NEXT-GENERATION BIOAEROSOL DETECTION & IDENTIFICATION

Zoonotic And Engineered Pathogen Identification From Metagenomic Reads

Ra'ad Mahmoud, Kromek   **Peter Rhodes,** Kromek   **Robert Taggart,** Kromek   **Katie Barr,** Kromek   **Jamie Marsay,** Kromek

Background information
The COVID-19 pandemic has shown that novel viruses with airborne transmission can have a catastrophic health, social and economic impact across the world. With tools for genetic modification becoming more widespread and easier to use, it has been highlighted by numerous sources that the release of a novel viral pathogen is likely to be an attractive proposition for terrorists. Even without human-directed intent, the rapid spread of COVID-19, Monkeypox, and other historical zoonoses have shown how animal diseases of minimal concern can easily undergo the mutations necessary to cause human disease on a societal scale.

Purpose & Objective
To combat the growing threat of gene-edited diseases and the persistent danger of zoonosis, bioinformaticians and machine learning (ML) engineers at Kromek are developing a system of algorithms for the recognition of such novel pathogens, both naturally occurring and artificially manipulated. This system aims to infer both human pathogenicity and the potential for zoonosis directly from the contents of metagenomic bins.

Methods & Results
Whilst ML has had a large impact in a range of fields, it has shown the best results in areas where a large amount of data for training is available. In this vein, Kromek has developed techniques to create huge quantities of simulated genomic data from reference genomes – both publicly available and proprietary. These training reads aim to encapsulate the threats presented by both naturally mutated and artificially engineered pathogens. In conjunction with this data synthesis, novel neural net architectures are also being developed by Kromek for the express purpose of extracting the key threat indicators from genomic reads. Our system has proven capable of characterising reads from known genomes, even under the application of significant noise (10% insertion/deletion rate) at a high rate of recall and precision (>96% and 94% respectively). Development is currently engaged in dataset expansion and assessment on unseen and engineered pathogens, and preliminary results are highly promising.

Impact to DTRA mission and war-fighter
Designed for future integration into the Kromek Automated Pathogen Scanner – Air Sequencing (KAPScan-AS) system, this decision system will provide forewarning of those emergent and engineered threats which alignment-based methods cannot identify, being that those methods rely on known and fully sequenced genomes. When deployed, these techniques for genome characterisation will provide actionable intelligence to authorities in the event of an aerosol-vectored release of engineered or mutant pathogens, protecting both the civilian and the combatant.