Microbiome and Metagenomics in Cancer

From Koch’s postulates to pan-cancer metagenomics

Sean Davis, MD, PhD

University of Colorado Anschutz School of Medicine

July 22, 2026

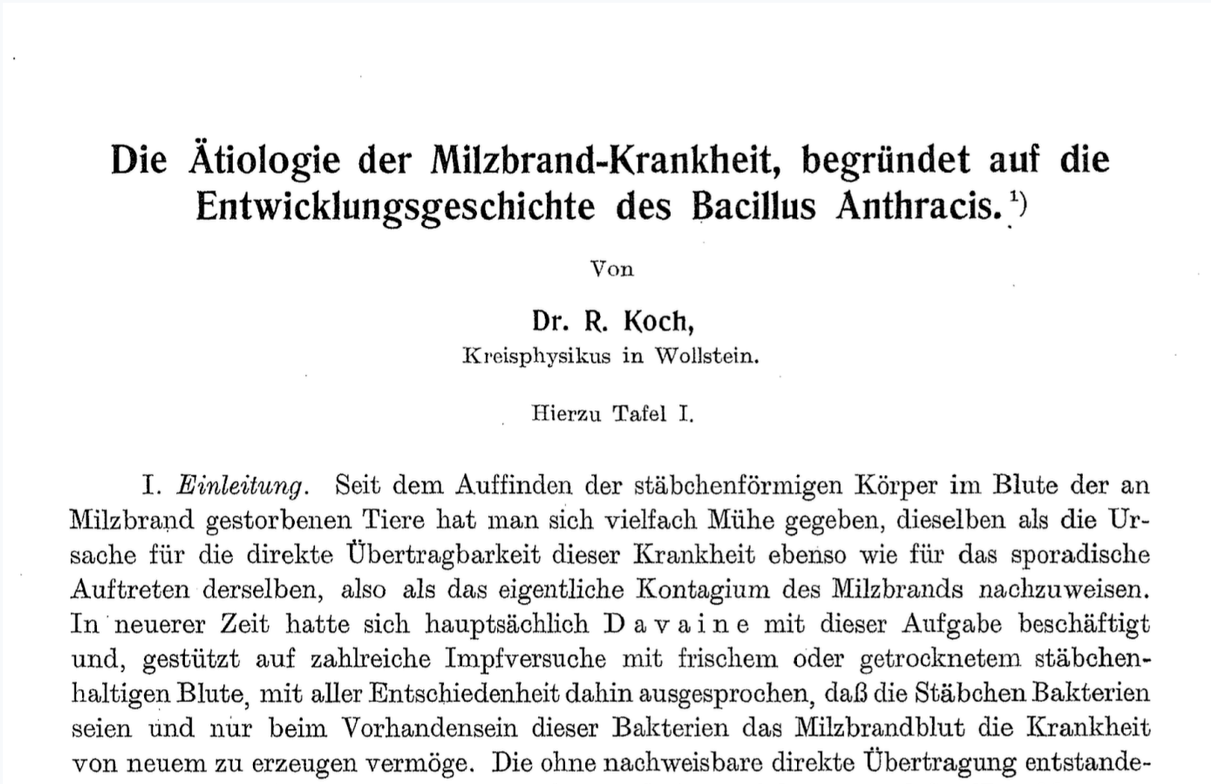
PART 01

Establishing causation

How do you prove a microbe causes a disease? The answer has been rewritten twice already — and the microbiome forces a third rewrite.

Koch’s postulates

In 1876, Robert Koch traced anthrax to *Bacillus anthracis* by following the organism through its entire life cycle — and in doing so gave medicine its first rigorous standard for **causation**, not mere association.



Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis.[1)]

Von

Dr. R. Koch,
Kreisphysikus in Wollstein.

Hierzu Tafel I.

I. *Einleitung.* Seit dem Auffinden der stäbchenförmigen Körper im Blute der an Milzbrand gestorbenen Tiere hat man sich vielfach Mühe gegeben, dieselben als die Ursache für die direkte Übertragbarkeit dieser Krankheit ebenso wie für das sporadische Auftreten derselben, also als das eigentliche Kontagium des Milzbrands nachzuweisen. In neuerer Zeit hatte sich hauptsächlich D a v a i n e mit dieser Aufgabe beschäftigt und, gestützt auf zahlreiche Impfversuche mit frischem oder getrocknetem stäbchenhaltigen Blute, mit aller Entschiedenheit dahin ausgesprochen, daß die Stäbchen Bakterien seien und nur beim Vorhandensein dieser Bakterien das Milzbrandblut die Krankheit von neuem zu erzeugen vermöge. Die ohne nachweisbare direkte Übertragung entstande-

Koch, R. *Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis.* Cohns Beiträge zur Biologie der Pflanzen **2**(2), 277–310 (1876). No DOI — the article predates DOI registration.

Koch’s postulates

1. The microorganism is found in abundance in **diseased** but not in healthy individuals
2. The microorganism can be **isolated** from the diseased host and grown in pure culture
3. The cultured microorganism **causes disease** when inoculated into a healthy host
4. The microorganism can be **re-isolated** from the inoculated host and is the same as the original
5. **Elimination** of the microbe from the host alleviates disease

> ⓘ **The load-bearing assumption**
>
> Every postulate assumes **one organism**, growable in **pure culture**, **sufficient on its own**. All three assumptions fail for the microbiome.

Koch’s postulates: the molecular age

Falkow recast causation in terms of **genes** rather than organisms — by the 1980s the interesting variation was between strains of the same species (Falkow, 1988).

1. The virulence gene is found in **pathogenic** but not nonpathogenic strains
2. **Deletion or inactivation** of the gene leads to loss of pathogenicity
3. **Reactivation or allelic replacement** of the gene restores pathogenicity

Falkow revisited these criteria fifteen years later, noting how often they bend in practice (Falkow, 2004).

Koch’s postulates: the sequencing age

Fredricks and Relman rewrote the postulates again once sequence, not culture, became the primary evidence — most pathogens of interest had never been cultured at all (Fredericks & Relman, 1996).

- Focus shifts to **nucleic acid sequences** rather than culture or whole genes
- **Individuals or communities**, as identified by sequencing, differ in abundance, organization, and/or function in diseased vs. healthy hosts
- Community virulence **may or may not** depend on specific, well-defined virulence factors
- **Modifying the community** — or removing specific members — alleviates disease

Why study the microbiome?

Humans are full of microorganisms — skin, gut, oral cavity, nasal cavity, eyes — and they affect health, drug metabolism, and treatment response.

3.8×10^{13}
bacterial cells in a reference 70 kg adult

≈ 1.3 : 1
bacteria-to-human-cell ratio

≈ 150 ×
more genes than the human complement — 3.3 M microbial genes

Correction to a familiar number

The often-repeated **“10× more microbial cells than human cells”** traces to a 1972 back-of-envelope estimate. Sender, Fuchs and Milo’s re-derivation puts the ratio near **1:1** — close enough that one defecation event shifts it. Most of the human-cell count is red blood cells, which is where the old estimate went wrong. The **gene** claim survived; the **cell** claim did not.

Cell counts from Sender et al. (2016); gene catalogue from Qin et al. (2010).

PART 02

Measuring the microbiome

Every step between the swab and the count matrix adds bias. Knowing which step added which bias is most of the skill.

Workflow

> **The thing to remember**
>
> **Every step can add bias or noise** — swab site, extraction chemistry, primer choice, PCR cycle number, sequencing platform, and reference database each leave a fingerprint on the final abundance table.

Two sequencing approaches

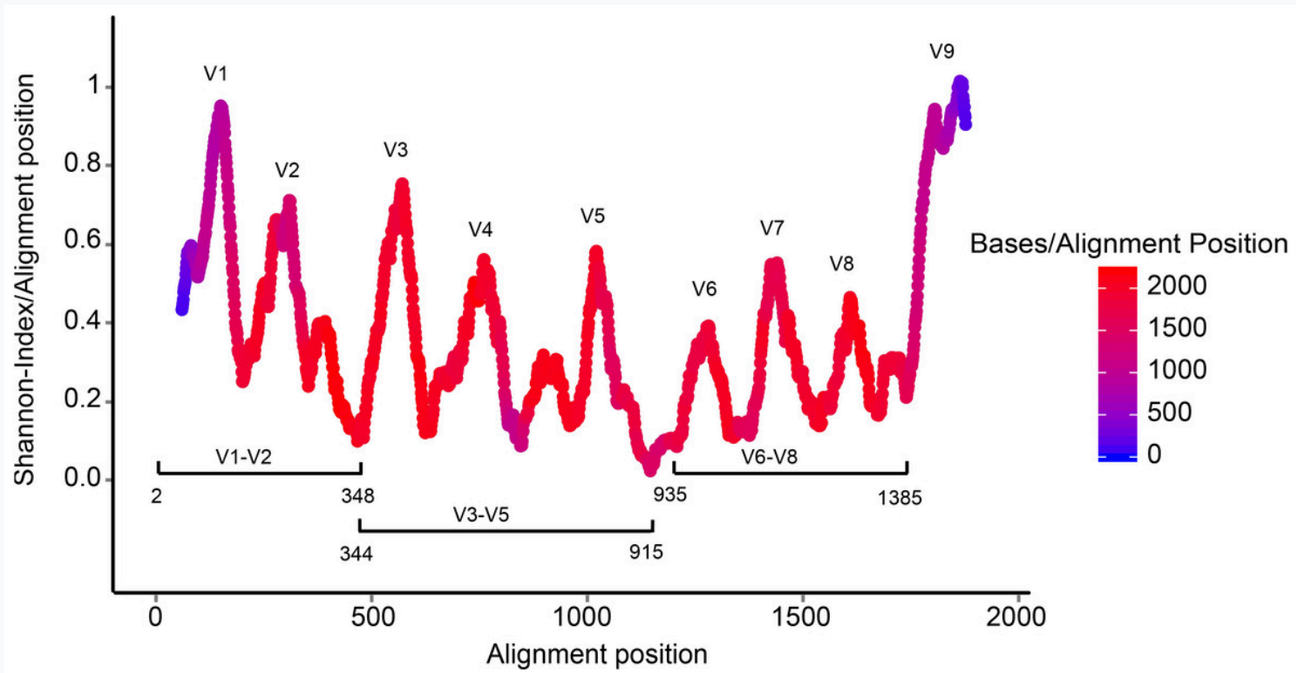
SHOTGUN

- Sequence **all** DNA in the sample
- More information — taxonomy *and* function
- Higher complexity
- Higher cost

