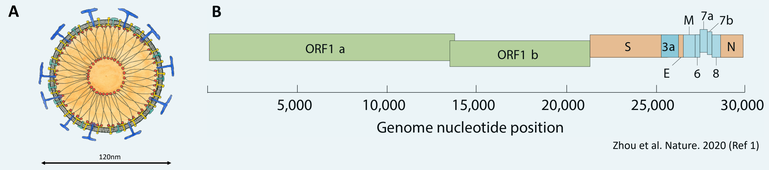
Leveraging Clinical Metagenomic Testing Against SARS-CoV-2

BENJAMIN BRIGGS MD, PHD[1,2], KENNETH RAND MD[3], NICOLE LOVINE MD, PHD[4], KARTIK CHERABUDDI, MD[4], ROBERT SCHLABERG MD,PHD[1,5]

1: IDbyDNA, Salt Lake City, UT, 2: University of California San Francisco, Division of Infectious Disease and Global Health, San Francisco, CA, 3: University of Florida Department of Pathology, Gainesville, FL
4: University of Florida, Division of Infectious Disease and Global Medicine, 5: University of Utah, Department of Pathology, Salt Lake City, UT

Introduction:

Molecular testing for SARS-CoV-2 is a major public health priority worldwide. Current PCR-based tests exhibit false-negative rates and have limitations in epidemiological investigations without additional genome sequencing. Clinical metagenomics has the potential for greater sensitivity, with the significant advantage of providing detailed sequence data and unbiased screening for coinfections, but its clinical utility for the detection of SARS-CoV-2 has not been evaluated.



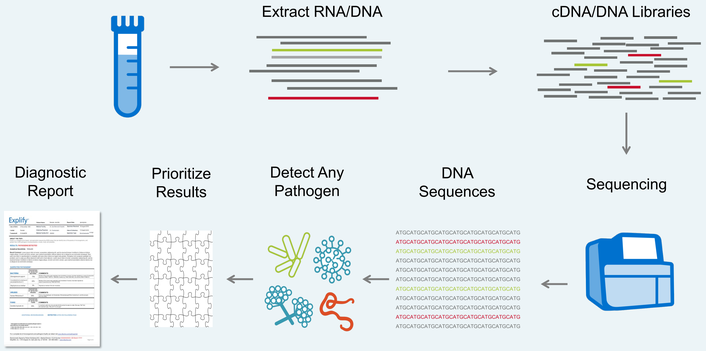
The SARS-CoV-2 viral particle has prominent Spike proteins (S) around its 120nm viral particle (A). The genome of the virus is just under 30kbp and is carried in a positive orientation (B). The RNA dependent RNA polymerase (RdRp) dominates the genome in the sub-genomic RNAs: ORF1a and ORF1b. However other sub-genomic RNAs are also encoded and are dynamic in natural infection (Ref 2). Clinical metagenomic evaluation is the most accurate method for evaluating sub-genomic RNAs since their abundance is directly measured from patient samples.

Clinical metagenomics is also the most effective way to rapidly evaluate molecular epidemiological data as viral genomes are assembled directly from patient samples without acquisition of mutations from passages in cell culture.

Methods:

Twenty respiratory samples (nasal swab, aspirates, and sputum) from suspected COVID-19 cases, including 10 positive and 10 negative for SARS-CoV-2 by directed PCR targeting three viral genes (RdRp, E, and N) were tested using the Explify® respiratory platform (IDbyDNA) using shotgun and targeted enrichment clinical metagenomics in a pilot feasibility study. Raw results for SARS-CoV-2 positivity, read counts, and percent coverage were compared with PCR positivity and Ct values.

A Laboratory Challenge Becomes a Computational One

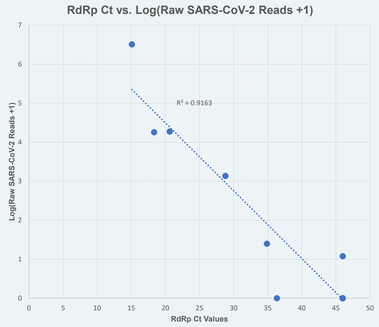


SARS-CoV-2 RT-PCR testing was performed using the EliTech InGenius platform (ELITech Group, Puteaux, France). This test has FDA EUA approval.

