Samir Rachid Zaim, PhD

samir.rachid.zaim@gmail.com · (520) 719-9224
LinkedIn · Github · GoogleScholar

Professional Summary

Biostatistician and research supervisor with over ten years of experience leading and contributing to more than ten completed research projects, publications, and a provisional patent. Demonstrated expertise as a biostatistician with impactful research, eight first author publications in biomedical informatics, public health, clinical, and scientific journals, six collaborative publications, and one provisional patent. I am a recipient of six professional awards and an internal research grant, an active scientific reviewer and contributor in leading biomedical journals, and I currently serve as a biostatistician supervisor leading publications, analysts, and biostatistics best-practices at the Allen Institute's Immunology moonshot.

Education

PhD, *Univ. of Arizona, Tucson, AZ* 8/2016 – 2/2021
Major: Statistics**; Minor:** Systems and Industrial Engineering
Thesis*:* Improving efficiencies of genomic feature engineering and selection through knowledge-anchored machine learning
PhD advisors: Yves A Lussier, MD, FACMI; Helen H Zhang, PhD

BA, *Carleton College, Northfield, MN* 2010 – 2014
Major: Mathematics/Statistics**; Concentration:** French & Francophone Studies
Thesis*:* Dynamic Models of Geologic Data from Ridge-transform Intersections
Major advisor: Robert Dobrow, PhD

Awards, Honors, and Recognitions

Editor's Pick Articles: Arthritis & Rheumatology, September 2025 Edition 07/2025
Commissioned Editorial, Arthritis & Rheumatology 04/2025
Post Baccalaureate Research Assistant Grant (> $57,500)
Allen Institute, Seattle WA 06/2023- 5/2024
Data Science Ambassador: ($1000)
Univ. of Arizona College of Medicine, Tucson AZ 9/2019- 5/2020
Candidate Best Paper: Bioinformatics and Translational Informatics
International Medical Informatics Association 2018
Center for Biomedical Informatics and Biostatistics Fellowship: (~$113,336)
Univ. of Arizona Health Sciences, Tucson AZ 6/2016- 5/2018
Frank Karel Fellowship in Public Interest Communications: ($3,000)
NonProfit Roundtable, Washington D.C. 6/2013
Kelloggs International Scholar: (~ **$180,000**)
Carleton College, Northfield MN 9/2010-6/2014

Languages

Spanish (Native), **English** (Fluent), **French** (Fluent), **Arabic/Darija** (Proficient)

Professional Experience

The Allen Institute, Immunology *Seattle WA*

Supervisor, Biostatistics 01/2025 – Present

Biostatistician II 10/2023 – 12/2024

Biostatistician I 3/2021 – 10/2023

In my role, I serve as a biostatistician research supervisor, leading methodology development & implementation in clinical studies, and mentor a team of master's-level analysts. As the study biostatistician, and have developed, authored, and reviewed over ten different data analysis plans across our institute, and have published three manuscripts as lead author, co-authored three immunology publications, and technically supervised three additional projects. My recent works 1) examining the impact of Long COVID in Washington state, and 2) a collaboration with the department of defense assessing the development of rheumatoid arthritis has been highlighted by journal editors & media.

- Led a multi-center collaboration with the Department of Defense identifying protein signatures in preRA
- Filed a provisional patent based on serum protein biomarkers that can help predict the onset of RA
- Co-led survey study with Univ of Washington assessing impact of Long COVID on Latinos as a co-investigator
- Provided statistical expertise and advisory in the design and analyses > ten in projects & publications
- Implemented processes for data analysis plan review across cross-disciplinary teams to align efforts
- Managed a team of 3 junior bioinformaticians across cross-functional teams and immunology studies
- Published a novel open-source statistical R package for functional genomic inference in scATAC
- Participated as an invited speaker at public health, biostatistics, and rheumatology conferences
- Won 1/6 Internal Allen Institute awards ('23) to study underrepresented populations in RA and Long COVID
- Provided study design for NIH grant (funded) for follow-up proteomics study of an RA prevention trial

Univ. of Arizona, Center for Biomedical Informatics and Biostatistics, *Tucson AZ* 8/2016 – 1/2021

Graduate Research Assistant

As a graduate researcher for the Center for Biomedical Informatics and Biostatistics (CB2), I was trained to conduct independent scientific research and communications using data, statistics, and technology to provide software solutions to biomedical and translational informatics. Collaborated with Biopharma on implementing personalized statistical software in clinical studies.

