ARP-22: Development of SARS-CoV-2 NGS Internal Standards to Ensure Sufficient Complexity Capture for Variant Detection and Diagnosis

TAP HERE TO RETURN TO KIOSK MENU

AccuGenomics
A higher standard of accuracy

Tom Morrison, Brad Austermiller & Colette Saccomanno, AccuGenomics

NIIMBL® The National Institute for Innovation in Manufacturing Biopharmaceuticals

Abstract

Improve standardization of epigenetic testing using Internal Standards

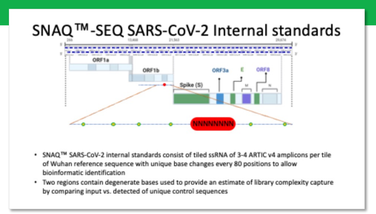
- Standardized viral abundance
- Complexity capture quality control
- Lyophilized for shelf and shipping stability

Introduction

NGS great potential as an epigenetic surveillance tool

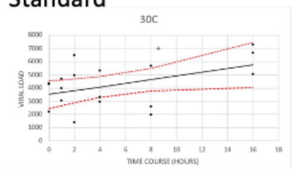
Lack of NGS standardization limits accuracy, especially as testing moves away from patient material to wastewater testing

Use of internal standards, which biochemically mimic NGS targets, will support the existing need for per sample quality controls to ensure standardization of NGS surveillance modeling.



Methods & Results

1. Shelf Stable ssRNA SARS-CoV-2 Internal Standard



- Shelf stable ssRNA SARS-CoV-2 internal standard for ARTIC NGS testing using elevated storage conditions
- Initial study met study milestone
- Further testing underway

❖ Next steps: test lyophilized IS performance in multi laboratory study

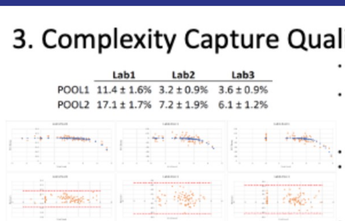
2. Standardized Viral Abundance

- Poor test results can arise from lower-than-expected genomic input or errors in NGS testing procedure
- SNAQ™ abundance directly measures genomic input in the NGS results by indicating sequence failure association with low sample input
- Results demonstrate SNAQ™ abundance measurements varied less than three-fold for samples ranging from >10^4 to 10 copy viral genomes sequenced in three different laboratories using different sequencing instruments and altered ARTIC NGS protocols

❖ SNAQ™ standardized viral abundance measurements

3. Complexity Capture Quality Control

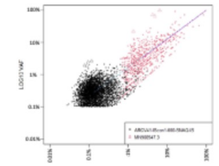
	Lab1	Lab2	Lab3
POOL1	11.4 ± 1.6%	3.2 ± 0.9%	3.6 ± 0.9%
POOL2	17.1 ± 1.7%	7.2 ± 1.9%	6.1 ± 1.2%



- How to more accurately detect NGS library preparation testing errors?
- SNAQ™ complexity capture (CC) acts as a full process control to measure how well each sample's viral genome was captured as sequencing reads
- Table indicates each lab's nominal CC rate
- A CC model was created from each lab's CC vs viral Load response; the residuals to this model indicate if a sample CC is nominal
- SNAQ™ could detect a >two-fold drop in complexity capture with high specificity.

❖ SNAQ™ CC indicated each sample met nominal testing performance

4. SNAQ™ Limit of Blank Calculation



- How to push COVID-19 NGS testing below 75% allele frequencies or call low viral load samples?
- SNAQ™ IS have a known sequence, therefore any detected variants arose from NGS testing errors
- A per sample limit of blank was created from the IS errors, which was then applied to each variant
- VAF vs LOB plot depicts the IS (black dots) used to set the 99.9% significance cutoff for the sample variants significantly above (red triangle) or below significantly (red dots) the limit of blank

❖ A per sample limit of blank was created from the IS sequencing errors

5. COVID-19 Variant Calling Quality Control

Table 1. Sensitivity and PPV of Lab Variant Callers with out or with SNAQ™ LOB

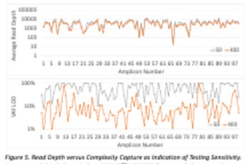
FILTER	ORIGINAL			SNAQ		
LAB ID	1	2	3	1	2	3
RECALL	630	630	630	630	630	630
CALLED	1611	186	475	783	141	368
TRUE	510	180	411	467	137	357
FALSE	1101	6	64	316	4	11
SENSITIVITY	81%	29%	65%	74%	22%	57%
PPV	32%	97%	87%	60%	97%	97%

