BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Ge, Aaron

eRA COMMONS USER NAME (credential, e.g., agency login): age1

POSITION TITLE: Medical Student

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Start Date MM/YYYY	Completion Date MM/YYYY	FIELD OF STUDY
University of Maryland, College Park	BS	08/2020	12/2022	General Biology
University of Maryland, School of Medicine	MD	08/2024	05/2028 (Expected)	Medicine

A. Personal Statement

My work integrates artificial intelligence (AI) and clinical medicine to address systemic healthcare inequities, particularly for underserved populations. My 4 years of experiences as an EMT responding to 911 emergencies through the COVID pandemic revealed how socioeconomic barriers–such as fear of medical costs–delay critical care and worsen outcomes. These observations drove me to pursue AI as a tool for scalable, equitable solutions.

At the National Cancer Institute, I developed privacy-preserving AI models to democratize access to precision medicine. I led FastImpute, an on-device machine learning tool for genotype imputation and breast cancer risk assessment, which enables secure genetic insights without cloud dependency. I also created TMA-Grid, a CNN-powered web application for tissue microarray analysis, now adopted by resource-limited labs to accelerate oncology research. For foundational AI training, I engineered a transformer neural network from scratch in JavaScript and led journal clubs analyzing architectures like Hyena and Mamba.

Driven by the critical need to bridge the gap between cutting-edge computational talent and pressing medical challenges, I co-founded Tensor Lab for Computational Medicine in June 2025. As a Medical Student Lead within a team of eight medical students, I am spearheading this highly selective summer research fellowship. Tensor Lab is designed to mentor and guide exceptionally talented computer science students in applying their advanced technical skills to devise AI-driven solutions for high-impact medical questions. The initiative has already demonstrated significant promise, attracting over 118 applications from students at top-tier universities such as Johns Hopkins and UC Berkeley, many of whom bring experience from internships at organizations like NASA and Google. Our goal for the inaugural summer is to facilitate the creation of approximately 15 independent research projects, closely mentored by leading physicians and scientists from renowned institutions including UCSF, the University of Maryland School of Medicine (UMSOM), and the NIH. This program aims to cultivate a new generation of innovators capable of working at the frontiers of medicine and AI.

As a medical student, my current research extends across multiple facets of computational medicine. Supported by the AIM-Ahead fellowship, one of my primary projects focuses on the interplay of clinical severity, social determinants of health (SDOH), and hospital care in predicting stroke outcomes. I am analyzing a large clinical dataset of hemorrhagic and ischemic stroke patients to understand how individual-level SDOH influences in-hospital mortality, independent of admission clinical severity indicators (like GCS, NIHSS), and whether these effects are mediated by differences in the type, timing, or intensity of diagnostic and therapeutic interventions received. Concurrently, in my summer research position at the Institute of Health Computing (June 2025 - Present), I am developing and evaluating an interpretable multimodal EHR-based 10-year cardiovascular disease (CVD) risk prediction model using LLM-generated embeddings from the MIMIC-IV dataset. This project focuses on robust model development, comprehensive evaluation, and in-depth analysis of interpretability and fairness.

Beyond technical innovation, my goal is to advance AI solutions that are clinically rigorous, socially accountable, and scalable for marginalized communities, both through my direct research and by empowering the next generation of innovators through programs like Tensor Lab.

B. Positions, Scientific Appointments and Honors

Positions and Scientific Appointments

- June - August 2019: Student researcher at the Comprehensive Liver Cancer Center
- June - September 2020: Lab Assistant at TopAlliance Biosciences Inc.
- January - June 2021: Research Intern at the National Institute on Alcohol Abuse and Alcoholism
- Summer 2021, 2022: Summer intern at the National Cancer Institute
- December 2022 – August 2024: Post-baccalaureate fellow at the National Cancer Institute
- August 2024 – Present: Medical Student at the University of Maryland, School of Medicine
- May 2025 – Present: Co-founder and Medical Student Lead, Tensor Lab for Computational Medicine
- June 2025 – Present: Summer Research Fellow, Institute of Health Computing

Honors

- August 2020 – December 2022: Presidential Scholarship at University of Maryland, College Park
- December 2022: Summa Cum Laude at University of Maryland, College Park
- December 2022: Integrated Life Sciences Honors Program Citation at University of Maryland, College Park
- June 2020 - August 2022; December 2022- August 2024: Cancer Research Training Award at the National Cancer Institute
- September 2024 – Maryland Delegate Scholarship
- January 2025 - AIM-Ahead Clinical Care Award
- April 2025 - PRISM Research Merit Award
- May 2025 - Maryland Senatorial Scholarship

C. Contributions to Science

1. **National Cancer Institute:** My research at the NCI has been dedicated to developing innovative, open-source software tools that address challenges in cancer genomics and histopathology while emphasizing data privacy, accessibility, and reproducibility. I achieved this by leveraging web technologies to allow computations to occur on the client-side and adhering to FAIR principles. I spearheaded four significant projects that demonstrate my contributions to the field**:**
    1. **FastImpute**: I led the development of FastImpute, a project focused on creating a lightweight, client-side genotype imputation model using deep learning. This model enables the prediction of missing genetic information directly on edge devices like smartphones, enhancing user privacy and democratizing access to personalized genetic risk information. Notably, FastImpute facilitates polygenic risk score calculations, crucial for assessing an individual's genetic predisposition to various diseases. Our research demonstrated that FastImpute achieved high accuracy, with an R^2 of 0.86 for PRS313, a polygenic risk score for breast cancer risk prediction. FastImpute's ability to perform complex genomic analyses on personal devices in real-time opens new avenues for point-of-care genetic testing and personalized medicine.

