Twist Pan-Cancer Reference Standards V2: Enhanced Precision and Reduced Errors in ctDNA Analysis

Lydia Bonar, Patrick Cherry, Michael Bocek, Derek Cai, Shawn Gorda, Derek Murphy, and Esteban Toro

Abstract

Liquid biopsies harness cell-free DNA (cfDNA) in the bloodstream, which can detect cancer in its early stages. However, the precision required to discern between malignant and benign genetic markers, and the sensitivity needed for disease monitoring, make their development challenging. The formulation and utilization of specific reference materials are crucial for the successful development of these assays.

Twist Bioscience initially introduced the Twist Pan-Cancer cfDNA Reference Standards to serve as a cfDNA analog reference material. This material featured a broad range of ctDNA variants and closely emulated the size and distribution of cfDNA fragments. It included synthetically-printed variants like single nucleotide variants (SNVs), insertion-deletions (INDELs), and structural variants (SVs). The reference standard covers over 400 variant sites across 84 genes, including literature-curated, clinically-relevant variant sites. Despite its comprehensive features, the first version had its drawbacks. The method used to derive the wild type (WT) background from donor derived cfDNA led to the introduction of artifacts, which limited the ability to test for low VAFs (<0.1%) at certain sites.

We have developed an enhanced version of the Twist Pan-Cancer Reference Standards, which offers a stable background genotype and a reduced background error rate. Evaluating this new reference standard with the Twist Standard V2 target enrichment system and custom capture panels targeting clinically-relevant variants, there is a clear separation between measured VAF of the lowest variant-positive standard and WT background. In addition, a side-by-side analysis showed that the error rate of this standard (V2) is markedly lower than both its predecessor (V1) and other market competitors. Importantly, its accuracy aligns closely with that of native cfDNA, and we demonstrate that it is an appropriate substrate for 'limit of blank' (LoB) studies of a given variant detection assay. The updated reference materials were created at multiple variant allele frequency (VAF) dilutions, and characterized using droplet digital PCR (ddPCR) and Next-Generation Sequencing (NGS).

In summary, the Twist Pan Cancer cfDNA Reference Standards V2 emerges as a valuable asset for assay developers and clinical labs using highly multiplexed NGS-based assays and qPCR/dPCR methods. Such a reference standard paves the way for more effective liquid biopsies in clinical research.

Reference Standard Variant Content

The reference standard is composed of both reference and variant DNA sequences. Twist's proprietary high-quality silicon DNA writer allows us to print synthetic variant sequences against desired sites from the genome. Variants of three types (CNVs, indels, and fusions) were designed using three principles: firstly, we curated a list of variants from cancer variant databases (*e.g.* COSMIC) based on a literature search of solid-tumor cancers (*e.g.* breast, lung, colorectal, renal, prostate, and bladder cancers, and melanoma, as well as literature on recurrently mutated genes in human cancers (i.e. TP53, EGFR, KRAS/NRAS, etc); secondly, we added panel-wide variants to the design in exons of highly-relevant cancer genes where variants were not already included, so that users can troubleshoot diverse target enrichment panels with a single control sample; thirdly, we included driver gene fusions where exact *DNA* breakpoints were available in the literature. In total the reference standards contain 456 variant sites in 84 unique cancer-associated genes: 225 SNVs, 216 indels (167 deletions, 49 insertions, see Fig. 1), and 15 structural variants.

The DNA is printed as a distribution with an average length of 167 bp and ± 5 bp of spread. The print extensively tiles over the variant site to which the DNA is designed, providing high coverage, fragment diversity, and location diversity of the variant site relative to the termini of the DNA molecule (Fig. 2).

Synthetic variant DNA sequences are pooled and combined with wild-type background material at the desired variant allele frequency (VAF). Target enrichment of the 0% (WT) and 5% VAF are used to verify the intended VAF across the variant sites. In addition to NGS, ddPCR is performed at each VAF across a subset of sites to provide quantitative QC at all VAFs.

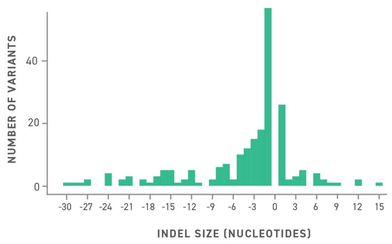


Figure 1. Distribution of indel sizes
A histogram shows the distribution of indel sizes curated for the reference standard, with positive values indicating an insertion and negative values indicating a deletion.

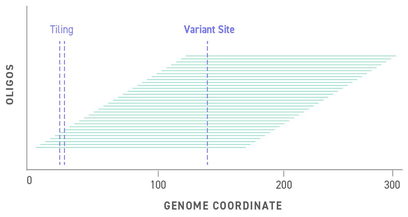


Figure 2. Variant DNA tiling
Variant DNA tiles across the genomic locus of the variant with extensive overlap. The location of the variant is marked with a dashed line.

