# Error Rate Reduction with the TruSight™ Oncology UMI Reagents for Variant Detection

Factors that impact analytical sensitivity and specificity of variant calling in next-generation sequencing (NGS).

## Introduction

Using NGS to detect rare variants, such as those found in circulating cell-free DNA (cfDNA), requires high sequencing depth and the ability to distinguish true low-frequency variants from noise. The TruSight Oncology UMI Reagents combine unique molecular identifiers (UMIs) with error correction software* to enable error rate reduction from 0.5% to ≤ 0.007%.[1] Noise reduction is achieved by filtering false variants based on alignment of UMI barcodes and subsequent read collapsing (Figure 1). This process removes false positives from the collapsed reads, enabling more accurate calling at low variant frequency.

Independent of UMI-based error correction, several other factors impact the accuracy of low-frequency variant calling, and the achieved limit of detection (LOD). This technical note reviews variables occurring during experimental setup that influence analytical sensitivity and specificity.

### DNA input

Sometimes due to scarcity of sample, users are faced with creating libraries with less than optimal input amounts. Decreasing DNA input results in a lower number of genome equivalents and reduced library diversity. Illumina recommends starting with 30 ng input, which represents approximately 9000 genome equivalents. Reducing the input to 10 ng reduces the genome equivalents present to ~3000. This has a marked impact on the ability to detect low-frequency variants accurately because a mutation present at 0.4% variant allele frequency (VAF) would be represented by only 12 starting copies at 10 ng versus 36 copies from 30 ng input.

### Conversion efficiency

Conversion efficiency describes the percentage of DNA input molecules that are successfully converted into a sequencing library. If the conversion efficiency is 50%, then from 10 ng input only 5 ng (1500 genome equivalents) are sequencable. From 1500 genome equivalents, a mutation present at 0.4% VAF would be represented by six individual starting molecules. If the conversion efficiency was only 25%, then that number would drop to three variant-containing molecules, which may produce inconsistent results.

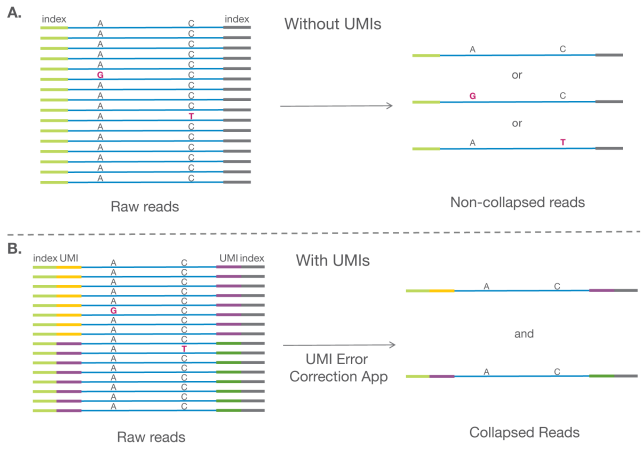


**Figure 1: Error correction with the TruSight Oncology UMI Reagents**—(A) Shown are 16 reads with two variants that could either be true rare variants or inherent errors. Without error correction, it is impossible to distinguish between true variants and false positives. (B) Integration of UMIs enables the UMI Error Correction App to recognize multiple reads from the same starting molecule and collapse them into a single read. Each set of reads contains one error. After error correction, only one correct sequence remains.

### Sequencing depth

When looking for low-frequency variants, increasing the sequencing depth maximizes the probability of their detection. To ensure accurate variant calling, Illumina recommends sequencing to 40,000× raw sequencing depth. With 30 ng DNA input and 25% conversion efficiency, this would result in a median target coverage (MTC) of 2500×. MTC is the median number of read families spanning a target region (Figure 2). This means that after read collapsing, the median collapsed fragment coverage for all bases in the panel is 2500×, and typically > 85% of reads have coverage of ≥ 1500×, which is recommended for calling variants at a frequency of 0.4%.

The simplest way to increase sequencing depth is to load fewer samples per sequencing run, which sacrifices throughout (samples per run) for increased depth (coverage per sample). Users can consult the provided simulated calculations as a guide to assess the potential outcomes when suboptimal DNA inputs or coverages are obtained (Table 1).

* The UMI Error Correction App is available in the cloud-based BaseSpace™ Sequence Hub, or for installation on a local pipeline.

