

DRAGEN Enrichment for MiSeq i100 v1.1.7

Table of Contents

1	App Highlights	3
1.1	DRAGEN core DNA Pipeline info on Gitbook	3
2	Pipeline Configuration	4
3	Modes of Operation	6
3.1	Cloud and Local Execution:	6
3.2	Requeues:	6
4	Output Files	7
4.1	Summary	7
4.2	Toplevel DRAGEN	7
4.3	Workflow Level Output	8
5	Reference Genome Usage Guidelines	9
5.1	Germline, Enrichment Germline Workflows	9
5.2	Enrichment Somatic, Somatic, and RNA Workflows	9
5.3	Reference Genome Version Compatibility with DRAGEN Versions	9
6	Available Resource Files	10
6.1	Available Reference Genomes:	10
6.2	Available Non-Reference Genome Resource Files:	11
7	How To Generate Custom Resource Files	12
7.1	Reference Genomes:	12
7.2	Noise Baseline and Panel of Normals Files:	12
8	How To Import Resource Files	13
8.1	Import Reference Files	13
9	How To Install DRAGEN and DRAGEN Applications	14
9.1	Install DRAGEN Versions:	14
9.2	Uninstall DRAGEN Versions:	14
9.3	Application Installation:	14
9.4	View Application Settings:	15
9.5	Uninstall Applications	15
	Release History	16

1 App Highlights

The DRAGEN Enrichment pipeline supports germline or somatic (tumor only) DNA samples and includes the following functional components:

- Decompression of input BCL data
- FASTQ conversion
- FASTQ compression to ORA or Gzip formats
- Mapping & Alignment (includes sorting and duplicate marking)
- BAM/CRAM Compression (optional)
- Variant Calling

The pipeline supports, via the `VariantCallingMode` parameter, algorithms for the following variant callers:

- None, or
- Small (SNV, with gVCF output for germline mode), or
- All variant callers:
 - Small
 - Structural
 - Copy Number

The following inputs are required:

- BCL Data generated from the MiSeq i100 instrument
- Sample Sheet
 - With `GermlineOrSomatic` setting
- BedFile (if Variant Calling Mode is not equal to None)

The following inputs are optional:

- `AuxBaselineNoiseFile`
- `AuxCnvPanelOfNormalsFile` (if Variant Calling Mode equals `AllVariantCallers`)
- `AuxCnvPopBAlleleVcfFile` (if Variant Calling Mode equals `AllVariantCallers` and `GermlineOrSomatic=somatic`)
- `AuxGermlineTaggingFile` (if Variant Calling Mode equals `AllVariantCallers` and `GermlineOrSomatic=somatic`)

1.1 DRAGEN core DNA Pipeline info on Gitbook

More information about the DNA pipeline can be found in core DRAGEN documentation:

<https://help.dragen.illumina.com/product-guide/dragen-v4.4/dragen-dna-pipeline>

2 Pipeline Configuration

DragenEnrichment_Settings

Parameter	Required	Description
SoftwareVersion	Yes	The version of the DRAGEN software used to process the DragenEnrichment pipeline, including conversion to FASTQ, specified using all three integers included in the DRAGEN version name. For example, 4.1.5.
AppVersion	Yes	The version of the workflow-specific application (i.e., DRAGEN Enrichment), using all three integers included in the version name. For example, 1.0.0.
KeepFastQ	Yes	Select whether FASTQs are saved (true) or discarded (false).
MapAlignOutFormat	Yes	Formatting of the output files. Accepted values are bam, cram, or none. Selecting none produces no map/align output. If MapAlignOutFormat is None, VariantCallingMode cannot be None for any sample.

DragenEnrichment_Data

Parameter	Required	Description
ReferenceGenomeDir	Yes	Genome name as an alphanumeric string with "_", "-", or ".". Can be placed in either settings or data section. For example, hg38-alt_masked.cnv.graph.hla.rna-10-r4.0-2.
Bedfile	Conditionally required	BED file to be used for analysis in text (*.txt) or gzip (*.gz) format. Only required if VariantCallingMode is not None. Must include the prefix DragenEnrichment/ before the BED file name. Alphanumeric string with "_", "-", or "." with no spaces allowed. The value must be na if BedFile is placed in the Data section and VariantCallingMode = None.
GermlineOrSomatic	Yes	Accepted values are germline or somatic.
AuxNoiseBaselineFile	No	Alphanumeric string with "_", "-", or "." and no spaces in text (*.txt) or gzip (*.gz) format. Applicable only if GermlineOrSomatic is Somatic and VariantCallingMode is not None. The value must be na if AuxNoiseBaselineFile is placed in the Data section and no AuxNoiseBaselineFile is provided for the sample."
AuxCnvPanelOfNormalsFile	No	Alphanumeric string with "_", "-", or "." and no spaces in text (*.txt) or gzip (*.gz) format. Optional if VariantCallingMode is AllVariantCallers. The value must be na if AuxCnvPanelOfNormalsFile is placed in the Data section and no AuxCnvPanelOfNormalsFile is provided for the sample.
VariantCallingMode	Yes	Variant calling mode for the run.

