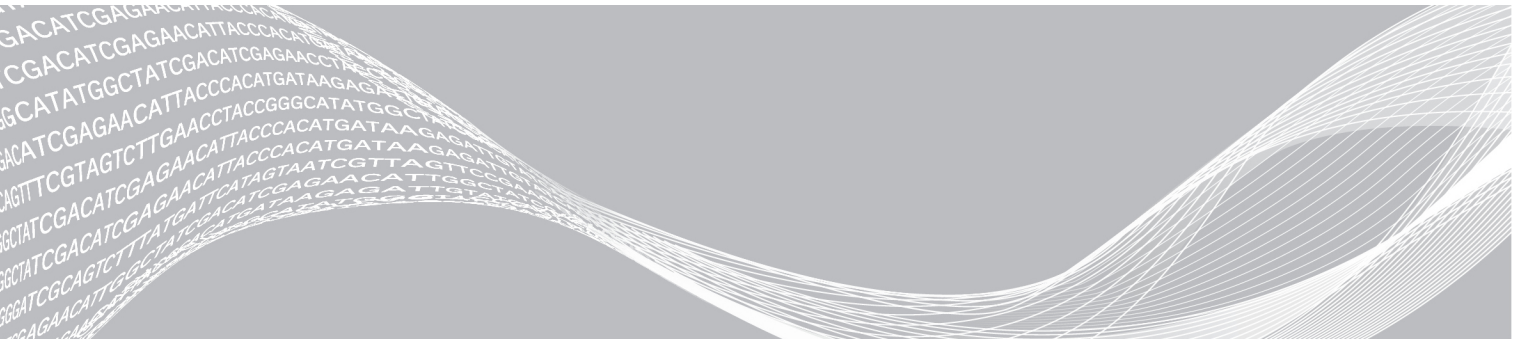
illumina®

VeriSeq NIPT Analysis Software (16 Samples)

User Guide



Document # 1000000012693 v05

April 2020

FOR IN VITRO DIAGNOSTIC USE

ILLUMINA PROPRIETARY

Revision History

Document	Date	Description of Change
Document # 1000000012693 v05	April 2020	Updated EU Authorized Representative address.
Document # 1000000012693 v04	July 2018	Added Limitations of the Procedure and Appendix B Method Comparison Study.
Document # 1000000012693 v03	January 2017	• Fetal Fraction Estimate - added further clarification regarding fetal fraction estimate. • Table 4 Normal Status Change Notifications and Action Requests - added note to the Email Notification Content Example for Invalid Sample IDs found in the Sample Sheet. • Sample Sheet Specification and Validation Rules - replaced the content of the second note. • Table 8 Sample NGS Option 1 Sheet Validation Rules (Data Section) - added "The Sample ID cannot contain spaces. Avoid combining multiple, adjacent underscores and dashes. As of version 1.4, the Sample_ID cannot begin with a 0 (zero)." to Validation Rules in Sample ID row. • Table 11 Sample NGS Option 2 Sheet Validation Rules (Data Section) - added "The Sample ID cannot contain spaces. Avoid combining multiple, adjacent underscores and dashes. As of version 1.4, the Sample_ID cannot begin with a 0 (zero)." to Validation Rules in Sample ID row.
Document # 1000000012693 v02	August 2016	Updated content for v1.4 release
Document # 1000000012693 v01	June 2016	Updated: • European Authorized Representative address and CE IVD marking on back cover. • System Overview • Sample_ID validation rules • Requeue Analysis for clarity and to provide troubleshooting information
Document # 1000000012693 v00	April 2016	Initial release.

Table of Contents

Overview

System Overview

VeriSeq NIPT Analysis Software (16 Samples) is available pre-installed on the VeriSeq NIPT Analysis Server (16 Samples), Illumina Catalog Number RH-400-1001. The server and pre-installed software provide:

- An analytical server with a capacity sufficient for analysis of sequencing data generated by up to 2 next-generation sequencing (NGS) instruments. The 2 NGS instrument options are:
  - A two flow cell sequencer utilizing 2-lane flow cells (NGS Option 1).
  - A single flow cell sequencer utilizing a 4-lane flow cell (NGS Option 2).
- A software suite capable of analyzing BCL formatted sequencing data generated by the sequencing software from libraries prepared according to the cfDNA Sequencing protocols to detect fetal aneuploidies based on chromosomal representation. The software suite contains 2 components:
  - Analysis Task Manager Service (ATMS)—A background service (daemon) that:
    - Monitors output paths for new run folders.
    - Analyzes metadata about the runs to compare sequencing run parameter configuration to a set of preconfigured analytical workflows.
    - Loads the sample sheet associated with each sequencing run, mapping identities of individual samples on a given flow cell to the indexes.
    - Prepares inputs for the analytical pipeline.
    - Executes the pipeline.
    - Tracks all the input and output data in a database.
    - Generates a run report for each of the individual samples on a flow cell.
  - cADAS—An analytical pipeline for detection of fetal aneuploidy from sequencing data generated from cfDNA isolated from maternal plasma.
    - Analyzes sequencing data by processing through alignment, coverage calculation, data normalization, and per-chromosome summarization.
    - Generates QC metrics and a pass, fail, or warning status for every sample.
    - Generates a score that characterizes over or underrepresented chromosomal material for each of the target chromosomes.

NOTE

The maximum number of failed samples allowed in a single batch is 4. Do not process batches through analysis that have fewer than 11 valid samples.

Intended Use

VeriSeq NIPT Analysis Software (16 Samples) generates quantitative scores to aid in the detection and differentiation of fetal aneuploidy status for chromosomes 21, 18, 13, X and Y by analyzing sequencing data generated from cell free DNA (cfDNA) fragments isolated from maternal peripheral whole blood specimens in pregnant women of at least 10 weeks gestation.

The quantitative scores are z-scores associated with under-or-over representation of a target chromosome relative to an expectation for a diploid genome.

Limitations of the Procedure

- The VeriSeq NIPT Analysis Software (16 Samples) is designed to be used as part of a screening test, which should not be considered in isolation from other clinical findings and test results. User defined cutoffs applied to the data outputs of this software should consider the relative benefits of increasing sensitivity at the cost of specificity and vice versa. No single cutoff achieves concurrent 100% sensitivity and 100% specificity. While rare, samples with a relatively low FF for the sequencing depth at which they have been processed can have data outputs near the threshold and may have lower accuracy.
- The VeriSeq NIPT Analysis Software (16 Samples) outputs data for use in reporting on the following:
  - Over representation of chromosomes 21, 18, and 13
  - The following sex chromosomal aneuploidies: XO, XXX, XXY, and XYY
- The VeriSeq NIPT Analysis Software (16 Samples) is not intended for use in reporting polyploidy.
- The algorithms used in the VeriSeq NIPT Analysis Software (16 Samples) can be confounded by certain maternal and fetal factors including, but not limited to, the following:
  - Recent maternal blood transfusion
  - Maternal organ transplant
  - Maternal surgical procedure
  - Maternal immunotherapy or stem cell therapy
  - Maternal malignancy
  - Maternal mosaicism
  - Confined placental mosaicism
  - Fetal demise
  - Disappearing twin
  - Fetal partial trisomy or partial monosomy
  - Fetal mosaicism

VeriSeq NIPT Analysis Software (16 Samples) Concepts

The following concepts and terms are common to VeriSeq NIPT Analysis Software (16 Samples).

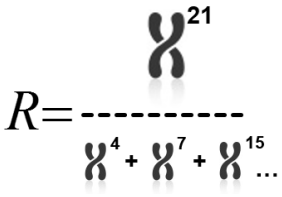
Concept	Description
cADAS	The analysis pipeline software. A server-side application used for sequencing data analysis and aneuploidy detection.
cfDNA	Cell-free DNA is DNA from both maternal and fetal origin circulating freely in the maternal blood stream. Analysis of cfDNA provides a method of noninvasive prenatal testing.
Run Folder	The folder structure generated by the NGS sequencing instrument and populated by RTA (Real-Time Analysis) primary data analysis.
Sample Sheet	A comma-separated values file (*.csv) that contains information required to set up and analyze a sequencing run, including a list of samples and their index sequences.
Workflow	An analytical process for analyzing sequencing runs performed by VeriSeq NIPT Analysis Software (16 Samples). The workflow for each run is specified in the sample sheet.