        Ge A, Balasubramanian J, Wu X, Kraft P, Almeida JS. FastImpute: A Baseline for Open-source, Reference-Free Genotype Imputation Methods—A Case Study in PRS313. arXiv preprint arXiv:2407.09355. 2024 Jul 12

    2. **TMA-Grid**: I spearheaded the development of TMA-Grid, a web-based, zero-footprint application for tissue microarray (TMA) de-arraying. TMA-Grid provides a user-friendly, interactive platform that allows researchers to extract individual cores from TMAs directly within their web browser, eliminating the need for downloads or installations. The application incorporates a convolutional neural network for precise tissue segmentation and a grid estimation algorithm for accurate core alignment. TMA-Grid's interactive design allows users to refine segmentation and gridding results, ensuring flexibility and accuracy. By enabling client-side computation, TMA-Grid prioritizes data privacy and promotes collaboration, allowing researchers to access and analyze TMA data from any device with a web browser.

        Ge A, Saha M, Duggan MA, Lenz P, Abubakar M, García-Closas M, Balasubramanian J, Almeida JS, Bhawsar PM. TMA-Grid: An open-source, zero-footprint web application for FAIR Tissue MicroArray De-arraying. arXiv preprint arXiv:2407.21233. 2024 Jul 30.

    3. **mSigSDK**: I led the development of mSigSDK, an in-browser Software Development Kit (SDK) designed to facilitate the analysis of mutational signatures. Mutational signatures, patterns of mutations in DNA, offer valuable insights into the causes of cancer and potential therapeutic targets. mSigSDK, created as a companion library to the NCI's mSig Portal, a web-based platform for mutational signature research, provides researchers with tools for data processing, visualization, and analysis of mutational signatures. By adhering to modern web computing standards, mSigSDK ensures secure, private, and scalable analysis without requiring any downloads or installations. All computations are performed within the user's browser, leveraging the computational resources of their machine. The open-source and modular design of mSigSDK promotes extensibility, enabling researchers to seamlessly integrate it into their data science workflows and potentially connect with other relevant API ecosystems.

        Ge A, Zhang T, Martins YC, Landi MT, Park B, Chen K, Balasubramanian J, Almeida JS. mSigSDK - private, at scale, computation of mutation signatures. ArXiv [Preprint]. 2024 Jan 19:arXiv:2308.02995v2.

    4. **Framework for FAIR Analytics**: I developed a framework for FAIR analytics, demonstrated through a case study in mutational signature detection, which focuses on creating a portable, privacy-preserving, in-browser framework for reproducible analysis of mutational signatures. This framework specifically tackles reproducibility challenges in the field by adhering to FAIR (Findable, Accessible, Interoperable, and Reusable) principles. By utilizing web technologies such as HTML, CSS, JavaScript, and WebAssembly, the framework ensures portability,

enabling researchers to run analyses on various devices with a web browser. The framework prioritizes privacy by keeping all computations local, within the browser, preventing sensitive genetic data from leaving the user's device.

Ge A, Zhang T, Bodelon C, García-Closas M, Almeida JS, Balasubramanian JB. A FAIR platform for reproducing mutational signature detection on tumor sequencing data. arXiv preprint arXiv:2306.01634. 2023 Jun 2.

5. **AI-generated mammographic texture feature**: This study investigated the biological basis of a novel AI-generated mammogram risk score (MRS) by examining its relationship with known breast cancer risk factors within the Nurses' Health Study II cohort. Using observational and genetic analyses, the research validated that a higher MRS is strongly associated with increased breast cancer risk, even after adjusting for traditional BI-RADS breast density. The findings revealed robust phenotypic and genetic links between MRS and mammographic density measures. Furthermore, after accounting for breast density, the study identified a significant association between a genetic predisposition for a higher waist-to-hip ratio and an increased MRS, suggesting that central obesity may influence breast tissue characteristics in ways not captured by density alone. My role in this project included preparing the risk factor GWAS data for analysis , conducting the data analysis , interpreting the results, and serving as a contributor in revising the final manuscript.