Results:

Six of the 10 PCR-positive samples (60%) were also positive by shotgun clinical metagenomics when compared to a respiratory virus database containing the SARS-CoV-2 reference sequence, based on minimum genome coverage of 5%. After enrichment, the same 6 samples were positive based on a coverage threshold of 25%, but read number and sequencing depth were markedly increased relative to those obtained by a shotgun approach. Additional microorganisms were also identified in samples, including potential coinfections with *S. marcescens*, *M. catarrhalis*, *H. parainfluenzae*.

UF Identifier	Sample Type	InGenious Result	RdRp Ct	E Ct	N Ct	Explify Results	SARS-CoV-2 Raw Reads	SARS-CoV-2 Coverage (%)
1 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
2 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
3 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
4 UF Health	NP	Not Detected	N/A	N/A	N/A	*H. parainfluenzae*	0	N/A
5 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
6 UF Health	Sputum, NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
7 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
8 UF Health	NP	Not Detected	N/A	N/A	N/A	*M. catarrhalis*	0	N/A
9 UF Health	NP	Detected	20.63	18.19	19.01	SARS-CoV-2	138840	100.0
10 UF Health	NP	Detected	15.09	13.21	14.94	SARS-CoV-2	3240851	100.0
11 UF Health	TA	Detected	18.35	15.6	17.13	SARS-CoV-2	18090	99.7
12 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
13 UF Health	NP	Not Detected	N/A	N/A	N/A	Negative	0	N/A
14 UF Health	TA	Detected (N gene only)	N/A	N/A	37.55	SARS-CoV-2	11	5.0
15 UF Health	NP	Detected (RdRp and N gene only)	36.34	N/A	34.54	*S. marcescens*	0	N/A
16 UF Health	NP	Detected (E and N gene only)	N/A	40.72	35.01	Negative	0	N/A
17 UF Health	NP	Detected (N gene only)	N/A	N/A	35.82	Negative	0	N/A
18 UF Health	NP	Detected (E and N gene only)	N/A	36.22	32.31	Negative	0	N/A
19 UF Health	NP	Detected	34.88	30.45	30.4	SARS-CoV-2, *M. catarrhalis*	24	9.2
20 UF Health	TA	Detected	28.8	26.28	27.27	SARS-CoV-2	1371	91.7



- Given the sensitive nature of the enrichment approach, level of detection studies were performed to further inform thresholds for positivity. Log transformation of read counts and the percent of genome coverage both showed strong inverse correlations with PCR Ct values.
- Comparison of RdRp Ct values to Log transform of total mapped SARS-CoV-2 reads, which represents closest correlation (Pearson = 0.92).

Clinical Metagenomics and SARS-CoV-2: Comparing Strains

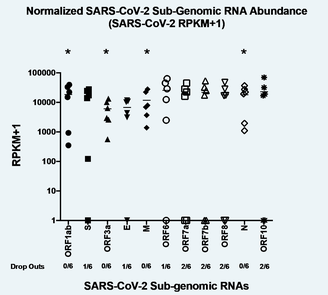
- Concern for nosocomial spread of COVID-19
- Samples sent from both healthcare worker and patient of healthcare worker
- 100% Pairwise identity between samples, could not rule out nosocomial transmission

Sample	Raw Reads	SARS-CoV-2 Reads	Genome Coverage	Mean Depth
9 UF	7,097,518	138,840	100%	573
10 UF	5,675,824	3,240,851	100%	11,771

Nucleotide Statistics:
Length: 29,903 bp
Sequences: 2
Identical Sites: 29,903 (100.0%)
Pairwise Identity: 100.0%

