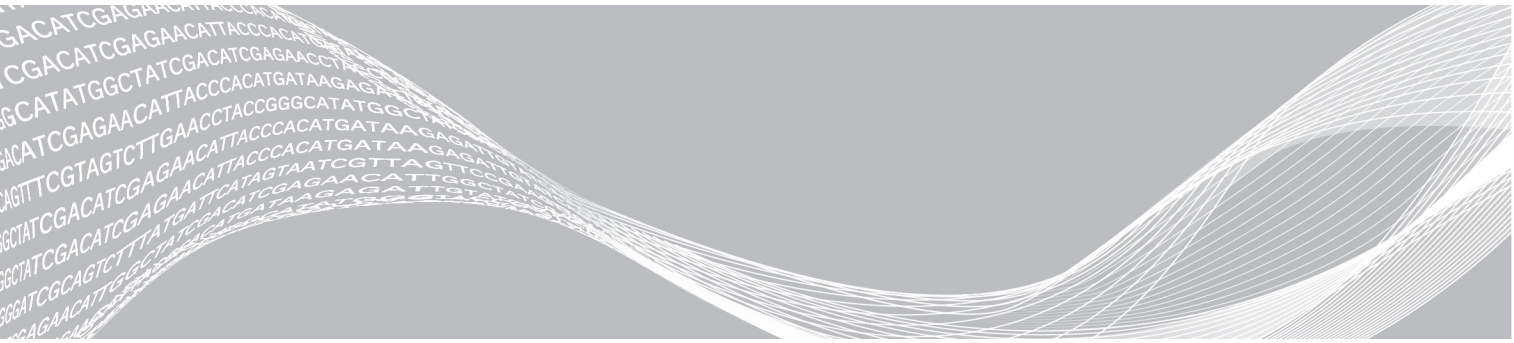
illumina®

Illumina DRAGEN COVIDSeq Test Pipeline

Software Guide



Document # 1000000128119 v00

ILLUMINA PROPRIETARY

August 2020

FOR IN VITRO DIAGNOSTIC USE

FOR USE UNDER AN EMERGENCY USE AUTHORIZATION (EUA) ONLY

FOR PRESCRIPTION USE ONLY

Table of Contents

Intended Use

The Illumina® COVIDSeq™ Test is a Next-Generation Sequencing (NGS) *in vitro* diagnostic test on the Illumina NovaSeq 6000 Sequencing System, NextSeq 500 Sequencing System, NextSeq 550 Sequencing System, or NextSeq 550Dx Instrument intended for the qualitative detection of SARS-CoV-2 RNA from nasopharyngeal (NP) swabs, oropharyngeal (OP) swabs, anterior nasal swabs, mid-turbinate nasal swabs, nasopharyngeal wash/aspirates, nasal aspirates, and bronchoalveolar lavage (BAL) specimens from individuals suspected of COVID-19 by their healthcare provider.

Testing is limited to laboratories certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), 42 U.S.C. §263a, to perform high complexity tests.

Results are for the identification of SARS-CoV-2 RNA. The SARS-CoV-2 RNA is generally detectable in respiratory specimens during the acute phase of infection. Positive results are indicative of the presence of SARS-CoV-2 RNA; clinical correlation with patient history and other diagnostic information is necessary to determine patient infection status. Positive results do not rule out bacterial infection or co-infection with other viruses. Laboratories within the United States and its territories are required to report all positive results to the appropriate health authorities.

Negative results do not preclude SARS-CoV-2 infection and should not be used as the sole basis for patient management decisions. Negative results must be combined with clinical observations, patient history, and epidemiological information. The Illumina® COVIDSeq™ Test is intended for use by qualified and trained clinical laboratory personnel specifically trained in the use of the NovaSeq 6000 Sequencing System, the NextSeq 500 Sequencing System, the NextSeq 550 Sequencing System, or the NextSeq 550Dx Instrument, as well as Next-Generation Sequencing workflows and *in vitro* diagnostic procedures. The Illumina® COVIDSeq™ Test is only for use under the Food and Drug Administration's Emergency Use Authorization.

Warnings and Limitations

- This test has not been FDA cleared or approved.
- This test has been authorized by FDA under an EUA for use by laboratories certified under the Clinical Laboratory Improvement Amendments (CLIA) of 1988, 42 U.S.C. §263a, to perform high complexity tests.
- This test has been authorized only for the detection of nucleic acid from SARS-CoV-2, not for any other viruses or pathogens.
- This test is only authorized for the duration of the declaration that circumstances exist justifying the authorization of emergency use of *in vitro* diagnostic tests for detection and/or diagnosis of COVID-19 under Section 564(b)(1) of the Act, 21 U.S.C. § 360bbb-3(b)(1), unless the authorization is terminated or revoked sooner.

Overview

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

Additionally, Illumina DRAGEN COVIDSeq Test Pipeline 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.

