illumina®

Local Run Manager
TruSight Oncology 500 Analysis Module

Workflow Guide

For Research Use Only. Not for use in diagnostic procedures.

For Research Use Only. Not for use in diagnostic procedures.

For Research Use Only. Not for use in diagnostic procedures.

Document # 10000000099600 v00

3

For Research Use Only. Not for use in diagnostic procedures.

Overview

The Illumina® TruSight™ Oncology 500 analysis module generates sequencing libraries for DNA and RNA from formalin-fixed, paraffin-embedded (FFPE) tissue samples. The module supports run setup, sequencing, and analysis for the prepared DNA and RNA libraries. DNA library analysis outputs include tumor mutational burden, variant call files for small and complex variants, microsatellite instability, and gene amplifications. RNA library analysis outputs include fusions and splice variant call files. The TruSight Oncology 500 Analysis Module module supports 3–8 DNA libraries, 3–16 RNA libraries, and select combinations of DNA and RNA libraries per run.

About This Guide

This guide provides instructions for setting up run parameters for sequencing and analysis for the TruSight Oncology 500 Analysis Module. Use of the software requires basic knowledge of the current Windows operating system and web browser-based user interface.

For more information about Local Run Manager, see the *Local Run Manager Software Guide (document # 1000000002702)*. This guide includes instructions on the following:

▶ Initiating sequencing

▶ Requeuing an analysis

▶ Editing a run

▶ Importing a run

Additional Resources

The Local Analysis Software support pages on the Illumina support website provide additional system resources. These resources include software, training, compatible products, documentation, and product files, including a sample sheet template. Always check support pages for the latest versions.

Resource	Description
Local Analysis Software SampleSheet.csv (document # 1000000105888)	Setup template for running the analysis.
Denature and Dilute Guide	Guidelines on the number of libraries and possible DNA/RNA combinations per sequencing run.

Local Run Manager Settings

Before initiating a TruSight Oncology 500 Analysis Module run, ensure that the output directory path you have configured for your instrument does not exceed 40 characters. Exceeding 40 characters can cause failure during results copy-out.

For further detail on configuring the output directory, see *Local Run Manager Software Guide (document # 1000000002702)*.

Document # 1000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

4

Run Setup and Viewing Results

Local Run Manager is the software used to set up an TruSight Oncology 500 run.

Enter run and sample setup information directly into the Local Run Manager TruSight Oncology 500 analysis module. View analysis results and additional module information from the Local Run Manager dashboard.

Set Run Parameters

1   Log in to Local Run Manager on the instrument or from a networked computer.

2   Select **Create Run**, and then select **TSO 500**.

3   Enter a run name that identifies the run from sequencing through analysis with the following criteria:
    ▶   1-40 characters.
    ▶   Only alphanumeric characters, underscores, or dashes.
    ▶   Underscores and dashes must be preceded and followed by an alphanumeric character.
    ▶   Unique across all runs on the instrument.

4   **[Optional]** Enter a run description to help identify the run with the following criteria:
    ▶   1-150 characters.
    ▶   Only alphanumeric characters or spaces.
    ▶   Spaces must be preceded and followed by an alphanumeric character.

Specify Samples for the Run

Specify samples for the run using one of the following options:

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

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

After you populate the samples table, you can export the sample information to an external file. This file can serve as a reference when preparing libraries or importing the file for another run.

> CAUTION
>
> Mismatches between the samples and index primers cause incorrect result reporting due to loss of positive sample identification. Enter sample IDs and assign indexes in Local Run Manager before beginning library preparation. Record sample IDs, indexes, and plate well orientation for reference during library preparation.

Enter Samples Manually

1   Enter a unique sample ID in the Sample ID field with the following criteria:
    ▶   Only alphanumeric characters, underscores, or dashes.
    ▶   Unique within the run setup.
    ▶   ≤ 25 characters.

2   **[Optional]** Enter a sample description in the Sample Description field.
    ▶   Use alphanumeric characters, dashes, underscores, or spaces.

3   Select an index for the sample.

i7 and i5 fields auto-populate after selecting an Index ID.
- ▶ If the sample is DNA, select a unique index ID from the DNA index ID drop-down list.
- ▶ If the sample is RNA, select a unique index ID from the RNA index ID drop-down list.

4 **[Optional]** Select **Export to Template** to export sample information to an external file.

5 Review the information on the Create Run Screen. Incorrect information can impact results.

6 Select **Save Run**.

Import Samples

1 Select **Import Template** and browse to the location of the sample information file. There are two types of files you can import.
- ▶ Select **Download Template** on the Create Run screen to download a new template file. The template file contains the required column headings and format for import. Enter sample information in each column for the samples in the run. Delete example information in unused cells, and then save the file.
- ▶ Use a file of sample information that was exported from the TruSight Oncology 500 Analysis Module analysis module using the Export to Template feature.

2 On the Create Run screen, review the imported information. Incorrect information can impact results.

3 **[Optional]** Select **Export to Template** to export sample information to an external file.

4 Select **Save Run**.

View Analysis Results

1 From the Local Run Manager dashboard, select the run name.

2 From the Run Overview tab, review the sequencing run metrics.

3 To change the analysis data file location for future requeues of the selected run, select **Edit**, and edit the output run folder file path.
The file path leading up to the output run folder name is editable. The output run folder name cannot be changed.
The file path for the output run folder must be ≤40 characters.

4 **[Optional]** Select **Copy to Clipboard** for access to the output run folder.

5 Select the Sequencing Information tab to review run parameters and consumables information.

6 Select the Samples & Results tab to view the directory where Analysis output files can be located.

7 **[Optional]** Select **Copy to Clipboard** to copy the Analysis folder file path.

For more information on the Run Overview and Sequencing Information tabs, and how to requeue analysis, see *Local Run Manager Software Guide (document # 1000000002702)*.

Modules and Manifests

To view more information about the Local Run Manager Module, select **Modules and Manifests** from the Tools drop-down menu. The module settings menu is only accessible to admin users or users with admin-specified access permission.

The Module settings screen displays the following information:

- ▶ Module name
- ▶ Module version
- ▶ RTA version used for primary analysis
- ▶ Last date modified on
- ▶ Regulatory label
- ▶ Sequencing run settings

Analysis Methods

The Local Analysis Software workflow performs the following analysis steps, and then writes analysis output files to the folder specified.

