NGS libraries from cell-free DNA containing molecular tags prepared with ThruPLEX® technology improve ability to detect rare alleles

Suvarna Gandlur[1*], Anthony Popkie[1], Emmanuel Kamberov[1], Matthew Carroll[1], Dmitry Goryunov[1], Konstantinos Charizanis[1], Fang Sun[1], Maria Dinkelmann[2], Edward Jan[1], Kamran Shazand[2], Shawn Quinn[3], John Langmore[1]

[1]Takara Bio USA, Inc., Mountain View, CA, 94043, USA • [2]Rubicon Genomics, Inc., Ann Arbor, MI, 48108, USA • [3]Curio Genomics, Inc. Dexter, MI, 48130, USA • *Corresponding author: suvarna_gandlur@takarabio.com

Abstract

Liquid biopsies provide a non-invasive method to acquire the genetic information provided in 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. ThruPLEX Tag-seq was developed with the addition of 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 ThruPLEX Tag-seq using 10–30 ng of Horizon Discovery's Multiplex I cfDNA Reference Standard Set containing six single nucleotide variants (SNV) for 4 different genes (EGFR, KRAS, NRAS, PIK3CA) present at 0.5–5% allele frequency. The libraries were enriched with either a 110 kb or 240 kb custom panel or the Agilent ClearSeq 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 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% MAF was increased from 30% to 95% reads and at 0.2% MAF, increased from 7% to 75% and false positive calls reduced by 32X. Therefore, use of UMTs in the preparation of NGS libraries from cfDNA enhances 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.

ThruPLEX chemistry

ThruPLEX Tag-seq library preparation is a simple, 3-step, 2-hour, 1-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 indexes added by high-fidelity amplification. (Figure 1).

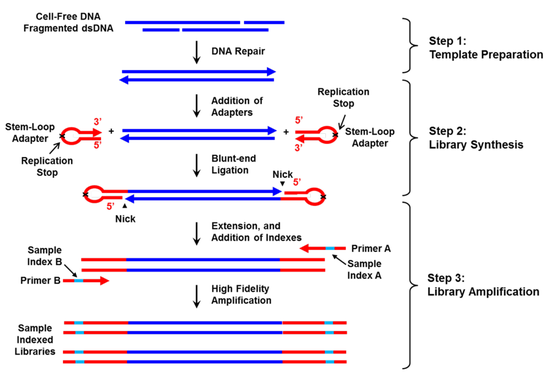

Figure 1. ThruPLEX chemistry. ThruPLEX Tag-seq starts with cell-free DNA or fragmented double-stranded DNA. In the first step, DNA fragments are repaired in a highly efficient process. Following repair, stem-loop adapters are ligated to DNA fragments to attach the Illumina® adapter sequence and a unique molecular tag (UMT) to each side of the DNA fragment. In the final reaction, indexing primers containing Illumina P5 and P7 sequences are used to complete the library structure and amplify the library fragments.

Unique molecular tags provide confident variant detection

To establish the variant detection performance of the 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).

Figure 2. Complete workflow from sample to data analysis.

Uniform distribution of unique molecular tags uniquely label fragments

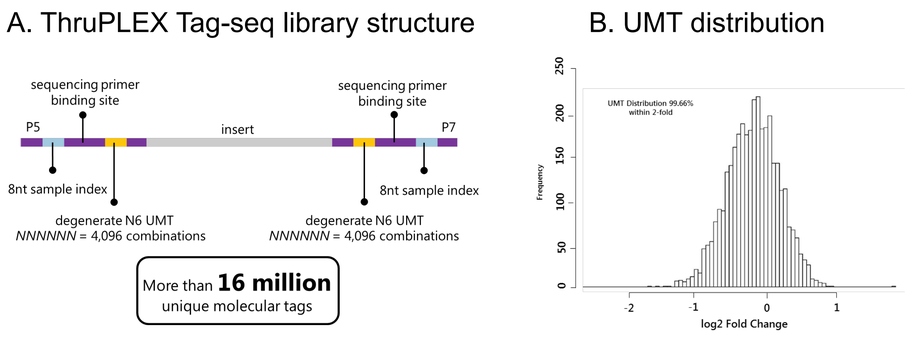

Figure 3. Incorporation of unique molecular tags. A.ThruPLEX Tag-seq provides >16 million unique combinations of molecular tags. More than 4,000 UMTs are available on each side of the fragment. **B.** After sequencing and counting of the reads, more than 99% of the UMT counts are distributed within twofold of the mean, demonstrating unbiased incorporation.