		Accepted values are None, SmallVariantCaller, AllVariantCallers. The option for all variant callers includes Small, Structural, and CNV callers (if panel of normals is provided).
AuxCnvPopBAlleleVcfFile	No	Alphanumeric string with "_", "-", or "." with no spaces allowed. Text or gzip format. Optional only if VariantCallingMode = AllVariantCallers and GermlineOrSomatic=somatic. Error otherwise. A note should be added to the UI to indicate that CNV output will only be generated if this file is provided.
AuxGermlineTaggingFile	No	Alphanumeric string with "_", "-", or "." with no spaces allowed. Binary file with .bin extension. Optional when SmallVariantCaller or AllVariantCaller option selected and GermlineOrSomatic=somatic. Error otherwise.
Sample_ID	Yes	See description in the BCL Convert section.

3 Modes of Operation

3.1 Cloud and Local Execution:

DRAGEN can be scheduled to operate in the following modes:

- Cloud:
 - Run planning:
 - Occurs in the Cloud via the Cloud User Interface (UI), or alternatively, sample sheets can be imported into the Cloud UI
 - Sequencing execution:
 - Runs are identified on the sequencer, and are executed locally on the instrument
 - DRAGEN Execution:
 - Occurs in the Cloud automatically after sequencing is complete
- Local:
 - Run planning:
 - Occurs on the instrument via the Instrument UI, or alternatively, sample sheets can be imported into the Instrument UI
 - Sequencing execution:
 - Runs are executed locally on the instrument
 - DRAGEN Execution:
 - Occurs locally on the instrument, automatically after sequencing is complete

3.2 Requeues:

After a sequencing run is complete, it may be desirable to run or re-run DRAGEN with, for example, an updated configuration. Requeues can occur on the instrument or in the Cloud. For on-instrument requeues, the following pre-requisites are necessary in Local Mode:

- The BCL's must be stored on the instrument
- The system must be idle

4 Output Files

4.1 Summary

- Aligned reads - BAM/CRAM format
- Variants - VCF format
- Optional: Structural Variants - VCF
- Optional: CNV - GFF3, VCF
- Gene counts - BED format
- Reports - HTML

4.2 Toplevel DRAGEN

<run_id>/Analysis/<no>/inputs

- <SampleSheet>.csv

<run_id>/Analysis/<no>/Data

- Secondary_Analysis_Complete.txt
- summary
 - <x.y.z> (Note: DRAGEN version)
 - highlevel_summary.json
 - detailed_summary.json
- AggregateReports
 - report.html
 - report_files
 - *Links to workflow level reports*
- Demux
 - AggregateReports
 - *Links to lower-level reports*
 - Demultiplex_Stats.csv
 - *other stats*
- Dragen<Workflow>
 - Data (including FASTQ) for samples configured for the workflow (see below)
- RunInstrumentAnalyticsMetrics
- logs
 - *high-level logs*

4.3 Workflow Level Output

<run_id>/Analysis/<no>/Data/DragenEnrichment

- Dragen<workflow>
 - AggregateReports
 - *Links to workflow specific sample level reports*
 - fastq (or ora_fastq)
 - <sample_ID>.S0_Rm_001.fastq.gz or *.fastq.ora (m=1-2)
 - *Additional samples*
 - Reports
 - *Adapter_Metrics.csv*
 - *Quality_Metrics.csv*
 - *Additional metrics files*
 - <sample_ID>
 - <workflow>_seq
 - <workflow>-specific logs
 - logs
 - *Sample specific logs*
 - logs
 - *Workflow specific logs*

5 Reference Genome Usage Guidelines

DRAGEN supports two types of genome hash tables: linear and pangenome. Pangenome reference genomes extend the reference genomes with alternative variant paths from a sample cohort used to construct the pangenome reference.

