Rapid next-generation sequencing of the complete SARS-CoV-2 genome using the Ion AmpliSeq SARS-CoV-2 Insight Research Assay

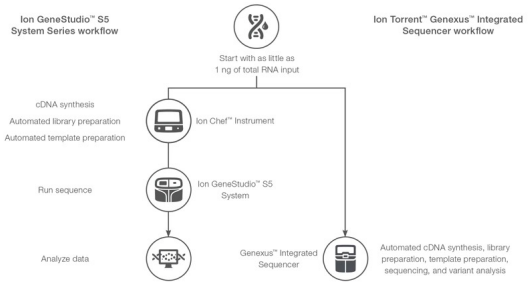
Anna McGeachy, Anupma Sharma, Na Li, Wally Zhang, Xiaoping Duan, Loni Pickle, Jingwei Ni, Xuehong Meng, Lei Guo, Lucius Zheng, Yingfu Zhong, Sheri Ambriz, Bethia Hsaw, Yvonne Lin, Steven Roman, Anjali Shah, Fiona Hyland, Thermo Fisher Scientific, 200 Oyster Pt Blvd, South San Francisco, CA, USA, 94080

ABSTRACT

The Ion AmpliSeq™ SARS-CoV-2 Research Panel is a targeted Next Generation Sequencing (NGS) solution that facilitates analysis and monitoring of the complete SARS-CoV-2 genome. We report the intelligent design for robust performance and utility in a wide variety of SARS-CoV-2 research applications. We demonstrate the ability of the panel to assemble whole genomes and track naturally occurring variants.

INTRODUCTION

SARS-CoV-2 has resulted in over >510 million infections and >6 million deaths globally[1]. Rapid tracking of the introduction, spread, and evolution of the virus is of paramount importance to international health concerns. The Ion AmpliSeq™ SARS-CoV-2 Research Panel was designed to provide a sample-to-answer analysis solution with as little as 2.5 hours for sequencing data. Automated workflows reduce hands-on time to as little as 45 minutes using our IonChef ™ system or 5 minutes with the turnkey Genexus™ sequencing system, reducing the risk of operator induced errors. Furthermore, the high sensitivity of the assay allows for sample inputs as low as 1 ng total RNA, increasing the eligibility of low viral load samples for testing. With the 540 chip on GeneStudio™ S5, up to 80 samples can be multiplexed, allowing for the high throughput necessary for epidemiological studies. Together, this uniquely positions the Ion AmpliSeq™ SARS-CoV-2 assay to provide rapid, high-quality results for tracking viral spread and evolution.



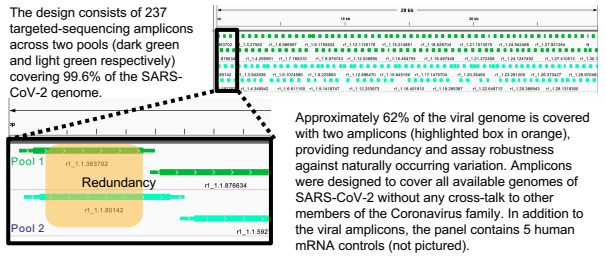
MATERIALS AND METHODS

Amplicons were designed to cover all serotypes of SARS-CoV-2 in GISAID[2] as of January 2020. The amplicons were designed with high analytical specificity and demonstrated no crosstalk to other Coronaviruses. Panel performance was evaluated using synthetic controls or heat-killed virus. Library preparation was performed in accordance with manufacturer's protocols manually or automatically on the Ion Chef™ System, sequenced on Ion GeneStudio™ S5 or Genexus™ System, and evaluated for sequencing uniformity, coverage analysis, variant calling, and consensus sequence assembly with built-in plugins.

RESULTS

Figure 1. Panel design and amplicon redundancy

The design consists of 237 targeted-sequencing amplicons across two pools (dark green and light green respectively) covering 99.6% of the SARS-CoV-2 genome.



Approximately 62% of the viral genome is covered with two amplicons (highlighted box in orange), providing redundancy and assay robustness against naturally occurring variation. Amplicons were designed to cover all available genomes of SARS-CoV-2 without any cross-talk to other members of the Coronavirus family. In addition to the viral amplicons, the panel contains 5 human mRNA controls (not pictured).

Figure 2. Panel size distribution and overlaps

The amplicons range in size from 99 to 275bp with a mean of 252bp. The insert length size ranges from 54 to 232bp with a mean of 201bp. We can further quantify our functional redundancy by measuring the overlap of our inserts. This insert overlap has led to exceptional performance over time against naturally occurring variation. Despite the many fold increase in the number of available genomes in GISAID over time, the panel continues to successfully amplify and assemble whole genomes of all current viral variants.