AMPLICON

- Sequence **one** marker gene
- No functional information
- Less complex to analyze
- Cheaper

The 16S rRNA gene

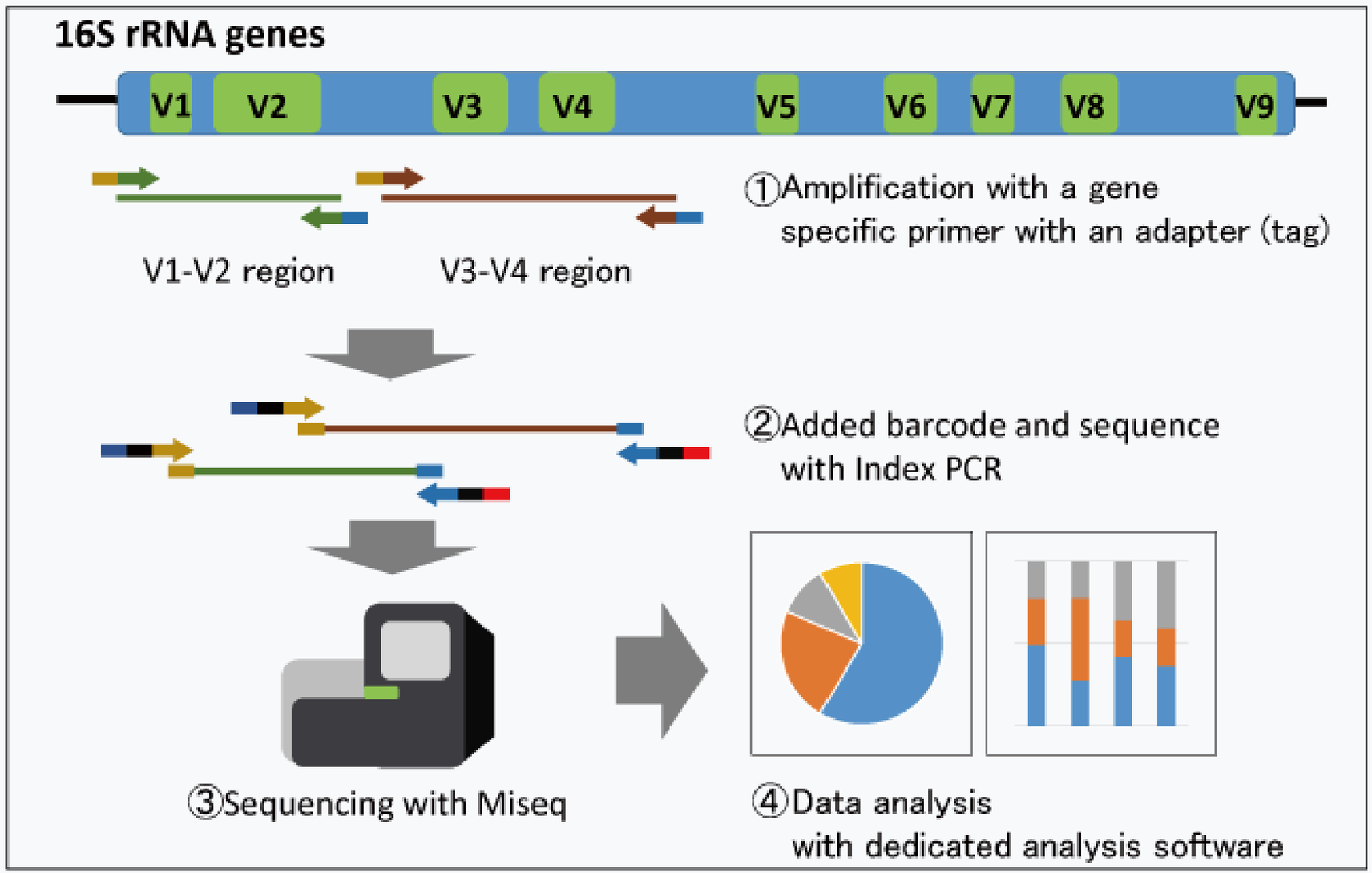


Nine **variable regions** (V1–V9) sit between highly conserved stretches. The conserved parts give you universal primer binding sites; the variable parts give you taxonomic signal.

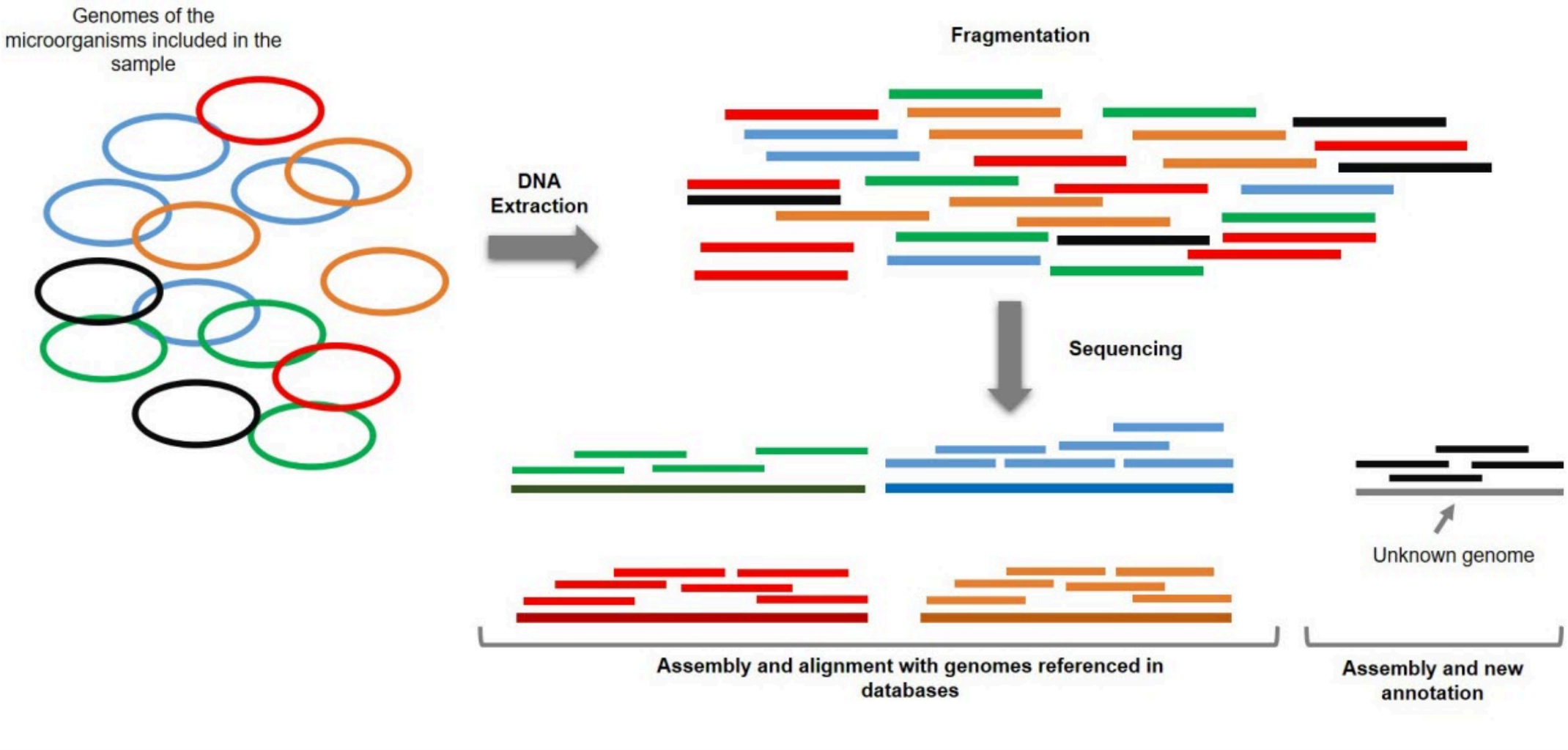
Which region you amplify determines what you can resolve — and what you systematically miss.

