Liquid biopsies provide a noninvasive method to acquire tumor DNA from cell-free DNA (cfDNA). Access to this genetic information through next-generation sequencing (NGS) identifies mutations and alterations that may play a role in cancer and other diseases. The key to identifying rare mutations is improved sequencing accuracy and the ability to distinguish between biological and PCR duplicates. SMARTer® ThruPLEX® Tag-seq was developed with unique molecular tags (UMTs) to improve sequencing accuracy by accounting for polymerase and sequencing errors and to increase confidence in rare allele identification. Libraries were prepared with SMARTer ThruPLEX Tag-seq using 10–30 ng of Horizon Discovery’s Multiplex I cfDNA Reference Standard Set containing six single nucleotide variants (SNVs) for four different genes (EGFR, KRAS, NRAS, PIK3CA) present at 0.5–5% allele frequency. The libraries were enriched with either a 110-ko or 240-ko custom panel or the Agilent ClearSeq CPCR Comprehensive Cancer Panel. Enriched libraries were sequenced with an average total read coverage of approximately 5,000X and analyzed with and without the UMTs. Deduplication without molecular tags reduced coverage to 295X; whereas, deduplication with UMTs allowed for a separation of biological duplicates from PCR duplicates and increased coverage to 2,110X, a significant reduction in false positives, 73% elimination of background noise, and a 10-fold increase in unique coverage compared to deduplication without UMTs. Employing UMT consensus reads, the sensitivity to detect 70 SNVs at 1% minor allele frequency (MAF) was increased from 30% to 95% reads and at 0.2% MAF, increased from 15% to 75% and false positive calls reduced by 32X. Therefore, use of UMTs in the preparation of NGS libraries from cfDNA vastly improves sequencing accuracy by distinguishing between biological duplicates and PCR duplicates, increasing read coverage and decreasing background noise, reducing false positives, and in more confident mutation calls.

**SMARTer ThruPLEX chemistry**

SMARTer ThruPLEX Tag-seq library preparation is a simple, three-step, two-hour, one-tube reaction process with no intermediate cleanups. The repair and ligation reactions use proprietary stem-loop adaptors with added molecular tags that ligate only to the 5′ ends of the cfDNA, leaving the 3′ ends to be extended and indexed by high-fidelity amplification. (Figure 1).

**Unique molecular tags provide confident variant detection**

To establish the variant detection performance of the SMARTer ThruPLEX Tag-seq Kit, the limits of variant detection were measured using cfDNA reference standards engineered with variants at different allele frequencies. This effort involved the steps of library preparation, target enrichment, sequencing, and data analysis (Figure 2).

**Uniform distribution of unique molecular tags uniquely label fragments**

SMARTer ThruPLEX Tag-seq libraries provide >16 million unique combinations of molecular tags. More than 4,000 UMTs are available on each side of the fragment. (Figure 3).

**Variant detection using Horizon standards**

Six Horizon cfDNA reference standards were tested, ranging from 0% (wild type) to 5% MAF. All six variants were called at their expected frequencies with sensitivity and specificity over 99% (Figure 4). By combining deep sequencing with the SMARTer ThruPLEX Tag-seq Kit, it was possible to detect mutations present at 0.5% allele frequency using a starting input of just 10 ng of DNA. Lower detection limits can be achieved, depending on sample quality, input amount, capture efficiency, sequencing depth, and data processing algorithms.

**Reduction of background noise**

REDUCE background errors for more confident identification of variants. Panel A. Aligned raw reads show that the expected EGFR T790M mutation (gold dot) is obscured by false positive-deep sequencing, making it difficult to distinguish the true mutation from false positives. Panel B. In contrast, by using UMTs in the same sample, the consensus reads show a dramatic reduction in the level of background errors, and as a result, clearer separation of the true mutation from the noise. (Figure 5).

**UMTs improve signal-to-noise ratio**

For each sample and processing method, the signal was calculated by taking the average of the allele frequency detected at six expected mutation positions, and the noise was calculated by averaging the allele frequency detected outside the six positions across the entire captured region. The results show a 3X to 6X improvement in signal-to-noise ratio when UMTs were utilized for error-correction during data processing.

**Variant detection using cell-free DNA spike-in samples**

Baylor Miraca Genetics laboratory prepared human cfDNA spike-in samples by combining plasma-derived cfDNA from two distinct donors. Libraries were prepared with SMARTer ThruPLEX Tag-seq using 30 ng of cfDNA isolated from plasma, enriched with a 250-ko custom panel and sequenced to 30,000X coverage. The data was analyzed using the Curio Genomics software, and 70 unique variants were identified.

**Conclusions**

Equipped with more than 16 million UMTs, the SMARTer ThruPLEX Tag-seq library preparation kit is a powerful tool for confident detection of low-frequency variants. SMARTer ThruPLEX Tag-seq technology’s highly efficient chemistry and single-tube workflow work together to preserve molecular complexity, allowing researchers to discover more from precious samples using just 1 to 50 ng of DNA. The kit provides the freedom to use any commercially available capture panels or to design custom capture panels to interrogate genomic regions of interest that span hundreds of genes. These panels also allow structural studies of the genomic regions. Lower detection limits can be achieved, depending on sample quality and input amount, capture efficiency, and sequencing depth.