Customer Release Notes

DRAGEN Germline v4.2.7

for NextSeq 1000/2000

Contents

INTRODUCTION

These Release Notes detail the key changes to the DRAGEN Germline Workflow v4.2.7 onboard the NextSeq 1000/2000 instrument, since the v3.10.12 release. If you are updating from an older version, please see all intermediate release notes.

NEW FEATURES

- DRAGEN™ v4.2 offers significant improvements in accuracy, added features for a more comprehensive solution, and efficiency improvements. For full extensive details on each feature of pipeline, please consult the latest Illumina DRAGEN™ Bio-IT Platform User Guide available on the support website at https://support.illumina.com/downloads/illumina-dragen-bio-it-platform-user-guide.html
- **Reference Genome**
  - Enhanced multigenome (graph) reference v3 and Machine Learning (ML) models improves small variant calling accuracy.
    - DRAGEN v4.2 workflows require the installation of new reference genome Hash Tables (v9).
    - Pre-built instrument reference genomes can be downloaded from the NextSeq 1000/2000 Sequencing System Software Downloads page, for offline installation.
    - Please reference the NextSeq 1000/2000 Compatible Products page on the Illumina support site, for the list of available reference genomes and recommended genomes for different workflows.
  - New reference updates for hg38 improves variant calling in the Challenging, Medically Relevant Genes (CMRG) regions.
    - GRCh38 reference includes 34 sequences from chm13 and hs37d5 as decoys.
    - 29 decoys identified as missing segmental duplications.
    - 5 decoys identified in acrocentric arms of chromosomes 13, 14, 15, and 22 of CHM13.
    - Yield accuracy improvements in the CMRG genes: FANCD2, MAP2K3, KCNJ18, and KMT2C, as well as in the Y chromosome.
  - Support for the CHM13 v2.0 reference is available for research purposes. Only small variants have been fully validated.
    - Introducing support for the Telomere-to-Telomere CHM13v2.0 reference.
    - Accuracy has been validated only for WGS samples and small variant ML calls.
    - Accuracy tested on CMRG truth set.
- **Mapper/Aligner**
  - DNA and RNA Alignment updated.
    - DRAGEN mapper implemented a much more rigorous method for determining split-read alignments and influencing primary alignments and MAPQs with split-read analysis.
    - This sophisticated new method can support up to a maximum of 4095 secondary alignments and 4095 supplementary alignments per read.

    - New skew normal insert size model and pairing penalty function
        - Skew normal insert model better fits observed real-world asymmetric insert size distributions.
        - New PDF-based pairing penalty to avoid excessively penalizing larger insert size proper pairs.
        - Enabled by default for DNA pipelines, resulting in increased proper pairing rate.
        - RNA insert model unchanged.
- **Germline Small Variant Caller**
    - ML model updates further improves small variant calling accuracy.
        - ML is enabled by default for human samples.
        - ML is supported on hg38, hg19, hs37d5, CHM13 references.
        - Supported for Germline and Enrichment workflows. Not supported for Somatic, RNA and Amplicon workflows.
        - ML will automatically be enabled/disabled for supported/unsupported workflows and reference types.
        - Accuracy improvements at high depths (> 100x) WGS.
        - DGT, DGQ and DQUAL for the non-ML variant caller are no longer included in the VCF when ML is enabled.
- **Mitochondrial variant calling updates**
    - Default mitochondrial allele frequency thresholds have been updated. AF call threshold 1%, filter threshold 2%
    - Accuracy improvements have been achieved by
        - Tuning of mito-specific columnwise detection parameters
        - Implementation of mito-specific De Bruijn graph algorithm and parameter updates to reduce FP
- Improved **CNV and Structural Variant** calling accuracy.
- **Structural Variant Caller**
    - Accuracy and runtime improvements
        - Optimizations across the SV caller leading to improved accuracy and analysis times.
        - Insertion recall improved by 2% for insertions. Recall now exceeds 72% and 64% for deletions and insertions respectively, while maintaining high precision, as measured on HG002 using NIST GIAB v0.6 tier1 SV truth.
        - Significant reduction in run times across all workflows. Reduced by 20-60% across the board.
    - Tumor-only scoring model
        - Tumor-only pipeline previously reported all putative variants in the final VCF file, leading to low precision.
        - Candidate events and now scored and assigned a QUAL score in the VCF for filtering and training, reducing the total number of reported SVs.
        - The scoring model is consistent with the small variant caller.
- **Targeted Calling**
    - New targeted callers for higher genotyping accuracy - GBA, LPA, HBA, RH, CYP2D6, CYP2B6, CYP21A2, SMN, and accuracy improvements in CYP2D6.
    - Consolidated output file(s), *.targeted.vcf and *.targeted.json. Output files from multiple targeted callers are output into one combined "targeted" output. The TSV outputs are now disabled.

