March 9th 2026
BIOL-8802
Reproducible Bioinformatics

Sanya Badole

Prospectus Presentation : RxBiome

A Reproducible Pharmacomicrobiomics Pipeline for Precision Drug Dosing

AGENDA

01 THE PROBLEM

02 THE GAP

03 THE SOLUTION

04 REPRODUCIBILITY CORE

05 THE VISION

06 IMPACT

YOUR GUT BACTERIA CAN SILENTLY INACTIVATE YOUR MEDICATION

Clinical Reality: 30-60% of patients fail to respond to standard drug therapies
Hidden Culprit: Gut microbiome metabolizes drugs before they reach therapeutic targets
Real Examples:

Digoxin (heart failure drug)

Eggerthella lenta inactivates it leading to treatment failure [1]

Irinotecan (chemotherapy drug)

Clostridium perfringens reactivates it leading to severe toxicity [2][3]

Levodopa (Parkinson's drug)

Gut bacteria degrade ***56%*** before reaching the brain [4][5]

Pembrolizumab (cancer immunotherapy drug)

Absence of *Akkermansia muciniphila* in the gut leads to primary resistance [6][7]

The Disconnect: Precision medicine ignores the 100 trillion microbes in your gut

DRUG-MICROBE RESEARCH IS TRAPPED IN A REPRODUCIBILITY CRISIS

The Current State:

- **Fragmented Tools**: Researchers chain 8-12 different tools manually
- **No Standards**: Every lab has custom scripts which is irreproducible across institutions
- **Data Black Holes**: No provenance tracking means we can't validate predictions
- **Clinical Translation Bottleneck**: Takes 5+ years from microbiome discovery to a clinical tool

The Numbers:

9/10

Cohorts failed to replicate a key microbiome finding when the same methods were applied to independent datasets [8]

$2.6-6B

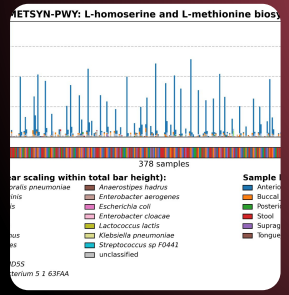
Cost to bring a single drug to market - yet microbiome interactions are ignore in trial design despite PK failures accounting for 10% of all trial failures [9][10]

Zero

End-to-end frameworks exist for pharmacomicrobics

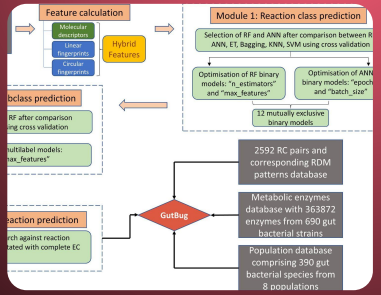
Module 1: QC and Taxonomic Profiling

Fastp, KneadData, Kraken 2, MetaPhlAn 4



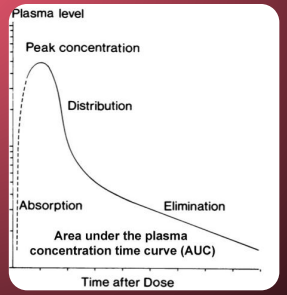
Module 2: Functional Profiling

HUMAnN3 -> microbial gene/pathway coverage



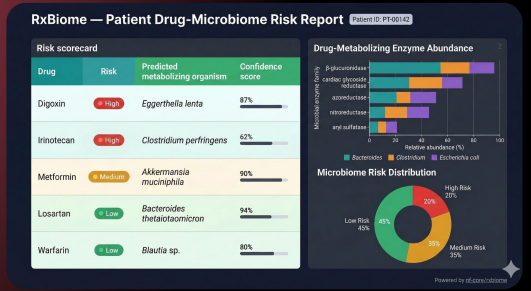
Module 3: Drug-microbiome Interaction Prediction

DrugBank + GutBug ML models



Module 4: PK impact Modelling

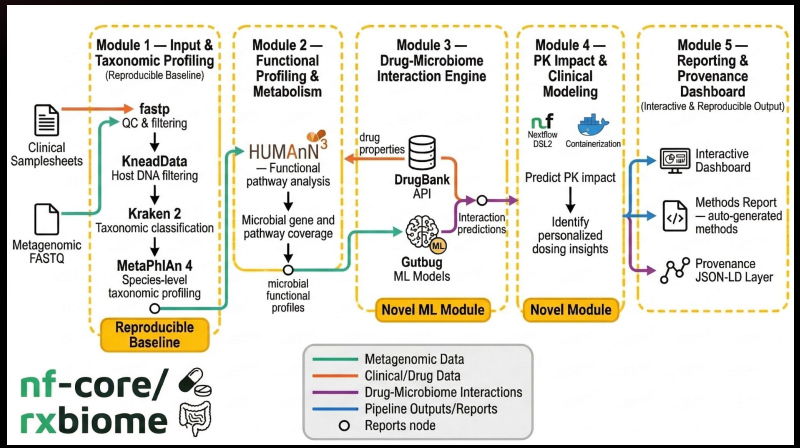
Predicts personalized dosing adjustments



