



Host-Microbe Metagenomics: a Lens To Refocus Our Perspective on Infectious and Inflammatory Diseases

Katrina L. Kalantar,^a  Charles R. Langelier^{b,c}

^aChan Zuckerberg Initiative, San Francisco, California, USA

^bDivision of Infectious Diseases, University of California San Francisco, San Francisco, California, USA

^cChan Zuckerberg Biohub, San Francisco, California, USA

ABSTRACT A dynamic relationship involving pathogen, host immune response, and microbiome characterizes the biological framework of many infectious and inflammatory diseases. Combined host/microbe metagenomics (mNGS) enables simultaneous assessment of all three features, enabling the study and diagnosis of diverse infectious and inflammatory processes ranging from pneumonia to sepsis to inflammatory diseases such as rheumatoid arthritis. Host/microbe mNGS holds promise for new mechanistic insights, diagnostic approaches, and precision medicine interventions.

Infections represent a dynamic relationship between pathogen, host immune response, and local microbiome. Despite this, each aspect has historically been evaluated individually—through studies focused on clinical microbiology (1), immunology (2), or the bacterial microbiome (3)—leaving gaps in understanding of pathogenesis mechanisms and providing a narrow perspective on the systems biology of a given disease. Combined host/microbe metagenomics (mNGS) addresses this gap by enabling broad profiling of all three key components of infectious diseases. Its utility spans across diverse disease profiles, ranging from lower respiratory tract infections (LRTI) to sepsis to gastrointestinal disease to meningitis, where the goal of clinical diagnostics is ultimately to determine which patients have an infection and what is causing the infection. This holds potential to inform precision treatment that can reduce the need for broad-spectrum antibiotics.

BACKGROUND ON mNGS

mNGS enables culture-independent, unbiased profiling of microbial and host-derived nucleic acid directly from biological samples. DNA-based mNGS has been used for over a decade to detect bacterial pathogens and study the human microbiome (3, 4). Transcriptome sequencing (RNA-seq) expands this utility by also enabling detection of RNA viruses (5) while still permitting detection of other pathogens (6).

As opposed to stool specimens used to study the gut microbiome, nucleic acid obtained from other anatomical sites, such as the respiratory tract, is derived mostly from host and not from microbial origin, even in the setting of active infection (5). Thus, both DNA sequencing (DNA-seq) and RNA-seq can be used to identify heritable single nucleotide polymorphisms (SNPs) associated with disease (7). A unique attribute of RNA-seq, however, is the capacity to detect variations in host gene expression levels, thereby enabling combined host and microbial profiling (5).

APPLICATION OF HOST/MICROBE mNGS TO COVID-19 AND OTHER LRTI

The challenges associated with determining the cause of LRTI are particularly well suited for the application of mNGS. Even before the coronavirus disease 2019 (COVID-19) pandemic, LRTI were the leading cause of death from infectious disease globally, and a top cause of death by any cause in low and middle income countries (8).

Citation Kalantar KL, Langelier CR. 2021. Host-microbe metagenomics: a lens to refocus our perspective on infectious and inflammatory diseases. *mSystems* 6:e00404-21. <https://doi.org/10.1128/mSystems.00404-21>.

Copyright © 2021 Kalantar and Langelier. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Charles R. Langelier, chaz.langelier@ucsf.edu.

Conflict of Interest Disclosures: K.L.K. has nothing to disclose. C.R.L. has nothing to disclose.

The views expressed in this article do not necessarily reflect the views of the journal or of ASM.

This article is part of a special series sponsored by Floré.

Published 17 August 2021

Difficulties identifying, and thus appropriately treating, the etiologic microbes contributes significantly to the burden of LRTI, as the responsible pathogens frequently remain undetected due to the limitations of existing clinical diagnostics, which rely heavily on bacterial culture. Further compounding this issue are challenges in accurately distinguishing LRTI from other inflammatory lung diseases, as many conditions mimic infection, especially in critically ill patients with systemic illness.

In the first proof-of-concept demonstration of host/microbe mNGS for diagnosis of LRTI (5), we identified host transcriptional signatures and features of the lung microbiome that enabled accurate identification of infection and differentiation from noninfectious acute respiratory diseases. In a complex population of critically ill patients requiring mechanical ventilation, traditional clinical metrics such as fever and white blood cell count were unable to distinguish LRTI from other causes of respiratory failure at the time of hospital admission, while in contrast, host and microbial mNGS classifiers could do so with high accuracy (2).

