Intellectual Merit: When I was in high school, I applied to an internship at UCSF, looking for a way to save money for college and pursue my interests. My experiences in the classroom made me think working in a lab might be monotonous, but the work I did was fascinating. The lab, deemed the "Asthma Collaboratory", combines epidemiological and computational genetic approaches to address health disparities in asthma. Under the supervision of Dr. Marquitta White and Dr. Esteban Burchard, I learned the fundamentals of genome-wide association studies (GWAS) and found myself working as a full-time researcher, repeatedly conducting statistical tests and rewriting scripts. I challenged myself to learn programming, basic statistics and advanced genetics to further my research and keep pace with the university students, post-docs and clinicians surrounding me. Ultimately, my work culminated in a research paper investigating common variants associated with asthma susceptibility in African American children (White and Risse-Adams et al 2016, Immunogenetics). Since the African American community has historically been underrepresented in large-scale biomedical asthma studies, my work focused on the generalizability of genetic risk factors identified in other racial-ethnic groups to African American youth. Of note, not only did we find that ~5% of previously reported asthma genetic associations replicated in African Americans, but we also identified a gene variant associated with asthma risk unique to African Americans. When I got the news that it was accepted to be published, I was absolutely stunned--I had transitioned from a freshman level biology student to someone with the potential to become a real scientist, affecting real change, real people. Before my time at UCSF, I had this inherent belief that all my liking science and mathematics would add up to was complicated words and complicated findings that wouldn't become interpretable or relevant for the next decade. But the contents of my paper described a current, actionable call to social justice. I learned not only that I could become a scientist, but also that the role of a scientist should be to help mend gaps of knowledge in their community. My mentors not only gave me the opportunity to conduct and be credited for my research but inspired me to think critically about the impact I want to make on the world.

Over the course of my high school education, I went on to author several additional publications with Dr. Burchard's group, all investigating asthma in admixed populations. Motivated by the research I had done, after graduating I began college at UC Santa Cruz as a Bioinformatics major, hoping to continue studying human genetics. I joined Dr. David Haussler's research group, working on a collaborative research team integrating modern artificial intelligence (AI) with cerebral organoids to study human neural circuits. Coming in with little wet-lab exposure, I focused on gaining experience in cell culture and basic benchwork, particularly characterizing cell growth in a fluidic setup. Wanting to learn and contribute more to the AI portion of this project, I took an internship position at Adobe Research in their audio division in the summer of 2019 through their associated Women-in-Technology scholarship program. After arriving at Adobe Research, I was noticeably the youngest and least knowledgeable person on my team, and one of only three women in the group of twenty plus. Although I felt the pressure of representing both my generation and my sex, I was able to successfully develop an audio watermarking program in collaboration with my mentor, using deep learning and signal processing. This work was accepted for patent in late 2021 (Jin et al, 2021). While my experience at Adobe was wonderful, it revealed to me just how little of the math I understood "under the hood". Having (at the time) only taken introductory differential calculus, jumping onto an in-depth deep learning project without linear algebra was like trying to sprint without learning to crawl. Because of this experience I decided to transfer to UC Berkeley as a Statistics/Mathematics major to challenge myself to learn those fundamental concepts.

I returned to UC Santa Cruz with renewed energy, and quickly enrolled in the remaining undergraduate mathematics courses available to me. Before my entry to UC Berkeley, I was accepted to the summer 2020 UCLA Bruins in Genomics program, where in Dr. Sriram's group I focused on characterizing methods of detecting natural selection. Most traditional methods of detecting natural selection do so through discrete assignment or set-based population identification, and then search for genetic variants with unusual differentiation between populations. In cases where there is admixture, these methods fall short. Because of this, statistics based on principal component analysis (PCA) have been developed as an alternative method to identify signals of natural selection that address such shortcomings. However, many PCA-based statistics are understudied in their sensitivity and power to detect variants under various models of natural selection. As part of my project, I assessed three previously proposed PC-based selection statistics using data simulated under common models of selection. Ultimately, we found that PCA-based statistics are generally underpowered, revealing a need for further developments in statistical methods to detect putative signals of selection.