Module 5: Interactive Dashboard + auto-generated methods + JSON-PD provenance

* Workflow diagram generated using Nano Banana 2

RxBiome Architecture



Built on Reproducibility-First Design Principles

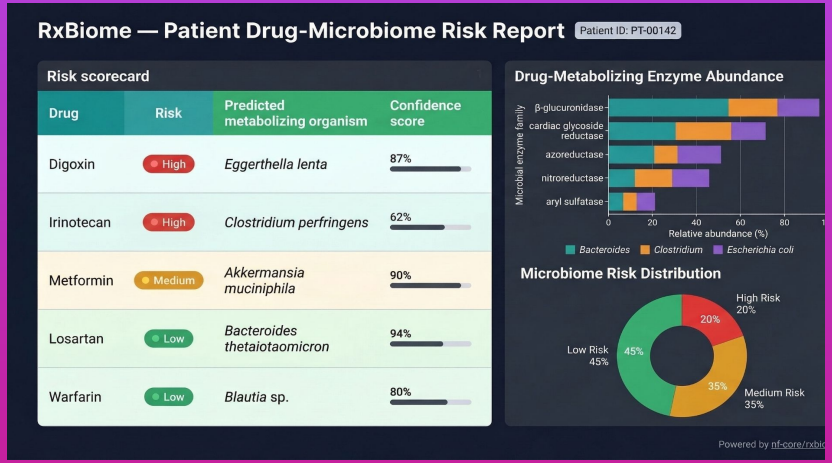
Challenge	RxBiome Solution
Environment Drift	Docker containers with version locked tools
Parameter Archaeology	Nextflow DSL2 with embedded parameter tracking
Data Provenance Mystery	JSON-LD provenance graph auto-generated for every run
Cross-platform Failure	Tested on HPC and local machines

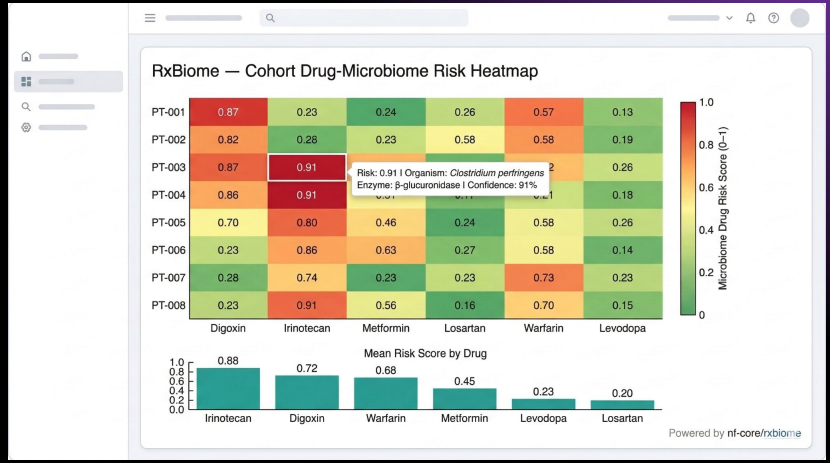
Standards Compliance:

- FAIR data principles (Findable, Accessible, Interoperable, Reusable)
- nf-core best practices, CI/CD, version control, automated testing)
- Containerized at every step

What Clinicians See

From Patient Sample to Actionable Clinical Intelligence in Hours





Patient Risk Report

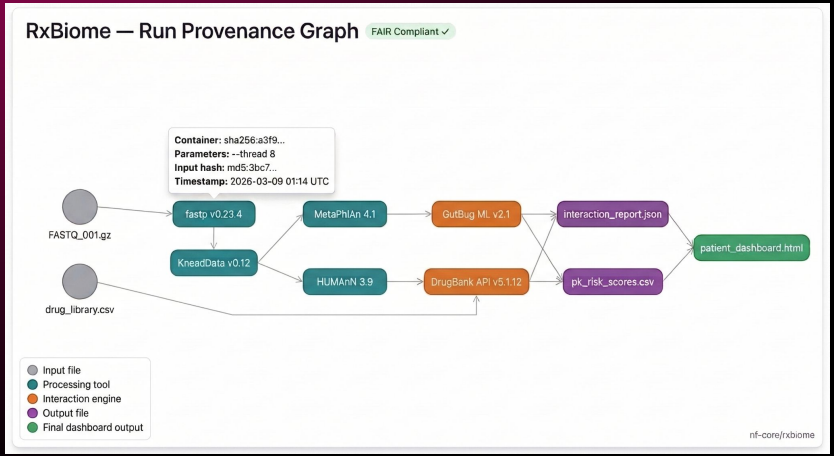
- Individual drug risk scores
- Predicted metabolizing enzymes
- Confidence intervals

