illumina®

Using TruSight™ Oncology UMI Reagents with TruSight Tumor 170 DNA Content for Detection of Low-Frequency Variants in Cell-Free DNA

Using error correction software with next-generation sequencing (NGS) to lower the limit of detection for rare variants in cell-free DNA (cfDNA).

Introduction

Analysis of cfDNA is increasingly used to obtain information about the genetic state of normal and tumor tissues. Combining liquid biopsy with NGS technology provides a noninvasive method to analyze numerous cancer-related genes in a single assay. However, the low percentage of circulating tumor DNA (ctDNA) within total cfDNA causes variant allele frequencies to exist near the limit of detection (LOD). To address this challenge, Illumina offers the TruSight Oncology UMI Reagents, with unique molecular identifiers (UMIs) and error correction software, to lower the rate of inherent errors in NGS data. With significant reduction of the background noise level, low-frequency variant calling in NGS data can be performed with higher confidence.

To explore the possibility of analyzing cancer-related genes in cfDNA, targeted molecular analysis was performed using the TruSight Oncology UMI Reagents paired with DNA content from TruSight Tumor 170. This application note demonstrates implementation of UMIs within the TruSight Oncology workflow, with error correction and resultant low LOD during variant analysis of cfDNA.

Methods

DNA samples

Four different sample types were used in the study. Blood collection was performed with two different methods to provide cfDNA samples, and two different control DNA samples were used.

cfDNA-Streck: Blood samples from healthy individuals were collected in Cell-Free DNA BCT blood collection tubes (Streck) and kept at room temperature for up to two days prior to processing.

cfDNA-EDTA: For analysis of cfDNA samples with variants, non-small cell lung cancer (NSCLC) research samples were provided via a collaboration with the University of Pennsylvania Perelman School of Medicine. Blood samples were collected in Vacutainer Plus Plastic K_2EDTA Tubes (BD Biosciences) and centrifuged within three hours of collection.

Nucleosomal control DNA: Because nuclease-treated chromatin results in DNA that resembles cfDNA in size distribution (~170 bp), nucleosomal control DNA was prepared from the LoVo cancer cell line (ECACC 87060101) containing two known single nucleotide variants (SNVs). A method using the EZ Nucleosomal DNA Prep Kit (Zymo Research) is described in more detail in a technical note provided by Illumina.[1]

Horizon control DNA: Horizon HD753 control DNA (Horizon) contains known variants including three SNVs, two insertions/deletions (indels), and two copy number variants (CNVs). Horizon control DNA was sheared during preparation. Because the library conversion efficiency of mechanically sheared DNA is lower than cfDNA, more Horizon DNA was used for library prep.

For both blood sample types, DNA was extracted with the QIAamp Circulating Nucleic Acid Kit (QIAGEN) according to manufacturer instructions , except that twice the recommended plasma volume was bound to the spin column. Quantification was performed on a Fragment Analyzer using the High Sensitivity Large Fragment Analysis Kit (Advanced Analytical). Sample concentration was measured using the 75–250 bp range for cfDNA and the 50–700 bp range for Horizon control DNA. For each library prep made with cfDNA or nucleosomal control DNA, 30 ng cfDNA was used. For Horizon control DNA, 75 ng input DNA was used.

Library preparation

Libraries were prepared according to the reference guides for the TruSight Oncology UMI Reagents[2] and TruSight Tumor 170[3] following the DNA workflow only. Because cfDNA is already small when extracted, the DNA shearing step was omitted from the protocol during library preparation for blood samples, and control DNA samples were sheared mechanically or enzymatically prior to use in the assay.

Sequencing

Instrument compatibility was demonstrated with three different Illumina sequencing systems. The indicated sample distribution for each flow cell resulted in ~40,000× coverage per sample. The HiSeq™ 2500 System was run in high-output mode, six samples per flow cell at 2 × 125 bp. The HiSeq 4000 System was run with eight samples per flow cell at 2 × 150 bp. The NovaSeq™ 6000 System was run with ten samples per S2 flow cell at 2 × 150 bp.