- ▶ FASTQ Generation
- ▶ DNA Analysis Methods
  - ▶ DNA Alignment and Realignment
  - ▶ Read Collapsing
  - ▶ Indel Realignment and Read Stitching
  - ▶ Small Variant Calling
  - ▶ Small Variant Filtering
  - ▶ Copy Number Variant Calling
  - ▶ Phased Variant Calling
  - ▶ Variant Merging
  - ▶ Annotation
  - ▶ Tumor Mutational Burden
  - ▶ Microsatellite Instability Status
  - ▶ Contamination Detection
- ▶ RNA Analysis Methods
  - ▶ Downsampling
  - ▶ Read Trimming
  - ▶ Alignment
  - ▶ Duplicate Marking
  - ▶ Fusion Calling
  - ▶ RNA Fusion Filtering
  - ▶ Splice Variant Calling
  - ▶ Annotation
  - ▶ Fusion Merging
- ▶ Quality Control
  - ▶ Run QC
  - ▶ DNA Sample QC
  - ▶ RNA Sample QC

FASTQ Generation

BCL files are demultiplexed and the software generates intermediate analysis files in the FASTQ format. FASTQ files use a text format to represent sequences. They contain reads for each sample and the

associated quality scores. Any controls used for the run and clusters that did not pass the filter are excluded. Each FASTQ file contains reads for only one sample, and the FASTQ file name includes the name of that sample.

The bcl-convert software uses the SampleSheet_dna.csv and SampleSheet_rna.csv generated by the analysis module for demultiplexing. The sample sheets can be found in the FastqGeneration step folder.

DNA Analysis Methods

Figure 1 TruSight Oncology 500 Analysis Module DNA Workflow

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

8
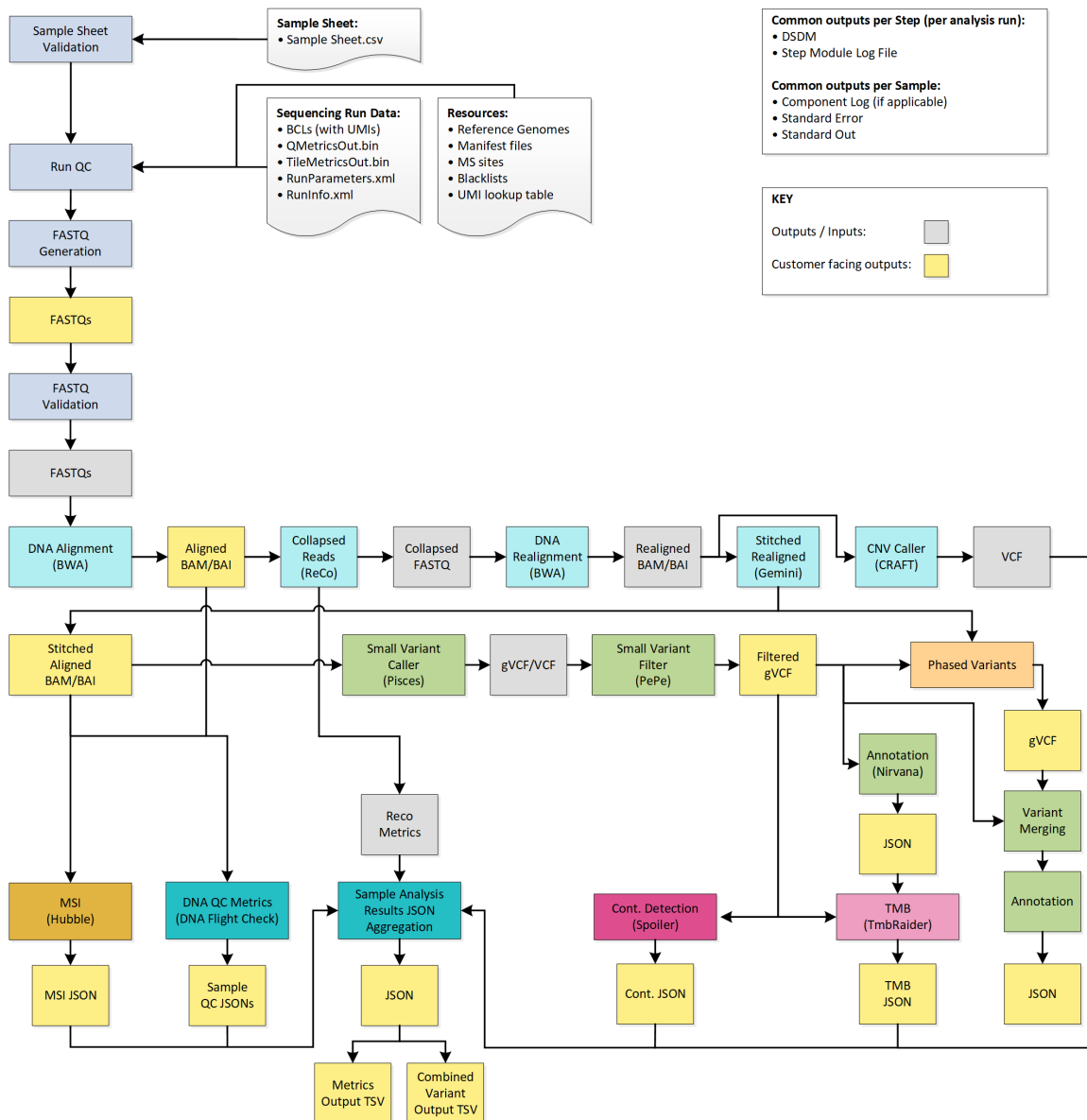
DNA Alignment and Realignment

The alignment step uses the BWA-MEM alignment with the SAM Tools utility to align DNA sequences in FASTQ files to the hg19 genome. This alignment step generates BAM files (*.bam) and BAM index files (*.bam.bai), which are saved to the DnaAlignment folder. A BAM file is the compressed binary version of a SAM file that is used to represent aligned sequences.

The software performs a second alignment on FASTQ files after the Read Collapsing step completes. The second alignment enables the realignment of sample reads using only unique molecular identifier (UMI) collapsed reads.

For more information on BWA-MEM, see the Burrows-Wheeler Aligner website. For more information on SAM and BAM files, see the Sequence Alignment/Map specification page on GitHub.

Read Collapsing

The read collapsing analysis step executes an algorithm that collapses sets of reads (known as families) with very similar genomic locations into representative sequences using UMI tags. This process allows for the accurate removal of duplicate reads without losing the signal of very low frequency sequence variations. Additionally, UMI collapsing further reduces FFPE deamination artifacts by utilizing duplex collapsing where information from complimentary strands are combined. The read collapsing step produces FASTQ files and associated metrics files in the CollapsedReads output folder. Read collapsing adds the following BAM tags:

▶ **RX/XU**—UMI.

▶ **XV**—Number of reads in the family.

▶ **XW**—Number of reads in the duplex-family, or 0 if not a duplex family.

Indel Realignment and Read Stitching

The Gemini software performs local indel realignment, paired-read stitching, and read filtering to improve small variant calling results. A stitched read is a single read that has been combined from a pair of reads. Reads near detected indels are realigned to remove alignment artifacts. It takes in a single BAM file and the genome FASTA used to align it and outputs a corresponding single BAM file with stitched, pair-realigned reads. Read pairs with poor map quality or supplementary and secondary alignments from the input BAM are ignored.

