias in terms of how successful “Nature” articles compare with other successful biomedical scientific journals like “Science,” “Cell,” etc. Finally, because this study searched for abstracts under a topic rising in popularity in the scientific field, The Human Microbiome, correlations, and results may be different for abstracts in fields that are not under a topic rising in popularity in the scientific field.

Overall, this study’s comparison between the Altmetric score and syntactical components of a biomedical science abstract using Bahadoran et al.’s 15 components [also average word count and total word length] has yielded a pure sample of highly impactful ab-

stracts. These results suggest that Bahadoran et al.’s 15 components of a successful abstract can potentially be referenced by anyone interested in writing a successful biomedical scientific abstract. With these initial findings, this study also sets up a potential framework for future studies on the syntactical trends within a journal article’s effect on the article’s impact factor.

References

1. Bahadoran Z, Mirmiran P, Kashfi K, Ghasemi A. The Principles of Biomedical Scientific Writing: Abstract and Keywords. *International Journal of Endocrinology and Metabolism*. 2020;18(1).
2. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*. 2003;91(1):42–46. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC141186/>
3. Lab S. Scimago Journal Rankings. *Scimagojr.com*. 2018. <https://www.scimagojr.com/journalrank.php>
4. Bookmarklet. Altmetric. 2015 Dec 29 [accessed 2022 Apr 11]. <https://www.altmetric.com/products/free-tools/bookmarklet/>
5. Trueger NS, Thoma B, Hsu CH, Sullivan D, Peters L, Lin M. The Altmetric Score: A New Measure for Article-Level Dissemination and Impact. *Annals of Emergency Medicine*. 2015
6. Altmetric. Altmetric Attention Score. Altmetric. 2021 Sep 21
7. National Academy of Sciences. On Being a Scientist: A Guide to Responsible Conduct in Research--Third Edition. National Academies Press.
8. Baylor College of Medicine. Human Microbiome Project. Baylor College of Medicine. 2011. <https://www.bcm.edu/departments/molecular-virology-and-microbiology/research/human-microbiome-project>
9. Grogan KE. Writing Science: What Makes Scientific Writing Hard and How to Make It Easier. *The Bulletin of the Ecological Society of America*.
10. Ghasemi A, Bahadoran Z, Mirmiran P, Hosseini F, Shiva N, Zadeh-Vakili A. The Principles of Biomedical Scientific Writing: Discussion. *International Journal of Endocrinology and Metabolism*. 2019
11. Aksnes DW, Langfeldt L, Wouters P. Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *SAGE Open*. 2019;9(1):
12. Cooper C, Booth A, Varley-Campbell J, Britten N, Garside R. Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies. *BMC Medical Research Methodology*. 2018;18(1).
13. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen M a. Y, Bolte LA, Andreu-Sánchez S, Chen L, Collij V, Hu S, et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature*. 2022 Apr 13
14. Paik D, Yao L, Zhang Y, Bae S, D’Agostino GD, Zhang M, Kim E, Franzosa EA, Avila-Pacheco J, Bisanz JE, et al. Human gut bacteria produce TH17-modulating bile acid metabolites. *Nature*. 2022
15. Li XV, Leonardi I, Putzel GG, Semon A, Fiers WD, Kusakabe T, Lin W-Y, Gao IH, Doron I, Gutierrez-Guerrero A, et al. Immune regulation by fungal strain diversity in inflammatory bowel disease. *Nature*. 2022 16. Needham BD, Funabashi M, Adame MD, Wang Z, Boktor JC, Haney J, Wu W-L, Rabut C, Ladowsky MS, Hwang S-J, et al. A gut-derived metabolite alters brain activity and anxiety behaviour in mice. *Nature*. 2022 Feb 14:1–7. <https://www.nature.com/articles/s41586-022-04396-8>. doi:10.1038/s41586-022-04396-8
17. Fluhr L, Mor U, Kolodziejczyk AA, Dori-Bachash M, Leshem A, Itav S, Cohen Y, Suez J, Zmora N, Moresi C, et al. Gut microbiota modulates weight gain in mice after discontinued smoke exposure. *Nature*. 2021 Dec 8
18. Sato Y, Atarashi K, Plichta DR, Arai Y, Sasajima S, Kearney SM, Suda W, Takeshita K, Sasaki T, Okamoto S, et al. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature*. 2021
19. Han S, Van Treuren W, Fischer CR, Merrill BD, DeFelice BC, Sanchez JM, Higginbottom SK, Guthrie L, Fall LA, Dodd D, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature*. 2021
20. Wu W-L, Adame MD, Liou C-W, Barlow JT, Lai T-T, Sharon G, Schretter CE, Needham BD, Wang MI, Tang W, et al. Microbiota regulate social behaviour via stress response neurons in the brain. *Nature*. 2021

