# Illumina DRAGEN COVIDSeq Test (RUO) v1.3
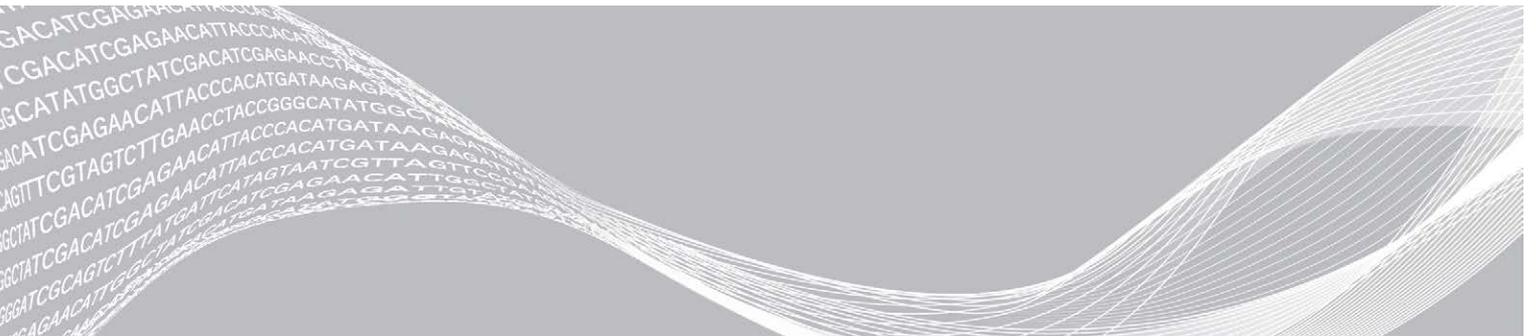
## App Guide

For Research Use Only. Not for use in diagnostic procedures.

**For Research Use Only. Not for use in diagnostic procedures.**

**For Research Use Only. Not for use in diagnostic procedures.**

# Revision History

| Document | Date | Description of Change |
|---|---|---|
| Document # 1000000129548 v01 | February 2021 | Added guidance for sample sheet to the Workflow Requirements section.<br>Added information about FAST mode and the Metrics and Logs Datasets options to the Set Parameters and Analysis Methods sections.<br>Added and updated quality control metrics for different instruments (including addition of NextSeq 2000) to the Lane Quality Control and Flow Cell Quality Control sections.<br>Updated software version numbers throughout the document from 1.2 to 1.3.<br>Updated the Output Folder Structure to reflect changes in the Sample_ Analysis folders.<br>Updated folder names and variant and region criteria in the Variant Calling and Consensus Sequence Generation for Research Use Only section.<br>Corrected and standardized the software naming convention throughout the document for the DRAGEN COVIDSeq Test (RUO) App. |
| Document # 1000000129548 v00 | July 2020 | Initial release. |

# Table of Contents

## Overview

The DRAGEN COVIDSeq Test (RUO) app analyzes sequencing reads of RNA libraries prepared using the Illumina COVIDSeq Test. The DRAGEN COVIDSeq Test (RUO) uses the Illumina DRAGEN Bio-IT Platform to perform analysis to determine the presence of SARS-CoV-2 as the output and generates results in PDF and tab-delimited formats.

Additionally, DRAGEN COVIDSeq Test (RUO) performs small variant calling for samples with at least 90 SARS-CoV-2 virus targets detected using the SARS-CoV-2 reference genome and generates a consensus sequence in FASTA format. Variant calls and consensus sequences are generated for informational purposes as research use only and not for patient reporting.

## Workflow Requirements

▶ A valid sample sheet (CSV) in BaseSpace Sequence Hub that was either:
  ▶ Used in a completed run.
    The completed run must use Run Monitoring and Storage as the Configuration option.
  ▶ Uploaded to a project folder in BaseSpace Sequence Hub via the GUI or BaseSpace CLI prior to running the app.

See the *Illumina COVIDSeq Test Reference Guide (document # 1000000126053)* for information on setting up your sample sheet.

