Saber Hafezqorani, PhD

saber.hafezqorani@gmail.com | saberhq.com | linkedin.com/in/saberhq | github.com/saberhq | Google Scholar

Bioinformatics scientist with 12+ years of R&D at the intersection of AI × Biology. Built solutions for NGS data, including a deep learning model for sequence representation (ntEmbd) and a multi-omics simulation framework (NanoSim). Developed generative and discriminative algorithms for high-throughput sequence data, multimodal integration, and statistical modeling of gene regulatory networks for causal inference and perturbation analysis. Mentored students, collaborated globally, and published 9 papers in top-tier journals. Passionate about applying AI to drive drug discovery.

EXPERIENCES

Doctoral Research Assistant Vancouver, BC, Canada
BC Cancer Genome Sciences Centre *Sep 2017 — Aug 2024 (6 years, 11 months)*

- Designed and implemented ntEmbd, a deep learning foundation model for nucleotide sequence representation using an unsupervised autoencoder architecture, outperforming state-of-the-art tools in RNA classification tasks
- Co-developed NanoSim, a Nanopore multi-omics long-read synthetic sequence generation pipeline (56,000 downloads, 270+ GitHub stars, 270+ citations), enabling tool benchmarking across diverse research areas
- Co-authored 4 research papers on single-cell transcriptome visualization, long-read sequencing assembly, simulation, and benchmarking, published in *Nature Communications* and *Nature Methods* (150+ citations)
- Contributed to the LRGASP consortium, collaborating with 85+ interdisciplinary researchers from cross-functional teams to benchmark long-read RNA-seq tools for transcript identification and quantification
- Maintained and documented pipelines using workflow management tools, deployed across distributed, HPC (Slurm) and cloud environments (GCP) with robust Git-based versioning, automated testing, and code reviews
- Mentored graduate students whose work was presented at international conferences and earned travel awards

Research Associate Ankara, Turkey
METU Informatics Institute *Sep 2015 — Aug 2017 (2 years)*

- Developed RBPSponge, a web tool for genome-wide identification of lncRNAs that sponge RNA-binding proteins (RBPs), integrating ENCODE eCLIP data, differential gene expression, and motif enrichment analysis
- Designed machine learning frameworks for patient stratification and clinical outcome prediction in neuroblastoma using multimodal data integration (RNA-seq, microarray and aCGH)
- Analyzed alternative polyadenylation-mediated 3′ UTR shortening and resulting miRNA binding site loss, using TCGA RNA-seq data, highlighting these events as prognostic biomarkers in lung squamous cell carcinoma

Master’s Research Assistant Ankara, Turkey
Middle East Technical University *Feb 2013 — Aug 2015 (2 years, 7 months)*

- Modelled RBP–microRNA interplay in post-transcriptional regulation using CLIP-seq assays, knockdown perturbation experiments, and RNA structure analysis (RNAplfold/Sfold), enabling causal inference of regulatory effects on mRNA stability and gene expression
- Served as Teaching Assistant for Python and Java programming courses, facilitating weekly tutorials, grading assignments, and mentoring students through hands-on support and regular office hours

EDUCATION

Doctor of Philosophy, Bioinformatics — University of British Columbia (UBC) | 2017 – 2024
Thesis: Machine learning in transcriptome analysis using long RNA sequencing data
Master of Science, Bioinformatics — Middle East Technical University (METU) | 2013 – 2015
Thesis: Modeling the combined effect of RBPs and microRNAs in post-transcriptional regulation

SELECTED PUBLICATIONS

For the full list of publications, please visit my Google Scholar.

1. ntEmbd: Deep learning embedding for nucleotide sequences
   Hafezqorani et al. **BioRxiv**, 2024 — [Paper] [Code] [Video Presentation] [Blog]
2. Trans-NanoSim characterizes and simulates nanopore RNA-seq data
   Hafezqorani et al. **GigaScience**, 2020 — [Paper] [Code] [Blog]
3. Reference-free assembly of long-read transcriptome sequencing data with RNA-Bloom2
   Nip, **Hafezqorani** et al. **Nature Communications**, 2023 — [Paper] [Code]
4. Systematic assessment of long-read RNA-seq methods for transcript identification and quantification
   LRGASP Consortium, et al. **Nature Methods**, 2024 — [Paper] [Code] [Website]
5. Modeling the combined effect of RNA-binding proteins and microRNAs in post-transcriptional regulation
   Hafezqorani et al. **Nucleic Acids Research**, 2016 — [Paper]

SELECTED TALKS & PRESENTATIONS

ISMB/ECCB - MLCSB (Machine Learning in Computational and Systems Biology) — Lyon, France - 2023
ISMB/ECCB – HiTSeq (High Throughput Sequencing Algorithms & Applications) — Basel, Switzerland - 2019
Genome Informatics — Hinxton, Cambridgeshire, United Kingdom – 2018
ISMB/ECCB – CAMDA (Critical Assessment of Massive Data Analysis) — Prague, Czech Republic - 2017
+ 8 other academic conference presentations, with attendances at NeurIPS and ICML. See here for full list.

TECHNICAL SKILLS

Bioinformatics & Omics: NGS (long/short RNA-seq, single-cell and bulk transcriptomics, metagenomics), multi-omics integration & visualization, perturbation screens, protein-binding assays (ChIP/CLIP-Seq), causal inference, RNA-protein interactions, RBPs, microRNAs, cancer genomics, clinical data integration, RNA secondary structure

ML & Deep Learning: Representation learning, transformers, LSTMs, autoencoders, state-space models, Bayesian optimization, hyperparameter tuning (Optuna), feature selection, GNN, VAE, PyTorch, TensorFlow, scikit-learn

Statistical & Probabilistic Modeling: Mixture models, kernel density, EM, Markov models, Bayesian inference

Programming & Scientific Software: Python, R, Bash & Shell scripting, Linux, Matplotlib, Pandas, NumPy, SciPy

Workflow & Infrastructure: Nextflow, Slurm, Google Cloud Platform (GCP), Docker, Conda, GitHub, CI/CD

SELECTED AWARDS & ACADEMIC SERVICES

Reviewed for Scientific Journals (Bioinformatics, NAR Genomics and Bioinformatics, Bioinformatics Advances)
Received over $250K in Scholarships, Fellowships, Travel Awards, and Research Fundings | **2017 — 2024**
Active Member of International Society for Computational Biology (ISCB) | **2015 — 2025**