Data analysis

For detection of variants at low allele frequency, the strategy was to reduce artifacts so that signals from true variants could be distinguished from noise. To analyze results, the following steps were followed:

Read collapsing and error correction: Data was streamed to BaseSpace™ Sequence Hub and FASTQ files were analyzed using the BaseSpace Sequence Hub UMI Error Correction App. After alignment with the HG19 reference genome, the UMI Error Correction App collapsed sequence data to unique reads based on the start and stop position of the read fragments (read duplicates) and the UMI barcodes. This resulted in collapsed reads (called simplex families) with low noise and high confidence. Read families with both the forward and reverse strands represented were further collapsed into a duplex family, which further reduced the error rate due to the low probability of having identical errors on both strands. The output of the UMI Error Correction App was a collapsed BAM file.

Small variant calling: Collapsed BAM files were analyzed using an internal Illumina pipeline. An optimized version of the Illumina Pisces Variant Caller was used to call low-frequency SNVs, multiple nucleotide variants (MNVs), and indels. The output of the Pisces Variant Caller is a genome VCF (gVCF) file.

Baseline construction and variant filtering: To reduce background noise containing false positive variant calls, results were compared with the baseline distribution of variant allele frequency (VAF) in a collection of 38 healthy cfDNA samples. Using this baseline, variant calling confidence was adjusted in genomic regions with varying background noise levels, improving the calling accuracy.

CNV detection: A CNV baseline was also generated and was used to remove coverage bias related to probe capture efficiency. For each CNV targeted by the TruSight Tumor 170 panel, the fold change was calculated by comparing the normalized coverage of the targeted gene to the rest of the genome.

Suggested quality control (QC) metrics: Following error correction and sequencing, individual samples were assessed for median target coverage (MTC) and mean family depth (MFD). MTC is the median number of read families spanning a target region, and the suggested MTC for calling variants at 0.4% VAF is ≥ 1500×. MFD is the average number of collapsed reads within each UMI family. MFD should be ≥ 10 to ensure sufficient depth for error correction. All results in this study met these criteria.

Concordance analysis

Samples with known variants were titrated to low expected VAF values, then aliquoted and analyzed with two independent methods. NGS was performed using TruSight Oncology UMI Reagents with Trusight Tumor 170 DNA oligos, along with associated UMI Error Correction software. For simplicity, this method will be referred to as TST170-UMI. As a comparator method, droplet digital PCR (ddPCR) was done with the QX200 Droplet Digital PCR System (Bio-Rad).

Results

Error correction using UMIs

A set of 31 cfDNA-Streck samples from healthy blood donors were sequenced on the HiSeq 4000 System, and the average error rate and standard deviation were calculated. The UMI sequences were trimmed and read collapsing was performed to reduce errors, then the adjusted error rate was calculated again for all 31 samples (Table 1).

Table 1: Error rates before and after error correction with UMI-based read collapsing

Metric	Error Rate (Uncorrected)	Error Rate (With Error Correction)
Average	0.050%	0.0024%
Standard deviation	0.020%	0.0005%
Maximum value	0.093%	0.0041%
Minimum Value	0.033%	0.0017%

Instrument compatibility

To demonstrate compatibility of the TruSight Oncology UMI Reagents with different sequencing instruments, cfDNA-Streck and nucleosomal control DNA samples were processed and sequenced on the HiSeq 2500, HiSeq 4000, and NovaSeq 6000 Systems. After error correction, the average reduced error rate was calculated (Figure 1). Results on all three instruments were below the maximum error rate of 0.007% that is expected for high confidence variant calling.

