## REVOLUTIONIZING BIOMEDICAL RESEARCH: INTEGRATING CUTTING-EDGE AI/ML TO UNLEASH INNOVATION IN DRUG DISCOVERY AND THERAPEUTICS DEVELOPMENT

# Predictive Genotype-to-phenotype Modeling And Applications In Antivirals

**Phillip Davis** MRIGlobal

The rapid expansion of machine learning applications in biological sequence analysis is revolutionizing our approach to medical and defense-related challenges. With the advent of next-generation sequencing technologies, a vast reservoir of genomic data is now available, providing unprecedented opportunities for predictive analytics in healthcare and biosecurity. Our project harnesses this potential to address urgent needs in pathogenic threat detection and mitigation, aligning closely with the strategic objectives of the Defense Threat Reduction Agency's Joint Science and Technology Office (DTRA JSTO).

Utilizing a robust framework of feature engineering and ensemble machine learning methods, we have developed a predictive model for human pathogenicity within the Orthocoronaviridae family of viruses, which includes the pathogens responsible for COVID-19. By analyzing codon usage frequencies and the Orf1ab region of the viral genome, our model achieves a remarkable 97.5% out-of-bag accuracy in predicting human pathogen phenotypes. We show how relaxing assumptions that constrain traditional codon analysis methods, such as Relative Synonymous Codon Usage (RSCU), facilitate this performance. Our breakthrough model is robust and facilitated by novel approaches in feature selection that maintain predictive power regardless of data subset variations.

Our findings are particularly significant for the joint forces, offering a strategic advantage in early detection and response to biological threats by surveilling specifically for phenotype-of-concern directly at the genomic level. The ability to predict pathogenicity from genomic data enables rapid identification of potential threats in diverse operational environments, enhancing the warfighter's situational awareness and readiness. Furthermore, correlation analysis between our model's predictions and changes in cellular tRNA abundances in SARS-CoV-2 infected cell lines has provided insights into viral behavior and host interaction, suggesting a universal fitness requirement for RNA viruses that could be exploited for therapeutic interventions. There may be tight requirements for host tRNA distributions that the virus requires for effective translation. For example, too little of a particular tRNA and translation slows too much to out-compete premature tRNA decay. Too much of the tRNA and translation proceeds to quickly, leading to failure of co-translational machinery. This concept is summarized in Figure1. We discuss the translational potential of tRNA-based therapeutic technologies, originally developed for rare genetic disorders, as novel antivirals. By repurposing these cutting-edge treatments, we propose a pathway towards broad-spectrum antiviral countermeasures that can be rapidly deployed against emerging viral threats, supporting the DTRA JSTO mission to safeguard the joint forces through enhanced biodefense capabilities. Our research not only advances the field of bioinformatics and machine learning in genotypic to phenotypic predictions but also provides actionable insights for 1) enhanced genome-level biosurveillance for specific phenotypes-of-concern, even if in previously-unseen organisms that are not in available reference databases, and 2) the development of novel, effective countermeasures against pathogenic threats. These capabilities align directly with the DTRA JSTO's goal of integrating cutting-edge science and technology to enhance the nation's defensive measures against chemical and biological threats, thereby bolstering the security and operational effectiveness of the joint forces.