illumina®

DRAGEN for Illumina DNA Prep with Enrichment Dx on NextSeq 550Dx

Application User Guide

ILLUMINA PROPRIETARY
Document # 200025238 v00
February 2023

FOR IN VITRO DIAGNOSTIC USE.

Revision History

Document	Date	Description of Change
200025238 v00	February 2023	Initial release.

Table of Contents

Overview

The DRAGEN for Illumina DNA Prep with Enrichment Dx Application (DRAGEN for IDPE Dx) is used to plan and perform secondary analysis of IDPE Dx libraries generated for sequencing on the NextSeq 550Dx.

DRAGEN for IDPE Dx supports sequencing to analysis when used with the Illumina DNA Prep with Enrichment Dx Library Prep, NextSeq 550Dx, and Illumina DRAGEN Server for NextSeq 550Dx.

Analysis Methods

DRAGEN for IDPE Dx performs demultiplexing, FASTQ generation, read mapping, alignment to a reference genome, and small variant calling depending on the selected workflows:

- FASTQ generation
- Germline FASTQ and VCF generation
- Somatic FASTQ and VCF generation

NOTE ORA compression is available for use with all three workflows. DRAGEN ORA Compression is a fully lossless compression software that creates a file with an Original Read Archive (*.ora) extension. The ora format is a reference-based compression format for FASTQ files and is designed for very fast compression/decompression and high compression ratio.

FASTQ Generation

The assembled sequences are written to FASTQ files per sample. FASTQ files are text files that contain sequencing data and quality scores for only one sample. For each sample, separate FASTQ files are generated per flow cell lane, per sequencing read. The name of the sample as specified during run setup is included in the FASTQ file name. FASTQ files are the primary input for alignment. The first step of FASTQ generation is demultiplexing. Demultiplexing assigns clusters that pass filter to a sample by comparing each index read sequence to the index sequences specified for the run. No quality values are considered in this step. Index reads are identified using the following steps:

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

The software includes ORA compression to compress FASTQ files. This format can be optionally enabled. When using the ORA (*.ora) format, the md5 checksum of the FASTQ content is preserved after a compression and decompression cycle to ensure a lossless compression.

DNA Mapping & Alignment

After FASTQ generation, reads are mapped and aligned to a reference genome. The first stage of mapping is to generate seeds from the read, and then look for exact matches in the reference genome. These results are then refined by running full Smith-Waterman alignments on the locations with the highest density of seed matches. This well-documented algorithm works by comparing each position of the read against all the candidate positions of the reference. These comparisons correspond to a matrix of potential alignments between read and reference. For each of these candidate alignment positions, Smith-Waterman generates scores that are used to evaluate whether the best alignment passing through that matrix cell reaches it by a nucleotide match or mismatch (diagonal movement), a deletion (horizontal movement), or an insertion (vertical movement). A match between read and reference provides a bonus on the score, and a mismatch or indel imposes a penalty. The overall highest scoring path through the matrix is the alignment chosen. The algorithm is hardware accelerated on the DRAGEN field-programmable gate array (FPGA) cards. The reference genome used in the app is created from the UCSC hg19 FASTA with the DRAGEN option to create a liftover based alt-aware hash table.

DRAGEN Germline Variant Calling

The DRAGEN Germline Small Variant Caller takes mapped and aligned DNA reads as input and calls single nucleotide polymorphisms (SNPs) and insertion or deletion (indels) through a combination of column-wise detection and local *de novo* assembly of haplotypes. To enable DRAGEN Germline Small Variant Caller, select the germline variant workflow.

Germline variant calling is typically used for germline samples where the ploidy is known to be two. Callable reference regions are first identified with sufficient alignment coverage. Within these reference regions, a fast scan of the sorted reads identifies active regions, which are centered on pileup columns with evidence of a variant. The active regions are padded with enough context to cover significant, nonreference content nearby. If there is evidence of indels, the active regions receive additional padding.

Aligned reads are clipped within each active region and assembled into a De Bruijn graph. The edges of the clipped reads are weighted by observation counts, with the reference sequence as a backbone. After some graph cleanup and simplification, all source-to-sink paths are extracted as candidate haplotypes. Each haplotype is Smith-Waterman aligned to the reference genome to identify the variants it represents. This set of events may be augmented by a position-based detection. For each read-haplotype pair, the probability $P(r|H)$ of observing the read, assuming the haplotype is the true starting sample, is estimated using a pair hidden Markov model (HMM).

