eGastroenterology

Blood microbial DNA signature differentiates hepatocellular carcinoma from metastatic lesions

Caitlin Guccione,[1,2,3] Ana Carolina Dantas Machado,[4] Fady Youssef ,[4,5] Isabella Angeli-Pahim,[6,7] Sergio Duarte,[6,7] Curtis Warren,[6,7] Sawyer Farmer,[3] Gregory Humphrey,[3] Roland Alexander Richter,[4] Daniel McDonald,[3] Yuhan Weng,[2,3] Adam Burgoyne,[8,9] Rohit Loomba,[4,8,10,11] Kit Curtius ,[1,8,12] Ali Zarrinpar ,[6,7,13] Rob Knight,[3,14,15,16,17] Amir Zarrinpar [4,5,8,14,15]

To cite: Guccione C, Dantas Machado AC, Youssef F, *et al*. Blood microbial DNA signature differentiates hepatocellular carcinoma from metastatic lesions. *eGastroenterology* 2025;**3**:e100193. doi:10.1136/egastro-2025-100193

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/egastro-2025-100193).

CG and ACDM contributed equally.

Received 06 February 2025
Accepted 11 July 2025

For numbered affiliations see end of article.

Correspondence to
Dr Amir Zarrinpar;
azarrinp@ucsd.edu

The microbiome, particularly the composition and diversity of microbial communities within various tissues, influences cancer development, progression and therapy response across various malignancies.[1–5] Understanding specific microbial signatures associated with cancers, termed the oncobiome, is crucial for advancing diagnostic and therapeutic strategies.

Analysis of microbial reads from whole-genome sequencing data of blood and tissue samples from The Cancer Genome Atlas (TCGA) identified microbial communities distinguishing hepatocellular carcinoma (HCC) from non-cancerous tissue and other cancers.[2] These findings are robust across multiple rigorous analysis pipelines. Additionally, circulating microbiome DNA has shown promise as a biomarker for early lung cancer diagnosis and recurrence, suggesting the viability of blood-based microbial detection platforms.[2] Furthermore, recent research on metastatic tissues indicated microbial communities align more closely with the metastatic site rather than the primary cancer origin, though this was not evaluated in blood.[1]

However, significant challenges remain for microbial cell-free DNA (cfDNA) to effectively distinguish primary from secondary liver tumours. Previous studies have not demonstrated if microbial cfDNA differentiates between HCC and metastatic liver tumours. Additionally, hepatitis B virus (HBV) DNA contamination limited the ability to clearly separate HCC from non-HCC samples. Thus, it remains uncertain whether liver cancers share a common microbial cfDNA profile or if primary liver cancers possess a distinct signature. In this study, we apply shotgun metagenomics to cfDNA to identify unique blood-based bacterial signatures that differentiate HCC from metastatic colorectal cancer (mCRC).

We assessed the metagenomic profile of tumour, adjacent tissue and plasma samples from participants diagnosed with HCC (n=16) or mCRC (n=11) (online supplemental table 1). There were no significant differences in age, sex, race or ethnicity between the participants in the two groups. 88% of patients with HCC had underlying advanced fibrosis/liver cirrhosis, while none of the patients with mCRC had cirrhosis. None of the patients in the study cohort had an underlying HBV infection or exposure. Only one patient with HCC had alpha-fetoprotein (AFP) levels that would warrant concern for HCC (ie, AFP>400 ng/mL), and most patients with HCC had AFP below the threshold of surveillance (ie, AFP<20 ng/mL).

We performed shotgun metagenomics on cfDNA from plasma and paired liver tissues using established protocols,[6] followed by stringent computational preprocessing to remove host-derived reads, laboratory-associated contamination and potential misclassification (figure 1a; see online supplemental methods). We observed a clear difference in plasma microbial cfDNA between the two groups on the Robust Principal Coordinates Analysis (RPCA)-Principal Coordinates Analysis (PCoA) biplot (figure 1b) (RPCA-Permutational Multivariate Analysis of Variance (PERMANOVA), p=0.015). However, despite many of the patients with HCC having liver disease and/or cirrhosis, we did not observe a difference based on tissue microbial signatures between the patient groups on the RPCA-PCoA biplot (online supplemental figure S1a,b) (RPCA-PERMANOVA, p=0.624