Shannon index per alignment position (50-bp moving average); per-base coverage shown by the heatmap; numbering follows the *E. coli* 16S rRNA gene. Figure 2 of the RIM-DB paper — computed over a **methanogen** alignment, so read the shape of the profile rather than exact peak heights (Seedorf et al., 2014).

16S amplicon sequencing



Shotgun metagenomics



Choosing between them

SHOTGUN

Pros

- Not biased by amplicon primer set
- Not limited by conservation of one gene
- Provides functional information

Cons

- Environmental and host contamination
- Expensive ($1000+/sample)
- Complex analysis; needs HPC, high memory

AMPLICON

Pros

- Well established, large body of prior data
- Inexpensive ($50–$100/sample)

Cons

- V-region choice biases results
- Built on a very well-conserved gene — hard to resolve species and strains

16S reference databases

- Berkeley lab
- August 2013
- 202,421 entries

green genes
16S rRNA gene database and workbench compatible with ARB
greengenes.lbl.gov

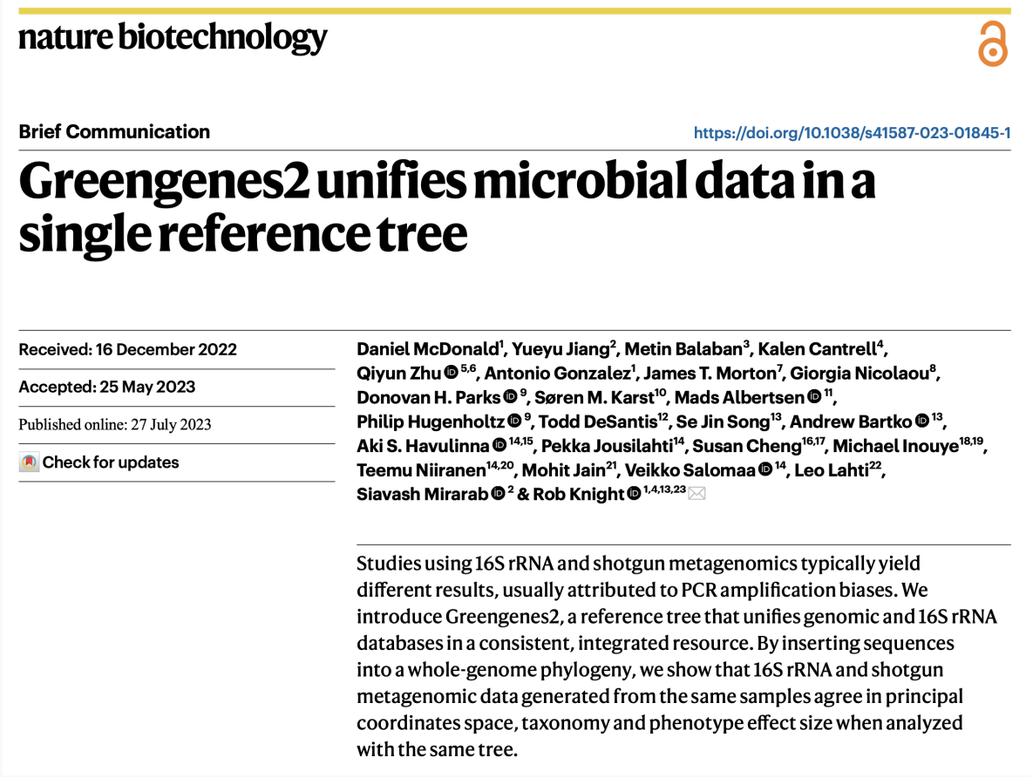
- Max Planck Institute
- July 2015
- 172,418 entries

silva
high quality ribosomal RNA databases

These snapshots are stale

Greengenes 13_8 and SILVA 128 are the versions many published pipelines still pin. Taxonomy has moved substantially since — results are **not comparable across database versions**, and “which database” is a real methods choice, not a detail (DeSantis et al., 2006; Quast et al., 2012).

Greengenes2 unifies the two worlds



nature biotechnology

Brief Communication

https://doi.org/10.1038/s41587-023-01845-1

Greengenes2 unifies microbial data in a single reference tree

Received: 16 December 2022

Accepted: 25 May 2023

Published online: 27 July 2023

Check for updates

Daniel McDonald[1], Yueyu Jiang[2], Metin Balaban[3], Kalen Cantrell[4], Qiyun Zhu[5,6], Antonio Gonzalez[1], James T. Morton[7], Giorgia Nicolaou[8], Donovan H. Parks[9], Søren M. Karst[10], Mads Albertsen[11], Philip Hugenholtz[9], Todd DeSantis[12], Se Jin Song[13], Andrew Bartko[13], Aki S. Havulinna[14,15], Pekka Jousilahti[14], Susan Cheng[16,17], Michael Inouye[18,19], Teemu Niiranen[14,20], Mohit Jain[21], Veikko Salomaa[14], Leo Lahti[22], Siavash Mirarab[2] & Rob Knight[1,4,13,23]

Studies using 16S rRNA and shotgun metagenomics typically yield different results, usually attributed to PCR amplification biases. We introduce Greengenes2, a reference tree that unifies genomic and 16S rRNA databases in a consistent, integrated resource. By inserting sequences into a whole-genome phylogeny, we show that 16S rRNA and shotgun metagenomic data generated from the same samples agree in principal coordinates space, taxonomy and phenotype effect size when analyzed with the same tree.

16S and shotgun studies of the *same samples* have historically disagreed — usually blamed on PCR amplification bias.

Greengenes2 inserts both data types into a **single whole-genome phylogeny**. Analyzed against the same tree, 16S and shotgun data agree in principal coordinates space, taxonomy, and phenotype effect size (McDonald et al., 2023).

Profiling shotgun data: MetaPhlAn 4

Metagenomic assembly finds novel organisms but recovers only the abundant ones. MetaPhlAn 4 combines **metagenome assemblies with isolate genomes** — built from a curated collection of ~1.01 M prokaryotic reference and metagenome-assembled genomes.

The practical consequence: a large fraction of reads that previously went unclassified now map to *uncharacterized* species (Blanco-Míguez et al., 2023).

Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4

Aitor Blanco-Miguez[1], Francesco Beghini[1], Fabio Cumbo[1], Lauren J. McIver[2,3], Kelsey N. Thompson[2,3], Moreno Zolfo[1], Paolo Manghi[1], Leonard Dubois[1], Kun D. Huang[1], Andrew Maltez Thomas[1], Gianmarco Piccinno[1], Elisa Piperni[1,4], Michal Punčochář[1], Mireia Valles-Colomer[1], Adrian Tett[1,5], Francesca Giordano[6], Richard Davies[6], Jonathan Wolf[6], Sarah E. Berry[7], Tim D. Spector[8], Eric A. Franzosa[2,3], Edoardo Pasolli[9], Francesco Asnicar[1], Curtis Huttenhower[2,3], Nicola Segata[1,4,#]

[1] Department CIBIO, University of Trento, Trento, Italy
[2] Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3] The Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4] Department of Experimental Oncology, IEO European Institute of Oncology IRCCS, Milan, Italy
[5] Centre for Microbiology and Environmental Systems Science, University of Vienna, Austria
[6] Zoe Global, London, UK
[7] Department of Nutritional Sciences, King's College London, London, UK
[8] Department of Twin Research, King's College London, London, UK
[9] Department of Agricultural Sciences, University of Naples, Naples, Italy
[#] Corresponding author: nicola.segata@unitn.it

Keywords: Metagenomics, microbiome taxonomic profiling, metagenomic assembly

Abstract

Metagenomic assembly enables novel organism discovery from microbial communities, but from most metagenomes it can only capture few abundant organisms. Here, we present a method - MetaPhlAn 4 - to integrate information from both metagenome assemblies and microbial isolate genomes for improved and more comprehensive metagenomic taxonomic profiling. From a curated collection of 1.01M prokaryotic reference and

What comes out: the count matrix

	CL3	CC1	SV1	M31Fcsw	M11Fcsw	M31Plmr	M11Plmr	F21Plmr	M31Tong	M11Tong	LMEpi24M	SLEpi20M	AQC1cm
549322	0	0	0	0	0	0	0	0	0	0	0	1	27
522457	0	0	0	0	0	0	0	0	0	0	0	0	0
951	0	0	0	0	0	0	1	0	0	0	0	0	0
244423	0	0	0	0	0	0	0	0	0	0	0	0	0
586076	0	0	0	0	0	0	0	0	0	0	0	0	0
246140	0	0	0	0	0	0	0	0	0	0	0	0	0
143239	7	1	0	0	0	0	3	0	0	0	0	0	0
244960	0	0	0	0	0	0	0	0	0	0	0	0	0
255340	153	194	0	0	0	0	0	0	0	0	0	0	1
144887	3	5	0	0	0	0	0	0	0	0	0	0	0
141782	9	35	0	0	0	0	0	0	0	0	0	0	0
215972	0	3	0	0	0	0	0	0	0	0	0	0	0
31759	0	1	0	0	0	0	0	0	0	0	0	0	0
30678	0	3	0	0	0	0	0	0	0	0	0	1	0
138353	0	0	19	0	0	0	0	0	0	0	0	0	0
406058	1	0	0	0	0	0	0	0	0	0	0	0	0
30405	808	11	3	0	0	0	0	0	0	0	0	0	0
1126	0	0	0	0	0	0	0	0	0	0	0	0	0
46043	0	0	0	0	0	0	0	0	4	0	0	0	0
512519	0	0	0	0	0	0	0	0	0	0	0	0	0
554594	0	0	0	0	0	0	0	0	0	0	0	0	0

Read the shape, not the numbers

Taxa in rows, samples in columns. **Sparse** — mostly zeros. **Compositional** — column sums are an artifact of sequencing depth, not biology. **Over-dispersed** — non-zero counts span orders of magnitude.