- **Star Allele Caller**
  - Star alleles are used for identification of genotypes and metabolism status of PGx genes that are included in FDA's PGx recommendations or have CPIC Level A designation.
  - Calls star alleles for 26 genes, including metabolism status of 21 genes.
  - Findings from PGx research can lead to better future outcomes for both individuals and healthcare providers through improved medication safety and efficacy and lowered medical costs.
  - Star alleles are haplotype patterns of a gene and can correlate with drug metabolism status for genes involved in drug metabolism.
  - Star Allele caller reports the optimal genotype along with the corresponding metabolism status* which is associated with that genotype.

Table 1 Target Callers Support

Targeted Callers	
Gene	**Application Area**
CYP21A2	Carrier Screening
HBA	
SMN silent carrier	
GBA	
LPA	Cardiovascular Disease
RH	Blood Typing
CYP2D6	PGx
CYP2B6	
HLA-A*	Transplant Matching
HLA-B*	

Table 2 PGx Gene Support

Caller	Gene	Supports hg38, hg19, hs37d5
Star Allele	CACNA1S	Y
	CFTR	
	CYP2C19	Y
	CYP2C9	Y
	CYP3A5	Y
	CYP4F2	Y
	IFNL3	Y
	RYR1	
	NUDT15	Y
	SLCO1B1	Y
	TPMT	
	UGT1A1	
	VKORC1	Y
	DPYD	Y
	G6PD	
	MT-RNR	
	BCHE *	Y
	ABCG2	
	NAT2 *	
	F5 *	Y

	UGT2B17 *	
Targeted	CYP2D6	
	CYP2B6	
	HLA-A *	
	HLA-B *	

- **Joint CNV/SV Detection**
  - Joint CNV/SV signal improves CNV detection.
    - Short CNVs in the genome are invisible or poor quality due to high variance in coverage. A combined depth and junction signal allows for base pair accurate CNV calling and refined breakpoint detection.
    - Enabled by default for Germline WGS analysis when both CNV and SV are configured.
    - New output VCF `*.cnv_sv.vcf.gz` which contains <DEL> and <DUP> records down to 1kbp.
    - Legacy VCFs still exist for backwards compatibility, though for CNV it is recommended to use the new VCF.
    - Improved recall and precision across all length scales
    - Recall for CNVs 1-10Kbp improves to >90%, alongside with precision gains 25%, when joint SV/CNV detection is employed.
- **Expansion Hunter**
  - Expansion Hunter v4 in DRAGEN v3.10 supported genotyping of 30 STR loci by default.
  - DRAGEN v4.2 updated to Expansion Hunter v5 and expands the default catalog supported to 60 pathogenic STR loci (including 30 from gnomAD)

NOTES

Reference Genome Usage Recommendations

Germline, Enrichment Germline, Amplicon Germline Workflows

- Use ALT-Masked mapping.
- Use Graph genomes.
- Use the Homo sapiens [1000 Genomes] hg38 Alt Masked Graph v3 reference genome.
- If not using hg38, the Homo sapiens [UCSC] hg19 Alt Masked Graph v3 or Homo sapiens [NCBI] hs37d5 v3 Graph are recommended.
- Non-graph is supported, but has reduced accuracy

Enrichment Somatic, Amplicon Somatic, RNA and scRNA Workflows

- Use ALT-Masked mapping.
- Use non-Graph genomes.
- Use the Homo sapiens [1000 Genomes] hg38 Alt Masked v3 reference genome or Homo sapiens [UCSC] hg19 Alt Masked v3 reference genome.

