DRAGEN v4.2.4 Software Release Notes

Introduction

These release notes detail the key changes to software components for the Illumina® DRAGEN™ Bio-IT Platform v4.2.4.

Changes are relative to DRAGEN™ v4.1.5 or v4.0.3. If you are upgrading from a version prior to DRAGEN™ v4.1.5 or v4.0.3, please review the release notes for a list of features and bug fixes introduced in subsequent versions. The bioinformatics features available in DRAGEN™ v4.1 were the same as DRAGEN™ v4.0.

DRAGEN™ Installers, User Guide and Release Notes are available here:
https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform.html

The software package includes downloadable installers for Phase 3 and Phase 4 on-site servers:
- DRAGEN™ SW for x86 Oracle 8 - dragen-4.2.4-9.el8.x86_64.run
- DRAGEN™ SW for x86 Centos 7 - dragen-4.2.4-9.el7.x86_64.run

The following configurations containing DRAGEN™ 4.2.4 are also available on request:
- Centos 7 Amazon Machine Images (AMI) for f1 instances, available in 12 regions
- Centos 7 Microsoft Azure Image (VM) available in West US 2
- Centos 7 and Oracle 8 RPM packages for use with Amazon Web Services (AWS) f1 instances, for customer generated AMIs or customer generated docker images
- DRAGEN™ Kernel drivers for el7 and el8, for use with customer generated AMIs or QuickStart
- Pre-built docker images with Centos 7 and Oracle 8 for on-site, AWS usage
- Pre-built docker image with Centos 7 for Microsoft Azure cloud usage

Deprecated platforms:
- Support for DRAGEN™ Server v1 FPGA cards have been deprecated since DRAGEN™ v3.10
- Support for Ubuntu has been deprecated since DRAGEN™ v3.9
- Support for x86 CentOS 6 has been deprecated since DRAGEN™ v3.8

Contents

Overview

Below is a summary of the changes included in v4.2.4. 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

Accuracy

- Enhanced multigenome (graph) reference and Machine Learning (ML) models improves small variant calling accuracy.
- Improved CNV and Structural Variant calling accuracy.
- New targeted callers for higher genotyping accuracy - HBA, LPA and RH, CYP21A2, SMN (silent carrier variant), and accuracy improvements in CYP2D6.
- Accurate star allele calling for 5 more pharmacogenes: BCHE, ABCG2, NAT2, F5, UGT2B17.

Comprehensiveness

- Support for CHM13 v2.0 reference.
- Germline with high sensitivity mode.
- Sex chromosome low allele frequency variant support.
- Imputation for haploid species and sex chromosomes.
- Enhanced Bulk-RNA QC metrics.
- Integrated pipelines for precision metagenomics analysis on RPIP/UPIP panels.

Efficiency

- Increase ORA compression speeds up to 30% when mapping/align step is enabled.
- Runtime improvements for joint genotyping pipelines compared to previous release.

Please review the section on Known Issues and limitations of the release.

Updated Resource Files

DRAGEN™ 4.2 requires updates to key resource files to function correctly and achieve the optimum performance. All resource files are available for download at the Illumina DRAGEN™ Product Files support site here: https://support.illumina.com/sequencing/sequencing_software/dragen-bio-it-platform/product_files.html

The following resource files are updated:

Resource	Description	File name(s)

Hash Tables v9	Pre-built v9 multigenome hash tables for hg38, hg19, hs37d5, CHM13. The hash table builds include DNA, RNA, CNV, HLA tables.	`hg38-alt_masked.cnv.graph.hla.rna-9-r3.0-1.tar.gz`
		`hg19-alt_masked.cnv.graph.hla.rna-9-r3.0-1.tar.gz`
		`hs37d5-cnv.graph.hla.rna-9-r3.0-1.tar.gz`
		`chm13_v2-cnv.graph.hla.rna-9-r3.0-1.tar.gz`

SNV Systematic Noise Baseline collection v1.1.0	A collection of noise baseline BED files for hg19, hs37d5, hg38 and	`systematic-noise-baseline-collection-1.1.0.tar`
		
		The tar archive contains the following files:
		`snv_wes_nextera_hg19_max_v1.1_systematic_noise.bed.gz`

	for WGS and WES respectively	snv_wes_nextera_hg19_mean_v1.1_systematic_noise.bed.gz snv_wes_nextera_hg38_max_v1.1_systematic_noise.bed.gz snv_wes_nextera_hg38_mean_v1.1_systematic_noise.bed.gz snv_wes_nextera_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wes_nextera_hs37d5_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg19_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg19_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg38_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hg38_mean_v1.1_systematic_noise.bed.gz snv_wes_truseq_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wes_truseq_hs37d5_mean_v1.1_systematic_noise.bed.gz snv_wgs_hg19_max_v1.1_systematic_noise.bed.gz snv_wgs_hg19_mean_v1.1_systematic_noise.bed.gz snv_wgs_hg38_max_v1.1_systematic_noise.bed.gz snv_wgs_hg38_mean_v1.1_systematic_noise.bed.gz snv_wgs_hs37d5_max_v1.1_systematic_noise.bed.gz snv_wgs_hs37d5_mean_v1.1_systematic_noise.bed.gz
SV Systematic Noise Baseline collection v2.0.0	A collection of noise baseline BEDPE files for hg19, hs37d5, hg38 for WGS	sv-systematic-noise-baseline-collection-2.0.0.tar The tar archive contains the following files: WGS_v2.0.0_hg19_sv_systematic_noise.bedpe.gz WGS_v2.0.0_hg38_sv_systematic_noise.bedpe.gz WGS_v2.0.0_hs37d5_sv_systematic_noise.bedpe.gz
Custom Multigenome Reference Builder resources v1.1.0	Fasta, graph BED, mask BED files for hg38, hg19, hs37d5, CHM13, needed for custom multigenome hash table building from own population VCFs	hg38-custom-reference-genome-1.1.0.tar.gz hg19-custom-reference-genome-1.1.0.tar.gz hs37d5-custom-reference-genome-1.1.0.tar.gz chm13_v2-custom-reference-genome-1.1.0.tar.gz
Imputation Reference Panels v1.2 and v2.0	Genetic maps and reference panels for hg38	irp-hg38-1.2.1.tar irp-hg38-2.0.0.tar

NOTES:
- ML Model files for DRAGEN™ v4.2 are now included in the installer by default and does not need to be downloaded.
- Multigenome references can now be built with the hash table builder. Pre-built hash tables are provided for reference.

Major Features and Updates

Reference Genome

- Hash Tables v9
  - Hash tables must be re-built to use DRAGEN™ 4.2. Existing hash tables built with v4.1 or older are not supported.
  - Improvements to the multigenome reference, expanded genome support, and changes to the mapper/aligner, resulted in interface changes to the reference hash tables. The hash table interface is updated to version 9 (HTv9).
- Hash table builder now supports building of multigenome references.
  - Multigenome reference contigs and decoys are now packaged in the installer.
  - The hash table builder auto-detects the human based references and applies the additional decoys to the provided fasta.

  - The option `--ht-apply-graph` can be enabled when executing the hash table builder, to automatically apply any graph resources that are relevant to the input fasta. It is supported for hg38, hg19, hs37d5, and chm13 references, and highly recommended for Germline workflows.
  - The option is disabled by default. See Table1 and Table 2 below for recommended usage.
- Multigenome reference introduced in v3.7 is now extended to global population samples.
  - The v4.2 multigenome reference extends the sample population to 32 samples of different ancestries around the globe.
  - Includes CHM13 population contigs.
  - Graph bed total bases increased by 20%.
  - Error reduction among non-European ancestry samples of 16.9% compared to the multigenome reference of v4.0 built on 16 samples of European ancestry.
- 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.
- These reference updates are denoted as reference v3.
- CHM13v2.0 reference support
  - Introducing support for the Telomere-to-Telomere CHM13v2.0 reference.
  - Accuracy has been validated only for WGS samples and small variant ML calls.
  - Available with the multigenome reference v3.
  - Accuracy tested on CMRG truth set.
- Pre-built hash tables for all supported human references are available at the Illumina DRAGEN™ Product Files support site.

Table 1 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	Nirvana	Yes	Yes	Yes	No	n/a

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

Table 2 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	Nirvana	Yes	n/a

Mapper/Aligner

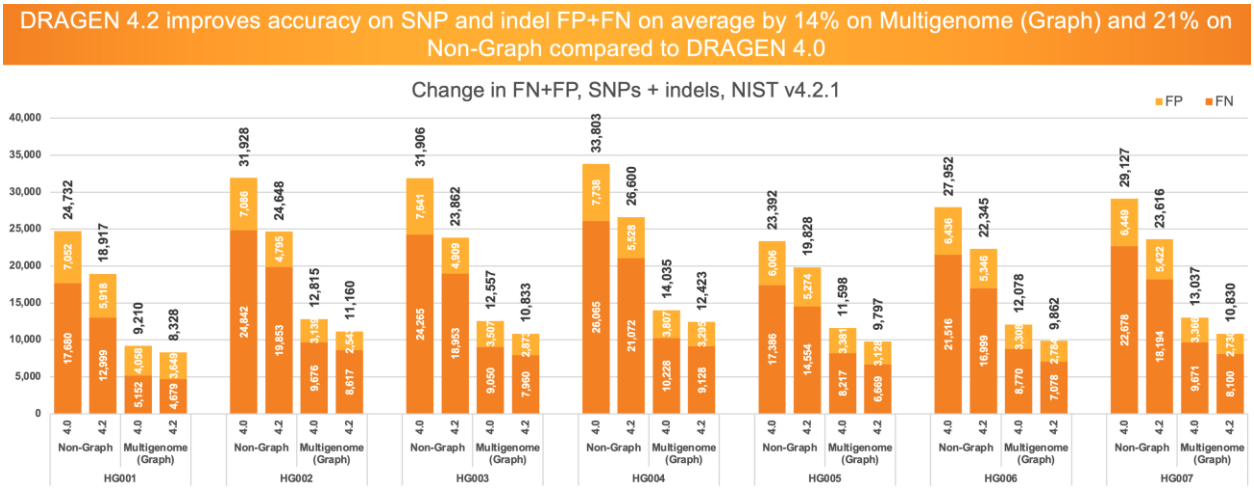
- New skew normal insert size model and pairing penalty function
  - Skew normal insert model better fits observed real-world asymmetric insert size distributions.
  - New PDF-based pairing penalty to avoid excessively penalizing larger insert size proper pairs.
  - Enabled by default for DNA pipelines, resulting in increased proper pairing rate.
  - RNA insert model unchanged.
- New mapper tags in BAM/CRAM/SAM output
  - New mate alignment tags: Mate cigar (MC:Z), Mate mapping quality (MQ:i), Mate sequence (R2:Z) and Mate base quality (Q2:Z) tags.
  - New pair score tag (ps:i): that reports the pair score used internally in the mapper to compare alignment candidates and report mapping quality (MAPQ).
  - Available options:

    `--generate-mc-tags` Generate mate cigar MC:Z tag (default=true)
    `--generate-mq-tags` Generate mate mapping quality MQ:i tag (default=true)
    `--generate-r2-tags` Generate mate sequence R2:Z tag (default=false)
    `--generate-q2-tags` Generate mate base quality (Q2:Z) tag (default=false)
    `--generate-ps-tags` Generate pair score ps:i tag (default=false)
  - Please see the User Guide for more details on usage.

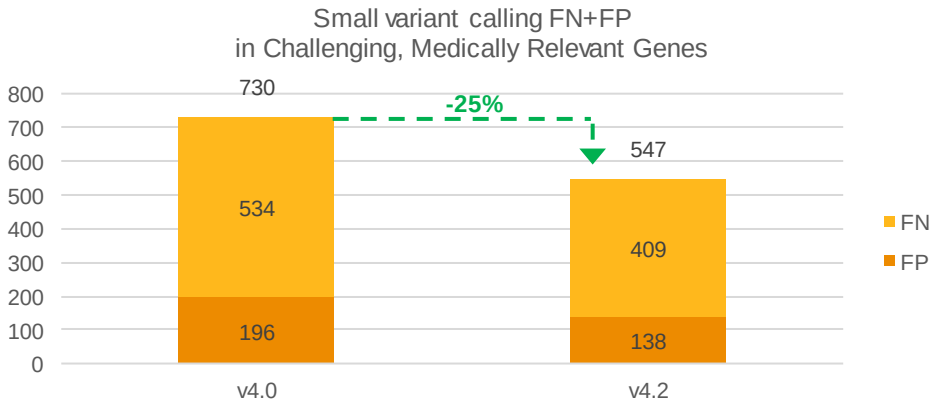
Germline Small Variant Caller

- 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.