Scanning by reference position over the active region, candidate genotypes are formed from diploid combinations of variant events (SNPs or indels). For each event (including reference), the conditional probability $P(r|e)$ of observing each overlapping read is estimated as the maximum $P(r|H)$ for haplotypes supporting the event. These are combined into the conditional probability $P(r|e1e2)$ for a

genotype (event pair) and multiplied to yield the conditional probability P(R|e1e2) of observing the whole read pileup. Using Bayes' Formula, the posterior probability P(e1e2|R) of each diploid genotype is calculated, and the winner is called.

DRAGEN for IDPE Dx applies automatic filtering. Refer to *Germline Workflow VCF File Annotations* on page 12 for more information.

DRAGEN Somatic Variant Calling

The DRAGEN Somatic Small Variant Caller takes mapped and aligned DNA reads as input and calls SNVs and indels through local *de novo* assembly of haplotypes in an active region. To enable DRAGEN Somatic Small Variant Caller, select a somatic variant application.

Somatic variant calling is typically used for tumor samples. With this workflow, DRAGEN does not make any ploidy assumptions, which enables detection of low-frequency alleles. For loci with coverage up to 100x in the tumor sample, DRAGEN has a detection threshold at variant allele frequencies of 5%. The limit scales with increasing depth on a per-locus basis and halves every time the coverage doubles beyond 100x. Callable reference regions are first identified with sufficient alignment coverage. Within these reference regions, a scan of the sorted reads identifies active regions, which are centered on pileup columns with evidence of a variant in the tumor reads. The active regions are padded with enough context to cover significant, nonreference content nearby. If there is evidence of indels, the active regions receive additional padding.

Aligned reads are clipped within each active region and assembled into a De Bruijn graph. The edges of the clipped reads are weighted by observation counts, with the reference sequence as a backbone. After some graph cleanup and simplification, all source-to-sink paths are extracted as candidate haplotypes. Each haplotype is Smith-Waterman aligned to the reference genome to identify the variants it represents. For each read-haplotype pair, the probability P(r|H) of observing the read is estimated using a pair hidden Markov model (HMM), assuming the haplotype is the true starting sample.

To determine the tumor limit of detection (TLOD) score, DRAGEN Somatic Small Variant Caller first scans by reference position for each candidate somatic event as well as the reference event over the active region. The conditional probability P(r|e) of observing each overlapping read is estimated as the maximum P(r|H) for haplotypes supporting the event. These are combined into the conditional probability P(r|E) for an event hypothesis, E, involving a mixture of the reference and candidate somatic allele over a range of possible allele frequencies and multiplied to yield the conditional probability P(R|E) of observing the whole read pileup. From there, a TLOD score is calculated as the evidence that an ALT allele is present in the tumor sample at a given locus.

DRAGEN for IDPE Dx applies automatic filtering. Refer to *Somatic Workflow VCF File Annotations* on page 14 for more information.

Create a Planned Run

Use the following steps to set up a run in Illumina Run Manager either on the NextSeq 550Dx or using a browser on a networked computer. Use a browser on a networked computer if importing sample data is desired. Refer to the Illumina Run Manager for NextSeq 550Dx Software Guide (document # 200025239) for instruction on accessing Illumina Run Manager from a networked computer.

There are two different ways to create a new planned run:

- **Import Run**—Use a Sample Sheet from an existing run as a template for a new run. Refer to the Illumina Run Manager for NextSeq 550Dx Software Guide (document # 200025239) for information on how to import a run.
- **Create Run**—Manually enter run parameters. The following instructions describe how to create a run.

NOTE The required input fields in the user interface are marked with an asterisk symbol (*).

Application

1. From the Runs screen Planned tab, select **Create Run**.
2. Select the DRAGEN for Illumina DNA Prep with Enrichment Dx application, and then select **Next**.

Run Settings

1. On the Run Settings screen, enter a unique run name. The run name identifies the run from sequencing through analysis.
2. **[Optional]** Enter a run description to further identify the run.
3. Select index adapter kit(s) used during library preparation.
4. Review the Read Length and modify if necessary. Read 1 and Read 2 have a default value of 151 cycles. Index 1 and Index 2 have a fixed value of 10 cycles and cannot be modified.
5. **[Optional]** Enter a library tube ID.
6. Select **Next**.

Sample Data

Sample data includes the Sample ID, Well Position (index plate well position), and Library Name. When using Index A&B, Well Position also includes Plate identifier.