This improved accuracy for identification of infection is driven by the fact that host and microbial data provide complementary information that can assist with contextualization that is critical to the ability to disambiguate cases of infection. Specifically, it is possible to interpret the set of microbes identified in a particular case within the context of host biology and vice versa to illuminate interactions that may be indicative of infection. Incidental carriage of potentially pathogenic microbes in the respiratory tract is common, and thus differentiating pathogens from commensals is a major challenge for existing microbiologic diagnostics. Profiling the host response provides context to microbes identified, informing the presence of infection versus just incidental pathogen colonization (5).

Complementing this host-based pathogen contextualization is the ability to assess lung microbiome composition, which becomes disrupted during LRTI, with loss of diversity and pathogen dominance correlating with infection (5, 9). Using a machine learning model trained on metagenomic data from patients with either proven LRTI or noninfectious respiratory illnesses, we found that mNGS could accurately identify pathogens—including viruses, bacteria, and fungi—amid a background of more abundant predominantly bacterial commensals. The importance of this contextualization is paramount given the increasing use of highly sensitive molecular assays for clinical infectious disease diagnosis.

Emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has highlighted the value of an unbiased diagnostic assay capable of detecting novel, emerging pathogens, as indeed it was mNGS that illuminated the etiology of an idiopathic pneumonia syndrome in Wuhan, China, in December 2019 (10). Further work has leveraged mNGS to not only identify coinfecting pathogens and track the emergence of new variants but has also illuminated unique aspects of the host immunologic response to SARS-CoV-2 that differ markedly from other viral pathogens (11, 12).

For instance, recent work by our group and others used host/microbe mNGS to demonstrate that early SARS-CoV-2 infection is characterized by attenuated innate immune signaling in the upper respiratory tract (11). One of the defining features of COVID-19 is peak viral load—and infectiousness—at or prior to the onset of symptoms, a feature that has made control of coronavirus transmission extremely difficult and likely contributed to the emergence of a pandemic far beyond the scale of what was experienced with the original SARS virus in 2003. Suppressed immune signaling provides a potential explanation for how peak viral load could exist in the absence of symptoms.

APPLICATION OF HOST/MICROBE mNGS TO OTHER INFECTIOUS AND INFLAMMATORY DISEASES

The potential of host/microbe mNGS to advance understanding of respiratory disease biology extends far beyond LRTI. Compelling work demonstrates that both dysregulated host response and disrupted microbial components characterize the acute respiratory distress syndrome (ARDS) (3), chronic obstructive pulmonary disease, asthma,

and interstitial lung disease. To date, immunologic and microbial aspects of these conditions have largely been examined independently. Host/microbe mNGS provides an opportunity to more holistically study and diagnose these conditions and to understand clinical subphenotypes (13) with unique clinical features or treatment responses.

While the respiratory tract is particularly well suited for host-microbe mNGS-based assessment, the utility of this approach without doubt extends to other systems and disease states. Sepsis, for instance, is a disease defined as a dysregulated host response to infection (14). Despite this, nearly all available sepsis diagnostics are designed to detect microbes, although recent advancements have led to the development of accurate host-based transcriptomic classifiers (15). No diagnostic yet, however, has integrated unbiased pathogen detection with host response profiling for sepsis diagnosis.

Meningitis is well studied in the context of mNGS pathogen diagnostics (16, 17), which have proven exceptional in terms of detecting rare and uncommon microbial explanations for persistent central nervous system disease. Infectious etiologies of meningitis and encephalitis, however, can often be clinically undifferentiable from autoimmune neurologic disease, despite increasing knowledge of differing cell types (and therefore host gene expression patterns) across these etiologies. A clear opportunity exists for host-microbe mNGS to advance our ability to diagnose neuroinflammatory disease.

Recent work has demonstrated that the human gut microbiome may play a role in modulating the onset and perpetuation of pathological inflammatory signaling in chronic autoimmune diseases, such as rheumatoid arthritis (18, 19). Combining assessment of systemic and local host immune responses with unbiased profiling of microbial taxa, including bacteria, viruses, and fungi, holds promise for characterizing this relationship at a higher resolution. Without question, opportunities exist to leverage host/microbe mNGS to further understand the systems biology of autoimmune diseases, identify unique patient subphenotypes, and predict clinical outcomes.

