Use of synthetic CNV fragments to mimic copy number alterations for ctDNA reference standards

Jason Corwin, Patrick Cherry, Shawn Gorda, Michael Bocek, Jean Challacombe, Derek Murphy, Esteban Toro

Abstract

Diagnostic assays of liquid biopsies hold great promise for cancer detection and monitoring. To aid the advancement of this technology, Twist Bioscience previously released the Pan-cancer Reference Standard, a cell-free DNA (cfDNA) mimic reference material with varying allelic fractions of synthetically-printed single nucleotide variations (SNVs), insertion-deletions (INDELs), and structural variants (SVs). However, many cancers are characterized by changes in the copy number of genes that support growth and survival of cancer tissue—dubbed copy number variants (CNVs). To aid assay development for detection of CNVs, we designed, synthesized, and characterized an additional control supplemented into the Pan-cancer Reference Standard: a CNV for the exonic regions of the ERBB2 gene (also known as HER2). ERBB2 is often amplified and overexpressed in invasive carcinomas, including breast, ovarian, stomach, bladder, salivary, and lung cancers, making it a desirable variant to include in reference materials.

Similar to the design of the synthetic variants in the Twist Pan-cancer Reference Standard, the ERBB2 CNV DNA is developed to closely mimic the size and content of native ctDNA while providing a wide selection of common and rare cancer targets for analytical validation. It mimics the length distribution of cfDNA, with a peak at 167 bp ± 5 bp for diversity. The synthetic CNV DNA tiles over the exonic intervals for uniform gene coverage. Variant fragments were printed, quantified, and uniformly pooled followed by spiking into DNA derived from a single, highly NGS-characterized donor as background. The final admixture of ERBB2 is ddPCR-verified for copy number amplification.

The Twist Pan Cancer Reference Standard continues to provide a valuable highly multiplexed solution to assay developers seeking reference materials for a wide array of cancer variants and is now further improved with the addition of copy number variants.

CNV Design

One major concern when working with synthetic standard references is the ability of that reference to mimic the behavior of cfDNA fragments from *in vivo* samples. To address this issue, the existing Twist Pan-Cancer Reference standard is designed to maintain a narrow size distribution of DNA fragments around 167 bp, which mimics native cfDNA from blood plasma (Udomruk, 2021). Similarly, we applied the same conditions and tiling strategy to the development of a novel CNV Spike-in as the oligos in the Pan-Cancer Reference Standard. However, the design was tiled with much deeper coverage of the ERBB2 CDS derived from the HG38 genomic reference.

To test the ability of Twist DNA printing technology to generate a novel CNV standard, we targeted the sequence of ERBB2 due to its role in a large variety of cancers and as a target of directed therapies. We utilized the coding sequence of the ERBB2 gene model listed in the MANE gene model from ENSEMBL to design a cfDNA CNV reference standard. The entire 3,965 bp of the coding sequence was tiled using 167 bp oligos with a total of 1,097 probes (**Figure 1**). To ensure even and deep coverage of all bases within the exon, the design extends some coverage over inter-exonic bases adjacent to the coding region. After standard filtering and post processing, a total of 3,906 bp of the coding region is covered within the design with each base covered at a depth of over 120 unique oligos (**Figure 2**).

Oligos were printed using Twist's proprietary silicon chip writer resulting in a pool of oligos with an average length of 167 bp +/- 5 bp. These pools of short oligos were diluted and spiked into the Twist Pan-Cancer 5% VAF Reference Standard at a concentration of 15X genomic equivalents of ERBB2 for testing and quality control.

Figure 1: CNV Oligos cover all exons of target ERBB2. Whole gene coverage for all coding sequences of the CNV allows for compatibility with multiple assay formats, including both amplicon based and target enrichment chemistries. All 29 exons of the MANE gene model were covered with some intergenic regions as a buffer.

