Academic records.

I’m currently a first-year master student in Biostatsitics at jhu. Since jhu medical campus has a quarter system, I just obtained my grade for the 1 term. My cumulative gpa right now is 4.0. And because I want to get more exposure in clinical genomics, my course focus is highly-programming based. I’m taking one advanced programming class in r which concentrates on optimization algorithm and I will take one computational biology class next term, which talks about analyzing genomics data in python. I obtained my bachelor’s degree at ucla double majoring in statistics and biochemistry. My cumulative gpa for my undergrad was 3.63 and my major gpa was 3.92

We know students have a lot of choices when considering summer internship opportunities. What about Illumina and the work we do inspired you to apply for a role with us? Why do YOU want to be a part of something bigger?

I wanted to work at healthcare industry ever since my undergraduate and that’s why I go to pursue my Biostats master’s degree. And then why do I choose to apply for illumina? Because I think illumina innovates genomics industry and is also constantly searching for new solutions for oncology genomics.

I believe that the future treatment for cancer and oncology and any other fields that we are currently having difficulties with will largely depend on clinical genomics. So I choose to work at healthcare field to help people, to save people’s life or at least improve their life and illumina provides an exact platform to achieve my goal.

In terms of illumina’s technology, I really love the idea of the basespace sequence hub. It makes genome data analysis more accessible, more understandable and more comprehensive. And this convenience is not only for research groups and companies, but most importantly for individual use. I think as personalized genetics treatment gets increasingly popular, more and more people with no biological background will be more interested in getting their genome sequenced. Of course at this moment, online cloud service like basespace sequence hub that can store and analyze large genome data will be widely used. And this accessibility of genomics to ordinary people may be the big step for clinical genomics to succeed in the future.

Additionally, illumina is a company that truly values potential talents. It provides an EEO for all kinds of candidates. For international students, I can see that illumina is very willing to provide sponsorship. If I do have a chance to work in such a diverse environment, I’m very honored that I can learn a lot from co-wokers with all kinds of background.

Please speak to school projects, academic experiences, or previous jobs you've had that would help us understand what you bring to this position. For example, what was your role, your contribution, any challenges you faced and how you overcame them, the outcome, what you learned from the experience, etc?

I used to work for sellmark corporation as business analyst intern for 6 months. Sellmark corporation is a retail-based company selling hunting optics products. It is located in texas so I worked remotely to conduct statistical consulting based on my employer’s demand. They wanted to figure out if there was certain pattern for the actual sales amount and how could we use this pattern to forecast future sales in terms of time.

Initially, I would say I had no clue at all since the data set was pretty small; they only gave me 5 years of data for a certain series of products. And due to the different time frame, some products got into the market earlier and the other were pretty new. So I had to also consider a strategy to deal with the missing data.

I got some advice from my statisitcs prof about handling time frame data and started my own research on establishment of time series model. I checked tons of books, papers and tutorials and finally came through a hybrid time series model combining arima and multivariate regression. Coefficients for this hybrid model were estimated by cross validation in r. Products with similar time frame were grouped together for better prediction. Therefore I didn't have to consider the difference in time frame to complicate my model.

Even though the variance of the resulting prediction was a little bit high, my model still successfully forecasted the general trend for the sales. Therefore I was able to provide business insight to my employer based on my data analysis and I was also able to predict the potentiality of products that were new into the market. And most importantly, I got involved into a brand new field that I had never heard before. This experience of working individually to find solutions starting from 0 was definitely not sth you could simply learn from classroom.

Illumina has a fast-paced culture where embracing change is essential to thinking the unthinkable, doing the un-doable, and unlocking the power of the genome. Expand on experience you have working in that environment. Tell us about a situation where you had to adjust quickly to a significant change. How did this affect you and how did you manage it?

Back to my undergraduate period, I was admitted as a biochemistry major since I wanted to pursue my study in pharmacy. I sticked onto this goal for my freshman and sophomore year, working at research group, contributing to poster making, volunteering at dental office in order to get exposure to healthcare industry. But gradually, I realized that I did not have passion any more. All I did was following other people, doing exactly what they were doing for preparation of phar school. It was not because I did not like healthcare industry any more, I still wanted to save people’s life, but probably in a different manner.