Gemini adds the following BAM tags:

▶ **XD**—For successfully stitched reads only. Directional support string indicating forward, reverse, and stitched positions.

▶ **XR**—For successfully stitched reads only. Pair orientation (FR or RF).

Small Variant Calling

Pisces software performs somatic variant calling to identify variants at low frequency in DNA samples. Pisces calls small variants in the BAM files that are generated from the StitchedRealigned analysis step.

For each variant candidate, Pisces adds an US field under the Format column in the genome.vcf for the mutant support of the following read type counts: duplex-stitched, duplex-nonstitched, simplex-forward-stitched, simplex-forward-nonstitched, simplex-reverse-stitched, simplex-reverse-nonstitched, followed

by total support of the same read type counts. The small variant calling genome.vcf at this step is only responsible to collect candidate and output corresponding read support information, and the final variant call is determined in following post-processing step.

The software component Psara is used to trim the gVCF based on the panel manifest. Variants are considered for inclusion if they overlap with the manifest or are contained within an overlapping indel. Small Variant Filtering determines the final variant call.

Small Variant Filtering

The software component, Pepe, performs post-processing on the small variant calling VCFs with the intended result of background polishing and quality score adjustment. The software filters out variants when their error rates do not meet quality thresholds. This analysis step produces gVCF files and associated error rate files. The minimum read depth for reference calls is 100 and the limit of detection for VAF is 0.05 at that depth.

Pepe computes two quality scores for each variant candidate that dynamically adjusts for the following conditions:

▶ Background noise

▶ Nucleotide change

▶ Read support type

For each variant candidate, background noise at the same site is estimated from normal baseline samples of varying qualities. A p-value is calculated using the observed mutant depth, total depth, and background noise using binomial distribution and then converted to a variant quality score (AQ). The sample-specific error rate of each nucleotide change is estimated from different support categories in each sample by using all the positions with an allele frequency less than 5%. For each variant candidate, a likelihood ratio (LQ) score is computed by the corresponding error rate of the observed total and mutant read.

For variants with a Catalogue of Somatic Mutations in Cancer (COSMIC) count > 50, the LQ and AQ thresholds are 20 and the remaining sites are 60. For indel, at least one stitched mutant support is required. The net effect of the read collapsing and variant filtering reduce false positives in a typical FFPE DNA sample from ˜1500 per Mb to < 5 per Mb.

Copy Number Variant Calling

The CRAFT copy number variant caller performs amplification, reference, and deletion calling for target CNV genes within the assay. The CRAFT software component counts coverage of each target interval on the panel, performs normalization, calculates fold change values for each gene, and determines the CNV status for each CNV target gene. During normalization steps, coverage biases are corrected using potential variables such as sequencing depth, target size, PCR duplicates, probe efficiency, GC bias, and DNA type. A collection of normal FFPE and genomic DNA samples is used to correct some of these variables. For each target CNV gene, *in silico* data is trained to determine a gene specific threshold for amplification and deletion. The inputs are collapsed read in BAM format and the outputs are VCF files. Amplification are annotated as DUP in the VCF file. Deletions status (DEL) are provided for information only and always be marked as LowValidation in the VCF file.

Phased Variant Calling

Scylla rapidly detects multiple nucleotide variants (MNVs) in a given sample. The workflow uses Scylla to detect specific, clinically relevant mutations in EGFR exon 19 that would otherwise be out of scope for the variant caller. Psara filters the small variant gVCF to a small region in exon 19 of EGFR.

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

10

Candidate SNPs, MNVs, and indels from this subset of the gVCF are given to Scylla along with the BAM output from Gemini. Scylla uses the original BAM to determine which of these small variants should be phased together into longer MNVs.

At a high level, Scylla identifies variants that are candidates for phasing in the input gVCF and arranges the variants into local neighborhoods. Scylla then mines the sample BAM file for any evidence that these small variants occur in the same clonal sub-populations with each other. This is done by clustering overlapping reads in the neighborhood into a minimal set of clusters, which contain the same variants.

Unlike Pisces, Scylla does not require that variants be on the same read to be phased. Once the phasing is complete, a new gVCF is generated.

Variant Merging

The software merges the phased variants with the other small variants generated from small variant filtering step and produces a gVCF file. In this process, exact duplicates that match chromosome, position, reference allele, and alternative allele are removed. Only the following Epidermal Growth Favor Receptor (EGFR) variants are added if found from Phased Variant Calling. All other EGFR variants are filtered out in variant merging.

Table 1   EGFR Variants

Chromosome	Position	Reference Allele	Alternative Allele
chr7	55242463	AAGGAATTAAGAGAAG	A
chr7	55242464	AGGAATTAAGAGA	A
chr7	55242464	AGGAATTAAGAGAAGC	A
chr7	55242465	GGAATTAAGAGAAGCA	G
chr7	55242467	AATTAAGAGAAGCAAC	A
chr7	55242469	TTAAGAGAAGCAACATCTC	T
chr7	55242468	ATTAAGAGAAGCAACATCT	A
chr7	55242466	GAATTAAGAGAAGCAA	G
chr7	55242465	GGAATTAAGA	G
chr7	55242469	TTAAGAGAAGCAA	T
chr7	55242462	CAAGGAATTAAGAGAA	C
chr7	55242466	GAATTAAGAGAAGCAACAT	G
chr7	55242482	CATCTCCGAAAGCCAACAAGGAAAT	C
chr7	55242465	GGAATTAAGAGAAGCAACA	G
chr7	55242467	AATTAAGAGAAGCAACATC	A
chr7	55242469	TTAAGAGAAG	C
chr7	55242467	AATTAAGAGAAGCAACATC	T
chr7	55242469	TTAAGAGAAGCAA	C
chr7	55242467	TTAAGAGAAGCAA	TTGCT
chr7	55242468	ATTAAGAGAAG	GC
chr7	55242469	TTAAGAGAAGCAACATCTCC	CA
chr7	55242465	GGAATTAAGAGAAG	AATTC
chr7	55242465	GGAATTAAGAGAAGCAAC	AAT

chr7	55242467	AATTAAGAGAAGCAAC	T
chr7	55242467	AATTAAGAGAAGCAACATCTC	TCT
chr7	55242469	TTAAGAGAAGCAACATCT	CAA
chr7	55242465	GGAATTAAGAGAAGCAA	AATTC
chr7	55242465	GGAATTAAGAGAAGCAACATC	AAT
chr7	55242468	ATTAAGAGAAGCAAC	GCA

Annotation

The Illumina Annotation Engine Nirvana performs annotation of small variants. The inputs are gVCF files and the outputs are annotated JSON files. Files are first annotated after the variant filter analysis step, and then again after the complex variant calling step.