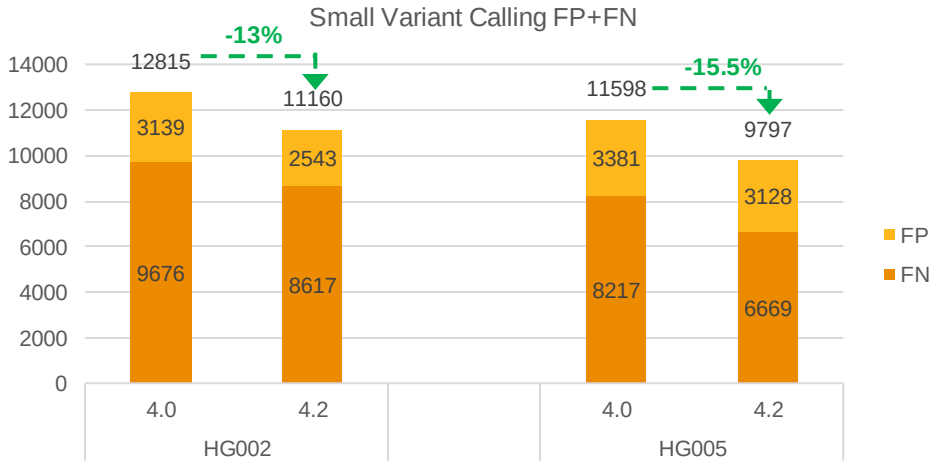
  - ML will automatically be enabled/disabled for supported/unsupported workflows and reference types.
  - Accuracy improvements at high depths (> 100x) WGS.
  - DGT, DGQ and DQUAL for the non-ML variant caller are no longer included in the VCF when ML is enabled.
- Enhanced multigenome reference v3.
  - Multigenome reference contigs and decoys are packaged as part of the installer and applied during hash table building.



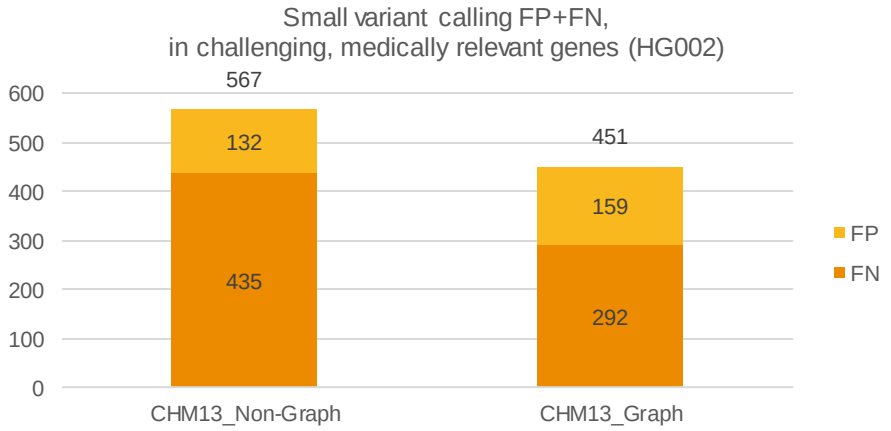
  - Updates to decoys on hg38 yields accuracy improvements on CMRG genes.



- Improved accuracy on non-European samples.
  - Error reduction among non-European ancestry samples compared to v4.0.



- T2T CHM13 support
  - Available with the multigenome reference v3.
  - Accuracy tested on CMRG truth set.
  - NOTE: Accuracy has been validated only for WGS samples and small variants ML calls.



- High Sensitivity Mode
  - Optionally enable high sensitivity mode with `--vc-enable-high-sensitivity-mode=true`
    - Processes regions which present only reads with mapping quality of 0.
    - Enables calling of variants with low allele frequency (AKA mosaic variants).
    - Runs regular small VC on those regions.
  - High sensitivity mode lowers FN at the cost of increased FP. Results in increased sensitivity for reads with low MAPQ and low allele frequency.
- Sex chromosome mosaic variants support
  - New option `--vc-enable-sex-chr-diploid` enabled by default.
  - New option `--vc-haploid-call-af-threshold=<af_threshold>` to control threshold.
    - Diploid model is applied to haploid (chrX/Y, non-PAR) regions in male samples.
    - Variants with only one alt allele and with AF>=85% are rewritten to haploid calls.
    - The potential mosaic calls with AF<85% will have GT of "0/1" and an INFO tag of "MOSAIC" will be added.
    - Example VCF output:

```
chrX    18622368        .   C    T    48.84
PASS    AC=1;AF=0.500;AN=2;DP=22;FS=4.154;MQ=248.02;MQRankSum=3.272;Q
D=2.27;ReadPosRankSum=2.671;SOR=1.546;FractionInformativeReads=1.000;
MOSAIC  GT:AD:AF:DP:F1R2:F2R1:GQ:PL:GP:PRI:SB:MB        0/1:9,13:0.59
09:22:1,8:8,5:48:84,0,51:4.8837e+01,7.4031e-
05,5.4007e+01:0.00,34.77,37.77:5,4,4,9:3,6,5,8
```

- Increased sensitivity toward germline SNV detection

SV Caller

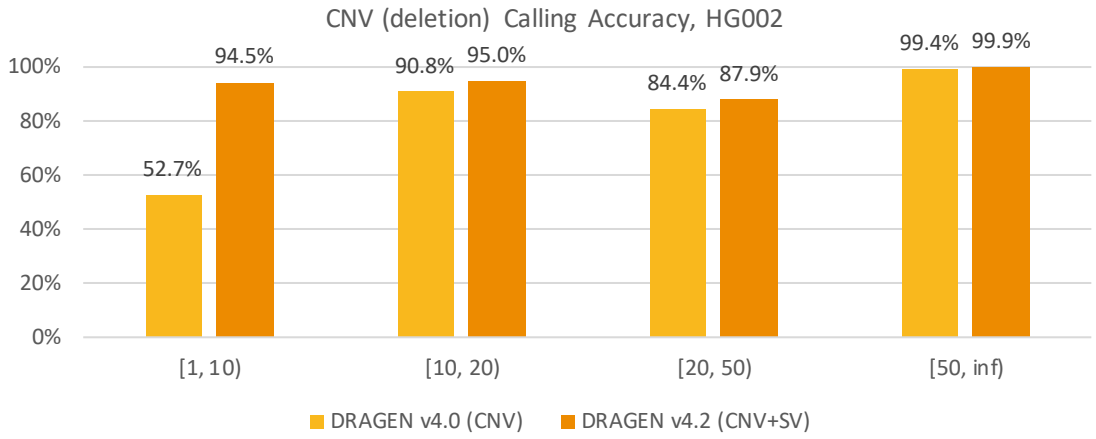
- Improved FLT3-ITD sensitivity in v4.2 with updates to SV hotspot handling
  - SV hotspots have been updated.
  - The combined accuracy of SNV+SV calling improves sensitivity of v4.2 and outperforms Pindel on SV calling.

Caller	Recall	Avg Variants
DRAGEN 4.2	**94.8%** (74/78)	4
DRAGEN 4.0	87.2% (68/78)	1
Pindel	93.6% (73/78)	350

- SV VCF header is updated to be consistent with other DRAGEN™ callers.
  - Added
    `##DRAGENVersion`
    `##DRAGENCommandLine`
    Changed
    `##source=DRAGEN_SV`
    `##fileformat=VCFv4.2`
    Removed
    `##cmdline`

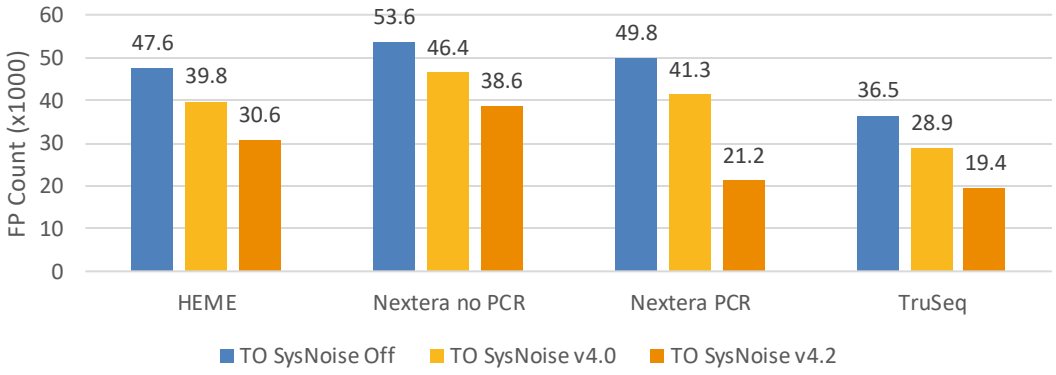
Joint CNV/SV Detection

- 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.



Somatic Small Variant Caller

- Better T/O and T/N accuracy with improved Systematic Noise Filter + FFPE Blocklists.
  - New systematic noise files use Nirvana germline annotation to ignore germline calls and to extract noise from panel of normals (PoN).
  - New noise files result in approximate 30% fewer FP calls.
  - The default noise files are updated, and settings for custom generated noise files are also improved.
  - User recipes include steps for filtering ALU regions that may represent 90% of FP calls in some FFPE samples.
  - The updated systematic noise files are available for download at the Illumina DRAGEN™ Product Files support site.

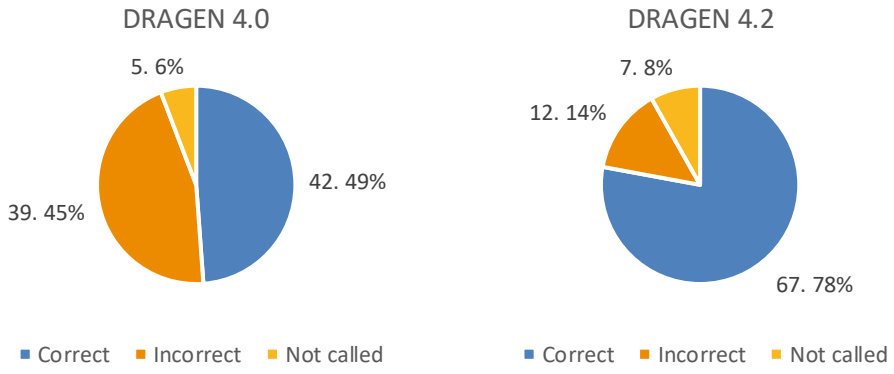


- Simplified usage for Somatic SNV
  - Better standardization on 10 internal settings across pipelines
  - New umbrella setting to specify the lowest allele frequency of interest:
    - `--vc-target-vaf=<float>` The variant caller will aim to detect variants with allele frequencies equal to and larger than this setting. This setting will not apply a hard filter and it is possible to detect variants with allele frequencies lower than the selected threshold.
      - Default setting is 0.03 (AF = 3%)
      - Set lower (~0.01) to improve low AF sensitivity in high quality samples.

        - Set higher (~0.04) to reduce low AF FP.
    - Easy to follow recipes have been added in the User Guide.
    - Simplified TMB usage
        - Improved standardization across pipelines: TN, TO + WGS, WES
        - Command line interface has been simplified.
        - Reference specific coding regions bundled & auto applied.
        - TMB analysis aims to harmonize with MSK-IMPACT and Foundation One TMB metrics.
- Other new Somatic SNV VC options:
    - `--vc-base-qual-threshold` replaces `--vc-min-base-qual` for specifying the minimum BQ that will be used at any stage in the SNV caller
    - Phasing and Phased Variants
        - Component SNVs/INDELs of MNV calls are output only if the VAF of the component call is greater than that of the MNV by more than 0.1. The VAF difference threshold for outputting component calls along with MNV calls can be controlled by the `--vc-combine-phased-variants-max-vaf-delta` option.
        - To output all component SNVs/INDELs of MNVs, regardless of VAF difference, when enabled, use the option `--vc-mnv-emit-component-calls`.
        - These two options are mutually exclusive and are only available for use in the somatic pipeline.