   Wu, X., Jiang, S., **Ge, A**., Turman, C., Colditz, G., Tamimi, R. M., & Kraft, P. (2025). Investigating the relationship between breast cancer risk factors and an AI-generated mammographic texture feature in the Nurses' Health Study II (p. 2025.02.18.25322419). medRxiv. https://doi.org/10.1101/2025.02.18.25322419

2. **AIM-Ahead Clinical Care Research Fellowship: Stroke Outcomes Research:** As an awardee of the AIM-Ahead Clinical Care Research Fellowship, my work focuses on understanding the multifaceted predictors of stroke outcomes. This research culminated in a **poster presentation**, "Predictors of Stroke Outcomes: The Interplay of Clinical Severity, Non Medical Health Factors, and Hospital Care," at the AIM-Ahead 2025 Conference. I analyzed a comprehensive stroke cohort dataset by creating a foundational analysis dataset, performing descriptive and exploratory data analysis, and utilizing multivariable modeling techniques such as logistic regression and generalized linear mixed models (GLMM). Mediation analysis is also planned to understand the pathways through which SDOH may affect outcomes. The overarching goal is to identify key predictors and potential disparities in stroke care and outcomes, providing insights that can inform more equitable healthcare delivery.

3. **National Institute on Alcohol Abuse and Alcoholism**: During my internship at the NIAAA, I explored how reduced dopamine D2 receptor (D2R) expression affects neurophysiology and gene regulation. I conducted bioinformatic analysis on iMSN gene expression data in mice engineered to express low D2Rs. I helped identify 43 differentially expressed genes and conducted a weighted gene correlation analysis to identify a D2R-related module enriched in synaptic and signaling genes. This research experience in bioinformatics honed invaluable skills in R, scientific communication, and collaboration, cementing my passion for computational disease research.

   Guerri L, Dobbs LK, da Silva E Silva DA, Meyers A, **Ge A**, Lecaj L, Djakuduel C, Islek D, Hipolito D, Martinez AB, Shen PH, Marietta CA, Garamszegi SP, Capobianco E, Jiang Z, Schwandt M, Mash DC, Alvarez VA, Goldman D. Low Dopamine D2 Receptor Expression Drives Gene Networks Related to GABA, cAMP, Growth and Neuroinflammation in Striatal Indirect Pathway Neurons. Biol Psychiatry Glob Open Sci. 2022 Sep 8;3(4):1104-1115.

4. **Comprehensive Liver Cancer Center**: At the Comprehensive Liver Cancer Center in Beijing, I investigated hepatocellular carcinoma (HCC) and explored the role of a protein called MARCH6 in HCC development. We analyzed data from a human tumor sample database and found that MARCH6 levels were significantly higher in HCC tissues than in normal liver, and patients with the highest MARCH6 levels tended to have poorer survival outcomes. We confirmed that MARCH6 protein levels were elevated in HCC versus adjacent normal tissue using patient tumor samples. Further, experiments in HCC cell lines showed that MARCH6 enhanced HCC cell proliferation, growth, DNA synthesis, and migration. Conversely, reducing MARCH6 levels hampered these cancer cell behaviors. Through this

experience, I learned to use R to create graphs and analyses and played a crucial role in translating and revising the manuscript from Mandarin to English

Sun J, Dong Z, Chang Z, Liu H, Jiang Q, Zhang D, Lu S, Jia X, Wu D, **Ge A**, Zhao P, Wang J, Lu Y. MARCH6 promotes hepatocellular carcinoma development through up-regulation of ATF2. BMC Cancer. 2021 Jul 17;21(1):827.

D. Scholastic Performance

Test Scores

- MCAT: 521 (98th Percentile)
- SAT: 1560 (99th+ Percentile)

Undergraduate GPA: 4.0

YEAR	COURSE TITLE	GRADE
	University of Maryland, College Park	
Fall 2020	Organic Chemistry I	A
Fall 2020	Organic Chemistry Lab I	A
Fall 2020	Academic Writing	A+
Fall 2020	Students in University: ILS	A+
Fall 2020	Principles of Biology III: Organismal	A
Spring 2021	Social Enterprise	A
Spring 2021	Organic Chemistry II	A
Spring 2021	Organic Chemistry Lab II	A+
Spring 2021	ILS Service-Learning	A+
Spring 2021	Genetics & Genomics	A
Spring 2021	Applied Probability & Statistics I	A
Fall 2021	Cell Biology & Physiology	A
Fall 2021	Insects + Public Health	A
Fall 2021	General Bioanalytical Chemistry Lab	A
Fall 2021	Becoming a Design Thinker	A+
Fall 2021	Introduction to MATLAB	A+
Fall 2021	Fundamental Physics of Life Sciences I	A+
Spring 2022	Biochemistry I	A
Spring 2022	Mammalian Physiology	A+
Spring 2022	Science Writing	A
Spring 2022	Introduction to UI/UX Design	A+
Spring 2022	Fundamental Physics of Life Sciences II	A+
Fall 2022	Bioinformatics & Genomics	A+
Fall 2022	General Virology	A+
Fall 2022	General Endocrinology	A+
Fall 2022	Discovering New Ventures	A+
Fall 2022	Self-Defense	A
	University of Maryland, School of Medicine	
Fall 2024	Foundations	Honors
Spring 2025	Bugs and Drugs	Honors
Spring 2025	Brain and Behavior	Honors
Spring 2025	Digestion and Hormones	Honors