## Set Parameters

1 Open DRAGEN COVIDSeq Test (RUO) from BaseSpace Sequence Hub as follows.

    a Select the Apps tab, and then select **DRAGEN COVIDSeq Test (RUO)**.
    b From the Version drop-down list, select **1.3.0**.
    c Select **Launch Application**.

2 To override the default name, enter a preferred name in the Analysis Name field.
  The default is the app name with the date and time the session started.

3 From the Save Results To field, select **Select Project**, and then select a project for the storage of app results.
  You can also create a new project by selecting **New** and then entering a name and description.

4 From the COVIDSeq Run to Analyze field, select **Select Run(s)** to select a completed COVIDSeq run to analyze.
  The completed run must contain a valid sample sheet.

5 **[Optional]** From the Override Sample Sheet field, select a sample sheet.
  Overriding the run's sample sheet allows you to fix any errors in the original sample sheet or adjust the scope of the analysis.

6 Set the Fast Mode option by selecting the Enable Fast Mode checkbox.
  In Fast Mode, only virus detection is performed. Alignment, variant calling and consensus generation is disabled.

7 To exclude run logs, QC metric files, and other file types, deselect the Create Metrics and Logs Datasets checkbox.

**For Research Use Only. Not for use in diagnostic procedures.**

Deselecting this option improves analysis speed, but only the results and sample analysis are generated under the `Logs_Intermediates_Lane_<Lane number or all>` folder. See *Analysis Methods on page 2* for more information on the subfolders excluded.

8    Select **Launch Application** to start the analysis.
When the analysis is complete, the status of the app session is automatically updated and you receive a confirmation email. When the app is launched through a workgroup account, the confirmation email is sent to the workgroup owner.

# Analysis Methods

The DRAGEN COVIDSeq Test (RUO) App performs analysis using the following steps. Each step creates a subfolder in Logs_Intermediates_Lane_<Lane number or all> subfolder under the analysis folder.

1    Validates the sample sheet fields.
This step generates the `SampleSheetValidation` subfolder. If the Metrics & Logs option is disabled, the subfolder is not available.

2    Performs run quality checks on the BCL data from the run folder.
This step generates the `RunQC` subfolder. If the Metrics & Logs option is disabled, the subfolder is not available.

3    Converts BCL data in the run folder to FASTQ sample data. All samples from the run are available as FASTQ files compressed in a gzip.
This step generates the `FastqGeneration` subfolder. If the Metrics & Logs option is disabled, the subfolder is not available.

4    For each sample, DRAGEN COVIDSeq Test (RUO) App determines the presence of SARS-CoV-2 and an internal (human) control. The read coverage per target is compared to a fixed target threshold to determine covered targets. The number of covered targets is then used to detect SARS-CoV-2 (≥ virusThreshold) and the internal control (≥ humanThreshold). The result is preliminary and undergoes quality control in later steps.
The step generates the `VirusDetection` subfolder. If the Metrics & Logs option is disabled, the subfolder is not available.

5    For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, DRAGEN COVIDSeq Test (RUO) App aligns FASTQ files to the SARS-CoV-2 reference genome (NC_045512.2) and the human control amplicon sequences.
This step generates the `MapAlign` subfolder. If the FAST mode is enabled or Metrics & Logs option is disabled, the subfolder is not available.

6    For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, DRAGEN COVIDSeq Test (RUO) App performs variant calling to determine any variants present in the sample with respect to the SARS-CoV-2 reference genome. This step produces VCF files containing detected variants for each processed sample. See *Variant Calling and Consensus Sequence Generation for Research Use Only* on page 8 for more information.
This step generates the `VariantCalling` subfolder. If the FAST mode is enabled or Metrics & Logs option is disabled, the subfolder is not available.

7    For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, DRAGEN COVIDSeq Test (RUO) App generates a consensus genome in FASTA format using variant calls and coverage metrics as input. See *Variant Calling and Consensus Sequence Generation for Research Use Only* on page 8 for more information.