Each variant entry that is processed by Nirvana will be annotated with available information from databases such as dbSNP, gnomAD genome and exome, 1000 genomes, ClinVar, COSMIC, RefSeq, and Ensembl. Version information and general details can be retrieved from the header, which is the first line of the JSON file. Each annotated variant is included as a nested dictionary structure in separate lines following the header. Version information for each annotation database is shown in the following table.

Database	Version
gnomAD	2.1
COSMIC	v84
ClinVar	2019-02-04
dbSNP	v151
100 Genomes	Phase 3 v5a
RefSeq	VEP build 91
Ensembl	VEP build 91

Tumor Mutational Burden

The tumor mutational burden (TMB) analysis step generates TMB metrics from the annotated small variant JSON file and the gVCF file that is generated from the small variant filtering analysis step. The annotated JSON file is used to retrieve information regarding individual variants, such as allele counts in public databases and resulting consequences at a transcript level. The gVCF is used to evaluate the effective panel size denominator.

To remove germline variants from the TMB calculation, the software uses a combination of public database filtering and post-database filtering strategy that uses allele frequency information and variants in close proximity.

First, the component excludes any variant with an observed allele count ≥10 in any of the GnomAD exome, genome, and 1000 genomes database. To filter germline variants that are not observed in the database, the software identifies variants on the same chromosome with an allele frequency within a certain range. If a given variant is not filtered out based on occurrence in the databases, variants on the same chromosome with similar allele frequencies will be grouped, and if 5 or more similar variants are found to have been filtered, the variant of interest is removed from the TMB Calculation. Additionally, variants with an allele frequency ≥ 90% are removed from the TMB calculation as well. The TMB is calculated as follows:

TMB = Eligible Variants / Effective panel size

Eligible Variants (numerator)	• Variants not removed by the filtering strategy. • Variants in the coding region (RefSeq Cds) • Variant Frequency >= 5% • Coverage >= 50X • SNVs and Indels (MNVs excluded) • Nonsynonymous and synonymous variants • Variants with COSMIC count >= 50 excluded
Effective Panel Size (denominator)	• Total coding region with coverage > 50X • Excluding low confidence regions in which variants are not called

Outputs are captured in a *_TMB_Trace.tsv file that contains information regarding variants used in the TMB calculation and a *.tmb.json file that contains the TMB score calculation and configuration details.

Microsatellite Instability Status

The microsatellite instability status step determines microsatellite instability from the BAM file from the read stitching analysis step and generates an MSI metric file. The software assesses microsatellite sites for evidence of instability, relative to a set of baseline normal samples that are based on information entropy metrics. The percentage of unstable MSI sites to total assessed MSI sites is reported as a sample-level microsatellite score.

Contamination Detection

The contamination analysis step detects contamination by foreign DNA in the VCF files that the small variant filter step generates. The software determines whether a sample has foreign DNA from the combination of contamination p-value (p-score) and contamination scores.

The contamination score is the sum of all the log likelihood scores across all positions. The p-score represents the significance that SNPs are distributed non-uniformly across the chromosomes. This could indicate a highly-rearranged genome and cause false positives for contaminaton.

In contaminated samples, there are SNPs that have variant allele frequency shifts from 0%, 50%, or 100%. The algorithm collects all the positions that overlap with common SNPs with variant allele frequencies of < 25% or > 75%. Then, the algorithm computes the likelihood that the positions are an error or a real mutation using the following qualifications:

▶ Estimates the error rate per sample.

▶ Mutation support.

▶ Total depth of each position selected.

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

13

RNA Analysis Methods

Figure 2   TruSight Oncology 500 Analysis Module RNA Workflow
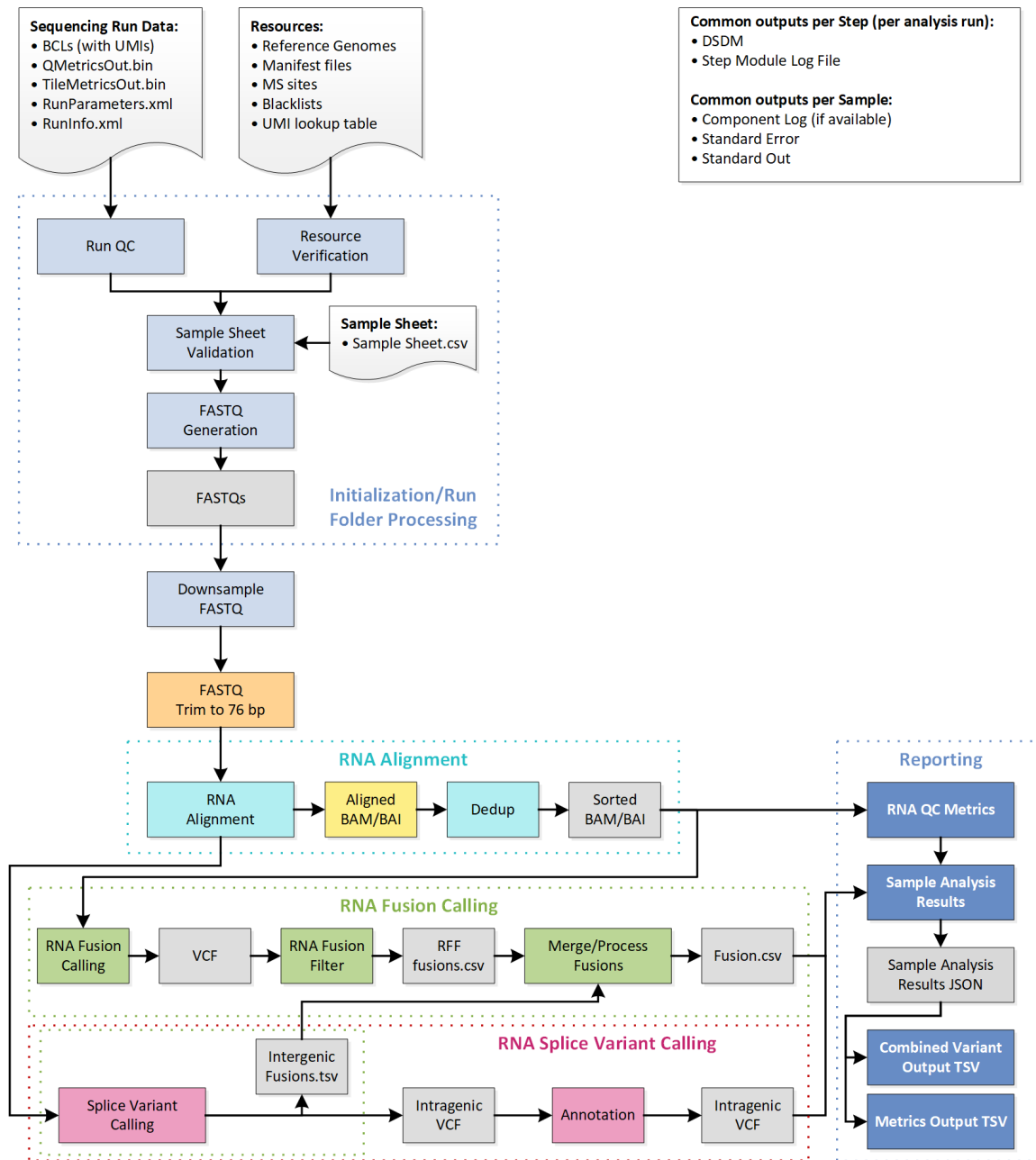
Downsampling

