Samir Rachid Zaim, PhD

samir.rachid.zaim@gmail.com · (520) 719-9224 · [LinkedIn](#) · [GitHub](#)

Applied scientist with deep expertise in machine learning, statistics, and high-dimensional biomedical data, with hands-on experience in the multi-omics space across immunology, infectious diseases, and oncology. Technical leader with hands-on IC and people-management experience, known for translating research into validated, scalable solutions. Proven collaborator across science, informatics, and product teams in high-velocity, ambiguous environments.

Professional Experience

Supervisor, Biostatistics *Jan 2025 - Present*

Allen Institute, Immunology. Seattle WA
- Owned statistical strategy for 10+ translational and clinical studies, helping define analysis plans, success metrics, and timelines across early- to late-stage research environments
- Acted as a thought partner on technical roadmap and strategic initiatives, aligning applied AI/ML research with organizational priorities
- Led end-to-end analysis of a high-impact oncology study; presented findings to leadership and advisory board; and delivered a priority project
- Managed and mentored a team of 3 bioinformaticians, driving execution across 5 cross-functional clinical research programs while maintaining high scientific standards

Biostatistician, II 10/2023-12/2024
- Served as senior IC providing biostatistics expertise across 10+ in-house clinical studies spanning proteomics, transcriptomics, flow cytometry, clinical data, and epigenetic datasets
- Collaborated with clinicians, wet-bench scientists, and external vendors to design validation pipelines for novel diagnostic proteomics kits supporting a biomarker patent application
- Led end-to-end peer-reviewed publication and patent application with the U.S. Department of Defense, developing predictive proteomics models for early disease onset

Biostatistician, I 3/2021-09/2023
- Secured funding and hired intern to evaluate 'omics-based subtyping in rheumatoid arthritis
- Conceptualized, designed, benchmarked and published epigenetic prediction software package
- Designed and analyzed a large-scale survey study (n = 1,500) assessing Long COVID outcomes in Latino populations, translating findings into actionable public-health insights

Center for Biomedical Informatics & Biostatistics Researcher *8/2016 - 1/2021*

Univ. of Arizona, Center for Biomedical Informatics and Biostatistics, Tucson AZ
- Led and contributed to 8 novel methods in statistical machine learning approaches in medicine
- Recognized by the [IMIA yearbook on Bioinformatics](#) as a top research contributor in 2018

Collegiate Fellow & Data Scientist *7/2014 - 5/2016*

Parkland Center for Clinical Innovations & Pieces Technologies, Dallas TX
- Designed a risk-surveillance dashboard to identify at-risk asthmatic children
- Co-developed a sepsis monitoring surveillance algorithm to triage patients and improve care
- Deployed a predictive model to identify 30-day readmissions to optimize resource allocation

Core Skills

Biostatistics and Machine Learning Experience:
- Cross-sectional & longitudinal data clinical studies, study design and power analysis, EHR data mining, statistical modeling, classifier development, mixed effect modeling, GLMMs, survival modeling, regression, survey design, survey data analysis, tensor decomposition, supervised & unsupervised learning, high-dimensional data clustering, high-dimensional feature selection, high-dimensional data visualization, statistical & ml software development

Programming Experience:
- Language: R (10+ yrs), Python (3+ yrs), SAS & STATA (1-2 yrs), SQL (1-2 yrs),
- Environment: Linux (> 5 yrs), HPC (5 yrs), Google Cloud Platform (3+ yrs)

Technical Leadership Experience:
- Statistical consulting, Statistical Analysis Plans, Software benchmarking, project management, cross-functional statistical collaborations

Technical Leadership Management:
- Requirements gathering, milestone tracking, resource & budget management
- Cross-functional leadership, technical project planning, technical
- Jira, Trello, Git/GitHub,Tableau, Slack, Confluence, Jupyter Notebooks, R Markdown

Patents

BIOMARKERS TO PREDICT ONSET OF CLINICAL RHEUMATOID ARTHRITIS, U.S. PROVISIONAL PATENT Application No. 63/597,631

Awards, Honors, and Recognition

Editor's Pick Articles: Arthritis & Rheumatology, September 2025 Edition	07/2025
Commissioned Editorial, Arthritis & Rheumatology	09/2025
Post-baccalaureate Research Grant (> $57,500), Allen Institute, WA	06/2023
Top Contributor in Bioinformatics, International Medical Informatics Association, France	2018
Center for Biomedical Informatics and Biostats Fellowship: (~$113,336), Univ. Arizona.	06/2016

Open-source Statistical and Machine Learning Software Packages: Lead Developer

MOCHA: a statistical algorithm for functional genomic inference: CRAN
binomialRF: a machine learning tool to identify genomic biomarker and interactions: CRAN, GitHub

Selected Publications (Google Scholar)

- "Serum proteomic signatures before the diagnosis of rheumatoid arthritis..." *Arthritis & Rheumatology* 77.9 (2025)**. Led biomarker study in collaboration with the Dept. of Defense**
- "MOCHA's advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human cohorts." *Nature Communications*(2024). **Novel statistical software**
- "binomialRF: interpretable combinatoric efficiency of random forests to identify biomarker interactions." *BMC bioinformatics*(2020). **Novel machine learning software development.**

Education

2021 Ph.D.: Statistics**,** *Univ. of Arizona**,** Tucson, Arizona*
2014 BA: Mathematics/Statistics, *Carleton College, Northfield, MN*