This step generates the `ConsensusFasta` subfolder. If the FAST mode is enabled or Metrics & Logs option is disabled, the subfolder is not available.

8    For all samples, the TSV Run Report Generator performs quality control of each sample and generates a report in TSV format. Quality control is performed at the lane, plate, and sample-level and incorporates information from NTC and positive controls before determining patient results.
This step generates the `JsonTSVReport` subfolder.

9    Generates a PDF report that contains the summarized information.

## Quality Control

Quality control is performed on each flow cell or flow cell lane, depending on your sequencing system, each index set, and each patient sample using the internal control, positive control, and NTC. If using the NovaSeq 6000, only the lanes and index set currently existing in the sample sheet are assessed.

### Lane Quality Control

If using the NovaSeq 6000, quality control is performed for each flow cell lane based on whether quality metrics pass predefined thresholds. If using the NextSeq 500/550 or NextSeq 550Dx, quality is performed on each flow cell. See *Flow Cell Quality Control* on page 3 for more information.

The flow cell lane must meet the following requirements to pass quality control. If the lane fails quality control, all index sets display a `N/A` QC status and all patient samples display `Invalid`.

| Flow Cell | % Q30 | Yield (Gb) |
|---|---|---|
| NovaSeq S4 | ≥ 90% | ≥ 88 |
| NovaSeq SP | ≥ 90% | ≥ 14 |

### Flow Cell Quality Control

If using the NextSeq 500/550, NextSeq 550Dx, or NextSeq 2000, quality is performed on each flow cell. If using the NovaSeq 6000, quality control is performed on each flow cell lane. See *Lane Quality Control* on page 3 for more information.

The flow cell must meet the following requirements to pass quality control. If the flow cell fails quality control, all index sets within the failed flow cell display a `N/A` QC status and all patient samples display `Invalid`.

Table 1    Flow Cell Quality Control Thresholds

| Flow Cell | % Q30 | Yield (Gb) |
|---|---|---|
| NextSeq 550/550 Dx HO | ≥ 80.0 | ≥ 12.0 |
| NextSeq 2000 | ≥ 85.0 | ≥ 14.0 |

### Index Set Quality Control

Quality control is performed on each index set based on the NTC and positive control samples. Each index set is required to have one NTC and one positive control sample. If the associate lane or flow cell failed QC, the index set is not assessed.

The index set fails QC if one of the following events occurs:

▶    The SARS-CoV-2 virus or internal control is detected in the NTC.

▶    The SARS-CoV-2 virus is not detected in the positive control.

▶    A software error occurs in either the NTC or positive control.

**For Research Use Only. Not for use in diagnostic procedures.**

If an index set fails QC, all patient samples in the index set display `Invalid`.

## Internal Control

An internal control is assessed for each patient sample. If the SARS-CoV-2 virus and the internal control are not detected in the patient sample, then the sample displays an Invalid result and the internal control is reported as `Fail`.

If the internal control is detected in the patient sample, then the internal control is reported as `Pass`.

If the SARS-CoV-2 virus is detected in the patient sample, but the internal control is not detected, the internal control is reported as `N/A`. The N/A internal control does not impact patient sample validity when the SARS-CoV-2 virus is detected.

## View Analysis Output

1  When analysis is complete, select **Analyses** from the My Data tab in BaseSpace Sequence Hub.

2  Select the analysis name.

3  Select any of the following tabs to view information on the analysis session.
   ▶  **Summary**—General information about the analysis session, including log files.
   ▶  **Reports**—Analysis reports.
   ▶  **Inputs**—The samples and settings specified for the analysis session.
   ▶  **Files**—Output files.