CNV Caller

- Improved tumor purity/ploidy estimation accuracy
    - Somatic CNV selection model is limited to essential genes.
    - Knowledge on cell-essential genes constrains model identification.
    - Somatic variant calling based on linear copy-ratio to for large somatic alterations (≥ 1 Mb). Thresholds modeled based on the sample's variance.
    - Features enabled by default for supported reference genomes (hg19, GRCh37, hg38)
    - Usage:
        - Essential Genes
          `--cnv-somatic-enable-lower-ploidy-limit=true`
          `--cnv-somatic-essential-genes-bed=<BEDFILE_PATH>`
        - Low model confidence
          `--cnv-filter-dup-mean=<DUP_threshold>`
          `--cnv-filter-del-mean=<DEL_threshold>`



- Detect ASCN on Somatic WES samples with T/N analysis.
    - Estimate purity/ploidy of sample to make precise copy number calling.

- Now supports Allele Specific Copy Number (ASCN) detection on somatic WES samples with tumor-normal analysis.
- Normalize copy numbers with matched normal samples along with panel of normals (PoN)
- Accepts FASTQ/BAM/CRAM/VCF formats.
- Usage:

```
--tumor-bam-input=<TUMOR_BAM>
--bam-input=<NORMAL_BAM>
--enable-cnv=true
--enable-variant-caller=true
--cnv-target-bed=<BED>
--cnv-normals-list=<PoN>
--cnv-use-somatic-vc-baf=true
```

- If PoN is not provided, the matched normal will be used as a single sample PoN (not recommended).

HCC1395	Metric	DRAGEN 4.0 WES	DRAGEN 4.2 WES	DRAGEN 4.2 WGS
Deletions	Recall	0.865	0.876	0.984
	Precision	0.044	0.967	0.964
	F-score	**0.083**	**0.919**	**0.975**
Duplications	Recall	0.446	0.982	0.990
	Precision	0.997	0.987	0.970
	F-score	**0.616**	**0.985**	**0.980**

- CNV caller options changes
  - Count duplicate marked alignments during target counts: `--cnv-count-duplicate-alignments` (Disabled by default)
  - Enable fixed window BAllele stat estimate: `--cnv-enable-fixed-size-ballele-stat-estimate` (Enabled by default)
  - Segment Mean (SM) threshold values used to hard filter DELs/DUPs in CNV VCF (Somatic WGS) can be adjusted: `--cnv-filter-del-mean`, `--cnv-filter-dup-mean`.
  - Generate PoN metric file when using WES/targeted panel: `--cnv-generate-pon-metric-file` (Enabled by default)
  - Removed CNV Ploidy option: `--cnv-ploidy`.
  - Enable check on lower ploidy limit based on essential genes: `--cnv-somatic-enable-lower-ploidy-limit` (Enabled by default)
  - BED file containing genes where the model should not predict HOMDEL: `--cnv-somatic-essential-genes-bed` (The software automatically selects a default file based on reference used for hg19, hs37d7, hg38, chm13, if none specified)

Targeted Callers and PGx

Overview of new features:

- New targeted callers: CYP21A2, HBA, LPA, RH, SMN silent carrier
  - Each new caller can be enabled/disabled with their own options: `--enable-cyp21a2`, `--enable-hba`, `--enable-lpa`, `--enable-rh`.
  - All targeted callers can be enabled together with one option `--enable-targeted=true`
- Star Allele caller expansion
  - BCHE, ABCG2, NAT2, F5, UGT2B17
- Consolidated JSON output file
  - Consolidate multiple TSV output files from targeted callers into one JSON file.
  - The TSV outputs are now disabled by default.

- Merge targeted caller small variants into the `hard-filtered` SNV VCF output.
  - Space separated list of targeted callers: `--targeted-merge-vc` Accepted values are [`none | all | hba lpa rh smn`] (Default=rh)
  - Targeted calls merged into the hard-filtered files are marked with a `TARGETED` INFO flag.

Targeted Callers in DRAGEN 4.2

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*	

New with DRAGEN 4.2
Improved with DRAGEN 4.2

Targeted callers' details:

- **Hemoglobin subunit alpha (HBA) genotype detection from WGS data**
  - HBA1 and HBA2 variants and copy number changes may cause Alpha thalassemia. Alpha thalassemia is a major type of hemoglobinopathy, the world's most common genetic disease. Carrier frequency can be as high as 30-40% in African populations.
  - HBA caller identifies HBA1 and HBA2 copy numbers from WGS data by analyzing read depth and 4 informative regions of sequence surrounding HBA 1/2.=
  - HBA genotype is reported along with the possible target variants detected.
  - Usage
    - To enable HBA caller: `--enable-hba=true`
      Example output from `*targeted.json` file:

```
{
  "sample": "HG00699",
  "hba": {
    "genotype": "--/aa",
    "genotypeFilter": "PASS",
    "genotypeQual": 96.70607775855007,
    ...
}
```

      ** Supported references: hs37d5, hg19, hg38*
      ** HBA caller additionally includes VCF output*
  - Highly concordant with orthogonal methods
    - Orthogonal validation methods include concordance with long read calls.
    - Calls have been validated against 247 samples from 1000 Genomes Project.

HBA Genotype	Total samples	Concordant
aa/aa	202	199 (**98.51%**)
-a3.7/aa	31	31 (**100%**)
--/aa	4	4 (**100%**)
-a3.7/-a3.7	4	4 (**100%**)

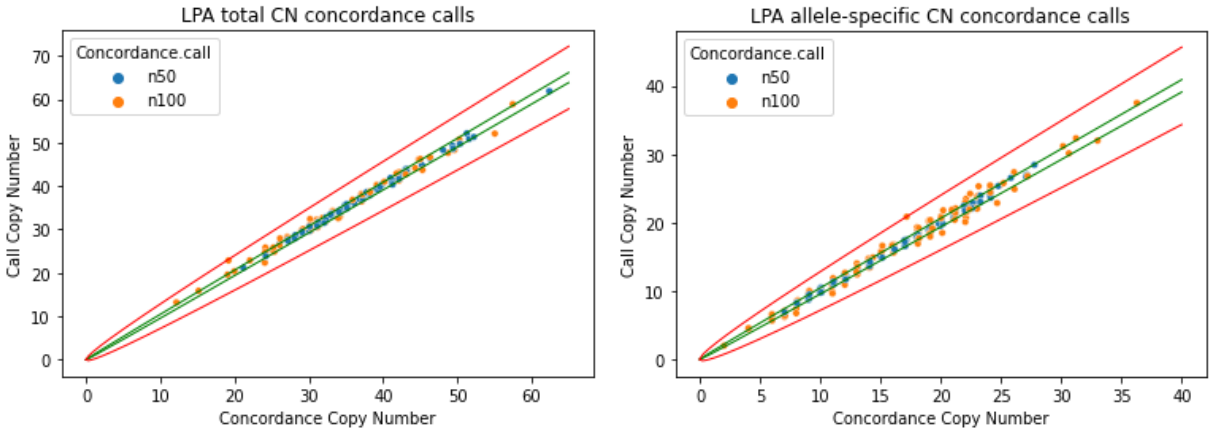
aaa3.7/aa	3	3 (**100%**)
-a4.2/aa	2	2 (**100%**)
aaa4.2/aa	1	1 (**100%**)
Total	247	244 (**98.79%**)

- **Lipoprotein(A) (LPA) Kringle-IV-2 copy number detection from WGS data.**
  - Lower copy number of LPA Kringle-IV-2 (KIV2) VNTR increases blood concentrations of the Lp(a) lipoprotein, which may cause or indicate increased cardiovascular disease risk.
  - LPA caller identifies KIV2 copy numbers from WGS data.
  - The KIV2 copy number is reported along with the possible target markers detected.
  - Usage:
    - To enable LPA caller: `--enable-lpa=true`

      Example output from `*targeted.json` file:

```
{
  "sample": "HG00096",
  "lpa": {
    "kiv2CopyNumber": 39.2360,
    "refMarkerAlleleCopyNumber": 17.8467,
    "altMarkerAlleleCopyNumber": 21.3892,
    "type": "Heterozygous markers call",
   ...
}
```

      ** Supported references: hs37d5, hg19, hg38*
      ** HBA caller additionally includes VCF output*
  - Highly concordant with orthogonal methods
    - Orthogonal validation methods include Bionano calls.
    - Validated against 144 samples from 1000 Genomes Project



- **RHCE*CE-D(2)-CE gene conversion detection from WGS data**
  - Rhesus antigens play an important role in Red Blood Cells (RBC) antigens phenotype.
  - RH caller identifies the RHCE*CE-D(2)-CE gene conversion from WGS data by analyzing read depth and 798 sites of difference between RHD and RHCE genes.
  - The RHCE*CE-D(2)-CE gene conversion is reported as a list of SNV variants in the VCF file.
  - Usage:
    - To enable RH caller: `--enable-rh=true`

Example outputs from `*targeted.json` file:

```
{
  "sample": "HG002",
  "rh": {
    "rhdCopyNumber": 2,
    "rhceCopyNumber": 2,
    "variants": [{
      "hgvs":
            "NC_000001.11g.25405596_25409676con25283766_25287797",
      "qual": 63.4236,
   ...
}
```

** Supported references: hs37d5, hg19, hg38*
** HBA caller additionally includes VCF output*

- Highly concordant with orthogonal methods
  - Orthogonal validation methods include concordance with long read Human Pangenome Reference Consortium (HPRC) Dipcalls pipeline.
  - Validated against 42 samples from 1000 Genomes Project

RHCE*CE-D(2)-CE gene conversion	Total samples	Concordant
Not present	24	24 (**100%**)
Heterozygous	10	9 (**90%**)
Homozygous	8	8 (**100%**)
Total	42	41 (**97.62%**)

- **CYP21A2 variant detection from WGS data**
  - Pathogenic variants in the CYP21A2 gene can cause 21-Hydroxylase-Deficient Congenital Adrenal Hyperplasia, an autosomal recessive disease.
  - Reads mapping to either CYP21A2 or CYP21A1P are used to detect variants in homology regions and read phasing is used to detect any recombination events between the two genes.
  - Usage:
    - To enable CYP21A2 caller: `--enable-cyp21a2=true`

Example fields from `*targeted.json` output file:

```
{
    "sample": "HG01801",
    "cyp21a2": {
            "totalCopyNumber": 3,
            "recombinantHaplotypes": [
                "NM_000500.9:c.518T>A",
                ""
     ...
}
```

** Supported references: hs37d5, hg19, hg38*

  - Highly concordant with orthogonal methods
    - Validation samples used:
      - 14 affected samples from Radboudumc (Long-range PCR, MLPA)
      - 66 unaffected 1KGP samples (some carriers)
      - 4 Coriell samples (2 affected and 2 carriers)

Benchmark set	Total samples	Concordant
Radboudumc	14	13 (**92.86%**)
1KGP	66	65 (**98.48%**)

Coriell	4	4 (**100%**)
Total	84	82 (**97.62%**)

- **Star Allele caller**
  - Star alleles are used in pharmacogenomics (PGx). 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.
  - Start Allele caller reports the optimal genotype along with the corresponding metabolism status* which is associated with that genotype.
  - v4.2 adds 5 more PGx genes to the Star Allele caller:
    - BCHE, ABCG2, NAT2, F5, UGT2B17
    - These genes have been assigned Level-A guidelines for gene-drug interactions by CPIC.
    - Now supports hg38, hg19*, hs37d5* references.(**Available for selected genes.)*
  - Usage:
    - Star allele caller can be run as standalone caller, with an input gVCF file; or enabled with other callers when starting from FASTQ or BAM; by using the option: `--enable-star-allele=true`
    - When run from gVCF input, the caller can take DRAGEN™ VCF/gVCF and DRAGEN™ CNV-VCF files as input.
      - Example standalone run: `dragen --star-allele-gvcf /staging/NA12878.vcf --output-directory /staging --output-file-prefix NA12878_dragen --enable-star-allele true.`
      - Use optional input `--star-allele-cnv-vcf=<CNV-VCF>` to skip calling star alleles that are detected through CNV analysis.

