



OPEN ACCESS

EDITED BY

Mauro Cataldi,
University of Naples Federico II, Italy

REVIEWED BY

Francisco Jose Roig,
Universidad San Jorge, Spain

*CORRESPONDENCE

Priyankar Dey

✉ priyankardey28@gmail.com;

✉ priyankar.dey@thapar.edu

RECEIVED 28 August 2024

ACCEPTED 18 October 2024

PUBLISHED 11 November 2024

CITATION

Dey P and Choubey S (2024) An urgent need for longitudinal microbiome profiling coupled with machine learning interventions.

Front. Microbiol. 15:1487841.

doi: 10.3389/fmicb.2024.1487841

COPYRIGHT

© 2024 Dey and Choubey. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An urgent need for longitudinal microbiome profiling coupled with machine learning interventions

Priyankar Dey^{1*} and Sandeep Choubey^{2,3}

¹Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala, India, ²The Institute of Mathematical Sciences, Chennai, India, ³Homi Bhabha National Institute, Mumbai, India

KEYWORDS

liver, microbiome, gut-liver axis, machine learning, artificial intelligence, microbiota, gut, intestine

1 Introduction: lack of universal gut microbial signatures associated with metabolic liver disease

Our understanding of the role of gut microbes in human health and disease has come a long way since John M. Whipps et al. first defined the term “microbiome.” Since the early 2000s, with the gradual lowering of the cost of commercial DNA sequencing, health science has been flooded with 16S rRNA data. Unfortunately, despite a plethora of pre-clinical and clinical publications on the gut–liver axis, the majority of our understanding of the microbiota–liver reciprocal interaction remains limited to correlation analysis. The most exciting part of such metagenomic studies is the sheer amount of “big data” generated, which is rather easy to correlate with physiological variables; the bigger the metagenomic data and number of independent variables, the more the chances to find “significant” associations. Notably, our limited understanding of the gut–liver axis is derived from the microbiome, rather than the microbiota. Therefore, the conclusions obtained through microbiome-related correlation studies often do not reflect causation and are not representative of a universal phenomenon, and there have been almost no true microbial markers of dysbiosis linked to chronic liver disease.

2 Discussion

2.1 The boundary between the known and the unknown

Translational potentials of pre-clinical microbiota–liver associations in clinical disease prediction and treatment have not been very successful despite enormous numbers of interventional and observational studies. In fact, the relevance of utilizing rodents’ gut microbial signatures in understanding human diseases has been criticized due to the massive difference in the gut microbes between both species, attributed to the gastrointestinal biogeography and genetic makeup (Nguyen et al., 2015). Especially in the high-fat diet model of metabolic disease, there is a lack of consistency between good and bad microbes. In fact, diseases have also been associated with the strains belonging to probiotics (e.g., *Lactobacillus*) and commensals (e.g., *Akkermansia* and *Faecalibacterium*) (Dey and Ray Chaudhuri, 2023). Furthermore, confounding results are obtained from chemical-induced metabolic disease models that are physiologically mostly irrelevant (Dey, 2020) and data from germ-free mice that possess innate defects

in various physiological processes (Jans and Vereecke, 2024). The clinical progression of metabolic liver disease is distinct from pre-clinical models of chronic liver disease in terms of timeline, cellular phenotype, pathological complexity, the difference in immune responses, and metabolic machinery (Liu et al., 2013). Furthermore, the innate and general differences between human and animal disease models, such as biological differences (e.g., genetic makeup, organ anatomy, and liver functional capacity), extent of disease complexity (e.g., model specificity and co-influence of comorbidities), ability to perform controlled experiments, and predictive validity, make it challenging to conclude on clinical gut microbial phenotypes based on pre-clinical data. Although there have been emerging reports of *in vitro* models of the human distal intestine (Qi et al., 2023), these models are physiologically irrelevant given their host-independent nature.

Due to the advent of culturomics techniques and controlled clinical studies, pre-clinical gut microbial patterns that were initially considered associated with disease conditions, such as Firmicutes-to-Bacteroides ratio, enrichment of energy-harvesting species, and specific metabolic functions, seem to be falling apart. For instance, γ -proteobacteria, due to the presence of lipopolysaccharide (LPS), were previously thought to simply cause hepatic inflammation. However, studies have identified that the LPS-TLR4-inflammation axis cannot be generalized due to huge differences in LPS structure dictating the extent of immune response (Picarello, 2022) and that the majority of the luminal LPS-supplying Bacteroides rather display immunosuppressive characteristics (d'Hennezel et al., 2017). Today, it is strongly recognized that the good, bad, and ugly nature of the gut microbiota is condition-specific. Factors such as the availability of preferred nutrients, pathoadaptive mutations, and potential to evade the mucosal immune response have been recognized as critical factors that define a specific microbial species as commensal or pathobiont (Dey and Ray Chaudhuri, 2023; Dey, 2024). The bottom line is that there is no proper clinical definition of dysbiosis and eubiosis in terms of specific microbial features. However, the only aspect almost universally accepted is that loss of gut microbial diversity is associated with chronic liver disease, and when we talk about the diversity, it is the community effects not disease causation by a single species.