## Summary

The Summary page provides access to analysis settings, execution details, and log files.

| Property | Definition |
| --- | --- |
| Name | Name of the analysis session. |
| Application | Application that generated the analysis. |
| Date Started | Date and time the analysis session started. |
| Date Completed | Date and time the analysis session completed. |
| Duration | Duration of the analysis. |
| Compute Charge | Cost in iCredits to compute the analysis. |
| Session Type | Number of nodes used, which can be multinode or single node. |
| Node Count | Total number of nodes. |
| Status | Status of the analysis session:<br>• Queued for Analysis<br>• Running<br>• Complete<br>• Aborted |
| Delivery | Delivery status of the analysis results. |

## Reports

The DRAGEN COVIDSeq Test (RUO) app provides a summary of key analysis metrics for the analyzed samples.

**For Research Use Only. Not for use in diagnostic procedures.**

## Inputs

The Inputs page lists the samples and settings specified for the analysis session.

## Files

The Files page provides access to the output files for each sample analysis. The output files are organized in the following folder structure. Key output files are shown below.

▶ Results

    ▶ COVIDSeq_RunReport.pdf — Contains only results for patient samples. See *PDF Report* on page 6.

    ▶ COVIDSeq_RunReport.tsv — Contains test results for both patient and control samples. See *TSV Run Report* on page 7

    ▶ Errors.tsv

▶ Logs_Intermediates_Lane_<Lane number> (on NovaSeq 6000) — See *Variant Calling and Consensus Sequence Generation for Research Use Only* on page 8. On NextSeq 500/550 and NextSeq 550Dx this folder is called Logs_Intermediates_Lane_all.

    ▶ ConsensusFasta

        ▶ <Sample ID>

            ▶ <Sample ID>.consensus_metrics.csv

    ▶ FastqGeneration

        ▶ Reports

            ▶ Demultiplex_Stats.csv

            ▶ Top_Unknown_Barcodes.csv

    ▶ FindSampleValidity

    ▶ JsonTsvReport

    ▶ MapAlign

    ▶ RunQc

    ▶ SampleSheetValidation

    ▶ VariantCalling

    ▶ VirusDetection

▶ Sample_Analysis

    ▶ <Sample ID>

        ▶ <Sample ID>.fasta

        ▶ <Sample ID>.fasta.md5sum

        ▶ <Sample ID>.bam

        ▶ <Sample ID>.bam.bai

        ▶ <Sample ID>.bam.md5sum

        ▶ <Sample ID>.hard-filtered.vcf.gz

        ▶ <Sample ID>.hard-filtered.vcf.gz.md5sum

        ▶ <Sample ID>.hard-filtered.vcf.gz.gz.tbi

        ▶ <Sample ID>.consensus_filtered_variants.vcf.gz

- ▶ <Sample ID>.consensus_filtered_variants.vcf.gz.md5sum
- ▶ <Sample ID>.consensus_filtered_variants.vcf.gz.gz.tbi

## PDF Report

The **COVIDSeq_RunReport.pdf** run report is located in the Results subfolder in the analysis folder.

The report contains the following sections:

- ▶ **Run Information**—Includes information on the following fields.

| Field | Description |
|---|---|
| Run ID | The unique ID associated with the sequencing run. |
| Run Date | The date of the sequencing run. |
| Instrument Serial | The unique serial number associated with the sequencing system. |
| Flow Cell ID | Unique ID for the sequenced flow cell. |
| Software Version | The software version used to perform analysis and generate reports. |

- ▶ **Quality control**—Includes information on the following fields.

| Field | Description |
|---|---|
| Lane 1, Lane 2, Lane 3, Lane 4 | The QC result for each lane. If using the NovaSeq 6000, values can include `PASS` or `FAIL`. If using the NextSeq 500/550 or NextSeq 550Dx, the value is `Lane 1,2,3,4`. |
| Index Set 1, Index Set 2, Index Set 3, Index Set 4 | The QC result for each index set within the associated lane. Values can include `PASS`, `FAIL`, or `N/A`. |

- ▶ **Invalid Results, SARS-CoV-2 Detected, SARS-CoV-2 Not Detected**—List of all patient samples with Invalid, SARS-CoV-2 Detected, or SARS-CoV-2 Not Detected results. The number of samples is displayed in each section's header.