- **CYP2D6 updates**
  - Improved the caller for samples with incorrect genotype.
  - Robustness improvements

- **Pharmacogenomics (PGx) grouping**
  - All PGx callers can be enabled/disabled through individual options as indicated above.
  - Alternately, all the targeted callers for PGx can be enabled together using: `--enable-pgx=true` (enables both start allele and targeted callers)
  - `--enable-pgx` will enable the following:
    - Star allele caller, which calls star alleles for 26 genes, including metabolism status of 21 genes.
    - Targeted callers for CYP2D6 and CYP2B6, which call star alleles and metabolism status for these difficult genes.
    - Targeted caller for HLA, which outputs the genotype.
  - PGx support summary

Caller	Gene	New with v4.2	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	Y
	ABCG2	Y	
	NAT2 *	Y	
	F5 *	Y	Y
	UGT2B17 *	Y	
Targeted	CYP2D6		
	CYP2B6		
	HLA-A *		
	HLA-B *		

**Outputs genotype only*

Amplicon

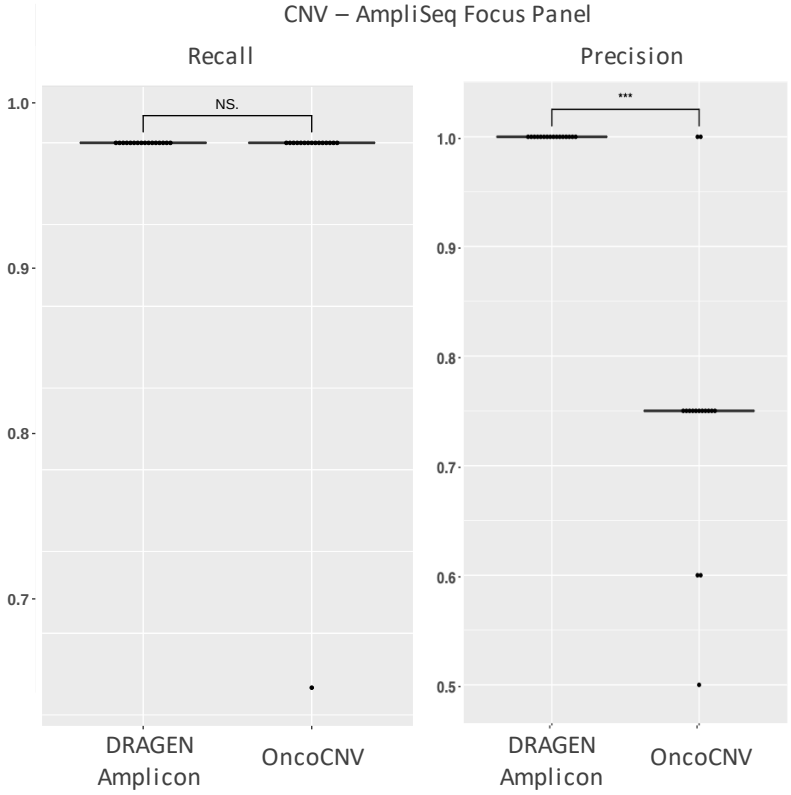
- Support for comprehensive amplicon analysis: SNV, RNA-fusion, CNV and SV.
- Amplicon CNV
  - Default to use bed segmentation (gene-level)
  - Auto-generate CNV segmentation bed based on the gene symbols at the fifth column of amplicon bed.
  - Usage:

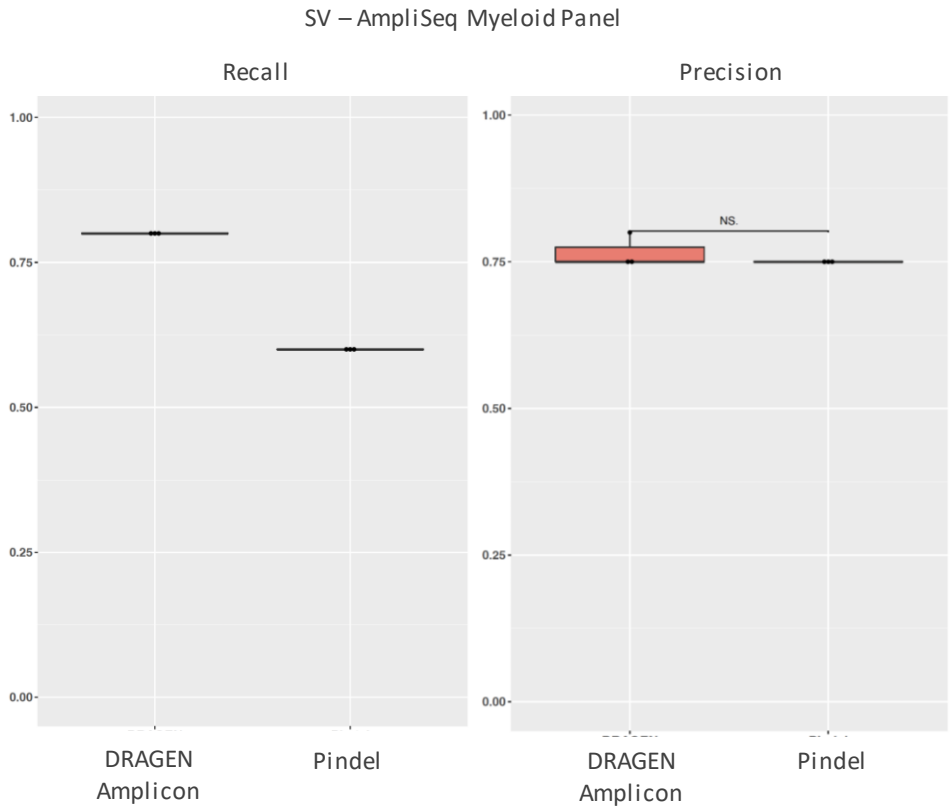
```
--enable-cnv=true
--enable-dna-amplicon=true
--amplicon-target-bed=<amplicon bed>
```

- Amplicon SV
  - Minimum candidate variant size is 10.
  - Increased sensitivity in hotspot regions
  - Usage:

```
--enable-sv=true
--enable-dna-amplicon=true
--amplicon-target-bed=<amplicon bed>
```

- DRAGEN™ Amplicon Outperforms OncoCNV and Pindel


CNV – AmpliSeq Focus Panel
Recall
Precision
NS.
***
1.0
0.9
0.8
0.7
0.6
0.5
DRAGEN Amplicon
OncoCNV
DRAGEN Amplicon
OncoCNV



RNA

- Metrics enhancements
  - Added mapping/aligning summary metrics to account for detected abundant sequences. New metrics are in `mapping_metrics.csv` file.
    - rRNA filtered metrics:
      ```
      Filtered rRNA read count and % of total reads:
      Filtered rRNA reads,n,%
      ```
    - Excluded ChrM read count and % of total reads:
      ```
      Mitochondrial reads excluded,n,%
      ```
    - Mapped reads adjusted by adding rRNA & ChrM counts:
      ```
      Mapped reads adjusted for filtered mapping,n,%
      Mapped reads adjusted for excluded mapping,n,%
      Mapped reads adjusted for filtered and excluded mapping,n,%
      ```
    - Unmapped reads adjusted by subtracting rRNA & ChrM counts:
      ```
      Unmapped reads adjusted for filtered mapping,n,%
      Unmapped reads adjusted for excluded mapping,n,%
      Unmapped reads adjusted for filtered and excluded mapping,n,%
      ```
    - Related command line options:
      rRNA metrics
      ```
      --rrna-filter-enable=true
      --rrna-filter-contig=CONTIG [optional]
      ```
      ChrM metrics
      ```
      --rna-mapping-metrics-exclude-chrm=true
      --rna-chrm-filter-contig=CONTIG [optional]
      ```

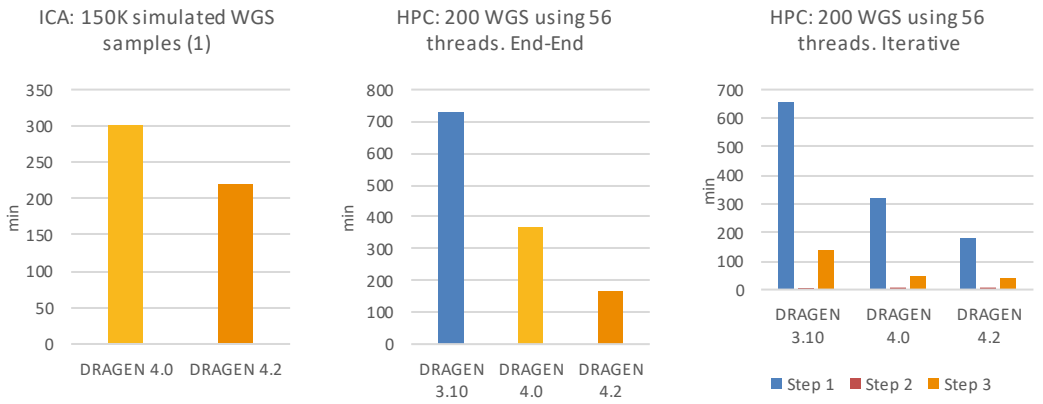
NOTE: If optional CONTIG name not specified, the default for current reference will be used if possible

- Added Quantification metrics to check RNA-seq coverage quality. New metrics below are in `quant_metrics.csv` file.
  - Gene counts (previously only had transcript counts)
    ```
    Total Genes
    Coding Genes
    Number of genes with coverage > 1x,n,%
    Number of genes with coverage > 10x,n,%
    Number of genes with coverage > 30x,n,%
    Number of genes with coverage > 100x,n,%
    ```
  - Fold coverage of genomic regions
    ```
    Fold coverage of all exons (i.e all biotypes)
    Fold coverage of introns
    Fold coverage of intergenic regions
    Fold coverage of coding exons
    ```
    (excludes non-coding genes and pseudogenes)
  - Transcript end-coverage biases (*)
    ```
    Median 5' coverage bias
    Median 3' coverage bias
    ```
  - Quantification metrics are output when `--enable-rna-quantification=true`
- Added Fusion and new Trimming metrics for functional QC. New metrics below are added to `fusion_metrics.csv` and `trimmer_metrics.csv` files respectively.
  - Gene fusion fusion statistics
    ```
    All fusion candidates (unfiltered)
    Final fusion candidates (passing filter)
    Unique passing gene fusions
    ```
  - Poly-A (X) soft-trimmer 5’ statistics (additional to 3’ stats)
    ```
    Poly-X soft trimmed reads unfiltered R1/R2 5prime,n,%
    Poly-X soft trimmed reads filtered R1/R2 5prime,n,%
    Poly-X soft trimmed bases unfiltered R1/R2 5prime,n,%
    Poly-X soft trimmed bases filtered R1/R2 5prime,n,%
    Poly-A (X) hard-trimmer 5’ statistics (additional to 3’ stats)
    Poly-X trimmed reads unfiltered R1/R2 5prime,n,%
    Poly-X trimmed reads filtered R1/R2 5prime,n,%
    Poly-X trimmed bases unfiltered R1/R2 5prime,n,%
    Poly-X trimmed bases filtered R1/R2 5prime,n,%
    ```
  - Fusion metrics are output when `--enable-rna-gene-fusion=true`, soft trimmer metrics are output when either soft or hard trimming on polyg and/or polya are enabled.