PART 03

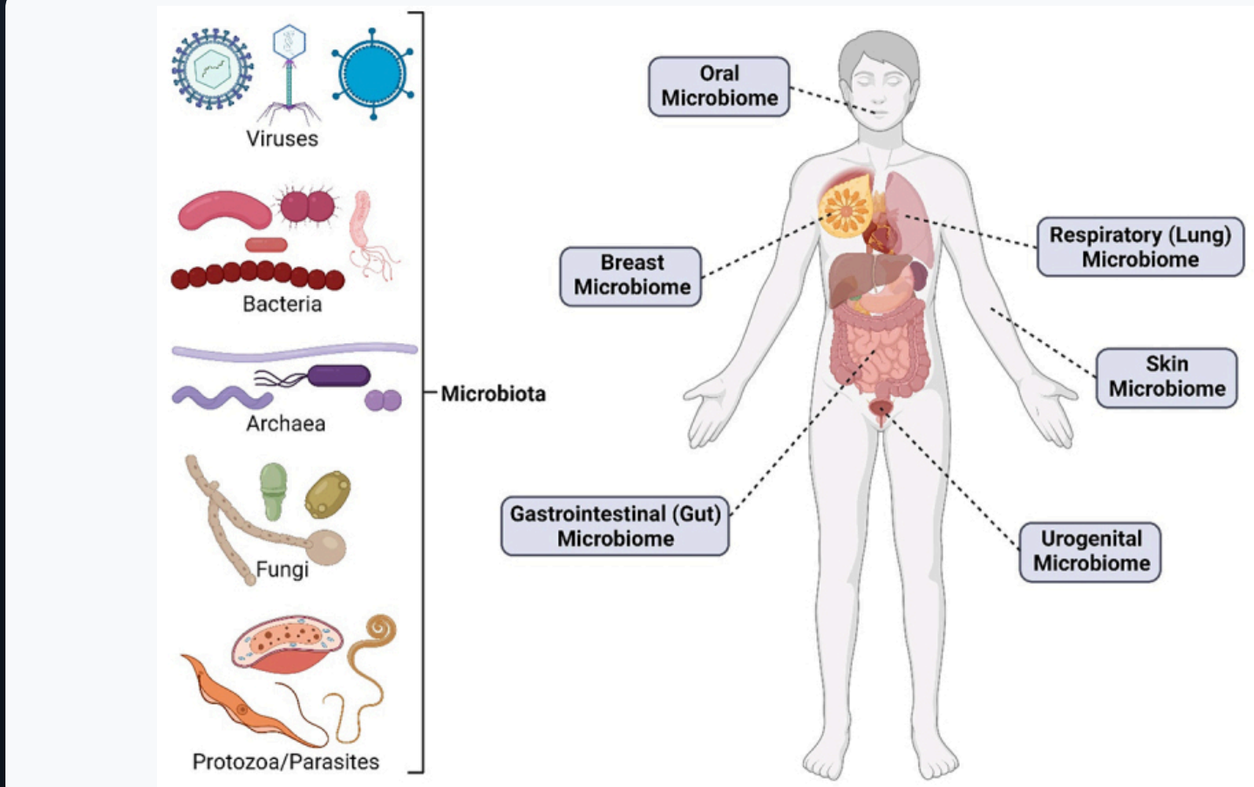
The microbiome in cancer

Some microbes cause cancer outright. Far more shape how it grows, how it is detected, and how it responds to treatment.

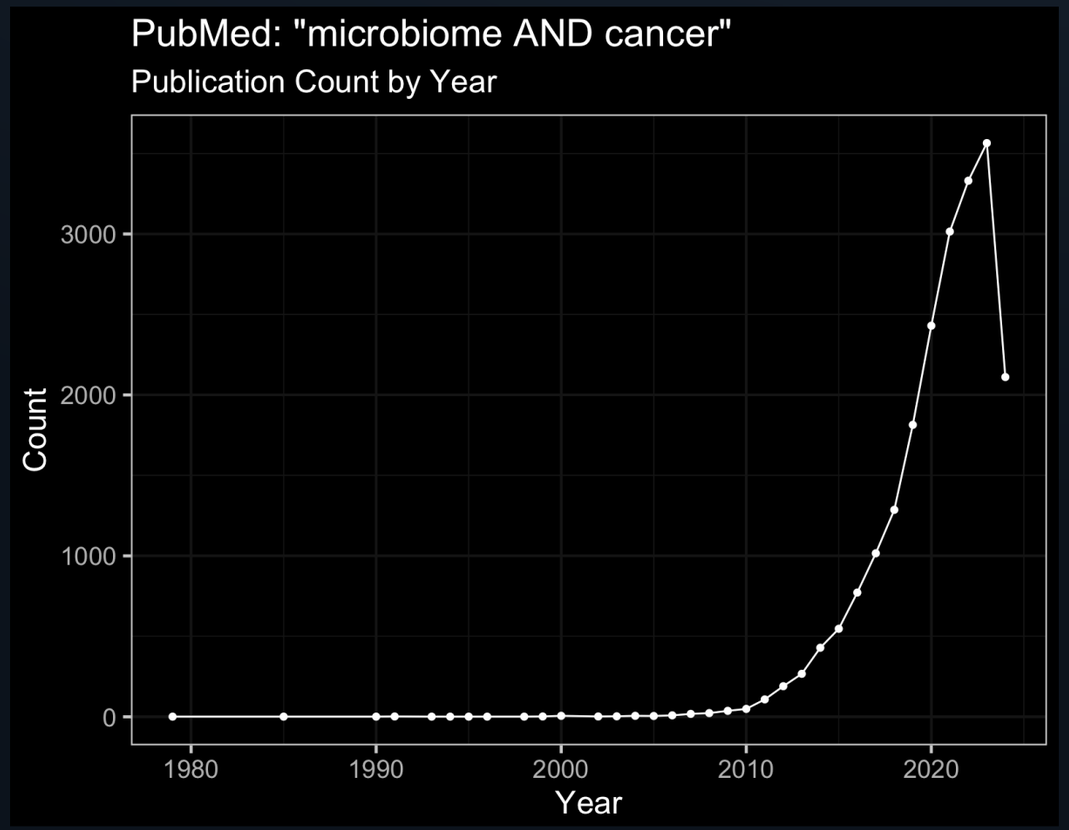
The human microbiota

The microbiota spans **viruses, bacteria, archaea, fungi, and protozoa/parasites**, organized into distinct site-specific communities — oral, respiratory, breast, gastrointestinal, skin, and urogenital.

Site matters more than almost anything else: two gut samples from different people resemble each other far more than a gut and a skin sample from the same person (Kandalai et al., 2023).



A field that grew very fast



Essentially nothing before 2010; roughly **3,600 papers in 2023 alone**.

That growth is the opportunity and the problem. A field this young, moving this fast, accumulates findings faster than it validates them — as the next few slides show.

PubMed query: `microbiome AND cancer`. The final year’s apparent drop is incomplete indexing, not a real decline.

Single microbes that cause cancer

These are the cases where something close to Koch’s postulates actually holds — a single agent, an established mechanism, and in several cases a vaccine or eradication therapy that measurably reduces incidence.

Cancer type	Causative microbe
Gastric cancer	*Helicobacter pylori*
Liver cancer	Hepatitis B virus, Hepatitis C virus
Biliary tree cancer	*Clonorchis sinensis*, *Opisthorchis viverrini*
Cervical cancer	Human papillomavirus (HPV)
Head and neck cancer	HPV
Urinary bladder cancer	*Schistosoma haematobium*
Lymphoma	Epstein–Barr virus
Merkel cell carcinoma	Merkel cell polyomavirus
Kaposi sarcoma	Kaposi sarcoma–associated herpesvirus

H. pylori and gastric cancer: Uemura et al. (2001). *Fusobacterium nucleatum* in colorectal carcinoma sits one rung below this table — robustly replicated, mechanism still open (Castellarin et al., 2011; Kostic et al., 2011).

Indirect effects of the microbiome

Adapted from Kandalai et al. (Kandalai et al., 2023). Intratumoral bacteria were confirmed by microscopy and culture, not sequencing alone (Nejman et al., 2020).

Microbiome-based diagnostics

Method	Cancer type	Marker(s)	AUROC
Salivary microbiome	Pancreatic	*N. elongata*, *S. mitis*	0.90
Salivary microbiome	Lung squamous cell	*Capnocytophaga*, *Veillonella*	0.86
Fecal microbiome	Colorectal	*C. symbiosum*	0.73
Fecal microbiome	Colorectal	*F. nucleatum*	0.86
Fecal microbiome	Lung	Various bacteria	0.76
Plasma cell-free DNA †	Various types	Various bacteria	0.90
Plasma cell-free DNA †	Various types	Various fungi	0.80
Plasma cell-free DNA †	Various types	Bacteria + fungi	0.92

† Do not quote the last three rows

The source of those numbers has been **retracted**. Next slide.

