

Local Run Manager Germline Variant Analysis Module

Workflow Guide for MiSeqDx

FOR IN VITRO DIAGNOSTIC USE

Overview	3
Enter Run Information	3
Analysis Methods	5
View Run and Sample Data	6
Analysis Report	6
Analysis Output Files	7
Revision History	13
Technical Assistance	14



This document and its contents are proprietary to Illumina, Inc. and its affiliates ("Illumina"), and are intended solely for the contractual use of its customer in connection with the use of the product(s) described herein and for no other purpose. This document and its contents shall not be used or distributed for any other purpose and/or otherwise communicated, disclosed, or reproduced in any way whatsoever without the prior written consent of Illumina. Illumina does not convey any license under its patent, trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

The instructions in this document must be strictly and explicitly followed by qualified and properly trained personnel in order to ensure the proper and safe use of the product(s) described herein. All of the contents of this document must be fully read and understood prior to using such product(s).

FAILURE TO COMPLETELY READ AND EXPLICITLY FOLLOW ALL OF THE INSTRUCTIONS CONTAINED HEREIN MAY RESULT IN DAMAGE TO THE PRODUCT(S), INJURY TO PERSONS, INCLUDING TO USERS OR OTHERS, AND DAMAGE TO OTHER PROPERTY.

ILLUMINA DOES NOT ASSUME ANY LIABILITY ARISING OUT OF THE IMPROPER USE OF THE PRODUCT(S) DESCRIBED HEREIN (INCLUDING PARTS THEREOF OR SOFTWARE).

© 2021 Illumina, Inc. All rights reserved.

Illumina, MiSeqDx, and the streaming bases design are registered or pending trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Overview

The Local Run Manager Germline Variant analysis module is for use with the Illumina TruSeq Custom Amplicon Kit Dx assay. When used with the Germline Variant module, the assay is intended for preparing libraries used for the sequencing of DNA from peripheral whole blood specimens.

The analysis module evaluates short regions of amplified DNA, or amplicons, for variants. Focused sequencing of amplicons enables high coverage of particular regions. See the package insert *TruSeq Custom Amplicon Kit Dx* (document # 1000000029772).

The Germline Variant analysis module requires the MiSeqDx Reagent Kit v3 sequencing consumables. See the package insert *MiSeqDx Reagent Kit v3* (document # 1000000030849).

About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis for the Germline Variant analysis module. For information about the Local Run Manager dashboard and system settings, see the *Local Run Manager Software Reference Guide for MiSeqDx* (document # 1000000011880).

Viewing Local Run Manager

Enter Run Information

Set Parameters

- 1 Log in to Local Run Manager.
- 2 Click **Create Run**, and select **Germline Variant**.
- 3 Enter a run name that identifies the run from sequencing through analysis.
Use alphanumeric characters, spaces, underscores, or dashes.
- 4 [Optional] Enter a run description to help identify the run.
Use alphanumeric characters, spaces, underscores, or dashes.

Import Manifest Files for the Run

- 1 Make sure that the manifests you want to import are available in an accessible network location or on a USB drive.
- 2 Click **Import Manifests**.
- 3 Navigate to the manifest file and select the manifests that you want to add.



NOTE

To make manifest files available for all runs using the Germline Variant analysis module, add manifests using the Module Settings feature. This feature requires admin user level permissions. For more information, see the *Local Run Manager Software Reference Guide for MiSeqDx* (document # 1000000011880).

Specify Samples for the Run


Specify samples for the run using one of the options and the directions that follow.

- **Enter samples manually**—Use the blank table on the Create Run screen.


- **Import samples**—Navigate to an external file in a comma-separated values (*.csv) format. A template is available for download on the Create Run screen.

After you have populated the samples table, you can export the sample information to an external file, and use the file as a reference when preparing libraries or import the file for another run.

Enter Samples Manually

- 1 Select the number of samples and index set from the drop-down list.
- 2 Enter a unique sample ID in the Sample ID field.
Use alphanumeric characters, dashes, or underscores.
- 3 [Optional] For positive or negative control samples, right-click and select the control type.
- 4 [Optional] Enter a sample description in the Sample Description field.
Use alphanumeric characters, dashes, or underscores.
Sample descriptions are associated with a sample name. Sample descriptions are overwritten if the same sample name is used again in a later run.
- 5 Select an Index 1 adapter from the Index 1 (i7) drop-down list.
- 6 Select an Index 2 adapter from the Index 2 (i5) drop-down list.
- 7 Select a manifest file from the Manifest drop-down list.
- 8 Choose an option to view, print, or save the plate layout as a reference for preparing libraries:
 - Click the  **Print** icon to display the plate layout. Select **Print** to print the plate layout.
 - Click **Export** to export sample information to an external file.
 Make sure that manifest and sample information is correct. Incorrect information can impact results.
- 9 Click **Save Run**.

