AI-POWERED DIAGNOSTICS

Recent Progress Towards Rapid, Automated Design Of Nucleic Acid-based Pathogen Diagnostic Assays

Jason Gans Los Alamos National Laboratory **Patrick Chain** Los Alamos National Laboratory **Bin Hu** Los Alamos National Laboratory **Jeffrey Koehler** United States Army Medical Research Institute of Infectious Diseases **Korey Delp** United States Army Medical Research Institute of Infectious Diseases **Po-E Li** Los Alamos National Laboratory **Chien-chi Lo** Los Alamos National Laboratory **Yan Xu** Los Alamos National Laboratory **Christopher Stefan** United States Army Medical Research Institute of Infectious Diseases

The development of nucleic acid-based assays for detecting and diagnosing pathogens is time-consuming and technically challenging. The large (and continuously) increasing number of publicly available microbial genome sequences can be difficult to use in the traditional assay design paradigm of a human expert identifying conserved (and target-unique) oligonucleotide binding sites from a multiple sequence alignment of pathogen target and near neighbor gene/genome sequences. In addition to the high computational burden of computing a multiple sequence alignment for thousands of sequences, some assay formats, like Loop mediated isothermal AMPlification (LAMP), require the identification of large numbers (six to eight) of highly conserved oligo binding sites, all within close proximity (< 400 bp) in the target genomes. Finally, assay designers must address the challenge of missing and incomplete input genome sequences. In addition to missing data within existing genomes (e.g., truncated sequences and runs of completely degenerate nucleotides within sequences), how should assay designs incorporate unknown genome sequences (including unobserved genomes that currently exist in nature and hypothetical, but highly likely, genomes that have yet to evolve)?

To help address these challenges, the DTRA-funded BioAI assay design software has been upgraded and enhanced to automate LAMP assay design, and to use graphical processing units (GPUs) to accelerate assay design calculations required for processing thousands of bacterial genomes. Experimental results will be presented to assess the sensitivity and specificity of BioAI-designed LAMP assays that target a variety of pathogens. A comparison of the time required to develop pathogen diagnostic assays using a traditional “manual” assay design process and the semi-automated BioAI assay design process will also be presented. Finally, computational results demonstrating the use of a language model to predict the future evolution of the SARS-CoV-2 genome sequence will be presented. These results suggest that it may be feasible to reduce the rate of assay signature erosion by including computationally generated, hypothetical sequences, representing possible (but currently unobserved) viral genome sequence variants, in the BioAI assay design process.

DTRA eBioAID project