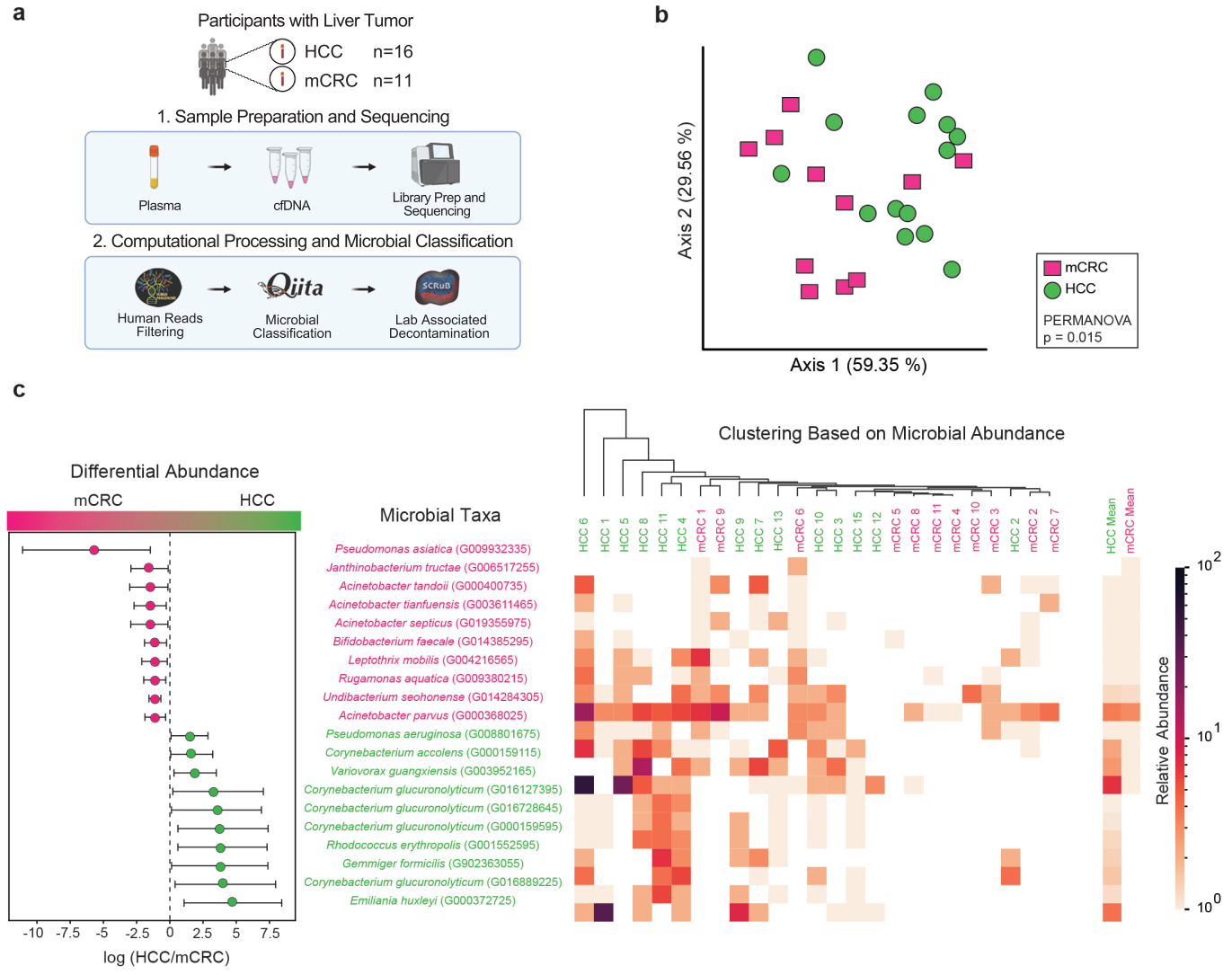

Figure 1 Plasma microbial signatures of patients with liver tumours. (a) Schematic representation of study design, experimental and computational processing including host filtration using the Human Pangenome Reference Consortium, microbial classification using Qiita and laboratory-associated decontamination with SCRub. (b) Differences in community structure represented by the RPCA-PCoA between the plasma of participants with primary liver tumour (HCC) or with metastatic liver tumour (mCRC) (RPCA-PERMANOVA, p=0.015). (c) Top and bottom 10 microbial taxa associated with primary liver tumour (HCC, green) or metastatic liver tumour (mCRC, pink).The left panel represents the differential abundance rankings from BIRDMAn. The right panel represents the relative abundances for each participant and average for each condition. BIRDMAn, Bayesian Inferential Regression for Differential Microbiome Analysis; cfDNA, cell-free DNA; HCC, hepatocellular carcinoma; mCRC, metastatic colorectal cancer; PCoA, Principal Coordinates Analysis; PERMANOVA, Permutational Multivariate Analysis of Variance; RPCA, Robust Principal Coordinates Analysis.