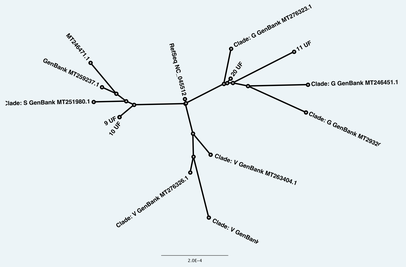
Ungapped lengths of 2 sequences:
Mean: 29903.0 Std Dev: 0.0
Minimum: 29903 Maximum: 29903

- Reads mapped to individual sub-genomic RNAs were normalized to total SARS-CoV-2 mapped reads and size of sub-genomic RNAs to give an RPKM statistic. Scatter of RPKM+1 sub-genomic reads are plotted on Log_{10} scale. Drop out (failure to detect sub-genomic RNA) is observed in 7/11 sub-genomic RNAs.
- Sub-genomic RNAs in which drop out is not observed (ORF1ab, ORF3a, M, N) may be more reliable for directed detection (RT-PCR).



UF Identifier	Explify Enrichment Results	Explify Enrichment Reads	Coverage	Fold Enrichment
1 UF Health	Negative	0	N/A	N/A
2 UF Health	Negative	0	N/A	N/A
3 UF Health	Negative	0	N/A	N/A
4 UF Health	Negative	0	N/A	N/A
5 UF Health	Negative	0	N/A	N/A
6 UF Health	Negative	0	N/A	N/A
7 UF Health	Negative	0	N/A	N/A
8 UF Health	Negative	0	N/A	N/A
9 UF Health	SARS-Cov-2	4343290	100	31.3
10 UF Health	SARS-CoV-2	12365719	100	3.8
11 UF Health	SARS-CoV-2	14919513	100	824.7
12 UF Health	Negative	0	N/A	N/A
13 UF Health	Negative	0	N/A	N/A
14 UF Health	SARS-CoV-2	18827	52.5	1711.5
15 UF Health	SARS-CoV-2*	3474	5.6	New Detection?
16 UF Health	Negative	0	0	N/A
17 UF Health	Negative	0	0	N/A
18 UF Health	SARS-CoV-2*	4525	8.1	New Detection?
19 UF Health	SARS-CoV-2	29149	39.5	1214.5
20 UF Health	SARS-CoV-2	1690613	100	1233.1

- Hybridization capture probes were designed across the SARS-CoV-2 genome.
- Enrichment of SARS-CoV-2 ranged from 3.8-1700-fold compared to shotgun data.
- Fold enrichment was more prominent in low abundance SARS-CoV-2 samples.
- Enrichment enabled full genome sequencing and potential new detections (UF 15, UF 18).



- Phylogenetic relationships of 4 SARS-CoV-2 isolates where enriched coverage is 100% were compared to reference sequences from representative clades.
- Unrooted tree with three major clades and RefSeq sample (NC_045512, Wuhan-Hu-1) shown.
- 2 samples appear to map within the G clade (11 UF and 20 UF) and 2 samples within the S clade (the identical 9 UF and 10 UF).

Conclusions:

A clinical metagenomic approach to SARS-CoV-2 diagnosis yielded agreement with directed PCR but was marginally less sensitive, especially in low positive (high Ct value) samples. Samples with low positivity for SARS-CoV-2 by PCR were missed by mNGS at preset thresholds but identified with optimization of thresholds. Refinements in clinical metagenomic workflows could enable a metagenomic approach to match or exceed the sensitivities of current PCR-based tests for SARS-CoV-2. Sequence data also have the potential to enhance surveillance, identify coinfections – which may become critical factors in clinical management – and facilitate early detection of variants, which may have epidemiological and laboratory implications when using fixed primers, in addition to providing laboratory evidence of SARS-CoV-2 infection.

Acknowledgements:

We thank the volunteers who participated in the study and the numerous technicians who processed these samples and generated data.

References:

1 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).
2 Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921 e910, doi:10.1016/j.cell.2020.04.011 (2020).