The Illumina DRAGEN COVIDSeq Test Pipeline requires a sample sheet. See the Illumina COVIDSeq Test Instructions for Use (document # 1000000128490) for information on creating a sample sheet.

Installation Requirements

Illumina DRAGEN COVIDSeq Test Pipeline contains the following minimum operating requirements.

The Illumina DRAGEN COVIDSeq Test Pipeline is compatible with a DRAGEN Server v2 and v3.

By default, the software includes the following items:

- Linux CentOS 7.3 operating system, or later.

The following additional software is required before installing Illumina DRAGEN COVIDSeq Test Pipeline.

- Docker version 18.09, or later.

Storage Requirements

The DRAGEN Server provides NVMe SSD located in `/staging` directory to use as the software output directory.

If using the DRAGEN Server v2, store sequencing run data in a network-attached folder to make sure the required disk space is available on the NVMe SSD drives for analysis output. Network-attached storage is required for long-term storage for both DRAGEN Server v2 and v3.

Analysis output is automatically written to the `/staging/covidseq_analysis_<timestamp>` to make sure the DRAGEN Server processes read and write data on the NVMe SSD. You can modify this location using the command-line.

Before beginning analysis, develop a strategy to copy data from the DRAGEN Server to a network-attached storage. Delete output data on the DRAGEN Server as soon as possible.

The following are the run and analysis output sizes for each sequencing system per 36 bp.

Sequencing System	Run Folder Output (GB)	Analysis Output (GB)
NovaSeq 6000 SP flow cell	20	60
NovaSeq 6000 S4 flow cell	225–240	860
NextSeq 500/550 and 550Dx HO flow cell	12	30

Install the Illumina DRAGEN COVIDSeq Test Pipeline

Use the instructions in this section to install the Illumina DRAGEN COVIDSeq Test Pipeline.

Illumina recommends running Docker as a non-root user by adding the user to the docker group. It is possible to run the Illumina DRAGEN COVIDSeq Test Pipeline as root but not recommended. For more information, see the Docker website.

The Illumina DRAGEN COVIDSeq Test Pipeline installation script uninstalls any existing DRAGEN software on the server. If you would like to use a different DRAGEN pipeline, you will need to uninstall the Illumina DRAGEN COVIDSeq Test Pipeline, and download a DRAGEN software installation package from the DRAGEN support page.

1 Contact your local Illumina Field Application Scientist to obtain the Illumina DRAGEN COVIDSeq Test Pipeline installer package.

2 Install Docker 18.09 or later using the install instructions for CentOS provided in the Docker documentation.

3 Install the DRAGEN Server license using the instructions provided in the Illumina DRAGEN Server Site Prep & Installation Guide.

4 Download the Illumina DRAGEN COVIDSeq Test Pipeline installation script provided in the email from Illumina.
The link expires after 72 hours.

5 Store the install script in the `/staging` directory.

6 To update the run script permissions, enter the following command:
```
chmod +x /staging/install_covidseq-EUA-1.2.0.run
```

7 To run the installation script, enter the following command:
```
/staging/install_covidseq-EUA-1.2.0.run
```
The script removes any previously installed DRAGEN software,

Running the System Check

Make sure that the system is functioning properly by running the `check_covidseq-1.2.0.sh` script. The self-test script checks the following functions:

- If all required services are running.
- If the proper Docker image is installed.
- If the Illumina DRAGEN COVIDSeq Test Pipeline successfully runs on a test data set.

The self-test runs for approximately five minutes. If the self-test prints a failure message, contact Illumina Technical Support and provide the `/staging/check_covidseq_<timestamp>.tgz` output file.

Running the Illumina DRAGEN COVIDSeq Test Pipeline

The Illumina DRAGEN COVIDSeq Test Pipeline is started by selecting the shell script using the command line, and then running the software with Docker. Analysis outputs are located in the `/staging/covidseq_analysis_<timestamp>` directory.

This location ensures that the server is on an optimized NVMe SSD.

Do not move files or press CRTL+C when the app is running. Moving files during the analysis can cause the analysis to fail or provide incorrect results. Pressing CRTL+C stops the analysis and might cause an error. If an error does occur, restart the server.

1 To run the Illumina DRAGEN COVIDSeq Test Pipeline, enter the following command-line argument:

```
covidseq.sh --runFolder <FULL_PATH_TO_RUN_FOLDER>
```

2 [Optional] Enter any of the other following available commands:
- **--analysisFolder**—Full path to the alternative analysis folder. For high performance, this folder must be on an NVMe SSD partition.
- **--sampleSheet**—Full path to the sample sheet. This command is required if your sample sheet is not named SampleSheet.csv.
- **--version**— Displays the version of the software, and then exits.
- **--fastMode**—Turns off alignment, variant calling, and consensus sequence FASTA generation to improve speed.
- **--help**—Displays a help screen, and then exits.