Import Samples

- 1 Click **Import Samples** and browse to the location of the sample information file. There are two types of files you can import.
 - Click **Template** on the Create Run screen to make a new plate layout. The template file contains the correct column headings for import. Enter sample information in each column for the samples in the run. Delete example information in unused cells, and then save the file.
 - Use a file of sample information that was exported from the Germline Variant module using the Export feature.
- 2 Click the  **Print** icon to display the plate layout.
- 3 Select **Print** to print the plate layout as a reference for preparing libraries.
- 4 [Optional] Click **Export** to export sample information to an external file.
Make sure that manifest and sample information is correct. Incorrect information can impact results.
- 5 Click **Save Run**.

Edit a Run

For instructions on editing the information in your run before sequencing, see *Local Run Manager Software Reference Guide for MiSeqDx (document # 1000000011880)*.

Analysis Methods

The Germline Variant analysis module performs the following analysis steps and then writes analysis output files to the Alignment folder.

- ▶ Demultiplexes index reads
- ▶ Generates FASTQ files
- ▶ Aligns to a reference
- ▶ Identifies variants

Demultiplexing

Demultiplexing compares each Index Read sequence to the index sequences specified for the run. No quality values are considered in this step.

Index reads are identified using the following steps:

- ▶ Samples are numbered starting from 1 based on the order they are listed for the run.
- ▶ Sample number 0 is reserved for clusters that were not assigned to a sample.
- ▶ Clusters are assigned to a sample when the index sequence matches exactly or when there is up to a single mismatch per Index Read.

FASTQ File Generation

After demultiplexing, the software generates intermediate analysis files in the FASTQ format, which is a text format used to represent sequences. FASTQ files contain reads for each sample and the associated quality scores. Clusters that did not pass filter are excluded.

Each FASTQ file contains reads for only one sample, and the name of that sample is included in the FASTQ file name. FASTQ files are the primary input for alignment. Two FASTQ files are generated per sample, one from Read 1, and one from Read 2.

Alignment

During the alignment step, the banded Smith-Waterman algorithm aligns clusters from each sample against amplicon sequences specified in the manifest file.

The banded Smith-Waterman algorithm performs semi-global sequence alignments to determine similar regions between 2 sequences. Instead of comparing the total sequence, the Smith-Waterman algorithm compares segments of all possible lengths.

Each paired-end read is evaluated in terms of its alignment to the relevant probe sequences for that read.

- ▶ Read 1 is evaluated against the reverse complement of the Downstream Locus-Specific Oligos (DLSO).
- ▶ Read 2 is evaluated against the Upstream Locus-Specific Oligos (ULSO).
- ▶ If the start of a read matches a probe sequence with no more than 3 differences (mismatches or shifts due to leading indels), the full length of the read is aligned against the amplicon target for that sequence.
- ▶ Indels within the DLSO and ULSO are not observed given the assay chemistry.

Alignments are filtered from alignment results based on mismatch rates over either the region of interest or the full amplicon, depending on the amplicon length. Filtered alignments are written in alignment files as unaligned and are not used in variant calling.

Variant Calling

Developed by Illumina, the Pisces Variant Caller identifies variants present in the DNA sample.



The Pisces variant caller identifies SNVs, MNVs, and small indels in 3 steps:

- ▶ Considers each position in the reference genome separately
- ▶ Counts bases at the given position for aligned reads that overlap the position
- ▶ Computes a variant score that measures the quality of the call using Poisson model. Variants with a quality score below Q20 are excluded.

If a variant passes all filters, the variant is marked as PASS in VCF.

For more information, see github.com/Illumina/Pisces/wiki.

View Run and Sample Data

- 1 From the Local Run Manager dashboard, click the run name.
- 2 From the Run Overview tab, review the sequencing run metrics.
- 3 [Optional] Click the **Copy to Clipboard**  icon to copy the output run folder path.
- 4 Click the Sequencing Information tab to review run parameters and consumables information.
- 5 Click the Samples and Results tab to view the analysis report location.
 - ▶ If analysis was repeated, expand the Select Analysis drop-down list and select the appropriate analysis.
- 6 Click the **Copy to Clipboard**  icon to copy the Analysis folder path.