Table adapted from Kandalai et al. (Kandalai et al., 2023).

A cautionary tale

In 2020, Poore *et al.* reported that microbial DNA read out of blood and tissue could discriminate dozens of cancer types with near-perfect accuracy (Poore et al., 2020). The result launched a company and an FDA breakthrough-device designation.

It was wrong

Reanalysis found the signal came from **contaminated reference genomes**, human reads **misclassified as microbial**, and a **batch-correction step that manufactured cancer-type structure**. The classifier was leaning on, among other things, a **shrimp virus** as a human cancer biomarker (Gihawi et al., 2023). *Nature* retracted the paper on **26 June 2024**, with all authors agreeing (Poore et al., 2024).

Two independent 2025 reanalyses — one reprocessing all of TCGA, one using a separate 8,908-patient cohort with contamination controls — found that when contamination is properly handled, **only colorectal cancer** carries a robust distinct microbial signature (Ge et al., 2025; Gihawi et al., 2025).

What to take from that

The tumor microbiome is real

H. pylori in gastric cancer, *F. nucleatum* in colorectal cancer, and Nejman's intracellular bacteria — confirmed by FISH, immunohistochemistry, electron microscopy, and culture, not sequencing alone — all stand (Nejman et al., 2020).

What collapsed was pan-cancer diagnostics from low-biomass sequencing

At these DNA concentrations, contamination from reagents, kits, and reference databases can **exceed the true signal**. Batch correction applied across confounded batches then converts that contamination into apparent biology. This failure mode generalizes well beyond microbiome work.

The practical lesson: when a machine-learning classifier reports an AUROC of 0.95 on low-biomass data, the first question is not "which taxa?" but **"what else differs between my batches?"**

Microbiome and cancer therapy

Therapy	Model	Microbe (site)	Finding
Radiotherapy	Melanoma, lung, cervical	Gram-positives (gut)	**Depleting** them with vancomycin *improves* RT response
Radiotherapy	Breast	Fungi vs. bacteria (gut)	Opposite directions: depleting **fungi** helps, depleting **bacteria** hurts
Cyclophosphamide	Melanoma, sarcoma	*L. johnsonii*, *E. hirae* (gut)	Translocation drives the Th17/Th1 response the drug needs
Gemcitabine	Pancreatic	Gammaproteobacteria (tumor)	Bacterial cytidine deaminase **inactivates the drug**
Dacarbazine	Melanoma lung mets	*L. rhamnosus* (**aerosolized**)	Pulmonary probiotic promotes anti-metastatic immunity

Rows in order: Uribe-Herranz et al. (2019); Shiao et al. (2021); Viaud et al. (2013); Geller et al. (2017); Le Noci et al. (2018). Note the direction of the first two — the benefit comes from **removing** microbes, and bacteria and fungi push opposite ways.

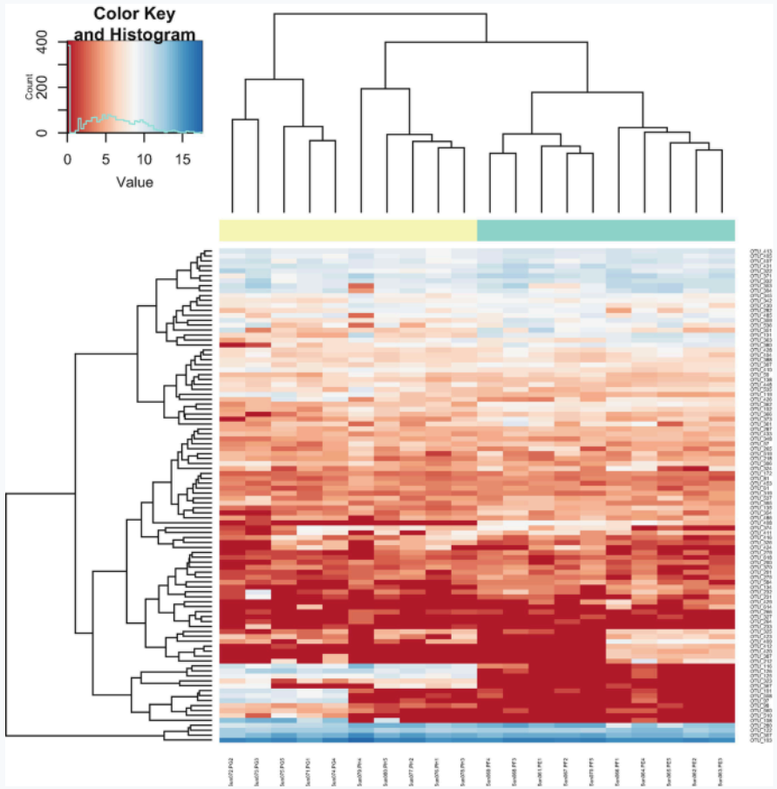
PART 04

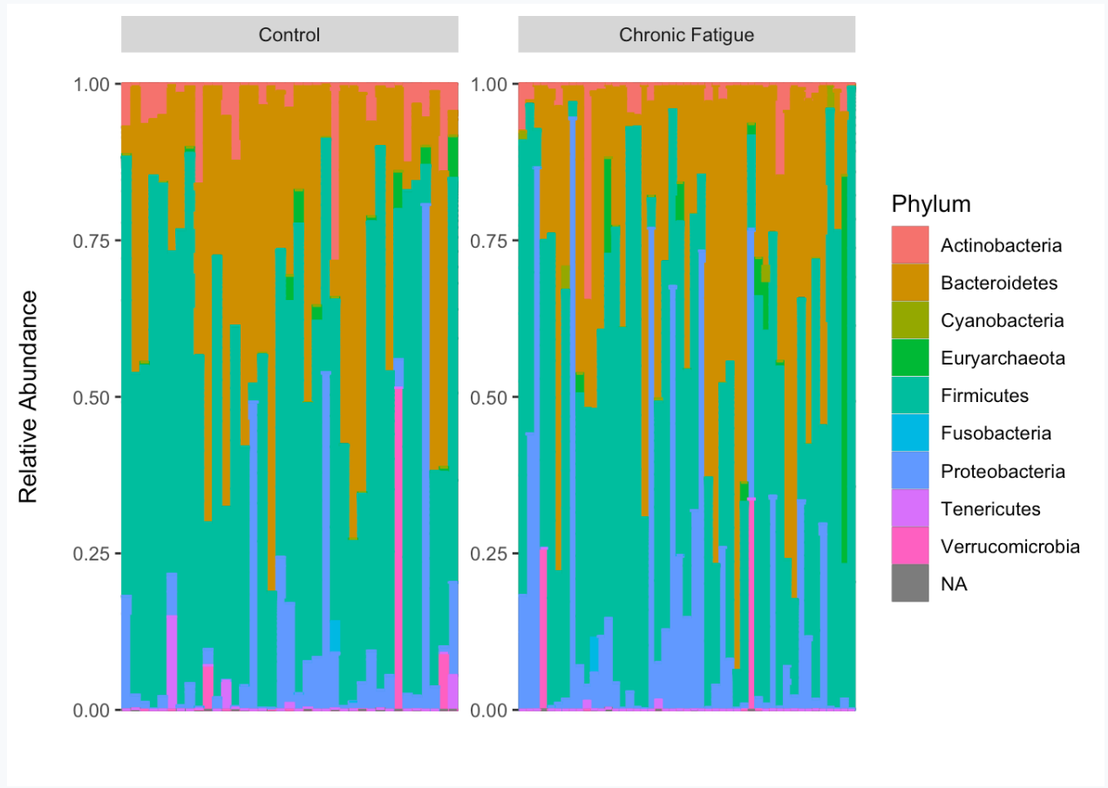
Analysis and data resources

The count matrix is where the biology stops
and the statistics start.

Back to the count matrix

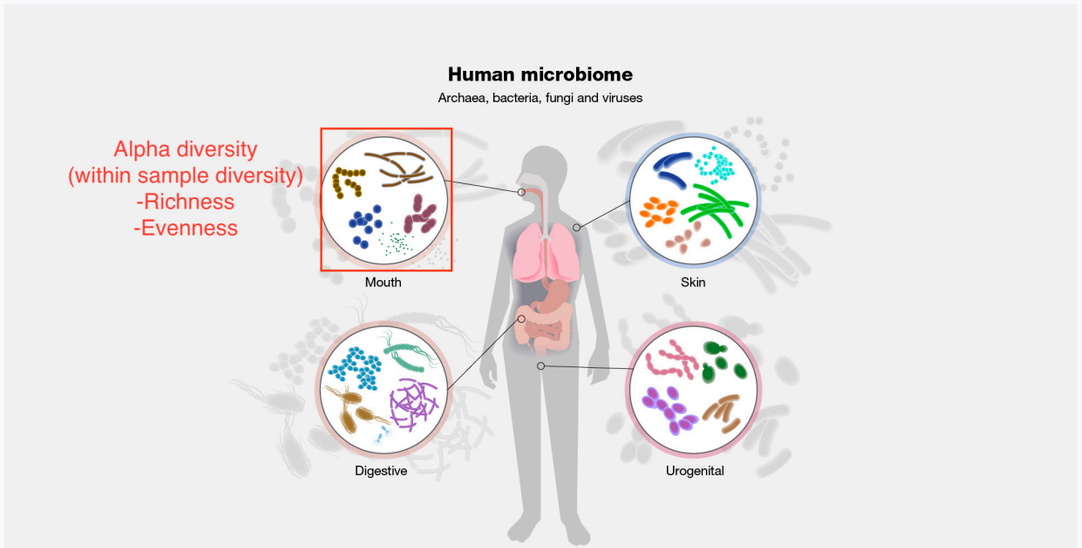
Differential abundance analysis





Naive *t*-tests on relative abundance are badly behaved on compositional, sparse, over-dispersed counts — hence a whole literature of dedicated methods.

