Twist pan-cancer synthetic reference materials for cell-free DNA (cfDNA) assay development

Patrick Cherry, Alonzo Lee, Michael Bocek, Jinfeng Shen, Derek Murphy, and Esteban Toro

1. Abstract

Liquid biopsies hold great promise in reducing the burden of the disease of cancer on patients. Early detection of cancer via cell-free DNA represents the broadest vision and requires high specificity. Ongoing disease monitoring is another important application of liquid biopsy and requires high sensitivity. Assay development requires analytical validation, and assays must be monitored for ongoing performance. These assays demand reference materials at specific variant allele frequencies for the assessment of specificity and sensitivity. Previously, obtaining reference materials was a laborious process of finding or engineering mutant cell lines. Here, we describe the design, manufacture, and quality control of a synthetic reference standard: the Twist Pan-cancer Reference Standards, developed using Twist's proprietary DNA printing platform to include a wide selection of both common and rare cancer targets for analytical validation. The reference material closely mimics the size distribution and content of cell-free DNA, with a primary peak at 167 bp and a secondary peak at 334 bp. The background cfDNA is derived from a single donor and highly characterized through NGS. The reference standard includes over 400 variant sites across 84 genes, including literature-curated, clinically-relevant variant sites. Additional panel-wide variants were included to aid in troubleshooting capture panels. The variant DNA is synthetically printed at 167 bp ± 5 bp and tiles over the site with extensive overlap, providing diversity of DNA termini relative to the position of the site of variation. Variant sites were printed and quantified independently so they can be pooled uniformly. NGS quality controls are implemented at multiple steps to ensure quality. The final products are available in a variant allele frequency (VAF) dilution series. The dilution series is quality controlled using droplet digital PCR (ddPCR) to ensure precision of variant allele frequencies. The Twist Pan-cancer Reference Standards provide a valuable solution for NGS-based assay developers seeking reference materials for a wide array of cancer variants, advancing adoption of liquid biopsy tests toward routine clinical practice.

2. Variant Content Design

The reference standard is composed of both reference and variant DNA sequence. We print the variant sequences synthetically, and so we can design them against desired sites in the genome. Variants of three types (CNVs, indels, and fusions) were designed in a stepwise manner using three principles: firstly, we curated a list of variants based on a literature search of solid-tumor cancers (*e.g.* breast, lung, colorectal, renal, prostate, and bladder cancers, and melanoma) and cancer variant databases (*e.g.* COSMIC); 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 target enrichment diverse panels with a single control sample; thirdly and lastly, we included driver gene fusions where *DNA* breakpoints were available in the literature. The total of this design is 458 variant sites in 84 unique cancer-associated genes: 228 SNVs, 215 indels (168 deletions, 47 insertions, see Fig. 1), and 15 structural variants.

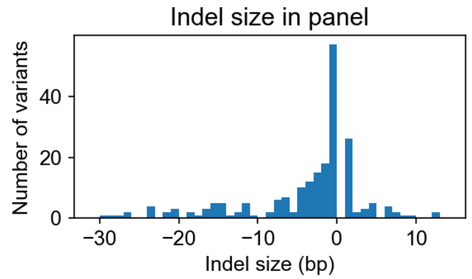

Fig. 1: Distribution of indel sizes in design
A histogram shows the counts of insertion/deletions in the design of the Twist Pan-cancer Reference Standard, with positive values indicating an insertion, and negative values indicating a deletion. SNVs (single-nucleotide variants) are not depicted in this plot.

3. Variant DNA

We print the variant DNA synthetically on Twist's proprietary high-quality silicon DNA writer. The DNA is printed at 167 bp in length and 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. 2A). All of the DNA fragments contain the variant site.

Additionally, we print the variant DNA such that the mode fragment length is 167 bp, and the fragment length diversity is further increased with a distribution of lengths that extend as far as ± 5 bp around 167 bp, meaning the variant DNA fragments range from 162 bp to 172 bp in length (Fig. 2B), with a plurality at 167 bp, the canonical cfDNA fragment length.

