Sasha Yousefi
CME291 - Xplore (Spring 2022)
Mentors: Alex Ioannidis
Number of units: 1

Mentor Background:

Dr. Alexander Ioannidis (PhD, MPhil) graduated summa cum laude from Harvard University in Chemistry and Physics and earned an M.Phil in Computational Biology and Diploma in Greek from the University of Cambridge. His Ph.D. from Stanford University was in Computational and Mathematical Engineering, where he still teaches machine learning and data science as an Adjunct Lecturer in the School of Engineering. He also has an M.S. in Mgmt. Sci. and Eng. (Optimization) from Stanford. Prior to Stanford, he worked in superconducting computing logic and quantum computing at Northrop Grumman. As a current research fellow in the Stanford School of Medicine (Department of Biomedical Data Science), his work focuses on applying computational methods to problems in genomics and population genetics.

Problem Statement:

The SARS-CoV-2 pandemic has differentially impacted populations of varied race, ethnicity, and socioeconomic status. Utilizing tools such as GWAS, scientists have identified several genetic loci that are contributors to susceptibility and severity of COVID-19 infections. However, epidemiologists have also shown that socioeconomic status, exposure risk, sex, and other comorbidities play a major role in suceptibility and severity. In order to get an accurate picture of which portions of the genomes affect infection severity, we must control for these socioeconomic and external factors. Admixture mapping and local ancestry inference are imperative tools in order to decouple genetic risk with confounding factors. We will use these methods in order to identify genetic loci for COVID-19 risk.

Proposal:

This project stems from the paper *"Deconcoluting complex correlates of COVID19 severity with local ancestry inference and viral phylodynamics: Results of multinomic pandemic tracking strategy"* Parikh et al. (2021), where the researchers randomly collected and sequenced 938 positive and 389 negative covid samples from the population near Stanford University. These tests were conducted between March 2020 and June 2020. In addition to collecting demographic data for each patient, the researchers collected a severity score based on the scale proposed by the WHO. The researchers aimed to identify regions of the genome that contribute to COVID19 illness severity. To eliminate confounders, the researchers used genetic ancestry inferences to identify subpopulations that were highly impacted by the pandemic. This genetic ancestry proportion (along with age, sex, and BMI) were adjusted for when assesing disease severity with genetic markers. This analysis revealed loci in chromosome 5 of individuals with African ancestry and chromosome 14 in individuals with Oceanic ancestry that were associated with increased disease severity.

However, there are some improvements to be made on this model. Admixture mapping association analyses were used to regress the residual of severity of COVID19 symptoms for each patient. In order to conduct this analysis, the researchers utilized a p-value threshold of <= 2.57 x 10^-5 as a test for significance. This is utilizing the Bonferroni corrected p-value with the assumption that the genome is divided between 19,474 *independent* windows for local ancestry assignment. However, due to crossover events, we cannot assume that these 19,474 locations on the genome are inherited independently of each other. This changes our initial assumptions of significance. Individuals with higher levels of genetic mixing (i.e. African descent) are likely to have a higher level of independence amongst the 19,474 windows. Additionally, this phenomenon suggests that the date of admixture plays an important role. Individuals with more recent admixture events are unlikely to have undergone enough recombination events for there to exist relatively independent genomic windows.

Throughout this quarter, our project will aim to identify more accurate p-value levels for the various ancestries presented in the Parikh et al. (2021) paper. Jessica and I will make use of the STEAM package (https://github.com/kegrinde/STEAM) in order to conduct our analyses. This package estimates genome-wide significance thresholds for admixture mapping studies. These results are crucial in determining which regions of the genome are associated with

COVID-19 severity. This study will be quite useful for resource management when it comes to COVID cases, and can assist analysis of different diseases in the future.