Research Statement

Shri Vishalini Rajaram
Interdisciplinary Graduate Program in Human Toxicology, University of Iowa

Problem and agenda

Human health is shaped by continuous exposure to diet, medications, and environmental chemicals. These exposures do not act on the host in isolation; they perturb multikingdom microbial communities that in turn modulate risk and resilience. The key control points inside these communities are mobile genetic elements and stress-response circuits. Plasmids, phages, transposons, and integrative elements redistribute functions horizontally, while systems like toxin–antitoxin modules and CRISPR–Cas regulate persistence, immunity, and ecological rearrangement. Today, the literature is rich in sequencing and thin in synthesis. Data are siloed by kingdom and platform, methods are inconsistently reproducible, and exposure signals are rarely linked to defensible mechanisms that travel across cohorts. I address this by integrating multikingdom data under pre-specified statistical contracts so exposure effects, uncertainty, and transport are explicit and reproducible.

Foundations that built the current research

I came to computation through the lab bench. At CSIR–CSMCRI, I learned to treat culture conditions as exposures with measurable consequences and to write modular, versioned analyses that others can rerun. At Madurai Kamaraj University, I moved from compounds to regulatory programs in cancer, which raised my bar for reproducibility and pushed me toward multiscale thinking from exposure to pathway to phenotype. At the Institute of Mathematical Sciences, I matured into systems biology and tool building, leading curation and modeling of secretion machinery with disciplined version control and reusable software, including contributions that became T9GPred. These experiences set my habits now: respect costly experimental data, model exposures as structured perturbations, and engineer analysis as infrastructure rather than an afterthought.

Current work: workflow and development

My PhD operationalizes an exposome–microbiome–host agenda centered on mobile genetic elements and stress-response circuits, with CRISPR and toxin–antitoxin systems as primary modules. I convert public and collaborator meta-omics into defensible evidence by building the statistical contract into the workflow so compositionality, sparsity, longitudinal structure, and uncertainty are handled up front. The goal is portable software, harmonized datasets, and interpretable signatures that transfer across oral-health and cancer testbeds. All projects run in containers on Argon with version-pinned references and audited notebooks.

Microbiome Normalization and Differential analysis for Statistics Enabled Testing (MiNDSET) treats normalization and differential analysis as a scored decision rather than a one-off method. It supports binary, multiclass, and longitudinal designs, including zero inflation and overdispersion routes. A post-analysis scoring function uses standardized report objects to rank pipelines on statistical robustness, biological coherence, and efficiency. I co-developed realistic simulations to stress-test these routes, and every release is auditable, containerized, and fully regenerates figures from a single script.

Multi-Organism Multi-Omics – Microbiome Analysis Pipeline (MoMo-MAP) converts raw short-read data into analysis-ready expression tables for bacteria, fungi, viruses, and archaea. Each kingdom follows a validated path, then merges into a shared schema that records provenance and quality metrics. For metaproteomics, outputs include protein, taxonomic, and pathway abundance tables; comparable modules exist for metagenomics and metatranscriptomics. The design lets collaborators add cohorts and obtain identical, audited outputs without re-engineering.

Bootstrap-Ensemble Attribute Framework (BEAR) delivers compact, stable signatures for translation. It ensembles feature selectors, ranks features by stability, and evaluates nested panels to find the smallest set that preserves

discrimination and calibration. Cross-cohort transfer tests, permutation checks, and performance interval estimates are included. Outputs are simple to adopt: a readable feature list with coefficients, an application script, and diagnostics that flag out-of-scope use.

I analyze CRISPR spaceromes and toxin–antitoxin activity as the transducers of exposure into community structure. For CRISPR, I parse repeat–spacer–repeat arrays, quantify spacer conservation and turnover, and assess sharing patterns in exposure-aware strata, with origin tracing when evidence permits. For toxin–antitoxin systems, I estimate differential activity from curated metatranscriptomes with models that handle compositionality and zero inflation, mapping effects to curated families; figures regenerate from the same fit objects for exact contrasts and labels. Two current hypotheses ground this work. First, sustained high sucrose or acid exposure increases expression of specific toxin–antitoxin families within Streptococcus-dominant consortia; I test this on curated oral metatranscriptomes with calibrated effect sizes. Second, recent antimicrobial exposure increases CRISPR spacer turnover and lowers network clustering while elevating plasmid markers; I quantify turnover and sharing metrics and link them to mobile-element readouts under exposure windows.

