Genetic characterization | **Detection of adventitious agents** | Custom solutions

Detection of Adventitious Agents

The benchmark of GenomeScan's bioinformatics pipeline

The detection of adventitious agents (DAA) (viruses, bacteria, mycoplasma, fungi), which may be present in the biologically derived samples, poses a great challenge. Although a number of in-vivo and in-vitro assays are routinely utilized to assess the purity of biologics, performing numerous tests on each product requiring clearance is laborious, lengthy, expensive and time consuming. Next Generation Sequencing (NGS)-based assay can circumvent the above challenges. NGS enables an unbiased detection of all potential contaminants, with unparalleled specificity, reliability, accuracy and speed. It is, therefore, used increasingly often in clinical and biopharmaceutical manufacturing setting, guiding diagnostics and supporting quality assurance.

The outcomes of NGS-based biosafety testing depend heavily on both the wet and dry-lab procedures. Particularly from the bioinformatics perspective, the choices of classification algorithms and databases are critical. Here, we tested the performance of our own DAA pipeline. To that end, we benchmarked our DAA pipeline against 13 pipelines frequently utilized in clinical virological laboratories. We used a recently published metagenomic datasets of 13 clinical samples from patients with encephalitis or viral respiratory infections [1]. Briefly, DNA and RNA was extracted and used for library preparation, which was then sequenced. Human reads from the output FASTQ files were removed by mapping them to the human reference genome with Bowtie2. These pre-processed datasets were used as input for our DAA analysis pipeline (Figure 1).

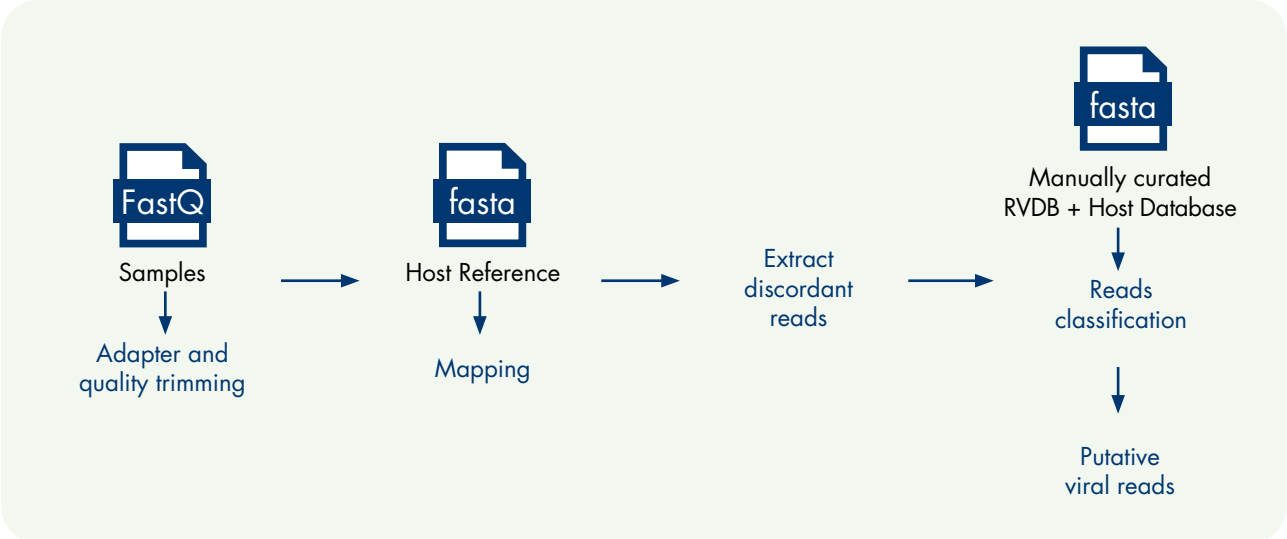

Figure 1.
DAA pipeline workflow. Adapters and low quality reads are trimmed and mapped to the host genome. Discordant reads are extracted and classified using a manually curated viral database (RVDB) and host protein sequences. Putative viral reads are reported using Krona.

Results

Sensitivity

Theoretically, pipelines for viral detection can reach a high sensitivity (low number of false negatives) at the expense of precision (increase in false positives). Figure 2 depicts the sensitivity of the tested pipelines. Notably, the least sensitive tools fail to identify approximately 25% of the clinically relevant viruses. In general, these are either low abundant viral pathogens or mixed virus infections. These were only detected by our DAA and three other pipelines out of all the evaluated tools.