- New RNA VC options
  - `--rna-vc-enable-homozygous-genotype` Enable/disable homozygous genotypes (e.g., 1/1) in RNA variant calling (somatic pipeline, Default=true)
  - `--rna-vc-homozygous-genotype-af-threshold` Allele frequency threshold to rewrite genotypes to homozygous in RNA variant calling (Default=0.85)

Gvcf Genotyper

- Significant run time improvements, enables larger cohort analysis and reduces cost.
  - 2x faster for non-iterative small cohort (<10k samples) analysis
  - 20-30% faster for iterative large cohort (>10k samples) analysis
  - 8x faster for targeted region analysis (e.g., WES)



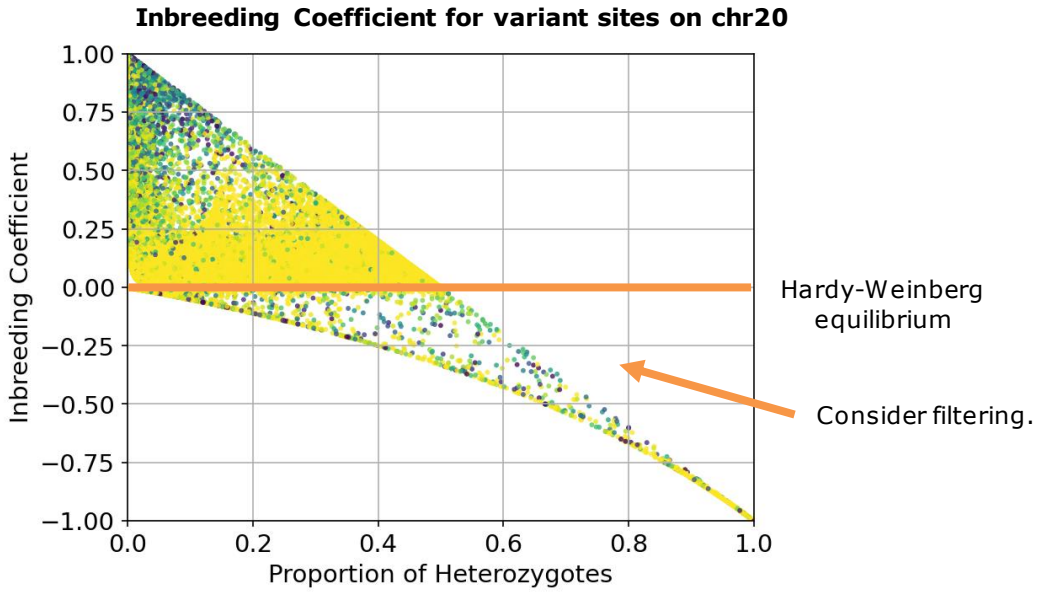
(1) Measured on Chr21 using 200 compute nodes and 150k simulated samples

Deployment options	Max cohort size	Description
Illumina Connected Analytics (ICA) *(1)*	>100K batch-wise	Using PopGen CLI *(2)* or Population Explorer WebApp. Example: 3K samples will take 1-2h using 900 ICA jobs and cost about 0.31 iCredits per sample.
Population Explorer *(1)*	>100K batch-wise	Population Explorer WebApp https://popex.dragen.illumina.com/ Input from ICA, BSSH or S3 buckets
DRAGEN™ server	Recommend for small cohorts only (10K samples hard limit), can run several batches in sequence.	~ 15h for 1K samples on a single server. (Requires fast I/O) Can distribute by genomic region using --shard command line option
HPC	>100K, batch-wise	HPC binary and template workflow available on request

(1) recommended and most scalable options for large cohorts.
(2) the PopGen CLI is available as a bundle in ICA.

- New input formats supported.
  - Multi-sample gVCF as written by DRAGEN™ Pedigree caller.
  - Support of GATK 4.0 gVCF input.
- Automatic renaming of duplicate samples.
- New QC metrics in the multi-sample VCF (msVCF) output
  - Inbreeding coefficient
    - Quantifies the excess heterozygosity at a genomic position. Strongly negative values can indicate read alignment artifacts.
    - Measures only heterozygous genotypes.
  - Hardy-Weinberg equilibrium and Excess Heterozygosity p-values
    - Test if the counts of called genotypes deviate from Hardy-Weinberg Equilibrium
    - Deviations from the expectation can indicate an artefact of variant calling Variant confidence score (QUAL) in msVCF.
    - Example INFO field in msVCF: (new fields in red)
      `AC=1,0;AN=2;NS=53;NS_GT=1;NS_NOGT=52;NS_NODATA=0;IC=-1;HWE=1,1;ExcHet=1,1;HWEc2=0.32`
  - Each of these metrics exists also in a "global" version. Example: INFO/GIC = inbreeding coefficient computed for the whole cohort (as opposed to only batch)

Name	msVCF field	Description
Inbreeding coefficient	INFO/IC	Negative values suggest an excess of heterozygous genotype calls, symptomatic of poor variant calling
Hardy-Weinberg Chi^2	INFO/HWEc2	Pearson's chi-squared method to test the distribution of heterozygote and homozygote genotype calls
Hardy-Weinberg p-value	INFO/HWE	Tests if the distribution of heterozygotes and homozygotes is close to that expected under Hardy-Weinberg
Excess Heterozygosity	INFO/ExcHet	Tests if number of heterozygotes is close to the number expected under HWE



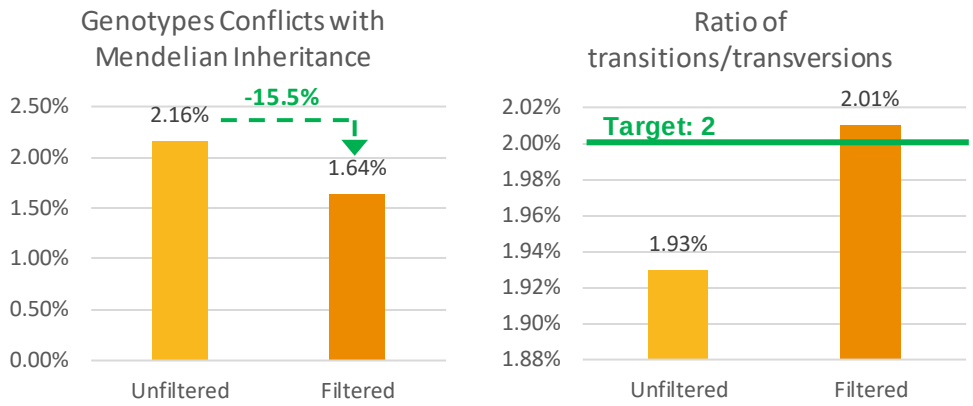
- Filtering of msVCF output
  - Filtering variants in msVCF output using user-defined expressions.
  - Filtering of inputs (QUAL and failed VC filters)
  - Usage:
    - Do not write filtered variant sites to msVCF: `--gg-skip-filtered-sites=true` (default=false)
    - Conditions to filter in the output msVCF:

```
--gg-hard-filter="InbreedingFail:all:GIC<0"
```

Applying filter tag

Filtering predicate

Filter applies to all variants.



Sample dataset of 50 samples from the 1,000-genome cohort.
Inbreeding coefficient <0.3 and where fewer than 90% of the samples had a called genotype.

- New options supported:
  - Input BED file containing regions to limit processing to. Only applicable when reading gVCFs: `--gg-regions-bed=<BED>`

Population Haplotyping (Beta)

- v4.2 implements a **beta** version of a Population Haplotyping tool. This tool supports the estimation of haplotypes from a population scale dataset via the packaging of the *SHAPEIT5* Software (*2022, Hofmeister RJ, Ribeiro DM, Rubinacci S., Delaneau O*).
- It is designed to phase common variants as well as rare variants in a step-by-step mode.
- A common use case of the Population Haplotyping tool is the generation of a custom reference panel to be used for the VCF Imputation pipeline.
- The tool supports autosomes and mixed ploidy chromosomes for diploid species only.
- Please see the User Guide for extensive details on usage and outputs.
- Known limitations.
  - The phase common step has a run-run variation in output in this Beta version.
  - Haplotyping of regions with no variants will crash.

Imputation

- Now supports haploid species, haploid/diploid chromosomes.
  - VCF imputation is enabled for haploid and diploid species on mixed ploidy chromosomes, including human sex chromosomes.
  - Added contig file that defines regions of mixed ploidy to impute variants in:
    - Haploid species
    - Diploid species
    - Human sex chromosomes
  - Falls back to a default all diploid setting for all chromosomes when config file is not present.
  - Example command line options to impute chrX nonPAR region:

```
--imputation-chunk-input-region <chrX_nonpar>
--imputation-phase-sample-type-list <sample_type_file.txt>
```

- Optionally disable inclusion of input samples to reference panel

- Samples can be imputed independently without contributing to the reference panel calculation causing batch effects.
- Usage:
  - Set `--imputation-phase-input-independently=true`. (Default is false).

- Imputation Reference Panel (IRP) updates
  - IRPv1.2 updates the initial release IRPv1.1, and corrects the following issues:
    - Added ##contig "length" field in the header.
    - Added the `IRPv1.2.forcegt.sites.all.vcf.gz` file that contains all the positions present in the reference panel IRPv1.2. It is recommended to use force genotyped input for imputation generated during variant calling using that file.
  - IRPv2 is a new reference panel that offers improved accuracy.
    - Multi-allelic SNPs positions added.
    - INDELs with AF>3% added.
  - The pre-built reference panels are available for download at the Illumina DRAGEN™ Product Files support site.

	IRPv2	**IRPv1**
Data source[1]	3,202 samples	2,504 samples
Multi-allelic SNP positions	Yes	No
INDELs[2]	Yes	No
Total Number of Variants	125,715,255	49,493,544

(1) Data from 1000 Genome Project, processed with Gvcf Genotyper
(2) Imputation covers indels with AF > 3%

ORA Compression

- Reduced runtime for ORA compression with map/align.
  - Speed-up 15-30% by executing ORA compression in parallel with mapping/aligning step. Both can now be enabled at the same time.
  - Output aligned reads (BAM/CRAM) and FASTQ.ora compressed files from the same run, starting from FASTQ input.
  - Requires compression add-on license. The compression license gets deducted at the same time as the Genome license.

	Pipeline	**Runtime**
DRAGEN 4.0	ORA compression standalone	520 Sec
	map/align phase	890 Sec
DRAGEN 4.2	ORA compression + map/align	**975 Sec**

- ORA Helper Suite for seamless integration of ORA compressed FASTQ
  - ORA Helper Suite makes FASTQ.ora compatible with 3rd party tools.
  - Available for download at the Illumina DRAGEN ORA support site https://support.illumina.com/sequencing/sequencing_software/DRAGENORA.html

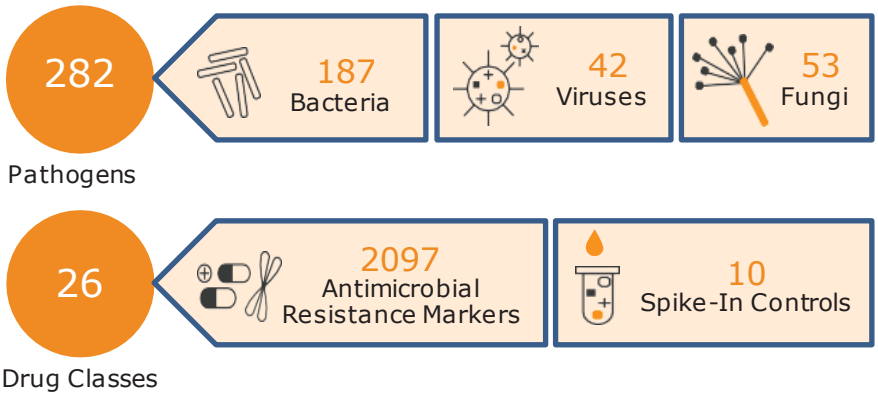
Precision Metagenomics Pipeline and Tools

- v4.2 introduces a new pipeline and tools for precision metagenomics to support infectious disease applications.
- The solution integrates the secondary analysis tools from of the Illumina Explify Software
- Two tools are supported:
  - Pipelines for analysis of data from the precision panels
    - Respiratory Pathogen ID/AMR Enrichment Kit (RPIP),

        - Urinary Pathogen ID/AMR Enrichment Panel Kit (UPIP).
    - A k-mer classification standalone tool.