Software Analysis Overview

The VeriSeq NIPT Analysis Software (16 Samples) evaluates the copy number of test chromosomes in experimental samples. The analysis input is 36-base reads generated by a next-generation sequencing instrument. Reads are aligned against the whole human genome. Only reads that align to a unique location or site in the genome are used for further analysis. Duplicate reads are removed from the analysis. Reads are further filtered to exclude the sites that are associated with high variation in coverage across euploid samples. Raw coverage is adjusted through normalization for GC content and other factors at subchromosomal level, and then summarized into chromosomal coverage by robust mean of coverage along the chromosome.

Test chromosomes include 21, 18, and 13, X, and Y. Normalized coverage on test chromosomes is normalized to predefined reference (denominator) chromosomes to create the test chromosomal ratio (R). The predefined denominator chromosomes are optimized to maximally reduce variance in the chromosomal ratios for euploid samples. The chromosomal ratios for test samples are converted to normalized chromosomal values (NCVs) using a correction to flowcell-adjusted ratio mean and scaling by predefined expected variation in normal euploid samples (estimated from the training data).

Figure 1 Test Chromosome Ratio (R) Example



The Normalized Chromosomal Value (NCV) is determined according to the equation shown in Figure 2. The NCV value is equivalent to a z-score. A z-score describes the difference between a value and the population mean in terms of the standard deviation. The threshold for calling a sample as unaffected or affected based on NCV is determined by customers prior to their clinical validation of the workflow and can be adjusted based on the outcome of the clinical validation study.

Figure 2 Normalized Chromosomal Value (NCV)

$$NCV_{ik} = \frac{R_{ik} - \overline{R_{Ui}}}{\sigma_{Ui}}$$

i - Chromosome

k - Sample

U - Unaffected sample

R_{ik} - Ratio of chromosome *i* in the k^{th} sample

$\overline{R_{Ui}}$ - Flowcell-adjusted mean chromosomal ratio

σ_{Ui} - Standard deviation for the ratio of chromosome *i* in the unaffected samples from the training data set

Fetal Fraction Estimate

Fetal fraction refers to the percent of cell-free, circulating DNA in a maternal blood sample that is derived from the placenta. VeriSeq NIPT Analysis Software calculates the fetal fraction estimate based on differences in genomic coverage between maternal and fetal cfDNA.[1]

The VeriSeq NIPT Analysis Software (16 Samples) uses statistics generated during sequencing to provide a fetal fraction estimation (FFE) for each sample. The FFE is the estimated fetal cfDNA component that is recovered by the assay and reported as a rounded percentage for each sample. The average standard deviation of this estimate across all samples is 2%. The FFE is not to be used in isolation to exclude samples when reporting results.

[1] Kim, S.K., et al, Determination of fetal DNA fraction from the plasma of pregnant women using sequence read counts, Prenatal Diagnosis Aug 2015; 35(8):810-5. doi: 10.1002/pd.4615

System Operation

Logging In

The analytical server is set up as a Linux CentOS 6.6 machine with an sbsuser account.

Logging in to the server is not part of normal operation. It is only required to initiate reboot or shutdown.

Log in to the server using a terminal or an ssh connection with the initial preset credentials:

- **User Name**—sbsuser
- **Password**—Email Illumina Technical Support for password.
- **Group**—sbsuser

Organizing Data

The analytical server has a network sharing service setup that allows access to the hard drive from Windows systems through a samba sharing protocol. The preset user name and initial password for the samba shares is 'sbsuser' and 'sbs123'. Disk sharing for this user account through the samba protocol allows access to the following shares:

Location on the Linux Server	Share Name	User Name	Initial Password	Access Rights
/data01/runs	runs	sbsuser	Email Illumina Technical Support for password.	Read/Write
/data01/analysis_output	analysis_output	sbsuser	Email Illumina Technical Support for password.	Read

During sequencing run setup, set the output to the runs directory. Navigate to the \\<SERVER.IP.ADDRESS>\runs through the sequencing instrument control software run setup screens, where <SERVER.IP.ADDRESS> is the IP address of the Onsite Server.

The analysis_output directory contains reports for all flow cells processed through the cfDNA analytical workflow. The system organizes reports by the original run folder name generated by sequencing software and appends them with the analysis date and time.

For example, the analysis of run 140806_SN7001227_0199_AHABHTADXX generates an output folder named 140806_SN7001227_0199_AHABHTADXX_140806_230337.

Use the default run folder name format provided by your sequencing system. The VeriSeq NIPT Analysis Software requires that the run folder name contain only the following alphanumeric characters: a–z, A–Z, 0–9, and underscores ("_"). No spaces or other characters are allowed.

Sequencing Run Compatibility

The server only analyzes sequencing runs that are compatible with the cfDNA analytical workflow.

Configure sequencing using compatible read parameters.

For NGS Option 1:

- Read 1—36 bases
- Index 1 (i7)—7 bases

For NGS Option 2:

- Read 1—36 bases
- Index 1 (i7)—6 bases

Use only compatible sequencing methods and software versions to generate base calls.

NOTE

Regularly monitor sequencing data performance metrics to make sure that the quality of the data are within specification.

Table 1 NGS Option 1 Compatible Sequencing Methods and Software Versions

Parameter	Compatible Value
SBS	TruSeq Rapid SBS Kit TruSeq Rapid SBS Kit v1 or HiSeq Rapid SBS Kit v2
Index	TruSeq Rapid SR Cluster Kit TruSeq Rapid SR Cluster Kit v1 or HiSeq Rapid SR Cluster Kit v2
Clustering Choice	OnBoardClustering
Application Name	HiSeq Control Software
Application Version	2.0.12 or 2.2.38 or 2.2.58
FPGA Version	3.10.3 or 7.7.2.5 or 7.9.7
RTA Version	1.17.21 or 1.18.61 or 1.18.64

Table 2 NGS Option 2 Compatible Sequencing Methods and Software Versions

Parameter	Compatible Value
Application Name	NextSeq Control Software
Application Version	1.3.0 or 2.0.0 or 2.1.0
RTA Version	2.1.3 or 2.4.6 or 2.4.11

Workflow Timeout and Storage Requirements

The cfDNA analytical workflow is subject to the following timeout and storage limitations.

Table 3 Workflow Timeout and Storage Requirements

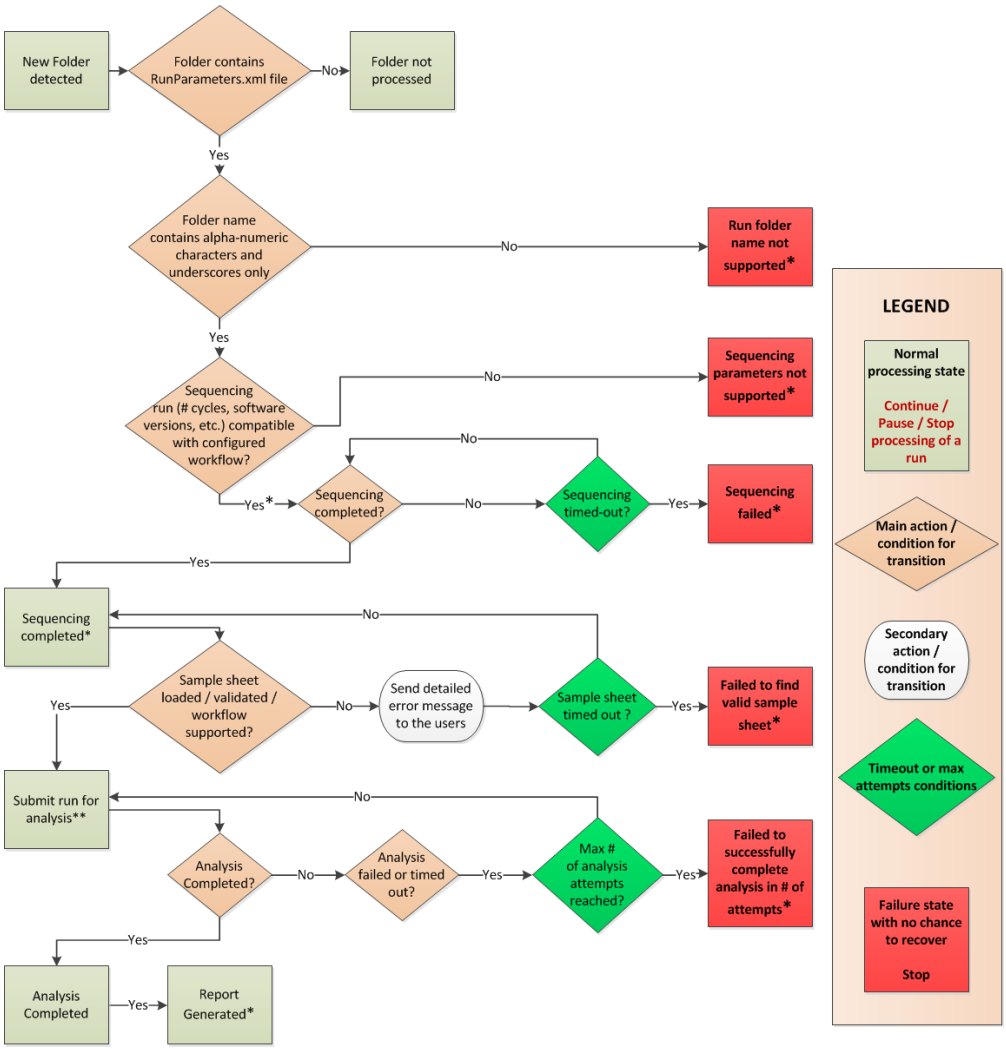
Parameter	Default Value
Maximum Run Parameters Wait Time	4 hours
Maximum Sequencing Time	20 hours
Maximum Sample Sheet Wait Time	96 hours
Maximum Analysis Time	3.5 hours
Minimum Scratch Space Storage	200 GB