- Published four first author publications in statistics methodology and applications in biomedical research
- Established individual nonprofit volunteering collaborations with Code for Venezuela and iMMAP
- Scientific communicator with more than 10 conference talks & posters, two career columns in *Nature Magazine*, and featured profile in *Hearts on Venezuela* as a Venezuelan academic and migrant
- Developed an open-source software package to identify biological interactions using gene expression data
- Conducted pilot studies to develop stronger statistical frameworks for precision medicine research
- Taught statistical and machine learning literacy workshops to make health data science research accessible

Parkland Hospital & Health System - Center for Clinical Innovations / Pieces Tech, *Dallas TX* 7/2014 – 5/2016

Collegiate Fellow & Data Scientist

In two years, I participated in the research, development, and deployment of over six risk-surveillance models across hospital systems in Texas and the USA, first as collegiate fellow for Parkland Hospital (the non-profit research branch) and later as a data-scientist for Pieces Technologies (sister-for-profit entity).

- Designed and implemented a risk-surveillance dashboard to identify asthmatic children prone to ER visits
- Collaborated with ER physicians to develop a sepsis monitoring surveillance algorithm to prevent deaths
- Deployed a predictive model to identify 30-day readmissions to optimize resource allocation at discharge

- Developed of a natural language processing tool that identified COPD complications using clinical notes

The Education Trust, *Washington D.C.* 6/2013 – 8/2013

Frank Karel Fellow in Public Interest Communications

- Awarded a public interest communications fellowship from the Non-Profit Roundtable in Washington D.C.
- Presented research findings and motivation for using data to better inform and improve education policy and programs for minority students at the Frank Harel Gathering at the University of Florida (video-link)

Scientific Reviewer

University of Washington Latino Center for Health, Grant Reviewer	02/2025
Allen Institute, Next Generation Scientific Leaders: Reviewer	7/2024
Nature Communications: Co-reviewer	11/2023
European Journal of Medical Research: Reviewer	6/2023
AMIA 2020 Informatics Summit: Reviewer	8/2019
Computers in Biology and Medicine: Reviewer	5/2019
AMIA Annual Summit 2019: Reviewer	3/2019
ISMB/ECCB 2019: Sub-reviewer	2/2019
International Conference on Health Informatics: Sub-reviewer	6/2016
International Conference on Health Informatics: Sub-reviewer	6/2015

Publications in Progress

([1] Denotes co-first author)

1. Chander, A[1],.; **Rachid Zaim, Samir[1]**, et al. “Longitudinal analysis of Newly Diagnosed Multiple Myeloma: Bone Marrow Recovery and Immune microenvironment dynamics in the context of Tumor Heterogeneity.”
    a. Target: Cancer Cell
    b. Expected submission: Dec 2025
2. Moss, Nicholas, …, **Rachid Zaim, Samir,** …, Kuan, Emma. “A new cell-type specific human interferon response atlas reveals differential contributions of INF-I and IFN-II in disease.”
    a. Target: Nature Immunology
    b. Expected submission: Dec 2025

Journal Publications

3. Gong[1],, Q., Sharma[1],, M., Kuan, E. L., Glass, M. C., Chander, A., Singh, M., ..., **Rachid Zaim, Samir**, … & Gustafson, C. E. "Multi-omic profiling reveals age-related immune dynamics in healthy adults." *Nature* (2025): 1-11.