Alpha diversity



One number per sample. *How diverse is this one community?*

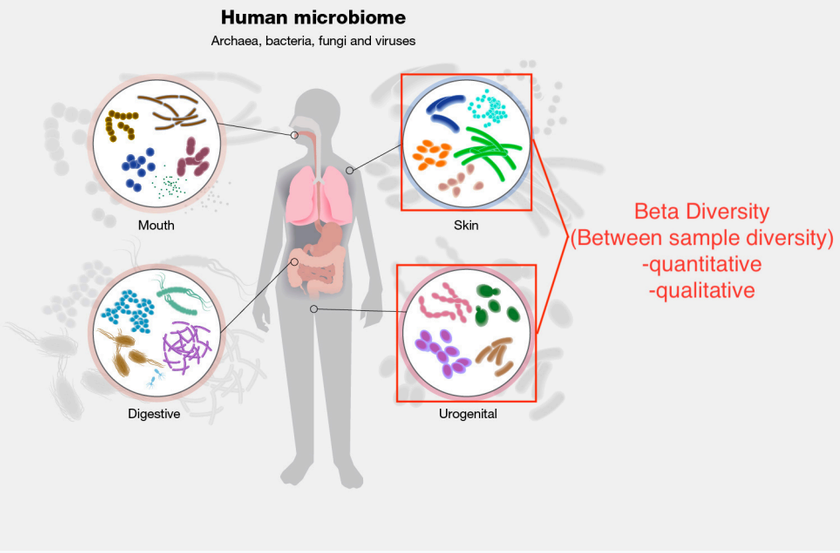
Built from **richness** (how many taxa) and **evenness** (how equally abundance is spread across them).

Questions it answers:

- Does gut diversity fall after antibiotics — and how fast does it recover?
- Does low pre-treatment diversity predict poor checkpoint-inhibitor response? (Gopalakrishnan et al., 2018)

Indices weight the two components differently: Shannon balances them, Simpson favours dominant taxa, Faith’s PD counts phylogenetic branch length. Richness rises with sequencing depth, so depth must be controlled before comparing.

Beta diversity



One number per *pair* of samples. *How different are these two communities?*

Questions it answers:

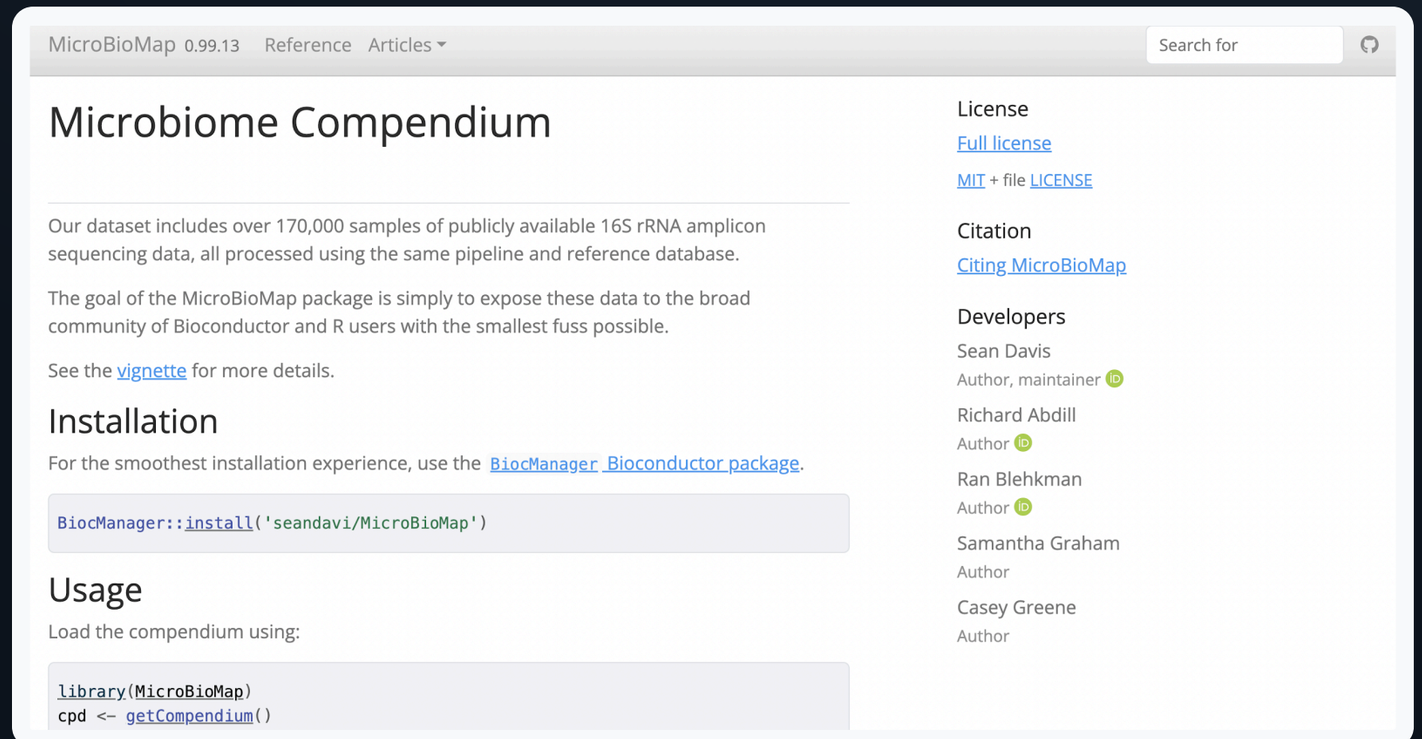
- Do responders and non-responders differ in overall composition?
- Do samples cluster by disease — or by sequencing batch?

The distinction in one line

Alpha summarizes **one sample**; beta compares **two**. Samples can have identical alpha diversity and still separate perfectly on beta — the same *number* of taxa, entirely different taxa.

Qualitative metrics use presence/absence (Jaccard, unweighted UniFrac); **quantitative** ones use abundance (Bray-Curtis, weighted UniFrac). UniFrac scores closely related taxa as more similar by using the phylogeny.

Accessing microbiome data: MicroBioMap



Over **170,000 publicly available 16S amplicon samples**, all processed through the same pipeline against the same reference database — so cross-study comparison is not confounded by pipeline choice.

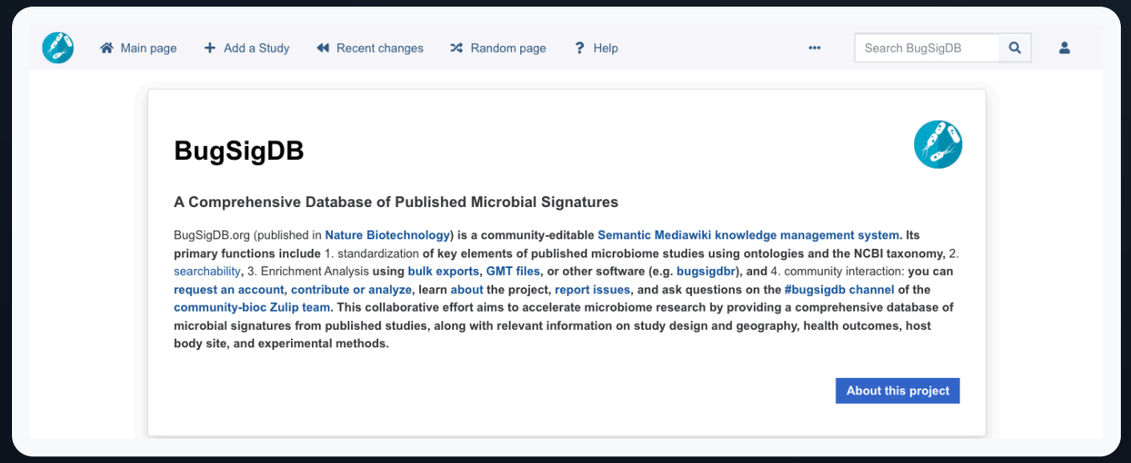
```
1 BiocManager::install('seandavi/MicroBioMap')
2 library(MicroBioMap)
3 cpd <- getCompendium()
```