Each sample is downsampled to 30 million RNA reads. This number represents the total number of single reads (ie, R1 + R2, from all lanes). When using the recommended sequencing configurations or plexing, the samples can have fewer reads than the downsampling limit. In these cases, the FASTQ files are left as-is.

Read Trimming

Reads are trimmed to 76 base pairs for further processing. From internal testing, fragment sizes in RNA FFPE samples hover around 100 bp, so the majority of reads at 76 bp are overlapping (see). While STAR alignment does perform stitching to handle overlapping reads, internal testing using simulated data indicates that performance is improved with fewer overlapping reads.
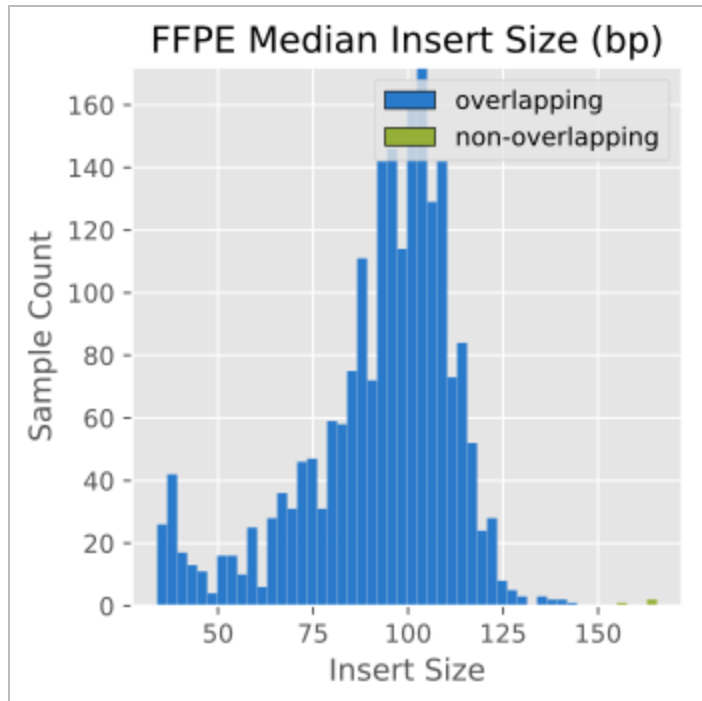
Alignment

The STAR Aligner aligns RNA reads to the human hg19 genome containing unplaced contigs (i.e., chrUn_gl regions) and uses GENCODEv19 to identify splice sites. STAR also includes stitching logic to handle overlapping reads.

Duplicate Marking

RNA does not contain UMIs, so duplicate marking is performed using an internally developed tool based on the Picard duplicate marking algorithm. The start and end coordinates of alignments (adjusted for soft clipping) are used to determine whether fragments are overlapping or not. Fusion and splice variant calling only use deduped fragments to score variants. Only primary alignments are considered, supplementary and secondary alignments are not. The alignment with the highest read score is chosen as the unique fragment.

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

15

Figure 3   Picard Alignment Based Duplicate Marking.
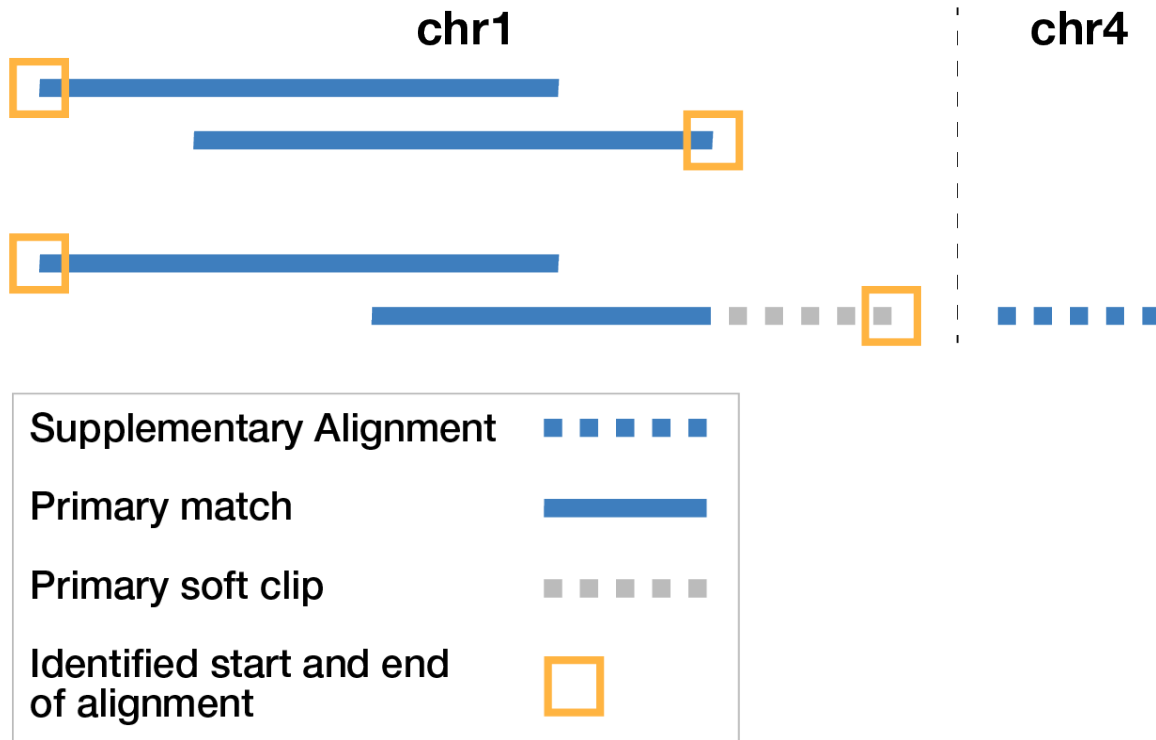
Fusion Calling

Manta is a fusion caller which discovers, assembles, and scores large-scale SVs. Manta only considers candidate fusions with at least 3 unique supporting reads, one of which must be a split read (a single read crossing the fusion breakpoint). The inputs are BAM files and the outputs are VCF files.