FILTER indicates either lab's *ORIGINAL* variants or what's left after *SNAQ*™ QC.
LAB ID indicates source of VCF.
RECALL indicates possible total number of "true positive" variants.
CALLED indicates total variants called.
TRUE indicates how many variants were "true positive."
FALSE indicates how many variants were not in "true Positive" pool.
SENSITIVITY indicates what fraction of "true positive" variants were detected.
PPV indicates positive predictive value ("true_positives"_count / Total_Variant_count)

- The positive predictive value of each lab's *ORIGINAL* variant calls could be improved (*SNAQ*) with a modest loss of sensitivity, more so for the Lab 1, the only lab whose test had sufficient sensitivity for low VAF detection.
- SNAQ™ LOB QC expected to improve
  - SNAQ™ LOB does not capture the genomic damage introduced when using RNA extracted from FFPE inactivated viral particles.
  - IS variants should be called using lab's pipeline

❖ SNAQ™ LOB PPV improvement hints benefit in NGS testing accuracy, but further optimization is required.

6. SNAQ™ VAF LOD Estimation



- Average read depth by amplicon for COVID-19 Alpha NIIMBL study samples represents current practice for sequencing sensitivity
- Results indicate similar read depth for the two samples

❖ Expectation would be very similar variant calling sensitivity, this expectation would be incorrect

- SNAQ™ coverage estimates how many unique templates were captured in each amplicon
- Amplicon coverage allows estimation of VAF LOD on a per sample per amplicon basis
- Plot depicts 95% confidence of detecting variant at or above the indicated VAF
- LOD model predicted >95% of variant detected in all low viral load samples

❖ SNAQ™ coverage provides a per sample per region variant sensitivity measurement

Results

Lyophilization of ssRNA COVID-19 internal standard reference material for ARTIC NGS testing.

Creation of novel Quality Controls:

- Standardized viral abundance
- Complexity Capture QC
- Improved specificity using SNAQ™ Limit of Blank
- Per amplicon variant limit of detection

Public Impact

NGS potential as an epigenetic surveillance tool is limited by lack of standardization.

SNAQ™ internal standards provide a needed per sample quality controls to ensure standardization of NGS surveillance modeling

Acknowledgements

Lilly Katumba, Kelly Rogers, NIIMBL; Megan Cleveland & Christopher Robert, NIST for their support and guidance.

Abstract

The COVID-19 pandemic led to the rapid buildup of sequencing laboratories to support epigenetic analysis of positive COVID-19 samples. While reference materials were developed to demonstrate a minimal level of proficiency testing, there remains testing performance uncertainty when test results yield low sample coverage. Did poor test performance arise from poor sample, low genomic input, variants dropouts, or a test failure? NIIMBL funding was used to support the development of a stabilized ssRNA Internal Standard (IS) and the IS use in multi-laboratory performance testing. The performance testing was based on AccuGenomics' Standardized Nucleic Acid Quantification (SNAQ™) technology, which creates internal standards that biochemically mimic target sequences and provides quality controls during sequencing. When the ssRNA IS are added to each sample, the resulting SNAQ™ analysis provides a standardized viral load measurement (viral abundance), a per sample measurement of how well the test captured sample genome (complexity capture), assesses dropouts (coverage analysis), and per sample – variant limit of blank (LOB) to improve variant detection accuracy. Compatible with the ARTIC primer pairs, the lyophilized IS were sent to three laboratories along with 73 blinded SARS-CoV-2 samples consisting of 5 strains and different viral loads. The resulting FASTQs were processed through a publicly available Docker container which creates SNAQ™ results and FASTQ stripped of human and IS sequences. The viral sequence only FASTQ were returned to the laboratory for their respective COVID-19 surveillance analysis.

The SNAQ-based viral loads measured from the FASTQ files in each lab were within three-fold of each other. This result supports the use of IS to provide a multi-laboratory standardized abundance measurement. The complexity capture profile for each laboratory was created and indicated all samples were consistently tested within the lab, but the detection sensitivity differed at the amplicon and whole genome level by over 10-fold. On a per sample basis, the IS indicated if the dropouts were caused by unoptimized test protocol, low genomic input or sample variants interfering with multiplex primer binding. The laboratories' variant detection correlation with SNAQ™ analysis will be presented.

Overall, the NIIMBL grant funded the development of a stable COVID-19 ARTIC NGS testing Internal Standard used to provide missing quality assurance metrics for NGS-based pandemic surveillance monitoring, as well as standardized metrics for comparing inter-laboratory performance.