Variant detection using Horizon standards

Six Horizon cfDNA reference standards were tested, ranging from 0% (wild type) to 5% minor allele frequency (MAF). All six variants were called at their expected frequencies with sensitivity and specificity over 99% (Table 1). By combining deep sequencing with the 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.

DNA Input	Enrichment	Sequencing	Sample MAF	Detected Variant Frequency						Sensitivity	Specificity
				EGFR L858R	EGFR T790M	KRAS G12D	NRAS A59T	NRAS Q61K	PIK3CA E545K		
30 ng	110 kb panel, Agilent SureSelect	~1000X mean unique coverage, NextSeq®500	5%	3.7%	5.7%	6.1%	4.4%	5.9%	5.6%	100.0%	99.8%
			1%	0.5%	1.4%	1.5%	1.7%	1.2%	1.0%	100.0%	99.9%
			WT**	0%*	0%*	0%*	0%*	0%*	0%*		
10 ng	240 kb panel, Roche NimbleGen	~500X mean unique coverage, HiSeq®2500	2.5%	2.3%	1.0%	2.5%	3.6%	2.3%	1.6%	100.0%	99.6%
			1%	1.4%	0.6%	1.3%	0.9%	0.4%	1.7%	100.0%	99.8%
			0.5%	1.4%	0.2%	0.9%	0.8%	1.3%	1.1%	100.0%	99.8%

*Not detected
**100% wild type negative control

Table 1. UMTs provide excellent variant detection. Horizon Multiplex I cfDNA Reference Standards (Horizon HD780) were used as is or titrated using the wild-type reference standard to generate samples at additional allele frequencies. Variants were detected at their expected mean allele frequencies (MAF) with high sensitivity and specificity.

Reduction of background noise

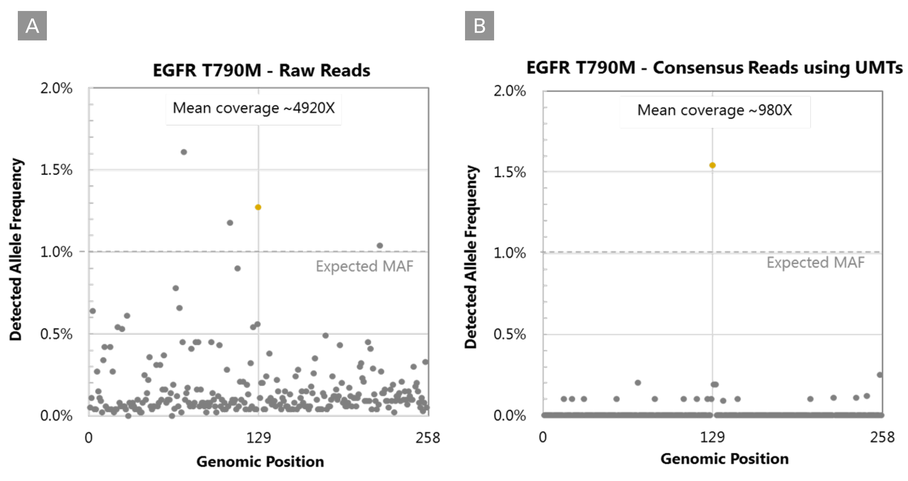

Figure 4. UMTs reduce background errors for more confident identification of variants. Panel A. Aligned raw reads show that the expected EGFR T790M mutation (gold dot) is obscured by false positive noise (grey dots), making it difficult to distinguish the true mutation from false positives. **Panel B**. 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, clear separation of the true mutation from the noise.

