Title: Identifying Thyroid Cancer Subtypes through Publicly Available Omics Data

Group members: Kun-Hsing Yu, Luke Pappas, Michael Fitzpatrick, Jessica Kung

Proposal abstract:

Thyroid cancer is one of the few cancers that have increased in incidence rates over recent years. The American Cancer Society estimates that in 2014, there will be about 60,220 new cases of thyroid cancer in the U.S. and about 1,850 people will die from this disease each year. Current subtyping system is largely based on microscopic characteristics, which doesn't always correlate with clinical outcomes. In addition, the biomolecular mechanisms leading to different clinical phenotypes have not been fully elucidated.

The recent availability of cancer omics data has created unique opportunities for the development and refinement of subtyping systems. Consortiums like The Cancer Genome Atlas (TCGA) profiled the genomic variation, DNA methylation status (epigenomics), gene expression (transcriptomics), and protein expression and modification status (proteomics) by next-generation sequencing / high-throughput modalities. Through careful analysis of the omics profiles, we can build a clinically relevant subtyping system and generate important insights into the genomic mechanisms contributing to clinical outcomes.

In this project, we aim to:
1. Obtain the data, which is freely downloadable from the TCGA website (https://tcga-data.nci.nih.gov/tcga/).
2. Identify patterns of genomic / epigenomic / transcriptomic / proteomic changes that predict clinical outcomes (tumor stage, survival) of thyroid cancer patients through supervised machine learning methods.
3. Determine the biological mechanisms associated with clinical outcomes through enrichment analysis (using GO analysis for genes and GREAT for non-coding genomic regions) and pathway analysis.
4. Analyze the extent to which the identified genes and enrichments make sense, in how they indicate relevant biological mechanisms.

We plan to implement software tools in Python and R to perform the data analysis to obtain our results. From our own perspective, we hope to practice using the analytic tools introduced in class and consolidate the concepts we learned through this project.