| Field | Description |
|---|---|
| Sample ID | The sample ID in the sample sheet. |
| Internal control | The quality control result for the internal (human) control in a patient sample. Values include `Pass`, `Fail`, or `N/A`. |
| Result | The result for the patient sample. Possible values include the following: `SARS-CoV-2 Detected`—The sample lane or flow cell and index set passed quality control and the SARS-CoV-2 virus is detected in the sample. `SARS-CoV-2 Not Detected`—The sample lane or flow cell and index set passed quality control, the internal (human) control was detected in the sample, and the SARS-CoV-2 virus is not detected. `Invalid`—The sample lane or flow cell or index set failed quality control, a software error occurred for the sample, or neither the internal control or the SARS-CoV-2 virus was detected. |
| Consensus Sequence | Indicates if the consensus SARS-CoV-2 sequence was generated for the sample. |
| Lane / Index Set | The lane and index set associated with the sample. For Lane, values can include `Lane 1`, `Lane 2`, `Lane 3`, or `Lane 4`. If using the NextSeq 500/550 or NextSeq 550Dx, the value is `Lane 1,2,3,4`. For Index Set, values can include `Index Set 1`, `Index Set 2`, `Index Set 3`, or `Index Set 4`. If the lane or index set failed quality control, `Fail` is included at the end of the field value. |

# TSV Run Report

The **COVIDSeq_RunReport.tsv** run report is located in the Results subfolder in the analysis folder.

The report contains the following sections:

▶ **Header**—Contains information on the test name, run ID, run date, report date/time, instrument serial number, flow cell ID, and software version.

▶ **Quality Control**—Contains information about the quality control status for each lane or flow cell and each index set. Lane values can be `PASS` or `FAIL`. Index set can be `PASS`, `FAIL`, or `N/A`.

▶ **Patient Sample Results**—The patient sample results include the following fields:

| Field | Description |
|---|---|
| Sample ID | The sample ID in the sample sheet. |
| Internal control | The status of the internal control in a patient sample. Possible values include `Pass`, `Fail`, or `N/A`. |
| Result | The result for the patient sample. Possible values include the following:<br>`SARS-CoV-2 Detected`—The sample lane or flow cell and index set passed quality control and the SARS-CoV-2 virus is detected in the sample.<br>`SARS-CoV-2 Not Detected`—The sample lane or flow cell and index set passed quality control, the internal control was detected in the sample, and the SARS-CoV-2 virus is not detected.<br>`Invalid`—The sample lane or flow cell index set failed quality control, a software error occurred for the sample, or neither the internal control or the SARS-CoV-2 virus was detected. |
| Consensus Sequence | Indicates if the consensus SARS-CoV-2 sequence in FASTA format was generated for the sample. |
| Lane | The flow cell lane associated with the sample. If using the NovaSeq 6000, values can include `1`, `2`, `3`, or `4`. If using the NextSeq 500/550 or NextSeq 550Dx, the value is `1,2,3,4` all together. If using the NextSeq 2000, the value is `1`. |
| Index Set | The index set/adapter plate associated with the sample using the values from the Index_ID or Index/Index 2 columns in the sample sheet. Values can be `1`, `2`, `3`, or `4`. |
| Index ID | The index ID associated with the sample. If the Index_ID column is specified in the sample sheet, the Index ID field displays the same value. If not specified, Index ID is derived from the Index and Index2 columns from the sample sheet. |

▶ **Control Sample Results**—The control sample results include the following fields:

| Field | Description |
|---|---|
| Sample ID | The sample ID specified in the sample sheet. |
| Control Type | The control sample type. Values can include `Positive` or `NTC`. |
| Human Control | Indicates if the internal (human) control is detected in a control sample. Values can include `Detected` or `Not Detected`. |
| SARS-CoV-2 | Indicates if SARS-CoV-2 is detected in the control sample. Values can include `Detected` or `Not Detected`. |
| Lane | The flow cell lane associated with the same. If using the NovaSeq 6000, values can include `1`, `2`, `3`, or `4`. If using the NextSeq 500/550 or NextSeq550Dx, the value is `1,2,3,4` all together. If using the NextSeq 2000, the value is `1`. |
| Index Set | The index set/adapter plate associated with the control sample using the values from the Index_ID or Index/Index 2 columns in the sample sheet. Values can be `1`, `2`, `3`, or `4`. |

| Field | Description |
|---|---|
| Index ID | The index ID specified in the sample sheet. If the Index_ID column is specified in the sample sheet, the Index ID field displays the same value. If not specified, Index ID is derived from the Index and Index2 columns from the sample sheet. |

