Sheri Sanders, Ph.D.

Bioinformatics | Computational Biology | Research Software Engineering
574.317.5932 | sheri.anne.sanders@gmail.com | kallistaindustries.com

Professional Summary

Over the last decade, I have built nationally recognized research support programs that combine software development, scalable computational workflows, and collaborative scientific consulting. My work has focused on developing reproducible software, scalable computational workflows, and cloud/HPC infrastructure that enable researchers to analyze complex genomic datasets. I have contributed to research in neurodevelopmental disease, cancer modeling, bioacoustics, and evolutionary theory.

Selected Achievements

- Built national bioinformatics infrastructure supporting over 3,000 researchers.
- Developed reproducible software and computational workflows supporting 18 manuscripts, 8 dissertations and 39 presentations.
- Managed 300+ scientific software packages with ~3 million annual uses.
- Increased sequencing core bioinformatics revenue by 300%.
- Trained 1,200+ researchers through national cloud-based workshops.
- *Full CV and linked products available at kallistaindustries.com*

Core Skills

Genomics

- NGS / single-cell / spatial transcriptomics
- Statistical genomics (GWAS, Multi-omics, Population Genetics Modeling)
- Machine learning and neural networks in biological data analysis (scikit-learn, PyTorch)

Computing

- Workflow engineering (Nextflow/Snakemake/custom)
- Containers and reproducible computing (Apptainer, docker, conda, git)
- HPC/cloud infrastructure (SLURM, ACLs, GCP, Jetstream2, AWS)
- Scientific software development (R, python, bash, SQL)
- Version control (Git)
- Linux Administration (Ubuntu, RHEL)

Leadership

- Grant and research collaboration (NIH, NSF, USDA, USGS)
- Project management, contract management

Experience

Owner and Consultant, Kallista Consulting
2021-current

- Provide contract-based bioinformatics consulting, research software engineering, and statistical analysis for academic collaborators.

- Continue development, documentation, tutorials, and deployment of dockerized comparative genomics platform for emerging model systems.
- Provide statistical analysis for complex analytical projects, including multi-variate decomposition, Bayesian statistical analysis of genomics, pooled population genomics, and multi-omics.
- Develop reproducible and flexible computational tools and workflows integrating containers, version-controlled code, and automated deployment.
- Manage client contracts and billing across six collaborating universities.

Associate Professor of the Practice – Bioinformatics, University of Notre Dame
Associate Director, Genomics and Bioinformatics Core Facility, University of Notre Dame
2022-2025

- Secured funding for dedicated high memory compute nodes to support genomics research
- Grew bioinformatic service income by 300% and managed bioinformatics requests for the core.
- Collaborated on 3-4 research projects annually as a contributing author or Co-PI, spanning projects in neurodevelopmental disorders, infectious disease, microbiomes, and soil ecology.
- Designed and implemented courses in CRISPR technologies, spatial and single-cell analyses, machine learning, and applied bioinformatics.
- Designed and built comparative genomics platform with integrated, GUI-based tools to allow exploration of published and novel genomic data on emerging systems.
- Designed, modernized, and managed data transfer, networking, and storage infrastructure supporting high-throughput sequencing workflows and client data delivery.

Director, National Center for Genome Analysis Support, Indiana University
Sn. Bioinformatic Analyst, Research Technologies, Indiana University
2016-2021

- Built and led national-scale bioinformatics support infrastructure serving 3000 genomic researchers in all 50 states and Puerto Rico.
- Resolved approximately 300 annual technical support requests involving bioinformatics software, HPC infrastructure, workflow deployment, and genomic data analysis.
- Generated and consulted on 4-6 research projects with over 80 billable hours each, including machine learning projects in ecology, remote sensing, and cancer treatment outcomes.
- Managed ~300 scientific software packages on HPC systems with ~3 million annual uses
- Developed containerized workflow, one of which supported 12 publications and 8 dissertations.
- Led cloud-based training initiatives reaching more than 1200 researchers nationally.
- Mentored twelve undergraduate and graduate researchers in bioinformatics, statistical analysis, and computing.
- Managed renewal, reporting, and administration of an approximately $1 million NSF-funded national bioinformatics support program.