For more information on the Run Overview and Sequencing Information tabs, and how to requeue analysis, see the *Local Run Manager Software Reference Guide for MiSeqDx (document # 1000000011880)*.

Analysis Report

Analysis results are summarized on the Samples and Results tab, and as an aggregate report in the Alignment folder. A report for each sample is also available in a PDF file format for each sample.

Sample and Details Tab Information

- 1 Click a sample in the list to see the sample report.

Table 1 Run Information Table

Column Heading	Description
Run Status	States whether the sequencing run passed or failed.
Total Yield (GB)	Number of bases called in the sequencing run. Shows passing threshold, and pass or fail status.
% ≥ Q30	The percentage of reads in the sequencing run with a quality score of 30(Q30) or greater. Shows passing threshold, and pass or fail status.
Sample Name	The sample name provided when the run was created.
Total PF Reads	The total number of reads passing filter.
Read 1% ≥ Q30	The percentage of reads in Read 1 with a quality score of 30(Q30) or greater for the sample.

Column Heading	Description
Read 2% \geq Q30	The percentage of reads in Read 2 with a quality score of 30(Q30) or greater for the sample.
Autosome call rate	The number of genomic positions across the autosomes (chromosomes 1 through 22) that meet a predefined confidence value threshold, divided by the total number of autosomal genomic positions interrogated. Call rate is described on a per-sample basis and reported as percentage that is calculated as 1 minus (number of autosomal positions with incomplete calls divided by the total number of autosomal positions sequenced).

Table 2 Sample Reports Table

Column Heading	Description
Sample	The sample ID provided when the run was created.
Report Date	Date the report was generated.
Sample Information	The sample ID that was provided when the run was created, total reads that passed filter in the sample, the percentage of reads for the sample with a quality score of 30(Q30) or greater, and the autosomal call rate.
Amplicon Summary	Total number of amplicon regions sequenced, and the total length in base pairs of sequenced amplicons in the target regions, for the sample and the manifest file. The manifest file specifies the reference genome and targeted reference regions used in the alignment step.
Read Level Statistics	Number and percent of reads for the sample that cover each position in the reference, for Read 1 and Read 2.
Variants Summary	Number of SNVs, insertions, and deletions detected for the sample that passed suggested values to determine if quality results are within an acceptable range,
Coverage Summary	The total number of aligned bases divided by the targeted region size, and the percentage of amplicon regions with coverage values greater than the low coverage threshold of $0.2 \times$ amplicon mean coverage, for the sample.
Coverage Plots	The Coverage by Amplicon Region plot show the coverage across amplicon regions for the sample. Regions with coverage values lower than the coverage threshold are highlighted in red. The average of all values is indicated by an orange line.
Software Versions	Software versions when the sample was sequenced. Includes MiSeq Operating Software (MOS), Local Run Manager Software, RTA Software, and the Germline Variant Module version.

Analysis Output Files

The following analysis output files are generated for the Germline Variant analysis module and provide analysis results for alignment and variant calling. Analysis output files are located in the Alignment folder.

File Name	Description
Demultiplexing (*.txt)	Intermediate files containing demultiplexing summary results.
FASTQ (*.fastq.gz)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step.
Alignment files in the BAM format (*.bam)	Contains aligned reads for a given sample.

File Name	Description
Variant call files in the genome VCF format (*.genome.vcf.gz)	Contains the genotype for each position, whether called as a variant or called as a reference.
Variant call files in the VCF format (*.vcf.gz)	Contains all variants called across the target region.
AmpliconCoverage_M1.tsv	Contains information about coverage per amplicon per sample for each manifest provided. M# represents the manifest number.

Demultiplexing File Format

The process of demultiplexing reads the index sequence attached to each cluster to determine from which sample the cluster originated. The mapping between clusters and sample number is written to a demultiplexing (*.demux) file for each tile of the flow cell.

The demultiplexing file naming format is s_1_X.demux, where X is the tile number.

Demultiplexing files start with a header:

- ▶ Version (4 byte integer), currently 1
- ▶ Cluster count (4 byte integer)

The remainder of the file consists of sample numbers for each cluster from the tile.

When the demultiplexing step is complete, the software generates a demultiplexing file named DemultiplexSummaryF1L1.txt.

- ▶ In the file name, **F1** represents the flow cell number.
- ▶ In the file name, **L1** represents the lane number.
- ▶ Demultiplexing results in a table with 1 row per tile and 1 column per sample, including sample 0.
- ▶ The most commonly occurring sequences in index reads.