for adjacent tissue, p=0.735 for tumour). Using Bayesian Inferential Regression for Differential Microbiome Analysis (BIRDMAn), we identified the top 10 microbial taxa differentiating HCC and mCRC (figure 1c). Microbes predominantly found in patients with mCRC include *Acinetobacter (A.) tandoii*, *A. tianfuensis*, *A. septicus*, *A. parvus*, *Pseudomonas asiatica* and *Bifidobacterium faecale*. *Pseudomonas aeruginosa* is more commonly found in the plasma of patients with HCC. Another genus more dominant in patients with HCC was *Corynebacterium (C.)*, which includes the *C. accolens* and *C. glucuronolyticum* species.

Our analysis identified bacterial taxa significantly differentiating mCRC from HCC, notably highlighting several species with known clinical relevance. The genus *Acinetobacter*, including species such as *A. septicus*, *A. parvus* and *A. lwoffii*, is often associated with nosocomial infections, bloodstream infections and gastrointestinal inflammation, suggesting possible clinical implications in mCRC patients. Additionally, *Pseudomonas asiatica*, identified predominantly in mCRC, has recently emerged as a potential pathogen in hospitalised patients, whereas *Bifidobacterium faecale* is related to antitumour immunity

in preclinical models. Conversely, the HCC group showed enrichment for *P. aeruginosa*, a pathogen associated with immunocompromised patients and liver transplantation complications, but intriguingly also implicated in tumour-suppressive pathways. Similarly, *C. accolens* and *C. glucuronolyticum*, also enriched in HCC, have recognised roles in antimicrobial defence and genitourinary infections, respectively. These findings suggest diverse functional roles of microbial species that may influence tumour progression or reflect the underlying host environment, underscoring the need for further exploration of their origin, functional impact and clinical relevance in liver malignancies.

Our study demonstrates that bacterial cfDNA in the blood distinguishes patients with HCC from those with mCRC. This finding supports prior work showing that microbial DNA can differentiate between tumour types and provides new evidence that it may distinguish primary hepatic malignancies from liver metastases. A key strength of our study is that the microbial signatures differentiating HCC from mCRC were not driven by viral hepatitis DNA, which limited prior studies involving large hepatitis B cohorts.[2] In contrast to the earlier TCGA study,[2] we did not observe distinct microbial signatures between tumour and adjacent tissue within groups or between HCC and mCRC tissue samples. One possible explanation is that plasma may carry a stronger microbial signal—particularly in patients with mCRC, who may shed microbes from both colorectal and liver tumours—compared with patients with HCC, whose signal likely derives only from liver tumours. Although plasma microbial DNA may originate from multiple sources (eg, gut, oral cavity or liver), the ability of blood-based profiles to distinguish HCC from mCRC—despite both being hepatic tumours—suggests the presence of a tumour-specific microbial signature that merits further investigation.

Several factors may explain the stronger microbial signal observed in mCRC. Disruption of the gut vascular barrier can allow bacterial translocation from colorectal tumours to the liver, promoting metastasis and potentially amplifying microbial cfDNA in circulation.[7] Additionally, more advanced cancer stages often exhibit higher levels of cfDNA,[8] which may also apply to microbial cfDNA and contribute to the clearer signal in patients with mCRC. The lack of tissue-level microbial differences between groups may also reflect limited sample size, as seen in the TCGA study, or methodological factors—such as our use of blood-specific protocols optimised for microbial read recovery. Together, these findings suggest that blood-derived microbial cfDNA may outperform tissue-based microbial profiling for detecting and distinguishing malignancies.

Our pilot study has several limitations. First, the cohort consists of individuals with resectable tumours, and it is unclear whether bacterial cfDNA is present in sufficient quantities to enable screening or surveillance in broader clinical populations. Second, the impact of prior treatments—such as locoregional or systemic therapy—on the tumour microbiome and corresponding plasma cfDNA remains unknown. Third, the depth of metagenomic sequencing required for detecting microbial cfDNA as a diagnostic marker remains costly. However, as sequencing costs decrease or more specific microbial signals are identified, this approach may become clinically viable. Fourth, the small sample size may limit statistical power and introduce bias. Our design was informed by prior TCGA-based work showing that blood-derived microbial DNA could distinguish HCC from other cancers with an Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.99.[2]