21. Delannoy-Bruno O, Desai C, Raman AS, Chen RY, Hibberd MC, Cheng J, Han N, Castillo JJ, Couture G, Lebrilla CB, et al. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature*. 2021
22. Surana NK, Kasper DL. Moving beyond microbiome-wide associations to causal microbe identification. *Nature*. 2018 [accessed 2022 May 4];552(7684):244–247. doi:10.1038/nature25019
23. Wibowo MC, Yang Z, Borry M, Hübner A, Huang KD, Tierney BT, Zimmerman S, Barajas-Olmos F, Contreras-Cubas C, García-Ortiz H, et al. Reconstruction of ancient microbial genomes from the human gut. *Nature*. 2021
24. Pruss KM, Sonnenburg JL. *C. difficile* exploits a host metabolite produced during toxin-mediated disease. *Nature*. 2021
25. Crunkhorn S. Applying the human metabolome. *Nature Reviews Drug Discovery*. 2019
26. Song X, Sun X, Oh SF, Wu M, Zhang Y, Zheng W, Geva-Zatorsky N, Jupp R, Mathis D, Benoist C, et al. Microbial bile acid metabolites modulate gut ROR γ ⁺ regulatory T cell homeostasis. *Nature*. 2019
27. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. *Nature*. 2020, Coley EJJ, Siegler EL, Qiu A, Kazantsev M, Wilson CJ, Rendon T, Hsiao EY. The maternal microbiome modulates fetal neurodevelopment in mice. *Nature*. 2020.
29. Wu S, Hashimoto-Hill S, Woo V, Eshleman EM, Whitt J, Engleman L, Karns R, Denson LA, Haslam DB, Alenghat T. Microbiota-derived metabolite promotes HDAC3 activity in the gut. *Nature*. 2020
30. Muller PA, Schneeberger M, Matheis F, Wang P, Kerner Z, Ilanges A, Pellegrino K, del Mármol J, Castro TBR, Furuichi M, et al. Author Correction: Microbiota modulate sympathetic neurons via a gut–brain circuit. *Nature*. 2020.
31. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature*. 2020
32. Pham VT, Lacroix C, Braegger CP, Chassard C. Lactate-utilizing community is associated with gut microbiota dysbiosis in colicky infants. *Scientific Reports*. 2017
33. Burberry A, Wells MF, Limone F, Couto A, Smith KS, Keaney J, Gillet G, van Gastel N, Wang J-Y, Pietiläinen O, et al. C9orf72 suppresses systemic and neural inflammation induced by gut bacteria. *Nature*. 2020
34. Campbell C, McKenney PT, Konstantinovskiy D, Isaeva OI, Schizas M, Verter J, Mai C, Jin W-B, Guo C-J, Violante S, et al. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature*. 2020
35. Redhai S, Pilgrim C, Gaspar P, Giesen L van, Lopes T, Riabinina O, Grenier T, Milona A, Chanana B, Swadling JB, et al. An intestinal zinc sensor regulates food intake and developmental growth. *Nature*. 2020
36. Von Schwartzberg RJ, Bisanz JE, Lyalina S, Spanogiannopoulos P, Ang QY, Cai J, Dickmann S, Friedrich M, Liu S-Y, Collins SL, et al. Caloric restriction disrupts the microbiota and colonization resistance. *Nature*. 2021
37. Quinn RA, Melnik AV, Vrbnac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, et al. Global chemical effects of the microbiome include new bile-acid conjugations. *Nature*. 2020
38. Duan Y, Llorente C, Lang S, Brandl K, Chu H, Jiang L, White RC, Clarke TH, Nguyen K, Torralba M, et al. Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. *Nature*. 2019
39. Ross BD, Verster AJ, Radey MC, Schmidtke DT, Pope CE, Hoffman LR, Hajar AM, S. Brook Peterson, Elhanan Borenstein, Mougous JD. Human gut bacteria contain acquired interbacterial defence systems. *Nature*. 2019
40. Chu C, Murdock MH, Jing D, Won TH, Chung H, Kressel AM, Tsaava T, Addorisio ME, Putzel GG, Zhou L, et al. The microbiota regulate neuronal function and fear extinction learning. *Nature*. 2019
41. Blacher E, Bashiardes S, Shapiro H, Rothschild D, Mor U, Dori-Bachash M, Kleimayer C, Moresi C, Harnik Y, Zur M, et al. Potential roles of gut microbiome and metabolites in modulating ALS in mice. *Nature*. 2019
42. Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, et al. Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature*. 2019
43. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, Weinberger A, Fu J, Wijmenga C, Zhernakova A, et al. Structural variation in the gut microbiome associates with host health. *Nature*. 2019
44. Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K, Lernmark Å, Hagopian WA, Rewers MJ, She J-X, et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature*. 2018
45. Stewart CJ, Ajami NJ, O’Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni H, Metcalf GA, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018
46. Schretter CE, Vielmetter J, Bartos I, Marka Z, Marka S, Argade S, Mazmanian SK. A gut microbial factor modulates locomotor behaviour in *Drosophila*. *Nature*. 2018
47. Shepherd ES, DeLoache WC, Pruss KM, Whitaker WR, Sonnenburg JL. An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature*. 2018
48. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature*. 2018
49. Zhu W, Winter MG, Byndloss MX, Spiga L, Duerkop BA, Hughes ER, Büttner L, de Lima Romão E, Behrendt CL, Lopez CA, et al. Precision editing of the gut microbiota ameliorates colitis. *Nature*. 2018

