Development of a targeted next generation sequencing panel for infectious diseases

Description of the Proposed Study:

Molecular detection of pathogens has been commonplace for diagnosis of infectious diseases. qPCR has now become the gold standard test for detection of many viral and bacterial pathogens, surpassing virus isolation and bacterial culture methods in sensitivity. However, while qPCR assays are highly sensitive and specific, their use can be costly, especially if testing for multiple pathogens. Routinely this is done with use of many single qPCR reactions or some multiplex qPCR reactions. Multiplex PCR cuts costs, but the method is limited to a small number of pathogens that can be detected at a time, potentially at the expense of the sensitivity for each pathogen (J Virol Methods 2015;217:28-35; J Med Virol 2015; doi: 10.1002/jmv.24413).

Next generation sequencing (NGS) is the newest sequencing method available. Unlike Sanger sequencing, which can only sequence a single portion of an organism at a time, NGS is massively parallel sequencing. This means that theoretically, all DNA/RNA in the sample can be sequenced in a single run (eg. not just a portion of an organism, but the whole organism). NGS was originally used to sequence the human genome, which reduced this feat down from a 13 year project costing $2.7 billion dollars (Initial sequencing and analysis of the human genome Nature 2001;409:860-921, with Sanger sequencing) to approximately $1000 within a period of days. This sequencing capability has revolutionized human medicine, now allowing comparison of human genomes to understand the source of genetic diseases and cancer development and with this, the implementation of personalized medicine.

Next generation sequencing is also useful for whole genome sequencing of bacteria and viruses, and the FDA has replaced an older technology with NGS for tracing foodborne pathogen outbreaks. Whole genome sequencing of clinical bacterial isolates is also being performed in veterinary diagnostic labs in conjunction with the USDA and FDA to trace antibiotic resistance genes in veterinary pathogens.

Generally, NGS for infectious diseases is limited to whole genome sequencing, mainly from organisms that have been cultured or isolated from a clinical sample, or deep sequencing (metagenomics) for detection of an unknown pathogen. Because NGS does not require use of primers specific for a pathogen, it can be used for discovery of new organisms. Use of NGS in both of these situations has limitations, which are due to the number of sequences required to obtain results and the related cost. In a clinical sample, the vast majority of DNA/RNA in the sample is from the host itself. For infectious disease testing, this is junk DNA/RNA that we are not interested in sequencing. So, because NGS is indiscriminate sequencing, this host material must be removed or reduced. This adds extra cost and time. Trying to detect pathogens in a host tissue is like looking for a needle in a haystack. The cost to do this is generally more than $1000 per sample vs. sequencing the genome of an organism that has been isolated from the tissue by culture or by virus isolation, which is about $200 per sample.

While whole genome sequencing of a single pathogen is overkill in most diagnostic situations, use of the technology for targeted sequencing, which is done for cancer detection in human samples (Int J Mol Sci 2015;16:28765-82), certainly has merit for infectious diseases also. Targeted NGS differs from other NGS methods because of the incorporation of primers and PCR into the assay prior to sequencing. Primers specific for bacteria, viruses, fungi, or parasites of interest that have been shown to be associated with syndromic diseases, whether it be respiratory disease, reproduction failure/abortion, or enteritis, could be incorporated into a next generation sequencing protocol. This allows for an increase in the amount of DNA/RNA from pathogens above the background host DNA/RNA in the sample. Therefore, a reduced number of sequences need to be generated with NGS to obtain actionable results (detection of known pathogens), which reduces the cost associated with testing. This test could sequence portions of a large number of pathogens in the same test directly from a clinical sample, rather than fully sequencing the whole genome of a single pathogen.

