**BENG203: Genomics, Proteomics, Network Biology**

**Bafna/Ideker Spring 2015**

**FINAL PROJECT: Variant Classification**

**Challenge:** You have obtained the genotype of a person X diagnosed with a genetic disorder. Think of it as a list of chromosomal positions of variants, where at least one of the two nucleotides in that person X is different from the reference. For example

Chr 17 A (*Reference*) A/T (*X’s genotype*)

: :

You need to find out if any of the (e.g. A🡪T mutation) variants could be deleterious. There are many clues you can use to do that. For example, you could check if the mutation changes the sequence of some critical protein; independently, you could check how likely A is to have mutated in mammals (If it is conserved, the mutation is more likely to be deleterious).

In this project, you can model the problem as a supervised classification problem. An exemplar data-set of positive (deleterious variants) and negative (benign or neutral) is provided as a list of coding variants on proteins (Uniprot ID). The different clues act as features that you can use, but will need to be compiled from public resources. Remember to separate the data into training and test data-sets. However, you can also identify other data-sets that focus on non-coding variants.

**Reading:** Read about other approaches, e.g., SIFT, Polyphen, GERP scores, eXtasy for coding variation.

**Data:** You are free to use any datasets that are publicly available and fit some aspect of the problem (input/output/priors/etc). Here are some data that might be useful to various aspects of the project:

Mutations/Variants

TCGA: [https://tcga-data.nci.nih.gov/tcga/](https://tcga-data.nci.nih.gov/tcga/" \t "_blank)

GWAS CATALOG: [https://www.ebi.ac.uk/gwas/](https://www.ebi.ac.uk/gwas/" \t "_blank)

Gold standard disease genes

COSMIC: [http://cancer.sanger.ac.uk/cosmic](http://cancer.sanger.ac.uk/cosmic" \t "_blank) (Cancer)

Polyphen Data, where they have curated mutations benign and deleterious:  
[http://genetics.bwh.harvard.edu/pph/data/](http://genetics.bwh.harvard.edu/pph/data/" \t "_blank)

Protein networks as prior knowledge

HumanNet: [http://www.functionalnet.org/humannet/about.html](http://www.functionalnet.org/humannet/about.html" \t "_blank)