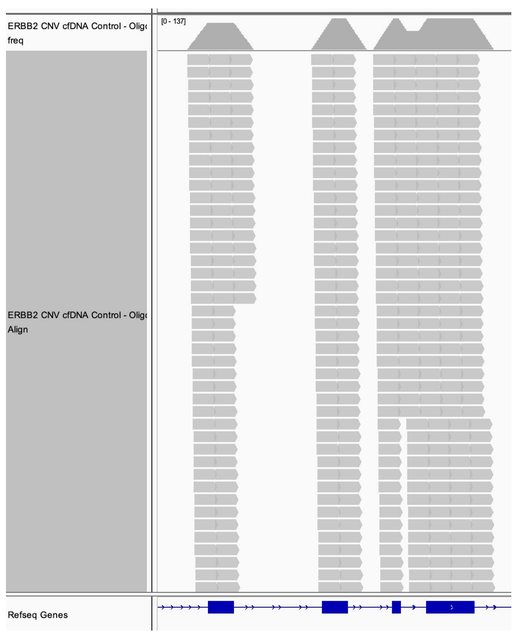
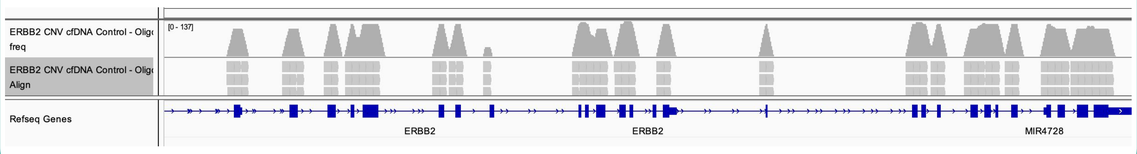


Figure 2: CNV oligo depth covers all exons of target ERBB2. An extended IGV view of the exons 6, 7, 8, and 9 illustrating the depth of oligo coverage for each base within the exon. Each base in the coding region of ERBB2 is covered by an average of 120 oligos.

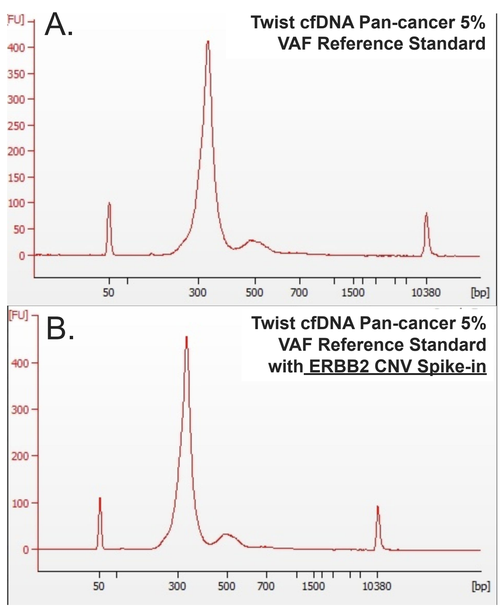


Uniformity in CNV design

When adding such a high concentration of a single gene with new oligos into an existing panel, it is vital to ensure that these fragments behave similarly under standard molecular protocols. While fragment size and design are similar to the Twist Pan-cancer Reference Standard, there is a possibility of aberrant behavior due to the relatively high concentration of a single locus.

To test if the ERBB2 CNV Spike-in chemically reacted similarly to synthetic cfDNA fragments found in the Twist Pan-cancer Reference Standard, we constructed Illumina short-read DNA sequencing libraries with and without the ERBB2 CNV Spike-in. BioA traces of the sequencing libraries with a 5% VAF Reference Standard in conjunction with Twist UMI Adapters show no aberrant spikes or unrelated peaks (**Figure 3**). Thus, the CNV reference standard mimics natural cfDNA from patients and can be added directly into established and developing cancer detection protocols.

Figure 3: Size distribution of Twist Pan-Cancer Reference Standard with and without ERBB2 CNV Spike-in. Size of cfDNA libraries using Twist Pan-cancer Reference Standard without (**A**) and with (**B**) the ERBB2 CNV Spike-in. Libraries were constructed using Twist UMI Adapters, which are 77 bp in length each. Total average length of cfDNA, 5% VAF Pan-Cancer reference, and the novel ERBB2 CNV Spike-in with adapters is 321 bp.



References

Udomruk S, Orrapin S, Pruksakorn D, Chaiyawat P. Size distribution of cell-free DNA in oncology. Crit Rev Oncol Hematol. 2021 Oct;166:103455. doi: 10.1016/j.critrevonc.2021.103455. Epub 2021 Aug 28. PMID: 34464717.

Conflict of interest statement:
All authors are employees and shareholders of Twist Bioscience

Increased CNV Raw Abundance

Successful detection of CNV variants relies on accurate and sensitive detection of gene copies above background relative to the number of genome equivalents. A well-accepted method of CNV detection is the use of quantitative PCR, including digital droplet PCR (ddPCR), using a single copy reference gene to normalize copy number.

We tested our ERBB2 CNV reference standard against the Twist Pan-cancer Reference Standard with 0% and 5% VAF using BioRad's ddPCR Copy Number Assay for ERBB2. The assay uses a small section of ERBB2 exon 21 and EIF2C1 as a reference for genomic copy number. Amplification of ERBB2 exon 21 showed approximately the same concentration of ERBB2 fragments with just over 1 genomic equivalent in both the 0% and 5% Pan-Cancer Reference Standard. Comparatively, two separate dilutions of the ERBB2 CNV Reference standard showed over an 8-fold enrichment of genomic equivalents relative to background (**Figure 4**). This illustrates that the CNV design can be successfully used to standardize accurate CNV detection.