Improved cfDNA Background

Wild type (WT) background material used in the V2 reference standard is composed of cell line DNA that is fragmented using a proprietary method to closely mimic native cfDNA, including a prominent mononucleosomal peak and a dinucleosomal peak (Fig. 3). This background represents an improvement over the donor-derived cfDNA used in the V1 reference standard as the original material went through an amplification process leading to the introduction of mutational artifacts. We evaluated the mutation rate in our new reference standard and compared it to native cfDNA, the V1 reference and a competitor's cfDNA standard (Fig. 4). The background error rate was determined across 6,135 alleles recurrently mutated in cancers, looking at both SNVs and Indels. Twist V2 reference standards had the lowest measured error rate at approximately 10 ppm and was comparable to native cfDNA. This is in contrast to competitor material and the original V1 reference standards which had mutation rates of 140 ppm and 160 ppm respectively. The low mutation rate makes these standards appropriate for extremely sensitive variant calling applications.

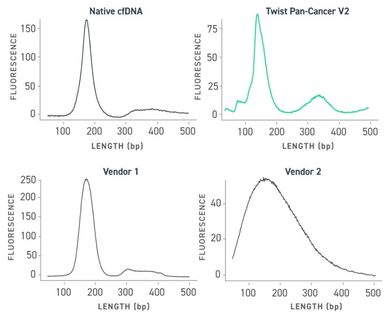
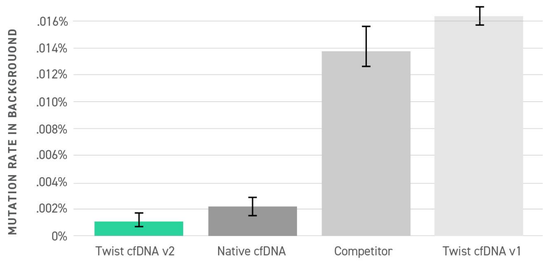


Figure 3. Fragment profile comparison of Twist's reference standards and competitors
DNA fragment analysis of the Twist cfDNA Pan-cancer Reference Standard v2, native cfDNA and competitor reference standards equalized for peak maximum height.

Figure 4. Background mutation rate comparison Analysis of background mutation rates is defined as the rate of non-germline variants detected in duplex consensus reads from a target-enriched WT sample. 30 ng of each reference standard and native cfDNA were prepared with the Twist mechanical fragmentation library prep kit and captured with the Twist Standard Hybridization v2 protocol using a custom 48kb panel. Libraries were sequenced to 80,000X raw coverage on an Illumina Nextseq 2000 P3 flow cell and error corrected using UMIs. Errors were assessed as the fraction of non-reference, non-germline-variant bases present in duplex consensus reads at 6135 recurrently mutated positions in human cancers that overlapped the capture panel.



Reducing Errors in ctDNA Analysis

The ability to accurately detect circulating tumor DNA (ctDNA) is a function of the selected assay parameters and the thresholds determined for positive variant calls. Limit of blank (LoB) is a measure of the technical noise of an assay generated by setting a threshold above the positive rate of an appropriate negative control. Limit of detection (LoD) is a measure of the lowest VAF that can be reliably distinguished from the background. Due to the improved background mutation rate, our reference standards can be used to determine these thresholds in a given NGS-based assay. To illustrate this use, standards from 0-1% VAF at variable mass inputs were taken through an example target enrichment workflow designed to target SNVs. When looking at aggregate data of 6 replicate WT and 0.1% VAF samples (Fig. 5) we observed a clear separation of observed VAF between the wild type and our lowest variant containing standard. Using a 95% confidence interval the LoD of this assay given a VAF of 0.1% was measured at 0.065%, the LoB using a 95% confidence interval was more than an order of magnitude lower at 0.0045% VAF. Per site variant calls (Fig. 6) are robust to drop-out when using higher mass input, however in 5 ng input samples, ≤0.5% VAF libraries show an increased drop-out rate and greater variability of observed VAF between sites. The observed drop outs in 0.1% VAF samples at all mass inputs have low reproducibility between technical replicates, highlighting the impact of sampling on variant detection in low VAF samples. Similarly positive calls in the WT material are mass dependent, and those calls are both rare and sporadic. For this analysis, a positive call was defined as one duplex consensus read supporting the variant (which illustrates the low number of positive calls in the wild type material even when using a permissive detection threshold). The clean background of the material enables high-confidence determination of variant detection thresholds.

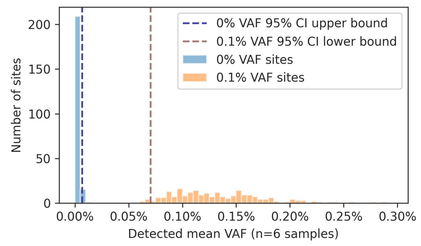


Figure 5. Histogram of observed VAF in WT and 0.1% VAF Reference Standards
A histogram of the mean measured VAF at 225 SNVs in 6 replicate 0% (WT) and 6 replicate 0.1% VAF reference standard samples. Dashed lines represent the 95% confidence interval of the 0% samples (blue) and the 0.1% VAF samples (red). 30 ng of each reference standard were used to generate each library.