4. He Z[1],, Glass MC[1], .., **Rachid Zaim, Samir**, …, Goldrath AW, Bumol TF, Li XJ, Holers VM, Skene PJ, Savage AK, Firestein GS, Deane KD, Torgerson TR, Gillespie MA. "Progression to rheumatoid arthritis in at-risk individuals is defined by systemic inflammation and by T and B cell dysregulation." *Science Translational Medicine* 17.817 (2025): eadt7214.
5. **Rachid Zaim S**, Savage AK, Gillespie MA, Castillo JD, Bennett C, Torgerson TR, Becker LA, Mahler M, Moss L, Feser ML, Edison JD, Mikuls TR, Holers VM, Li XJ, Deane KD. Serum Proteomic Signatures Before the Diagnosis of Rheumatoid Arthritis: Evolving Biologic Pathways and Specific Periods of Disease Development. Arthritis Rheumatol. 2025 Sep;77(9):1166-1178. doi: 10.1002/art.43175. Epub 2025 May 13. PMID: 40189919.
6. **Rachid Zaim, S.,** Castillo, J.D., Cabrera, A. *et al.* Long COVID among Latino Patients of Two Federally Qualified Health Centers in Washington State. *J GEN INTERN MED* (2025). https://doi.org/10.1007/s11606-025-09732-y
7. **Rachid Zaim[1], Samir,** Pebworth[1], MP**,** et al. "MOCHA: Advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human disease cohorts." *Nature Communications*: 2023-06.

8. Thomson, ZJ, …, **Rachid Zaim, Samir,** et al. "Tri-modal single cell profiling reveals a distinct pediatric CD8αα T cell subset and broad age-related molecular reprogramming across the T cell compartment." *Nature Immunology* (2023): 1-13.
9. Dillon Aberasturi, Nima Pouladi**, Samir Rachid Zaim**, Colleen Kenost, Joanne Berghout, Walter W Piegorsch, Yves A Lussier, 'Single-subject studies'-derived analyses unveil altered biomechanisms between very small cohorts: implications for rare diseases, Bioinformatics, Volume 37, Issue Supplement_1, July 2021, Pages i67–i75, https://doi.org/10.1093/bioinformatics/btab290
10. **Rachid Zaim**, S.; Kenost, C.; Zhang, H.H.; Lussier, Y.A. Personalized Beyond Precision: Designing Unbiased Gold Standards to Improve Single-Subject Studies of Personal Genome Dynamics from Gene Products. *J. Pers. Med.* **2021**, *11*, x. https://doi.org/10.3390/xxxxx
11. **Rachid Zaim, S.**, Kenost, C., Berghout, J. et al. binomialRF: interpretable combinatoric efficiency of random forests to identify biomarker interactions. *BMC Bioinformatics **21,** 374 (2020).*
12. **Rachid Zaim, S**.; Kenost, C.; Berghout, J.; Zhang, HH; Lussier, YA. "Evaluating single-subject study methods for personal transcriptomic interpretations to advance precision medicine." *BMC Medical Genomics 2019,12 (Suppl 5):96.*
13. **Zaim, Samir Rachid**, Qike Li, A. Grant Schissler, and Yves A. Lussier. "Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous transcriptomic responses." In *Pac. Symp. Biocomput*, vol. 23, pp. 484-495. 2018.
14. Kim, Ahyoung Amy, **Rachid Zaim, S**, and Vignesh Subbian. "Assessing Reproducibility and Veracity across Machine Learning Techniques in Biomedicine: A Case Study using TCGA Data." *International Journal of Medical Informatics* (2020): 104148

Peer-reviewed Conference Proceedings

15. Boyd, Andrew D., Colleen Kenost Jianrong' John' Li, **Samir Rachid Zaim**, Jacob Krive, Manish Mittal, Richard A. Satava, Michael Burton, Jacob Smith, and Yves A. Lussier. "ICD-10 procedure codes produce transition challenges." *AMIA Summits on Translational Science Proceedings* 2017 (2018): 35.B
16. Li, Qike[1];**Rachid Zaim, Samir**[1], Aberasturi, Dillon; Berghout, Joanne; Li, Haiquan; Vitali, Francesca ; Kenost, Colleen ; Zhang, Helen Hao; Lussier, Yves A. "Interpretation of Omics dynamics in a single subject using local estimates of dispersion between two transcriptomes." *AMIA Annual Symposium on Translational Science Proceedings* (2019).