There are two ways to enter sample data:

- **Import Samples**—Use a template file available for download on the Sample Data screen.
- **Manually**—Enter the sample data directly into the table on the Sample Data screen.

Import Samples

When planning a sequencing run using a browser on a networked computer, a template file (`*.csv`) is available for download on the Sample Data screen. The template file is not available for download when accessing Illumina Run Manager through the NextSeq 550Dx operating system software. To enter sample data using the Import Samples feature, do the following steps.

NOTE Complete Run Settings steps before proceeding.

1. Select **Download Template** to download a blank CSV file.
2. From the template file, enter sample data, and then save the file. Library Name is optional.

NOTE When using Index A&B, the data for column B must include both plate and well position (index plate well position). Example: A-A01, A-A02, A-A03.

3. Select **Import Samples** and browse to the template file containing the sample data information from the previous step.
4. Select **Open**, **Proceed**, and then **Next**.

NOTE Changing Sample ID before selecting Next may result in an error. Finish setting up the run before making changes to avoid errors.

Enter Samples Manually

Use the table on the Sample Data screen to enter sample data manually.

1. Enter a unique sample ID in the Sample ID field.
2. Use **Well Position** (Index A or Index B) or **Plate - Well Position** (Index A&B) to select the associated index for the samples.
   The i7 Index, Index 1, i5 Index, and Index 2 fields populate automatically.
3. **[Optional]** Enter a library name.
4. Add rows and repeat steps 1–3 as needed until all samples have been added to the table. You can add multiple rows at one time by first entering the number of rows to add, and then selecting the + icon. You can also remove rows by selecting the box next to the row number, and then clicking the trash icon.
5. Select **Next**.

Analysis Settings

1. Select the desired analysis workflow:
   • FASTQ generation

- FASTQ and VCF generation for a germline workflow (Manifest File required)
- FASTQ and VCF generation for a somatic workflow (Manifest File required)

2. **[Optional] Generate ORA compressed FASTQs** is enabled by default. FASTQ ORA compression losslessly compresses FASTQ files up to 5x compared with fastq.gz. Uncheck **Generate ORA compressed FASTQs** if uncompressed data (fastq.gz) is preferred.
3. For germline and somatic workflows, a manifest file is required. Use the **Manifest File Selection** dropdown menu to select a manifest file. The manifest is a tab-delimited BED(`*.bed`) file that specifies the names and locations of targeted reference regions. For more information, refer to *Manifest File* on page 7.
4. **[Optional]**For Somatic workflows, use the **Noise File Selection** dropdown menu to select a systematic noise file.
   A BED(`*.bed.gz`) file with site-specific noise level can be specified for filtering out systematic noise. For more information, refer to *Noise Filtering (Optional)* on page 7.
5. Select **Next**.

Run Review

1. On the Review screen, review the information for Run Settings, Sample Data, and Analysis Settings.
2. Select **Save**.
   The run is saved on the Planned tab on the Runs screen.

Settings

To view or change DRAGEN for IDPE Dx application settings, first select the Applications icon from the main screen. Then select the application you want to view or change. An Administrator account is required to change settings.

Configuration

The configuration screen displays the following application settings:

- **Library Prep Kits**— Displays the default library prep kit for the app. This setting cannot be changed.
- **Index Adapter Kits**— Displays the default index adapter kit for the app. This setting cannot be changed.
- **Read lengths**— Read lengths are set to 151 for the app by default, but can be changed during run creation.
- **Manifest and Noise Files**— Upload and change settings for manifest and noise files.
  - Select **Upload File** to upload files for use in analysis.
  - Select the **Default** radio button to set the file as the default manifest or noise file selected during run creation when the application is selected.

– Select the **Enabled** checkbox to set the file to display in the dropdown menu during run creation.

Permissions

Use the checkboxes on the Permissions screen to manage user access for the app.

Manifest File

When using DRAGEN for IDPE Dx, a manifest file is required input for the following workflows :

- FASTQ and VCF generation for a germline workflow
- FASTQ and VCF generation for a somatic workflow

The manifest file is a tab-delimited text file using the the BED format (*.bed) that specifies the names and locations of targeted reference regions. The main section of the manifest file is the Regions section and should contain the following data columns:

Column	Description
Name	Unique user-specified name for the target
Chromosome	Chromosome location (eg, chr10, chr5, etc.)
Start	1-based index for the start position of the target
Stop	1-based index for the stop position of the target
Upstream Probe Length	The length of the upstream probe. For the DRAGEN for IDPE Dx app, this should be set to 0.
Downstream Probe Length	The length of the downstream probe. For the DRAGEN for IDPE Dx app, this should be set to 0.

