Translational AI Center (TrAC) Seminar Fall 2022

Henri Chung
September 23 at 12:00 noon (US Central Time)
Physical location: 2206 Student Innovation Center
Zoom: https://iastate.zoom.us/j/92178103551?pwd=dINCa2l0ckVBTEVyR1JEN2Y3b21XQT09

Predicting Antimicrobial Resistance with Machine Learning: A Cautionary Tale of Missing Data

Abstract

Understanding the genetic mechanisms of antimicrobial resistance (AMR) is important in the early detection of emerging AMR infections, a pressing global health threat in human and veterinary medicine. The United States Department of Agriculture (USDA) launched the National Animal Health Laboratory Network antibiotic resistance pilot project in 2018 to monitor AMR trends in pathogens routinely isolated from sick livestock and companion animals. We examine whole genome sequencing and antibiotic susceptibility data from Escherichia coli (E. coli) to identify isolates' AMR genotypes and phenotypes. We then tested different potential modes of AMR genotype translation into resistance phenotypes using logistic regression models and evaluated the ability of random forest models to predict antimicrobial resistance. Our regression results show that AMR genotype data can be used to predict antibiotic resistance with varying levels of accuracy: individual AMR genes are strong predictors of AMR, followed by the binary presence of any gene within an antibiotic group. Random forest models predict resistance phenotypes with excellent recall but low negative predictive value, irrespective of a priori assumptions about drug resistance conferred by each gene. We acknowledge limitations in our machine learning results due to class imbalances and biases in our data and suggest potential guidelines to prevent these issues in future studies. Overall, we show that there are several contributing factors which significantly affect how AMR genes contribute to a resistant phenotype and suggest that careful and deliberate research is necessary to understand these differences as we begin to apply machine learning methods to genotype-phenotype association in AMR infections.

Short Bio

Henri Chung is a fourth year PhD candidate in the Bioinformatics and Computational Biology program, and is performing his research in Dr. Iddo Friedberg's lab. His research primarily covers methods to elucidate protein function, with special interest in bacterial systems. He most recently presented a project, "Identifying functional relationships between proteins with phylogenetic profiles", at the Intelligent Systems for Molecular Biology Conference in Madison, Wisconsin. Personal interests include making plots, cooking, and thrifting.