Table 3 v4.2 Reference Support and Recommended Use for Human Data

Human		hg19	hs37d5	hg38	chm13	Recommended Reference Type
Germline	SNV	Yes	Yes	Yes	Yes	Graph
	CNV	Yes	Yes	Yes	Yes*	Graph
	SV	Yes	Yes	Yes	Yes*	Graph
	Expansion Hunter	Yes	Yes	Yes	No	Graph
	Targeted Callers	Yes	Yes	Yes	No	Graph
	RNA	Yes	Yes	Yes	Yes*	Non-Graph
	De Novo	Yes	Yes	Yes	Yes*	Graph
	Joint Genotyping	Yes	Yes	Yes	Yes*	Graph
	Biomarkers (HLA)	Yes	Yes	Yes	Yes*	Graph
	Gvcf Genotyper	Yes	Yes	Yes	Yes*	Graph
Somatic	SNV	Yes	Yes	Yes	Yes*	Non-Graph
	UMI SNV	Yes	Yes	Yes	Yes*	Non-Graph
	CNV	Yes	Yes	Yes	Yes*	Non-Graph
	SV	Yes	Yes	Yes	Yes*	Non-Graph
Methylation	Methylation	Yes	Yes	Yes	No	Non-Graph
Annotation	Illumina Annotation Engine (Nirvana)	Yes	Yes	Yes	No	n/a

(*) DRAGEN™ supports the component execution; however, the component's accuracy has not been established.

Table 4 v4.2 Reference Support and Recommended Use for Non-Human Data

Non-Human		Supported	Recommended Reference Type
Germline	SNV	Yes	Non-Graph
	CNV	No	n/a
	SV	Yes	Non-Graph
	Expansion Hunter	No	n/a
	Targeted Callers	No	n/a
	RNA	Yes	Non-Graph
	De Novo	Yes	Non-Graph
	Joint Genotyping	Yes	Non-Graph
	Biomarkers (HLA)	No	n/a
	Gvcf Genotyper	Yes	Non-Graph
Somatic	SNV	No	n/a
	UMI SNV	No	n/a
	CNV	No	n/a
	SV	No	n/a
Methylation	Methylation	No	n/a
Annotation	Illumina Annotation Engine (Nirvana)	Yes	n/a

How to Update Instrument Reference Genomes

You can only import new reference genomes from the administrator account.
Steps:

1. Download the desired genome package tar.gz from the NextSeq 1000/2000 Software Downloads page, OR
2. Create a reference genome using the Reference Builder for Illumina Instruments BaseSpace Sequence Hub app. For more information, refer to Reference Builder for Illumina Instruments v1.0.0 App Online Help.
3. Select the control software menu, and then select ***Process Management***. Make sure that there are no sequencing runs or on-instrument secondary analyses in progress.
4. From the control software menu, select ***Minimize Application***.
5. Log into ilmnadmin.
6. Select the control software menu, and then select ***DRAGEN***.
7. In the Genome section, select ***View Installed Genomes*** to view a list of all currently installed genomes.
8. Close the window.
9. Under Import New Reference Genomes, select Choose.
10. Navigate to the reference genome file (*.tar.gz) on the portable or mounted network drive, and then select ***Open***.