Supplementary Information

	Parameter	Code	Description	Conditions
Syntactical Components of First 1/3 of Abstract	Background Information	C-1	Short description that highlights any contextual information that is relevant to the study.	1 if true, 0 if false
	Stress Statement	C-2	Call to action statement with keywords like "However", "But", etc. to address the field's gap in understanding	1 if true, 0 if false
	Hypothesis/Question Statement	C-3	A statement that focuses on what the study is aiming to test	1 if true, 0 if false
Syntactical Components of Middle 1/3 of Abstract	Explanation of Design	C-4	Explanation of the study design	1 if true, 0 if false
	Explanation of Study Setting	C-5	Explanation of the study parameters	1 if true, 0 if false
	Evaluation of Participants/Materials	C-6	Phrases that include the participants, materials, or other resources used in the study	1 if true, 0 if false
	Description of Standard Technique	C-7	Phrases that go into detail on the standard testing techniques	0 if true, 1 if false
Syntactical Components of Ending 1/3 of Abstract	Specific Pertinent Numerical Data as Results	C-8	Specific data values that summarize the significance of findings at the end of the study	1 if true, 0 if false
	Non-inclusion of P-values Without Providing μ , n , and σ	C-9	Phrases that contain P-Values without the mean, sample size, and standard deviation	0 if true, 1 if false
	Statement of Broader Importance	C-10	Phrase or sentence explaining potential applications or follow-up studies done	1 if true, 0 if false
General Syntactical Components	Logical Order of Presentation	C-11	Logical flow of wording, matches the ordering of initial claims with the ordering of the conclusion statement(s)	1 if true, 0 if false
	Broad Claims	C-12	Overly generalized claims	0 if true, 1 if false
	Verb Tense A	C-13	Present tense for question	1 if true, 0 if false
	Verb Tense B	C-14	Past/present tense when answering question	1 if true, 0 if false
	Signaling to Topics	C-15	Usage of signaling transition words ("Hypothesis", "As a result", "We conclude", etc.)	1 if true, 0 if false

Supplementary Table 1 | The 15 parameters used from Baradoran et al.'s Abstract Guide

The 15 parameters, broken down into 4 sections: the first 1/3, 2/3, 3/3 of the abstract, and and grammatical structures. Contains conditions that are used for the weighted average score.