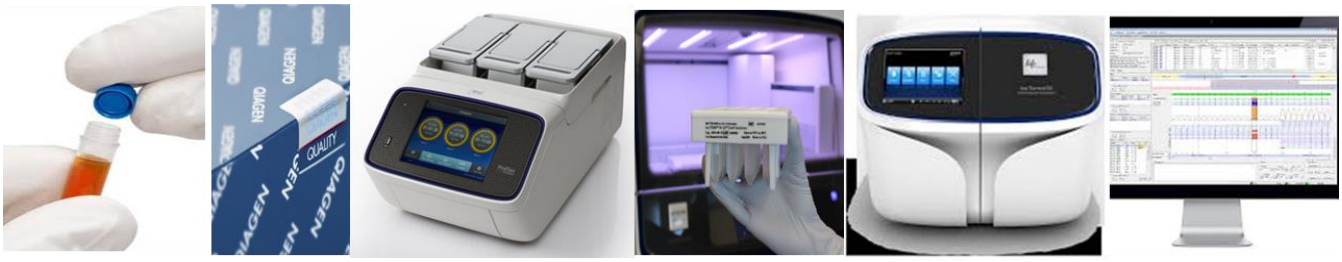
We propose use of a next generation sequencing platform for multiplex pathogen sequencing for syndromic testing. This type of comprehensive panel will remove the guesswork that is usually needed in these cases to determine appropriate test(s) to use for diagnosis. The comprehensive nature of this panel also lends itself as a useful tool for surveillance of infectious diseases in our animal populations. These primers could all be incorporated into the same assay because of the virtually unlimited amount of sequencing that can be done in parallel in a single next generation sequencing run. The assay could also be designed to incorporate primers for sequencing of specific areas that are useful for strain typing or detection of specific bacterial toxins. Additionally, because of the parallel nature of the sequencing, all strains of a particular organism in a sample, whether it be modified live virus from a vaccine in combination with a wild-type strain or multiple wild-type strains in a sample, could be detected with this test. The turnaround time for the targeted NGS method is as few as two days, providing both detection and characterization of the organisms. Limiting the sequencing to pathogens of interest by introduction of primers into the assay cuts the costs associated with this technology by allowing at least 10X more samples to be tested in a single run. As an example, 50 different pathogens can be targeted in a single run with 20 different samples or less/more (this is scalable) tested at the same time.

Specific Aims:

1. Develop primer pools to detect and characterize pathogens associated with syndromic diseases to be used in a targeted next-generation sequencing testing platform.
2. Validate the primers/method with characterized/reference strains of bacteria, viruses, and parasites. (analytical specificity)
3. Evaluate the limit of detection with plasmid DNA or in vitro transcribed RNA commercially purchased, if available, or prepared in our laboratory representative of a subset of the most common pathogens. (analytical sensitivity)
4. Diagnostic evaluation will be performed with clinical samples representing different syndromic diseases by comparing results from the targeted NGS method with routine methods used in the laboratory (eg. qPCR, culture, virus isolation, necropsy and histopathology).

Experimental Protocol:

Workflow:



Steps in the workflow include:
Day 1- (1) Receipt of appropriate clinical sample and extraction of all nucleic acid from the sample (DNA and RNA). (2) Multiplex PCR using primers specific for pathogens, (3) library preparation to add sample barcodes to distinguish samples from one another and to add adaptors needed for the sequencing protocol, and (4) loading of the samples on the Ion Chef for automated chip loading, which occurs overnight.
Day 2- (1) Chip moved from Ion Chef to the Ion S5 machine for sequencing, which takes approximately 4 hours per chip and finally, (2) sequence analysis. Quality control is performed to remove poor sequences and to remove primer and adaptor sequences. The sequences are then assembled and compared to a file that contains sequences of the pathogens that are potentially in the sample (based on the primers used). The sequences are also compared to sequences for pathogens in a national database to confirm the results.

Budget:

(Including justification for all expenses)
Primers $2000
Sequencing chips $2040
Library prep $14,300
Nucleic Acid extraction kits $1000
Chef kit/Sequencing kit- $5300
Bar codes for multiplexing samples- $400
Total $25,040

The requested costs are associated with next generation sequencing (primers, chips, library preparation, extraction kits, library equalizer kit, bar codes, sequencing kit). We have factored in additional expenses for purchase of multiple extraction kits to determine the best kit to use with clinical samples. The bar codes are necessary for multiplexing samples, which allow for identification of each individual sample during sequence data analysis. Funding for the additional testing will be paid by our clients as part of the normal sample testing done at the diagnostic lab. These costs cover the development of a single targeted NGS panel (eg. either for poultry or porcine).