https://seandavi.github.io/MicroBioMap/

Accessing microbiome data: curatedMetagenomicData

Pasolli et al. (2017) · https://waldronlab.io/curatedMetagenomicData/

Signatures, not samples: BugSigDB



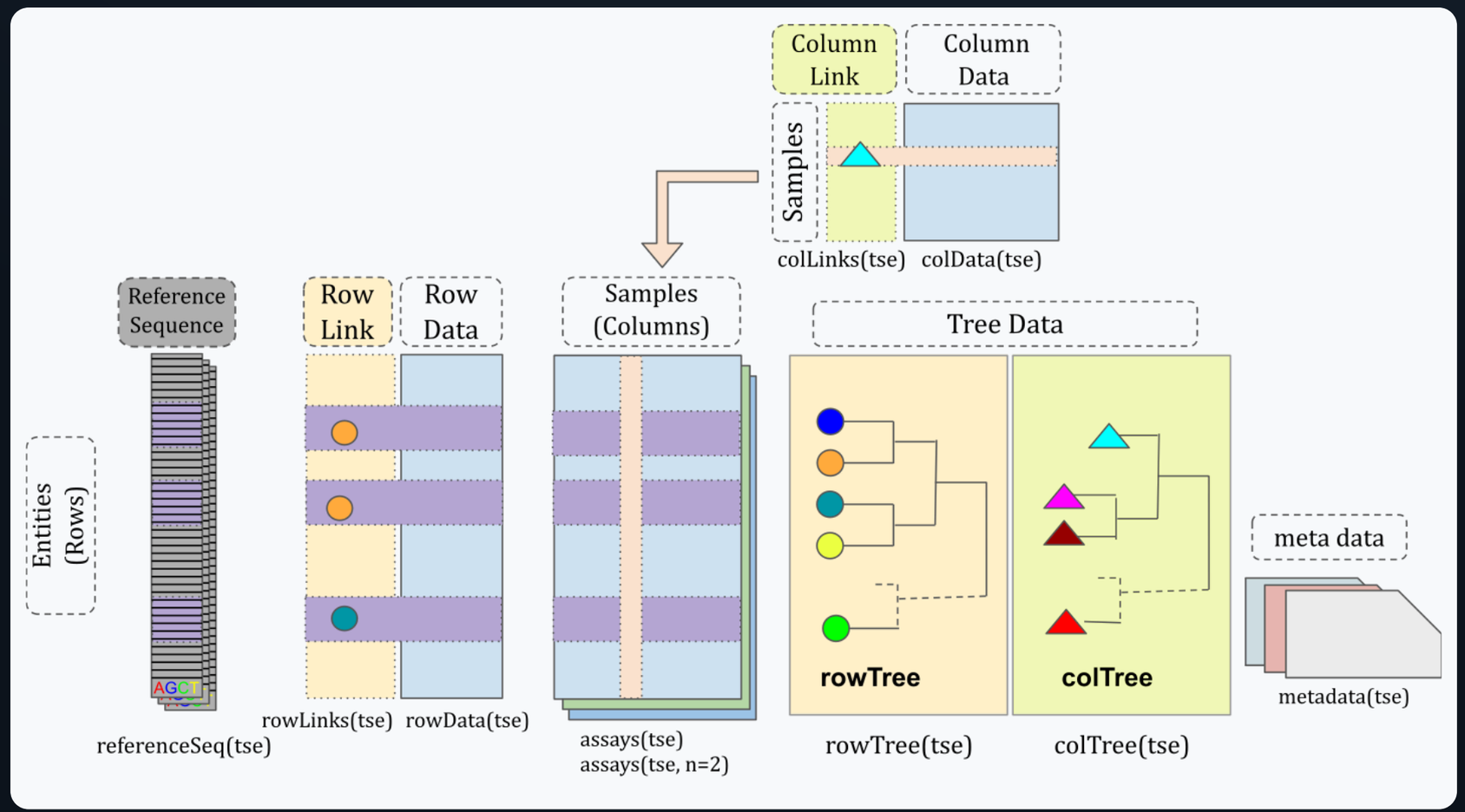
The other two resources give you **samples**. BugSigDB gives you **published results** — curated signatures of differentially abundant taxa, with study design, body site, outcome, and statistical method in controlled vocabulary.

>2,500 signatures from >600 studies at release; community-editable since (Geistlinger et al., 2023).

Gene set enrichment analysis, for microbes

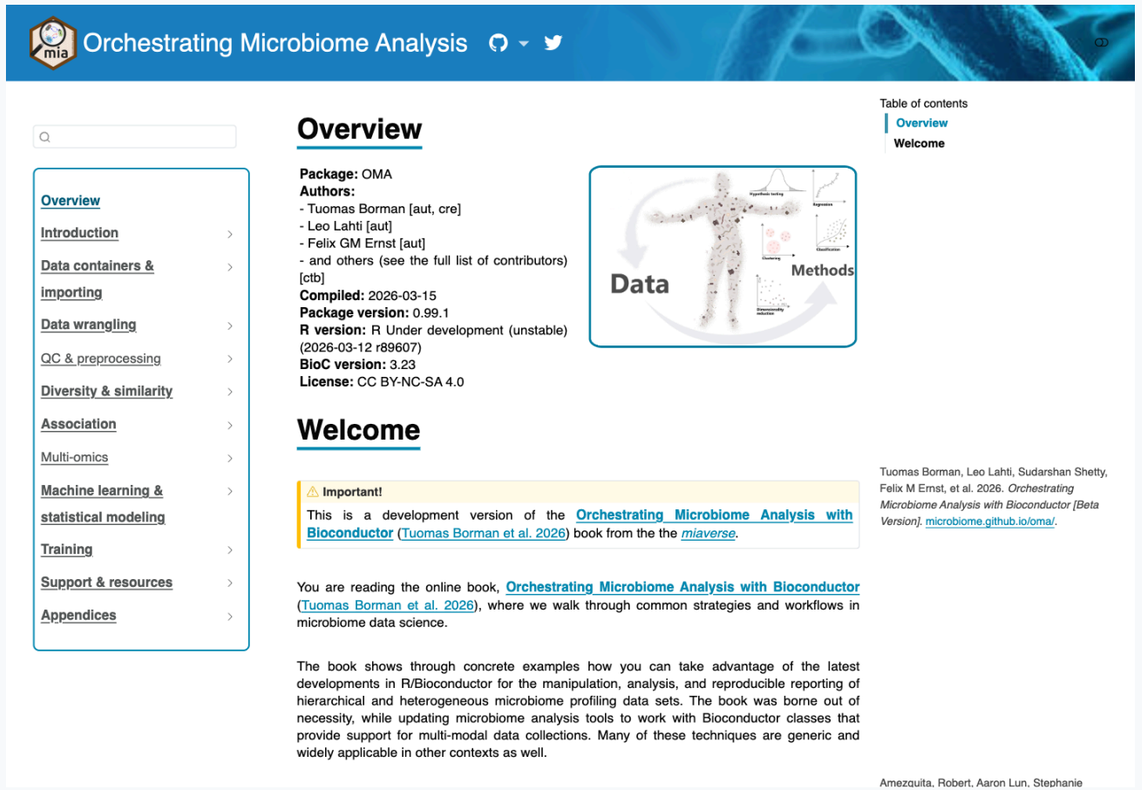
Signatures export as **GMT files**, the same format as MSigDB gene sets — so your taxa become a query set and you ask which published signatures they overlap more than chance. `bugsigdbr` does it in R.

One object to hold it all



Counts, sample metadata, taxon metadata, reference sequences **and the phylogeny** in one object — so the tree travels with the data through subsetting and filtering rather than drifting out of sync (Huang et al., 2021).

Where to go next: the OMA book



Orchestrating Microbiome Analysis with Bioconductor — free, online, and maintained by the `miaverse` project. It is built on exactly the object from the previous slide.

Importing and containers, QC, diversity, differential abundance, multi-omics, machine learning — each chapter is **runnable R against example data**, so it works as a course as much as a reference.

https://microbiome.github.io/OMA/

Takeaways

1. **Causation is the hard part.** Koch's postulates have been rewritten twice; community-level causation is still an open methodological problem.
2. **Amplicon vs. shotgun is a real trade-off** — cost and simplicity against resolution and function. Neither is the default right answer.
3. **The reference database is a methods choice**, not a detail. Results are not comparable across database versions.
4. **A handful of microbes cause cancer; many more modulate it** — through metabolites, inflammation, barrier disruption, and drug metabolism.
5. **Low-biomass diagnostics deserve skepticism.** Contamination and batch effects can manufacture an excellent-looking AUROC out of nothing.
6. **The count matrix is sparse, compositional, and over-dispersed** — and every downstream method lives or dies on handling that.

