

TruSight[®] Tumor 15 Data Analysis

On-instrument data analysis enables simple and efficient variant calling for novice users.

Introduction

TruSight Tumor 15 is a robust DNA-to-data workflow designed to detect variants in 15 genes that are commonly mutated in solid tumors (Figure 1). The assay uses an amplicon-based, multiplexed PCR approach to detect low frequency somatic variants from formalin fixed, paraffin embedded (FFPE) tissue. To avoid primer dimers, 2 distinct primer pools are prepared and sequenced as separate libraries. The assay supports 1-24 samples (2-48 libraries) with a 2 × 150 bp sequence run configuration.

The sequencing workflow is compatible with both the MiSeq[®] and MiniSeq[™] Systems (Figure 1). Data analysis and reporting are supported on instrument and on the Illumina BaseSpace[®] platform. This technical note illustrates the ease of use with the software workflow, and outlines the analysis steps that the software performs automatically.

Software Methodology and Usage

For local analysis of TruSight Tumor 15 results, software is available using MiSeq Reporter version 2.6 for the MiSeq System, and using Local Run Manager for the MiniSeq System. TruSight Tumor 15 analysis modules can be downloaded, free of charge, onto the sequencing system. Following installation, software analysis begins automatically after selected sequencing runs. Researchers can also use the TruSight Tumor 15 BaseSpace App, available on the BaseSpace Sequence Hub.

During analysis, the TruSight Tumor 15 Software processes the sequencing files produced by the instrument and generates per-sample reports. The core analysis pipeline (Figure 2) is consistent across each of the software options available for TruSight Tumor 15, from read alignment through the sample QC metrics report. With the

analysis steps automated, the user is not required to do anything after loading the samples onto the sequencer and starting the run. (Figure 2).

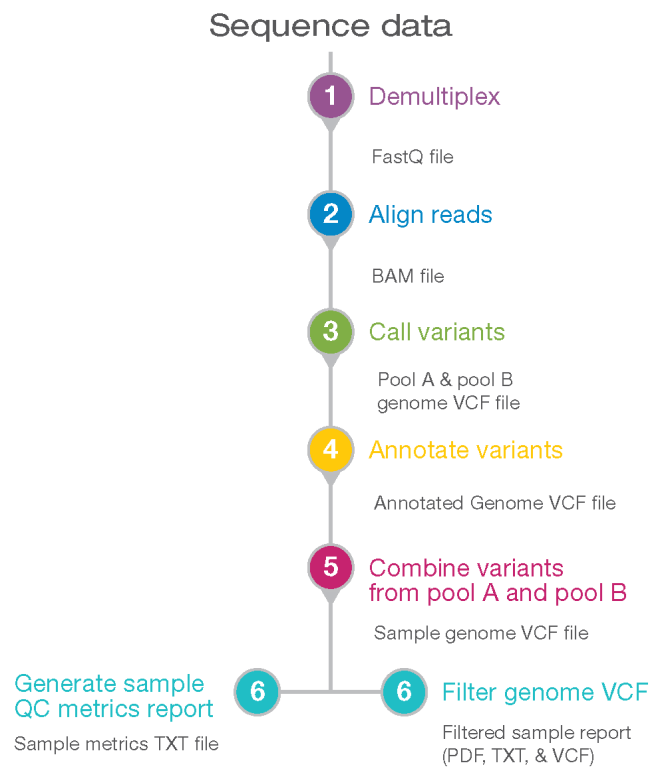


Figure 2: TruSight Tumor 15 Software Steps—Once the samples are loaded and the run is selected on the sequencing system, each step in the data analysis pipeline is automated.

