



AI Genomics Hackathon

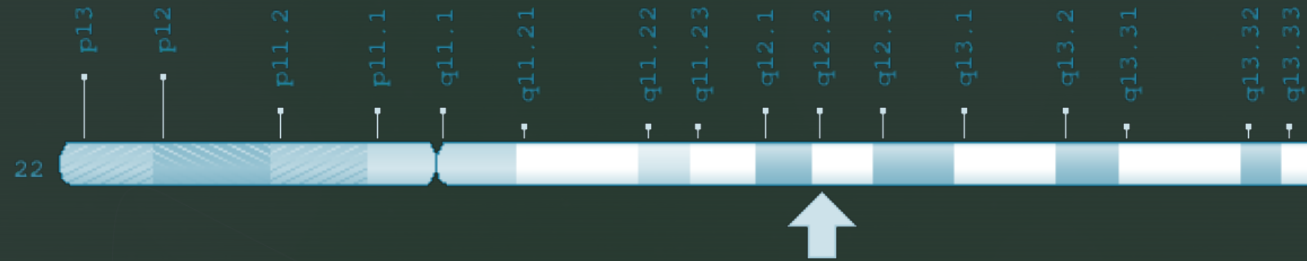
Project: SNP and Indel Density Cluster Ranking

Team Edward: Joshua Levy, Hide Shidara , Jordan Levy,
Kevin Tran, Mustafa Naik

Problem

- **Is NF2 driven solely by a mutation in the “NF2 gene”?**
- Is inactivation of NF2 gene the only reason for people to develop schwannomas, or might there be other gene alterations that influence tumor behavior? What other affected genes could play a role in the formation and growth of these tumors?

Goal:

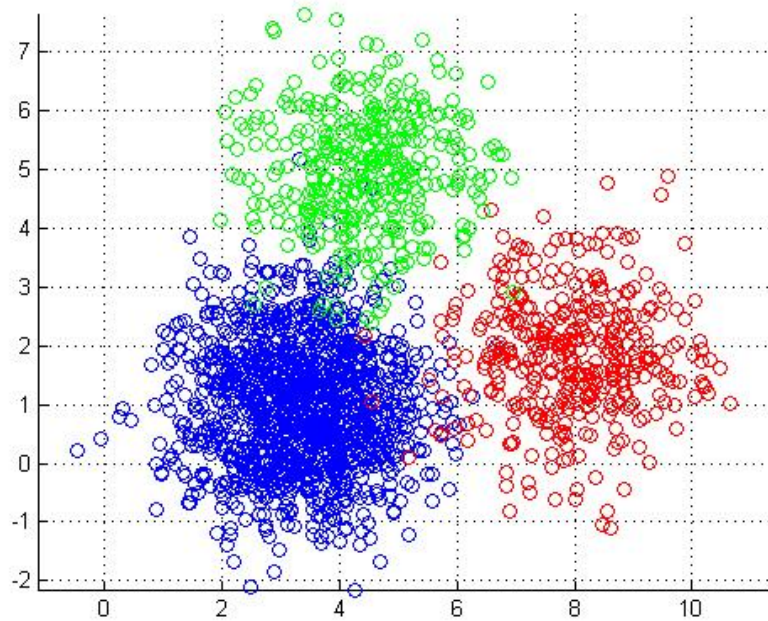


- Locate the NF2 regions
- Measure SNP and indel densities in the NF2 gene
- Find a pattern in the densities
- Compare this to genes without NF2 to see if they have similar characteristics

Retrieving Data

- Training Data: NF2 gene data from clinVar database hg19. Get the SNP and inDel data from vcf files.
- Testing Data: Onno's tumor mutation data SUBTRACT Onno's brother's non-mutation data + ENSEMBL. (Thanks Onno & bro!)

Methodology



- Find SNP and inDel density in test/train data set
- Trained a K means Clustering algorithm to characterize a gene by its distribution of SNP densities
- Cluster SNP and inDel density features → NF2 cluster model
- Used NF2 cluster model to identify genes with similar characteristics

Current Progress & Result

- Trained kmeans cluster model (SNP+indel density clusters) using NF2 training data
- Sparsely sampled Onno's test data and ranked genes
- Result: trained cluster model, but have not ranked test data

Conclusion

- Goal: find other genes that may cause NF2 tumor to develop
- Used NF2 variant data from ClinVar to train K-means clustering models, thus identifying preliminary SNP/indel density characteristics
- Attempted to use the model to understand non-NF2 genes
- Can use positional SNP density and indel density for annotation
- Can be used as a future metric to study NF2-like genes

Future Direction

- Use other metrics other than SNP and Indel densities
 - SNP density 100 genomes or allele frequency
- Study the rest of Onno's chromosomes
- Create a better fit model
- Explore integration with neural models