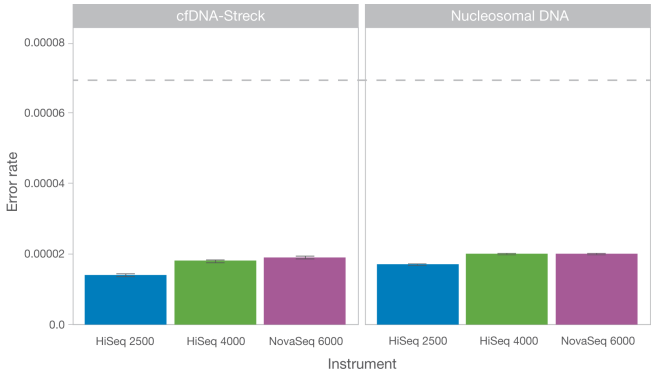


Figure 1: Error correction as indicated by the error corrected error rate for samples run on different Illumina sequencing instruments—The average of three libraries (HiSeq 2500 System), four libraries (HiSeq 4000 System), and five libraries (NovaSeq 6000 System) are shown for cfDNA-Streck and nucleosome prepared DNA, respectively. Average data from two runs on each instrument are shown. Error rate is expected to be below 0.007%, indicated by the dashed line.

Limit of detection

To examine the impact of error reduction on the LOD in variant calling, DNA samples with known variants were analyzed. Horizon control DNA contains three known SNVs, two indels, and two CNVs. With VAF values provided by the manufacturer, the Horizon control DNA was titrated with wild-type DNA (NA12878, Coriell Institute) to achieve an expected VAF range of 0.2–2%, then prepped and sequenced on the HiSeq 2500 System. With MTC values above the recommended metrics, the observed analytical sensitivity for the five small variants

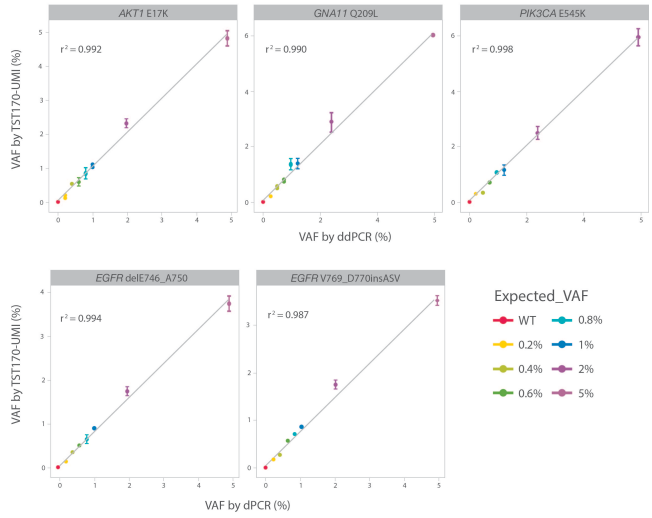
was 100% down to 0.4% expected VAF, and 93.33% at 0.2% expected VAF (Table 2). To correlate measured VAF values with an independent method, Horizon control DNA samples were titrated to an expected VAF range of 0.2–5% and analyzed with both TST170-UMI and ddPCR. High correlation values were observed ($r^2 \geq 0.987$) between the two methods for detection of SNVs, indels, and CNVs (Figure 2, Tables 3–5).

Table 2: Median target coverage and analytical sensitivity of titrated Horizon control DNA

Expected VAF	MTC	Analytical Sensitivity
2.0%	3768	100% (15/15)
1.0%	4250	100% (15/15)
0.80%	3912	100% (15/15)
0.60%	4103	100% (15/15)
0.40%	3334	100% (15/15)
0.20%	3999	93.33% (14/15)
0%	4179	0% (0/15)

Expected VAF calculated according to variant with the lowest frequency (*EGFR* indel). Analytical sensitivity represents detection of five variants across three replicates for each titration point. MTC value of ≥ 1500× is recommended for low-frequency variant calling.

A.



B.

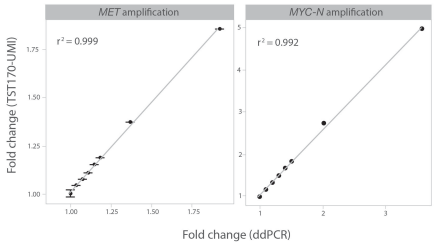


Figure 2: Correlation between TruSight Tumor 170 (TST170-UMI) and digital droplet PCR (ddPCR) with titrated Horizon control DNA—(A) VAF values are plotted between TST170-UMI and ddPCR for three SNVs and two indels. (B) Fold-change values are plotted for two CNVs. Samples for each titration level were measured in triplicate by TST170-UMI.

