PhD application statement of purpose

Kevin Stachelek

November 4, 2017

I want to pursue a PhD in bioinformatics because of interests in analysis of gene regulatory networks, rare disease variants, cancer genomics, and single cell sequencing. These interests have developed through experiences as an undergraduate at the University of Southern California (USC), a technician at Children's Hospital Los Angeles (CHLA) and a master's student in Translational Genomics at USC. In each of these roles I have consistently pursued training in computational methods and statistics to deal with problems in biological and medical research.

As an undergraduate I was exposed to bioinformatics through a seminar focusing on gene-environment interaction in gene regulatory networks instructed by Dr. Sergey Nuzhdin at USC. I worked with graduate students and post docs in the lab to investigate regulatory networks in the insulin/TOR signaling pathway using drosophila transcriptome data. I was exposed to the basics of sequence alignment and shell scripting through this work. This gave me the confidence to learn more complicated programming skills. In a similar way, I credit my exposure to gene network interactions as an undergraduate with giving me a sense of what analyses are possible when faced with complex genomic data, so that with a stronger theoretical foundation, I might contribute to advances in genomic analysis. I also volunteered as a technician in the lab of Dr. John Tower, where the focus was drosophila genetics in aging. Through this exposure to the wet lab at the bench and data analysis at the terminal I gained an appreciation for the demands at each level of research.

After graduating I began work in the laboratory of Dr. David Cobrinik at Children's Hospital Los Angeles. I was hired as a technician to perform cell culture and drug testing in partnership with a pharmaceutical company. I sought out ways to apply my knowledge of bioinformatics and saw a chance to make analysis more efficient and visualization more intelligible, so I taught myself survival analysis and plotting using the R programming language. I was then given the chance to work on a study of retinoblastoma tumor progression using exome sequencing data, a challenging task given the lack of expertise in the lab. However, I embraced the challenge and developed my skill working in the shell, Python and R for building biological pipelines, scripting and visualization. In a few months, I had a working pipeline with preliminary data on single nucleotide variants and somatic copy number alterations in retinoblastoma disease progression. Our manuscript is currently in preparation for submission late 2017. Of course, progress was not always straightforward. Analysis of next-generation sequencing data remains challenging. Even with perfect experimental design the discovery of variants that contribute to disease is hampered by sequencing error and false positive variants. My experience has given me the knowledge of what is possible and the resources to take on interesting and challenging sequencing analysis projects.

As Dr. Cobrinik’s lab has grown, I have continued to develop bioinformatics skills working with single-cell sequencing data. Single cell RNA sequencing (scRNA-seq), in particular, provides a view of gene expression with much greater resolution than cell population data but has many challenges for interpretation. Working with several lab members and collaborators on experiments involving scRNA-seq data, I have built up fundamental skills in genomic analysis including unsupervised learning as applied to dimensionality reduction for clustering and identification of developmental trajectories, as well as supervised methods like differential expression and classification for biomarker identification. I am interested in statistical methods to draw knowledge from sequence data; however, I want to move from implementation of established methods to development of bioinformatics applications for genomic analyses.

My career aspiration is to lead a bioinformatics research group in industry or academia creating more intuitive, accessible tools for genomic analysis that allow researchers to get more out of their data. I am amazed every day to interact with online communities of computational biologists and statistical programmers pushing bioinformatics forward. I have directly benefited from these communities in my research. But access to these resources can still depend on a first exposure to the possibilities of bioinformatics. In my own limited sample of research labs, I have seen the unmet need that I as a student can help address. Both bench and command-line skills are essential for progress in biology. I can make the biggest impact by applying my programming skills and by teaching others. As an undergraduate and masters student I have learned how to apply bioinformatic tools. As a PhD student I want to get the depth of knowledge to build bioinformatic tools and to be a bridge between programming and bench research.

The work of several professors at USC has been directly relevant to my own projects. Michael Bonaguidi's work on single cell developmental trajectories is the kind of technical development and collaborative process that I would want as a graduate trainee. The work of Kim Siegmund on modeling genetic and epigenetic networks of cancer has been an inspiration for me. And the bioinformatic tool development of Kai Wang has been directly relevant to my own projects. I am open to other research areas to push forward bioinformatics and further my own training and development.