Gemma Freeman

Exploring the Landscape of Mutations in PTPN11

As genetic testing becomes more and more common, the importance of accurately interpreting the results of these tests increases. Point mutations in the gene PTPN11 can manifest a range of outcomes. Genetic testing on PTPN11 mutations can be inconclusive in determining how severe, if at all, the effects will be. LEOPARD's Syndrome and Noonan's Syndrome are two phenotypically similar genetic diseases associated with PTPN11. Despite their similarities, the two diseases manifest and are treated in distinctly different ways. This project investigated the differences between LEOPARD's Syndrome and Noonan's Syndrome through the use of computational biology techniques. The goal was to create a way to classify mutations based on their effect on a certain feature of the protein.

PTPN11 expresses the protein SHP2, a key regulator in cell proliferation. SHP2's role is to dephosphorylate signal proteins. Dephosphorylation activates the signal proteins and causes downstream pathway activation. SHP2 consists of three distinct domains: the PTP domain, the N-Sh2 domain, and the C-SH2 domain. The closed state is formed when the N-Sh2 region folds over into the PTP domain blocking the catalytic site causing the protein to be inactive. In normal protein function, there is a balance in how often SHP2 adopts the open and closed states.

When SHP2 is mutated in Noonan Syndrome, it tends to favor its open conformation. These mutations lead to an increased rate of dephosphorylation which dysregulates the associated signal pathways, and therefore, the cell's proliferation. This dysregulation can result in heart defects, short stature, low set ears, and other symptoms. In LEOPARD's Syndrome, SHP2's catalytic site begins to favor the signal protein GAB1 as well as adopting an even more open

conformation. The signal pathway associated with GAB1 becomes overactivated leading to similar symptoms as Noonan syndrome as well as cafe au laite spots on the skin.

I was given access to a data set from a sample of patients describing the location of the mutated amino acid, the resulting amino acid change, and which syndrome this mutation manifested. The manifestation of either LEOPARD's Syndrome or Noonan Syndrome could be explained by a variety of different causes related to the mutations. In this project, I investigated the change in the change of enthalpy, the solvent accessible surface area, and the distance from certain structural features of the protein.

The folding energy of proteins can be measured by the change of enthalpy (delta g). The energy change between normal protein folding and mutated protein folding is the change in the change of enthalpy (delta delta g). If the mutated protein has a positive delta delta g it is more prone to misfolding whereas a negative delta delta g is a more stable protein. One proposed explanation for the difference between Noonan and LEOPARD mutations is that mutations associated with one syndrome could have positive ddg while the other has a negative ddg.

The surface area of a protein can be affected by mutations that change the polarity and hydrophobicity. Since negatively charged amino acids cluster towards the hydrophobic core, a change in polarity can change the way the protein folds and hence the surface area. The solvent accessible surface area (sasa) is represented by rolling a ball across the surface of the protein and measuring the area created by the path the center point of the ball followed. If the set of mutations for one syndrome increased the sasa while the other decreased it, sasa could be a classifier for which mutation would cause which disease.

Within the domains, SHP2 is made up of loops, alpha helices, and beta pleated sheets. The catalytic loop is the location of the catalytic activity which cleaves the phosphate from the

substrate. The E loop helps geometrically orient the phosphate to be dephosphorylated by forming hydrogen bonds with the other loops. The Q loop extends into the catalytic cavity of the PTP domain and ensures the catalytic loop is stabilized. The D'E loop interacts with the catalytic cleft and allows for the inactive form of SHP2. Since mutations near loops may cause destabilization and impairment of their function, I investigated whether the position of the mutations about these structural features was a classifier for disease. The distance from the mutation and the loop was calculated using a closed 3D structure model of SHP2.

To determine which of these was the best classifier for Noonan vs. LEOPARD's Syndrome, I ran a logistic regression with each variable against the true outcomes to see which proposed explanation best predicted if a mutation was Noonan's or LEOPARD's Syndrome. The lowest p-value and therefore the best predictor ended up being the distance between the mutation and the catalytic site.

A receiver operating characteristic (ROC) curve was then created to find a cutoff distance that optimized the number of true positives while keeping a low percentage of false positives. Mutations within 11.3 Angstroms of the catalytic site were classified as LEOPARD's syndrome while mutations further than this were classified as Noonan's syndrome.

Although this distance from the catalytic site made for a good classifier, around twenty percent of the data remained unexplained. After filtering the data so that only the unexplained remained, I ran the same process over again on the unexplained data. The lowest p-value for this set was the distance from the D'E Loop. After creating a second ROC curve, mutations within 17.6 Angstroms of the D'E loop were concluded to be LEOPARD's Syndrome.

Although mutations on PTPN11 can result in Noonan Syndrome, LEOPARD's syndrome, or even cancer, some mutations can manifest as benign. I reran the above steps on Noonan's

syndrome versus benign mutations as well as LEOPARD's syndrome versus benign mutations. It turned out that mutations that caused Noonan Syndrome tended to cluster more towards the Q loop than benign mutations. Mutations that caused LEOPARD's Syndrome were closer to the D'E loop than benign mutations.

Since LEOPARD's Syndrome mutations cluster near the catalytic site, this accounts for the change in specificity of the catalytic site. Several tests could be run to determine whether these cutoffs are accurate. The first way would be to access another large dataset with the numbered amino acid mutation, the change in amino acid, and the resulting syndrome. The algorithm would then be run on the new dataset and the output of the predicted syndrome compared with the actual reported syndrome to determine the accuracy of this classifier. Another way is to create proteins with all possible point mutations. These created proteins could then be placed in an assay with and without GAB1. Proteins with mutations within 11.3 Angstroms of the catalytic site would be predicted to have low activity without GAB1 and heightened activity in the presence of GAB1.

Throughout the course of this project, I had the chance to learn about how research is conducted, how labs are run, and how research is conveyed. I began the research project with around a week dedicated to reading literature devoted to the subject of PTPN11, genetic diseases, and protein structure and function. After another week of adapting the code to fit my protein of interest, I began to formulate hypotheses as to what differences these mutations had. After doing another round of reading research and tweaking my hypotheses, I realized how cyclical the research process is and how important it is to continue making changes to the original ideas and plans.

During my time researching I had the chance to attend weekly team lab meetings as well as sit in on another principal investigator's lab meeting. I learned how interconnected the research conducted by members of the same lab is. While I worked on the gene PTPN11, another member of the lab team was investigating other genes responsible for Noonan Syndrome. While discussing my career goals, my supervisor, Dr. Hall, told me that while you give up the "hands-on" research aspect by becoming a principal investigator you get the opportunity to execute your own research vision. He further told me how my research project could be incorporated into further research projects or grant proposals. I also learned the importance of discussing and presenting research to peers. While presenting my research to the lab I was able to gain insight into areas of my project that could use improvement. Working in the lab also gave me the chance to speak with a geneticist at UCLH who works with parents whose fetuses have genetic mutations in PTPN11. She helped me to understand how important this research is to prospective parents who want to understand the risks these mutations pose to the fetus.

During my project, I attended a seminar, viewed posters, and created a poster. Since all the information that I generated during the project could not fit into the poster, I had to learn how to pick which data would most succinctly and accurately convey my project. I was advised to use pictures and data representations while using fewer words. My supervisor further gave me guidance on how to best present and answer questions, including how to best take a pause and recollect my thoughts.