Table 3: Correlation between TruSight Tumor 170 (TST170-UMI) and digital droplet PCR (ddPCR) for SNV detection

	GNA11 Q209L		*AKT1* E17K		*PIK3CA* E545K	
Expected VAF	TST170-UMI	ddPCR	TST170-UMI	ddPCR	TST170-UMI	ddPCR
5.0%	6.02%	5.97%	4.84%	4.90%	6.05%	5.63%
2.0%	2.91%	2.39%	2.33%	1.96%	2.61%	2.25%
1.0%	1.43%	1.19%	1.08%	0.98%	1.19%	1.13%
0.80%	1.40%	0.95%	0.87%	0.78%	1.13%	0.90%
0.60%	0.82%	0.72%	0.62%	0.59%	0.75%	0.68%
0.40%	0.56%	0.48%	0.57%	0.39%	0.38%	0.45%
0.20%	0.24%	0.24%	0.25%	0.20%	0.32%	0.23%
0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 4: Correlation between TruSight Tumor 170 (TST170-UMI) and digital droplet PCR (ddPCR) for indel detection

	EGFR V769_D770insASV		*EGFR* delE746_A750	
Expected VAF	TST170-UMI	ddPCR	TST170-UMI	ddPCR
5.0%	3.51%	4.86%	3.73%	4.84%
2.0%	1.75%	1.95%	1.75%	1.93%
1.0%	0.90%	0.97%	0.95%	0.97%
0.8%	0.74%	0.78%	0.66%	0.77%
0.6%	0.62%	0.58%	0.53%	0.58%
0.4%	0.25%	0.39%	0.36%	0.39%
0.2%	0.20%	0.19%	0.17%	0.19%
0%	0.00%	0.00%	0.00%	0.00%

Table 5: Correlation between TruSight Tumor 170 (TST170-UMI) and digital droplet PCR (ddPCR) for CNV detection

MET amplification		*MYC-N* amplification	
TST170-UMI fold change	ddPCR fold change	TST170-UMI fold change	ddPCR fold change
1.856	1.920	4.966	3.558
1.372	1.368	2.712	2.024
1.186	1.184	1.840	1.512
1.151	1.147	1.677	1.410
1.109	1.110	1.499	1.307
1.075	1.047	1.325	1.205
1.043	1.037	1.162	1.102
1.001	1.000	0.995	1.000

Nucleosomal control DNA from the LoVo cancer cell line containing two known SNVs (*KRAS* G13D and *FBXW7* R505C) was titrated with nucleosome-prepared DNA from a normal cell line to achieve an expected VAF range of 0.14–1%, prepped and sequenced on the HiSeq 4000 System. With three replicates for each titration level, analytical sensitivity observed was 100% down to 0.14% VAF (Table 6). Nucleosomal control DNA samples were analyzed with both TruSight Tumor 170 and ddPCR, demonstrating a strong correlation (R2 ≥ 0.92) between observed VAF values for both methods (Figure 3 and Table 7).

Table 6: Analytical sensitivity and median target coverage of titrated nucleosomal control DNA

Expected VAF (*FBXW7*)	MTC	Analytical Sensitivity
1.00%	1786	100% (6/6)
0.70%	1745	100% (6/6)
0.54%	1568	100% (6/6)
0.40%	1687	100% (6/6)
0.27%	1288	100% (6/6)
0.14%	2097	100% (6/6)

Nucleosomal control DNA was titrated to expected VAF for the lower frequency of the two known variants. Analytical Sensitivity represents detection of two variants across three replicates for each titration point. MTC value of ≥ 1500× is recommended for low frequency variant calling.

