Jonah Poczobutt

ASI Assignment #1

As a student in the Data Science MS program, I’m clearly new to all of this. That being said, I’m most excited to learn about epigenetics/epitranscriptomics. Studying the processes by which gene expression is regulated seems to have the potential for huge impact, both in drug development and in enhancing our understanding of epigenetic risk factors. Being aware of epigenetic variation in crop breeding seems that it could also yield huge gains in this field. I could be wrong in this assessment, but it seems to me that epigenetics is not as well understood, thought not for lack of importance. I was surprised to learn that “the total amount of regulatory regions is likely greater than that of protein coding regions” as mentioned in the HTS technology reading for lecture one. This seems to emphasize the importance of epigenetics.

Designing algorithms to analyze such data is a daunting though worthwhile task. Given my current understanding, this seems to be a bottleneck in the scientific process, so I’m interested in contributing to solutions here. While the specifics of addressing any one problem will clearly vary, I’m interested in learning about design paradigms or general best practices in this space. I thought the “Rules to live by” section of the first lecture was useful for this reason.

As mentioned in the last paragraph, I’m interested in working with genetic data of different crops. It doesn’t seem like this course is necessarily oriented toward this application, but I’m optimistic about the prospect of learning methods to apply generally. Obviously, learning about and navigating different primary databases that might contain this material would be important for me.

I’m definitely lacking in experience and my knowledge of the lay of the land in the bioinformatics world, so the first thing I am hoping to gain is an understanding of what seem to me to be the basics: What kind of data is out there, what are the tools available to access/analyze it, what kind of hardware is gathering this data in the first place and what implications does that have for my specific project. The first lecture was particularly helpful with regards to this last question, so it seems like I’m in the right place. Getting more familiar with HPC environments is another goal of mine. I think I should be set up well to that this semester between this course and my “Big Data” course as part of the Data Science MS program.

How exactly I will apply the sequencing informatics knowledge I gain through this course is more of an open question for me. I imagine many of the Bioinformatics MS students have more concrete plans given they have spent more time in the space. At this point, my plan is to foster my interest in the bioinformatics world and get my feet wet through this course. I plan to be mindful of potential areas for further work/research based on what we cover in this class and select a final project with this goal in mind.