

A Next Generation Sequencing (NGS) Approach to Influenza Vaccine Development



AUTHORS
Wayne Hogrefe, Ph.D., D (ABMLI) and Patrick Hurban, Ph.D.

VACCINE EFFICACY STUDIES

DETECTION

Influenza vaccine efficacy studies entail monitoring Influenza like Illness (ILI) in vaccine trial participants to determine the efficacy of the vaccine in preventing influenza infection. Samples collected from trial participants with ILI are tested using molecular techniques such as PCR to identify the pathogen causing the ILI. If influenza is detected using these methods, the infecting influenza strain is isolated and expanded in cell culture.

GENETIC CHARACTERIZATION

The influenza isolates can be serotyped for strain identity but can also undergo sequencing by Sanger or Next-Generation Sequencing (NGS) methods, which allow a more granular view of the hemagglutinin and neuraminidase genes, or even the entire genome.

NGS can provide insight, including characterization of mixed populations. Genome characterization illuminates viral evolution, thus providing insight into the efficacy of vaccines and therapeutics. It supports epidemiological studies and informs the development of future treatments.

Over 20,000 influenza genomes have been sequenced through the Influenza Genome Project. The Global Initiative on Sharing All Influenza Data (GISAID) promotes sharing of sequence, clinical and epidemiological data and comprises over 40,000 isolates.

INFLUENZA GENOME STRUCTURE

The influenza genome is very compact and contained in virus particles that are ~ 100 nm in diameter. The genome is just over 13 kilobases, comprising eight RNA segments ranging from about 800 to 2,500 nucleotides packaged into ribonucleoprotein complexes containing the RNA segment, a nuclear protein and the polymerase complex, which are both encoded by the viral genome. Variations in the HA and NA genes determine the subtype and strain.

COMPARING SANGER AND NEXT-GENERATION SEQUENCING

It is possible to sequence the entire viral genome of influenza via reverse transcription of all eight RNA segments. Gene-specific primers are used to amplify the viral genome, which can then be sequenced bi-directionally using Sanger technology. The genome structure introduces complications, however. For example, the NS segment is small, producing a single PCR product that is less than 1000 bases, while the 2.3 to 2.5 kilobase PD2 segment needs to be split into several smaller amplicons to enable Sanger sequencing.

Sanger sequencing is broadly available and well understood; however, this approach does not work well with mixed populations. If the isolate is mixed, it may not be possible to detect minority subtypes or strains. There's also the possibility of false negatives: if primers fail to land due to sequence drift, they will fail to amplify cDNA from that

isolate. Due to these limitations with Sanger sequencing, NGS approaches are becoming the standard.

NGS uses a single tube reaction, taking advantage of the conserved sequences on the 5' and 3' ends of each viral RNA. PCR products are rather large, but by employing the Nextera Tagmentation method, all amplified fragments can be converted to sequencing library simultaneously. During this process, a sample barcode is incorporated to allow assignment of each resulting sequence to a specific sample. Individual fragments are clonal, and thus the sequence represents an individual molecule. This in turn enables identification of multiple subtypes or strains within a single isolate with outstanding sensitivity.

BIOINFORMATICS

A custom bioinformatics pipeline assembles sequences from all viral segments and compares them against reference sequences for strain identification. The bioinformatics pipeline includes FASTQ processing and contig assembly; contigs are matched

against existing sequences from the GISAID and other sources. A scoring matrix determines the type and strain of the isolate, with top-scoring strains subjected to pairwise competitive alignment, which is then used to make the final determination. We have accurately identified 100 percent of validated strains with this method.

Although the focus is on HA and NA genes, characterization of the entire viral genome allows detailed insight into strain evolution outside of those genes.

Vaccines are developed using specific strains. If a vaccine-inoculated patient develops flu symptoms, a sample can be collected and identified using NGS. Identifying a strain identical to the vaccine for that patient is evidence of poor efficacy. Similarly, when a different strain is observed, we gain insight into breakthrough strains or potential escape mechanisms that can inform future vaccine strategies. Thus, we identify genetic changes and incorporate this knowledge into the analysis of efficacy, leading to development of new anti-viral therapies.

Table 1. Comparison of Sanger and NGS Sequencing.

NGS	Sanger Sequencing
Sequencing the entire viral genome	Sequencing the entire viral genome.
Full sequence of HA and NA genes	Genome structure requires multiple amplicons and final sequence is amalgam of individual PCR product sequences
Reduced chance of false negatives	False negatives occur if primers not land due to sequence drift
Allows evaluation of mixed infections with more than one serotype	Only works if single serotype present
Streamlined workflow: nonspecific amplification is filtered out bioinformatically	Requires multiple primer sets and detailed analysis to derive final sequence