NOTE A valid manifest file format is required for analysis. DRAGEN will stop analysis if the manifest file is invalid.

Noise Filtering (Optional)

The systematic noise filter is available for somatic variant calling, and can be used to reduce false positive calls by accounting for site-specific noise. The systematic noise file is generated by first collecting approximately 50 normal samples (preferably specific to the panel, library prep, and sequencer) and then the sum of allele frequencies under 30% at each site with sufficient coverage is divided by the total number of samples (allele frequencies over 30% are assumed to be germline variants and not noise). Once the noise values are generated, somatic variants detected at that site will be filtered.

The filter can be used in Tumor-Normal mode, but is especially useful for Tumor-Only runs where a matched normal is not available. The systematic noise file must use a BED file that has a (`*.bed.gz`) file extension and must include four columns: Chromosome, Start, End and site-specific noise levels for each row. Systematic noise filtering is optional.

Analysis Outputs

Runs currently in progress are displayed in the Active tab. Completed runs are displayed in the Completed tab. DRAGEN for IDPE Dx creates a uniquely named analysis folder for each analysis, which is separate from the folder containing sequencing data. The analysis folder includes the following information:

- Manifest file used
- Software version
- Sample IDs
- Total aligned reads
- Percent of aligned reads per sample
- Number of SNVs called per sample
- Number of indels called per sample
- Coverage statistics

Analysis Output Files

The analysis folder location is specified by the External Storage for Analysis Results setting. Refer to the Illumina Run Manager for NextSeq 550Dx Software Guide (document # 200025239) for more information about the External Storage for Analysis Results setting.

On the Run Details screen, the External Location field provides the path for sequencing data. The unique analysis folder name is provided in the Analysis Output Folder field on the Run Details screen. The exact files generated depend on which analysis workflow is used. The following analysis output files are generated by the application.

NOTE If a maximum file path length limitation error occurs when accessing analysis output files, try moving the file to a shorter path location or use a different method to open the file.

Output File	Description
Variant summary report (`*.pdf`)	Contains a summary of the file information, software versions, sample information, read level statistics, and SNV, insertions, deletions, and coverage summaries. Only the germline and somatic workflows produce a variant report.

Output File	Description
FASTQ (`*.fastq.gz` or `*.fastq.ora`)	Intermediate files containing quality scored base calls. FASTQ files are the primary input for the alignment step. When ORA compression is selected, the `*.fastq.ora` file extension is used.
Alignment BAM files (`*.bam`)	Contains aligned reads for a given sample.
Genome VCF files (`*.gvcf.gz`)	Contains the genotype for each position, whether called as a variant or called as a reference.
VCF files (`*.vcf.gz`)	Contains variants called at each position.
Run metrics report (`*.csv`)	Contains quality metrics about the run, including non-indexed total yield and Q30 score.

FASTQ Files

FASTQ (*.fastq.gz, *.fastq.ora) is a text-based file format containing base calls and quality values per read. Each file contains the following information:

- The sample identifier
- The sequence
- The Phred quality scores in an ASCII + 33 encoded format

The sample identifier is formatted as follows:

```
@Instrument:RunID:FlowCellID:Lane:Tile:X:Y
ReadNum:FilterFlag:0:SampleNumber
Example:
@SIM:1:FCX:1:15:6329:1045 1:N:0:2
TCGCACTCAACGCCCTGCATATGACAAGACAGAATC
+
<>;##=><9=AAAAAAAAA9#:<#<;<<<????#=
```

BAM Files

A BAM file (*.bam) is the compressed binary version of a SAM (sequence alignment map) file that is used to represent aligned sequences up to 128 Mb. BAM files use the file naming format of `SampleName_S#.bam`. # is the sample number determined by the order that samples are listed for the run. In multinode mode, the S# is set to S1, regardless of the order of the sample.

BAM files contain a header section and an alignment section:

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

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

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

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

VCF Files

Variant call format (*.vcf) files contain information about variants found at specific positions in a reference genome.

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

Table 1 VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For single nucleotide variants (SNVs), this position is the reference base with the variant. For indels, this position is the reference base immediately preceding the variant.
ID	The rs (reference SNP) number for the SNP obtained from `dbSNP.txt`, if applicable. If multiple rs numbers exist at this location, the list is delimited by semicolons. If a dbSNP entry does not exist at this position, a missing value marker ('.') is used.
REF	The reference genotype. For example, a deletion of a single T is represented as reference TT and alternate T. An A to T single nucleotide variant is represented as reference A and alternate T.