QUESTIONS

Thank you

Sean Davis, MD, PhD · University of Colorado
Anschutz School of Medicine

References

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., … Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, *41*(11), 1633–1644. https://doi.org/10.1038/s41587-023-01688-w

Castellarin, M., Warren, R. L., Freeman, J. D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R. A., & Holt, R. A. (2011). *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Research*, *22*(2), 299–306. https://doi.org/10.1101/gr.126516.111

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. https://doi.org/10.1128/aem.03006-05

Falkow, S. (1988). Molecular Koch's Postulates Applied to Microbial Pathogenicity. *Clinical Infectious Diseases*, *10*(Supplement 2), S274–S276. https://doi.org/10.1093/cid/10.supplement_2.s274

Falkow, S. (2004). Molecular Koch's postulates applied to bacterial pathogenicity — a personal recollection 15 years later. *Nature Reviews Microbiology*, *2*(1), 67–72. https://doi.org/10.1038/nrmicro799

Fredericks, D. N., & Relman, D. A. (1996). Sequence-based identification of microbial pathogens: A reconsideration of Koch's postulates. *Clinical Microbiology Reviews*, *9*(1), 18–33. https://doi.org/10.1128/cmr.9.1.18

Ge, Y., Lu, J., Puiu, D., Revsine, M., & Salzberg, S. L. (2025). Comprehensive analysis of microbial content in whole-genome sequencing samples from The Cancer Genome Atlas project. *Science Translational Medicine*, *17*(814). https://doi.org/10.1126/scitranslmed.ads6335

Geistlinger, L., Mirzayi, C., Zohra, F., Azhar, R., Elsafoury, S., Grieve, C., Wokaty, J., Gamboa-Tuz, S. D., Sengupta, P., Hecht, I., Ravikrishnan, A., Gonçalves, R. S., Franzosa, E., Raman, K., Carey, V., Dowd, J. B., Jones, H. E., Davis, S., Segata, N., … Waldron, L. (2023). BugSigDB captures patterns of differential abundance across a broad range of host-associated microbial signatures. *Nature Biotechnology*, *42*(5), 790–802. https://doi.org/10.1038/s41587-023-01872-y

Geller, L. T., Barzily-Rokni, M., Danino, T., Jonas, O. H., Shental, N., Nejman, D., Gavert, N., Zwang, Y., Cooper, Z. A., Shee, K., Thaiss, C. A., Reuben, A., Livny, J., Avraham, R., Frederick, D. T., Ligorio, M., Chatman, K., Johnston, S. E., Mosher, C. M., … Straussman, R. (2017). Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science*, *357*(6356), 1156–1160. https://doi.org/10.1126/science.aah5043

Gihawi, A., Ge, Y., Lu, J., Puiu, D., Xu, A., Cooper, C. S., Brewer, D. S., Pertea, M., & Salzberg, S. L. (2023). Major data analysis errors invalidate cancer microbiome findings. *mBio*, *14*(5). https://doi.org/10.1128/mbio.01607-23

Gihawi, A., Wood, H. M., Clark, J., O'Grady, J., Eeles, R. A., Wedge, D. C., Jakobsdottir, G. M., Magiorkinis, G., Schache, A. G., Masterson, L., Lechner, M., Fenton, T. R., Jones, T. M., Flanagan, A. M., De Noon, S., Rubinsteyn, A., Hurst, R., Cooper, C. S., & Brewer, D. S. (2025). The landscape of microbial associations in human cancer. *Science Translational Medicine*, *17*(814). https://doi.org/10.1126/scitranslmed.ads6166

Gopalakrishnan, V., Spencer, C. N., Nezi, L., Reuben, A., Andrews, M. C., Karpinets, T. V., Prieto, P. A., Vicente, D., Hoffman, K., Wei, S. C., Cogdill, A. P., Zhao, L., Hudgens, C. W., Hutchinson, D. S., Manzo, T., Petaccia de Macedo, M., Cotechini, T., Kumar, T., Chen, W. S., … Wargo, J. A. (2018). Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients. *Science*, *359*(6371), 97–103. https://doi.org/10.1126/science.aan4236

Huang, R., Soneson, C., Ernst, F. G. M., Rue-Albrecht, K. C., Yu, G., Hicks, S. C., & Robinson, M. D. (2021). TreeSummarizedExperiment: A S4 class for data with hierarchical structure. *F1000Research*, *9*, 1246. https://doi.org/10.12688/f1000research.26669.2

Kandalai, S., Li, H., Zhang, N., Peng, H., & Zheng, Q. (2023). The human microbiome and cancer: A diagnostic and therapeutic perspective. *Cancer Biology & Therapy*, *24*(1). https://doi.org/10.1080/15384047.2023.2240084

Kostic, A. D., Gevers, D., Pedamallu, C. S., Michaud, M., Duke, F., Earl, A. M., Ojesina, A. I., Jung, J., Bass, A. J., Tabernero, J., Baselga, J., Liu, C., Shivdasani, R. A., Ogino, S., Birren, B. W., Huttenhower, C., Garrett, W. S., & Meyerson, M. (2011). Genomic analysis identifies association of *fusobacterium* with colorectal carcinoma. *Genome Research*, *22*(2), 292–298. https://doi.org/10.1101/gr.126573.111

Le Noci, V., Guglielmetti, S., Arioli, S., Camisaschi, C., Bianchi, F., Sommariva, M., Storti, C., Triulzi, T., Castelli, C., Balsari, A., Tagliabue, E., & Sfondrini, L. (2018). Modulation of Pulmonary Microbiota by Antibiotic or Probiotic Aerosol Therapy: A Strategy to Promote Immunosurveillance against Lung Metastases. *Cell Reports*, *24*(13), 3528–3538. https://doi.org/10.1016/j.celrep.2018.08.090

McDonald, D., Jiang, Y., Balaban, M., Cantrell, K., Zhu, Q., Gonzalez, A., Morton, J. T., Nicolaou, G., Parks, D. H., Karst, S. M., Albertsen, M., Hugenholtz, P., DeSantis, T., Song, S. J., Bartko, A., Havulinna, A. S., Jousilahti, P., Cheng, S., Inouye, M., … Knight, R. (2023). Greengenes2 unifies microbial data in a single reference tree. *Nature Biotechnology*, *42*(5), 715–718. https://doi.org/10.1038/s41587-023-01845-1

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., … Straussman, R. (2020). The human tumor microbiome is composed of tumor type–specific intracellular bacteria. *Science*, *368*(6494), 973–980. https://doi.org/10.1126/science.aay9189

Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., Beghini, F., Malik, F., Ramos, M., Dowd, J. B., Huttenhower, C., Morgan, M., Segata, N., & Waldron, L. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nature Methods*, *14*(11), 1023–1024. https://doi.org/10.1038/nmeth.4468

Poore, G. D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., Kosciolek, T., Janssen, S., Metcalf, J., Song, S. J., Kanbar, J., Miller-Montgomery, S., Heaton, R., Mckay, R., Patel, S. P., Swafford, A. D., & Knight, R. (2020). RETRACTED ARTICLE: Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature*, *579*(7800), 567–574. https://doi.org/10.1038/s41586-020-2095-1

Poore, G. D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., Kosciolek, T., Janssen, S., Metcalf, J., Song, S. J., Kanbar, J., Miller-Montgomery, S., Heaton, R., Mckay, R., Patel, S. P., Swafford, A. D., & Knight, R. (2024). Retraction Note: Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature*, *631*(8021), 694–694. https://doi.org/10.1038/s41586-024-07656-x

Qin, J. and, Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Seedorf, H., Kittelmann, S., Henderson, G., & Janssen, P. H. (2014). RIM-DB: A taxonomic framework for community structure analysis of methanogenic archaea from the rumen and other intestinal environments. *PeerJ*, *2*, e494. https://doi.org/10.7717/peerj.494

Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biology*, *14*(8), e1002533. https://doi.org/10.1371/journal.pbio.1002533

Shiao, S. L., Kershaw, K. M., Limon, J. J., You, S., Yoon, J., Ko, E. Y., Guarnerio, J., Potdar, A. A., McGovern, D. P. B., Bose, S., Dar, T. B., Noe, P., Lee, J., Kubota, Y., Maymi, V. I., Davis, M. J., Henson, R. M., Choi, R. Y., Yang, W., … Underhill, D. M. (2021). Commensal bacteria and fungi differentially regulate tumor responses to radiation therapy. *Cancer Cell*, *39*(9), 1202–1213.e6. https://doi.org/10.1016/j.ccell.2021.07.002

Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., Taniyama, K., Sasaki, N., & Schlemper, R. J. (2001). *Helicobacter pylori* Infection and the Development of Gastric Cancer. *New England Journal of Medicine*, *345*(11), 784–789. https://doi.org/10.1056/nejmoa001999

Uribe-Herranz, M., Rafail, S., Beghi, S., Gil-de-Gómez, L., Verginadis, I., Bittinger, K., Pustylnikov, S., Pierini, S., Perales-Linares, R., Blair, I. A., Mesaros, C. A., Snyder, N. W., Bushman, F., Koumenis, C., & Facciabene, A. (2019). Gut microbiota modulate dendritic cell antigen presentation and radiotherapy-induced antitumor immune response. *Journal of Clinical Investigation*, *130*(1), 466–479. https://doi.org/10.1172/jci124332

Viaud, S., Saccheri, F., Mignot, G., Yamazaki, T., Daillère, R., Hannani, D., Enot, D. P., Pfirschke, C., Engblom, C., Pittet, M. J., Schlitzer, A., Ginhoux, F., Apetoh, L., Chachaty, E., Woerther, P.-L., Eberl, G., Bérard, M., Ecobichon, C., Clermont, D., … Zitvogel, L. (2013). The Intestinal Microbiota Modulates the Anticancer Immune Effects of Cyclophosphamide. *Science*, *342*(6161), 971–976. https://doi.org/10.1126/science.1240537