System Data Flow

Under normal conditions, the ATMS sends sequencing run and analysis status notifications to users through an email system. Figure 3 shows the flow of data through the system and states with associated email notifications.

- **Gray Rectangles**—Normal processing states
- **Diamonds**—Primary conditions for transitioning to the next state
- **Oval**—Secondary condition for transitioning to the next state
- **Red Rectangles**—Failure states

Figure 3 Data Flow Diagram



* System generates email notification.

** If inadequate storage space is available on server, the system generates email notification.

During normal processing, the **ATMS**:

- Monitors its default directory (/data01/runs) for new sequencing runs. New sequencing runs are defined as folders containing a runParameters.xml file **[NGS Option 1]** or a RunParameters.xml file **[NGS Option 2]**.

- Verifies the compatibility of sequencing run parameters with predefined analysis workflows.
- Loads the sample sheet.
- Schedules and executes analytical processing to generate final reports.

Analysis is performed on one flow cell at a time. Additional flow cells awaiting analysis are queued on the server and proceed through analysis in the order they are loaded.

System Notifications

The system sends email notifications to individuals or email distribution groups set up during the server installation process. Illumina recommends using email distribution groups, which the email administrator can modify. If configured using individual addresses, the analysis server email configuration requires modification in the event of user changes. The email notifications indicate the status during normal operation and alert the user to any errors generated during analysis.

Table 4 describes the various email notifications sent by the system. The naming conventions in the table are required by the VeriSeq NIPT Analysis Software to import the NGS output files.

NOTE

Make sure that your email spam settings allow email notifications from the server. Email notifications are sent from an account named atms@<customer email domain>, where the <customer email domain> is specified by your local IT team when the server is installed.

Table 4 Normal Status Change Notifications and Action Requests

Condition	Normal/Warning/Error	Email Notification Content Example
Sequencing started. This notification is sent when the server detects a new fun folder. The run folder contains the run parameters file, which indicates that the sequencing has started with appropriate sequencing parameters. Run parameters file name: [NGS Option 1] runParameters.xml [NGS Option 2] RunParameters.xml	Normal Operation	Sequencing Run Folder Name: 140207_D00409_0027_ AH8HT6ADXX Sequencing Run Status: Sequencing started Sequencing Start Time: 2014-05-12 08:15 PDT Sequencing Complete Time: NA Workflow Name: NA Analysis Scheduled Time: NA Analysis Start Time: NA Analysis Finish Time: NA Analysis Output Directory: NA
Sequencing run completed.	Normal Operation	Sequencing Run Folder Name: 140207_D00409_ 0027_ AH8HT6ADXX Sequencing Run Status: Sequencing completed Sequencing Start Time: 2014-05-12 08:15 PDT Sequencing Complete Time: 2014-05-12 08:16 PDT Workflow Name: NA Analysis Scheduled Time: NA Analysis Start Time: NA Analysis Finish Time: NA Analysis Output Directory: NA
Sequencing run parameters not supported.	Error (not recoverable)	Sequencing run parameters for sequencing run '140207_D00409_0027_AH8HT6ADXX' are not supported by any of the configured workflows. This sequencing run folder will not be processed further. See the following errors: Workflow Name: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Mismatching Sequence Run Parameters found: NumCycles2, NumIndexed2 Found NumCycles2 value: 10, expected value: 7 Found NumIndexed2 value: 10, expected value: 7

Condition	Normal/Warning/Error	Email Notification Content Example
Incorrect flow cell barcode found in the sample sheet.	Warning (recoverable within 96 hours)	The sample sheet for sequencing run '140207_D00409_0027_AH8HT6ADXX' found in the sequencing run folder generated the following error: The flow cell ID (barcode) recorded in the sample sheet ('Experiment Name' slot) is ''. This barcode is required to be identical to the barcode associated with the run folder 'H8HT6ADXX'. Please correct the error in order to proceed with analysis. The sample sheet will be uploaded again in approximately 1 minute. The sample sheet is located in the run folder '/data01/runs/140207_D00409_0027_AH8HT6ADXX.
Unsupported workflow specified in the “Description” header row of the sample sheet.	Warning (recoverable within 96 hours)	The sample sheet for sequencing run '140207_D00409_0027_AH8HT6ADXX found in the sequencing run folder generated the following error: The workflow indicated in the sample sheet 'NIPT template1' is not supported by any of the configured workflows. The supported workflow names are: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Please correct the error in order to proceed with analysis. The sample sheet will be uploaded again in approximately 1 minute. The sample sheet is located in the run folder '/data01/runs/140207_D00409_0027_AH8HT6ADXX'.
Missing SampleSheet.csv file in the sequencing run folder.	Warning (recoverable within 96 hours)	The sample sheet for sequencing run '140207_D00409_0027_AH8HT6ADXX' in the sequencing run folder generated the following error: '/data01/runs/140207_D00409_0027_AH8HT6ADXX/SampleSheet.csv (No such file or directory)'. Please correct the error in order to proceed with analysis. The sample sheet will be uploaded again in approximately 1 minute. The sample sheet is located in the run folder '/data01/runs/140207_D00409_0027_AH8HT6ADXX'.

Condition	Normal/Warning/Error	Email Notification Content Example
Invalid Sample IDs found in the Sample Sheet	Error (recoverable by correcting Sample IDs)	Parsing the sample sheet for sequencing run '160217_NS500208_0021_AHK5NKBGXX' in the sequencing run folder generated the following error(s): Error: Invalid Sample IDs found (contain characters other than alpha-numerics / dashes / underscores). Invalid Sample ID values are: Plasma Control. Correct the error to proceed with analysis. The sample sheet will be uploaded again in approximately 1.0 minutes. Sample sheet should be located in the run folder '/data01/runs/160217_NS500208_0021_AHK5NKBGXX'. Note: This error is generated if any invalid characters, including spaces, are included in the sample sheet.
Missing header row in the sample sheet.	Warning (recoverable within 96 hours)	Attempt to load sample sheet for sequencing run '140207_D00409_0027_AH8HT6ADXX' generated the following error: Error: Invalid Sample Sheet Header. Missing required fields: Description Please correct the error in order to proceed with analysis. The sample sheet will be uploaded again in approximately 1 minute. The sample sheet is located in the run folder '/data01/runs/140207_D00409_0027_AH8HT6ADXX'.