KNOWN ISSUES

Known issues of the DRAGEN™ v4.2.7 Germline Workflow

Comp	ID	Summary	Resolution/Workaround
CNV VC	DRAGEN-28695	Up to v4.0, if we cannot detect a purity or ploidy (DEGENERATE_DIPLOID) then we would FAIL all records in our VCF. In v4.2, we changed this behavior for WGS, but not for WES.	No workaround for WES. A fix has been made to future version.
Multi genome reference	DRAGEN-26308	WGS runtime increased with multigenome vs legacy genome	5% longer run time. For information only.
Paralog Caller	DRAGEN-25971	GBA reports a single recombinant haplotype with RecNciI+RecNciI instead of two recombinant haplotypes with RecNciI each for NA20273	Not a bug in the caller, but output formatting may be misleading. Format change planned for future version
SNV Germline	DRAGEN-23801	Joint Calling in Mito is not giving proper VAF's, when one or more samples have a variant, but other samples have a homref call at the same position.	For some alleles, the AD values in the joint VCF are not accurate. Looking at the corresponding single sample gVCF can resolve the inconsistency.
SNV Germline	DRAGEN-26359	Small regression in INDEL sensitivity in v4.2 compared to v4.0	Single sample SNV has major accuracy improvements due to graph and reference and machine learning updates. In some cases, such as INDEL, there are very minor changes where precision may improve at cost of recall, and vice versa. For information only.

SNV VC	DRAGEN-25905	Hang observed on high depth samples, when target BED is used to run the SNV caller over regions which are close to the end of a chromosome.	Have more BED regions throughout the chromosome or increase bin memory.
SNV VC	DRAGEN-25933	VCF GQ values may not match VCF specification	In most positions, the probability that the position is a variant is very close to 1 and the impact is negligible. In corner cases where p(0/0) is not negligible, we have the wrong value in the GQ field. For information only.

INSTALLATION INSTRUCTIONS

DRAGEN v4.2.7 workflows are compatible with Control Software 1.7.x. For information regarding compatibility with other Control Software versions, please reference the NextSeq 1000/2000 Compatible Products page on the Illumina support site. If you would like to update the control software, please follow the steps detailed in the NextSeq 1000/2000 Control Software Suite v1.7.x Release Notes on the Illumina Support Site.

Online Installation of Workflows

If the instrument is connected to the internet, you can install DRAGEN workflows directly from the Control Software. Online installation of workflows is available since Control Software v1.3 or later.

Steps:

1. Make sure that you have the password to the *ilmnadmin* account.
2. Log in to the *ilmnadmin* account:
    a. If you are logged in as *ilmnuser,* and in control software, select the control software menu, and then select **Exit Application** to access the desktop.
    b. Select the power button icon in the upper right corner and log out of *ilmnuser*.
    c. After you are on the login screen, select *ilmnadmin*, and then enter the password to log in.
    d. The control software automatically launches once you are logged in.
3. Make sure that there are no sequencing runs or on-instrument secondary analysis in progress.
4. On the control software menu, select **DRAGEN**. Under Version, the Available Workflows section lists the workflows currently installed on the system.
5. Select **Check Online**. Not all DRAGEN versions and workflows are compatible with online installation. Use offline installation for additional workflows.
6. Select the checkbox for the workflows that you would like to install. **NOTE: For Online Updates, DRAGEN BCL Convert must be installed before or with other workflows of the same DRAGEN Version.** You can view information about the latest version of a workflow in the release notes.
7. Select Install to start installation.
8. Enter *ilmnadmin* for the system password, and then select Authenticate.
9. After installation is complete, you will be navigated back to the DRAGEN screen and can view the updated list of installed DRAGEN workflows.

Offline Installation of Workflows

Steps:

1. When a DRAGEN workflow update is available, download the installer (*.tar.gz) from the NextSeq 1000/2000 Sequencing System support page. Save the installer to a local or portable drive.
2. If you saved the installer to a portable drive, plug the drive into a USB 3.0 port, located on both the side and back of the instrument. Gently move the instrument as needed to access the back.
3. Follow steps 1-3 above to log in to *ilmnadmin*.
4. Select the control software menu, and then select **DRAGEN**.
5. Under Version, select **Browse for New Version** to navigate to the installer.
6. Select Install to start installation.
7. Enter *ilmnadmin* for the system password, and then select Authenticate.
8. After installation is complete, you will be navigated back to the DRAGEN screen and can view the updated list of installed DRAGEN workflows.

Note that starting with control software v1.5, it is possible to uninstall previous versions of DRAGEN workflows.

RELEASE HISTORY

Revision	Release Reference	Originator	Description of Change
00	CN 1103441	Yi Lian	Initial release