RNA Fusion Filtering

The RNAFusionFilter scores fusions and applies filters based on contig alignment to the genome and other features. It also determines which gene is on the 3' end and which gene is on the 5' end based on gene annotations and aligning the contig to the genome. The inputs are BAM files and VCF files and the outputs are .csv format.

Table 2   Scored Features

Score Component	Fusion Feature	Scored Range	Coefficient
Split reads	SplitAlt	0-10	.15
Paired reads	PairedAlt	0-5	.15
Alt allele fraction	(splitAlt + PairAlt) / (SplitRef + PairRef)	0-.01	.1
Fusion contig align length (bp)	ContigAlign1 ContigAlign2	12-76	.4
Breakend homology (bp)	BreakpointHomology	2-20	-.2

Fusion contig align length (bp) (For fusions on different chromosome: dist- DistMax)		100000 - 2000000	.1
Coverage after breakend (bp)	min(CoverageGene1, CoverageGene2)	0-500	.1

Splice Variant Calling

Splice variant calling is performed using internally developed software. The inputs are BAM files and SJ.out.tab files from STAR. Junctions from SJ.out.tab are filtered first using splice annotations from GENCODEv19, and then further filtered using a baseline from a cohort of non-tumor FFPE samples of varying tissue types. Splice junctions appearing on a whitelist are not filtered. The whitelist contains ARv7, MET exon 14 skipping, and EGFRvIII. The outputs are VCF files, which are the final output, and TSV file containing intergenic variants, which are used in fusion merging. Splice variants are scored from 0-10 as shown in the table below.

Table 3   Scored Features in Splice Variant Caller

Score Component	Splice Feature	Scored Range	Coefficient
Split reads	split_unique_reads_alt	0-10	1

Annotation

The Illumina Annotation Engine performs annotation of splice variants. The inputs and outputs are VCF files.

Fusion Merging

Fusions identified through the fusion calling and RNA fusion filtering are combined with the intergenic calls made during splice variant calling. Additionally, each precise fusion event from the RNA fusion filter is recalculated for read count support. The inputs are intergenic splice variant TSV files, fusion filter CSV files, and BAM files. The outputs are CSV files.

Quality Control

The TruSight Oncology 500 Analysis Module includes several quality control analyses.

Run QC

The Run Metrics report provides suggested values to determine if run quality results are within an acceptable range using InterOp files from the sequencing run folder. For Read 1 and Read 2, the report provides the average percentage of bases ≥ Q30, which is a quality score (Q-score) measurement. A Q-score is a prediction of the probability of an incorrect base call.

DNA Sample QC

The inputs for DNA Sample QC are DNA alignment, read collapsed BAM, indel realignment, read stitching BAM, and CRAFT normalized BinCount.tsv files. The metrics and guideline thresholds can be found in the MetricsOutput.tsv file.

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

17

Metric	Description	Recommended Guideline Quality Threshold	Variant Class
CONTAMINATION_ SCORE and CONTAMINATION_ P_VALUE	The contamination score from based on VAF distribution of SNPs. The contamination p-value is used to assess highly rearranged genomes and only needed when contamination score is above USL. A p-score less than 0.05 suggest that the sample has likely large-scale rearrangements that could lead to high contamination scores without actual sample contamination.	Contamination Score ≤ 3106' OR Contamination Score > 3106 and Contamination p-value ≤ 0.049	All
MEDIAN_EXON_ COVERAGE	Median exon fragment coverage across all exon bases.	≥ 150	Small variant Tmb
PCT_EXON_50X	Percent exon bases with 50X fragment coverage.	≥ 90.0	Small variant Tmb
MEDIAN_INSERT_ SIZE	The median fragment length in the sample.	≥ 70	Small variant Tmb
COVERAGE_MAD	Median Absolute Deviation. Represents the median normalized deviation across all regions used for CNV calling.	≤ 0.210	CNV
MEDIAN_BIN_ COUNT_CNV_ TARGET	The median raw bin count per CNV target.	≥ 1.0	CNV
USABLE_MSI_ SITES	The number of MSI sites usable for MSI calling.	≥ 40	MSI

RNA Sample QC

The inputs for RNA Sample QC are RNA alignment. Metrics and guideline thresholds can be found in the MetricsOutput.tsv file.

Metric	Description	Recommend Guideline Quality Threshold	Variant Class
MEDIAN_ CV_ GENE_ 500X	The median CV for all genes with median coverage > 500x. Genes with median coverage > 500x are likely to be highly expressed. Higher CV median > 500x indicates an issue with library preparation (poor sample input and/or probes pull down issue).	≤ 93	Fusion Splice
MEDIAN_ INSERT_ SIZE	The median fragment length in the sample.	≥ 80	Fusion Splice
TOTAL_ ON_ TARGET_ READS	The total number of reads that map to the target regions.	≥ 9000000	Fusion Splice

Analysis Output

To view analysis output, complete the following steps.

1   Navigate to the directory that contains the analysis output folder.

2   Open the folder, and then select the files that you want to view.

Combined Variant Output

The combined variant output file contains the variants and biomarkers in a single file that is based on a Paired Sample (if using PairID) with the file name *_CombinedVariantOutput.tsv. The following variant types and biomarkers are present in the output:

▶   Small Variants (including EGFR complex variants)

▶   Gene Amplifications

▶   TMB

▶   MSI

▶   Fusions

▶   Splice Variants

The combined variant output file also contains Analysis Details and Sequencing Run Details sections. The content of each is listed in the following table.

Analysis Details	Sequencing Run Details
• Pair ID • DNA Sample ID (if DNA is run) • RNA Sample ID (if RNA is run) • Output Date • Output Time • Module Version • Pipeline Version (Docker Image Version #)	• Run Name • Run Date • DNA Sample Index ID (if DNA is run) • RNA Sample Index ID (if RNA is run) • Instrument ID • Instrument Control Software Version • Instrument Type • RTA Version • Reagent Cartridge Lot Number

Variant Filtering Rules

▶   **Small Variants**—All variants with the FILTER field marked as PASS in the merged genome VCF are present in the combined variant output.
   ▶   Gene information is only present for variants belonging to canonical transcripts that are within the Gene Whitelist-Small Variants.
   ▶   Transcript information is only present for variants belonging to canonical transcripts that are within the Gene Whitelist-Small Variants.

▶   **Copy Number Variants**—All copy number variants in the copy number variant VCF are only present if they meet the following conditions:
   ▶   FILTER field marked as PASS.
   ▶   ALT field is <DUP>.

