

Ahmad Abdel-Azim

5382 Windtree Dr, Doylestown, PA 18902

(920) 540-2092 | ahmad_abdelazim@college.harvard.edu | GitHub: <https://github.com/ahmadazim>

EDUCATION

Harvard University | Graduate School of Arts and Sciences (GSAS) Cambridge, MA
Master's degree (A.M.) in Statistics Expected 05/2024

Harvard University | Faculty of Arts and Sciences (FAS) Cambridge, MA
A.B. Joint degree in Statistics, Molecular and Cellular Biology – GPA: 3.85 Expected 05/2024
Teaching Fellow for the following courses: Statistics 111 (Statistical Inference), Statistics 110 (Probability), Statistics 185 (Unsupervised Learning), MCB 112 (Biological Data Analysis), Statistics 117 (Biostatistics)

RESEARCH EXPERIENCE

Harvard T.H. Chan School of Public Health | *Researcher, Xihong Lin laboratory* Boston, MA | 09/2022 – present

- Developing robust polygenic risk scores (PRS) for early disease diagnosis and reliable patient stratification, leveraging mixed effects model framework. Our method correctly accounts for longitudinal data and cryptic/familial relatedness.
- Validating method via simulation and empirically using UK Biobank data to show that predictions produced are consistently more accurate than current state-of-the-art approaches.

Brigham and Women's Hospital | *Researcher, David Kwiatkowski laboratory* Boston, MA | 09/2020 – 04/2022

- Led broad scientific mission to unravel the genomic landscape of a rare, aggressive tumor type known as PEComa and provide diagnostic markers (extensive experience analyzing WES/WGS, RNA-seq, and ChIP-seq data).
- Co-authored 5 manuscripts reporting new somatic genetic events (SNVs, indels, CNVs) and mutational signatures that contribute to the development and progression of malignant PEComa with metastatic potential.

Massachusetts Institute of Technology (M.I.T.) | *Manolis Kellis laboratory* Cambridge, MA | 09/2021 – 12/2021

- Developed multi-omics machine learning model to predict Alzheimer's disease (AD) onset and progression from genetic (XGBoost), neuroimaging (3D CNN), and clinical data (MLP) to facilitate rapid and precise diagnosis.

Harvard Medical School | *Research Science Institute (RSI), Debora Marks laboratory* Boston, MA | 06/2019 – 04/2020

- Selected as one of 82 students around the world to conduct research at the RSI 2019 program.
- Integrated publicly available single-cell RNA sequencing data sets to investigate differentiation pathways and cell fate determination signals. Explored effects of simultaneous biological systems on fate decision-making during differentiation by modeling organelle development relative to cell development.

Lawrence University | *Researcher, Brian Piasecki laboratory* Appleton, WI | 09/2017 – 06/2018

- Studied evolutionary behavior of bacteria in antibiotics and Manuka honey. Characterized the bactericidal properties of honey and the role of osmotic stress. Modeled genetic basis and patterns of gaining antibiotic resistance.

WORK EXPERIENCE

Regeneron | *Analytical Genetics Intern, Regeneron Genetics Center (RGC)* Tarrytown, NY | 06/2023 – 09/2023

- Developed efficient methods for computing polygenic overlap at biobank-scale from GWAS summary data. Implementing Monte Carlo and Bayesian methods for rapid and efficient computation. Validating method empirically in UK Biobank.
- *Manager: Goncalo Abecasis, VP RGC Chief Genomics & Data Sciences Officer*

Biotia | *Bioinformatics Consultant* New York, NY | 05/2022 – 09/2023

- Developed and tested AMR (antimicrobial resistance) ML identification pipeline for metagenomic diagnostic samples.
- Developed unsupervised machine learning model to predict AMR from protein structural disruptions in protein variants. AlphaFold was leveraged for protein structure prediction and *de novo* AMR gene discovery.

Preverna | *Data Scientist, Computational genomics* San Francisco, CA | 06/2021 – 05/2022

- Built end-to-end machine learning pipeline to identify kinase drug targets and extract relevant features from multi-omics (e.g. proteomic) data associated with disease onset, progression, and survival for several cancer types.
- Identified novel biomarker candidates to support drug discovery using in-house and publicly available datasets.
- Continually presented methods to internal and external stakeholders and made a case for advancing discoveries from *in-silico* to *in-vitro* and *in-vivo* trials.

LEADERSHIP & ACTIVITIES

Harvard Data Analytics Group (HDAG) | Chief Consulting Officer (CCO) Cambridge, MA | 09/2020 – 04/2023

- Led and organized sourcing process for HDAG data consulting services. Pitched to hundreds of Fortune 500 companies, NGOs, and local startups across several industries (Roche, WHO, NBA) and collectively sourced engagements worth \$250k+ each semester.
- Organized and managed case teams of 100+ competitively recruited undergrads to conduct data projects for signed clients.

Harvard Faculty of Arts and Sciences (FAS) | Teaching Fellow Cambridge, MA | 09/2022 – present

- Served as a Teaching Fellow for several courses in the statistics, computer science, and biology departments, including Statistics 111 (Statistical Inference), Statistics 110 (Probability), Statistics 185 (Unsupervised Learning), MCB 112 (Biological Data Analysis), CS 181 (Machine Learning), and Statistics 117 (Biostatistics). Led weekly review sections for students and hosted weekly office hours.