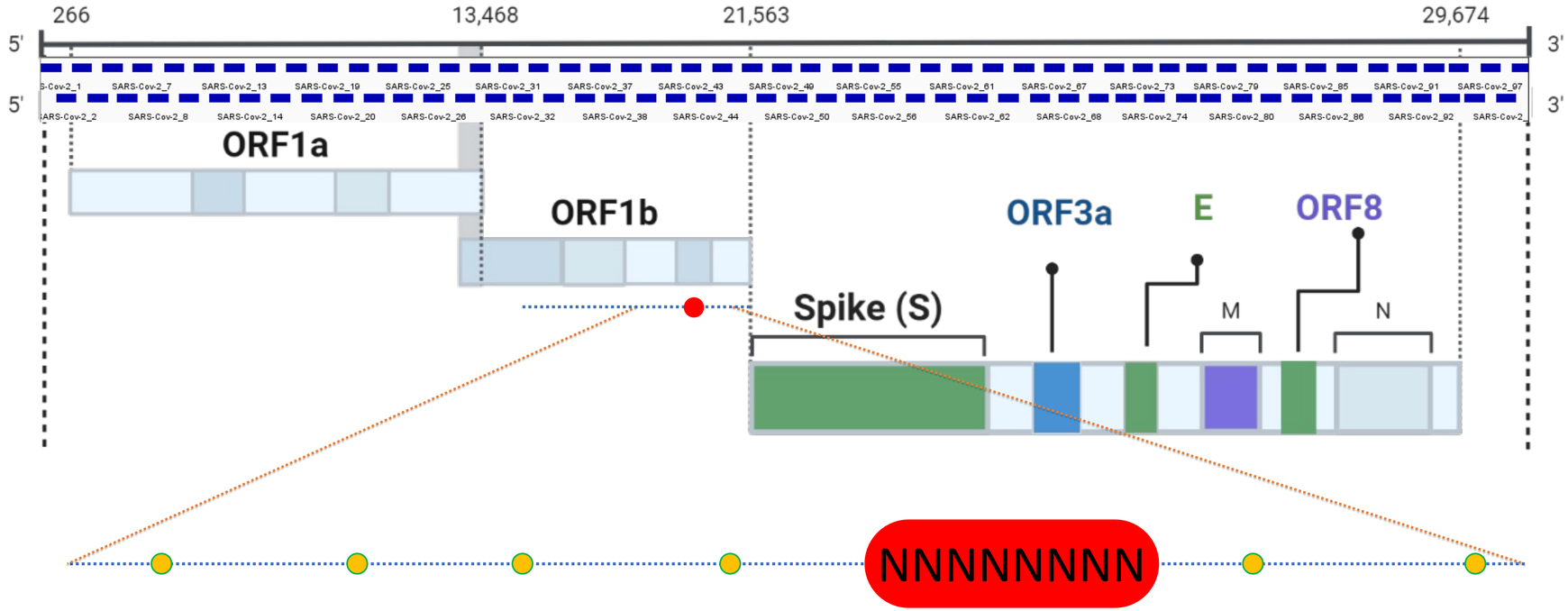
AccuGenomics

Introduction

The COVID-19 pandemic led to an unprecedented expansion of laboratories using high throughput next generation sequencing (NGS) for epidemiologic surveillance of emerging COVID-19 strains. The resulting surveillance network was sufficient for high titer patient samples, but as surveillance moved into the very low titers of wastewater samples, a need emerged to ensure more robust testing performance. NIIMBL funded the creation of a shelf stable ssRNA SARS-CoV-2 internal standard for ARTIC NGS testing which, when added to each sample, indicates if each sample achieves sufficient analytic sensitivity and specificity.

AccuGenomics

SNAQ™-SEQ SARS-CoV-2 Internal standards



- SNAQ™ SARS-CoV-2 internal standards consist of tiled ssRNA of 3-4 ARTIC v4 amplicons per tile of Wuhan reference sequence with unique base changes every 80 positions to allow bioinformatic identification
- Two regions contain degenerate bases used to provide an estimate of library complexity capture by comparing input vs. detected of unique control sequences

