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

Mixed CNN-attention Machine Learning Model For Predicting Gene Regulatory Relationships Across Fungal Species Towards Developing Computational Methods For Defending Against Emerging Pathogenic Fungi.

Laura Weinstock Sandia National Laboratories   **Jenna Schambach** Sandia National Laboratories   **Anna Fisher** Sandia National Laboratories   **Elizabeth Koning** Sandia National Laboratories   **Wittney Mays** Sandia National Laboratories   **Cameron Kunstadt** Sandia National Laboratories   **Raga Krishnakumar** Sandia National Laboratories

Background: Disease-causing fungal pathogens present a major challenge to both the warfighter and domestic populace health. In particular, drug resistance and climate-driven pressure are accelerating the potential for emerging fungal pathogens are exacerbating the issue, particularly in parts of the world that our warfighters are deployed to and are likely to harbor unknown and evolving pathogenic species. Moreover, the fungal kingdom is immensely diverse and largely uncharacterized, which presents a challenge for the rapid discovery, development, and efficient repurposing of countermeasures to combat fungal infection.

Developing reliable interventions for addressing pathogenicity in diverse fungi at scale remains a significant challenge. Determining ground rules of gene regulation and their applicability across fungal species is critical for enhancing preparedness by narrowing the search-space for countermeasure development, particularly in novel emerging species. Epigenetic modifications play a crucial role in regulating gene expression and elucidating conserved and divergent patterns in epigenetics regulation across fungal species will significantly improve the identification and development of novel therapeutics in existing and emerging fungi. Recently, there have been efforts to better understand the relationship between gene sequence, epigenetic modifications, and gene expression within fungal species through the use of machine learning and deep learning (ML/DL) methods that are able to ingest the high-dimensional, complex sequencing data. However, there remains limited understanding of how similar epigenetic modifications impact function and fitness across fungal species.

Objective: The discovery of conserved epigenetic modification design rules that control gene expression would greatly improve the efficiency of therapeutic discovery and countermeasure development in preparation for threats spanning the fungal kingdom. In this study, we aimed to predict gene expression levels based on combinatorial epigenetic modification expression within and across fungal species. By predicting the effect of epigenetic modifications on gene expression, we can a) predict more accurately how a given strain changes when modulating epigenetics and b) determine how to target critical genes and cellular functions as a means to identify targets for antifungal therapeutics.

Methods & preliminary results: We trained and tested ML/DL models within and across fungal species (both pathogenic and non-pathogenic) to identify the relationship between epigenetic modification features and gene expression. We tested a battery of models, ranging from regression to deep neural networks, on increasingly complex engineered data features to not only predict gene expression based epigenetic modifications, but also to understand the degree of complexity required to make accurate predictions. Our model out-performs off-the-shelf solutions for cross-species applicability, suggesting that the architecture that we have developed is specifically suited to identify common design rules across species, a critical finding for establishing a computational line of defense against novel emerging fungi.

Impact: The multi-omic data analysis and predictive modeling capability developed in this effort may support the rapid identification of novel countermeasures for diverse emerging fungal diseases, including in areas of the world accessed by our warfighters. It also provides a framework for evaluating how to exert more precise control of pathogenic function and fitness that may support reduced drug resistance.