FASTQ File Format

FASTQ is a text-based file format that contains base calls and quality values per read. Each record contains 4 lines:

- ▶ The identifier
- ▶ The sequence
- ▶ A plus sign (+)
- ▶ The Phred quality scores in an ASCII + 33 encoded format

The identifier is formatted as:

@Instrument:RunID:FlowCellID:Lane:Tile:X:Y ReadNum:FilterFlag:0:SampleNumber

Example:

```
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAAA9#:<#<;<<<????#=#
```

BAM File Format

A BAM file (*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences up to 128 Mb. SAM and BAM formats are described in detail at samtools.github.io/hts-specs/SAMv1.pdf.

BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed for the run.

BAM files contain a header section and an alignment section:

- ▶ **Header**—Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- ▶ **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and the match descriptor string.

The alignments section includes the following information for each read or read pair:

- ▶ **AS:** Paired-end alignment quality
- ▶ **RG:** Read group, which indicates the number of reads for a specific sample.
- ▶ **BC:** Barcode tag, which indicates the demultiplexed sample ID associated with the read.
- ▶ **SM:** Single-end alignment quality.
- ▶ **XC:** Match descriptor string
- ▶ **XN:** Amplicon name tag, which records the amplicon ID associated with the read

BAM index files (*.bam.bai) provide an index of the corresponding BAM file.

VCF File Format

Variant Call Format (VCF) is a common file format developed by the genomics scientific community. It contains information about variants found at specific positions in a reference genome. VCF files end with the .vcf suffix

The VCF file header includes the VCF file format version and the variant caller version and lists the annotations used in the remainder of the file. The VCF header also includes the reference genome file and BAM file. The last line in the header contains the column headings for the data lines. Each of the VCF file data lines contains information about one variant.

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTQ file.
POS	The single-base position of the variant in the reference chromosome. For SNPs, this position is the reference base with the variant; for indels or deletions, this position is the reference base immediately before the variant.
ID	The rs number for the variant obtained from dbSNP.txt, if applicable. If there are multiple rs numbers at this location, the list is semi-colon delimited. If no dbSNP entry exists at this position, a missing value marker ('.') is used.
REF	The reference genotype. For example, a deletion of a single T is represented as reference TT and alternate T. An A to T single nucleotide variant is represented as reference A and alternate T.

Heading	Description
ALT	The alleles that differ from the reference read. For example, an insertion of a single T is represented as reference A and alternate AT. An A to T single nucleotide variant is represented as reference A and alternate T.
QUAL	A Phred-scaled quality score assigned by the variant caller. Higher scores indicate higher confidence in the variant and lower probability of errors. For a quality score of Q, the estimated probability of an error is $10^{-(Q/10)}$. For example, the set of Q30 calls has a 0.1% error rate. Many variant callers assign quality scores based on their statistical models, which are high in relation to the error rate observed.

VCF File Annotations

Heading	Description
FILTER	If all filters are passed, PASS is written in the filter column. <ul style="list-style-type: none"> • LowDP— Applied to sites with depth of coverage below 150x. For amplicon positions covered by both the forward and reverse read, this is equivalent to 300x single read coverage. • q20—Quality score < 20. • MultiAllelicSite— Variant does not conform to diploid model. • R5x9— The number of adjacent repeats (of length 1 to 5 bp) to the variant calls ≥ 9. • SB— The strand bias is more than the given threshold.
INFO	Possible entries in the INFO column include: <ul style="list-style-type: none"> • AC— Allele count in genotypes for each ALT allele, in the same order as listed. • AF— Allele Frequency for each ALT allele, in the same order as listed. • AN— The total number of alleles in called genotypes. • CD— A flag indicating that the SNP occurs within the coding region of at least 1 RefGene entry. • DP— The depth (number of base calls aligned to a position and used in variant calling). • Exon— A comma-separated list of exon regions read from RefGene. • FC— Functional Consequence. • GI— A comma-separated list of gene IDs read from RefGene. • QD— Variant Confidence/Quality by Depth. • TI— A comma-separated list of transcript IDs read from RefGene.
FORMAT	The format column lists fields separated by colons. For example, GT:GQ. The list of fields provided depends on the variant caller used. Available fields include: <ul style="list-style-type: none"> • AD— Entry of the form X,Y, where X is the number of reference calls, and Y is the number of alternate calls. • DP— Approximate read depth; reads with MQ=255 or with bad mates are filtered. • GQ— Genotype quality. • GT— Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available. • NC— Fraction of bases which were uncalled, or with base call quality below the minimum threshold. • NL— Noise level; an estimate of base calling noise at this position. • SB— Strand bias at this position. Larger negative values indicate less bias; values near 0 indicate more bias. • VF— Variant frequency; the percentage of reads supporting the alternate allele.
SAMPLE	The sample column gives the values specified in the FORMAT column.