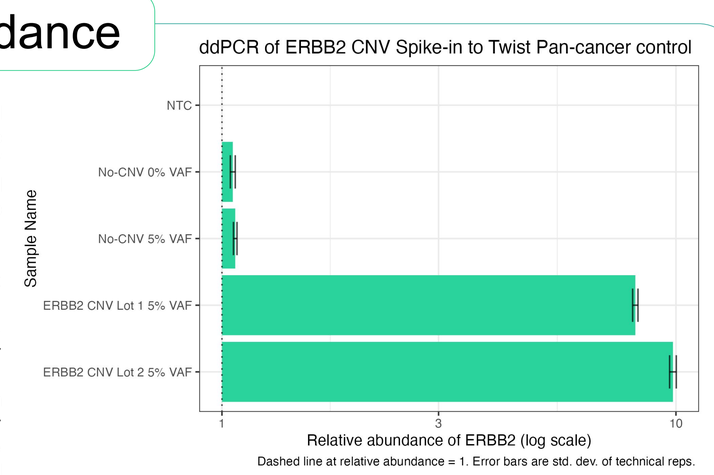


Figure 4: Relative abundance of ERBB2 detected by ddPCR. Bar graph illustrating the increase in the relative abundance of ERBB2 relative to a non-template control. Samples spiked with the ERBB2 control consistently show a relative abundance of 8-10 fold increases relative to our native 0% VAF and 5% VAF Pan-cancer Standard.

Coverage of ERBB2 CNV

While qPCR detection assays are often considered the gold standard for detection of individual CNV loci, many researchers are developing sequencing-based assays (NGS) to simultaneously detect multiple cancer-diagnostic loci, including SNPs, INDELs, and CNVs. To test our ERBB2 CNV reference standard in a defined NGS assay, we sequenced three libraries of the NGS control and two libraries of the ERBB2 CNV Spike-in constructed with the Twist UMI Adapter System.

The ERBB2 CNV Standard Reference showed superior sequencing depth without sacrificing coverage. Both the native Twist cfDNA Pan-cancer 5% VAF Reference Standard samples and samples with the ERBB2 CNV spike-in show full coverage of all exons within the ERBB2 coding sequence. However, all exons in samples with the ERBB2 CNV Spike-in showed 3-5X increased depth of sequencing over the native Twist cfDNA Pan-cancer 5% VAF Reference Standard (**Figure 5**). Additionally, a frequency histogram regarding depth of coverage for all targeted exons within the Twist Exome 2.0 panel shows a slight increase in the representation of the 24 ERBB2 exons, with deep coverage (> 300 reads) observed relative to all other targeted exons (**Figure 6**). These results indicate that the Twist CNV Standard reference can serve as a predictable reference for development and detection of diagnostic CNVs.

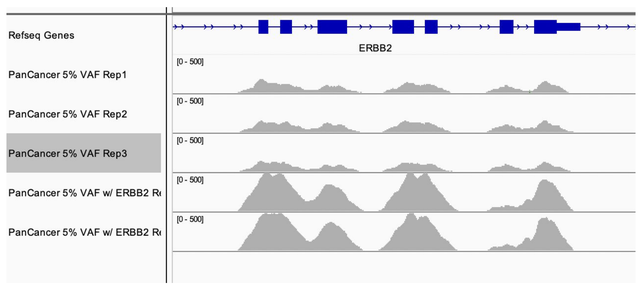


Figure 5: Increased depth of coverage of ERBB2 with CNV Spike-in. IGV trace illustrating the increased depth of coverage for ERBB2 when CNV is spiked into the Twist PanCancer 5% VAF Standard Reference.