2.2 A critical lack of liver-specific longitudinal studies

A recent systematic review and meta-analysis of 54 clinical studies have indicated substantial inter-study heterogeneity in gut microbial taxonomic identification, in which the enrichment of inflammation-inducing genera was more closely associated with non-alcoholic fatty liver disease, but no genera were identified to provide long-term disease risk-predictive value (Su et al., 2024). To date, there have been more than 100 cross-sectional studies to identify the predominant gut microbes associated with metabolic liver disease, yet no absolute core microbiome related to progressive liver disease has been identified. Although longitudinal studies are considered superior to cross-sectional studies and that informed understanding of disease pathogenesis can only

be obtained through the former, there have been only a few longitudinal studies undertaken to link the gut microbiome with liver health. One study from Kyoto (Japan) evaluated the gut microbial alterations from pre-transplantation to 2 months post-surgery in 38 liver transplant patients (Kato et al., 2017). Data show an initial decline and later increase of microbial diversity, along with the overall increase of Bacteroides, Enterobacteriaceae, Streptococcaceae, and Bifidobacteriaceae, while a depletion in the abundance of Lactobacillaceae, Enterococcaceae, Clostridiaceae, Peptostreptococcaceae, and Ruminococcaceae was noted. The authors acknowledged that variations in antibiotic regimes, food, synbiotics, and patient heterogeneity (e.g., various donors and underlying diseases) likely influence the study's findings. Confounding variables and data on relative microbial abundance were among the limitations. In line, a relatively recent study from the National Institutes of Health, USA, investigated the gut-liver axis in hepatitis C patients, taking into account varied degrees of fibrosis severity (Ali et al., 2023). The investigation was performed 6 months after HCV was undetectable ($n = 23$) and before ($n = 29$) attaining a durable virologic response. Data suggest that increased hepatic fibrosis was correlated with *Anaerostipes hadrus*, while *Bacteroides vulgatus* with portal inflammation in HCV. A prolonged virologic response suggests that *Methanobrevibacter smithii* may have a beneficial effect on indicators of the severity of liver disease. Although this longitudinal investigation made it possible to compare HCV-infected individuals with a supposedly improved viral clearance state, they were unable to report gut microbial patterns in healthy controls due to the unavailability of samples. Another recent study from Stanford examines the gut microbial composition at various body sites (including gut) and correlated with host multi-omics, immunological, and clinical indicators (including hepatic) ($n = 86$) over 6 years to comprehend the dynamic interaction between the human microbiomes and host throughout health and illness (Zhou et al., 2024). Despite identifying microbial compositional and diversity patterns associated with host physiological parameters and metabolites, no temporal associations were derived between the gut microbiota and the measured liver-specific parameters (e.g., transaminases). Emerging studies claim the causative effects of oral microbiota in the pathogenesis of chronic metabolic disease, including liver disease (Gupta and Dey, 2023), and there has been no longitudinal study undertaken in this line to date. Thus, understanding the true nature of gut microbial dynamics under the course of hepatic disease pathogenesis and remission remains largely unknown.

2.3 Machine learning in pattern recognition along the gut-liver axis

Beyond the availability of longitudinal gut microbial data, a fundamental difficulty in analyzing large-scale microbiome big-data lies in their high dimensionality (Advani and Ganguli, 2016). In classically designed experiments, a small number of carefully selected variables (V) are measured to test a specific hypothesis, with a large number of measurements (M) for each variable. Thus, the measurement density is very large (M/V

→ ∞). Such datasets are referred to as low-dimensional, and much of classical statistics operates within this framework. In contrast to this classical scenario, the recent technological capacity for high-throughput sequencing has led to a different statistical regime. It is commonplace to simultaneously measure many variables (V), such as the abundance of hundreds of taxa at the individual level. However, due to constraints on time or resources, often it is possible only to make a limited number of simultaneous measurements. Thus, while both M and V are large, the measurement density (M/V) is much smaller than in conventional experiments. Such datasets are referred to as high dimensional, that is, they consist of a small number of points in a high-dimensional space. Microbiome datasets consist of the composition of thousands of microbial taxa in an individual gut. Hence, understanding the role of gut microbiome in liver health requires us to use machine learning (ML) approaches to dissect such a high-dimensional dataset. Utilizing these approaches enhances our understanding of the complex host-microbe reciprocal, helping in tracking disease progression over time or monitoring treatment responses, which is valuable for personalized medicine.