Condition	Normal/Warning/Error	Email Notification Content Example
Duplicated index values listed in the Sample Sheet	Error (recoverable by correcting the Sample Sheet)	Parsing the sample sheet for sequencing run '140220_D00409_0041_AH8P5EADXX_COPY2' in the sequencing run folder generated the following error(s): Error: Duplicate Index value found: ACTGAT (A025) for Lane: 1 Invalid sample record found: S109_S109__A7_A025_ACTGAT___Test_62 for Index: ACTGAT Duplicate Index value found: ATTCCT (A027) for Lane: 1 Invalid sample record found: S113_S113__B7_A027_ATTCCT___Test_62 for Index: ATTCCT Duplicate Index value found: ACTGAT (A025) for Lane: 2 Invalid sample record found: S109_S109__A7_A025_ACTGAT___Test_62 for Index: ACTGAT Duplicate Index value found: ATTCCT (A027) for Lane: 2 Invalid sample record found: S113_S113__B7_A027_ATTCCT___Test_62 for Index: ATTCCT Correct the error to proceed with analysis. The sample sheet will be uploaded again in approximately 1.0 minutes. Sample sheet should be located in the run folder '/data01/runs/140220_D00409_0041_AH8P5EADXX_COPY2'.
Missing or invalid Lane value (NGS Option 1 only)	Error (recoverable by correcting Sample IDs)	Parsing the sample sheet for sequencing run '140220_D00409_0041_AH8P5EADXX_COPY' in the sequencing run folder generated the following error(s): Error: Invalid Lane value found at row: 47. Invalid value: Invalid Lane value found at row: 47. Invalid value: Invalid Sample IDs found (contain characters other than alpha-numerics / dashes / underscores). Invalid Sample ID values are: <blank> Correct the error to proceed with analysis. The sample sheet will be uploaded again in approximately 1.0 minutes. Sample sheet should be located in the run folder '/data01/runs/140220_D00409_0041_AH8P5EADXX_COPY'.

Condition	Normal/Warning/Error	Email Notification Content Example
Sequencing run failed. No RTA Complete file. This notification is sent when the RTA Complete file is not found after 20 hours.	Error (not recoverable—RTAComplete.txt file after maximum wait time of 20 hours)	Sequencing Run Folder Name: 140207_D00409_0027_AH8HT6ADXX_D12_NO_RTAComplete_TC_SC_3 Sequencing Run Status: Failed sequencing Sequencing Start Time: 2014-05-12 19:45 PDT Sequencing Complete Time: NA Workflow Name: NA Analysis Scheduled Time: NA Analysis Start Time: NA Analysis Finish Time: NA Analysis Output Directory: NA
Analysis started. This notification is sent when analysis starts. It appears after the RTA Complete appears, which triggers the analysis. Analysis takes 1—2 hours to run.	Normal Operation	Sequencing Run Folder Name: 140207_D00409_0027_AH8HT6ADXX Sequencing Run Status: Analysis started Sequencing Start Time: 2014-05-12 19:45 PDT Sequencing Complete Time: 2014-05-12 19:55 PDT Workflow Name: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Analysis Scheduled Time: 2014-05-12 20:05 PDT Analysis Start Time: 2014-05-12 20:06 PDT Analysis Finish Time: NA Analysis Output Directory: NA
Analysis failed The system automatically reprocesses the run 3 times.	Warning (recoverable by attempting to rerun analysis—ATMS requeues for processing up to 3 times)	Sequencing Run Folder Name: 140207_D00409_0027_AH8HT6ADXX Sequencing Run Status: Analysis failed. It will automatically be restarted to reprocess the run. Sequencing Start Time: 2014-05-11 08:26 PDT Sequencing Complete Time: 2014-05-11 08:27 PDT Workflow Name: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Analysis Scheduled Time: 2014-05-11 08:47 PDT Analysis Start Time: 2014-05-11 08:57 PDT Analysis Finish Time: 2014-05-11 08:59 PDT Analysis Output Directory: NA

Condition	Normal/Warning/Error	Email Notification Content Example
Maximum number of analysis attempts failed. This notification is sent after the third unsuccessful analysis attempt.	Error (not recoverable)	Sequencing Run Folder Name: 140207_D00409_0027_AH8HT6ADXX_TC_A_3 Sequencing Run Status: Maximum number of analysis attempts were exhausted. Please contact Illumina Technical Support. Sequencing Start Time: 2014-05-13 07:00 PDT Sequencing Complete Time: 2014-05-13 07:01 PDT Workflow Name: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Analysis Scheduled Time: 2014-05-13 07:09 PDT Analysis Start Time: 2014-05-13 07:11 PDT Analysis Finish Time: 2014-05-13 07:12 PDT Analysis Output Directory: NA
Run folder name has illegal characters.	Error (recoverable by removing illegal characters)	Invalid Sequencing Run Folder name found: '140207 D00409 0027 AH8HT6ADXX' The Sequencing Run Folder name can only contain the following alphanumeric characters: a–z, A–Z, 0–9, and underscores ("_"). No spaces or other characters are allowed. This sequencing run folder will not be processed further. Correct the run folder name to requeue for analysis.
cfDNA Sequencing report generated.	Normal Operation	Sequencing Run Folder Name: 140207_D00409_0027_AH8HT6ADXX Sequencing Run Status: Reports generated Sequencing Start Time: 2014-05-12 19:45 PDT Sequencing Complete Time: 2014-05-12 19:55 PDT Workflow Name: [NGS Option 1] cfDNAHiSeqv1.0 [NGS Option 2] cfDNANextSeqv1.0 Analysis Scheduled Time: 2014-05-12 20:05 PDT Analysis Start Time: 2014-05-12 20:06 PDT Analysis Finish Time: 2014-05-12 21:24 PDT Analysis Output Directory: /data01/analysis_output/140207_D00409_0027_AH8HT6ADXX_140512_200514

System Shutdown

Recovering from Unexpected Shutdown

In the event of a power outage or accidental shutdown by the user during analysis run, the system:

- Automatically restarts the software upon reboot.
- Recognizes the latest running analysis at the time of shutdown as failed and resubmits it to the queue for processing.
- Generates output when analysis successfully completes.

NOTE

If analysis fails, the software allows the system to resubmit the run for analysis up to 3 times.

Analysis and Reporting

Sample Sheet Specification and Validation Rules

This section provides instructions for creating the sample sheet, which is required for analysis of a run folder by the VeriSeq NIPT Analysis Software. Follow the instructions for the NGS option you are using.

NOTE

Confirm that sample ID mapping to the associated indexes is accurate. Accurate mapping is required to maintain sample integrity. Have a second person, other than the person who created it, verify the sample sheet prior to starting the sequencing run. Any errors in matching the samples to the proper indexes can lead to potentially incorrect results reported for misidentified samples.

NOTE

Always include a process and negative (no template) control in the sample batch. The process control (but not the negative control) should be added to the library pool, and identified as sample type Control on the sample sheet. Do not add the negative control to the sample batch or the sample sheet.

NGS Option 1

The VeriSeq NIPT Analysis Software (16 Samples) requires a sample sheet for each flow cell. For the NGS Option 1 workflow, the sample sheets are uploaded during sequencing run setup and placed in the output folder as “SampleSheet.csv”. The sample sheet is a comma-separated file that contains 2 sections: a header that captures run level information and a data section that captures sample-specific attributes. NGS Option 1 uses a 2-lane flow cell. The same sample pool is run in both lanes (1 and 2). When entering the sample information in the sample sheet, each Sample_ID, well and index combination must be listed in both lanes 1 and 2. The Sample_ID, well and index combination must be unique within a lane.

Confirm that mapping of the sample ID to the associated indexes is accurate. Accurate mapping is required to maintain sample integrity.

See Table 5 and Table 6 for example sample sheet header and data sections.

NOTE

The naming conventions in the following table are required by the VeriSeq NIPT Analysis Software to import the NGS output files.

Table 5 Example NGS OPtion 1 Sample Sheet (Header Section)

[Header]	
IEMFileVersion	4
Investigator Name	
Experiment Name	H9KY7ADXX

Date	
Workflow	GenerateFASTQ
Application	HiSeq FASTQ Only
Assay	TruSeq LT
Description	cfDNAHiSeqv1.0
Chemistry	Default
[Reads]	
36	
[Settings]	

NOTE

The header section of the sample sheet must have the exact flow cell ID (all uppercase letters) in the Experiment Name field, and the Description field must contain "cfDNAHiSeqv1.0".