Process Lane Subsets or Multiple Flow Cells

If using the NovaSeq 6000 Sequencing System, Illumina DRAGEN COVIDSeq Test Pipeline supports processing subsets of lanes in a flow cell because quality control is performed at the lane-level.

To analyze a subset of lanes, create a copy of the sample sheet, and then remove all samples that are not in the lanes to process. Specify this new sample sheet on the command line.

To analyze multiple flow cells, perform multiple, serial executions of the software. Only initiate a new analysis after the previous is completed. Running multiple executions of the software concurrently on the same server can cause the analysis to fail or produce incorrect results.

Each flow cell includes a separate run folder.

Analysis Methods

The Illumina DRAGEN COVIDSeq Test Pipeline performs analysis using the following steps. Each step creates a subfolder in Logs_intermediates subfolder under the analysis folder.

1 Validates the sample sheet fields.
This step generates the `SampleSheetValidation` subfolder.

2 Performs run quality checks on the BCL data from the run folder.
This step generates the `RunQC` subfolder.

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.

4 For each sample, Illumina DRAGEN COVIDSeq Test Pipeline 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.

5 For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, Illumina DRAGEN COVIDSeq Test Pipeline 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.

6 For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, Illumina DRAGEN COVIDSeq Test Pipeline 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 10 for more information.
This step generates the `VariantCalling` subfolder.

7 For each sample with a result of "SARS-CoV-2 Detected" and at least 90 SARS-CoV-2 targets detected, Illumina DRAGEN COVIDSeq Test Pipeline 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 10 for more information.
This step generates the `ConsensusFasta` subfolder.

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.

Output Structure

The Illumina DRAGEN COVIDSeq Test Pipeline outputs results in the following folder structure.

- Results
  - COVID-Seq_RunReport.pdf
  - COVID-Seq_RunReport.tsv
  - Errors.tsv
- Logs_Intermediates
  - ConsensusFasta
    - <sampleID>
    - <sampleID>
  - FastqGeneration
  - FindSampleValidity
  - JsonTsvReport
  - MapAlign
  - VariantCalling
  - VirusDetection
- Sample_Analysis
  - <sampleID1>
    - <sampleID1>.fasta
    - <sampleID1>.bam
    - <sampleID1>.bam.bai
    - <sampleID1>.bam.md5sum
    - <sampleID1>.hard-filtered.vcf.gz
    - <sampleID1>.hard-filtered.vcf.gz.md5sum
    - <sampleID1>.hard-filtered.vcf.gz.tbi
    - sample_name.*.metrics.csv
  - <sampleID2>

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 7 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	≥ 85%	≥ 70
NovaSeq SP	≥ 85%	≥ 12

Flow Cell Quality Control

If using the NextSeq 500/550 or NextSeq 550Dx, 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 7 for more information.

The flow cell must have a %Q30 ≥ 85% or a total yield ≥ 12 Gb 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`.

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.

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.

Analysis Outputs

The Illumina DRAGEN COVIDSeq Test Pipeline generates the tab-separated values (TSV) and PDF report. The TSV report contains test results for both patient and control samples. The PDF report contains only results for patient samples.

TSV Run Report

The COVID-Seq_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: `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 control was detected in the sample, and the SARS-CoV-2 virus is not detected. `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 NextsSeq 550Dx, the value is `1,2,3,4`.
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`.

Field	Description
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 550Dx, the value is `1,2,3,4`.
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`.
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.

PDF Report

The COVID-Seq_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`.

Field	Description
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`. 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.

Variant Calling and Consensus Sequence Generation for Research Use Only

Illumina DRAGEN COVIDSeq Test Pipeline 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/Logs_Intermediates/VariantCalling/<SAMPLE ID>/<SAMPLE ID>.hard-filtered.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 0.7.
- Depth is greater than 15.

Regions of sequence with coverage below 15 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/Logs_Intermediates/ConsensusFasta/<SAMPLE ID>/<SAMPLE ID>.fasta.

Uninstall Illumina DRAGEN COVIDSeq Test Pipeline

The Illumina DRAGEN COVIDSeq Test Pipeline includes an uninstall script located in the `/usr/local/bin` called `uninstall_covidseq-1.2.0.sh`.

The uninstall script removes the following assets:

- All scripts (covidseq.sh, check_covidseq-1.2.0.sh, uninstall_covidseq-1.2.0.sh).
- The Illumina DRAGEN COVIDSeq Test Pipeline Docker image.
- Test data stored in /staging/illumina/covidseq.

The script does not uninstall Docker.

To uninstall the Illumina DRAGEN COVIDSeq Test Pipeline, enter the following command as root.

```
/usr/local/bin/uninstall_covidseq-1.2.0.sh
```

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.

Document # 1000000128119 v00

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

IVD

For Emergency Use Only

Prescription Use Only

FOR IN VITRO DIAGNOSTIC USE

FOR USE UNDER AN EMERGENCY USE AUTHORIZATION (EUA) ONLY

FOR PRESCRIPTION USE ONLY

illumina®