In recent years, ML approaches have been widely used to shed light on how gut microbiome impacts various liver diseases. These studies have identified microbiome biomarkers associated with metabolic liver diseases (Ruuskanen et al., 2021; Zhang et al., 2021; Liu et al., 2022; Park et al., 2024). However, not only that majority of these studies are mostly cross-sectional in nature but also these ML approaches suffer from a number of limitations. First, it is extremely difficult to detect patterns in microbiome datasets that can provide useful biological insight with translational value. High-dimensional data analysis techniques such as PCA, ICA, or t-SNE are useful for reducing dimensionality and detecting patterns. However, since the resulting axes represent linear combinations of a large number of features (e.g., taxa abundance), interpreting the analysis or making experimental predictions is often difficult (Donoho and Tanner, 2009; Furchtgott et al., 2017). Identifying patterns becomes even more challenging as the proportion of relevant features decreases at higher taxonomic levels (Donoho and Tanner, 2009). In fact, decades of research related to the gut microbiome have shown that specific combinations of a few microbes can be associated with a disease. This indicates that the composition of a small subset of microbes may be most relevant for making accurate computational inferences. Therefore, there is a need to develop novel methods to detect sparse patterns in high-dimensional datasets.

The second challenge is to extract meaningful models from high-dimensional datasets that can predict interventional outcomes and guide translational implications of research findings. Building complex models describing microbial networks involves hundreds of parameters. Unfortunately, in most cases, the parameters remain completely unknown. As Von Neumann once said: “with four parameters I can fit an elephant, and with five I can make him wiggle his trunk.” The challenges involving microbial networks are clearly exacerbated since even the simplest of models would require hundreds of parameters. Hence, the available data are always limited and

can never constrain the space of models completely. Given the underconstrained nature of the problem, there exist an infinite number of combinations of the parameters that can successfully replicate the data. Hence, making predictions based on such underconstrained gut microbial networks has proven to be challenging. Moreover, most of the studies trained their models on specific datasets, such as Western or Chinese datasets, that may not generalize to different populations, limiting their overall applicability. Often these studies used a small patient population size. Some of these studies suffered from the absence of population characteristics such as the lack of lifestyle information (e.g., diet, socioeconomic status, tobacco, and alcoholism). Furthermore, in some cases, the lack of patient clinical, metagenomic, and metabolomic profiles hampered garnering a comprehensive understanding of the various causal aspects of the disease.

3 Conclusion

The complexity of the gut microbial dynamics and its function in metabolic liver disease remains largely unknown. The progressive character and course of liver disorders are not well captured by pre-clinical and cross-sectional investigations. We suggest using cutting-edge ML methods in conjunction with longitudinal research to address these issues. Hence, it is imperative to first develop methods to detect sparse relevant features that can then be used to find patterns in the data and train predictive models. By addressing these limitations and building new computational approaches, we can fully harness the potential of ML approaches to deepen our understanding of the gut microbiome's role in liver disease and other areas. By doing this, we can pinpoint certain gut microbial patterns associated with the advancement of liver disease, resulting in the development of more potent preventative and therapeutic approaches. For gut-liver axis research to reach its full potential, machine learning must be included.

Author contributions

PD: Conceptualization, Writing – original draft, Writing – review & editing. SC: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors are thankful to the Indian Council of Medical Research (ICMR).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