Table 6 Example NGS Option 1 Sample Sheet (Data Section)

[Data]										
Lane	Sample_ID	Sample_ Name	Sample_ Plate	Sample_ Well	I7_Index_ ID	Index	Sample_ Project	Description	SampleType	Library_ nM
1	Sample1	Sample1		A1	A002	CGATGT			Test	80.87774
1	Sample2	Sample2		B1	A005	ACAGTG			Test	75.3396
1	Sample3	Sample3		C1	A007	CAGATC			Test	87.35632
1	Sample4	Sample4		D1	A012	CTTGTA			Test	68.02508
1	Sample5	Sample5		E1	A013	AGTCAA			Test	97.49216
1	Sample6	Sample6		F1	A014	AGTTCC			Test	93.20794
1	Sample7	Sample7		G1	A018	GTCCGC			Test	63.63636
1	Sample8	Sample8		H1	A019	GTGAAA		Failed Library	Test	5.2
1	Sample9	Sample9		A2	A001	ATCACG			Test	84.6395
1	Sample10	Sample10		B2	A003	TTAGGC			Test	81.5047
1	Sample11	Sample11		C2	A008	ACTTGA			Test	78.78788
1	Sample12	Sample12		D2	A010	TAGCTT			Test	83.17659
1	Sample13	Sample13		E2	A020	GTGGCC			Test	79.62382
1	Sample14	Sample14		F2	A022	CGTACG			Test	62.59143
1	Control-ID	Control-ID		G2	A025	ACTGAT			Control	65.20376
2	Sample1	Sample1		A1	A002	CGATGT			Test	80.87774
2	Sample2	Sample2		B1	A005	ACAGTG			Test	75.3396
2	Sample3	Sample3		C1	A007	CAGATC			Test	87.35632
2	Sample4	Sample4		D1	A012	CTTGTA			Test	68.02508
2	Sample5	Sample5		E1	A013	AGTCAA			Test	97.49216
2	Sample6	Sample6		F1	A014	AGTTCC			Test	93.20794
2	Sample7	Sample7		G1	A018	GTCCGC			Test	63.63636
2	Sample8	Sample8		H1	A019	GTGAAA		Failed Library	Test	5.2
2	Sample9	Sample9		A2	A001	ATCACG			Test	84.6395
2	Sample10	Sample10		B2	A003	TTAGGC			Test	81.5047
2	Sample11	Sample11		C2	A008	ACTTGA			Test	78.78788

2	Sample12	Sample12	D2	A010	TAGCTT	Test	83.17659
2	Sample13	Sample13	E2	A020	GTGGCC	Test	79.62382
2	Sample14	Sample14	F2	A022	CGTACG	Test	62.59143
2	Control-ID	Control-ID	G2	A025	ACTGAT	Control	65.20376

Sample sheet validation rules for header and data sections are outlined in Table 7 and Table 8. The data in each cell of the Sample Sheet cannot exceed 100 characters.

NOTE

The naming conventions in the following table are required by the VeriSeq NIPT Analysis Software to import the NGS output files.

Table 7 Sample Sheet Validation Rules (Header Section)

Field	Required	Validation Rules
IEMFileVersion	Yes	Must be 4.
Investigator Name	Yes	No validation rules.
Experiment Name	Yes	Must be the flow cell ID (all uppercase letters). Validated against the barcode from runParameters.xml.
Date	Yes	No validation rules.
Workflow	Yes	No validation rules.
Application	Yes	No validation rules.
Assay	Yes	No validation rules.
Description	Yes	Must be cfDNAHiSeqv1.0
Chemistry	Yes	No validation rules.

Table 8 Sample NGS Option 1 Sheet Validation Rules (Data Section)

Column Name	Interpretation	Class	Valid Entries	Required	Validation Rules
Lane	Lane on which the sample is located	Integer	1, 2	Yes	Must be 1 or 2.
Sample_ID	Sample ID (used for cADAS output reporting)	Character string	Unique per index within flow cell	Yes	For a given Sample ID, all the sample sheet data values must be identical other than lane. Sample ID can only contain alphanumeric characters including a–z, A–Z, 0–9, underscores, and dashes ("-"). The Sample ID cannot contain spaces. Avoid combining multiple, adjacent underscores and dashes. As of version 1.4, the Sample_ID cannot begin with a 0 (zero).
Sample_Name	Sample Name	Character string	Ignored	No	This field can be blank. No validation rules apply. Sample Name is truncated to 100 characters.
Sample_Plate	Sample Plate ID	Character string	PXXXX, where XXXX are numeric	No	This field can be blank. No validation rules apply. Sample Plate ID is truncated to 100 characters.

Column Name	Interpretation	Class	Valid Entries	Required	Validation Rules
Sample_Well	Sample Well ID	Character string	A01–A08 B01–B08	Yes	Both A1 and A01 formats are supported. Values are validated against a regular expression. First character A–H and next 2 can be 1–12 or 01–12.
I7_Index_ID	Index ID	Character string	A001–A024	Yes	First character is always A and then 3 numeric digits; see Table 12.
Index	Index composition	Character string		Yes	Any of the index sequences found in Table 12 are allowed. Total number of Index values within a given lane must be 8 or more. If less than 8, an error is generated. Additional validation is done to match the I7_Index_ID and Index value pairs. For a given Lane value, all the Index values must be unique.
Sample_Project	Project name	Character string	Ignored	No	This field can be blank.
Description	Sample Description	Character string	Ignored	No	This field can be blank. If the word “failed” is included in this field, the sample is marked as failed and no results are reported for it.
SampleType	Sample Type	Character string	‘Patient’, ‘Test’, ‘Control’	Yes	Must be Patient, Test, or Control. (Validation is case sensitive.)
Library_nM	Library concentration	Real	Numeric values	Yes	Must be numeric.

The user can exclude a sample from analysis by indicating failed (case insensitive) in the description field for the sample in the sample sheet. Doing so tracks samples throughout the entire workflow that do not go through sequencing due to presequencing QC failure. The value in the sample description field is included in the output file and the data fields contain blank values.

NGS Option 2

The NGS Option 2 run setup workflow does not include an option to upload a sample sheet manually at run setup. Instead, after a new run is detected, the user places the sample sheet named samplesheet.csv into the output run folder in the runs folder on the analysis server. ATMS sends the user an email stating a new run has been detected after the RunParameters.xml file is written to the run folder in the analysis server /data01/runs directory, and after sequencing begins. The sample sheet must be placed into the run folder before the end of the sequencing run (before the RTAComplete.txt file is written to the run folder).

NOTE

If the samplesheet.csv file is not present in the output run folder by the time the RTAcomplete.txt file is written, the analysis software will send a notification. See Chapter 2 System Operation, *System Notifications*, Table 4 on page 9.

When using NGS Option 2, the same sample pool is run across the entire flow cell. Lane numbers are not specified in the sample sheet. When entering the sample information in the sample sheet, each Sample_ID, well and index combination will be listed once in the data section of the sample sheet. Each Sample_ID, well and index combination should be unique.

Confirm that mapping of the sample ID to the associated indexes is accurate. Accurate mapping is required to maintain sample integrity.

See Table 9 and Table 10 for example sample sheet header and data sections.

NOTE

The naming conventions in the following table are required by the VeriSeq NIPT Analysis Software to import the NGS output files.

Table 9 Example NGS Option 2 Sample Sheet (Header Section)

[Header]	
IEMFileVersion	4
Investigator Name	Name
Experiment Name	FlowCellID
Date	2/4/2014
Workflow	GenerateFASTQ
Application	FASTQ Only
Assay	TruSeq LT
Description	cfDNANextSeqv1.0
Chemistry	Default
[Reads]	
36	
[Settings]	
ReverseComplement	0

NOTE

The header section of the sample sheet must have the exact flow cell ID (all uppercase letters) in the Experiment Name field, and the Description field must contain "cfDNANextSeqv1.0".

Table 10 Example NGS Option 2 Sample Sheet (Data Section)

[Data]									
Sample_ID	Sample_Name	Sample_Plate	Sample_ Well	I7_Index_ ID	Index	Sample_Project	Description	SampleType	Library_ nM
Sample1	Sample1		A2	A002	CGATGT			Test	53.2
Sample2	Sample2		B2	A005	ACAGTG			Test	51
Sample3	Sample3		C2	A007	CAGATC			Test	83.3
Sample4	Sample4		D2	A012	CTTGTA			Test	79
Sample5	Sample5		E2	A013	AGTCAA			Test	67
Sample6	Sample6		F2	A014	AGTTCC			Test	44.3
Sample7	Sample7		G2	A018	GTCCGC			Test	61.9
Sample8	Sample8		H2	A019	GTGAAA			Test	62.9
Sample9	Sample9		A4	A001	ATCACG			Test	76.8
Sample10	Sample10		B4	A003	TTAGGC			Test	71.1
Sample11	Sample11		C4	A008	ACTTGA		Failed_QC	Test	5
Sample12	Sample12		D4	A010	TAGCTT			Test	71.1
Sample13	Sample13		E4	A020	GTGGCC			Test	55
Sample14	Sample14		F4	A022	CGTACG			Test	88.6
Control-ID	Control-ID		G4	A025	ACTGAT			Control	64.7

Sample sheet validation rules for data sections are outlined in Table 11. The data in each cell of the Sample Sheet cannot exceed 100 characters.