5.1 Germline, Enrichment Germline Workflows

- Use pangenome reference genomes.
- Use the hg38 assembly. The recommended reference genome is [Homo sapiens \[1000 Genomes\] hg38 v5 Multigenome](#)
- If not using hg38, the [Homo sapiens \[UCSC\] hg19 v5 Multigenome](#) or [Homo sapiens \[NCBI\] hs37d5 v5 Multigenome](#) are recommended.
- Linear reference is supported, but has reduced accuracy

5.2 Enrichment Somatic, Somatic, and RNA Workflows

- Use linear reference genomes.
- Use the [Homo sapiens \[1000 Genomes\] hg38 v5](#) reference genome or [Homo sapiens \[NCBI\] hs37d5 v5](#) reference genome.

5.3 Reference Genome Version Compatibility with DRAGEN Versions

	HT10	HT11
4.3.13		
4.4.6		

6 Available Resource Files

6.1 Available Reference Genomes:

A reference genome is an input to a DRAGEN pipeline. Reference genomes are available at:

- https://support.illumina.com/sequencing/sequencing_instruments/miseq-i100-plus/product-files.html

and are summarized below.

DRAGEN Reference Genomes Provided with the Instrument (DRAGEN 4.4.6)

Directory Name	Display Name (Species [Source] Assembly)	DN A	RN A	GT F	CN V	Graph	Met h
hg19-alt_masked.cnv.hla.methyl_cg.methylated_combined.ma-11-r5.0-2	Homo sapiens [UCSC] hg19 v5	Yes	Yes	Yes	Yes	No	No
hg38-alt_masked.cnv.hla.methyl_cg.methylated_combined.ma-11-r5.0-2	Homo sapiens [1000 Genomes] hg38 v5	Yes	Yes	Yes	Yes	No	No
mm10-methyl_cg.methylated_combined.ma-11-r5.0-2	Mus musculus [UCSC] mm10 v5	Yes	Yes	Yes	No	No	No
m6-ma-11-r5.0-2	Rattus norvegicus [UCSC] m6 v5	Yes	Yes	Yes	No	No	No
Bacillus_cereus_ATCC_10987-cnv.ma-11-r5.0-2	B. cereus [NCBI] ATCC_10987 v5	Yes	Yes	No	No	No	No
Rhodobacter_sphaeroides_2.4.1_12905-cnv.ma.seed_len19-11-r5.0-2	R.sphaeroides [NCBI] 2.4.1 v5	Yes	Yes	No	No	No	No
eschColi_K12_1-cnv.rna-11-r5.0-2	E. coli [NCBI] K12 MG1655 v5	Yes	Yes	No	No	No	No
phix-cnv.ma-11-r5.0-2	Microvirus Escherichia virus	Yes	Yes	No	No	No	No

	phiX174 [NCBI] v5						
Cereibacter_sphaeroides_ATCC_17023-rna-11-r5.0-2	C. sphaeroides [ATCC] 2.4.1 v5	Yes	Yes	No	No	No	No
Staphylococcus_aureus_NCTC_8325-rna-11-r5.0-2	S. aureus [NCBI] v5	Yes	Yes	No	No	No	No
Mycobacterium_tuberculosis_H37Rv-rna-11-r5	M. tuberculosis [NCBI] v5	Yes	Yes	No	No	No	No
Streptococcus_pneumoniae_Hu17-rna-11-r5.0-2	S. pneumoniae [NCBI] v5	Yes	Yes	No	No	No	No
Salmonella_enterica_LT2-rna-11-r5.0-2	S. enterica [NCBI] v5	Yes	Yes	No	No	No	No
Shigella_sonnei_SE6-1-rna-11-r5.0-2	S. sonnei [NCBI] v5	Yes	Yes	No	No	No	No
Saccharomyces_cerevisiae_S288C-rna-11-r5.0-2	S. cerevisiae [NCBI] v5	Yes	Yes	No	No	No	No
Human_immunodeficiency_virus_1-rna-11-r5.0-2	HIV 1 [NCBI] v5	Yes	Yes	No	No	No	No
Influenza_A_virus_H1N1-rna-11-r5.0-2	Influenza A H1N1 [NCBI] v5	Yes	Yes	No	No	No	No
Influenza_A_virus_H3N2-rna-11-r5.0-2	Influenza A H3N2 [NCBI] v5	Yes	Yes	No	No	No	No
Influenza_A_virus_H5N1-rna-11-r5.0-2	Influenza A H5N1 [NCBI] v5	Yes	Yes	No	No	No	No
SARS_coronavirus_2-rna-11-r5.0-2	SARS-CoV2 [NCBI] v5	Yes	Yes	No	No	No	No
Drosophila_melanogaster_Fruitfly-rna-11-r5.0-2	D. melanogaster [NCBI] v5	Yes	Yes	No	No	No	No

6.2 Available Non-Reference Genome Resource Files:

DRAGEN requires resource files to function correctly and achieve the optimum performance for certain workflows. All resource files are available for download at the Illumina DRAGEN Product Files support site at:

- https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/product_files.html

These resources can be imported to the instrument.