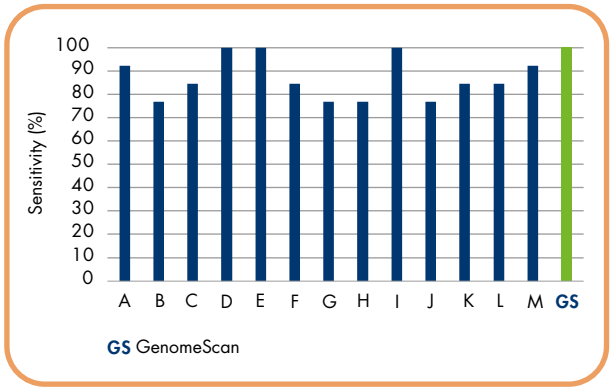

Figure 2. Sensitivity of the tested pipelines.

Taxonomic level of classification

The taxonomic levels of classification and typing of pathogenic viruses by the benchmarked pipelines are shown in Figure 3. Out of all the pipelines with 100% sensitivity (Figure 2), our methodology resulted in the most detailed level of classification. Although another tool (DAMIAN1) had a more fine-grained level of taxonomic classification, it also demonstrated much lower sensitivity of only 77%.

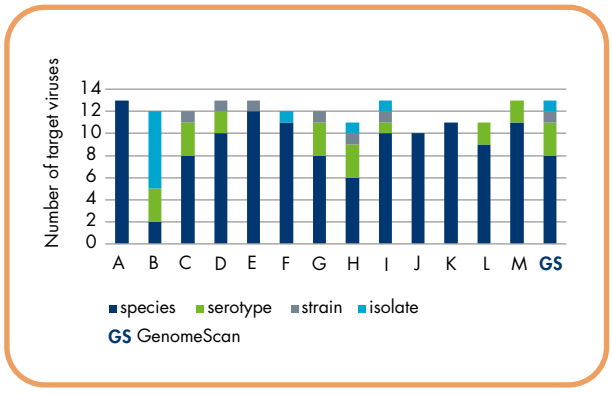

Figure 3. Taxonomic level of classification.

Precision

To determine the precision (ability to distinguish true positives from false positives) of the different pipelines, a reporting threshold of four or more reads was regarded as a positive hit in the metagenomic classification. If hits failed to be detected by PCR test, they were categorized as false positives. Results show that our DAA pipeline reached the highest precision (100%) possible (Figure 4).

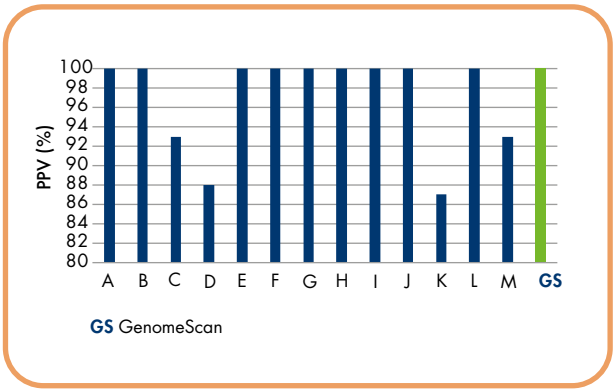

Figure 4. Precision of the tested pipelines.

Summary

Our DAA method identified most clinically relevant viruses. Moreover, unlike other pipelines, no PCR-verified false positives were reported. Notably, our pipeline was the only one with a 100% sensitivity and precision, with an astonishing isolate level of taxonomic classification.
Let us help you update your quality control process to the NGS era. Contact us to discuss your potential project with us.

References

1. de Vries JJC, Brown JR, Fischer N, Sidorov IA, Morfopoulou S, Huang J, Munnink BBO, Sayiner A, Bulgurcu A, Rodriguez C, Gricourt G, Keyaerts E, Beller L, Bachofen C, Kubacki J, Samuel C, Florian L, Dennis S, Beer M, Hoeper D, Huber M, Kufner V, Zaheri M, Lebrand A, Papa A, van Boheemen S, Kroes ACM, Breuer J, Lopez-Labrador FX, Claas ECJ. Benchmark of thirteen bioinformatic pipelines for metagenomic virus diagnostics using datasets from clinical samples. J Clin Virol. 2021 Aug;141:104908. doi: 10.1016/j.jcv.2021.104908. Epub 2021 Jul 8. PMID: 34273858.

genomescan.nl
+31 71 5681050
geneticQC@genomescan.nl