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

Table 2 Germline Workflow VCF File Annotations

Heading	Description
FILTER	If all filters pass, PASS is written in the filter column. Possible FILTER entries include: • **DRAGENSnpHardQUAL**—Applied if SNP variant QUAL score does not meet threshold • **DRAGENIndelHardQUAL**—Applied if indel variant QUAL score does not meet threshold • **LowDepth**—Site filtered because depth of coverage does not meet threshold • **LowGQ**—Site filtered because genotype quality does not meet threshold • **PloidyConflict**—Genotype call from variant caller not consistent with chromosome ploidy • **base_quality**—Site filtered because median base quality of alt reads at this locus does not meet threshold • **filtered_reads**—Site filtered because too large a fraction of reads has been filtered out • **fragment_length**—Site filtered because absolute difference between the median fragment length of alt reads and median fragment length of ref reads at this locus exceeds threshold • **low_depth**—Site filtered because the read depth is too low • **low_frac_info_reads**—Site filtered because the fraction of informative reads is below threshold • **low_normal_depth**—Site filtered because the normal sample read depth is too low • **long_indel**—Site filtered because the indel length is too long • **mapping_quality**—Site filtered because median mapping quality of alt reads at this locus does not meet threshold • **multiallelic**—Site filtered because more than two alt alleles pass tumor LOD • **non_homref_normal**—Site filtered because the normal sample genotype is not homozygous reference • **no_reliable_supporting_read**—Site filtered because no reliable supporting somatic read exists • **panel_of_normals**—Seen in at least one sample in the panel of normals vcf • **read_position**—Site filtered because median of distances between start/end of read and this locus is below threshold • **RMxNRepeatRegion**—Site filtered because all or part of the variant allele is a repeat of the reference • **strand_artifact**—Site filtered because of severe strand bias • **str_contraction**—Site filtered due to suspected PCR error where the alt allele is one repeat unit less than the reference • **too_few_supporting_reads**—Site filtered because there are too few supporting reads in the tumor sample • **weak_evidence**—Somatic variant score does not meet threshold

Heading	Description
INFO	Possible INFO entries include: • **AC**—Allele count in genotypes for each ALT allele, in the same order as listed. • **AF**—Allele Frequency for each ALT allele, in the same order as listed. • **AN**—The total number of alleles in called genotypes. • **DB**—dbSNP Membership. • **FS**—Phred-scaled p-value using Fisher's exact test to detect strand bias. • **QD**—Variant Confidence/Quality by Depth. • **R2_5P_bias**—Score based on mate bias and distance from 5 prime end. • **SOR**—Symmetric Odds Ratio of 2x2 contingency table to detect strand bias. • **DP**—Approximate read depth (informative and non-informative); some reads may have been filtered based on mapq etc. • **END**—Stop position of the interval. • **FractionInformativeReads**—The fraction of informative reads out of the total reads. • **MQ**—RMS Mapping Quality. • **MQRankSum**—Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities. • **ReadPosRankSum**—Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias. • **SOMATIC**—At least one variant at this position is somatic.
FORMAT	The format column lists fields separated by colons. For example, GT:GQ. Available fields include: • **AD**—Allelic depths (counting only informative reads out of the total reads) for the ref and alt alleles in the order listed. • **AF**—Allele fractions for alt alleles in the order listed. • **DP**—Approximate read depth (reads with MQ=255 or with bad mates are filtered). • **F1R2**—Count of reads in F1R2 pair orientation supporting each allele. • **F2R1**—Count of reads in F2R1 pair orientation supporting each allele. • **GT**—Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available. • **MB**—Per-sample component statistics to detect mate bias. • **PS**—Physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group. • **SB**—Per-sample component statistics, which comprise the Fisher's Exact Test to detect strand bias. • **SQ**—Somatic quality.
SAMPLE	The sample column gives the values specified in the FORMAT column.