▶   **Combined Variants**—All fusion variants that meet the following criteria:
   ▶   Passing Variant Call (KeepFusion field is true).
   ▶   Contains at least one gene on the fusion whitelist.

▶   **Biomarkers TMB/MSI**—Always present when DNA sample is processed.

▶   Fusion Variants—All fusion variants in the fusions csv are present if they meet the following condition:
   ▶   At least one gene partner is on the Gene Whitelist Fusion Section.

▶ **Splice Variants**—Passing splice variants that are contained on genes EGFR, MET, and AR.

Metrics Output

The MetricsOutput.tsv file contains the following quality control metrics for all of the samples: QC metrics for small variant calling, TMB, MSI, CNV, and Fusion, along with RunQc analysis status and contamination. This TSV file also includes expanded DNA library QC metrics per sample, based on total reads, collapsed reads, chimeric reads, and on-target reads.

DNA Output

Merged Small Variant gVCF

File name: {SAMPLE_ID}_MergedSmallVariants.genome.vcf

The merged variant genomeVCF combines the small variant genome VCF (output of Variant Filtering step) and clinically relevant variants in EGFR exon 19 from Phased Variant calling. This contains information on all candidate small variants evaluated. The variant status is determined by the FILTER column in the genome VCF as follows:

ALT	Filter	Note
.	PASS	WT
., A, C, G and etc.	LowDP	No call (DP < 100X, insufficient depth to confidently detect variants with VF >= 5%)
A, C, G and etc.	PASS	PASS variants
A, C, G and etc.	LowSupport	Filtered variant candidate: • Fail AQ or LQ • 0 stitched support for indel or variant in homopolymer context
A, C, G and etc.	Blacklist	Position with high background noise. Not available for variant detection
A, C, G and etc.	LowVarSupport	Filtered variant candidate with mutant support <1.

Merged Small Variant Annotated JSON

File name: {SAMPLE_ID}_MergedSmallVariantsAnnotated.json.gz

The merged small variants annotated file provides variant annotation information for all non-reference positions from the Merged genome VCF including Pass and non-Pass variants.

TMB Trace

File name: {Sample_ID}_TMB_Trace.tsv

The TMB trace file provides comprehensive information on how the TMB value is calculated for a given sample. All passing small variants from the small variant filtering step are included in this file. To calculate the numerator of the TmbPerMb value in the TMB JSON, set the TSV file filter to use the IncludedInTMBNumerator with a value of True.

The TMB trace file is not intended to be used for variant inspections. The filtering statuses are exclusively set for TMB calculation purposes, and setting a filter does not translate into the classification of a variant as somatic or germline.

Document # 1000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

20

Column	Description
Chromosome	Chromosome
Position	Position of variant
RefCall	Reference base
AltCall	Alternate base
VAF	Variant allele frequency
Depth	Coverage of position
CytoBand	Cytoband of variant
GeneName	Name of gene if applicable. A semicolon delimited list is used for multiple genes.
VariantType	Type of the variant: SNV, insertion, deletion, MNV
CosmicIDs	Cosmic IDs, if multiple concatenated by ";"
MaxCosmicCount	Maximum Cosmic study count
AlleleCountsGnomadExome	Variant allele count in gnomAD exome database
AlleleCountsGnomadGenome	Variant allele count in gnomAD exome database
AlleleCounts1000Genomes	Variant allele count in 1000 genomes database
MaxDatabaseAlleleCounts	Maximum variant allele count over the three databases mentioned above
GermlineFilterDatabase	TRUE if variant was filtered by the database filter
GermlineFilterPRoxi	TRUE if variant was filtered by the proxi filter
CodingVariant	TRUE if variant is in the coding region
Nonsynonymous	TRUE if variant has any transcript annotations with nonsynonymous consequences
IncludedinTMBNumberator	TRUE if variant is used in the TMB calculation

Copy Number VCF

File name: {Sample_ID}_CopyNumberVariants.vcf

This file contains CNV calls for DNA libraries of the amplification genes targeted by TruSight Oncology 500 Analysis Module. The CNV call indicates fold change results for each gene classified as reference, deletion, or amplification.

The value in the QUAL column of the VCF is a Phred transformation of the p-value where Q=-10xlog10 (p-value). The p-value is derived from the t-test between the fold change of the gene against rest of the genome. Higher Q-scores indicate higher confidence in the CNV call.

In the VCF notation, <DUP> indicates the detected fold change (FC) is greater than a predefined amplification cutoff. <DEL> indicates the detected fold change (FC) is less than a predefined deletion cutoff for that gene. This cutoff can vary from gene to gene.

<DEL> calls have only been validated with *in silico* data sets. As a result, all <DEL> calls have LowValidation filter in the VCF.

Each copy number variant is reported in the form of fold change on normalized read depth in a testing sample relative to the normalized read depth in diploid genomes. Given tumor purity, you can infer the ploidy of a gene in the sample from the reported fold change.

Given tumor purity X%, for a reported fold change Y, the copy number n can be calculated by:

$n=[(200 *Y)-2*(100-X)]/X$

For example, a tumor purity at 30% and a MET with fold change of 2.2x indicates that 10 copies of MET DNA are observed.

RNA Output

Splice Variant VCF

File name: {Sample_ID}_SpliceVariants.vcf

Contains all candidate splice variants targeted by the Analysis panel identified by the RNA analysis pipeline. The following filters can be applied for each variant call:

Filter Name	Description
LowQ	Splice Variant Score is < the Passing Quality Score threshold value of 1.
PASS	Splice Variant Score is ≥ to the Passing Quality Score threshold value of 1.

See the headers in the output for more information about each column.

Splice Variant Annotated JSON

Each splice variant is annotated using the Illumina Annotation Engine and the following information is captured in the JSON if available:

▶ HGNC Gene

▶ Transcript

▶ Exons

▶ Introns

▶ Canonical

▶ Consequence

All Fusions CSV

File name: {Sample_ID}_AllFusions.csv

This file contains all candidate fusions identified by the RNA analysis pipeline. Candidate fusions from the splice variant caller are listed in this output with relevant supporting information but are not considered high confidence. Two key output columns in the file describe the candidate fusions: Filter and KeepFusion.

The following table describes the output found in the Filter columns. The output is either a confidence filter or information only as indicated. If none of these filters are triggered, the Filter column displays PASS.

Table 4 Filter Column Output

Filter	Description
Imprecise	**(Confidence filter)** A low-resolution candidate, not an assembled fusion call.
RepeatOverlap	**(Confidence filter)** The fusion is tagged as overlapping with a repeat region. Only used as a confidence filter for nonuniquely mapping fusion candidates, otherwise information only.
WeakBreakend	**(Confidence filter)** The read/alignment evidence on one side of the fusion is weak. Usually this filter indicates that the reads only overlap the fusion by a few base pairs. Alternatively, it can indicate too much homology (no unique sequence).