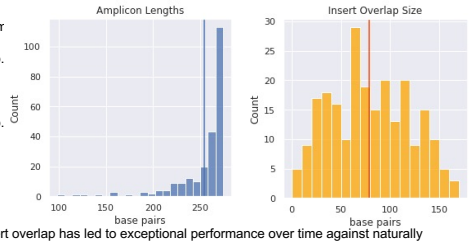


Figure 3. Panel performance metrics

We assayed the performance of the panel using synthetic controls (Twist 102019 [MT007544.1] and 102024 [MN9098947.3]) with 5ng of human lung RNA or heat-killed virus (ATCC#VR-1986HKTM). Libraries were constructed manually or using the IonChef™ system. These libraries were then sequenced using the GeneStudio™ S5 530 chip or Genexus™ system. More than 95% 1X coverage was obtained on viral copies as low as 20 and 100X for copies as low as 200.

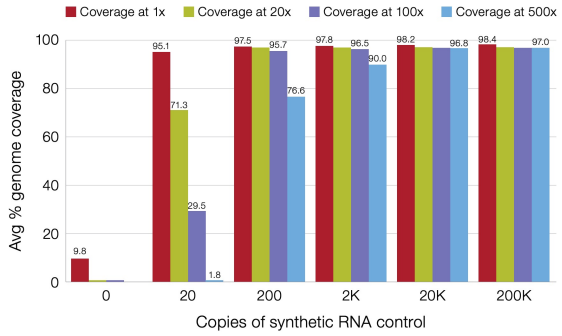
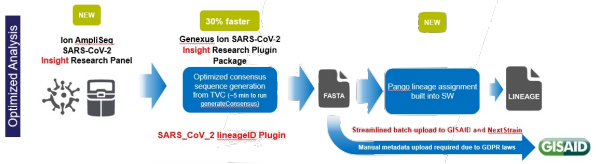


Figure 4. Sample-to-answer pipeline for SARS-CoV-2



The new Ion AmpliSeq ™ SARS-CoV-2 Insight Research plugin delivers a sample-to-answer solution by taking the output of the assay as sequenced on Genexus ™ or GeneStudio ™, performing genome assembly, annotating sample variants, and assigning the constellation of variants to lineages as described by pangolin[cite?]. This process, which uses Torrent Variant Caller (TVC), is optimized for AmpliSeq chemistry generates a consensus 30% faster than the prior research panel plugin. The resulting genome file (FASTA) with lineage assignment can then be uploaded to international, public databases (GISAID or NextStrain) through a streamlined batch upload.

Figure 5. Genome assembly of synthetic viral control

To test panel performance on emerging Variants of Concern (VOC), we constructed and sequenced libraries using 1,000-copy input of synthetic Omicron (BA.1) genome from Twist (Control 48, part #105204). This synthetic genome consists of six, non-overlapping 5kb fragments corresponding to a real-world isolate of Omicron (EPI_ISL_6841980), including mutations in the genome since initial panel design. Here we demonstrate the robustness of the assay to natural variation as we still demonstrate high coverage across the entire 29kb of the SARS-CoV-2 genome, excluding the gaps between synthesized viral fragments.

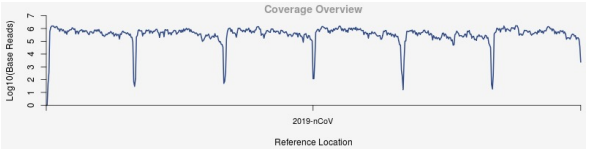


Figure 6. Variant calling of synthetic viral controls

Figure 6a. Overlap of 58 expected (known) and 56 observed (detected) mutations for Twist Control 48

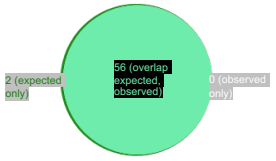
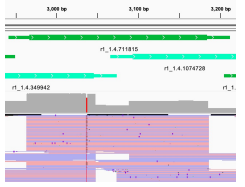
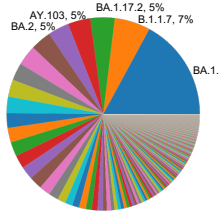


Figure 6b. IGV view Insight assay results for selected Twist Control 48 mutation



Gold-standard lineage assignment software such as pangolin[3] rely on the presence of key mutations to determine lineage. We used the Ion AmpliSeq ™ SARS-CoV-2 Insight Research assay and downstream plugin to call variants in triplicates of 1,000 copy Twist Control 48. Twist Control 48 carries 58 mutations, 33 of which are in the Spike protein used for cell-access and vaccine development. We observe 56 out of 58 mutations in each of the three replicates (fig. 6a). Observed mutations are well supported, with information from forward and reverse reads (red and blue respectively, fig. 6b). Moreover, these mutations are often covered by amplicons from multiple pools (green and light green, fig. 6b), providing redundancy in case of a loss of amplification efficiency of a single target. Even with the loss of two mutations due to template truncation and breaks, the Insight can accurately assign BA.1 to all three replicates for Twist Control 48.