While promising, these findings require validation in larger, independent cohorts to confirm generalisability, particularly in surveillance populations. Future studies should also investigate the origin of microbial cfDNA by characterising faecal, rectal or oral microbiomes alongside blood. Lastly, because our study compared patients with HCC (mostly with cirrhosis) to patients with mCRC (without liver disease), the observed differences could reflect underlying liver pathology. However, similar microbial patterns in adjacent tissues across both groups argue against this. To clarify, future studies should examine patients with cirrhosis or advanced metabolic dysfunction-associated steatotic liver disease (MASLD) prior to HCC development and compare them to those with HCC at earlier disease stages. This will help determine whether circulating microbial signatures reflect liver disease, tumour presence or both.

Although we do not present a specific bacterial DNA biomarker, predictive model or diagnostic test, our study lays groundwork for understanding the importance of microbial cfDNA profiles in blood. These profiles hold potential clinical applications, such as early cancer detection or monitoring high-risk patients, which can be validated in independent cohorts. Overall, these findings indicate a unique, disease-associated microbiota evident in blood, potentially serving as biomarkers for cancer diagnosis, prognosis, development, progression and therapy response. Furthermore, the distinct microbial signatures may offer insights into novel, microbiome-based therapeutic strategies to manipulate the tumour microenvironment.

Author affiliations
[1]Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, California, USA
[2]Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, California, USA
[3]Department of Pediatrics, University of California San Diego, La Jolla, California, USA
[4]Division of Gastroenterology, University of California San Diego, La Jolla, California, USA
[5]Division of Gastroenterology, Jennifer Moreno Department of Veterans Affairs Medical Center, La Jolla, California, USA
[6]Department of Surgery, University of Florida, Gainesville, Florida, USA
[7]Department of Biochemistry and Molecular Biology, University of Florida, Gainesville, Florida, USA
[8]Moores Cancer Center, University of California San Diego, La Jolla, California, USA

[9]Division of Hematology-Oncology, University of California San Diego, La Jolla, California, USA
[10]NAFLD Research Center, University of California San Diego, La Jolla, California, USA
[11]Division of Epidemiology, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA
[12]VA San Diego Healthcare System, La Jolla, California, USA
[13]J. Crayton Pruitt Family Department of Biomedical Engineering, Herbert Wertheim College of Engineering, University of Florida, Gainesville, Florida, USA
[14]Center for Microbiome Innovation, University of California San Diego, La Jolla, California, USA
[15]Shu Chien-Gene Lay Department of Bioengineering, University of California San Diego, La Jolla, California, USA
[16]Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, USA
[17]Halicioğlu Data Science Institute, University of California San Diego, La Jolla, California, USA

X Amir Zarrinpar @zarrinparlab

Acknowledgements We thank Gail Ackermann for compiling the metadata. Figure 1a was created in BioRender. Curtius, K. (2025) https://BioRender.com/o80z834.

Contributors Conceptualisation: CG, ACDM, AmZ. Methodology, investigation and validation: CG, ACDM, IA-P, SD, CW, SF, GH, DM, YW. Formal analysis: CG, ACDM, RAR. Resources: AIZ, AmZ. Data curation: CG, ACDM, RAR, IA-P, SD, CW. Writing—original draft preparation: CG, ACDM, FY. Writing—reviewing and editing: AB, RL, KC, RK, AIZ. Visualisation: CG, ACDM, AmZ. Supervision: KC, RK, AmZ. Project administration: AmZ. Funding acquisition: AmZ. We used ChatGPT to enhance the clarity of certain sections of our manuscript by refining language and improving sentence structure as well as reducing manuscript length. Importantly, ChatGPT was not employed to generate any new scientific content, data or ideas. All substantive contributions, including the study design, data analysis and scientific conclusions, were made by the authors. ChatGPT's role was limited to language editing to ensure clear and effective communication of our findings.