Homology	(**Information only**) The fusion contig is a substring of another fusion contig.
DuplicateContig	(**Confidence filter**) The two contigs of the fusion are the same contig.
ContigIntragenic	(**Confidence filter**) The realignment of half-contigs produces alignments that map to the same gene on both sides (or within 1 kb if unannotated).
LowQ	(**Confidence filter**) Fusion supporting reads (unique) < 5 (+ 1 for every 10 million reads over 16 million reads).
LowDupReads	(**Information only**) Fusion supporting reads (duplicate) < 5.
NonExonic	(**Information only**) Fusion breakpoint does not fall within an exon.
LocalContigAlign	(**Information only**) Contig realignment found a nonfusion alignment for this contig.
LowFusionRatio	(**Information only**) Few strong evidence reads compared to wild type reads.
NoReferenceReads	(**Information only**) No reads on either side of the presumed breakpoint are marked as reference (structurally normal) reads.

The KeepFusion column of the output has a value of True when the RNAFusionFilter score is ≥ 0.45, none of the confidence filters are triggered, and fusions called by the splice variant caller have a score of 1.

See the headers in the output for more information about each column.

Table 5 Fusion Columns

Fusion Object Field	Source
Caller	(**Either RNAFusionFilter or SpliceGirl**) The algorithm used to identify the fusion.
Gena A	The gene associated with the A side of the fusion. A semicolon delimited list is used for multiple genes.
Gene B	The gene associated with the B side of the fusion. A semicolon delimited list is used for multiple genes.
Gene A Breakpoint	The chromosome and offset of the Gene A side of the fusion.
Gene B Breakpoint	The chromosome and offset of the Gene B side of the fusion.
Score	The quality of fusion as determined by the respective caller. Results from different callers are not equivalent.
Filter	The filter associated with the fusion as determined by the respective caller. Results from different callers are not equivalent.
Precise/Imprecise	(**RNAFusionFilter Only**) Whether the algorithm could identify the precise fusion coordinates. Coordinates listed for imprecise fusions are based on strongest statistical evidence.
Intragenic Call	(**SpliceGirl Only**) List any genes associated with a splice overlapping the fusion. A semicolon delimited list will be used for multiple genes.
Ref A Split	Gene A uniquely mapping reads spanning the junction. Does not support fusion. Duplicate reads included.
Ref A Pair	Gene A uniquely mapping reads paired across junction. Does not support fusion. Duplicate reads included.
Ref B Split	Gene B uniquely mapping reads spanning the junction. Does not support fusion. Duplicate reads included.
Ref B Pair	Gene B uniquely mapping reads paired across junction. Does not support fusion. Duplicate reads included.

Alt Split	Uniquely mapping reads split by the junction. Supports fusion. Duplicate reads included.
Alt Pair	Uniquely mapping reads paired across junction. Supports fusion. Duplicate reads included.
CandidateAlt	**(RFF Imprecise Reads Only)**The number of reads and pairs that potentially supports this candidate before refinement and scoring.
Contig	Sequence of fusion. Can be used to determine fusion directionality. (RFF Only)
ContigAlign1	**(RFF Only)** Length of Gene A in contig.
CntigAlign2	**(RFF Only)** Length of Gene B in contig.
KeepFusion	The determination whether the fusion should be kept or dropped from the list of fusions.
Ref A Dedup	Gene A uniquely mapping reads paired across or split by the junction. Does not support fusion. Duplicate reads are not included.
Ref B Dedup	Gene B uniquely mapping reads paired across or split by the junction. Does not support fusion. Duplicate reads are not included.
Alt Split Dedup	Uniquely mapping reads split by the junction. Supports fusion. Duplicate reads are not included
Alt Pair Dedup	Uniquely mapping reads paired across junction. Supports fusion. Duplicate reads are not included.
Fusion Directionality Known	Whether Fusion Directionality is known and indicated by gene order.

When using Microsoft Excel to view this report, genes that are convertible to dates (such as MARCH1) automatically convert to dd-mm format (1-Mar) by Excel. Fusion whitelist genes are listed below:

▶ ABL1

▶ AKT3

▶ ALK

▶ AR

▶ AXL

▶ BCL2

▶ BRAF

▶ BRCA1

▶ BRCA2

▶ CDK4

▶ CSF1R

▶ EGFR

▶ EML4

▶ ERBB2

▶ ERG

▶ ESR1

▶ ETS1

▶ ETV1

▶ ETV4

▶ ETV5

- ▶ EWSR1
- ▶ FGFR1
- ▶ FGFR2
- ▶ FGFR3
- ▶ FGFR4
- ▶ FLI1
- ▶ FLT1
- ▶ FLT3
- ▶ JAK2
- ▶ KDR
- ▶ KIF5B
- ▶ KIT
- ▶ KMT2A
- ▶ MET
- ▶ MLLT3
- ▶ MSH2
- ▶ MYC
- ▶ NOTCH1
- ▶ NOTCH2
- ▶ NOTCH3
- ▶ NRG1
- ▶ NTRK1
- ▶ NTRK2
- ▶ NTRK3
- ▶ PAX3
- ▶ PAX7
- ▶ PDGFRA
- ▶ PDGFRB
- ▶ PIK3CA
- ▶ PPARG
- ▶ RAF1
- ▶ RET
- ▶ ROS1
- ▶ RPS6KB1
- ▶ TMPRSS2
- ▶ TP53

Technical Assistance

For technical assistance, contact Illumina Technical Support.

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

Illumina Customer Support Telephone Numbers

Region	Toll Free	Regional
North America	+1.800.809.4566	
Australia	+1.800.775.688	
Austria	+43 800006249	+43 19286540
Belgium	+32 80077160	+32 34002973
China	400.066.5835	
Denmark	+45 80820183	+45 89871156
Finland	+358 800918363	+358 974790110
France	+33 805102193	+33 170770446
Germany	+49 8001014940	+49 8938035677
Hong Kong, China	800960230	
Ireland	+353 1800936608	+353 016950506
Italy	+39 800985513	+39 236003759
Japan	0800.111.5011	
Netherlands	+31 8000222493	+31 207132960
New Zealand	0800.451.650	
Norway	+47 800 16836	+47 21939693
Singapore	+1.800.579.2745	
South Korea	+82 80 234 5300	
Spain	+34 911899417	+34 800300143
Sweden	+46 850619671	+46 200883979
Switzerland	+41 565800000	+41 800200442
Taiwan, China	00806651752	
United Kingdom	+44 8000126019	+44 2073057197
Other countries	+44.1799.534000	

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

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

Document # 10000000099600 v00

For Research Use Only. Not for use in diagnostic procedures.

26

illumina®