I wanted to get involved into fields that were more quantative based. I wanted to know how programming skills and math theory could be used to solve healthcare problem. So I decided to add one more major about statistics at the beginning of my junior year. Of course this means that I had to finish all of the stats class, both upper and lower and the remaiing of the biochem class within 6 quarters. I filled up basically all of schedules with classes and also sacrified the majority of my spare time for studying.

As I got more involved into statistics, I realized that the theoretical knowledge from undergraduate was probably not enough to solve real-life problem. So I also made determination to apply for grad school in biostats in order to get more research experience. Therefore, I had to adjust my experience completely to fit biostats. I had to build strong connection with statistics prof in order to get reference letter. I had to find relevant internship, get relevant certificate in order to make my resume looks better. It was really tough back in those days. But this was my choice. I never regret for that. And my determination and adjustment towards changes shape me into the track I am today.

As part of both our team and the iAspire program, you will be asked to collaborate with people from varying backgrounds and differing opinions. Tell us about your most relevant experience in collaborating with others in a similar situation. Why did you have to collaborate? What was the most challenging aspect of collaborating? What did you learn from the experience?

I used to attend a Kaggle competition for one of my data mining class. The project was about predicting the outcome for shelter animals. So we had to predict either this animal was adopted, euthanized, fostered or etc. We were given 3 weeks for this project since it was highly-computational based.

My group including me had 5 ppl. We haven’t really got started until the second week. By that time point, the majority of other groups probably had ideas about data cleaning and modeling but we still had no clue at all. We tried to build some very basic models and the results were extremely bad. Since kaggle had ranking for each project and our result was of course at the very bottom of the ranking. I could tell that the morality of our team was badly affected.

We started to have conflict about modeling the data. Basically each person had different opinions. And nobody was willing to compromise. As I mentioned previously, this project was high-programming based and both training and testing data set had over 110K observations. So at least 6 hours was needed for each model training not including cross validation for choosing coefficient. Of course the only way to improve our model performance was collaboration.

In order to be more efficient but meanwhile respect everyone’s idea, we decided that each team member would follow their own idea about cleaning and training the data first. But as long as one person made some progress, other people had to give up their ideas and collaborate together. By this approach, we narrowed down our model selection from support vector machine, multinomial regression, knn into only boosing and random forest. And this efficiency gained us a 7th place out of 60 teams eventually.

I think team work involves with sacrifice. For me in the kaggle project, I was trying support vector machine algorithm but the results were not positive and it was highly time consuming so I had to abandon my idea in the end. But team work also means contribution. Even though my idea didn't work well, I could still help teammates and tried everything I could to improve their models. So I think the most important things to consider in team work is not your own benefit but the benefit for the whole team.

Our iAspire program attracts "best and brightest" students across the globe. Why should you be chosen?

I think this customer solution analyst posttion is looking for someone who have strong analytical background and can also provide innovative business strategy. I’m very confidant about my analytical background. I have over 2 years experience with R and I have obtained sas advanced and base certificate. I also had exposure in establishing customer segmentation model during datafest completion, which was a programming-oritend completion sponsored by American statistical association.

I’m also curious and passionate towards field that I’m not familiar with. I think this curiocity helps me to move forward towards new challenges that I have never faced before. I’m now interested in clinical gemomics and trying to find a research assistant position relavent to it. So I read related papers on analyzing genomics data in statistical programming language such as R.

I know now the approach scientist normally use to find heterogeneity, to reduce batch effect. I also figure out the reason people use matrix algebra to decompose genomics data in order analyze the covarice. And most importantly, how can we utilize this analysis to help patients, to predict disease and even to innovate new treatments.

In a word, based on my quantative background and my euthasi towards gemonics, I’m very confidant that I can make contribution to illumina, to people in this customer solution department