1. Shelf Stable ssRNA SARS-CoV-2 Internal Standard

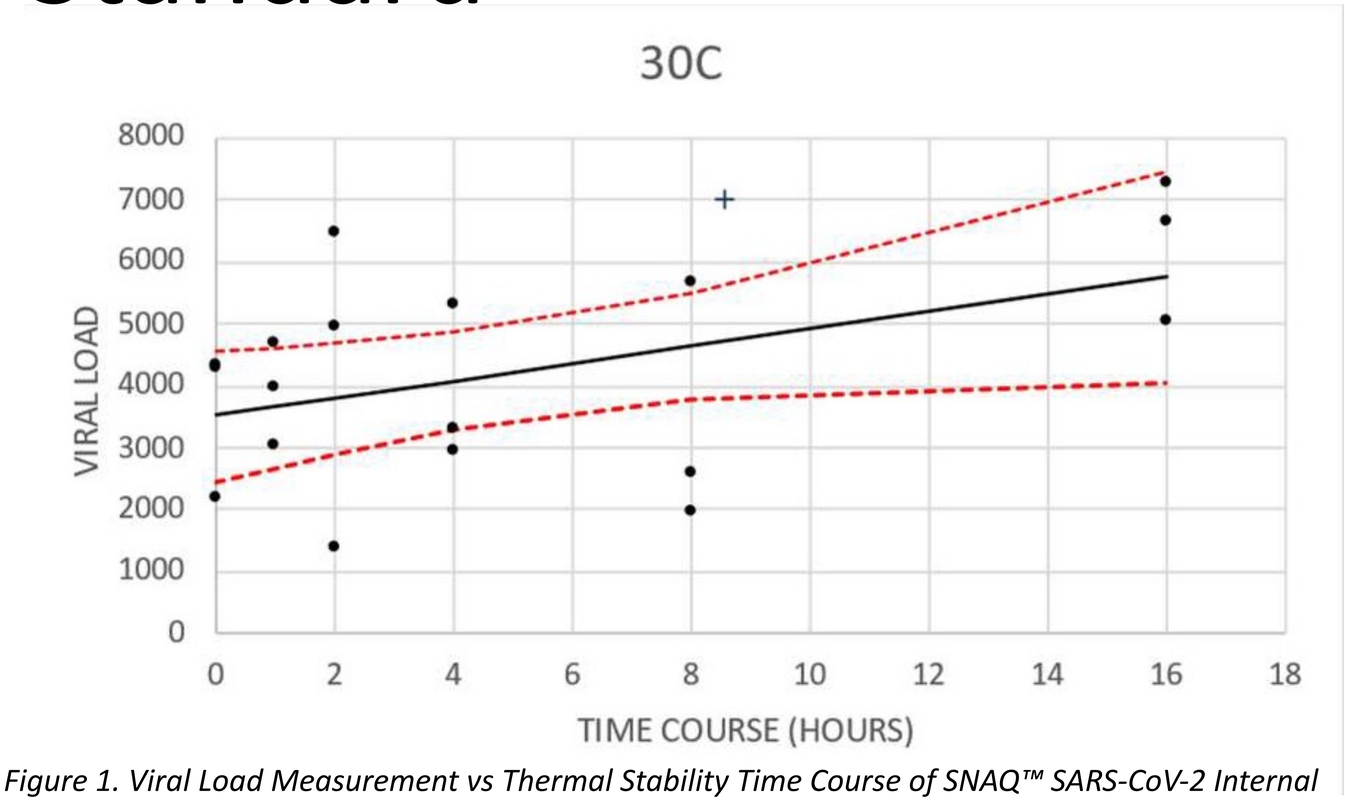


*Figure 1. Viral Load Measurement vs Thermal Stability Time Course of SNAQ™ SARS-CoV-2 Internal Standard. An accelerated aging study places the reference material in higher than ideal storage (-80C) conditions to predict how it will behave under normal storage conditions over a longer period. As a full process test, the viral load measurement (y-axis) of three replicates at 30C storage were collected at the indicated times (x-axis). The premise was IS degradation would lead to a rise in viral load measurement due to the viral load calculation (viral load = NT_reads / IS_reads * IS_input_copies). The observed low replicate repeatability resulted from sub-optimal pellet resuspension and was absent in future experiments. The solid line is the linear regression flanked by the 95% confidence intervals (dashed lines) with the "+" symbol indicating the threshold for 2-fold degradation at 2 years storage.*

- Shelf stable ssRNA SARS-CoV-2 internal standard for ARTIC NGS testing using elevated storage conditions
- Initial study met study milestone
- Further testing underway