Figure 7. Pango lineage distributions of Ion Torrent based submissions to GISAID



WHO lineage	pango lineage	# of sequence submissions	% of Ion Torrent sequence submissions
All lineages	*All lineages*	*261,551*	*100%*
Omicron	BA.1.1	44377	17.09%
Alpha	BA.1.1.7	15555	7.22%
Omicron	BA.1.17.2	10580	5.30%
Delta	AY.103	9810	5.19%
Omicron	BA.2	9125	5.09%

The international database, GISAID, has become central to the sharing and study of sequence data related to SARS-CoV-2. As of April 2022, there are 261,551 GISAID submissions tagged as Ion Torrent. These submissions date back to the start of the pandemic in December 2019. These Ion Torrent based sequences cover 792 distinct pangolin lineages. The top five represented pangol lineages correlate to the largest waves, Alpha (B.1.17), Delta (AY.103), and Omicron (BA.1.1, B.1.17.2, BA.2). Moreover, as BA.1.1 and BA.1.17.2 differ by only two mutations (ORF1a:1887I and S:701V), we demonstrate the high information density and quality of the Ion Torrent submissions.

CONCLUSIONS

The new Ion AmpliSeq ™ SARS-CoV-2 Insight Research assay uses a targeted sequencing approach for a sensitive, culture-free solution to whole SARS-CoV-2 genome sequencing. The assay requires minimal sample input and as little as 24 hours turnaround time with the Genexus™ system and custom assay plugin for genome assembly, variant annotation, and lineage assignment. With a high degree of functional redundancy, the panel has proven to be robust to natural variation, including recent VOC Omicron and Delta. We demonstrate the ability of the assay to assemble full genomes, discover underlying variation, and differentiate between closely related lineages. These features give the application exception flexibility, making it usable in a wide array of applications including contact-tracing and viral spread[4,5], epidemiology[6], and subgenomic RNAs[7] with a diversity of sample types such as nasopharyngeal swabs, wastewater[8], and post-mortem FFPE tissue[9].

REFERENCES

1. World Health Organization. (2020) Coronavirus disease (COVID-19) pandemic https://www.who.int/emergencies/diseases/novel-coronavirus-2019
2. Elbe, S., and Buckland-Merrett, G. (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges, 1:33-46. DOI:10.1002/gch2.1018 PMCID: 31565258
3. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L & Pybus OG (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology DOI:10.1038/s41564-020-0770-5
4. McNamara, Ryan P. et al. (2020) High-Density Amplicon Sequencing Identifies Community Spread and Ongoing Evolution of SARS-CoV-2 in the Southern United States. Cell Reports. https://doi.org/10.1016/j.celrep.2020.108352
5. Kaplan, Sarah et al. (2020) The code: How genetic science helped expose a secret coronavirus outbreak. The Washington Post. https://www.washingtonpost.com/graphics/2020/national/genetic-science-coronavirus-outbreak-travel/
6. Rapid COVID-19 Genome Sequencing Aids Outbreak Investigations. https://genomeweb.com/2020-05-06-Rapid-COVID-19-Genome-Sequencing-Aids-Outbreak-Investigations
7. Alexandersen, Soren, Chamings, Anthony, Bhatta, Tarka Raj. (preprint) SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. https://doi.org/10.1101/2020.06.01.20119792
8. Rimoldi, Sara G. et al. (2020) Presence and infectivity of SARS-CoV-2 virus in wastewaters and rivers. Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2020.140911
9. Sekulic, Miroslav, et al. (2020) Molecular Detection of SARS-CoV-2 Infection in FFPE Samples and Histopathologic Findings in Fatal SARS-CoV-2 Cases. American Journal of Clinical Pathology. https://doi.org/10.1093/ajcp/aqaa091

ACKNOWLEDGEMENTS

This assay is the combined effort of many people at Thermo Fisher Scientific, not all of whom we could list in the authors section of this poster. We would like to extend special thanks to the many members of the RND teams that worked to develop the biological and software components of this panel.

TRADEMARKS/LICENSING

ion torrent ThermoFisher SCIENTIFIC