Selected Technical Projects

1. Dockerized Comparative Genomics Platform (ongoing)
Designed and developed a fully containerized comparative genomics platform integrating JBrowse2, BLAST, CRISPR guide design, genome annotation, and downstream analysis workflows into a unified web-based resource. The platform automates deployment, configuration, and data integration from NCBI, enabling researchers to rapidly publish and explore new genomic resources. Documentation, tutorials, and source code are available at

- github.com/kallistaconsulting/genomic_resources_clogmia

Associated Publications
Amiri et al. (2026). PLOS Biology. https://doi.org/10.1371/journal.pbio.3003896.
Tenger-Trolander et al. (2025). Development. https://doi.org/10.1242/dev.204732

2. Bayesian Modeling and Deconvolution of Complex Population Genomics Data (ongoing)
Developed a probabilistic framework to identify genetic factors contributing to survival and adaptation in experimentally evolved populations using pooled sequencing data. The project addressed a common challenge in genomic analysis: separating biological signals from uncertainty caused by sequencing depth, sampling variation, and mixed populations.

Implemented Bayesian allele frequency models using beta-binomial frameworks to estimate genomic changes with uncertainty rather than relying on single point estimates. Developed Monte Carlo approaches to evaluate allele frequency changes while incorporating confidence intervals and measurement uncertainty.

Extended the workflow with Bayesian mixture deconvolution to estimate the relative contribution of individual genetic lineages within pooled samples. Combined clone-level genotype information with population-level sequencing data to partition changes in population composition and evaluate contributions from selection, drift, laboratory effects, and experimental conditions.

This work demonstrates the application of Bayesian statistics, mixture modeling, and reproducible computational workflows to extract interpretable biological signals from noisy, high-dimensional genomic datasets. Products from this project are in preparation.

3. Containerized Transcriptome Assembly Workflow and Workshop (2018-2022)
Developed and optimized a combined de novo transcriptome assembly pipeline that merges multiple assemblers and multiple k-mers to increase accuracy and reduce individual biases. The workflow integrates quality control, assembly, annotation, and analysis of differential expression into an end-to-end analysis framework driven by automatically installed biocontainers. The workflow was used in dozens of projects and collaborations as it substantially decreased computational and analysis time for first-time analysts.

The workflow became the foundation of a national workshop series, training 82 participants from 43 institutions over 26 states/territories and two international institutions. Code, presentations, and documentation available at

- https://github.com/kallistaconsulting/Transcriptome_Assembly_Workshop_Spring_2019

- https://ittraining.iu.edu/explore-topics/titles/denovo/index.html

Select Associated Publications

Lomheim et al. (2025). Evolution and Development. https://doi.org/10.1111/ede.70014
McManuset al. (2025). Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2025.1709239
Brooks et al. (2024). Association for Computing Machinery. https://doi.org/10.1145/3626203.3670605
Wenninger et al. (2025). bioRxiv. https://pubmed.ncbi.nlm.nih.gov/40964355/

4. Modeling White Blood Cell Levels to Prevent Radiation-induced Leukopenia (2020)
Developed a proof-of-concept machine learning model to predict the risk of clinically significant white blood cell depletion in cancer patients undergoing radiation therapy series. Identified relevant clinical features (e.g. point in series, anatomical target site, and exposure time) to evaluate patient-specific risk using predictive modeling techniques. The project was completed in collaboration with a medical researcher as the capstone for my graduate certificate in Data Science. Source code and data are under embargo.

Education

Graduate School

- Graduate Certificate in Data Science (Machine Learning), Indiana University (2021)
- Ph.D. Biology, University of Notre Dame (2016)
- M.S Biology, University of Texas at Tyler (2010)

Additional Education

- Sustaining Biological Infrastructure, ESA (2020)
- Advanced Linux System Administration, LCI (2018)
- Communication Training for Scientific and Technology Issues, COMPASS (2015)

Selected Grants and Research Leadership

NSF ABI 1759906 (PI - 2020, Co-PI 2018-2019) - National Center for Genome Analysis Support Sustaining award to continue NCGAS services ($962k)

NIH R01GM127366 Subaward (PI) - Genetic variation and function of body axis determinants in midges and other flies (~$136k) (moved to contract at ND, then KC)

NSF DBI-2022049 Subaward (Co-PI) - Genomics and Eco-evolution of Multi-Scale Symbioses (GEMS), to train students on bioinformatic and genomic analyses ($123k)

Equipment Renewal and Replacement Grant, Notre Dame Research (PI) – funding to purchase and administer three dedicated 2TB memory nodes for genomic research support ($50k)