❖ Next steps: test lyophilized IS performance in multi laboratory study

AccuGenomics

2. Standardized Viral Abundance

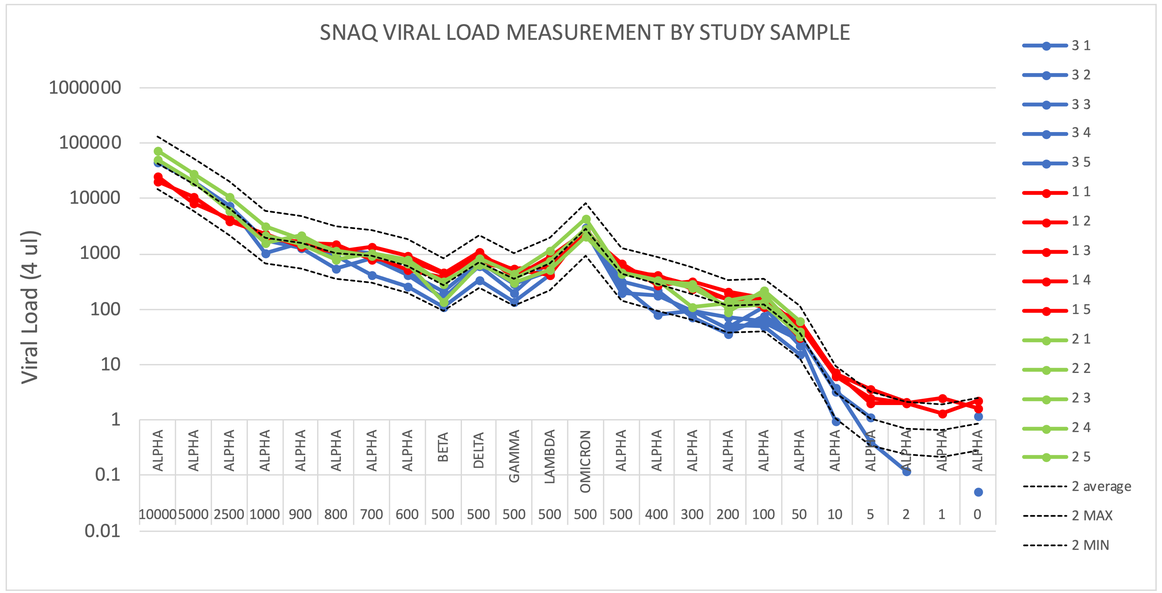


Figure 2. SNAQ™ Based Viral Load Measurement of Study Samples. Study FASTQ were analyzed using SNAQ™-VSOFT v1.2beta to estimate viral load (y-axis) of each study sample (x-axis). X-axis indicates COVID-19 genome and approximate genomic input; legend indicates lab and replicate number. The mean (central dashed line) and three-fold difference from mean (outer dashed lines) indicate that 163 of 165 samples with greater than 100 genomic abundance were within 3-fold of mean.

- Poor test results can arise from lower-than-expected genomic input or errors in NGS testing procedure
- SNAQ™ abundance directly measures genomic input in the NGS results by indicating sequence failure association with low sample input
- Results demonstrate SNAQ™ abundance measurements varied less than three-fold for samples ranging from >10^4 to 10 copy viral genomes sequenced in three different laboratories using different sequencing instruments and altered ARTIC NGS protocols

❖ SNAQ™ standardized viral abundance measurements

AccuGenomics

3. Complexity Capture Quality Control

	Lab1	Lab2	Lab3
POOL1	11.4 ± 1.6%	3.2 ± 0.9%	3.6 ± 0.9%
POOL2	17.1 ± 1.7%	7.2 ± 1.9%	6.1 ± 1.2%

