Tyler Shimko
Statement of Purpose

I first came to appreciate the power of biological research in my high school biology class. Through the instruction of my incredibly devoted teacher, I recognized that genomic research had the power to make predictions about and alter the traits of living creatures. Upon graduating from high school, I chose to attend the University of Utah, where, with the help of the school's Office of Undergraduate Research, I had an immediate opportunity to become deeply involved in research and pursue my interests to the fullest. Since beginning my undergraduate studies, I have been fortunate to take part in research at three universities across the United States and participate in projects ranging from molecular biology to neurobiology to quantitative genetics. Through these experiences my appetite for discovery has only grown. I now seek to undertake the next step in my scientific education, the pursuit of a PhD, to prepare myself for a career in research.

During my freshman year at the University of Utah, I joined the laboratory of Dr. Erik Jorgensen to assist in the construction of universal transgene insertion sites within the genome of *C. elegans*. My project eventually resulted in the creation of three distinct transgene landing sites. It was exhilarating to discover, create, and share knowledge with others. However, my exposure to the broader research community, through presentations at university-level events, alerted me that there were other opportunities to learn new skills and make significant contributions. To expand my research skill set, I sought summer internships in laboratories focused on computational methods in addition to molecular biology. I obtained an offer from Dr. Leonid Kruglyak, of Princeton University at that time.

In Dr. Kruglyak's lab I worked with Dr. Erik Andersen, a post-doctoral fellow. Dr. Andersen had previously completed a genomic mapping experiment wherein he had determined a region of the *C. elegans* genome that conferred resistance to the herbicide paraquat and enlisted my help to construct strains in an attempt to identify the causal genomic variations in that region. I explored different techniques, using both modern molecular biology and classical genetic crosses, to construct strains with which we could test the hypothesis that we had successfully identified causal genetic variants. Throughout the course of the summer, I supplemented my hands-on laboratory experience with instruction and practice with the computational methods that Dr. Andersen had employed in mapping experiments. This basic training in computer science and statistics eventually led me to take formal classes in these subjects at the University of Utah. These courses greatly expanded my research potential and prepared me to take on large projects with significant computational components. Dr. Andersen and I continued our collaboration in his new laboratory at Northwestern University in the following summers.

In the fall of 2012, I began a new project in the Jorgensen lab to identify suppressors of the phenotype associated with a mutant protein involved in synaptic vesicle endocytosis. I learned new techniques for the design and implementation of genetic screens and applied my new computational skills whenever possible. However, by the end of the academic year, it became apparent that the phenotype of interest was too weak for our suppressor screens to yield any useful information. While I was originally upset that our project and a year's worth of work were for naught, I quickly realized that failure is more of the rule than the exception in biological research. To continue toward a career in research, I would need to learn from failure and to fail gracefully. In fact, I learned that failure gives us many answers as well. In this respect, the year had not been wasted.

During my past two summers in Dr. Erik Andersen's lab, I have had the opportunity to explore how genetic variation dictates the way in which organisms respond to their environment. I sought to determine the ways in which the genetic variation present in the worldwide

population of the model nematode *C. elegans* affects responses to different chemicals including herbicides, pesticides, chemotherapeutic agents, and anthelmintics (compounds used to treat infections of parasitic nematodes). I helped to construct and optimize a high-throughput screening technique that allows us to measure the effects of the aforementioned compounds on nematodes in a multitude of ways. This pipeline has allowed us to conduct genome-wide association and linkage mapping studies. Over the past summer, I designed and built software to clean and process the data from our screening experiments and run statistical tests to map phenotypic differences to genetic variants. This software, *COPASutils*, has been published in the journal *PLoS ONE* and made freely available on the Comprehensive R Archive Network. My hope is that this software will gain widespread use in the model organism research community.

During my junior and senior years, I have worked in the lab of Dr. Gillian Stanfield at the University of Utah. My project has focused on identifying mutations that suppress premature sperm activation in male *C. elegans*. I carried out a series of crosses and positive phenotype selections as part of a genetic mapping scheme, and, at the end of my first year, several strains that displayed the suppressed phenotype were sent for sequencing in the hopes of being able to identify causal mutations. Now, in my second year in the lab, I am beginning to utilize raw genetic sequence and mapping data to identify variants implicated in suppression. Eventually this project may help to identify the causal mutation for each of the individual suppressed strains. This work will culminate in the publication of my honors thesis.

Most recently, I have become deeply interested in the large-scale collection and analysis of genomic data. I want to develop and refine the methods used to assemble and analyze whole-genome sequences. The discovery and annotation of novel genetic variants is important for the connection of phenotype back to genotype and the exploration of gene-environment interactions that govern susceptibility to disease. Additionally, the analysis of large genomic data sets can uncover the ways in which genomes have evolved to combat disease in the past. During my tenure in a PhD program, I hope to become well-versed in the methods of genomic data collection, storage, and analysis with cutting-edge computational tools. Through research and coursework in genetics, computer science, and statistics, I will work to confront the theoretical, logistical, and scientific challenges related to genomic sequence analysis.

I have selected the University of Washington, and the Department of Genome Sciences in particular, because of the department's commitment to the integration of disciplines from computer science, mathematics, and biology in the analysis of genomic data. In particular, I am interested in the work of Dr. Jay Shendure. Dr. Shendure's research team has recently developed the Combined Annotation–Dependent Depletion system for the assessing the potential health-related effects of various genetic variants in humans. I believe that with Dr. Shendure's guidance, I would be able to extend my knowledge of quantitative genetics and begin to predict the cumulative effect of a genetic background on an organism's overall health. Additionally, I am interested in the work Drs. Gail Jarvik and Mary-Claire King who have both applied the principles of quantitative genetics to identify alleles implicated in a multitude of human diseases. The work of Drs. Christine Queitsch and James Thomas in the study phenotypic adaptation to dynamic environments and the evolution of gene families, respectively, also interest me as unique ways to address to the consequences of specific genetic variants. The faculty and research opportunities that I will be exposed to at the University of Washington will prepare me exceptionally well to begin a career in genomic research.