Carnegie Mellon University

Siddharth **Reed**

Computational Biology Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

cmu.edu

slreed@andrew.cmu.edu | DJSiddharthVader | Sid-Reed

March 11, 2021

GlaxoSmithKline
14200 Shady Grove Road
Rockville, MD, 20850

Dear Hiring Manager,

I am an first year MSc. student in Computational Biology at Carnegie Mellon University applying for the Data Analytics internship.

Curating and cleaning data are often the most difficult and time consuming parts of any analytics pipeline, as the saying goes "garbage in, garbage out". I encountered this first hand during an internship in the Michael Hoffman lab at the University of Toronto under the supervision of Post-Doc and placenta expert Samantha Wilson. Despite the fact that I was working with public data from NCBI there was still a significant amount of curating which data to use as the placenta contains both fetal and maternal cells. I spent time reading papers to see how they handled their samples and writing R scripts to filter out data that did not meet our inclusion criteria. After curating the data I had about 20,000 gene expression profiles and preforming analyses looking differentially expressed genes, relevant GO term and survival outcomes among others. I also visualized my results and presented them to computational and wet-lab scientists to better understand what the data was saying about the relationship between the placenta and cancer.

I also have some experience with pipeline development from my undergraduate thesis. I was designing an end-to-end pipeline to explore the relationship between rates of horizontal gene transfer and the presence of CRISPR-Cas systems in bacteria. I wrote scripts to retrieve the data for all complete bacterial genomes on NCBI, produce gene and species trees for each species and preform network analysis. Since I was working with several thousand genomes I also needed to make sure that the code was relatively efficient and produced clear progress reports and organized output files. This involved significant python, R and shell scripting, working with and generating structured and unstructured data in an organized and reproducible way.

My formal education in Genetics and Molecular biology in my undergrad was complimented by my work and extracurricular experience writing code and working with biological data. I am now furthering my skills and knowledge by pursuing an MSc. in Computational Biology, shoring up my quantitative skills and gaining more experience with analyzing biological data.

Thank you for your time and consideration.
Sincerely,
Siddharth Reed
MSc. Student in Computational Biology
Carnegie Mellon University