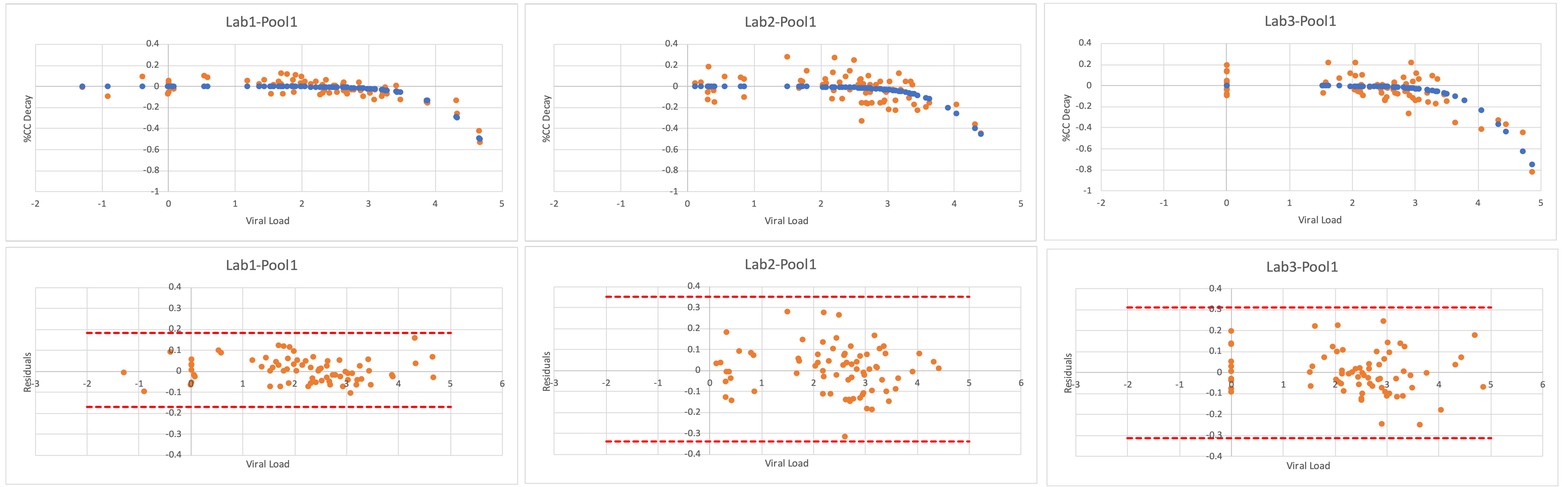


Figure 3. SNAQ™ Complexity Capture QC Profile for each Study Laboratory. The loss of unique Complexity Control Count (output / input) may be modeled as a function of viral load. The top table indicates the nominal complexity capture for the indicated labs (column heading) and indicated ARTIC multiplex PCR pool (rows). Flow cell sample load normalization leads to competition between NT and IS reads. With increasing viral load (x-axis), the Unique CC count drops. The %CC response was normalized using the lab's average %CC capture in low viral load samples (Pool1 %CC table in previous slide). The log10 CC change with log10 viral load (y-axis, top row plots) was modeled as Frac_CC = (IS x A) / (viral_load + IS x A), where A is lab & pool specific fudge factor influenced mostly by the read duplication rate. The residuals of the model (y-axis, lower plot rows) were found to be normally distributed by the Jarque-Bera test, from which 99.9% confidence intervals (red dashes) were calculated using Excel norm.inv function. For brevity, Pool 2 data not shown. CC QC will indicate a significant per sample drop in test sensitivity if %CC deviates more than 1.5, 2, 2-fold from model for Lab1,2,3, respectively. Of note, SNAQ™ was designed to QC <10,000 genomic input samples because these samples will be more challenging to obtain good sequence.

- How to more accurately detect NGS library preparation testing errors?
- SNAQ™ complexity capture (CC) acts as a full process control to measure how well each sample's viral genome was captured as sequencing reads
- Table indicates each lab's nominal CC rate
- A CC model was created from each lab's CC vs viral Load response; the residuals to this model indicate if a sample CC is nominal
- SNAQ™ could detect a >two-fold drop in complexity capture with high specificity.