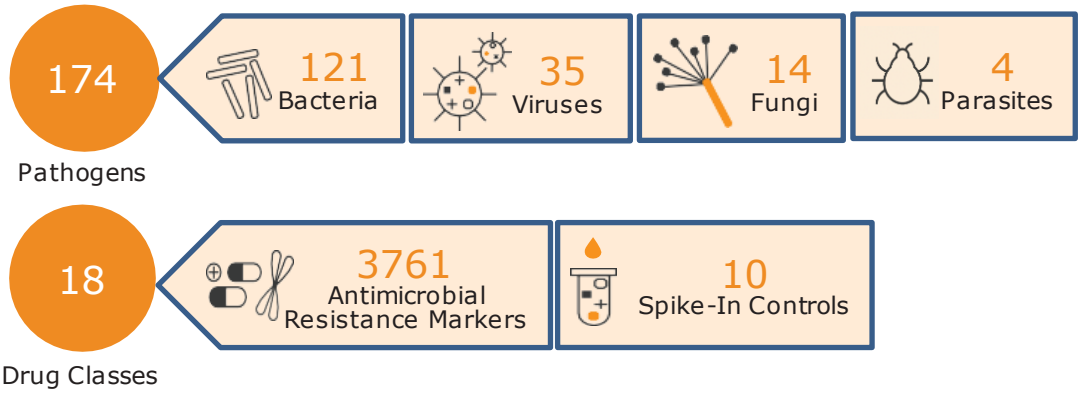
- **Pipelines for human pathogen detection and surveillance**
    - The pipeline applies a precision metagenomics approach, where probes are used to preferentially select specific fragments of DNA / RNA. This allows for greater organism detection sensitivity in complex sample types.
        - Detection of hundreds of pathogenic organisms.
        - Detection and assembled sequences for thousands of AMR markers.
        - Complete, assembled viral genomes for detected virus.
        - Sample composition.

    - RPIP Panel Overview
        - Comprehensive and simultaneous sequencing and quantification of
            - >280 respiratory pathogens and
            - >2,000 AMR markers
        - Target capture of known and emerging pathogens causing respiratory tract infections
        - Whole genome sequencing (WGS) of SARS-CoV-2 and Influenza A subtypes, Influenza B and Influenza C
        - Easily deployable in laboratory



    - RPIP Data Analysis Example:
        - `dragen --enable-explify=true --explify-sample-list /sample/list/tsv --explify-test-panel-name RPIP --explify-test-panel-version <VERSION> --explify-ref-db-dir /db/dir --explify-load-db-ram=true --output-file-prefix <PREFIX> --output-directory <OUTPUT_DIR> --intermediate-results-dir <OUTPUT_DIR> --explify-ncpus=32`
        - See the User Guide for comprehensive details.

    - UPIP Panel Overview
        - Comprehensive and simultaneous sequencing and quantification of
            - >170 genitourinary pathogens and
            - >3,700 AMR markers
        - Target capture of pathogens causing complicated and recurrent urinary tract infections (UTIs), sexually transmitted microorganisms, and fastidious, slow-growing, anaerobic uropathogens.
        - Easily deployable in laboratory

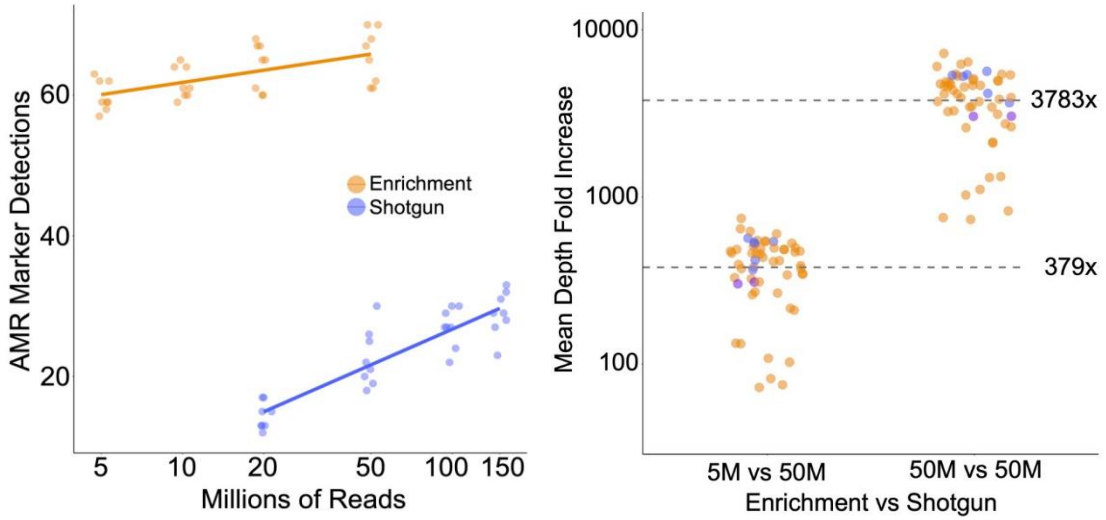


- UPIP Data Analysis Example:
  - `dragen --enable-explify=true --explify-sample-list /sample/list/tsv --explify-test-panel-name UPIP --explify-test-panel-version <VERSION> --explify-ref-db-dir /db/dir --explify-load-db-ram=true --output-file-prefix <PREFIX> --output-directory <OUTPUT_DIR> --intermediate-results-dir <OUTPUT_DIR> --explify-ncpus=32`
  - See the User Guide for comprehensive details.

- Pipeline input databases
  - Reference Databases are required to run the analysis pipeline in DRAGEN™. They are stored remotely and must be downloaded prior to running an analysis.
  - A database download shell script is provided to facilitate the downloads. The script can be downloaded with the following command:
    - `wget -O explify-dbs.sh https://illumina-explify-databases.s3.us-east-1.amazonaws.com/explify-dbs.sh`
    - See the User Guide for comprehensive details on usage.

- Pipeline output
  - The output of the analysis pipeline is a single raw results `report.json` of detection results written to the specified output directory. It contains below threshold targeted microorganism & AMR marker detection results.
  - Additional metrics / interpretation information such as aligned read count to targeted regions.
  - Coverage vectors for read alignments to organism targeted regions, viral full genomes (RPIP only), and AMR genes (UPIP only).
  - See the User Guide for a comprehensive description of the output.

- RPIP/UPIP Bacterial AMR Surveillance: Example
  - UPIP probes target 75% of the bacterial AMR marker Reference Sequences in The Comprehensive Antibiotic Resistance Database (CARD)
  - Gene presence-based and point mutation-based AMR markers are both reported.
  - For mutations in endogenous bacterial genes that are important in drug resistance, probes are designed to cover critical genomic regions that include specific point mutations.
  - Wastewater AMR surveillance study: Shotgun vs UPIP enrichment:



- **k-mer Classification standalone tool.**
  - The metagenomics classifier uses a k-mer based classification algorithm to classify each query sequence (usually a read) against a collection of reference sequences. There are two logical steps to this process: 1) reference sequences need to be indexed into a searchable database 2) reference sequence database will be searched using query sequences and classified to taxid(s) associated with the reference sequences.
  - The flexible feature set accommodates a wide variety of metagenomic analysis needs.
  - Drop-in replacement for tools like Kraken, but with additional features.
  - Usage example:

```
dragen --enable-kmer-classifier=true \
  --kmer-classifier-input-read-file /path/to/fastq.gz \
  --kmer-classifier-db-file /path/to/database \
  --kmer-classifier-min-window 1 \
  --kmer-classifier-ncpus=2 \
  --kmer-classifier-output-read-seq=false \
  --kmer-classifier-output-taxid-seq=false \
  --output-file-prefix <PREFIX> \
  --output-directory <OUTPUT_DIR>
```

  - See the User Guide for a comprehensive description of the available options and outputs.

- Notes about differences between the DRAGEN™ v4.2 tool and other platforms
  - The Explify Analysis Pipelines for RPIP and UPIP are available in BaseSpace.
  - The following inputs available in the BaseSpace applications will not be available in the v4.2 implementation:
    - The "Report AMR Only With Pathogen" option will not be available in v4.2. (It defaults to False).
    - The "Microorganism Reporting List" selector will not be available in v4.2 (It defaults to "All").
  - The v4.2 JSON output is similar to, but not exactly the same as the report JSONs available in the BaseSpace applications. In general, the DRAGEN™-based JSON will have more information. In contrast with the BaseSpace apps, there will be no PDF, no consensus fasta files, or variant TSV file. All information is included in the JSON.
  - Some items currently available in BaseSpace reports will not be available in the v4.2 implementation, including the pangolin lineage, some variant capability, and metadata that falls under tertiary analysis (e.g., drug names).

BCL

- Support for the output format of legacy *bcl2fastq2* stats, using the option `--output-legacy-stats=true` (disabled by default). Enabling the option will produce the following files in the `<output directory>/Reports` folder:
  - DemultiplexingStats.xml
  - ConversionStats.xml
  - AdapterTrimming.txt
  - FastqSummaryF#L#.txt
  - DemuxSummaryF#L#.txt
  - Stats.json
- Support both 1 / 0 and True / False for command-line options and sample-sheet settings, for the `TrimUMI` and `CreateFastqForIndexReads` settings to improve backwards compatibility with *bcl2fastq2*.
- Per sample settings introduced with the NovaSeq-X instrument
  - DRAGEN and bcl-convert v4.1 and later support the following settings as columns in the `[BCLConvert_Data]` section, allowing them to be specified differently for each sample: `OverrideCycles`, `BarcodeMismatchesIndex1`, `BarcodeMismatchesIndex2`, `AdapterRead1`, `AdapterRead2`, `AdapterBehavior`, `AdapterStringency`.
  - These per-sample settings can be specified by omitting the setting from the `[BCLConvert_Settings]` section and instead adding a column to the `[BCLConvert_Data]` section with that setting name. Settings that do not apply to a sample (e.g., 'index2' if i5 is masked out for that sample) must be blank or 'na' in the entry for that sample.
  - This feature is only supported on version two (v2) sample sheets, and no setting can be specified both globally and per-sample. Specifying `OverrideCycles` differently per-sample allows mixing of different pools into the same lane but must still obey barcode mismatch constraints for all cycles that are used for demultiplexing by any sample in that lane.
  - The software will detect all conflicts between samples at the beginning of the conversion run, even between different pools.
  - Different strategies such as UMI indexes and dual-index inputs can be combined, provided `IndependentIndexCollisionCheck` is not enabled.
- Make the combined index collision checking default to enabled for all lanes. Implement a new `IndependentIndexCollisionCheck` option to replace `CombinedIndexCollisionCheck`.
  - This important change reverts a strict check on dual index collisions added to BCL based on customer feedback. With this change, the default behavior matches bcl2fastq2 and adds an option to change the behavior.

DRAGEN™ version	Index collision check behavior
3.9.x	Relaxed by default. No option to change. Matches bcl2fastq2
3.10.x and 4.0.x	Strict by default. No option to change.
4.1.5	Strict by default. New option `CombinedIndexCollisionCheck` introduced to optionally relax the strictness
4.1.7 and 4.2.x	Relaxed by default. Remove `CombinedIndexCollisionCheck` option, add new `IndependentIndexCollisionCheck` option to allow optional strict checking. Default matches bcl2fastq2

- Various BCL issue fixes
  - Fix for index sequences missing from fastq headers when using `--no-sample-sheet` setting.
  - Fix for BCL behavior being different than bcl2fastq2 with respect to "`Sample_Name`" and "`Sample_Project`". In the special case of "`Sample_Name`" == "`Sample_ID`", bcl2fastq2

does not create a "`Sample_ID`" subdirectory. This change makes bcl-convert behavior the same.
  - Fix for false barcode collision reports when one sample's index is entirely trimmed out and another sample's index exists.
  - Fix for BCL not aborting when single-index datasets have barcode collisions.
  - Fix for incorrect yieldQ30/qscoresum stats when there is UMI in the first part of a read and TrimUMI is enabled (true by default).
  - Fix for a false error when using global BarcodeMismatchesIndex2 and a sample does not use index BarcodeMismatchesIndex1, when the sample sheet contains both single & dual-index samples.
  - Fix for BCL failing with a "`vector::reserve`" message for mixed index strategies.
  - Fix for BCL outputting many duplicate error messages for missing CBCL files.
  - Fix for BCL convert abort with an empty [BCLConvert_Settings] section in v2 sample sheets.
  - Fix an issue where Undetermined FASTQ files are still created even after setting `--bcl-only-matched-reads` to true.
  - Fix for failing validation check "No more than 27 total bases can be used as index bases".

CheckFingerprint

- CheckFingerprint outputs a LOD score to indicate whether all the genetic data between two files are from the same individual or not.
- The feature is broadly based on Picard CheckFingerprint
- Usage:
  - Enable CheckFingerprint output: `--enable-checkfingerprint=true`
  - Expected genotypes from VCF for comparison: `--checkfingerprint-expected-vcf=<VCF>`
  - `--checkfingerprint-enable-vcf-comparison=<VCF>`
    - When `true`, expected VCF must be used as input sample for comparison at each variant site.
    - When `false`, reads are used for comparison.