Genome VCF Files

Genome VCF (gVCF) files are VCF v4.1 files that follow a set of conventions for representing all sites within the genome in a reasonably compact format. The gVCF (*.genome.vcf.gz) files include all sites within the region of interest in a single file for each sample.

The gVCF file shows no-calls at positions that do not pass all filters. A genotype (GT) tag of `./.` indicates a no-call.

For more information, see sites.google.com/site/gvcftools/home/about-gvcf.

Amplicon Coverage File

An amplicon coverage file is generated for each manifest file. The M# in the file name represents the manifest number as it is listed in the samples table for the run.

Each file includes a header row that contains the sample IDs associated with the manifest. Under the header row are 3 columns that list the following information:

- ▶ The Target ID as it is listed in the manifest.
- ▶ The coverage depth of reads passing filter.
- ▶ The total coverage depth.

Supplementary Output Files

The following output files provide supplementary information, or summarize run results and analysis errors. Although these files are not required for assessing analysis results, they can be used for troubleshooting purposes. All files are located in the Alignment folder unless otherwise specified.

File Name	Description
AnalysisLog.txt	Processing log that describes every step that occurred during analysis of the current run folder. This file does not contain error messages. Located in the Alignment folder.
AnalysisError.txt	Processing log that lists any errors that occurred during analysis. This file will be empty if no errors occurred. Located in the Alignment folder.
DemultiplexSummaryF1L1#.txt	Reports demultiplexing results in a table with 1 row per tile and 1 column per sample. The # represents lane 1, 2, 3, or 4 of the flow cell. Located in the Alignment folder.
AmpliconRunStatistics.xml	Contains summary statistics specific to the run. Located in the Alignment folder.

Analysis Folder

The analysis folder holds the files generated by the Local Run Manager software.

The relationship between the output folder and analysis folder is summarized as follows:

- ▶ During sequencing, Real-Time Analysis (RTA) populates the output folder with files generated during image analysis, base calling, and quality scoring.
- ▶ RTA copies files to the analysis folder in real time. After RTA assigns a quality score to each base for each cycle, the software writes the file RTAComplete.txt to both folders.
- ▶ When the file RTAComplete.txt is present, analysis begins.
- ▶ As analysis continues, Local Run Manager writes output files to the analysis folder, and then copies the files back to the output folder.

Alignment Folders

Each time that analysis is requeued, the Local Run Manager creates an Alignment folder named **Alignment_N**, where N is a sequential number.

Folder Structure

 **Alignment**—Contains *.bam, *.vcf, FASTQ files, and files specific to the analysis module.

 **Date and Time Stamp**—Date_time stamp of analysis as YYYYMMDD_HHMMSS

 AnalysisError.txt

 AnalysisLog.txt

 AmpliconRunStatistics.xml

 Sample1.genome.vcf.gz

 Sample1.coverage.csv

 Sample1.report.pdf

 Sample1.summary.csv

 Sample1.vcf.gz

 Sample1.bam

 **FASTQ**

 **Sample1**

 **Stats**

 DemuxSummaryF1L1.txt

 FastqSummaryF1L1.txt

 **Data**

 **Intensities**

 **BaseCalls**

 **L001**—Contains one subfolder per cycle, each containing *.bcl files.

 **L001**—Contains *.locs files, 1 for each tile.

 **RTA Logs**—Contains log files from RTA software analysis.

 **InterOp**—Contains binary files used to report sequencing run metrics.

 **Logs**—Contains log files describing steps performed during sequencing.

 RTAComplete.txt

 RunInfo.xml

 runParameters.xml

Revision History

Document	Date	Description of Change
Document # 1000000012180 v01	August 2021	Updated EU Authorized Representative address. Added Revision History table.

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	400.635.9898	
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	

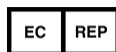
Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

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

CE



Illumina Netherlands B.V.
Steenoven 19
5626 DK Eindhoven
The Netherlands

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

FOR IN VITRO DIAGNOSTIC USE

© 2021 Illumina, Inc. All rights reserved.

illumina®