Brian Wiley

MS Bioinformatics Johns Hopkins University

Computational Biology and Bioinformatics Internship 2020

1. What biological problem would you like to solve computationally?

From a high level I would like to be to understand differential RNA expression and differential epigenetic modifications in children with hematological cancers compared to other children with these types of cancers and samples from normal children. I would like to compare and analyze the differentials to come up with targets to test small molecule drug therapy to eradicate not only the cancer cells but also therapy to treat the underlying cause in the stem and progenitor cell populations. I am also interested in doing protein analysis to discover hotspot target in fusion proteins which are frequent in these diseases and include solving crystalized structures of these molecule interacting with the proteins and the adverse effects when testing them with chimeric mice.

1. Our work involves assigning patients to clinical trials base on genomic variation. Attention to detail is critical for accurate assignment, how would you ensure your work is accurate?

I would make sure that I can retrace the steps completed for my analysis so that when reviewing it with mentor and client we will be able to nail down the point where errors could have arisen. I would also repeat the process 1 or 2 more times with and without the reproducible code and protocols to make sure the results are similarly expected. Where appropriate I would create Bash script to streamline the workflow process for creating variant representation from sample reads.

1. We are focused on delivering on client needs. What would you do if you don’t agree with the client?

I would first try to ask some questions if I did not agree because very often, I have noticed for myself, if I am in disagreement with another party, it may so happen to be that there is a communication issue between that other party and myself or I have not full understood the situation or facts coming from that party and the process or data. Depending on the situation, it could also be the fact that both the client and I are both correct and I think the best way in delivering the best service to the client is that everyone is on the same page. And it is also obvious that I may also bring my ideas back to my manager before explaining my viewpoint to the client.

1. What are some challenges with variant representation? Consider exchange of variant data between organizations and how you can ensure the groups are discussing the same variant.

I have read that one problem with variant representation is making sure it normalized. That is that each variant in the VCF and left aligned and parsimonious to not have any redundant variants in the files and keep the file sizes as small as possible without loosing any important variant information. This can be done with a tools like ‘gatk’ and ‘vt’.

With respect to variant exchange, I would make sure specific information is added to the VCF file header like a date and time or hash code reference id. Also I would make sure there is documentation on how the variant files are called, i.e. the steps in creating the Fast5/Fastq files or where, when, and who the reads came from, the reference genome from which the alignment was done and the aligner or assembler used for the assembly of the sample(s) as well as the caller and method used for the VCF generation. I would also keep a log for each of these processes.