## Variant Calling and Consensus Sequence Generation for Research Use Only

DRAGEN COVIDSeq Test (RUO) App performs variant and consensus sequence generation for each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 virus targets detected. Variant calls and consensus sequences are for information purposes only and should not be used for patient reporting.

Variant calling and consensus sequence generation is not performed for invalid samples.

The variant calling output file is generated in VCF 4.2 file format and located in Sample_Analysis/<Sample ID>/<Sample ID>.hard-filtered.vcf.gz.

The consensus filtered variant calling output file is located in Sample_Analysis/<Sample ID>/<Sample ID>.consensus_filtered_variants.vcf.gz.

To generate a consensus sequence in FASTA format, detected sequence variants that meet the following criteria are applied to the SARS-CoV-2 reference sequence (NCBI Accession NC_045512.2).

▶ All DRAGEN quality filters pass.

▶ Allele frequency is greater than or equal to 0.5.

▶ Depth is greater than 10.

Regions of sequence with coverage below 10 are masked as low-confidence. Hard-masking is applied, and all bases in low-confidence regions are converted to "N". A soft-masked sequence is also provided and indicates all low-confidence regions with lower case characters.

The hard-masked consensus FASTA is available in Sample_Analysis/<Sample ID>/<Sample ID>.fasta.

## Download Project Files with BaseSpace CLI

Due to the size and sheer number of files contained in a BaseSpace Sequence Hub project, attempting to download project files via the BaseSpace Sequence Hub GUI may be problematic. It is recommended that you use BaseSpace CLI to download all the files in a project.

Enter the following command in BaseSpace CLI to download all the files in a BaseSpace project.

```
bs project download --id <project ID> -o <output directory>
```

# Technical Assistance

For technical assistance, contact Illumina Technical Support.

| | |
|---|---|
| **Website:** | www.illumina.com |
| **Email:** | techsupport@illumina.com |

Illumina Customer Support Telephone Numbers

| Region | Toll Free | Regional |
|---|---|---|
| North America | +1.800.809.4566 | |
| Australia | +1.800.775.688 | |
| Austria | +43 800006249 | +43 19286540 |
| Belgium | +32 80077160 | +32 34002973 |
| China | 400.066.5835 | |
| Denmark | +45 80820183 | +45 89871156 |
| Finland | +358 800918363 | +358 974790110 |
| France | +33 805102193 | +33 170770446 |
| Germany | +49 8001014940 | +49 8938035677 |
| Hong Kong, China | 800960230 | |
| Ireland | +353 1800936608 | +353 016950506 |
| Italy | +39 800985513 | +39 236003759 |
| Japan | 0800.111.5011 | |
| Netherlands | +31 8000222493 | +31 207132960 |
| New Zealand | 0800.451.650 | |
| Norway | +47 800 16836 | +47 21939693 |
| Singapore | +1.800.579.2745 | |
| South Korea | +82 80 234 5300 | |
| Spain | +34 911899417 | +34 800300143 |
| Sweden | +46 850619671 | +46 200883979 |
| Switzerland | +41 565800000 | +41 800200442 |
| Taiwan, China | 00806651752 | |
| United Kingdom | +44 8000126019 | +44 2073057197 |
| Other countries | +44.1799.534000 | |

**Safety data sheets (SDSs)**—Available on the Illumina website at support.illumina.com/sds.html.

**Product documentation**—Available for download from support.illumina.com.

Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

**For Research Use Only. Not for use in diagnostic procedures.**

illumina®