Harvard Open Data Project (HODP) | Nonprofits Director Cambridge, MA | 12/2020 – 05/2022

- Organized and managed teams of trained undergraduate data scientists and journalists to conduct pro bono data projects for local nonprofits across a variety of industries (e.g. government, computer software).

Harvard Student Agencies (HSA) | Tutor Cambridge, MA | 10/2021 – 05/2022

- Subjects taught include statistics, computer science, biology, chemistry, and mathematics.

TECHNICAL SKILLS

- **R (high proficiency):** ShinyR, ggplot2, caret, Seurat, tidyverse, dplyr, devtools
Extensive experience with package development, simulation, statistical modeling, and data visualization.
- **Python (high proficiency):** Pytorch, TensorFlow, Keras, Sklearn, Scipy, OpenCV, NumPy, Pandas, Scanpy, Matplotlib
- **Julia, SQL, C/C++, HTML/CSS, MATLAB, Git, Markdown, Linux**
- Cloud computing platforms: Google Cloud Platform (GCP), Amazon Web Services (AWS), Microsoft Azure

REFEREED ARTICLES

- Giannikou, K., Martin, K., **Abdel-Azim, A.G.**, Hougard, T.R., Tang, Y., McKeigan, J., Kwiatkowski, D.J., Lam, H.C. (2023). Spectrum of Germline and Somatic Mitochondrial DNA Variants in Tuberous Sclerosis Complex. *Frontiers in Genetics*, 13, p. 917993. doi: 10.3389/fgene.2022.917993.
- Abdel-Azim, S.G., **Abdel-Azim, A.G.**, Piasecki, B.P., and Abdel-Azim, G.A. (2019). Characterization of the Gain and Loss of Resistance to Antibiotics Versus Tolerance to Honey as an Antimutagenic and Antimicrobial Medium in Extended-Time Serial Transfer Experiments. *Pharmacognosy Research*, 11(2), p.147-154. doi:10.4103/pr.pr_175_18.
- **Abdel-Azim, A.G.**, Abdel-Azim, S.G., and Abdel-Azim, G.A. (2019). Determining the Contribution of Osmotic Stress to the Antibacterial Properties of Honey. *Journal of Emerging Investigators*, 2(4), p.1-8.
- **Abdel-Azim, A.G.**, Abdel-Azim, G.A. (2016). The Impact of Antibiotic Exposure and Concentration on Resistance in Bacteria. *Journal of Emerging Investigators*. p.1-6.

ARTICLES IN PREPARATION

- **Abdel-Azim, A.G.**, Lin, X. Joint estimation of millions of genetic variants to account for correlated observations in Biobank-scale data for predicting disease risk and progression.
- Giannikou, K., **Abdel-Azim, A.G.**, Tsuji J., Adib, E., Zhu, Z., Wu C., Martignetti, G., Henske, E.P., Ha, G., Getz, G., Hornick, J., Wagner, A.J., Kwiatkowski, D.J. Comprehensive genomic and clinicopathologic characterization of advanced PEComa tumors with metastatic potential.
- Giannikou, K., **Abdel-Azim, A.G.**, Adib, E., Schaefer, I.M., Kesten, N., Taing, L., Zhu, Z., Hornick, J.L., Hirsch, M.S., Long, H.W., Wagner, A.J., Hemming, M.L., Kwiatkowski, D.J. Distinct oncogenic signatures in malignant PEComa and leiomyosarcoma identified by integrative RNA-seq and H3K27ac ChIP-seq analysis. *Journal of Clinical Oncology*.
- **Abdel-Azim, A.G.**, Giannikou, K., Kwiatkowski, D.J. Characterization of the germline and mitochondrial mutational landscapes in advanced PEComa tumors with metastatic potential.
- **Abdel-Azim, A.G.**, Green, T. Using Single-Cell RNA Sequencing Data to Study Plastid Differentiation Dynamics with Nucleus-Encoded Plastid Gene Expression in *Arabidopsis thaliana*.

SCIENTIFIC PRESENTATIONS

- Giannikou, K., **Abdel-Azim, A.G.**, Tsuji J., Adib, E., Zhu, Z., Wu C., Martignetti, G., Henske, E.P., Ha, G., Getz, G., Hornick, J., Wagner, A.J., Kwiatkowski, D.J. *Comprehensive genomic and clinicopathologic characterization of advanced PEComa tumors with metastatic potential*. Discover Brigham 2021.
- **Abdel-Azim, A.**, Green, T. (2019). *Using Single-Cell RNA Sequencing Data to Study Plastid Differentiation Dynamics with Nucleus-Encoded Plastid Gene Expression in Arabidopsis thaliana*. Presented at Research Science Institute Program. Received top 5 papers and top 10 oral presentations awards.

HONORS AND AWARDS

Harvard College Research Program, Term-time Research Funding	Awarded Spring 2021, Fall 2021, Spring 2023
Herchel Smith Undergraduate Research Fellow	06/2021 – 08/2021
Program for Research in Science and Engineering (PRISE)	Awarded 2021, Gratefully Declined