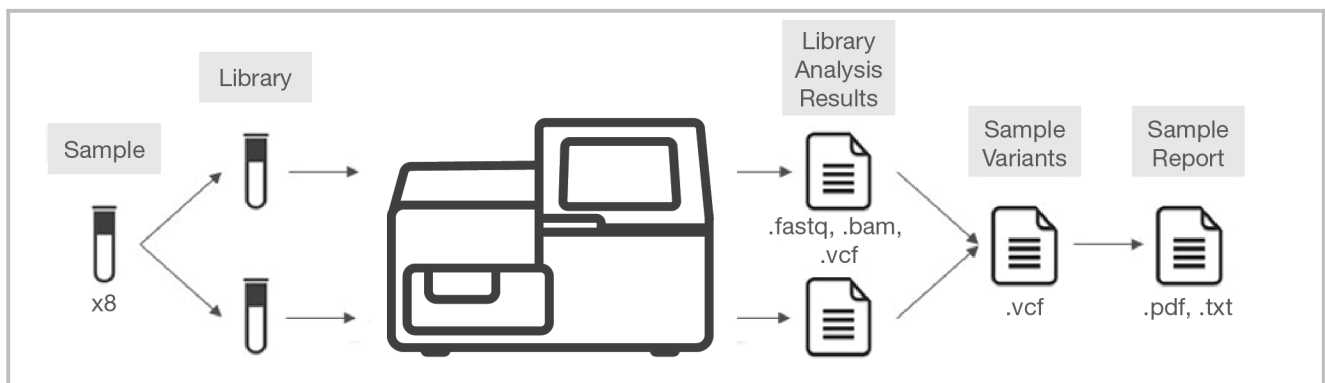


Figure 1: The TruSight Tumor 15 Analysis Workflow—TruSight Tumor 15 is optimized for seamless integration into current lab workflows, going from extracted DNA to data in approximately 36 hours. The assay can be run on the MiSeq or MiniSeq Series of sequencing systems.

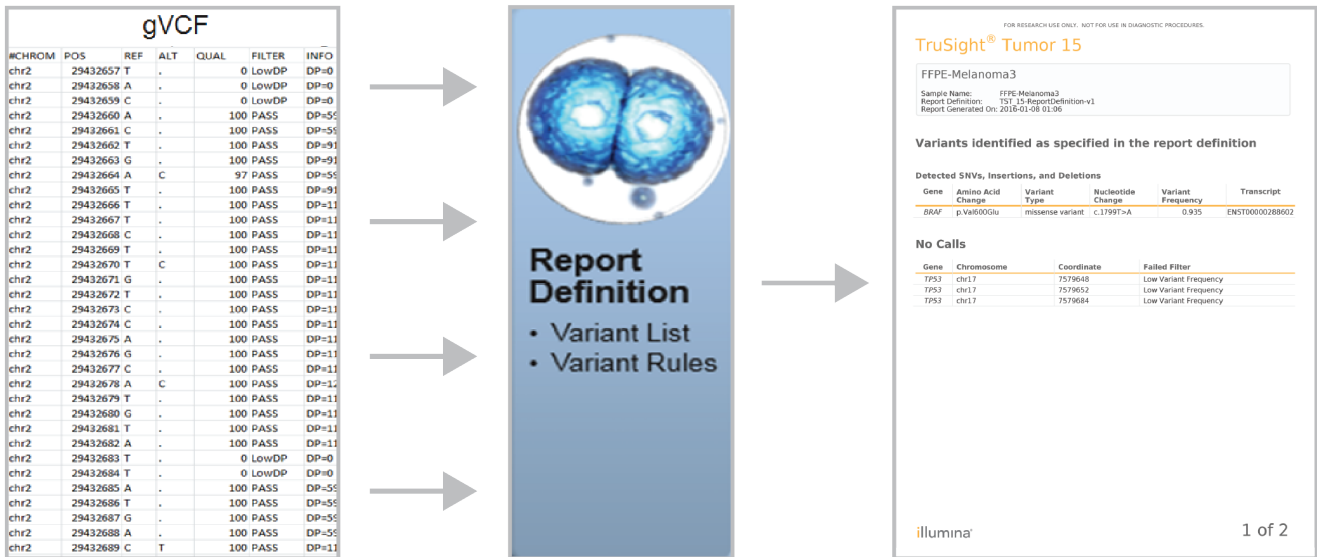


Figure 3: Variant Filtering—The TruSight Tumor 15 Software uses a static list of variants and variant rules to generate a sample report.