Table 11 Sample NGS Option 2 Sheet Validation Rules (Data Section)

Column Name	Interpretation	Class	Valid Entries	Required	Validation Rules
Sample_ID	Sample ID (used for cADAS output reporting)	Character string	Unique per index within flow cell	Yes	Sample ID can only contain alphanumeric characters including a–z, A–Z, 0–9, underscores, and dashes ("-"). The Sample ID cannot contain spaces. Avoid combining multiple, adjacent underscores and hyphens. As of version 1.4, the Sample_ID cannot begin with a 0 (zero).
Sample_ Name	Sample Name	Character string	Free text	No	This field can be blank. No validation rules apply. Name is truncated to 100 characters.
Sample_ Plate	Sample Plate ID	Character string	PXXXX, where XXXX are numeric	No	This field can be blank. No validation rules apply. Sample Plate ID is truncated to 100 characters.
Sample_ Well	Sample Well ID	Character string	A01–A08 B01–B08	Yes	Both A1 and A01 formats are supported. Values are validated against a regular expression. First character A–H and next 2 can be 1–12 or 01–12.
I7_Index_ID	Index ID	Character string	A001–A024	Yes	First character is always A and then 3 numeric digits.
Index	Index composition	Character string		Yes	Any of the index sequences found in Table 12 are allowed. Total number of Index values within a given lane must be 8 or more. If less than 8, an error is generated. Additional validation is done to match the I7_ Index_ID and Index value pairs. For each sample sheet, all the Index values are unique. They cannot be duplicates.
Sample_ Project	Project name	Character string	Ignored	No	This field can be blank.
Description	Sample Description	Character string	Ignored	No	This field can be blank. If the word “failed” is included in this field, the sample is marked as failed and no results are reported for it.
SampleType	Sample Type	Character string	'Patient', ‘Test’, ‘Control’	Yes	Must be Patient, Test, or Control. (Validation is case sensitive.)
Library_nM	Library concentration	Real	Numeric values	Yes	Must be numeric.

The user can exclude a sample from analysis by indicating failed (case insensitive) in the description field for the sample in the sample sheet. Doing so tracks samples throughout the entire workflow that do not go through sequencing due to presequencing QC failure. The value in the sample description field is included in the output file and the data fields contain blank values. See Table 12 for valid index values.

Valid Index Values

Table 12 Valid Index Values

i7_Index_ID	Index
A001	ATCACG
A002	CGATGT
A003	TTAGGC
A004	TGACCA
A005	ACAGTG
A006	GCCAAT
A007	CAGATC
A008	ACTTGA
A009	GATCAG
A010	TAGCTT
A011	GGCTAC
A012	CTTGTA
A013	AGTCAA
A014	AGTTCC
A015	ATGTCA
A016	CCGTCC
A018	GTCCGC
A019	GTGAAA
A020	GTGGCC
A021	GTTTCG
A022	CGTACG
A023	GAGTGG
A025	ACTGAT
A027	ATTCCT

Demultiplexing and FASTQ Generation

NGS Option 1 uses a custom demultiplexer. NGS Option 2 uses the bcl2fastq v2 converter for demultiplexing and FASTQ generation. Both analysis options produce an additional Sample Sheet related file in the run folder, in addition to the original SampleSheet.csv.

- SampleSheet.csv—The original sample provided by the user.
- sample_sheet_processed_YYYY_MM_DD_hh-mm-ss.txt—A file generated by ATMS after reading in the Sample Sheet provided by the user. This file contains the information passed to subsequent data analysis steps.

NOTE

Do not access sample sheet while analysis is running unless instructed to do so during the sample sheet validation step.

Requeue Analysis

NOTE

Attempt a requeue ONLY after an email notification is received from the server concerning a Sample Sheet error.

You can requeue your run for analysis if your sample sheet contains errors that do not affect validation or analysis. Changes to the sample sheet as outlined below should only be made after an email notification is received from the server which indicates an error in the Sample Sheet. For example:

- Empty rows or columns
- Missing header row
- Unsupported workflow in the Description header row
- Incorrect flow cell barcode

Run Folder Located on the Server

This procedure describes how to requeue analysis when your run folder is located on the server.

1. From a computer on the same network as the analysis server, open Windows Explorer, and browse to the /runs directory.
2. Locate the run folder that you want to requeue for analysis.
3. Right-click the run folder, and click **Copy**.
4. Right-click anywhere in the /runs directory, and click **Paste**.
   A copy of the run folder is created with " - Copy" appended to the end of the folder name. For example, Run_Folder_Name - Copy.
   The system sends an email notification about illegal characters in the folder name, which you can disregard.

   NOTE

   Do not proceed to the next step until the run folder is completely copied, approximately 30 minutes.
5. Open the copied run folder and delete the following file:
   sample_sheet_processed_YYYY_MM_DD_hh-mm-ss.txt
6. Working in the copied run folder, edit the SampleSheet.csv file to correct the errors. Remove any empty rows or columns.
7. Save the sample sheet to the copied run folder as SampleSheet.csv to overwrite the existing file.
   Make sure that the file remains in CSV (comma separated value) format. Some spreadsheet software packages can modify the file format without warnings and overwrite commas with other symbols. Do not modify the sample sheet after you have saved it to the copied run folder.
8. To initiate analysis, rename the copied run folder as follows:
   a. Right-click the copied run folder, and click **Rename**.
   b. Replace the spaces and dash with an underscore (_). For example, Run_Folder_Name_Copy.

NOTE

Do not add characters to the front of the folder name. For example, Copy_Run_Folder_Name. Add characters only to the end of the run folder name, using only the following alphanumeric characters: a–z, A–Z, 0–9, and underscores ("_"). Spaces, dashes, and other characters are not allowed.

The system automatically analyzes Run_Folder_Name_Copy.

9 If sample_sheet_processed_YYYY_MM_DD_hh-mm-ss.txt is not created within 30 minutes, see *Requeue Analysis Troubleshooting* on page 29.

Copy a Completed Run to the Server and Queue for Analysis

This procedure describes how to manually copy a run folder to the server and queue analysis.

NOTE

Follow the procedure in the exact sequence provided below.

Steps 1-5 must be completed before copying the run folder to the analysis server.

1 Open the run folder, and move the **RTAcomplete.txt** file to a location outside of the run folder.

2 Delete the following file from the run folder:
sample_sheet_processed_YYYY_MM_DD_hh-mm-ss.txt

3 If needed, edit your original sample sheet to correct errors or make other changes. Remove any empty rows or columns.

4 Save the sample sheet to the run folder as SampleSheet.csv to overwrite the existing file.
Do not modify the sample sheet after you have saved it to the run folder.

5 Make sure that the run folder still does not contain the RTAComplete.txt file.

6 Right-click the run folder, and click **Copy**.

7 From a computer on the same network as the analysis server, open Windows Explorer, and browse to the /runs directory.

8 Right-click anywhere in the /runs directory, and click **Paste**.

NOTE

Do not proceed to the next step until the run folder is completely copied, approximately 30 minutes or longer depending on network speed.

Do not add characters to the front of the folder name. For example, Copy_Run_Folder_Name. Add characters only to the end of the run folder name, using only the following alphanumeric characters: a–z, A–Z, 0–9, and underscores ("_"). Spaces, dashes, and other characters are not allowed.

9 To initiate analysis, copy the **RTAcomplete.txt** file from the location you moved it to, and paste it in the run folder.
The system automatically reanalyzes the run folder.

10 If sample_sheet_processed_YYYY_MM_DD_hh-mm-ss.txt is not created within 30 minutes, see *Requeue Analysis Troubleshooting* on page 29.

Requeue Analysis Troubleshooting

1 Check for an error notification email.

2 Review the email for information about errors in the sample sheet.
Review the entire email message because the error pertinent to the issue may be listed at the very end of the message.

3 If the errors are ones you can correct, repeat the requeue analysis procedure that applies to your run folder.

4 Contact Illumina Technical Support if the following occurs:
- You do not receive an error notification email.
- The analysis does not run.
- The sample sheet does not contain errors

Mention NIPT16 in the call or include it in the email subject line.