Article Title	C-1 – C-15 Values	Weighted Avg	Altmetric Score	Total # of Words	Average Words Per Sentence
Reconstruction of ancient microbial genomes from the human gut	1 1 1 1 1 1 1 1 1 1 0 1 1 1 1	93.33%	1264	195	21.67
Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	1181	232	14.5
The microbiota regulate neuronal function and fear extinction learning	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	947	278	27.8
Environment dominates over host genetics in shaping human gut microbiota	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	915	159	31.8
Temporal development of the gut microbiome in early childhood from the TEDDY study	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	759	324	24.92
Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	699	159	22.71
Caloric restriction disrupts the microbiota and colonization resistance	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	660	142	23.67
Potential roles of gut microbiome and metabolites in modulating ALS in mice	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	572	191	31.83
The human gut microbiome in early-onset type 1 diabetes from the TEDDY study	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	533	324	24.92
The maternal microbiome modulates fetal neurodevelopment in mice	1 1 1 1 1 1 1 1 0 1 1 0 1 1 1	86.67%	527	224	22.4
Structural variation in the gut microbiome associates with host health	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	417	158	26.33
Microbiota regulate social behaviour via stress response neurons in the brain	1 0 1 1 1 1 1 1 0 1 1 1 1 1 0	80.00%	404	220	24.44
Statin therapy is associated with lower prevalence of gut microbiota dysbiosis	1 0 1 1 1 1 1 1 1 1 1 1 1 0 1	86.67%	352	293	32.56
A gut-derived metabolite alters brain activity and anxiety behaviour in mice	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	336	248	20.67
Microbiota-derived metabolite promotes HDAC3 activity in the gut	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	335	197	21.89
A gut microbial factor modulates locomotor behaviour in Drosophila	1 1 1 1 1 1 1 1 0 1 1 0 1 1 1	86.67%	331	215	21.5
Host variables confound gut microbiota studies of human disease	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	313	218	36.33
Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	305	222	22.2
Gut microbiota modulates weight gain in mice after discontinued smoke exposure	1 0 1 1 1 1 1 1 0 1 1 1 1 0 1	80.00%	299	232	25.78
Longitudinal multi-omics of host–microbe dynamics in prediabetes	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	298	224	28
C9orf72 suppresses systemic and neural inflammation induced by gut bacteria	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	287	264	29.33
A reference map of potential determinants for the human serum metabolome	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	286	268	26.8
An exclusive metabolic niche enables strain engraftment in the gut microbiota	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	269	258	25.8
Global chemical effects of the microbiome include new bile-acid conjugations	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	260	220	24.44
Environmental factors shaping the gut microbiome in a Dutch population	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	250	208	26
Microbiota modulate sympathetic neurons via a gut–brain circuit	1 1 1 1 1 1 1 1 0 1 1 0 1 0 1	80.00%	246	197	28.14
Immune regulation by fungal strain diversity in inflammatory bowel disease	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	245	239	29.99
Combinatorial, additive and dose-dependent drug–microbiome associations	1 0 1 1 1 1 1 1 0 1 1 1 1 1 1	86.67%	241	206	29.43
Human gut bacteria contain acquired interbacterial defence systems	1 0 1 1 1 1 1 1 0 1 1 0 1 1 1	80.00%	237	223	22.3
Precision editing of the gut microbiota ameliorates colitis	1 1 1 1 1 1 1 1 0 1 1 0 1 1 1	86.67%	214	119	23.8
Human gut bacteria produce TH17-modulating bile acid metabolites	1 1 1 1 1 1 0 1 0 1 1 1 1 0 1	80.00%	194	196	21.78
A metabolomics pipeline for the mechanistic interrogation of the gut microbiome	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	100.00%	179	214	26.75
Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	144	198	22
An intestinal zinc sensor regulates food intake and developmental growth	1 1 1 1 1 1 1 1 0 1 1 0 1 1 1	86.67%	139	190	21.11
Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells	1 0 1 1 1 1 1 1 0 1 1 0 1 1 1	80.00%	130	231	23.1
Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans	1 1 1 1 1 1 1 1 0 1 1 1 1 1 1	93.33%	122	210	23.33
C. difficile exploits a host metabolite produced during toxin-mediated disease	1 1 1 1 1 1 1 1 0 1 1 0 1 1 1	86.67%	106	178	29.67
A metabolic pathway for bile acid dehydroxylation by the gut microbiome	1 1 1 1 1 1 1 1 1 1 0 1 0 1 1	86.67%	84	265	29.44

Supplementary Table 2 | Comparison and Raw Data Before Statistical Tests
 Contains all of the data that is used for the statistical tests on Graphpad Prism 9