DEMOCRATIZATION OF HOST/MICROBE mNGS COMPUTATIONAL PIPELINES

Despite the potential for host/microbe mNGS to advance our understanding of infectious and inflammatory diseases, use has been limited by both the cost of sequencing and the computational infrastructure needed for computationally intensive bioinformatics analysis. Advancements in cloud-based open source pipelines such as the IDseq pipeline (20) have democratized access to, and analysis of, mNGS data. Now anyone in the world with Internet connectivity can access the expanding repositories of publicly accessible host and microbial sequencing data to study their disease of interest. As the cost of sequencing continues to decline precipitously, the utility of culture-independent host/microbe mNGS will continue to grow in resource-limited areas, potentially enabling a leap beyond traditional microbiology laboratories and their requisite infrastructure of incubators, culture media, sterilization instruments, and microscopes.

FUTURE OUTLOOK

The COVID-19 pandemic has brought infectious disease systems biology to the forefront and has highlighted key roles for microbial metagenomics and host transcriptional profiling in both epidemiologic surveillance and in the study of emerging infections. When the pandemic subsides, we will have a renewed opportunity to refocus our traditional perspectives on other major infectious and inflammatory diseases—from pneumonia to sepsis to rheumatoid arthritis—through the lens of host/microbe metagenomics. This approach holds promise for new mechanistic insights, diagnostic approaches, and precision medicine interventions.

REFERENCES

1. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, Reed C, Grijalva CG, Anderson EJ, Courtney DM, Chappell JD, Qi C, Hart EM, Carroll F, Trabue C, Donnelly HK, Williams DJ, Zhu Y, Arnold SR, Ampofo K, Waterer GW, Levine M, Lindstrom S, Winchell JM, Katz JM, Erdman D, Schneider E, Hicks LA, McCullers JA, Pavia AT, Edwards KM, Finelli L, CDC EPIC Study Team. 2015. Community-acquired pneumonia requiring hospitalization among U.S. adults. *N Engl J Med* 373:415–427. <https://doi.org/10.1056/NEJMoa1500245>.
2. Sweeney TE, Wong HR, Khatri P. 2016. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med* 8:346ra91. <https://doi.org/10.1126/scitranslmed.aaf7165>.