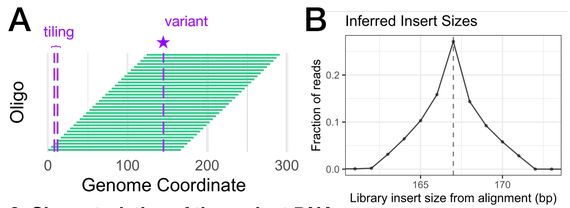

Fig. 2: Characteristics of the variant DNA.
A: Schematic of the tiling of the variant DNA. Variant DNA tiles across the genomic locus of the variant with extensive overlap. The location of the variant is marked in a purple dashed line and star. **B**: Variant DNA length distribution. A plot of of each fragment length in an Illumina library prepared from a pure pool of the variant DNA. The distribution shows a mode at 167 bp and tails that extend to ± 5 bp surrounding the mode, from 162 bp to 172 bp.

Financial Disclosures: All authors are employees and shareholders of Twist Bioscience

4. Background cfDNA

The background DNA that makes up the standard is derived from native cell-free DNA. The sample of cfDNA used to make the standard is derived from a single healthy donor and has been screened to be compatible with the set of variants included in the reference standard. Sequencing results of the background DNA are available to customers upon request. The background DNA of the standard has a size distribution that closely mimics patient-isolated cell-free DNA, including a mononucleosome peak at 167 bp and a dinucleosome peak at 334 bp. Sometimes, additional nucleosomal peaks can be visualized (Fig. 3A).

We also characterize the background DNA of the reference standard by non-targeted NGS (Fig. 3B). Upon alignment to the human genome (hg38), the inferred insert sizes show striking similarity between the Twist Pan-cancer Reference Standard product and the native cfDNA sample, each displaying a sharp mononucleosomal peak at 167 bp (dashed line) with "notches" in the trace leading up to that size from smaller sizes.

The insert size plots also resolve a dinucleosomal peak in both samples, zoomed in to in the inset. The Pan-cancer Reference Standard closely mimics the fragment size distribution of native cell-free DNA.

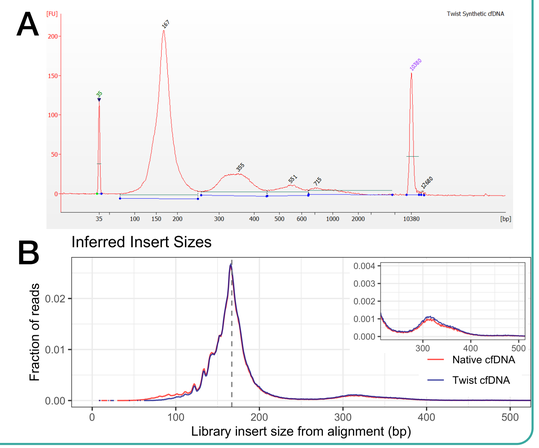

Fig 3: Characteristics of the background cfDNA
A: Size distribution of the background cfDNA. A Bioanalyzer High Sensitivity chip output shows the fragment profile of the Twist Pan-cancer Reference Standard background cfDNA. The electropherogram shows a primary mononucleosomal peak at 167 bp, a secondary dinucleosomal peak at 355 bp, and additional periodic peaks beyond those lengths. The HS chip result does not show any signal above baseline above 2 kb other than the marker. **B**: Inferred insert sizes of Twist and native cfDNA. Samples were Illumina library prepared and sequenced on a NextSeq550 instrument. Reads were aligned to the reference human genome, and library insert lengths were plotted. A dashed at 167 bp is provided for reference. Native cfDNA is in red and Twist background cfDNA is in blue. An inset in the upper-right corner zooms in on the dinucleosomal peak from 200 bp to 500 bp.