Exposome framework and modeling stance

I model exposures as structured perturbations that act through defense and persistence programs to reshape multikingdom function and phenotype. Parsimonious directed acyclic graphs pre-specify identification; estimators match estimands and data-generating processes; partial pooling stabilizes multi-study designs; and uncertainty is propagated into all summaries. Diet, therapeutics, environmental chemicals, and behaviors enter as measured exposures; demography and clinical indices as confounders; batch and site as nuisance; biologic mediators are modeled by design. Data use follows IRB and data-use-agreement requirements with de-identification and a public analysis manifest for transparency. Oral health is my primary testbed with colorectal cancer as the secondary domain so the same methods travel while near-term validation remains concrete.

Future research: Defense ecology under real exposures

Environmental and therapeutic exposures reorganize communities through mobile genetic elements and stress responses, yet the field lacks a harmonized, uncertainty-aware map of these responses. I will assemble raw metagenomes, metatranscriptomes, and metaproteomes with usable exposure windows; in clean-room pipelines detect plasmids, phages, transposons, and integrative elements using assembly graphs, hallmark gene profiles, coverage linkage, and read-pair evidence; parse repeat–spacer–repeat arrays to quantify CRISPR conservation and spacer turnover; and model toxin–antitoxin family activity from expression counts. Multilevel models with partial pooling will separate study effects from exposure contrasts, uncertainty will be propagated to all summaries, and results will be stratified by clinically meaningful exposure windows. Deliverables include a versioned atlas, a public metadata schema, and fully reproducible code and containers, with database pinning, consensus callers, exclusion criteria, leave-one-study-out checks, and negative controls to ensure transportability and independence from any prior lab software.

Causal targets and transportability

Many exposure–microbiome findings remain associative and brittle across sites because estimands are undefined and identification is untested. I will formalize questions with prespecified causal diagrams that distinguish confounders, mediators, and modifiers; define estimands such as the effect of sustained dietary sugar load on toxin–antitoxin activity or the effect of recent antimicrobial exposure on CRISPR spacer turnover; and estimate effects with targeted learning or double, debiased machine learning with cross-fitting, using generalized or zero-inflated mixed models for count outcomes and Bayesian partial pooling when data are sparse or hierarchical. Transportability will be audited by quantifying distribution shift and reweighting to the target site or by hierarchical transport formulas, with calibration and conformal prediction to bound deployment risk and principled missing-data handling recorded in a public manifest. Outputs include a lab-agnostic causal blueprint with reusable templates,

worked case studies on independent testbeds, and a reproducibility paper detailing identification checks, sensitivity analyses, and failure modes.

Translation to compact, interpretable signatures

Clinical and public-health use requires small, stable panels that survive domain shift. I will construct mechanistically anchored features, for example toxin–antitoxin activity scores, CRISPR diversity and network metrics, plasmid or phage markers, and core pathway readouts; combine them with exposure summaries aligned to decision points; and apply stability-led selection with resampling to identify the smallest panel that preserves discrimination and calibration. Models will favor sparse generalized linear frameworks or monotone gradient boosting with calibrated probabilities and transparent coefficients, and will undergo external validation, leave-one-study-out testing, and shift diagnostics with guardrails such as distance-to-training distributions and conformal risk controls. Each testbed will yield at least one panel with no more than twenty features, confidence intervals and decision-curve analyses, a containerized command-line tool for application, and a one-page usage memo that states assumptions and scope, enabling straightforward governance and deployment without dependence on current-lab software.

Milestones

Within 12–18 months I will release the defense-ecology atlas workflow and a first methods manuscript on exposure-linked toxin–antitoxin and CRISPR modeling, including transport diagnostics. Within 24–36 months I will complete the causal module and publish at least one validated compact signature with independent replication. At each stage success is measured by external re-runs that reach the same conclusions, adoption by other groups, and clear documentation of limits when transport fails.

Conclusion

This program treats the exposome as perturbations that move through mobile genetic elements and CRISPR to reorganize multikingdom function, and it delivers the infrastructure to measure, test, and translate those effects. It will leave the field with a harmonized, uncertainty-aware atlas of defense ecology under real exposures, a rigorously specified causal framework with transport diagnostics, and compact, interpretable signatures designed for clinical and public-health use. All artifacts will be released as clean-room, containerized, version-pinned resources that reproduce outside my group and remain independent of current-lab software. The work is collaboration-ready and disease-focused, so partners can slot datasets into audited workflows and obtain defensible, portable results. For the department, the value is immediate: reusable software and datasets, rapid and transparent collaboration, and a clear path to foundation and consortium support for open, translational methods. The outcome is straightforward and durable: turn messy meta-omics into reliable knowledge and practical tools others can adopt without reverse-engineering.