3. Dickson RP, Singer BH, Newstead MW, Falkowski NR, Erb-Downward JR, Standiford TJ, Huffnagle GB. 2016. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol* 1:16113. <https://doi.org/10.1038/nmicrobiol.2016.113>.
4. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. 2013. Metagenomics for pathogen detection in public health. *Genome Med* 5:81. <https://doi.org/10.1186/gm485>.
5. Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, Belzer A, Bolourchi S, Caldera S, Fung M, Jauregui A, Malcolm K, Lyden A, Khan L, Vessel K, Quan J, Zinter M, Chiu CY, Chow ED, Wilson J, Miller S, Matthay MA, Pollard KS, Christenson S, Calfee CS, DeRisi JL. 2018. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* 115:E12353–E12362. <https://doi.org/10.1073/pnas.1809700115>.
6. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY. 2014. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 370:2408–2417. <https://doi.org/10.1056/NEJMoa1401268>.
7. Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, Monnin D, Filali AE, Carareto CM, Vieira C, Picard F, Kremer N, Vavre F, Sagot M-F, Lacroix V. 2016. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Res* 44:e148. <https://doi.org/10.1093/nar/gkw655>.
8. World Health Organization. 2020. The top 10 causes of death. World Health Organization, Geneva, Switzerland.
9. Kalantar KL, Moazed F, Christenson SC, Wilson J, Deiss T, Belzer A, Vessel K, Caldera S, Jauregui A, Bolourchi S, DeRisi JL, Calfee CS, Langelier C. 2019. Metagenomic comparison of tracheal aspirate and mini-bronchial alveolar lavage for assessment of respiratory microbiota. *Am J Physiol Lung Cell Mol Physiol* 316:L578–L584. <https://doi.org/10.1152/ajplung.00476.2018>.
10. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, Chen H-D, Chen J, Luo Y, Guo H, Jiang R-D, Liu M-Q, Chen Y, Shen X-R, Wang X, Zheng X-S, Zhao K, Chen Q-J, Deng F, Liu L-L, Yan B, Zhan F-X, Wang Y-Y, Xiao G-F, Shi Z-L. 2020. Addendum: a pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 588:E6. <https://doi.org/10.1038/s41586-020-2951-z>.
11. Mick E, Kamm J, Pisco AO, Ratnasiri K, Babik JM, Castañeda G, DeRisi JL, Detweiler AM, Hao SL, Kangelaris KN, Kumar GR, Li LM, Mann SA, Neff N, Prasad PA, Serpa PH, Shah SJ, Spottiswoode N, Tan M, Calfee CS, Christenson SA, Kistler A, Langelier C. 2020. Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nat Commun* 11:5854. <https://doi.org/10.1038/s41467-020-19587-y>.
12. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Möller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA, tenOever BR. 2020. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 181:1036–1045.e9. <https://doi.org/10.1016/j.cell.2020.04.026>.
13. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, NHLBI ARDS Network. 2014. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2:611–620. [https://doi.org/10.1016/S2213-2600\(14\)70097-9](https://doi.org/10.1016/S2213-2600(14)70097-9).
14. Ranzani OT, Prina E, Menéndez R, Ceccato A, Cilloniz C, Méndez R, Gabarrus A, Barbeta E, Bassi GL, Ferrer M, Torres A. 2017. New sepsis definition (sepsis-3) and community-acquired pneumonia mortality. A validation and clinical decision-making study. *Am J Respir Crit Care Med* 196:1287–1297. <https://doi.org/10.1164/rccm.201611-2262OC>.
15. Sweeney TE, Perumal TM, Henao R, Nichols M, Howrylak JA, Choi AM, Bermejo-Martin JF, Almansa R, Tamayo E, Davenport EE, Burnham KL, Hinds CJ, Knight JC, Woods CW, Kingsmore SF, Ginsburg GS, Wong HR, Parnell GP, Tang B, Moldawer LL, Moore FE, Omberg L, Khatri P, Tsalik EL, Mangravite LM, Langley RJ. 2018. A community approach to mortality prediction in sepsis via gene expression analysis. *Nat Commun* 9:694. <https://doi.org/10.1038/s41467-018-03078-2>.
16. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, Shah MP, Richie MB, Gorman MP, Hajj-Ali RA, Calabrese LH, Zorn KC, Chow ED, Greenlee JE, Blum JH, Green G, Khan LM, Banerji D, Langelier C, Bryson-Cahn C, Harrington W, Lingappa JR, Shanbhag NM, Green AJ, Brew BJ, Soldatos A, Strnad L, Doernberg SB, Jay CA, Douglas V, Josephson SA, DeRisi JL. 2018. Chronic meningitis investigated via metagenomic next-generation sequencing. *JAMA Neurol* 75:947–955. <https://doi.org/10.1001/jamaneurol.2018.0463>.
17. Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, Federman S, Stryke D, Briggs B, Langelier C, Berger A, Douglas V, Josephson SA, Chow FC, Fulton BD, DeRisi JL, Gelfand JM, Naccache SN, Bender J, Dien Bard J, Murkey J, Carlson M, Vespa PM, Vijayan T, Allyn PR, Campeau S, Humphries RM, Klausner JD, Ganzon CD, Memar F, Ocampo NA, Zimmermann LL, Cohen SH, Polage CR, DeBiasi RL, Haller B, Dallas R, Maron G, Hayden R, Messacar K, Dominguez SR, Miller S, Chiu CY. 2019. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N Engl J Med* 380:2327–2340. <https://doi.org/10.1056/NEJMoa1803396>.
18. Zaiss MM, Wu H-JJ, Mauro D, Schett G, Ciccio F. 2021. The gut–joint axis in rheumatoid arthritis. *Nat Rev Rheumatol* 17:224–237. <https://doi.org/10.1038/s41584-021-00585-3>.
19. De Luca F, Shoenfeld Y. 2019. The microbiome in autoimmune diseases. *Clin Exp Immunol* 195:74–85. <https://doi.org/10.1111/cei.13158>.
20. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J, Holmes OB, Juan Y-F, King R, Kislyuk A, Lin MF, Mariano M, Morse T, Reynoso LV, Cruz DR, Sheu J, Tang J, Wang J, Zhang MA, Zhong E, Ah Yong V, Lay S, Chea S, Bohl JA, Manning JE, Tato CM, DeRisi JL. 2020. IDseq—an open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* 9:giaa111. <https://doi.org/10.1093/gigascience/giaa111>.