Multigenome Reference Builder

- The software can prepare a custom multigenome hash table from a set of population VCFs. Please see the User Guide for full details. Updates to the tool is summarized below.
- Added checking on inputs to avoid errors. Inputs must:
  - Include diploid GT calls.
  - Be in positionally sorted order.
  - FORMAT fields shall include GT, GQ, DP, MIN_DP, AD, VAF, PL. FORMAT phase set field ID shall be "PS".
  - No phased records detected in input VCF.
- Output hash table statistics in a .csv file `multigenome_reference_metrics.csv` to provide useful statistics information on the variants used to build the custom multigenome hash table.
- Support for CHM13 reference added.

Other Updates

- Software for DRAGEN™ server has been re-architected.

  - Communication with the FPGA card is now handled by a daemon process called `dragend`
  - `dragend` runs as a Linux system service. It is registered and started automatically when the DRAGEN™ server software is installed. The `dragen` application communicates with the daemon to access the FPGA card.
  - `$ ps aux | grep dragen` shows the new process. This process must run for dragen to function correctly:

```
/opt/edico/bin/dragend_hp -d
/opt/edico/bin/dragend -d
/opt/edico/bin/dragen_licd -d everyday -s 1800 -e 0800 -r 5 -z
```

  - **No new user actions or interactions are required for using dragen or switching between versions. This note is for information only.**
    - There are some behavioral changes:
    - `$ dragen -V` and dragen stdout will now return the client version of the `dragen` application as well as the daemon version of `dragend`

```
dragen Version 4.2.4
dragend Version 07.031.676.4.2.4
```

    - `dragen_reset` for cloud is unchanged.
    - `dragen_reset` on-server is now automatic, and no longer need to be called. For backwards compatibility with existing user scripts, tt can be called and will return 0 (without error).
- RPMs now require platform-python. Users on Centos7 whose version of the python RPM is lower than 2.7.5-92.el7_9 will encounter installation failure due to /usr/libexec/platform-python not being visible to the RPM database. Users can resolve this issue by upgrading python with `yum install python`
- HLA: Class II typing can be enabled in addition to Class I HLA typing, by setting `--hla-enable-class-2=true` (Default false)
- Cloud licensing: Credentials can be passed to the software via a separate config file using `--lic-credentials=<FILE>`, to avoid any logging of the command line in a shell environment. The option takes a path to the file.
- Down-sampler:
  - FASTQ output option is renamed from `--enable-down-sampler-output` to `--enable-down-sampler-fastq`, to reflect that the setting enables only down sampled FASTQ output.
  - Target setting method has been updated to use a quantity of fragments instead of reads. The option `--down-sampler-reads` has been replaced by `--down-sampler-fragments` to make it clearer.
- Hash Table Builder: When running from an input fasta that contains alt contig with standard names, but custom user modified base content, it is recommended to suppress automatic masking by setting `ht-suppress-mask=true` or by passing a custom mask bed file to `ht-mask-bed` option.
- Systematic Noise Builder: Added options to refine.
  - `--build-sys-noise-min-sample-cov` Min coverage at a site for a sample to be used towards noise estimation. At low coverages estimated allele frequencies become less reliable, but low coverage sites also tend to be noisy and useful for inclusion in the noise file.
  - `--build-sys-noise-min-supporting-samples` Min number of samples with noise at a position for a position to be considered systematic-noise.
- TMB: A reminder that in v3.10 we renamed a setting `tmb-skip-proxi-filter` to `tmb-enable-proxi-filter`. In both cases the default value is false. So effectively (by default) we disabled this feature in v3.10 onwards. In T/O mode the DB germline filter may not be able to detect all germline variants, especially for ethnicity groups that are not well represented in germline databases. The proxi filter uses allele frequency information to help remove germline variants missed by the DB and can help to obtain more accurate (lower) TMB values on samples with low tumor purity.

Issues Resolved

Issues found on DRAGEN™ v4.0.3 or older that are fixed in v4.2.4

Component	Defect ID	Issue Description
Azure cloud	DRAGEN-21794	Fix a race condition in XRT driver causing incorrect 64-bit reads, leading to incorrect FASTQC metrics on Azure cloud platforms
BCL	DRAGEN-18636	Fix for bcl-convert legacy stats output differing between default and 'gzip' setting of FastqCompressionFormat
BCL	DRAGEN-19157	Fix issue where filenames for interleaved fastqs that are Ora compressed, are not the same as the original file names. Fix preserves the lane identifier e.g.,"R_1.fastq" to be "R1_001.fastq".
BCL	DRAGEN-19292	Fix BCL FASTQ file paths in fastq_list.csv when ORA-interleaved compression format is used. Previously the fastq_list.csv file contained two files (instead of one single interleaved file) under the "Read1File" and "Read2File" columns and the files were not named correctly.
BCL	DRAGEN-19296	BCL outputs many duplicate error messages for missing CBCL files
BCL	DRAGEN-19376	A crash can occur when using per-sample-settings with higher sample counts in a lane due to a hash-table pre-size using a signed integer as input that is overflowed. The fix is to use a larger integer size in the pre-size calculation.
BCL	DRAGEN-19496	Fix a hang when "--output-legacy-stats" enabled when running BCL with maximum allowed samples
BCL	DRAGEN-20064	Fix for BCL convert abort with an empty [BCLConvert_Settings] section in v2 sample sheets
BCL	DRAGEN-20310	Fix for Index_Hopping_Counts.csv containing incorrect Same_{Name,Project}
BCL	DRAGEN-20513	Log clearer error messages when BCL errors on per-sample-settings
BCL	DRAGEN-21299, SET-4248	Fix an issue where Undetermined FASTQ files are still created even after setting --bcl-only-matched-reads to true.
BCL	DRAGEN-21443	Some versions of RTA3 outputs cbcl with 0 qual + non ero base for 0/”N#”, in other words masked nibbles (0 qual) does not zero out the base. Bcl2fastq2 has a masking step, but DRAGEN and bcl-convert does not. This fix adds masking so that bcl-convert matches bcl2fastq2 for those RTA3 outputs.
BCL	DRAGEN-21718	Fix for BCL crashes with threading error on 150k sample dataset

BCL	DRAGEN-22480	Fix to remove BCL conversion thread settings limit of 64, a regression in v3.10
BCL	DRAGEN-22935	BCL Sample_Name + Sample_Project behavior differs from bcl2fastq2
BCL	DRAGEN-23363	Fix for no-sample-sheet setting omits index sequences from fastq headers
BCL	DRAGEN-23388	Fix for BCL crash on contrived corner case: --no-sample-sheet true & 0 indexes
BCL	DRAGEN-23825	Fix to place FASTQ.ora files in Sample_Nume subdirectories
BCL	DRAGEN-25409	Fix for incorrect yieldQ30/qscoresum stats when UMI in first part of a read
BCL	DRAGEN-25718	Fix for failing validation check "No more than 27 total bases can be used as index bases"
Biomarkers	DRAGEN-22110	Add error when TMB is provided with invalid Nirvana JSON, instead of completing with no output.
Build, CI/CD	DRAGEN-23075	Enable el8 package builds for all platforms and variants (cloud platforms), so that they are available
CNV VC	DRAGEN-18956	In CNV VAF modeling, fix estimation of scale factor in presence of outliers
CNV VC	DRAGEN-19836, SET-3816	Skip unmatched gender removal process when "cnv-enable-gender-matched-pon=false" with "cnv-enable-linear-regression-noise-removal=true" is set.
CNV VC	DRAGEN-23857	Remove unwanted assert in options parser, to avoid putting system in state where reset is required
CNV VC	DRAGEN-24599	Fix for CNV assertion when sample coverage is low, on NTC samples
CNV VC	DRAGEN-25041	Updated threshold to detect sex genotyper appropriately for samples with chrY deletion.
Compression	DRAGEN-14390	Fix a segfault when CRAM input is processed with mismatching reference
DNA Alignment	DRAGEN-19361	Fix for potential crash in read_group_list, caused by timing race condition
DNA Mapper	DRAGEN-23631	Fix for incorrect CIGAR string produced by mapper, leading to crash in Variant Caller.

Down-sampling	DRAGEN-18531	Fix crash in HLA typer when downsampling enabled
Down-sampling	DRAGEN-20563	Fix hang when downsampling to zero reads
DRAGEN Apps	DRAGEN-21912	Fix Segmentation fault in tumor Only SV caller due to long assembly size causing 32bit integer overflow
Gene Fusion	DRAGEN-19957	Fix for very long run time and high memory usage on very large Amplicon RNA samples (10x larger than expected)
Gene Fusion	DRAGEN-20991	Add mitochondrial genes filter to fusion VCF header
Gene Fusion	DRAGEN-20998	Make Gene fusion VCF report PR:SR reads similar to DRAGEN SV
Gene Fusion	DRAGEN-21241	Fix for Gene fusion VCF qual score of "inf"
Gene Fusion	DRAGEN-23858	Fix RNA Gene Fusion run-run variation in candidates.features.csv output.
GVCF Genotyper	DRAGEN-19853	Fix an issue where the output has duplicate FILTER entries for chrM, due to ALT alleles at the same position on different gVCF lines.
GVCF Genotyper	DRAGEN-20094	Fix issue where records are dropped if an input file contains calls for same location on separate lines, common in chrM
GVCF Genotyper	DRAGEN-21282	Fix segfault on GATK gVCF inputs
GVCF Genotyper	DRAGEN-21458	Fix a memory leak in VCF file reading
GVCF Genotyper	DRAGEN-21775	Fix for unnormalized ALT alleles in the msVCF output, when "--gg-discard-ac-zero=true" in Iterative Gvcf Genotyper
GVCF Genotyper	DRAGEN-21922	Fix for incorrect LPL and LAA values in msVCF
GVCF Genotyper	DRAGEN-22446	Fix for unnormalized ALT alleles in the msVCF output, when "--gg-discard-ac-zero=true" in legacy Gvcf Genotyper
GVCF Genotyper	DRAGEN-23457	Fix to trim remaining alleles after AC=0 removal, if necessary. Enable trimming on lone REF allele after AC=0 discard
GVCF Genotyper	DRAGEN-24932	Fix incorrect processing of non-ref allele counts and frequencies when global ref allele is different from batch ref allele

GVCF Genotyper	DRAGEN-25395	Fix crash in Gvcf Genotyper when intermediate-results-dir option is set
GVCF Genotyper	DRAGEN-25621	Input gVCF files generated by DRAGEN versions prior to v3.5 were missing an AF value corresponding to the NON_REF allele. This caused an error in processing of PL arrays of unequal length, leading to inconsistent values written to the msVCF output.
Hash Table Builder	DRAGEN-23206	Fix non-deterministic output in the custom genome builder, caused by sort order variability.
HLA	DRAGEN-19524	Fix for incorrect HLA genotyping output format when minor allele has insufficient support
HLA	DRAGEN-25237	Fix crash in downsampler when HLA is enabled
Indel Realignment	DRAGEN-21404, SET-4227	Fix indel re-aligner crash due to mismatch in number of realigned reads.
Infrastructure	DRAGEN-11011	Make dragen more robust to incomplete command lines
Infrastructure	DRAGEN-20550	Excessive watchdog logs fills up /var/ partition
Infrastructure	DRAGEN-25897	Send periodic watchdog logs only to the syslog
Installer	DRAGEN-22634	Fix for EL8 DKMS 3.0 breaking networking on boot
Installer	DRAGEN-23268, SET-4412	Fix duplicate logrotate.d definition for NextSeq1k2k on-instrument /var/log/dlm.log rotation.
Joint Calling	DRAGEN-24805	Update qc-indel-denovo-quality-threashold for ML from 0.03 to 0.04
Joint Genotyping	DRAGEN-23609	Fix some VCF/gVCF discrepancies when ML is enabled. Re-compute GQ to be used as an MLfeature in gVCF mode. Fix PL and GP update for 0/0 calls
Joint Genotyping	DRAGEN-24604	Improve denovo SNV INDEL performance. Restrict use of genotyper PLs in pedigree caller
Methyl-Seq	DRAGEN-24516	Fix an issue where Methyl CX report file may contain trailing data from prior run, when same output folder is used.
Paralog Caller	DRAGEN-24404	Fix to allow complete loss variant for CYP21A2
Paralog Caller	DRAGEN-24833	Fix for GBA regression for LB-01223 with map/align and CYP21A2 regression for homozygous recombinant variants