References

- Advani, M., and Ganguli, S. (2016). Statistical mechanics of optimal convex inference in high dimensions. *Phys. Rev. X* 6:031034. doi: 10.1103/PhysRevX.6.031034
- Ali, R. O., Quinn, G. M., Umarova, R., Haddad, J. A., Zhang, G. Y., Townsend, E. C., et al. (2023). Longitudinal multi-omics analyses of the gut-liver axis reveals metabolic dysregulation in hepatitis C infection and cirrhosis. *Nat. Microbiol.* 8, 12–27. doi: 10.1038/s41564-022-01273-y
- Dey, P. (2020). The role of gut microbiome in chemical-induced metabolic and toxicological murine disease models. *Life Sci.* 258:118172. doi: 10.1016/j.lfs.2020.118172
- Dey, P. (2024). Good girl goes bad: understanding how gut commensals cause disease. *Microb. Pathog.* 190:106617. doi: 10.1016/j.micpath.2024.106617
- Dey, P., and Ray Chaudhuri, S. (2023). The opportunistic nature of gut commensal microbiota. *Crit. Rev. Microbiol.* 49, 739–763. doi: 10.1080/1040841X.2022.2133987
- d’Hennezel, E., Abubucker, S., Murphy, L. O., and Cullen, T. W. (2017). Total lipopolysaccharide from the human gut microbiome silences toll-like receptor Signaling. *mSystems* 2, 10–1128. doi: 10.1128/mSystems.00046-17
- Donoho, D., and Tanner, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. A Math. Phys. Eng. Sci.* 367, 4273–4293. doi: 10.1098/rsta.2009.0152
- Furchtgott, L. A., Melton, S., Menon, V., and Ramanathan, S. (2017). Discovering sparse transcription factor codes for cell states and state transitions during development. *Elife* 6:e20488. doi: 10.7554/eLife.20488
- Gupta, U., and Dey, P. (2023). The oral microbial odyssey influencing chronic metabolic disease. *Arch. Physiol. Biochem.* 2023, 1–17. doi: 10.1080/13813455.2023.2296346
- Jans, M., and Vereecke, L. (2024). A guide to germ-free and gnotobiotic mouse technology to study health and disease. *FEBS J.* 2024:17124. doi: 10.1111/febs.17124
- Kato, K., Nagao, M., Miyamoto, K., Oka, K., Takahashi, M., Yamamoto, M., et al. (2017). Longitudinal analysis of the intestinal microbiota in liver transplantation. *Transpl. Dir.* 3:e144. doi: 10.1097/TXD.0000000000000661
- Liu, Y., Méric, G., Havulinna, A. S., Teo, S. M., Åberg, F., Ruuskanen, M., et al. (2022). Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting. *Cell Metab.* 34, 719–730. doi: 10.1016/j.cmet.2022.03.002
- Liu, Y., Meyer, C., Xu, C., Weng, H., Hellerbrand, C., Ten Dijke, P., et al. (2013). Animal models of chronic liver diseases. *Am. J. Physiol.-Gastroint. Liver Physiol.* 304, G449–G468. doi: 10.1152/ajpgi.00199.2012
- Nguyen, T. L., Vieira-Silva, S., Liston, A., and Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Dis. Model. Mech.* 8, 1–16. doi: 10.1242/dmm.017400
- Park, I.-G., Yoon, S. J., Won, S.-M., Oh, K.-K., Hyun, J. Y., Suk, K. T., et al. (2024). Gut microbiota-based machine-learning signature for the diagnosis of alcohol-associated and metabolic dysfunction-associated steatotic liver disease. *Sci. Rep.* 14:16122. doi: 10.1038/s41598-024-60768-2
- Picarello, D. M. (2022). Investigating how the LPS structure affects the immune response of gram-negative bacterial outer membranes. *Biophys. J.* 121:487a. doi: 10.1016/j.bpj.2021.11.344
- Qi, Y., Yu, L., Tian, F., Zhao, J., and Zhai, Q. (2023). In vitro models to study human gut-microbiota interactions: Applications, advances, and limitations. *Microbiol. Res.* 270:127336. doi: 10.1016/j.micres.2023.127336
- Ruuskanen, M. O., Åberg, F., Männistö, V., Havulinna, A. S., Méric, G., Liu, Y., et al. (2021). Links between gut microbiome composition and fatty liver disease in a large population sample. *Gut Microbes* 13:1888673. doi: 10.1080/19490976.2021.1888673
- Su, X., Chen, S., Liu, J., Feng, Y., Han, E., Hao, X., et al. (2024). Composition of gut microbiota and non-alcoholic fatty liver disease: a systematic review and meta-analysis. *Obes. Rev.* 25:e13646. doi: 10.1111/obr.13646
- Zhang, T., Zhang, S., Jin, C., Lin, Z., Deng, T., Xie, X., et al. (2021). A predictive model based on the gut microbiota improves the diagnostic effect in patients with cholangiocarcinoma. *Front. Cell. Infect. Microbiol.* 11:751795. doi: 10.3389/fcimb.2021.751795
- Zhou, X., Shen, X., Johnson, J. S., Spakowicz, D. J., Agnello, M., Zhou, W., et al. (2024). Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. *Cell Host Microbe.* 32, 506–526. doi: 10.1016/j.chom.2024.02.012

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.