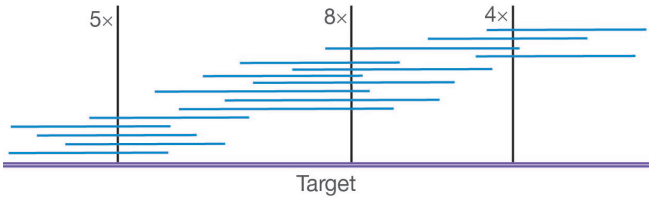


**Figure 2: Median target coverage**—Horizontal lines represent sequencing reads for the targeted region. Vertical lines indicate which reads cover three specific bases. Sequencing depths and positions shown are for illustration only. Every base across a target is considered. The MTC for the region shown is 5×. Recommended MTC for 0.4% LOD following error correction is ≥ 1500×.

## Analytical sensitivity and specificity

Analytical sensitivity is defined as the ability to identify a variant correctly that is present (the true positive rate). Analytical specificity is defined as the ability not to call a variant when it is absent (the true negative). When sequencing depth decreases, the accuracy of low-frequency variant calling is also reduced.

Increasing sequencing depth has a significant impact on the analytical sensitivity of the assay. For example, with 30 ng input and a 25% conversion rate, the analytical sensitivity at 0.2% VAF is 54.77% at 10,000× raw coverage depth, but increasing raw depth to 40,000× raises the analytical sensitivity to 80.02%. Similarly, at 0.4% LOD, the analytical sensitivity increases from 92.60% at 10,000× raw depth to 99.12% at 40,000× raw depth. (Table 1).

In contrast to analytical sensitivity, analytical specificity decreases with higher coverage depth because increasing the number of reads also increases the possibility of false positives. However, analytical specificity decreases at a significantly lower rate and scope than the associated increase in analytical sensitivity. For example, for 30 ng input at 0.2% LOD, increasing the raw coverage from 10,000× to 40,000× raises analytical sensitivity from 54.77% to 80.02%, while analytical specificity only drops from 99.99% to 99.98% (Table 1).

## Summary

Accuracy is an important consideration with variant detection. Error correction methods, such as the TruSight Oncology UMI Reagents, can help to remove noise, thereby increasing the accuracy of variant calling at low VAF. Analytical sensitivity and analytical specificity vary according to factors such as DNA input and sequencing depth. Users are sometimes faced with experimental inputs that do not meet recommended guidelines, and this technical note can be used as a guide to consider such factors when making decisions regarding experimental designs.

## Learn more

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

## References

1. Illumina (2018) TruSight Oncology UMI Reagents. (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/trusight-oncology-umi-reagents-datasheet-1000000050425.pdf).

**Table 1: Expected analytical specificity and analytical sensitivity from varied DNA inputs and coverage depths**

| Input | Raw Coverage Depth | Analytical Specificity | Analytical Sensitivity at Noted LOD | | | |
|---|---|---|---|---|---|---|
| | | | LOD = 0.20% | LOD = 0.40% | LOD = 0.60% | LOD = 0.80% |
| 10 ng | 10,000× | 100% | 7.26% | 30.06% | 54.86% | 73.74% |
| | 20,000× | 100% | 13.72% | 46.97% | 73.65% | 88.48% |
| | 30,000× | 100% | 15.63% | 51.06% | 77.34% | 90.82% |
| | 40,000× | 100% | 17.27% | 54.32% | 80.07% | 92.42% |
| 30 ng | 10,000× | 99.99% | 54.77% | 92.60% | 99.17% | 99.92% |
| | 20,000× | 99.98% | 73.62% | 98.22% | 99.92% | 100% |
| | 30,000× | 99.98% | 77.28% | 98.79% | 99.96% | 100% |
| | 40,000× | 99.98% | 80.02% | 99.12% | 99.98% | 100% |
| 50 ng | 10,000× | 99.93% | 85.64% | 99.61% | 99.99% | 100% |
| | 20,000× | 99.90% | 95.33% | 99.97% | 100% | 100% |
| | 30,000× | 99.92% | 96.54% | 99.99% | 100% | 100% |
| | 40,000× | 99.93% | 97.32% | 99.99% | 100% | 100% |
| 70 ng | 10,000× | 99.83% | 96.31% | 99.98% | 100% | 100% |
| | 20,000× | 99.74% | 99.35% | 100% | 100% | 100% |
| | 30,000× | 99.79% | 99.59% | 100% | 100% | 100% |
| | 40,000× | 99.82% | 99.72% | 100% | 100% | 100% |
| 90 ng | 10,000× | 99.65% | 99.16% | 100% | 100% | 100% |
| | 20,000× | 99.48% | 99.92% | 100% | 100% | 100% |
| | 30,000× | 99.58% | 99.96% | 100% | 100% | 100% |
| | 40,000× | 99.63% | 99.97% | 100% | 100% | 100% |

The listed values for analytical sensitivity and specificity are simulated calculations that assume 25% conversion efficiency and include UMI error correction. Analytical specificity assumes a fixed cutoff of two supporting fragments to call a variant. Shaded cells include values from 0–75% and 75–95%.

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



illumina®