Table 3 Somatic Workflow VCF File Annotations

Heading	Description
FILTER	If all filters pass, PASS is written in the filter column. Possible FILTER entries include: • **base_quality**—Site filtered because median base quality of alt reads at this locus does not meet threshold • **filtered_reads**—Site filtered because too large a fraction of reads have been filtered out • **fragment_length**—Site filtered because absolute difference between the median fragment length of alt reads and median fragment length of ref reads at this locus exceeds threshold • **low_depth**—Site filtered because the read depth is too low • **low_frac_info_reads**—Site filtered because the fraction of informative reads is below threshold • **low_normal_depth**—Site filtered because the normal sample read depth is too low • **long_indel**—Site filtered because the indel length is too long • **mapping_quality**—Site filtered because median mapping quality of alt reads at this locus does not meet threshold • **multiallelic**—Site filtered because more than two alt alleles pass tumor LOD • **non_homref_normal**—Site filtered because the normal sample genotype is not homozygous reference • **no_reliable_supporting_read**—Site filtered because no reliable supporting somatic read exists • **panel_of_normals**—Seen in at least one sample in the panel of normals vcf • **read_position**—Site filtered because median of distances between start/end of read and this locus is below threshold • **RMxNRepeatRegion**—Site filtered because all or part of the variant allele is a repeat of the reference • **strand_artifact**—Site filtered because of severe strand bias • **str_contraction**—Site filtered due to suspected PCR error where the alt allele is one repeat unit less than the reference • **too_few_supporting_reads**—Site filtered because there are too few supporting reads in the tumor sample • **weak_evidence**—Somatic variant score does not meet threshold • **systematic_noise**—Site filtered based on evidence of systematic noise in normals

Heading	Description
INFO	Possible INFO entries include: • **DP**—Approximate read depth (informative and non-informative); some reads may have been filtered based on mapq etc. • **END**—Stop position of the interval. • **FractionInformativeReads**—The fraction of informative reads out of the total reads. • **MQ**—RMS Mapping Quality. • **MQRankSum**—Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities. • **ReadPosRankSum**—Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias. • **AQ**—Systematic noise score. • **hotspot**—Known somatic site, used to increase confidence in call. • **SOMATIC**—At least one variant at this position is somatic.
FORMAT	The format column lists fields separated by colons. For example, GT:GQ. Available fields include: • **AD**—Allelic depths (counting only informative reads out of the total reads) for the ref and alt alleles in the order listed. • **AF**—Allele fractions for alt alleles in the order listed. • **DP**—Approximate read depth (reads with MQ=255 or with bad mates are filtered). • **F1R2**—Count of reads in F1R2 pair orientation supporting each allele. • **F2R1**—Count of reads in F2R1 pair orientation supporting each allele. • **GP**—Phred-scaled posterior probabilities for genotypes as defined in the VCF specification. • **GQ** —Genotype quality. • **GT**—Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available. • **MB**—Per-sample component statistics to detect mate bias. • **PL**—Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification. • **PRI**—Phred-scaled prior probabilities for genotypes. • **PS**—Physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group. • **SB**—Per-sample component statistics, which comprise the Fisher's Exact Test to detect strand bias. • **SQ**—Somatic quality.
SAMPLE	The sample column gives the values specified in the FORMAT column.

Genome VCF Files

Genome VCF (*.gvcf.gz) files follow a set of conventions for representing all sites within the genome in a reasonably compact format. The gVCF files include all sites within the region of interest in a single file for each sample. The gVCF file shows no-calls at positions that do not pass all filters. A genotype (GT) tag of ./. indicates a no-call.

Requeue Analysis

You might requeue analysis if analysis was stopped, if analysis was unsuccessful, or if you want to reanalyze a run with different settings. To requeue analysis, do the following steps:

1. From the Run screen, select the Completed tab, and then select the run name to reanalyze. If Requeue Analysis was previously performed, select the run name of the Parent Run.
2. From Run Details screen, after Sequencing Information, select **Requeue Analysis**.
3. Select an option:
    - Requeue analysis with no changes
    - Edit run settings and requeue analysis
    - Requeue analysis with a different application
4. Confirm that the location where the sequencing data currently resides is provided in the **Sequencing data file path** field.

NOTE The path to the sequencing data should match the path in the External Storage for Analysis Results setting. Refer to the Illumina Run Manager for NextSeq 550Dx Software Guide (document # 200025239) for information about changing the external storage path.

5. Enter a Reanalysis Reason.
6. Select **Requeue Analysis**.
7. Edit the desired changes to the Run Settings, Sample Data, and Analysis Settings.
8. Select **Save**. Analysis begins using current analysis parameters.

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Website:	www.illumina.com
Email:	techsupport@illumina.com

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

Product documentation—Available for download from support.illumina.com.

Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com

IVD

EC REP

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

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

FOR IN VITRO DIAGNOSTIC USE.

illumina®