PhenoHRD	DRAGEN-20471	Report QC failure rather than error for high tumor fraction LOH
PhenoHRD	DRAGEN-20639	Fix issue that required Genome license for PhenoHRD in TSO500 analysis
PhenoHRD	DRAGEN-20937	Fix for PhenoHRD outputting TSV files with CSV suffix in TSO500 analysis
PhenoHRD	DRAGEN-22810	Fix for PhenoHRD fatal error when running on NTC samples in TSO500 analysis
QC Metrics	DRAGEN-21857	An invalid check for 10 required columns for the --qc-cross-cont-vcf header leads to an exception. Fixed the check to require 8 columns. Also improved error handling for invalid qc-cross-cont-vcf input, with clearer messages.
QC Metrics	DRAGEN-21921	Fix for mapper metrics being double counted when HLA is enabled.
scATAC	DRAGEN-22854, SET-4565	Fix for Feature/Peak ID missing in scRNA/scATAC output
scATAC	DRAGEN-23486, SET-4836	scATAC produces empty outputs (barcode list, matrix) when using combinatorial barcodes
scRNA	DRAGEN-20879	Fix for Single cell scRNA.barcodeSummary.csv reporting incorrect UMI counts
scRNA	DRAGEN-24741	Fix to allow pre-signed URLs for UMI files in FASTQ list.csv
SNV Germline	DRAGEN-19863	New ML model fixes an ML accuracy regression vs non-ML, observed on some test samples.
SNV Germline	DRAGEN-23028	Fix Evidence BAM output failing when using UL Streaming on AWS
SNV Germline	DRAGEN-23390	Fix for Incorrect PL values in SNV VCF/gVCF when ML is enabled. hethom calls where ML prediction does not match VC call, the computation of PL from GP and PRI is missing. This leads to a regression in the accuracy when looking at the single sample in a joint called msVCF.
SNV Germline	DRAGEN-22600, SET-3794	Fix for missing SNV call on chrX, via sex chromosome mosaic variants support added in v4.2.
SNV Somatic	DRAGEN-18660	Fix joint genotyping mito variant always PASSing

SNV Somatic	DRAGEN-18823	Fix to correctly recompose phased complex variants in the VCF
SNV Somatic	DRAGEN-18826	Tumor Only analysis with GVCF mode + TMB + germline filtering aborts in BSSH with memory issue.
SNV Somatic	DRAGEN-19052	Improve run time in Somatic SNV with NTD error estimation enabled
SNV Somatic	DRAGEN-19218	Improve elevated INDEL FP+FN seen in v4.0.3 compared to v3.10. SNP FP is rescued by up to 3% and INDEL FP is reduced by up to 7%.
SNV Somatic	DRAGEN-19234, SET-4548	Fix for issue where some variants are not emitted when evidence BAM is enabled.
SNV Somatic	DRAGEN-19721	Fixed an issue where unphased SNVs that are part of the MNV are dropped from filtered VCF
SNV Somatic	DRAGEN-22216, SET-4433	Fix an issue where a somatic SNV was missed due to a bias where fwd reads have a systematically higher BQ than reverse reads. DRAGEN already excludes agreeing overlapping mate pairs from the strand bias model when there is no clear preference between the reads. The fix is to extend the exclusion to cover all agreeing overlapping mate pairs, even when there is a clear preference for one of the mates due to higher BQ.
SNV Somatic	DRAGEN-22828, SET-4597	Fix mapper slow down due to snperror estimation when Somatic SNV is enabled.
SNV VC	DRAGEN-20271	Add additional checks for reads with cigar length of 0
SNV VC	DRAGEN-22841	Fix an issue where MNVs are wrong when merging distance is greater than graph tlen
SNV VC	DRAGEN-23414	Fix for VC hard filtering: "!=" comparator operator not working. This is now used with Gvcf Genotyper filtering.
SNV VC	DRAGEN-23494	Update (lower) amount of bin-memory RAM used by Evidence BAM output, to prevent OOM
SNV VC	DRAGEN-24547	Fix bug introduced with Evidence BAM change. Fix FGT bug where all reads are disqualified incorrectly in regions with FGT only events
SNV VC	DRAGEN-21052, SET-4106	Fix a watchdog hang observed when using certain bin_memory settings, but not others, due to a race condition in the order which regions were being freed.
SNV VC, ML	DRAGEN-18342	Handle 'N' bases correctly in ML processing, and check for mates

SNV VC, ML	DRAGEN-19386	Add ML files for hs37d5, adds ML accuracy improvements for hs37d5/hg19
SNV VC, ML	DRAGEN-19435	Fix overlapping mate handling in ML processing
SNV VC, ML	DRAGEN-19829	Fix run-to-run variation in SNV VCF/gVCF due to ML
SNV VC, ML	DRAGEN-18807	Add ML files for hs37d5, adds ML accuracy improvements for hs37d5/hg19
Somatic SNV	DRAGEN-21798	Fix for issue where MNV length overflows a variable, leading to a corrupted TAG and a downstream component (germline filter) that asserts.
Somatic SNV	DRAGEN-23644	Increase memory headroom in somatic mode by lowering maximum bin-memory usage from 100GB to 80GB
Star Allele Caller	DRAGEN-23886	Fix a run time regression with Star Allele caller
SV	DRAGEN-19631	Fix some FFPE samples where the sv.vcf.gz file contains only 1 entry for a MantaBND. All MantaBND events should have 2 entries.
SV	DRAGEN-20267	After the introduction of T/O scoring model and filters, it's possible that regions marked as "hot spots" are still being filtered out. Need to check this behavior and ensure hotspot feature still works for T/O workflow.
SV	DRAGEN-20389, SET-3913	Fix crash in SV caused by tiny candidate contig size that is generated in SV assembly stage
Systematic Noise	DRAGEN-24388, SET4848	Fix for OOM on cloud, by limiting the systematic noise to max 2 threads. This should never exceed 140GB memory consumption. Two minor updates to sys noise default settings for the VAF threshold and decimal precision. Based on new studies these settings slightly improve accuracy
TMB	DRAGEN-20793	Fix rounding of VAF being different between TMB trace and VCF/gVCF in TSO500 analysis
UMI	DRAGEN-19092	New option "umi-parse-only" to enable the UMI parser for regular Map/Align without UMI (enable-umi=false). If user specifies the "umi-source", it is saved to the output bam with RX tag.

Known Issues

Known issues of the DRAGEN™ v4.2.4 release

Component	Summary	Resolution/Workaround
BCL	bcl-convert does not output all FASTQs when CFFI is on, and some samples have fully masked indices	Corner case. If the same SampleID uses different library preps in the same lane, and per-sample-settings is used with different settings between them. Make fix to the sample sheet.
BCL	BCL aborts without an error message when --bcl-only-lane is set to a lane not included in the Sample Sheet or RunInfo.xml, instead of printing error message.	Occurs when user specifies a lane subset that does not exist (invalid). Make fix to the sample sheet.
BCL	BCL compute performance suffers for very high sample counts (150K)	No workaround. Fix planned for future version
Compr	CRAM decompress & map/align with different references, can falsely run into an alt contig error check and crash, when hash table is used for cram decompression.	Alt contigs are erroneously counted on both references and can exceed a threshold. Use fasta for CRAM decompression instead of hash table
Compr	Runs on Azure occasionally crash with "corrupted size" message after streaming of ora compression/decompression finishes.	No workaround
Down-sampling	Downsampling from BAM input has a chromosome coverage bias. This is not the case when the input is FASTQ. The average coverage is the same. This impacts accuracy when using BAM input and downsampling.	No workaround. Fix planned for future version
GVCF Genotyper	GG does contig name truncation on HLA* alt contigs to the first colon. This could lead to incorrect outputs for those contigs	It is a long-standing issue we are highlighting. No workaround. Fix planned for future version
GVCF Genotyper	When a site is missing in the input gVCF file for a sample and the site is output to the msVCF file, the genotype is coded as missing using '.', i.e., haploid	No workaround. Fix planned for future version
GVCF Genotyper	There are some additional FN indels in the msVCF that are not in the input gVCF, due to unnormalized indel variants for indels of certain type.	Very small % of indel FN affected. No workaround
GVCF Genotyper	Iterative GVCF Genotyper v4.2 will fail with GVCF inputs from pre-3.3 DRAGEN, due to catching a bug in the pre-v3.3 DRAGEN output GVCFs	No workaround.
Hash Table Builder	Hash table decompression error on some fasta input files during build.	Use option to write the hash table uncompressed. The uncompressed hash table is valid.
HLA Mapping	**Runs with HLA enabled** could end up with partially incorrect reference loaded for mapping after the run is finished. With back-to back runs, and without reference re-loading, the first mapping stage of subsequent runs fail to re-load the reference and lead to potentially incorrect mapper outputs.	HLA user will have to force reload the reference to guarantee correct mapping output.
Map/Align	**Runs with --pe-overhang-trimming enabled** may cause read pairs with PNEXT field of one mate disagreeing with the POS field of the other mate by a few bases, due to one of these two positions being adjusted by overhang clipping, and the other not	Such BAM files fail Picard validate. No workaround.

	adjusted. The alignments themselves are reasonable, with the POS consistent with CIGAR.	
Inputs	Map/align errors out if R2 FASTQ file contain more reads than R1 (not expected), but runs ok if R1 FASTQ contains more reads than R2 (expected)	Trim the FASTQ pair to contain the same number of reads.
Joint Genotyping	Higher number of denovo SNP calls observed in some trios.	Minor change. No workaround
Multigenome reference	WGS runtime increased with multigenome vs legacy genome	5% longer run time. For information only.
Paralog Caller	GBA reports a single recombinant haplotype with RecNciI+RecNciI instead of two recombinant haplotypes with RecNciI each for NA20273	No workaround
Population Haplotyping	Non-deterministic output. Different output VCF PREFIX.ph_phase_common.vcf.gz every time it is run.	Use --num-threads=1 for the phase common step
Population Haplotyping	Crash when a region contains no variants	Increase region size
RNA Quantification	RNA quant - SJ.saturation.txt has minor differences with different num-threads value	No workaround
RNA Gene Fusion	RNA GF output file fusion_candidates.features.csv has small and infrequent differences minor (in the thousandth digit, i.e., delta less than 0.01) between local and cloud platforms.	Differences should not lead to any difference in output of passing fusion calls. A borderline score of exactly 0.500 might get pushed down in cloud vs. local. No workaround
QC metrics	COVERAGE SUMMARY printout is missing in stdout for Germline	Coverage metrics are available in the output file(s).
SNV Germline	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, SNV Somatic, SNV VC	Phased calls with same PS and GT and within distance threshold are not getting combined into MNVs	No workaround
SNV Somatic	Minor increase (<3% change) in INDEL FP rate for somatic SNV on specific sample.	No workaround
SNV Somatic	Small INDEL FP regression across most T/O WGS and WES test datasets, with some WES datasets having larger FP regression (>5%)	No workaround
SNV Somatic	Some T/N and T/O samples have >5% runtime regression relative to v4.0.3	No workaround
SNV Somatic, MNV	Somatic SNV T/O MNV failing to merge two MNV calls, in the edge case where we have a deletion upstream of another co-phased variant with an out-of-phase SNP in between them that is covered by the REF allele of the upstream deletion.	No workaround
SNV VC	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. Pre-existing issue. No workaround

SNV VC	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	In rare cases some tabix files produced by DRAGEN can't be opened with htslib	No workaround

SW Installation Procedure

- Download the desired installer from the Illumina support website and unzip the package.
- The archive integrity can be checked using: `./<DRAGEN 4.2.4 .run file> --check`
    - For Centos7 users, a failed dependency on /usr/libexec/platform-python should be resolved by running `sudo yum install python`
- Install the appropriate release based on your Linux OS with the command: `sudo sh <DRAGEN 4.2.4 .run file>`

1. Release History

Revision	Release Reference	Originator	Description of Change
00	1088629	Cobus De Beer	Initial release
01	1090150	Cobus De Beer	Updated installer package names to reflect correct release version Updated header
02	1091016	Cobus De Beer	Additional python dependency notes