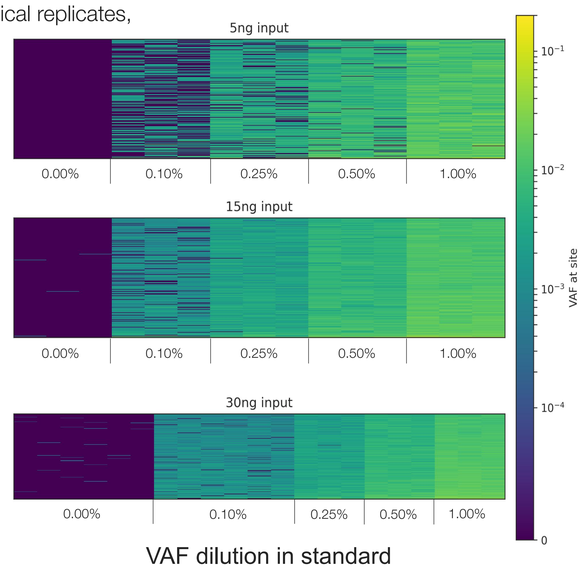


Figure 6. Observed VAF across SNVs using variable mass input
Heat map shows the observed VAF of 225 SNVs (rows) present in the reference standard and covered by the custom capture panel used for this experiment. Libraries of each reference standard VAF were prepared with the Twist cfDNA library preparation kit at variable mass inputs (5, 15 and 30 ng) and captured using the Twist cfDNA target enrichment protocol. All libraries were sequenced on an Illumina Nextseq 2000 P3 flow cell and downsampled to 80,000x raw coverage before UMI duplex collapse. A positive call was determined as one duplex consensus read supporting the mutant allele. Replicates are arranged by VAF dilution in the standard and observed VAF at each site is represented by color on a log scale.

Reference Standards Utility in Assay Development

Developing a robust liquid biopsy assay requires careful optimization of experimental conditions that can impact assay sensitivity and precision. Using high quality reference standards with known VAFs across diverse variants of interest can assist in benchmarking assay performance. Target enrichment based assays can be impacted by factors like sample quality, library conversion efficiency, panel content, and sequencing depth. One significant limitation in liquid biopsy assays is the abundance of cfDNA found in plasma or other body fluids. Here we modulated reference standard mass input at various VAFs into a target enrichment workflow optimized for use with low mass cfDNA samples (Fig. 7A) to interrogate the impact of cfDNA mass input on sensitivity. Mass input had little effect on recall in high VAF samples (≥1%), but had a significant positive impact on recall at lower VAFs, demonstrating how variable cfDNA input affects recall rate. Furthermore, the relationship between VAF level and recall rates appears non-linear, emphasizing the importance of including multiple VAF level samples in benchmarking tests.

In a second experiment, two library preparation and target enrichment workflows were tested head-to-head on a wide range of VAF levels (Fig. 7B). The Twist cfDNA library preparation workflow showed significantly higher recall at low VAFs (≤0.25%) compared to a legacy workflow, highlighting the impact of assay conditions—including library preparation and capture reagents—on performance.

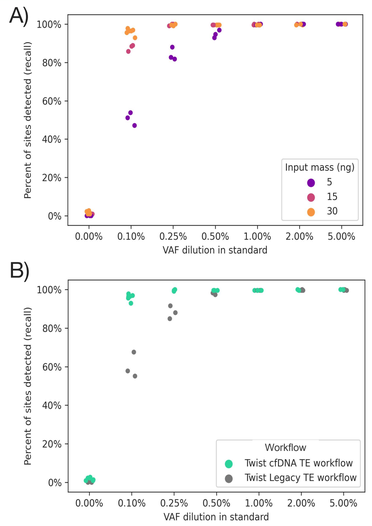


Figure 7 Recall rate across unique assay conditions
A) Recall rate for libraries of each reference standard VAF prepared with the Twist cfDNA library preparation kit at variable mass inputs and captured with a custom panel using the Twist cfDNA target enrichment protocol. B) Recall rate for libraries of each reference standard VAF prepared with 30 ng mass input using two target enrichment workflows. One workflow pairs the new Twist cfDNA library preparation kit with the Twist target enrichment protocol optimized for cfDNA. The second legacy workflow pairs the Twist Mechanical Fragmentation Library Preparation kit with the Twist Standard Hybridization v2 protocol. For panels A and B, a positive call was determined as one duplex consensus read supporting the mutant allele. Calls were made across 225 SNVs covered by the capture panel utilized. All libraries were sequenced on an Illumina Nextseq 2000 P3 flow cell and downsampled to 80,000x raw coverage before UMI duplex collapse.

Conclusions

Twist cfDNA Pan-Cancer Reference Standards v2 are composed of an expansive number of variants—spanning 84 cancer associated genes—and an improved background material that closely mimics the size profile of native cfDNA while maintaining a low background mutation rate. This clean background enables the high-confidence determination of an assay's analytical limit of blank, and facilitates systematic testing of mass input on sensitivity. The standards are available at multiple VAF levels and can be used to benchmark performance in NGS-based assays during analytical development and validation.

Conflict of Interest Statement

All authors are employees and shareholders of Twist Bioscience. Twist products mentioned are intended for Research Use Only.

Twist Bioscience and the Twist logo are trademarks of Twist Bioscience Corporation.