TECHNICAL ABSTRACT

Our project aligns with the FY23 PRMRP Portfolio Category: Infectious Diseases, FY23 PRMRP Topic: proteomics, and FY23 PRMRP Strategic Goal: Epidemiology: Identify strategies for surveillance or develop modeling tools and/or biomarkers to predict outbreaks or epidemics.

Influenza A viruses, with their segmented genome and high prevalence in animal hosts, can recombine genes from animal strains and (re)emerge as novel human pathogens. Pandemics triggered by animal Influenza A strains spilling over into humans have occurred multiple times in the past century. Mitigating such risk involves identifying animal strains that have high odds of spilling over into humans and rapidly achieving human-to-human transmission capability. While global surveillance efforts collect wild specimens annually, our ability to reliably and scalably risk-rank individual strains according their pandemic potential remains limited. CDC's current solution to this problem is the Influenza Risk Assessment Tool (IRAT), which is based on subject matter experts scoring strains-of-concern on ten different phenotypic aspects related to transmission and pathogenicity on the basis of multiple experimental assays. Thus scoring each strain can take weeks to months. With tens of thousands of strains being collected annually, this presents a scalability bottleneck.

Here we propose to develop the BioNORAD platform to proactively identify the risks posed by emerging Influenza A strains with high pandemic potential, akin to the strategic function of NORAD in defending the North American airspace. Powered by novel pattern discovery algorithms, our proposed algorithms will automatically parse emergent evolutionary constraints on Influenza A viruses in the wild. This platform will provide a less-heuristic, theory-backed, experimentally validated, scalable solution to emergence prediction, preempting strains expected in future human circulation and approximating IRAT scores of non-human strains a million times faster (seconds as opposed to weeks/months) without experimental assays or SME scoring.

Our central insight here is that to preempt strains expected to be in future circulation, we need to reliably estimate the non-heuristic numerical probability of a strain x spontaneously giving rise to strain y in the wild. To accomplish this, we aim to learn the complex cross-dependencies that constrain what a "valid alteration" of an amino acid (AA) sequence looks like (random perturbations are expected to be large deleterious), by first analyzing variations of residue sequences of key proteins implicated in cellular entry/exit, namely HA and NA, and then expanding the analysis to the complete viral genome. Our core algorithm, the Emergenet, leverages the inferred cross-dependencies to estimate the odds of a specific mutation arising in the future, and consequently the probability of a specific strain spontaneously evolving into another.

We will validate our ability to predict future mutations by showing that Emergenet predicted HA variants express correctly, are functional, and maintain fitness unlike random modifications, via demonstrating proper folding, cell surface expression of variants by flow cytometry as well as evaluate the fitness of recombinant viruses in 1:1 competition experiments with the unperturbed parental strain.

Finally these models will be assembled into a integrated platform (the BioNORAD) that shows global near-time pandemic threat events as they emerge (from observation of new animal strains, as they are uploaded to the platform), along with estimates of the threat severity and estimated time to emergence.

Current surveillance paradigms, while crucial for mapping disease ecosystems, fail to address the challenge that a higher edit-similarity between strains does not not imply a high likelihood of jump. Also, recent advances in predicting seasonal strains do not generalize to predicting emergence events, especially strains that do not yet circulate in humans, underscoring the problem of unknown unknowns. This project innovates and envisions a path to acquiring this transformative capability well-beyond the state-of-art.

The BioNORAD platform is crucial to US national interest, particularly in the context of protecting DoD assets and personnel deployed in potentially high-risk centers of emergence. Additionally, the platform will enable preemptive action, such as the inoculation of animal reservoirs before the first human infection, potentially eliminating the pandemic before it has a chance to trigger. The investment in the BioNORAD platform is thus a strategic step towards ensuring the health and safety of military personnel and the success of the DoD's mission in a world where pandemic threats are a growing concern. The platform's ability to automatically factor in the evolving host immunity and the current background environment sets it apart from current state-of-the-art methods, making it a vital asset in the fight against pandemic threats and safeguarding military personnel, assets, and global health security.