By the time I started at UC Berkeley in the fall, I essentially had a crash course in applied statistical methods. While I was exposed to population genetics research during my time at UCSF, my work with Dr. Sriram solidified my interests in research focusing on admixture, methods development and underserved populations. After starting at UC Berkeley, I began working with Dr. Priya Moorjani, whose lab focuses on how human evolutionary history affects current genetic variation and disease risk. Similar to the work I conducted with Dr. Sriram but with notably different methods and simulation frameworks, I designed a series of simulations using a genetic simulator, SLiM, under a range of demographic models to assess the performance of a novel admixture-sensitive selection method and compared it against classical methods. Namely, the integrated haplotype score (iHS), Population Branch Statistic (PBS), pcadapt, and others.

Despite my enthusiasm for the project, for health reasons I withdrew from school temporarily and was unable to continue my research. While planning my timeline for recovery and returning to Berkeley, I was offered a remote full-time internship position at Genentech and saw it as an opportunity to use my downtime productively. During the first half of my internship, I developed web app features to assist researchers in tracking their cell-culture experiment data. In the second half, I analyzed patient claims data to conduct drug-drug interaction analyses and collaborated on written proposals to the CF Foundation for related studies. In parallel, I worked at Novo Nordisk as a Bioinformatics Intern in their Computational Drug Discovery division. Under the mentorship of Dr. Deibler I studied protein-peptide complex structures, in particular benchmarking their various properties, such as predicting secondary structure and interface metric scoring using BASH shell scripting, Python, GitLab, Rosetta, and PyMOL. The culmination of my research efforts was a computational review of various protein-peptide datasets, currently in review. These industry positions allowed me to exercise the skills I had developed during my time in academic research, including but not limited to picking up new computational skills on the fly, becoming competent in domain specific knowledge and language, and writing technical proposals, all in very short periods of time.

My exposure to different research fields, within academia and industry, has reinforced my belief that I want to study evolutionary biology and population genetics during my graduate career, specifically admixed populations. Now, as we move away from eurocentric data and the United States becomes "majority minority", studying admixed populations is at the forefront of the field of human genetics, and area I intend to drive forward and lead. As someone who

identifies as admixed, of mixed Asian background, I am particularly interested in the evolutionary histories of Asian ethnic groups, which have been historically underrepresented in genetics research. Currently in collaboration with Dr. Jazlyn Mooney at the University of Southern California (USC), I am working on a novel method to jointly model IBD segments and ancestry switches (see Research Statement). In this project, I am particularly excited for the applications to demographic inference in diverse admixed populations.

Broader Impacts: During my high school and undergraduate career, I have benefited enormously from close relationships with my mentors. As an NSF GRFP funded graduate student, I intend to dedicate my time to mentoring the next generation of future scientists. I have excelled and matured as a researcher and am confident in my ability to perform and communicate science. More than that, having been involved in research at different stages of my life, I know how to convey concepts at a high and low-level in creative and engaging ways.

I have already mentored in various settings–at UC Santa Cruz, I served as the only general chemistry labs tutor on campus. I met with students regularly for approximately 10-15 hours a week, organizing regular check-ins and forming study groups. During the pandemic, when we transitioned to remote learning, I created a website for students highlighting additional resources and designed worksheets meant to prepare students for lab writeups and quizzes. I continue to mentor students in the science research program and STEM classes at my former high school, giving presentations every semester about the research and academic opportunities in college, as well as providing feedback on ongoing student projects. At USC, I will recruit high school students and university undergraduates to participate in laboratory projects. I will also join Dr. Mooney for introductory research science and coding workshops at the LA County Natural History Museum (see Research Plan Statement). I intend to use my multi-ethnic background to connect with and mentor other mixed students, particularly women. Through these activities I hope to influence young students in the same way I was influenced, by empowering them to see themselves as real scientists. I am passionate about promoting inclusivity and accessibility within the field of computational biology, and these values are at the core of my desire for outreach work. I have come a long way from being a high school freshman who expected lab work to be overwhelmingly technical and borderline inaccessible. I want to give back by encouraging students to think of population genetics research as a tool to understand their own histories and mend gaps of knowledge within their communities.

With the aid of the NSF GRFP, I will create a novel method for inferring the timing and strength of admixture within admixed groups. I am fully committed to an academic career publishing free and open-source software, and so this work will be packaged as software that is freely available to the larger scientific community. Throughout my graduate studies I am determined to continue developing inclusive computational methods to gain insight into human genetic variation. I will use my platform as a graduate student (and eventually as a future principal investigator) to continue mentoring students. I am incredibly lucky to have had mentors that believed in me intellectually and personally and desire to do the same for my future students. The NSF GRFP will provide me with the resources I need to achieve these goals and the funding necessary to pursue novel research within population genetics research.