Cohort Heatmap

- Population-level risk patterns
- Drug-prioritization by mean risk
- Clinical decision support at scale

*Mockups generated using Nano Banana 2

Every Prediction is Traceable



Provenance Graph

- Full computational lineage
- FAIR-compliant audit trail
- Reproducible in any lab

Key Clinical Insights

- ➔ Drug-specific risk stratification
  - ◆ Not generic “microbiome health scores”
- ➔ Mechanistic attribution
  - ◆ Which bugs? Which enzymes? Why?
- ➔ Actionable dosing guidance
  - ◆ Quantitative PK impact predictions
- ➔ Exportable for EHR integration
  - ◆ Structures JSON-LD outputs

*Mockups generated using Nano Banana 2

The Value Proposition

For Research Labs:

- Cut analysis time from 3 weeks to hours
- Guarantee cross-cohort reproducibility
- Auto-generate publication-ready figures

For Pharma:

- De-risk clinical trials by stratifying patients by microbiome drug metabolism
- Reduce adverse event rates (irinotecan toxicity hospitalizes 1 in 3 patients at $8,200 per admission) [18-20]
- Accelerate pharmacomicrobiome biomarker discovery

For Precision Medicine:

- Enable true personalized dosing (not just genetics - microbiomics)
- Bridge the 5-year gap between microbiome research to clinical translation
- Support regulatory-grade decision making

References

[1] Haiser, H.J. et al. (2013). "Predicting and Manipulating Cardiac Drug Inactivation by the Human Gut Bacterium Eggerthella lenta." Science, 341(6143): 295–298.
[2] Gong, J. et al. (2021). "Microbiota-Host-Irinotecan Axis: A New Insight Toward Irinotecan Chemotherapy." Frontiers in Cellular and Infection Microbiology.
[3] Ünlü, S. et al. (2014). "The Role of Intestinal Microbiota in Development of Irinotecan Toxicity." Toxicology Reports, 1: 268–280.
[4] Perez-Pardo, P. et al. (2019). "Gut bacterial tyrosine decarboxylases restrict levels of levodopa in the treatment of Parkinson's disease." Nature Communications, 10: 310.
[5] Rekdal, V.M. et al. (2019). "Discovery and inhibition of an interspecies gut bacterial pathway for levodopa metabolism." Science, 364(6445): eaau6323.
[6] Routy, B. et al. (2018). "Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors." Science, 359(6371): 91–97.
[7] Derosa, L. et al. (2022). "Intestinal Akkermansia muciniphila predicts clinical response to PD-1 blockade in advanced non-small cell lung cancer patients." Nature Cancer.
[8] Citation: Schloss, P.D. (2018). "Identifying and Overcoming Threats to Reproducibility, Replicability, Robustness, and Generalizability in Microbiome Research." mBio 9(3):e00525-18.
[9] DiMasi, J.A., Grabowski, H.G., Hansen, R.W. (2016). "Innovation in the pharmaceutical industry: New estimates of R&D costs." Journal of Health Economics, 47: 20–33. Reported by the Tufts Center for the Study of Drug Development.

[10] Waring, M.J. et al. (2015). "An analysis of the attrition of drug candidates from four major pharmaceutical companies." Nature Reviews Drug Discovery, 14: 475–486.
[11] Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018). "fastp: an ultra-fast all-in-one FASTQ preprocessor." Bioinformatics, 34(17): i884–i890.
[12] Huttenhower Lab, Harvard T.H. Chan School of Public Health. KneadData v0.12. bioBakery suite. Available at: huttenhower.sph.harvard.edu/kneaddata
[13] Wood, D.E., Lu, J., Langmead, B. (2019). "Improved metagenomic analysis with Kraken 2." Genome Biology, 20: 257.
[14] Blanco-Míguez, A., Beghini, F., et al. (2023). "Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4." Nature Biotechnology, 41: 1633–1644.
[15] Beghini, F., McIver, L.J., et al. (2021). "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3." eLife, 10: e65088.
[16] Wishart, D.S., et al. (2018). "DrugBank 5.0: a major update to the DrugBank database for 2018." Nucleic Acids Research, 46(D1): D1074–D1082.
[17] Malwe, A.S., Srivastava, G.N., et al. (2023). "GutBug: A Tool for Prediction of Human Gut Bacteria Mediated Biotransformation of Biotic and Xenobiotic Molecules." Journal of Molecular Biology, 435(14): 168079.
[18] Arbuckle, R.B., et al. (2004). "The cost of hospitalization secondary to severe chemotherapy-induced diarrhea (CID) in patients with colorectal cancer." *Journal of Clinical Oncology*, 22(14_suppl): 6111.
[19] Quinlan, M.R., et al. (2005). "Estimating the cost of illness in colorectal cancer patients who were hospitalized for severe chemotherapy-induced diarrhea." *Value in Health*, 8(2): A174.
[20] Managing Irinotecan-Induced Diarrhea: A Comprehensive Review. *PMC* 2025.