Next Generation Sequencing Methods for SARS-CoV-2 Viral RNA: A Performance Assessment

Brittany Peterson[1], Nicholas Gustafson[1], Brittany Knight[1], Jennifer Stone[1]

[1]MRIGlobal – Infectious Diseases Surveillance and Diagnostics, Kansas City, MO

Introduction:

Next Generation Sequencing (NGS) has proven itself to be a critical tool during the COVID-19 pandemic, allowing researchers to internationally share sequences of the SARS-CoV-2 virus at a rapid pace. This technology prompted the characterization and epidemiological surveillance of multiple variants of concern – such as B.1.1.7 (UK), B.1.617.2 (Delta) and BA.1 (Omicron). Yet not all NGS techniques have equivalent suitability, data quality, or practicality for the sequencing of SARS-CoV-2. The three main approaches discussed for the NGS of SARS-CoV-2 viral RNA are Whole Genome Sequencing (WGS also known as Shotgun Metagenomics), Hybrid Capture Enrichment, and Amplicon Sequencing (ARTIC). These methods were assessed from a practical standpoint; such as samples per run, turnaround time for sample to data, cost, best suited application, scalability, throughput, and if they are targeted or non targeted protocols. Each method has advantages and weaknesses associated, thus the NGS method used is dependent on the characteristics of the samples and data needed.

Library Preparation Overview:

Three library preparation methods were evaluated for detection and lineage determinations:

WGS/Shotgun Metagenomics – Sequences the whole genome of SARS-CoV-2 as well as any other DNA found in the sample. Has longer hands-on time in addition to overall turnaround time, but is capable of detecting previously unknown variants. WGS requires a higher sample target input to achieve breadth and depth of coverage needed, thus insufficient sample input will lead to poor detection and poor coverage.

Hybrid Capture Enrichment – Captures genomic regions of interest via hybridization to target-specific probes. Allows for sensitive detection without the read depth needed for WGS, and has more comprehensive profiling of target regions than amplicon sequencing. Is also robust enough to detect other Coronaviruses/respiratory viruses. However has the longest turnaround/hands-on time and is the most expensive to conduct.

Amplicon Sequencing – (ARTIC protocol) Uses 98 primer pairs to detect and amplify the 30kb SARS-CoV-2 genome in ~400bp amplicons . Rapidly scalable with the fastest turnaround time of all three methods. Can easily detect presence and more specific mutations, but is not cost-effective if few samples are sequenced. A more targeted approach can be used with amplicon sequencing where only a specific gene region of SARS-CoV-2 is amplified with custom designed primers.

Metric	WGS/Shotgun Metagenomics	Hybrid Capture Enrichment	Amplicon Sequencing (ARTIC Protocol)
Samples Per Run (Illumina MiSeq)	10	25	30-48
Turnaround Time	16 Hours	22 Hours (46 hours if 24 hour incubation used)	12 Hours
Cost	$$	$$$	$
Scalability/Throughput	Low	Medium	High
Targeted/Non-Targeted	Non-Targeted	Targeted: Respiratory Pathogens	Targeted: SARS-COV-2 Only
Best Suited Application	Detect Novel Virus	Detection and Surveillance of Respiratory Pathogens	Surveillance of SARS-COV-2

Figure 1: Comparison table of three library preparation methods for the NGS of SARS-CoV-2.

Workflow Diagram:

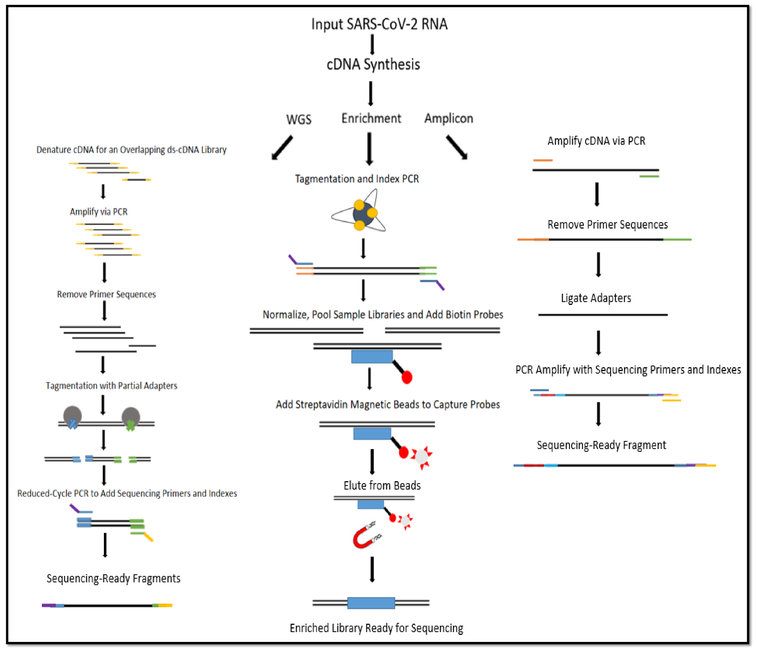

Figure 2: Workflow diagrams of all three library preparation methods for SARS-CoV-2 NGS. RNA is denatured and reverse-transcribed into complementary DNA (cDNA synthesis) for all methods. Diagrams made with reference to Illumina's workflow graphics. (www.illumina.com)