Comprehensive Somatic Variant Calling Solution

TruSight Tumor 15 software analysis uses a banded Smith Waterman algorithm for read alignment. The reference genome used for alignment is the UCSC hg19 human build. Targets for each pool are specified in the amplicon manifest files included with the software. To detect variants better at the edge of reads, the workflow configures the aligner to stitch overlapping paired end reads.

Variants are called using the Illumina Somatic Variant Caller, and algorithm developed by Illumina and built into the software.¹ Candidate variants are assessed with a Poisson model using minimum emit frequency of 1% and minimum emit quality of 20. Variants are then run through quality control filters for coverage, strand bias, Q-score, variant frequency, and reference repeat regions for insertions/deletions (indels). By consolidating calls for known variants that are reported as smaller variants, the workflow reduces potential for false negatives due to alternate alignments. The Somatic Variant Caller outputs a genome VCF (gVCF) file that describes all passing and filtered reference and variant calls for each targeted coordinate. Using the Illumina Annotation Engine, variants contained in the Report Definitions file are annotated with gene name, nucleotide change, amino acid change, and consequence. Annotations are aligned with the canonical Ensembl transcript.²

For each sample in a run, 2 libraries (pools A and B) are analyzed in parallel. Because pool A and pool B have some overlapping regions, creating 2 pools eliminates primer dimers and avoids extra clean-up steps. To simplify the workflow, the paired gVCFs are automatically combined into 1 sample output after variant calling (Figure 1). The software combines the results with a lookup file that identifies which

library is the source of variant calls for each coordinate. The final output is a VCF file ({sample name}.sample.genome.vcf), and intermediate pool-specific files are stored in the libraries subfolder.

User-Friendly Analysis and Reporting

The sample gVCF file contains all variant-calling data that can be used for downstream user filtering and reporting. The TruSight Tumor 15 Software provides a filtered view of the sample data by using a static list of variants (Report Definitions file) and variant rules to generate a sample report in *.pdf, *.txt, and *.vcf formats (Figure 3). End users cannot currently configure the list of variants and rules in the report definition filter. Variants of interest are defined in the report definition filter by specified chromosome, coordinate, reference allele, and alternate allele. Variant rules of interest are defined in the report definition filter by specified start/end region and predicted consequence type. When a sample gVCF contains the specified consequence in the specified region, the associated variant is added to the sample report. In the cases where a variant of interest is present in the sample gVCF, but flagged for not passing QC analysis, it is added to the list of “No Calls” in the report.

TruSight Tumor 15 software is designed to execute all steps in the workflow, providing a sample gVCF and a sample report for each sequencing run. To inform the user on the quality of resulting outputs, a Run Metrics Report and Sample Metrics Report are available (Figure 4). At a minimum for sample analysis success, it is recommended that the sequencing run meets Illumina guidelines for 80% Q30 score, and 85% reads passing filter. After secondary analysis, each sample is then individually assessed for quality using the Sample Metrics report. (Figure 4).

