

Week 3 Homework: TCGA Scavenger Hunt

Home page: National Cancer Institute; genomic, transcriptomic

Program history: TCGA has bolstered the field of computational biology by providing vast, rich datasets that can be utilized for various purposes such as for predicting genes of prognostic significance; October 2012; The paper I reviewed was “Human Subjects Protection and Data Access Policies”. One approach to protecting human subjects outlined was that the donors of tissues (and their associated generated data) would have to be contacted and consented to maintain their contribution to the TCGA datasets. This policy of informed consent ensures that contributors to the TCGA are fully aware of where and for what purpose their data is being collected.

TCGA Cancers Selected for Study: Poor prognosis, overall public health impact, Availability of samples meeting standards for patient consent; Men can also have breast cancer, although it is rare. In fact, 390 American men died from breast cancer in 2010.

Publications by TCGA: GATA3, PIK3CA, MAP3K1, HER2, PTEN, AKT1, TP53, CDH1, RB1, MLL3, CDKN1B, TBX3, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1 and CCND3.

Using TCGA: 72 projects; 33 projects; TP53, 35%; Based on the Overall Survival Plot from the Mutations tab, we can see that approximately 50% of breast cancer patients survive past 11 years since their diagnosis. After 20 years, only about 30% of patients survive.

Discussion:

1. TCGA serves as a one-stop source for relevant clinical and multi-omic data sets pertaining to various cancer types. By providing access to such data, TCGA aims to accelerate cancer and computational biology research across the world since vast amounts of rich data are within reach for

researchers and scientists. Given the difficult nature of obtaining access to patient clinical and genomic data, TCGA is very unique for its accessible and granular datasets.

2. Given that it contains both patient demographic and genomic data, TCGA is phenomenal for identifying genes of interest across different races that can serve as prognostic markers. For instance, we can investigate the prevalence of mutated genes across different races to see if there are different mutation rates and patterns for each race, making their individuals more susceptible to specific cancers.
3. A very evident benefit of TCGA is the degree of accessibility that it offers to researchers, clinicians and scientists around the world. With the uniform distribution of computational resources around the world, the next innovation in cancer and/or computational biology can frankly happen in any nation and by any individual. Such a statement is corroborated by the breakthroughs in the protein folding problem by the Google DeepMind team, which was more of an industry think tank than an expert cancer biology research group. Therefore, akin to how PDB provided the data imperative for the Google team to innovate, TCGA can be the crux of global research in cancer biology. However, a subtle yet equally important detriment of TCGA may be that faulty analyses of the openly available data by certain parties can lead to the dissemination of erroneous conclusions that can set back research progress.