

I want to pursue a career as a computational biologist because it combines biology and math, two of my favorite subjects, into a personally fulfilling career where I help biologist utilize their data which has increased exponentially in quantity and complexity. As such, I am excited to apply to Genome Sciences at the University of Washington because the graduate program is a place where I can build upon my degree in mathematics and my bioinformatic research experience to gain expertise in developing computational tools for single cell sequencing and in computational structural biology by working with the faculty associated with the program.

Currently, I am a research technician in Dr. X's laboratory at UNC Chapel Hill School of Medicine. The X's group's research focus is to develop innovative approaches to regenerate or repair an injured heart. The two projects where I have contributed the most involve cardiac reprogramming and improving our understanding of heart development.

Adult humans fail to regenerate their hearts following injury, and this failure to regenerate is a leading cause of heart failure and death worldwide. Cardiac reprogramming is a technique that can restore cardiac function by generating cardiomyocytes (CMs) from cardiac fibroblasts to replace scar tissue. We discovered that the combination of Mef2c + Ascl1 (MA) transcription factors (TFs) could reprogram fibroblasts to CM like cells (iCMs). The MA cocktail is smaller than existing cardiac reprogramming cocktails of Mef2c, Gata4, and Tbx5 which increases its ability to be delivered in a clinical setting to restore heart function. I helped quantify the cardiac reprogramming efficiency of the MA cocktail by carrying out immunofluorescence microscopy for CM markers on cells treated with MA that I cultured. I then used the microscopy data to quantify cardiac reprogramming with a software called Cell Ranger. I was also responsible for determining how chromatin accessibility predicts cell fate decisions in the context of MA cardiac reprogramming. This is important because Ascl1 is an epigenetic factor that manipulates chromatin and understanding its impact on chromatin gives us greater insights on the driving factors of MA cardiac reprogramming, and perhaps how to improve it. For this aim, I analyzed single-cell-ATAC-RNA-sequencing data from MA reprogrammed cells to uncover domains of regulatory chromatin sets affected in MA cardiac reprogramming. Since the publishing of the paper I have analyzed MA single cell datasets and have identified TFs that could be potentially causing off-target effects that decrease the efficiency of MA cardiac reprogramming. We are currently in the process of designing a screen to determine if silencing these TFs could increase the MA cardiac reprogramming efficiency.

For the second project, we are studying the postnatal period of heart development, a key window for the transition of CMs from an immature to a mature phenotype, in order to improve the maturity of iCMs produced from cardiac reprogramming. We have created the first high resolution single nucleus data set of postnatal mouse heart development to identify developmentally important genes during this time period that could uncover how to improve the maturity of iCMs. I played large roles in the preparation, integration, annotation, and analysis of single nucleus RNA-seq libraries from the mouse hearts of postnatal day P0, P7, P14, and P21 mice to compile the dataset. To interrogate it we needed a method to select developmentally important genes that change throughout development. However, it was infeasible to choose individual genes because the total number of genes that change expression in CMs throughout development is in the thousands. After trying many methods, I successfully sorted genes into smaller groups with similar expression patterns by using Ward's hierarchical clustering with log-fold change calculated with pseudocounts that, elegantly, prevent edge cases from dominating the analysis where gene expression changes from or to very

low levels throughout development. To compliment this analysis I used CellChat, a program able to quantitatively infer and analyze intercellular communication networks, to reveal potentially novel behavior of gene pathways throughout postnatal development. My computational efforts has allowed us to prioritize genes for analysis and to identify biological processes that occur at different times of development. To follow up with this analysis we are currently developing a method to knock out multiple genes with an in-vivo CRISPR mouse heart model to study the genes I find that are potentially developmentally important. I have carried out molecular cloning to incorporate gRNAs of dozens of genes into CRISPR aav-seq plasmid vectors for the experiment. As of this moment, we are waiting for the sequence results for a preliminary in-vivo heart Perturb-seq experiments that I will analyze.

My time in the X lab, has reinforced my decision to become a computational biologist. While learning the techniques I used to assist my fellow lab members, I was inspired by how researchers creatively applied math to answer biological questions and I tried to do this for my own projects. Though I have been successful in doing so, in order for me to continue to grow as a bioinformatician I need to matriculate into a graduate program. As a PhD student I can gain exposure to the frontiers of biology where computational tools could make an impact and then formally learn the computational tools and mathematics used in computational biology. The Genome Sciences program has the mentorship and structure to allow me to effectively acquire this knowledge and to grow as computational biologist.

In graduate school I want to develop methods to overcome many of the challenges I have become familiar when performing data analysis. In particular, for single cell analysis there is a need for methods of comparing different trajectories obtained from the same data type but across individuals or conditions, in order to highlight unique and common aspects. My postbaccalaureate research experience would perfectly translate into the Trapnell lab to achieve this goal. I have spent a year analyzing many single-cell cellular reprogramming and developmental biology timeseries datasets to elucidate gene function in context of pathology of cardiac reprogramming. I have first-hand research experience in almost every single area of focus in the lab and I have even extensively used and published pseudotime analysis data using their Monocle3 package. In the Trapnell lab I would be beyond excited to leverage my single-cell data analysis experience to develop new methodologies and software to help researchers make sense of their single-cell data.

Another interest of mine is to learn how machine learning can be used to model proteins and their interactions. Specifically, I want to study if ligand-receptor interactions can be modeled successfully now that intrinsically disordered regions of the binding site that have relatively fixed surrounding structure are solved with a high degree of accuracy thanks to programs like Rosetta and AlphaFold. At the University of Washington's Institute for Protein Design and with its associated researchers, such as those in the Baker lab and the King lab, I would have the opportunity to learn in a world-class environment the cutting edge of computational protein design and modeling. The recently released ProteinMPNN is an example of the software I would like to learn how to design, it beautifully uses Monte Carlo modeling, graph theory, machine learning, and protein biology in conjunction with AlphaFold2 to generate protein structures.