❖ SNAQ™ CC indicated each sample met nominal testing performance

AccuGenomics

4. SNAQ™ Limit of Blank Calculation

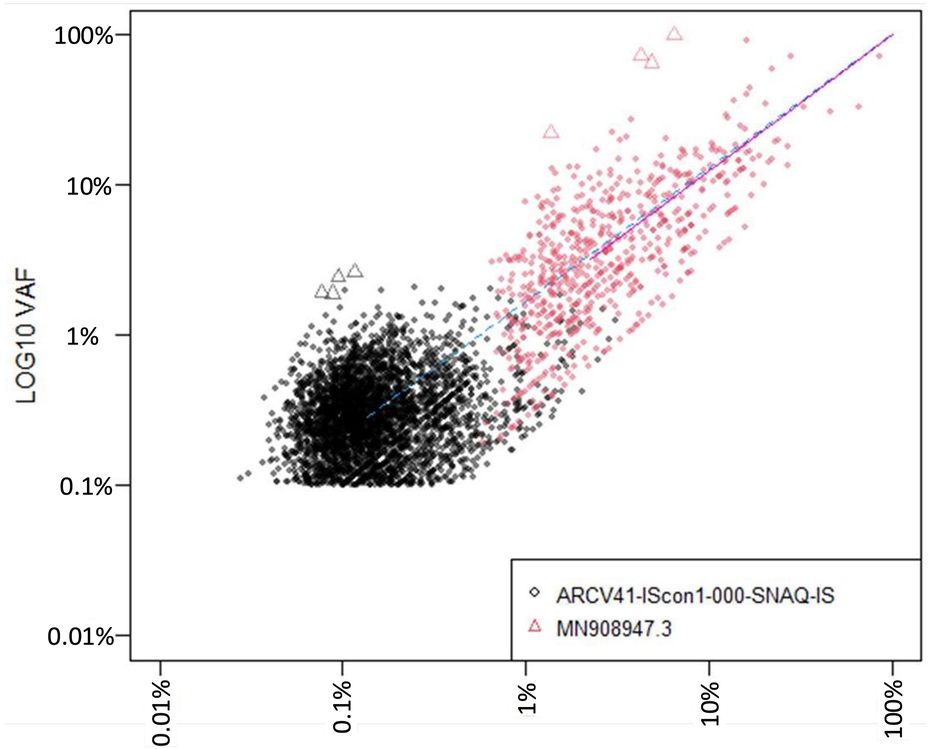


Figure 4. Visualization of SNAQ™ Variant Calling Quality Control. SNAQ™ uses the Poisson Exact Test to calculate a positional significance between a variant and the background sequencing error. The resulting p-value is converted into a limit of blank (LOB) and plotted against variant allele frequency (VAF) for each IS (black) and NT (red) variant. Each of the lines run through the cluster center, described by two points: the center of IS or NT mass and a point at 100%,100% (blue green for IS and dark blue for NT). Distances away from the lines provide a surrogate for significance. The IS variants with distance >99.9% (black triangles) were considered significant and used to set a distance cutoff for NT variants (large red).

- How to push COVID-19 NGS testing below 75% allele frequencies or call low viral load samples?
- SNAQ™ IS have a known sequence, therefore any detected variants arose from NGS testing errors
- A per sample limit of blank was created from the IS errors, which was then applied to each variant
- VAF vs LOB plot depicts the IS (black dots) used to set the 99.9% significance cutoff for the sample variants significantly above (red triangle) or below significantly (red dots) the limit of blank

❖ A per sample limit of blank was created from the IS sequencing errors

AccuGenomics

5. COVID-19 Variant Calling Quality Control

Table 1. Sensitivity and PPV of Lab Variant Callers with out or with SNAQ™ LOB

FILTER	ORIGINAL			SNAQ		
LAB ID	1	2	3	1	2	3
RECALL	630	630	630	630	630	630
CALLED	1611	186	475	783	141	368
TRUE	510	180	411	467	137	357
FALSE	1101	6	64	316	4	11
SENSITIVITY	81%	29%	65%	74%	22%	57%
PPV	32%	97%	87%	60%	97%	97%

FILTER indicates either lab's *ORIGINAL* variants or what's left after *SNAQ*™ QC.
LAB ID indicates source of VCF.
RECALL indicates possible total number of "true positive" variants.
CALLED indicates total variants called.
TRUE indicates how many variants were "true positive."
FALSE indicates how many variants were not in "true Positive" pool.
SENSITIVITY indicates what fraction of "true positive" variants were detected.
PPV indicates positive predictive value ("true_positives"_count / Total_Variant_count)

- The positive predictive value of each lab's *ORIGINAL* variant calls could be improved (*SNAQ*) with a modest loss of sensitivity, more so for the Lab 1, the only lab whose test had sufficient sensitivity for low VAF detection.
- SNAQ™ LOB QC expected to improve
  - SNAQ™ LOB does not capture the genomic damage introduced when using RNA extracted from FFPE inactivated viral particles.
  - IS variants should be called using lab's pipeline

❖ SNAQ™ LOB PPV improvement hints benefit in NGS testing accuracy, but further optimization is required.

