# Battling Superbugs: A Machine Learning Framework for Identifying Novel Antibiotic Resistance Factors

Bhullar, Harkirat (School: Evan Hardy Collegiate Institute)

Antimicrobial resistance (AMR) is a major threat to global public health with 10 million people at risk of dying from AMR-related causes by 2050. Understanding the underlying biological factors of resistance is crucial in the battle against AMR driven “superbugs”. Unfortunately, existing computational techniques used to uncover these factors lack novel insights and have inadequate biological interpretability. The objective of this work was to develop a fast, accurate and interpretable computational framework that utilizes machine learning to identify novel (and validate known) resistance factors from whole-genome sequencing (WGS) data. The developed framework used decision tree models to identify resistance factors. Genomic features (like single nucleotide polymorphisms [SNPs]) were extracted from WGS data. These features were used as input into multiple decision tree models to predict AMR. Each model’s performance was evaluated with 10-fold cross-validation. The structure of the model with the highest performance was probed to identify genomic features predictive of resistance. To demonstrate the power of the developed framework, it was used to identify resistance factors for 5 antibiotics in Neisseria gonorrhoeae (bacterium responsible for the sexually transmitted infection gonorrhea). The alternating decision tree model reported the best performance with an average accuracy of 95% and 90.9% for resistant and susceptible strains respectively. Overall, the framework uncovered 45 novel (and validated 6 known) resistance factors. This first-ever interpretable machine learning framework has the potential to provide unprecedented insight into the underlying mechanisms of AMR. Ultimately, it can also facilitate targeted antimicrobial treatment and drive drug development.

**Awards Won:**