5. Quality Control

During manufacturing of the Twist Pan-cancer Reference Standard, the inputs, intermediates, and final products are extensively quality controlled. The complexity of the reference standard requires us to QC using multiple approaches. To verify that we attained the intended variant allele frequency at many sites, we use target-enrichment of the 0% (wild-type) and 5% VAF samples followed by high-coverage NGS. Sequencing information generated is then k-mer searched for the variant and reference sequence, and a distribution with a mean can be visualized (Fig. 4A). However, because NGS is at best a semi-quantitative method, we also run droplet digital PCR (ddPCR) to ascertain the allele frequency at select sites among the product's variants using wet-lab validated TaqMan primer/probe sets from the manufacturer. All primer/probe sets were optimized for annealing temperature and verified to produce droplets with sufficient fluorescent signal separation. Assays using ddPCR show highly linear dilution series across all VAFs of the Reference Standard, from 5% to 0.1%, with r-squared values 0.986 or greater, and show slopes of the regression line close to 1, varying from 0.75 to 1.09. Noise/signal variability increases in the lower VAF levels, as expected in any mutation detection experiment. The 0% VAF (wild-type) samples show very few positive droplets, indicating the background DNA is an effective negative control compared to the other VAF samples in the series, which have variant DNA spiked in.

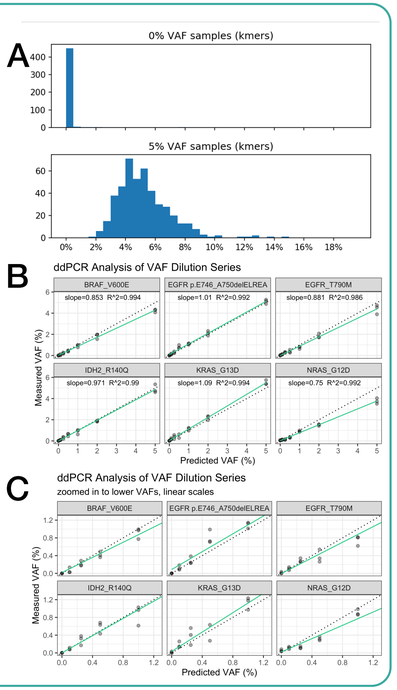

Fig 4: Quality Control Results for the Pan-cancer Reference Standard
A: Distribution of variant allele frequencies in 0% and 5% levels. Histograms for the 0% VAF (wild-type) and 5% VAF show the TE-NGS result of each included variant in the product. Samples were library prepped, target-enriched, sequenced on a NextSeq550 instrument, and k-mer counted for the variant and reference sequences for each site in the product design. 5% VAF sample has a mean VAF of 5.25%. **B**: ddPCR Analysis of VAF dilution series. Material from each VAF level of the product was used in ddPCR mutation detection assay for each assay indicated in the facet labels in triplicate. Sample and wet-lab validated primer/probe were mixed, divided into droplets on the Bio-Rad QX-200 automated droplet generator, thermocycled according to the manufacturer's instructions, read on the QX-200 automated droplet reader, and analyzed in the Quantasoft package. Predicted VAF on the *x*-axis is the intended VAF, as seen on the product label; Measured VAF on the *y*-axis is the result of ddPCR. The dotted line shows the line of equality $y = x$, is provided for reference. The green solid line is the regression line for the samples in that assay. The slope and r-squared values of that assay's regression is printed at the top of the facet. **C**: ddPCR Analysis of VAF Dilution Series, zoomed to low VAFs. The same data in B is displayed in C, but zoomed into the region 0% to 1% VAF in order to better show the lower VAF levels of the product. Axes are linear.

6. Conclusions

The Twist Pan-cancer Reference Standard is composed of a highly multiplexed set of useful variants synthetically printed and available in independent, individual VAF levels. The reference standard is also composed of background cfDNA that closely mimics the size profile of native cfDNA, and is derived from a healthy donor. The final product is highly QC'd using TE-NGS and ddPCR. The Twist Pan-cancer Reference Standard provides a valuable solution for NGS-based assay developers seeking reference materials for a wide array of cancer variants, advancing adoption of liquid biopsy tests toward routine clinical practice.