Peer-reviewed Conference Abstracts and Posters

1. **Rachid Zaim, Samir,** Cabrera, Andrea, et al. "Determining the Health and Economic Impact of Long COVID in WA Latinos." (2024) NWRPCA Community and Migrant Health Conference.
2. Rachid Zaim, Samir, et al. "Proteomic Signatures in Pre-Rheumatoid Arthritis Suggest Evolving Biological Pathways in Different Stages of Disease Development That May Inform Prediction and Prevention Strategies." (2023) ACR Conference.
3. **Rachid Zaim, Samir,** Cabrera, Andrea, et al. "Determining the Health and Economic Impact of Long COVID in WA Latinos." (2023) WSPHA.
4. Pebworth, Mark-Phillip, **Rachid Zaim, Samir**, et al. MOCHA: advanced statistical modeling of scATAC-seq data enables functional genomic inference in large human disease cohorts. ICSB 2023.
5. Gillespie, M., **Rachid Zaim, Samir,** et al. "OP0185 IMMUNOPHENOTYPIC ABNORMALITIES REMINISCENT OF ESTABLISHED RHEUMATOID ARTHRITIS ARE PRESENT IN ACPA (+) AT-RISK INDIVIDUALS." (2023): 122-122.

6. Fan, Jung-wei, **Samir Rachid Zaim**, Walter W. Piegorsch, Jianrong Li, Colleen Kenost, and Yves A. Lussier. "Systems of Phenome-Exposome Associations Unveiled by Mining Practice-Based Evidence with Environmental Exposures." In *AMIA*. 2018.
7. **Zaim, Samir Rachid**; Kim Ahyoung Amy; Zhang, Helen Hao; Lussier, Yves A; Subbian, Vignesh. "DSL-TEACH: Data Science Literacy Training to Enhance Approaches for Clinical decision-making in Healthcare." *AMIA Annual Symposium on Translational Science Proceedings* (2019)

Scientific, Volunteering, and Media Outreach

1. Seattle Researchers say more resources are needed for Latinos with Long COVID. **Washington Latino News Network** 06, 2024
2. Johanna, Figueira, and **Rachid Zaim Samir**. "How a Team of Venezuelan Expats Is Fighting COVID-19 at Home." *Nature (2020)*
3. **Rachid Zaim, S**. "Venezuelan migrants | Samir Rachid Zaim: "Better times will come and this crisis will help us learn to value life and what we have". *Hearts on Venezuela (2020)*
4. Rachid Zaim, Samir. "The Leaders of Tomorrow". https://vimeo.com/98656142 Frank Karel Fellowship in Public Interest Communications (2014)
5. Rachid Zaim, Samir. "A story of Cookies n'crème". Frank Karel Fellowship in Public Interest Communications (2016)
6. **Rachid Zaim, S**. "Seeking mental-health help was the best thing I did at grad school." *Nature* (2020)

Teaching Experience

Univ. of Arizona Health Sciences, Workshop Instructor, Tucson, AZ	4/2019 – 1/2020
Univ. of Arizona Sports Analytics, Graduate Mentor, Tucson, AZ	1/2018 – 5/2018
Carleton College French Department, Teaching Assistant, Northfield, MN	9/2012– 6/2014
Carleton College Spanish Department, Teaching Assistant, Northfield, MN	9/2011– 9/2013
Carleton College Math Department, Mathematics Tutor, Northfield, MN	1/2011 – 12/2011

Panel Talks

WNAR 2025 – Mentorship Programs for the Next Generation of Statisticians, Whistler BC	06/2025
Allen Institute – Fred Hutch EPIC 2024, Seattle, WA	05/2024
Univ. of Arizona Research Bazaar (virtual), Tucson, AZ	05/2020

Invited Talks

Northwest Primary Care Association: Community and Migrant Health 2024 02/2024
- Determining the Health and Economic Impact of Long COVID in WA Latinos

Latine Long COVID Symposium—study findings & pathways to health 11/2023
- Determining the Health and Economic Impact of Long COVID in WA Latinos

American College of Rheumatology 11/2023
- Proteomic Signatures in Pre-Rheumatoid Arthritis Suggest Different Stages of Disease Development That May Inform Prediction and Prevention Strategies

Washington State Public Health Association 10/2023
- Determining the Health and Economic Impact of Long COVID in WA Latinos

