To Kill a Mycobacterium:
Novel Antibiotic Discovery in *Mycobacterium tuberculosis H37Rv*

[1]http://www.who.int/mediacentre/factsheets/fs104/en/
[2]http://www.cdc.gov/drugresistance/biggest_threats.html
[3]https://www.collaborativedrug.com/buzz/2012/10/16/cdd-releases-the-first-free-mobile-app-for-tuberculosis-researchers-tb-mobile/
[4] https://pubchem.ncbi.nlm.nih.gov/bioassay/1332
[5] https://pubchem.ncbi.nlm.nih.gov/bioassay/1626#section=Top

INTRODUCTION

Pathogenic evolution is an arms race. Humans flood the planet with antibiotics, and bacteria are forced to cope with the onslaught, do or die, relying solely on the grace of genetic mutation and population dynamics to survive. Be that as it may, modern medicine is falling short. Evolution has, of course, risen to the task and hyper-impervious, antibiotic-resistant strains continuously emerge. As such, today's post-penicillin World is facing a medical crisis. Humanity's current library of antibiotics is no longer panaceaic, and new antibiotics are required to bastion the public's health.

One of the most notorious antimicrobial-resistant strains is M. tuberculosis H37Rv, a resilient mycobacterium invulnerable to all clinically-available antibiotics. This poses a major issue, seeing as one-third of Earth's population already carries some form of M. tuberculosis in its many guises, 2 million of which die yearly[1]. Furthermore, 10,042[2] Americans develop treatment-resistant tuberculosis. How many deaths could we expect once natural selection enriches the global presence of H37Rv? The effects could be catastrophic and citizen scientists have aligned with The WHO and other organizations[3] to fulfill the gaping need for new antibiotics.

This experiment is one such an attempt. Thereby, publically available data was curated from high-throughput compound screens in H37R and scrutinized in hopes of identifying pharmacophores (e.g. the molecular motifs and fragments that define a chemical's active properties) capable of predicting bactericidal activity in uncharacterized compounds.

METHODS AND RESULTS

DATA COLLECTION

PubChem hosts a wealth of publicly available datasets targeted for antitubercular discovery. For the purposes of this study, BioAssays AID1332[4] and AID1626[5] were web-scraped to substantiate the training dataset. The amalgamated datasets amassed a suite of 215,000compounds, all of which were screened *in vitro* with H37Rv. Activity of each compound was classified by observing compound lethality over a range of concentrations and timepoints. Chemicals that incurred >90% fatality amongst its initial bacterial-population were labeled "Active", whereas all other compounds were annotated as "Inactive." All activity assays were executed robotically in a highly controlled experiment as previously described [4, 5].

MOL-TO-VEC FEATURE REPRESENTATION

Each activity screen was annotated with molecular-structure data encoded as SMILES strings--a streamlined system for representing chemicals in a single line of text. SMILES strings were then vectorized in RDKit using the Morgan Fingerprint convention, as parsimonious strategy that leverages binomial trees and hash-function compression. This form of molecular

[1] http://www.who.int/mediacentre/factsheets/fs104/en/
[2] http://www.cdc.gov/drugresistance/biggest_threats.html
[3] https://www.collaborativedrug.com/buzz/2012/10/16/cdd-releases-the-first-free-mobile-app-for-tuberculosis-researchers-tb-mobile/
[4] https://pubchem.ncbi.nlm.nih.gov/bioassay/1332
[5] https://pubchem.ncbi.nlm.nih.gov/bioassay/1626#section=Top

representation is highly fidelic, capable of resolving molecular characteristics down to an enantiomeric level.

MODEL DEVELOPMENT
*****Enumerate all Putative models and optimisation of final (XGBLogistic Classifier)*****

MODEL VALIDATION
*****+Train and hit PCA*****
*****Superposition Electron Density Maps*****

DISCUSSION

[1]http://www.who.int/mediacentre/factsheets/fs104/en/
[2]http://www.cdc.gov/drugresistance/biggest_threats.html
[3]https://www.collaborativedrug.com/buzz/2012/10/16/cdd-releases-the-first-free-mobile-app-for-tuberculosis-researchers-tb-mobile/
[4] https://pubchem.ncbi.nlm.nih.gov/bioassay/1332
[5] https://pubchem.ncbi.nlm.nih.gov/bioassay/1626#section=Top