Archiving and Backing Up Data

Illumina recommends archiving the /data01/runs and /data01/analysis_output directories in agreement with local IT site archiving policy. The software monitors the remaining disk space in the /data01/runs directory and notifies users by email when the remaining storage capacity goes below 200 GB.

The VeriSeq NIPT Analysis Server should not be used for data storage. Data should be transferred off the analysis server and archived on a regular schedule.

A typical sequencing run that is compatible with the cfDNA analysis workflow requires approximately 11–13 GB for NGS Option 1 and approximately 11–16 GB for NGS Option 2. The actual run folder size depends on final cluster density. The server provides more than 4 TB of storage space, which is enough space for more than 200 sequencing runs.

Only archive data when the system is idle and no analysis or sequencing runs are in progress.

Report Specifications and Metrics Interpretation

The cfDNA Sequencing analysis output folder contains 2 text files in comma-separated format. The first file, <Run_Folder_Name>_NIPT_Results.csv, contains all the sample and flow cell data and QC metrics. This file also identifies the version of the software used to generate the results. The second file, <Run_Folder_Name>_Misindexed_Results.csv, tabulates number of reads on the flow cell for the indexes identified during demultiplexing that are not specified in the sample sheet. A third .txt file, REPORT.Complete.txt, is located in the results output folder. This file contains information about the analysis configuration, analysis time, location of the output files and MD5 checksum values for the NIPT_Results.csv and MISINDEXED_Results.csv files. A full list of the QC metrics and other values can be found in *QC Metrics and Upper and Lower Boundaries (NGS Option 1)* on page 34 and *QC Metrics and Upper and Lower Boundaries (NGS Option 2)* on page 38.

CAUTION

To avoid unintentionally modifying the original analysis output, copy <Run_Folder_Name>_NIPT_Results.csv and <Run_Folder_Name>_Misindexed_Results.csv to another computer before opening and editing the files.

NOTE

Illumina recommends that the output files generated by the cfDNA analysis/VeriSeq NIPT Analysis Software be integrated into a laboratory information management system, where the information provided can then be used to generate patient reports for subsequent review by clinical laboratory staff.

Table 13 Reported Sample Sheet Annotation Values (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Sample Sheet Source Field
SampleID	Sample_ID
SampleType	SampleType
Flowcell ID	Experiment Name
IndexID	I7_Index_ID
Well	Sample_Well
Library_nM	Library_nM

Table 14 Reported Per-sample Scoring Metrics (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Interpretation
Ratio_13	Chromosomal Ratio 13
Ratio_18	Chromosomal Ratio 18
Ratio_21	Chromosomal Ratio 21
Ratio_X	Chromosomal Ratio X
Ratio_Y	Chromosomal Ratio Y
NCV_13	Normalized Chromosomal Value (z-score) 13
NCV_18	Normalized Chromosomal Value (z-score) 18
NCV_21	Normalized Chromosomal Value (z-score) 21
NCV_X	Normalized Chromosomal Value (z-score) X
NCV_Y	Normalized Chromosomal Value (z-score) Y
FF_Formatted	Estimated fetal component of cfDNA which is recovered by the assay. Reported as a discreet, rounded percentage that provides additional information for each sample.

Table 15 Reported Per-sample QC Metrics (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Interpretation	Reasons for Failure
QCFlag	Overall indicator of QC pass (0), warning (1), failure (2)	See Table 20.
QCWarning	Concatenation of all reasons for sample warning (";" separated)	See Table 20.
QCFailure	Concatenation of all reasons for sample failure (";" separated)	See Table 20.
Clusters	Total number of clusters across lanes (Reported per flow cell)	Low/high cluster density
TotalReads2Clusters	Ratio of recovered reads to number of clusters across lanes (Reported per flow cell)	Corrupted BCL files
MaxMisindexedReads2Clusters	Ratio of misindexed reads across lanes to clusters in a virtual lane (Reported per flow cell)	Reads with unexpected indexes found across lanes
IndexedReads	Total number of indexed reads per sample across lanes	Index Read technical issues; wrong samples on the sequencing lanes
TotalIndexedReads2Clusters	Ratio of indexed reads to clusters (Reported per flow cell)	Index Read technical issues
Tags	Number of reads mapped to a unique place in the genome	High PCR or sequencing error rate; bias introduced during library construction
NonExcludedSites	Number of tags excluding filtered genome regions and duplicate reads mapping to the same location	Low cluster number, sequencing errors, low library complexity, typically recoverable upon rerun
NonExcludedSites2Tags	Ratio of NonExcludedSites to tags	Library Complexity
Tags2IndexedReads	Ratio of tags to indexed reads	Higher than expected number of reads not aligning to genome
PerfectMatchTags2Tags	Ratio of perfectly mapped tags to all tags	High sequencing or PCR error rate
GCBias	Residual GC bias in the read distribution after correction	Preanalytical failure in sample collection/ handling; sequencing artifacts
GCR2	R2 of the GC correction (percentage of variance explained by GC correction)	
NCD_13	Likelihood score for chromosome 13 denominators	Unexpected profile for chr 13 denominator chromosomes
NCD_18	Likelihood score for chromosome 18 denominators	Unexpected profile for chr 18 denominator chromosomes
NCD_21	Likelihood score for chromosome 21 denominators	Unexpected profile for chr 21 denominator chromosomes
NCD_X	Likelihood score for chromosome X denominators	Unexpected profile for chr X denominator chromosomes
NCD_Y	Likelihood score for entire chromosomal profile	Unexpected profile for all chromosomes

Table 16 Reported Per-sample Scoring Metrics (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Interpretation
Chr1, ..., Chr22, ChrX, ChrY	Total number of NonExcludedSites used for analysis of a corresponding chromosome (integer value)
Chr1_Coverage, ..., Chr22_Coverage, ChrX_ Coverage, ChrY_Coverage	Normalized coverage of each chromosome used in evaluation of chromosomal ratios

Table 17 Reported Per-batch Scoring Metrics (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Interpretation
Median_13, Median_18, Median_21, Median_X, Median_Y	Batch median of chromosomal ratios for putative diploid samples Note: chrX and chrY based on putative female samples only
Stdev_13, Stdev_18, Stdev_21, Stdev_X, Stdev_Y	Batch standard deviation of chromosomal ratios for putative diploid samples Note: chrX and chrY based on putative female samples only

Table 18 Reported Per-sample, More Fields from the Sample Sheet (<Run_Folder_Name>_NIPT_Results.csv)

Column Name	Sample Sheet Source Field
SampleProject	Sample_Project
Description	Description
Index	index

Table 19 Reported Per Flow Cell Misindexed Reads (<Run_Folder_Name>_Misindexed_Results.csv)

Column Name	Interpretation
Flow Cell	Flow Cell ID
Lane	Lane ID
IndexID	Index ID Note: Index ID A000 – is any sequence except the 24 indexes found in Table 12
IndexedReads	Number of indexed reads within flow cell/lane/index

Verifying that ATMS is Running

Starting the system automatically launches the background ATMS process for monitoring sequencing and analysis runs.

To make sure that ATMS is running:

1 Execute the command to connect to the analytical server as sbsuser (assuming $HOSTNAME is the name of the server as set up during the initial install process):
ssh –l sbsuser $HOSTNAME

2 Execute the command to check the ATMS process:
ps aux | grep jsvc

If the output contains jsvc.exec, the ATMS process is running in the background. There are 3 output lines: 1) one indicating an instance running from root user, 2) one indicating an instance from the ATMS user, and 3) one indicating an instance running from whatever user the command is run under.

If the ATMS process is not running, then ATMS does not monitor or process new runs until the service is restarted. A shutdown or reboot of the machine triggers an automatic restart of the service. An Illumina Service Engineer can restart the service using root privileges on the machine.

NOTE

If unexpected shutdown occurs, the system attempts to restart the ATMS on its own.