Epic Sciences, Sci & Tech Round Table, San Diego, CA 12/2020
- Ontology-anchored Machine learning for Precision Medicine

Pieces Technologies, Dallas, Tx 11/2020
- Ontology-anchored Machine learning for Precision Medicine

The Allen Institute for Immunology, Seminar Talk, Seattle, WA 10/2020
- Ontology-anchored Machine learning for Precision Medicine

Univ. of Arizona Research Bazaar, Tucson, AZ 5/2020
- Building R packages: best practices for reproducible, open-science

AMIA 2019 Annual Symposium, Washington D.C. 11/2019
- Interpretation of Omics dynamics in a single subject using local estimates of dispersion between two transcriptomes.

Arizona State University Machine Learning Day, Phoenix, AZ 4/2019
- binomialRF: Scalable Feature Selection and Interaction Screening for Random Forests

Univ. of Arizona College of Medicine Research Day, Tucson, AZ 2/2019
- Pathway biomarkers as starting points for reproducible genomics research

Translational Bioinformatics Conference*, Seoul, South Korea 10/2018
- Evaluating single-subject study methods for personal transcriptomic interpretations to advance precision medicine

Pacific Symposium on Biocomputing Conference*, Hawaii, Hawaii 1/2018
- Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous transcriptomic responses

Carleton College Research Symposium, Northfield MN 10/2012
- Examining race relations in France through soccer and sports immigration

Invited Poster Presentations

Univ. of Arizona GIDP Showcase, Tucson AZ 12/2019
- Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous transcriptomic responses

AMIA 2019 Annual Symposium, Washington D.C. 11/2019
- **DSL-TEACH:** Data Science Literacy Training to Enhance Approaches for Clinical decision-making in Healthcare

Univ. of Arizona Information Technology Summit, Tucson, AZ 10/2019
- Improving interpretation in feature selection & interactions in Random Forests

Univ. of Arizona TRIPODS Data Science Conference, Oracle, AZ 5/2019
- binomialRF: Scalable Feature Selection and Interaction Screening for Random Forests

Univ. of Arizona Discover BIO5 Research to Innovation Showcase, Tucson AZ 4/2019
- Assessing single-subject pathway biomarkers for reproducible genomics research

Univ. of Arizona College of Medicine Molecular Biomarker Symposium, Phoenix AZ 12/2018
- Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous transcriptomic responses

Master Student Theses Directed

1. Colleen Kenost (MPH), College of Public Health, University of Arizona 2019

Guest Lectures, Workshops, and Department Seminars

Univ. of Arizona College of Medicine, Machine Learning Workshop, Tucson AZ 4/2019 – Present
- Lead instructor for “Orange for Beginners”

Univ. of Arizona, Statistics GIDP Student Seminars Tucson AZ 10/ 2018
- Website Workshop for Academics & Data Scientists using GitHub Pages

Univ. of Arizona, **Department of Mathematics,** Sports Analytics**,** Tucson AZ 04/2018

- Introduction to Decision Trees and Random Forests in Sports Analytics

Service, Volunteering & Leadership

Member, WNAR, WNAR Engagement Committee 01/2025 – present
Translator, Hearts on Venezuela, Venezuela 10/2019 – 03/2021
Data Science Ambassador, Data Science Institute, Univ. of Arizona, Tucson, AZ 9/2019 – 6/2020
Founder, Mu Sigma Rho Chapter (Statistics National Honor Society), Univ. of Arizona, Tucson, AZ 10/2018 – 02/2021
Program Co-founder, Buenos Aires (Latino Outreach Initiative), Carleton College 9 /2013 – 6/2014
Program Director, Adult ESL, Carleton College 9 /2011 – 6/2014
Volunteer, Firebellies (Science of Cooking), Carleton College 9 /2013 – 6/2014

Open-source Software Packages

MOCHA: a statistical algorithm for modeling regions of accessible chromatin in the genome
binomialRF: a machine learning tool to identify genomic biomarker and interactions available on CRAN and GitHub
referenceNof1: a bioinformatics tool to optimize reference standards in N-of-1 studies; available on GitHub
binomialRF: a python implementation of binomialRF; available on GitHub and PyPi