7 How To Generate Custom Resource Files

7.1 Reference Genomes:

Custom reference genomes can be imported to the instrument.

- Use the [Reference Builder \(Instruments\) BaseSpace App](#), which accepts a fasta file and generates a .gz file.
- Reference Builder App version 2.0.0 generates reference genomes for DRAGEN versions 4.1.* (v8 hash table format)
- Reference Builder App version 3.0.0 generates reference genomes for DRAGEN versions 4.3.* (v10 hash table format)

7.2 Noise Baseline and Panel of Normals Files:

The noise baseline file is built using normal samples that do not match to the subject that the samples are from. To generate a custom noise baseline file, use one of the following methods:

- Use the DRAGEN Bio-IT Platform server. Refer to the [Custom Systematic Noise Files](#) and [Generating Panel of Normals \(Combined Counts\)](#) for instructions.
- Use [DRAGEN Baseline Builder BaseSpace App](#) that generates CNV baseline files (*.target-counts or *.target-counts.gc-corrected, and *.combined.counts.txt.gz), SNV baseline files (*v2.0systematic_noise.bed.gz), and SV baseline files (*v3.0.0systematic_noise.sv.bedpe.gz).
- The recommended number of normal samples is 50.

8 How To Import Resource Files

You can add and delete reference files on the Reference Files tab in the Resources settings screen. The Reference Files tab displays the reference file name, file type, file description, and the number of related reference genomes.

8.1 Import Reference Files

- Make sure that there are no sequencing runs or on-instrument secondary analyses in progress.
- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select Resources.
- On the Reference Files tab, select Import reference file.
- Navigate to the reference file, and then select Open.
- [Optional] Enter a description for the reference file.
- Select a file type from the drop-down list.

9 How To Install DRAGEN and DRAGEN Applications

9.1 Install DRAGEN Versions:

- When a new DRAGEN version is available, download the DRAGEN installer (*.ires) from the MiSeq i100 Series support page. Save the installer locally or to a network drive.
- Make sure that there are no sequencing runs or on-instrument secondary analysis in progress.
- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select DRAGEN.
- On the Versions tab, select Install version.
- Navigate to the installer, and then select Open.
- Select Install. A message indicates if the installation was successful or failed

9.2 Uninstall DRAGEN Versions:

- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select DRAGEN.
- To uninstall a previous DRAGEN version, do as follows.
 - a. On the Versions tab, select the ellipsis icon in the Actions column.
 - b. Select Uninstall.
 - c. Select Yes, uninstall.
- To uninstall the latest DRAGEN version, do as follows.
 - a. On the Versions tab, select the ellipsis icon in the Actions column.
 - b. Select Uninstall all.
 - c. Select Yes, uninstall all

9.3 Application Installation:

- Download the application (*.iapp) from the MiSeq i100 Series support page. Save the installer to a network drive.
- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select Applications.
- Select Install application.
- Navigate to the application file, and then select Open. After the file uploads, information about the application displays.

- Select Install. After the application installs, you can review the application configuration.

9.4 View Application Settings:

The DRAGEN application provides a default library prep kit, index adapter kit, read information, index information, and permissions. Some applications also provide settings and configuration for secondary analysis.

- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select Applications.
- Select the application to view. After you install an application, the Configuration screen opens automatically.
- Edit any of the following information:
 - Library prep kits
 - Index adapter kits
 - Index reads
 - Read type
 - Index lengths
 - Read length
- Select Save.

9.5 Uninstall Applications

Administrators can uninstall applications.

- Select the instrument icon to open the global navigation menu.
- Select Settings, and then select Applications.
- Select the application to uninstall.
- Select Uninstall.
- Confirm to uninstall the application.

Release History

Revision	Release Reference	Originator	Description of Change
00	1130656	Mark Bilstad	Initial release