QC Metrics

QC Metrics and Upper and Lower Boundaries (NGS Option 1)

Table 20 NGS Instrument Option 1: Two flow cell position, 2-lane flow cell—The QC Metrics, Upper and Lower Boundaries, Designation as Failure or Warning, Expected Rate of Failure/Warning and the Potential Causes.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Counting QC	Clusters	250,000,000	450,000,000	Warning		<5% flow cells	Low (more likely) or high (highly unlikely) cluster density.
Counting QC	Reads2Clusters	0.95	1	Warning		<1% flow cells	Software failed to recover more than 5% of reads recorded by the instrument.
Counting QC	MaxMisindexedReads2Clusters	0	0.0002	Warning		<0.1%	
Counting QC	TotalIndexedReads2Clusters	0.7	1	Warning		<0.1%	Indexing sequence failure.
Counting QC	NonExcludedSites	8000000	100000000	Failure		<=2%	Poor library or incorrect library quantitation; low cluster numbers; possibly recoverable upon rerun from plasma.
Counting QC	NonExcludedSites2Tags	0.8	1	Warning		<0.1%	Poor library diversity; possibly recoverable upon rerun from plasma.
Counting QC	Tags2Reads	0.75	0.9	Warning		<0.1%	High error rate in sequencing or PCR; possibly recoverable upon resequencing of the same library.
Counting QC	PerfectMatchTags2Tags	0.7	1	Warning		1%	High error rate in sequencing or PCR; possibly recoverable upon resequencing of the same library.
Median of Chromosomal Ratios	Median_13	0.1986891	0.2012977	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_18	0.2483363	0.2517526	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Median of Chromosomal Ratios	Median_21	0.2476093	0.2524342	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_X	0.3260502	0.3396256	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_Y	0	1.47E-08	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Standard Deviation of Chromosomal Ratios	Stdev_13	0	6.73E-04	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_18	0	1.37E-03	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_21	0	1.33E-03	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_X	0	3.27E-03	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Standard Deviation of Chromosomal Ratios	Stdev_Y	0	4.94E-09	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Likelihood Score for Chromosome Denominators	NCD_13	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_18	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_21	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_X	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_Y	-100	1000	Failure		<0.5%	Unexpected chromosomal representation somewhere in the genome; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
NCV of Control Samples	NCV_13	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).
NCV of Control Samples	NCV_18	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).
NCV of Control Samples	NCV_21	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
NCV of Test Samples	NCV_13	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_18	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_21	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_X	-100	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_Y	-6	2000	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
GC Bias of Control Samples	GCBias	-0.5	0.5	Warning	Control		Remaining GC bias after GC correction (expected to be centered around 0, informational only).
GC Bias of Test Samples	GCBias	-0.5	0.5	Warning	Test		Remaining GC bias after GC correction (expected to be centered around 0, informational only).
GC R2 of Control Samples	GC R2	0	0.9999	Warning	Control		R^2 associated with the GC correction (informational only).
GC R2 of Test Samples	GC R2	0	0.9999	Warning	Test		R^2 associated with the GC correction (informational only).

QC Metrics and Upper and Lower Boundaries (NGS Option 2)

Table 21 NGS Instrument Option 2: Single flow cell position, 4-lane flow cell—The QC Metrics, Upper and Lower Boundaries, Designation as Failure or Warning, Expected Rate of Failure/Warning and the Potential Causes.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Counting QC	Clusters	300,000,000	800,000,000	Warning		<5% flow cells	Low (more likely) or high (highly unlikely) cluster density.
Counting QC	MaxMisindexedReads2Clusters	0	0.0002	Warning		<0.1%	
Counting QC	TotalIndexedReads2Clusters	0.7	1	Warning		<0.1%	Indexing sequence failure.
Counting QC	NonExcludedSites	8000000	100000000	Failure		<=2%	Poor library or incorrect library quantitation; low cluster numbers; possibly recoverable upon rerun from plasma.
Counting QC	NonExcludedSites2Tags	0.8	1	Warning		<0.1%	Poor library diversity; possibly recoverable upon rerun from plasma.
Counting QC	Tags2Reads	0.75	0.9	Warning		<0.1%	High error rate in sequencing or PCR; possibly recoverable upon resequencing of the same library.
Counting QC	PerfectMatchTags2Tags	0.7	1	Warning		1%	High error rate in sequencing or PCR; possibly recoverable upon resequencing of the same library.
Median of Chromosomal Ratios	Median_13	0.1991238	0.2008629	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_18	0.2489057	0.2511832	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Median of Chromosomal Ratios	Median_21	0.2484135	0.25163	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_X	0.329444	0.3362317	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Median of Chromosomal Ratios	Median_Y	0	1.236665e-08	Warning		<0.1%	Unexpectedly high / low median chromosomal ratio across the entire flow cell; strong uncorrected batch effect associated with either extraction or library batch. Try reprocessing samples from plasma to resolve the problem.
Standard Deviation of Chromosomal Ratios	Stdev_13	0	0.0008695377	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_18	0	0.00113876	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_21	0	0.001608292	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Standard Deviation of Chromosomal Ratios	Stdev_X	0	0.005090769	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
Standard Deviation of Chromosomal Ratios	Stdev_Y	0	3.454837e-09	Warning		<0.1%	Unexpectedly high standard deviation of chromosomal ratios, which signifies extra sources of previously unseen variance; watch for trend over time.
Likelihood Score for Chromosome Denominators	NCD_13	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_18	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_21	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_X	-50	1000	Failure		<0.1%	Unexpected chromosomal representation of denominator (reference) chromosomes; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
Likelihood Score for Chromosome Denominators	NCD_Y	-100	1000	Failure		<0.5%	Unexpected chromosomal representation somewhere in the genome; unlikely to get resolved by rerunning the sample; suggest "data outside of expected range."
NCV of Control Samples	NCV_13	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).
NCV of Control Samples	NCV_18	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).
NCV of Control Samples	NCV_21	-5	4	Warning	Control		NCV bounds for controls (no monosomy, no trisomy).

Category	Metric	Lower Bound	Upper Bound	Failure/ Warning	Sample Type	Expected failure/ warning rate	Potential Causes
NCV of Test Samples	NCV_13	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_18	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_21	-5	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_X	-100	200	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
NCV of Test Samples	NCV_Y	-6	2000	Warning	Test		NCV bounds for test samples (no monosomy, fetal fraction within (generous) expected range).
GC Bias of Control Samples	GCBias	-0.5	0.5	Warning	Control		Remaining GC bias after GC correction (expected to be centered around 0, informational only).
GC Bias of Test Samples	GCBias	-0.5	0.5	Warning	Test		Remaining GC bias after GC correction (expected to be centered around 0, informational only).
GC R2 of Control Samples	GC R2	0	0.9999	Warning	Control		R^2 associated with the GC correction (informational only).
GC R2 of Test Samples	GC R2	0	0.9999	Warning	Test		R^2 associated with the GC correction (informational only).

Method Comparison Study

Method Comparison Data

For this study, previously prepared libraries of 105 plasma samples were resequenced and processed with the VeriSeq NIPT Analysis Software (16 Samples). These samples were previously run on the Verifi® test and were multiplexed into 7 libraries each consisting of 14 maternal plasma samples, 1 pooled positive control maternal sample, and 1 non-template control or NTC. Table 22 shows the sample composition of each library.

All 98 individual non-control samples passed QC and were analyzed for concordance with Verifi results. Each sample was classified based on the NCV values for trisomy 13 / 18 / 21 (using a threshold of NCV = 4), for presence of chromosome Y (using a threshold of NCV = 10), and for monosomy X (using a threshold of NCV_X = -4 and chromosome Y not present). The overall percent agreement between VeriSeq NIPT and Verifi is shown in Table 23.

There were two observed discrepancies. The first observed discrepancy was for Chromosome 13 which was classified as trisomy 13 by the Verifi test and classified as negative by Veriseq NIPT Analysis Software (16 Samples). Clinical information for this sample was later provided as negative for Trisomy 13. Another observed discrepancy was for Trisomy 18 and no clinical outcome information was available for this sample.

Table 22 Distribution of Samples Across Libraries

Library	Control	MX	T13	T18	T21	Unaffected
01	1				2	12
02	1			1	1	12
03	1	1			1	12
04	1		1	1	1	11
05	1	1			1	12
06	1		1		1	12
07	1				1	13
Total	7	2	2	2	8	84

Table 23 Overall Percent Agreement Between VeriSeq NIPT and Verifi

	Overall Agreement
Class 13	98.98%
Class 18	98.98%
Class 21	100%
ChrY Present/Absent	100%
Class Monosomy X	100%

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	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	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan	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 in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.

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 IN VITRO DIAGNOSTIC USE

CE 2797

IVD

EC REP

Illumina Netherlands B.V.
Freddy van Riemsdijkweg 15
5657 EE Eindhoven
The Netherlands

Australian Sponsor
Illumina Australia Pty Ltd
Nursing Association Building
Level 3, 535 Elizabeth Street
Melbourne, VIC 3000
Australia

illumina®