WGS vs Enrichment SARS-CoV-2 Detection:

Here we present differences based on quantity of SARS-CoV-2 reads generated for each library preparation method by MRIGlobal:

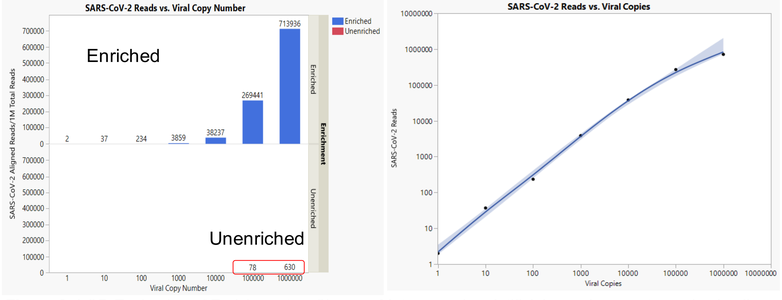

Figure 3: LOD Evaluation of Enrichment vs Shotgun Metagenomics. Artificial samples constructed using live WT SARS-CoV-2 virus spiked into VTM, then used to establish LOD for each sequencing method. Top histogram on the right shows enriched reads, bottom histogram reflects unenriched. Enrichment shown to improve SARS-CoV-2 detection 1000-fold even with low viral titer. Right graph reflects the known linear relationship between viral copies in sample and reads produced.

WGS vs Enrichment SARS-CoV-2 Variant Calling:

Here we present differences in the variant-calling capability for each library preparation method by MRIGlobal:

	Sample	PCR Ct	Lineage	SARS-CoV-2 Fraction of Targets detected	Coverage >= 30x
WGS	A2	32.11	BA.1	92.86%	9.03%
	A6	**30.6**	**Unassigned**	95.92%	**1.33%**
	A8	**30.06**	**Unassigned**	7.14%	**N/A**
	A10	**30.09**	**Unassigned**	72.45%	**N/A**
Enrichment	C2	32.22	P.1	100.00%	18.48%
	C4	29.41	B.1.1.7	100.00%	99.69%
	C15	29.31	B.1.1.7	100.00%	99.74%
	C18	**34.04**	**Unassigned**	98.98%	**1.64%**
	C3	**34.89**	**Unassigned**	93.88%	**0.08%**

Figure 4: Breadth and Depth of SARS-CoV-2 coverage comparison between WGS and Enrichment library preparation methods when low viral titer (Ct > 29) clinical samples used as input. Despite comparable Ct values and low depth of coverage, enrichment allows for lineage calls (Sample C2), whereas WGS can only reliably detect presence of SARS-CoV-2. Enrichment begins to be unable to identify lineage at almost a 10-fold lower concentration than WGS, demonstrating superior sensitivity.

Discussion:

Amplicon Sequencing is ideal for the surveillance of SARS-CoV-2 because of the high-throughput nature, speed of turnaround time, low input material required, and low cost per sample. However it cannot detect coinfections or novel pathogens and becomes drastically less cost effective if a small amount of samples are sequenced. Omicron variant was found to require additional primers to achieve similar coverage as previous variants.

WGS can detect coinfections and novel pathogens in addition to presence of SARS-CoV-2, however a larger concentration of target input or fewer samples per run are required in order to achieve sufficient genome coverage for more specific lineage calls. If low viral titers are used as input relative to host background, the reads produced will not be sufficient to characterize the virus. This is shown to be true for the artificial samples and clinical samples tested. (Figures 3 and 4). A cutoff Ct of 30 is used for samples when performing WGS to ensure adequate input concentration.

Hybrid Capture Enrichment is not only more robust than amplicon sequencing in identifying various respiratory viruses in addition to SARS-CoV-2, but it is also more sensitive than WGS for detection and characterization. It generates more reads and can adequately determine lineage/variants even with poor depth of coverage and low viral titer. Therefore more samples can also be sequenced per run than WGS as it requires less viral input, with LODs comparable to PCR detection. Hybrid Capture Enrichment is overall a superior method for characterization of SARS-CoV-2 including lineage detection and variant calling, but it is cost prohibitive and time intensive.

Contact Information

Brittany Peterson
T: 816-326-5953
E: bpeterson@mriglobal.org

MRIGlobal
425 Dr. Martin Luther King Jr Blvd
Kansas City, MO, 64110