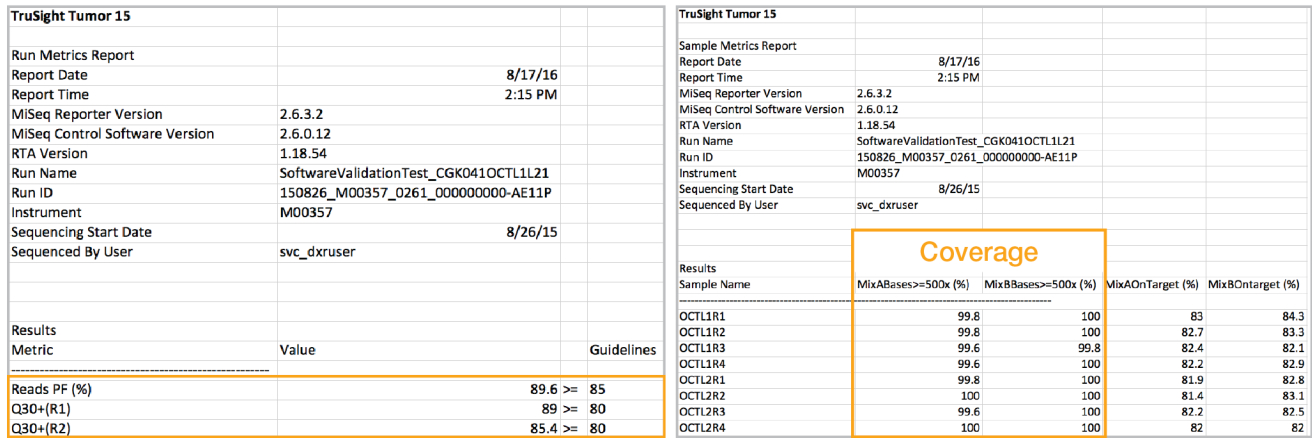


Figure 4: Run Metrics Reports and Sample Metrics Reports—Users can access the success of each run, and for each sample. Each run should meet the specifications of 80% Q30 score, and 85% reads passing filter. Each library should meet the coverage specification of 93.5% of bases covered at $\geq 500x$ minimum. If libraries fall below this spec, libraries should either be repeated or analyzed more thoroughly.

Sensitive, High-Confidence Variant Detection

TruSight Tumor 15 Software is built under Illumina software development life cycle procedures, with emphasis on traceability and quality; and rigorously tested for bioinformatics accuracy. Testing activities include low-level software code unit tests, nightly regression testing, integration testing, and validation testing. For example, custom FASTQ files are built to simulate large numbers of variants of different types: single nucleotide variants (SNVs), multiple nucleotide variants (MNVs), and indels. These custom FASTQ files are run through the product analysis pipeline, and then resulting variants and annotations are verified for accuracy. Data sets include thousands of simulated variants in regions targeted by the TruSight Tumor 15 Assay. The simulated variants are derived from the COSMIC database.¹ In addition to simulation testing, software testing includes runs on the MiSeq and MiniSeq Sequencing Systems using DNA extracted from FFPE tumor samples with known mutations. As a result, TruSight Tumor 15 Software is optimized to perform seamlessly and confidently in the product workflow.

To provide end users with the option to perform further interpretation of data as needed, BaseSpace Variant Interpreter (Beta) and VariantStudio software are both capable of loading and displaying VCF files for custom filtering, annotation, and reporting. The standard VCF output is also compatible with other third-party interpretation tools. TruSight Tumor 15 is capable of scaling for inclusion of new features in the future, with no need for resequencing.

Summary

The TruSight Tumor 15 Assay analyzes solid tumor samples for the detection of variants at low frequency. Compatible with the MiSeq and the MiniSeq Systems, the assay is part of an integrated workflow that includes library prep, sequencing, and data analysis (Figure 1). Following the library prep stage, the sequencing and data analysis steps require little user intervention. With automated on-instrument

data analysis, users simply have to select the run, start, and then print the final report. The TruSight Tumor 15 Software platform is designed to minimize the amount of time required from the user for data analysis, while also providing high-quality results. With numerous automated steps, the software platform sorts indexed samples (demultiplexing), aligns reads to the reference genome, performs variant calling, and assesses sequence data quality for each sample. The resulting variant report is simple to interpret (Figure 3). Together these features build confidence with reproducible and accurate results in an easy-to-use format.

Learn More

For more information about targeted sequencing with solid tumors, visit: www.illumina.com/OncologyPanel.

References

1. Illumina (2014) Somatic Variant Caller. (www.illumina.com/documents/products/technotes/technote_somatic_variant_caller.pdf).
2. Ensembl. uswest.ensembl.org/index.html. Accessed August 31, 2016.
3. Catalogue of Somatic Mutations in Cancer. cancer.sanger.ac.uk/cosmic. Accessed August 31, 2016.