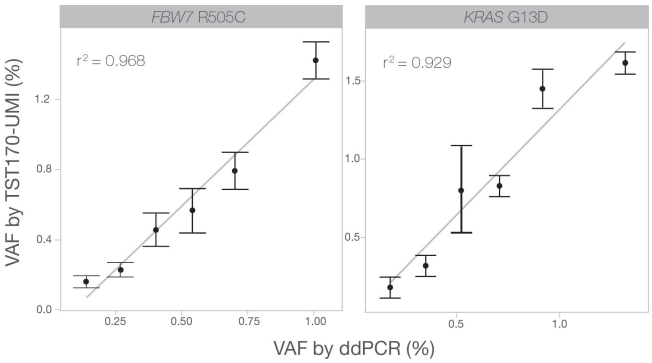


Figure 3: Correlation of observed VAF values between TruSight Tumor 170 (TST170-UMI) and ddPCR in titrated nucleosomal control DNA—Samples were measured by TST170-UMI in triplicate at each titration point.

Table 7: Variant calling results in titrated nucleosomal control DNA with TruSight Tumor 170 (TST170-UMI) and ddPCR

SNV					
FBXW7 R505C			*KRAS* G13D		
Expected VAF	TST170-UMI	ddPCR	Expected VAF	TST170-UMI	ddPCR
1.00%	1.42%	0.93%	1.30%	1.61%	1.56%
0.70%	0.79%	0.70%	0.92%	1.45%	1.21%
0.54%	0.57%	0.51%	0.71%	0.83%	0.87%
0.40%	0.46%	0.54%	0.52%	0.80%	0.84%
0.27%	0.23%	0.32%	0.35%	0.32%	0.53%
0.14%	0.16%	0.14%	0.18%	0.18%	0.30%

VAF values represent the averages of triplicate samples.

Demonstration of low-frequency variant detection in NSCLC research samples

To demonstrate variant detection in blood samples, a pool of 15 NSCLC research samples were collected (cfDNA-EDTA) and extracted DNA was used for library preparation similarly to the cfDNA–Streck samples, and sequenced on the HiSeq 4000 System. Three samples were identified with mutations at < 10% VAF, including three SNVs and two CNVs. The detection and frequency of each mutation was confirmed with ddPCR (Table 8).

Table 8: Detection of SNVs and CNVs in NSCLC research samples

SNV Detection			
Sample	Variant	VAF (TST170-UMI)	VAF (ddPCR)
NSCLC-1	*PIK3CA* E545K	0.10%	0.02%
NSCLC-1	*KRAS* G12V	0.49%	0.41%
NSCLC-2	*TP53* R282W	6.09%	5.77%
CNV Detection			
Sample	Variant	Fold Change (TST170-UMI)	Fold Change (ddPCR)
NSCLC-3	*FGF10* Duplication	1.71	1.63
NSCLC-3	*RICTOR* Duplication	1.62	1.57

Summary

Analysis of cfDNA holds promise for the development of noninvasive cancer diagnostics and monitoring. By integrating the TruSight Oncology UMI Reagents with the DNA workflow of TruSight Tumor 170, the benefit of UMIs for error reduction in cfDNA sequencing was demonstrated with cfDNA from two different blood collection methods, and across three Illumina sequencing platforms. A significant reduction in error rate enabled detection of SNVs, indels, and CNVs with analytical sensitivity consistently at 0.4% VAF or lower. High concordance was observed when comparing VAF values and fold-change value with ddPCR for detected SNVs and CNVs. TruSight Tumor 170 provides comprehensive coverage of cancer-related genes, enabling detection of unknown variants in a single assay. With the implementation of the TruSight Oncology UMI Reagents, improvements in accuracy and analytical sensitivity bring the value of NGS into the hands of translational researchers using cfDNA analysis.

Learn more

For more information about the TruSight Oncology UMI Reagents, visit www.illumina.com/UMI-Reagents

For more information about TruSight Tumor 170, visit www.illumina.com/TruSightTumor170

References

1. Illumina (2018) Options for Control Materials for NGS Analysis of Circulating Cell-Free DNA. (www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/trusight-umi-cfDNA-control-technote-1170-2017-020.pdf).
2. Illumina (2018) TruSight Oncology UMI Reagents Workflow Integration Guide. (support.illumina.com/sequencing/sequencing_kits/trusight-oncology-umi-reagents.html).
3. Illumina (2017) TruSight Tumor 170 Reference Guide. (support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/trusight/tumor-170/trusight-tumor-170-reference-guide-1000000024091-01.pdf).

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com