UMTs improve signal-to-noise ratio

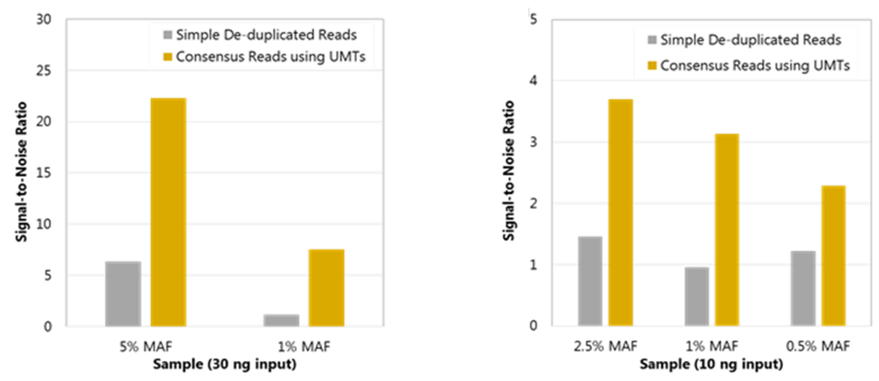

Figure 5. UMTs greatly increase signal-to-noise ratio. For each sample and processing method, the signal was calculated by taking the average of the allele frequency detected at six expected mutation positions and noise was calculated by averaging the allele frequency detected outside the six positions 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 ThruPLEX Tag-seq using 30 ng of cfDNA isolated from plasma, enriched with a 250 kb custom panel and sequenced to 30,000X coverage. The data was analyzed using the Curio Genomics software and 70 unique variants were identified.

UMTs increase unique coverage and sensitivity while reducing false positives

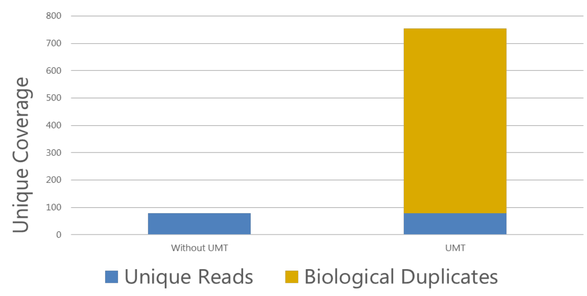

Figure 6. UMTs increase unique coverage. ThruPLEX Tag-seq libraries with UMTs show a 10-fold increase in unique coverage by utilizing biological duplicates (yellow) compared to libraries analyzed without UMTs.

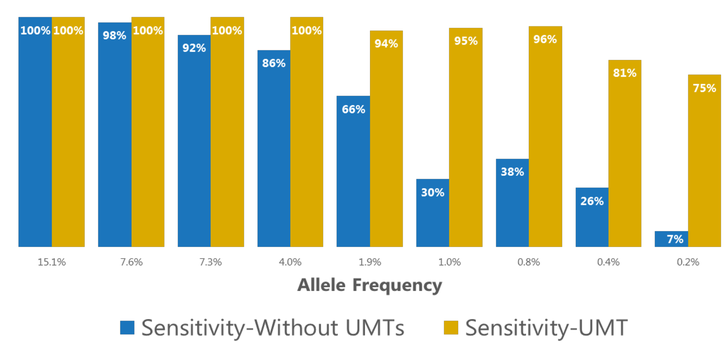

Figure 7. UMTs increase sensitivity. ThruPLEX Tag-seq libraries with UMTs show higher sensitivity when detecting low-frequency alleles (<1%) compared to libraries without UMTs.

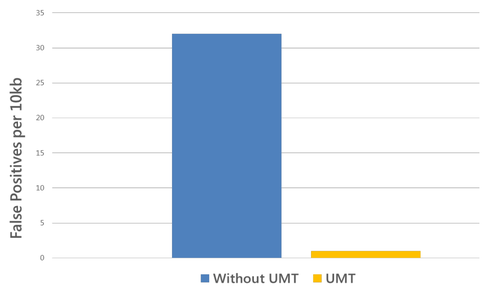

Figure 8. UMTs reduce false positives. ThruPLEX Tag-seq libraries with UMTs show higher specificity when compared to libraries without UMTs.

Conclusions

Equipped with more than 16 million UMTs, the ThruPLEX Tag-seq library preparation kit is a powerful tool for confident detection of low-frequency variants. ThruPLEX Tag-seq'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.

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

Scan to download your copy of this poster or visit takarabio.com/ABRF-2017-Tag-seq

800.662.2566
Visit us at takarabio.com

Clontech TaKaRa cellartis