AccuGenomics

6. SNAQ™ VAF LOD Estimation

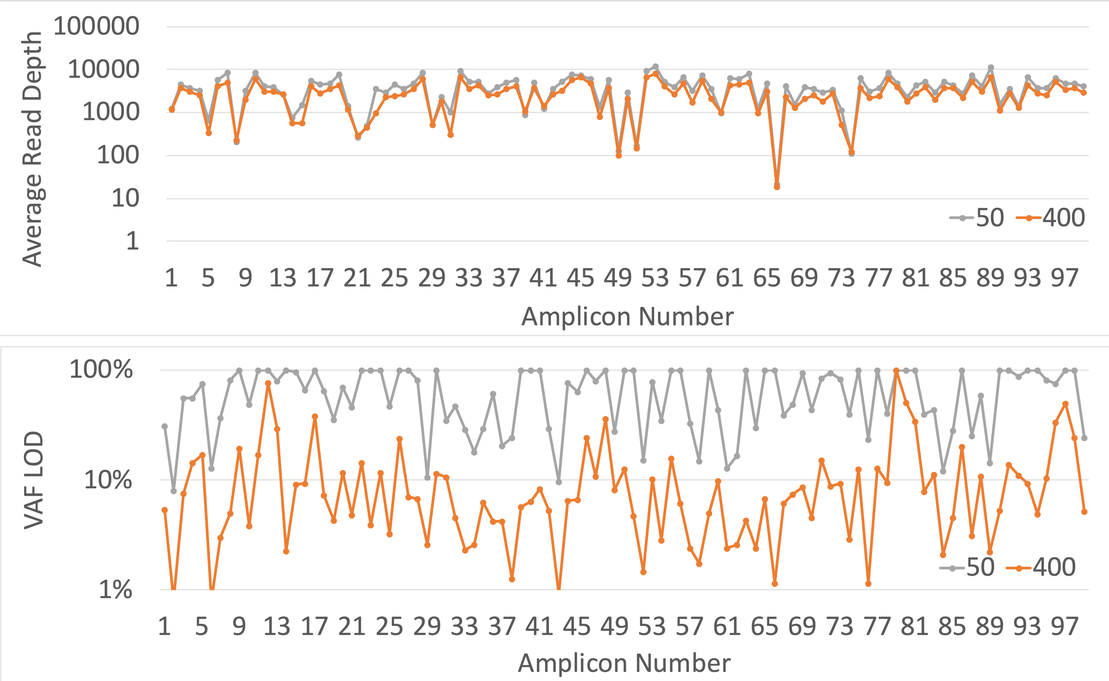


Figure 5. Read Depth versus Complexity Capture as Indication of Testing Sensitivity. Average read depth (top plot) and SNAQ™ VAF LOD (bottom plot) per amplicon were calculated using snaq-split v1.1 software for two samples at indicated 50 and 400 Alpha COVID-19 genomic input (legend) by lab 1. Artic COVID-19 NGS testing uses 98 tiled amplicons to which each read pair was mapped, base count extracted, and average read depth calculated. Read depth for read pairs mapping to two adjacent amplicons were split between adjacent amplicons in proportion to each amplicons mapped reads. VAF was calculated by first using the read depth ratio for the matching NT and IS amplicons to calculate NT yield, complexity capture was used to estimate how many of the NT reads were unique, and then a Poisson calculation was used to indicate the 95% confidence of detecting a given VAF based on the number of unique captured NT amplicon templates.

- Average read depth by amplicon for COVID-19 Alpha NIIMBL study samples represents current practice for sequencing sensitivity
- Results indicate similar read depth for the two samples
  - ❖ Expectation would be very similar variant calling sensitivity, this expectation would be incorrect

- SNAQ™ coverage estimates how many unique templates were captured in each amplicon
- Amplicon coverage allows estimation of VAF LOD on a per sample per amplicon basis
- Plot depicts 95% confidence of detecting variant at or above the indicated VAF
- LOD model predicted >95% of variant detected in all low viral load samples
  - ❖ SNAQ™ coverage provides a per sample per region variant sensitivity measurement

AccuGenomics

Results Summary

The two major goals of this project were to create a shelf stable SNAQ SARS-CoV-2 Internal Standard material and demonstrate Standardized Nucleic Acid Quantification (SNAQ™) analysis benefits in an interlaboratory study.

- Global deployment depends on the reference material remaining stable during shipment and storage.
- NIIMBL funds supported the creation of a lyophilized ssRNA SNAQ SARS-CoV-2 product with improved shelf-life stability.

A three laboratory COVID-19 NGS study tested SNAQ™ analysis to improving viral load estimation, create a sensitivity QC, and a specificity QC.

- SNAQ™ viral abundance and complexity capture were used to indicate if a poor test results arose from lower-than-expected genomic input or a poor test performance.
- Variant calling QC demonstrated how SNAQ™ generated limit of blank improve the positive predictive value with little loss of sensitivity.

AccuGenomics

Looking Forward

- The impact on society of the COVID-19 pandemic may never be fully comprehended. The logistics of per patient surveillance sample testing required a large shift in public funding and activities.
- It is now recognized that wastewater testing is 10-to-100-fold more efficient for surveillance monitoring, which translates to a more rapid build out response with less public funding.
- Public confidence in this new approach can be bolstered by incorporating SNAQ™-SEQ technology into the new protocols.
- The work performed under this grant showed that per sample quality control will support future wastewater testing and meets the required per sample testing accuracy.

AccuGenomics