Funding ACDM is supported by R01 HL148801-02S1. AmZ is supported by the Prevent Cancer Foundation Grant, and NIH grants R21 CA259921, R01 AI163483, R01 EB030134 and U01 CA265719. This material is based upon work supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development (VA Merit BLR&D Award I01 BX005707 to AmZ). RL receives funding support from NIH grants U01 DK061734, U01 DK130190, R01 DK106419, R01 DK121378, R01 DK124318 and P01 HL147835, as well as the John C Martin Foundation (RP124). KC is supported by AGA Research Foundation (AGA Research Scholar Award AGA2022-13-05). KC and CG are supported by NIH P30 CA023100 and R01 CA270235. AIZ has received funding support related to this study from NIH UH2 TR002087 and UH3 TR002087. RK is supported by NIH U24 CA248454, Pioneer DP1 AT010885 and R01 CA241728. All authors receive institutional support from NIH P30 DK120515, P30 DK063491, P30 CA014195, P50 AA011999 and UL1 TR001442. The funders had no role in the study design; in the collection, analysis or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. This work includes contributions made as part of official duties by a US Department of Veterans Affairs employee and, under 17 U.S.C. § 105, is considered a work of the US Government and not subject to copyright protection in the USA.

Competing interests AmZ is a co-founder, acting chief medical officer, and equity-holder in Endure Biotherapeutics. RK is a scientific advisory board member, and consultant for BiomeSense, Inc., has equity and receives income. He is a scientific advisory board member and has equity in GenCirq. He is a consultant and scientific advisory board member for DayTwo, and receives income. He has equity in and acts as a consultant for Cybele. He is a co-founder of Biota, Inc., and has equity. He is a co-founder of Micronoma, and has equity and is a scientific advisory board member. DM is a consultant for BiomeSense, Inc. and has equity. KC has research grant support from Phantom Pharmaceuticals. RL serves as a consultant to Aardvark Therapeutics, Altimmune, Arrowhead Pharmaceuticals, AstraZeneca, Cascade Pharmaceuticals, Eli Lilly, Gilead, Glympse bio, Inipharma, Intercept, Inventiva, Ionis, Janssen Inc., Lipidio, Madrigal, Neurobo, Novo Nordisk, Merck, Pfizer, Sagimet, 89 bio, Takeda, Terns Pharmaceuticals and Viking Therapeutics. RL has stock options in Sagimet biosciences. In addition, his institution received research grants from Arrowhead Pharmaceuticals, AstraZeneca, Boehringer-Ingelheim, Bristol-Myers Squibb, Eli Lilly, Galectin Therapeutics, Gilead, Intercept, Hanmi, Intercept, Inventiva, Ionis, Janssen, Madrigal Pharmaceuticals, Merck, Novo Nordisk, Pfizer, Sonic Incytes and Terns Pharmaceuticals. Co-founder of LipoNexus Inc. The terms of all these arrangements have been reviewed and approved by the University of California, San Diego, California, USA, in accordance with its conflict of interest policies.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study, which is a secondary analysis of biobanked samples, was reviewed and exempted by the Institutional Review Board at the University of California, San Diego. Specimens were derived from the University of Florida Transplantation and Hepatobiliary Surgery Tissue Biorepository (bank UF IRB #201700650). The original study that collected the biobanked samples was reviewed and approved by the Institutional Review Board at the University of Florida (IRB201802364).

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Fady Youssef http://orcid.org/0000-0003-1020-0869
Kit Curtius http://orcid.org/0000-0002-2678-0960
Ali Zarrinpar http://orcid.org/0000-0002-6812-1816
Amir Zarrinpar http://orcid.org/0000-0001-6423-5982

REFERENCES

1 Battaglia TW, Mimpen IL, Traets JJH, *et al*. A pan-cancer analysis of the microbiome in metastatic cancer. *Cell* 2024;187:2324–35.
2 Sepich-Poore GD, McDonald D, Kopylova E, *et al*. Robustness of cancer microbiome signals over a broad range of methodological variation. *Oncogene* 2024;43:1127–48.
3 Sepich-Poore GD, Zitvogel L, Straussman R, *et al*. The microbiome and human cancer. *Science* 2021;371:eabc4552.
4 Dohlman AB, Klug J, Mesko M, *et al*. A pan-cancer mycobiome analysis reveals fungal involvement in gastrointestinal and lung tumors. *Cell* 2022;185:3807–22.
5 Nejman D, Livyatan I, Fuks G, *et al*. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. *Science* 2020;368:973–80.
6 Minich JJ, Zhu Q, Janssen S, *et al*. KatharoSeq Enables High-Throughput Microbiome Analysis from Low-Biomass Samples. *mSystems* 2018;3.
7 Bertocchi A, Carloni S, Ravenda PS, *et al*. Gut vascular barrier impairment leads to intestinal bacteria dissemination and colorectal cancer metastasis to liver. *Cancer Cell* 2021;39:708–24.
8 Gao Q, Zeng Q, Wang Z, *et al*. Circulating cell-free DNA for cancer early detection. *Innovation (Camb)* 2022;3:100259.