International Alliance for Biological Standardization

4th Conference on Next Generation Sequencing for Adventitious Virus Detection in Biologics for Humans and Animal

Frankfurt, Germany
December 3-5, 2024

BLOODVIR – Surveillance system for novel viruses based on next generation sequencing and artificial intelligence

Authors:
Machyna M.[1,*], **Braun M.**[1,*], **Rubio Quintanares G.H.**[2,*], **Brückmann J.**[2], **Margiotta E.**[1], **Schatzl J.**[1], **Spekhardt D.**[1,3], **Santos P.**[3], **Miskey C.**[3,#], **Childs L.**[1,#], **Blümel J.**[2,#], **König R.**[1,#]

[1]Paul-Ehrlich-Institut, Host-Pathogen-Interaction, Langen, Germany
[2]Paul-Ehrlich-Institut, Viral Safety, Langen, Germany
[3]Paul-Ehrlich-Institut, HZG6 NGS Core, Langen, Germany

Continuous population growth and intensity of long-distance travel contribute to emergence and spread of viruses. Many of these viruses are bloodborne and a pose an immediate risk to receivers of blood transfusions. Therefore, it has become vital to establish a reliable and high-throughput surveillance system for screening of blood and blood-derived products for novel and re-emerging viruses. Technologies such as next-generation sequencing and artificial intelligence possess the capability to analyze samples in a massively parallel fashion. The surveillance system will allow continuous screening for infection risks in a representative cross-section of the population to best prepare for future virus outbreaks.

Here we report a novel metagenomic pipeline based on next-generation sequencing and machine learning capable of detecting a broad range of known as well as novel viruses. Sequencing reads from blood samples undergo removal of host sequences and are cross-referenced to a curated virus database. Simulated reads and viral spike-in reads were used to evaluate several state-of-the art classification tools for detection of known viruses and identified MiCoP as the most accurate classifier. A high sensitivity (1.0) and precision (0.875) of sequencing reads classification result in low limit of detection of 10 viral genome copies per milliliter of sample.

Unclassified reads from the previous step could originate from previously unobserved viruses and are used to predict presence of novel virus species. Selected deep neural network models were evaluated on their ability to successfully predict viral origin of sequences using a leave-one-out benchmarking strategy. We identified VirHunter architecture trained on human, bacterial and viral genome data to have the highest prediction capabilities (precision = 0.97, recall = 0.99) of novel viral sequences.

BLOODVIR detections system thus holds a potential to significantly enhance and synergize with conventional detection methods.