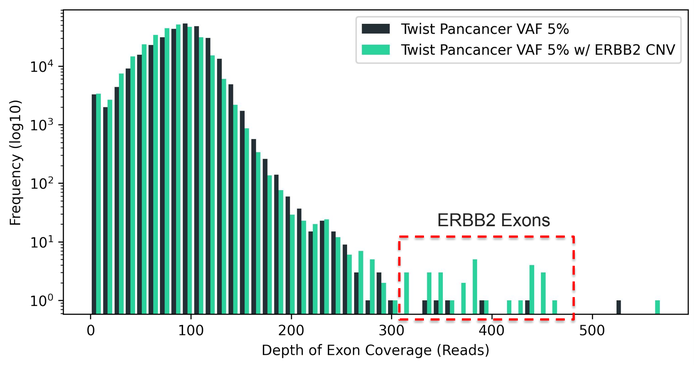
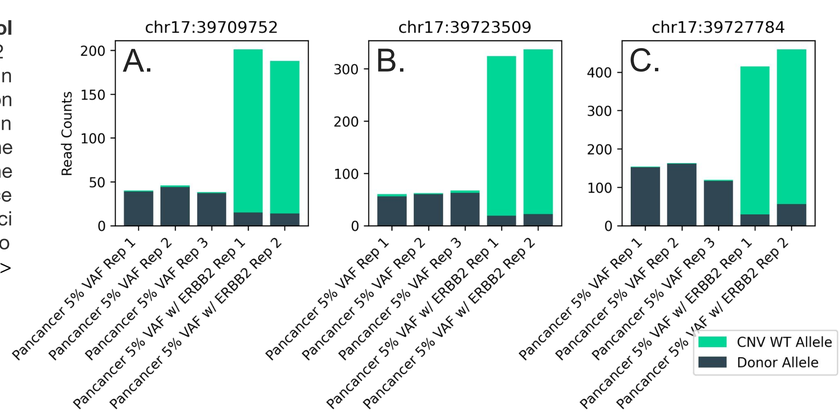


Figure 6: Histogram illustrating the increased depth of coverage of ERBB2 with CNV Spike-in. A dodged histogram showing the frequency (log axis) of depth of coverage between a sample negative for the ERBB2 spike and a sample positive for the ERBB2 CNV Spike-in.

Proportion of ERBB2 CNV Relative to Background

As described above in **Figures 4 and 5,** both ddPCR and NGS TE assays detect the presence of the synthetic ERBB2 cfDNA spiked in to the 5% VAF Pan-Cancer Control. Despite this positive result, the relative abundance of the amplification event observed in the assay results is lower than the targeted 15-fold amplification level. These results highlight the importance of assay selection and optimization for the detection of genomic variants as neither assay was optimized for the quantitative detection of the ERBB2 copy number event. In the case of the ddPCR assay, an off-the shelf, validated probe-set targeting ERBB2 copy number alterations was selected for analysis. It was discovered post-analysis that the reverse primer of the ddPCR amplicon lies within a region of reduced coverage depth in an intergenic region between exon 21 and 22 of ERBB2. This sub-optimal probe design results in a deflated estimate of the relative abundance of the contrived copy number event. However, we can use the proportion of background ERBB2 SNPs from the donor relative to the number of wild type (WT) alleles from the synthetic CNV spike-in to more accurately assess the quality of enrichment over the background within the NGS data. Using three SNPs from the donor background (chr17:39709752, chr17:39723509, and chr17:39727784), we found a dramatic increase in the number of NGS reads covering that SNP (**Figure 7**). Relative to stock Twist Pan-Cancer Reference Standard samples, samples containing the synthetic ERBB2 CNV Spike-in showed a 3.85x increase in enrichment of ERBB2. However, the number of raw reads containing the donor SNPs at these positions in the ERBB2 CNV spiked-in samples are much lower than that of the stock Twist Pan-Cancer controls. This is likely due to the saturation of ERBB2 probes contained within the standard protocol for Twist Exome 2.0 panel used for target enrichment. When looking within the ERBB2 CNV Spike-in samples, we see an average enrichment of 12.6x reads covering the ERBB2 CNV allele over the background donor SNP. Thus, a more accurate estimate of CNV enrichment can be achieved when normalizing to the number of background reads from the donor.

Figure 7: Proportion of ERBB2 CNV Control relative to background. Spiked ERBB2 increases the total count of ERBB2 reads within NGS assay as well as increases the proportion of WT reads. Shown are the relative proportion of read counts that contain either the background SNP from a cfDNA donor or the HG38 WT reference in the CNV reference standard for three loci within ERBB2. These loci include: chr17:39709752 (C > T; CNV WT to Background Donor SNP), chr17:39723509 (G > A), and chr17:39727784 (C > G)



Materials & Methods

A CNV reference standard of ERBB2 was designed using proprietary Twist design approaches to tile the exonic regions of the ERBB2 gene with reference sequences. The design was printed on the Twist silicon DNA writer, processed through the Twist Pan-cancer cfDNA production process, and pooled to equimolarity of oligos. The pooled ERBB2 reference standard DNA was quantified and spiked into the Twist Pan-cancer Reference Standard VAF 5% (SKU: 104569). Two distinct pools of the ERBB2 copy number variant were produced with a goal of a 15x amplification over the background ERBB2 copies. One sample of the same 5% VAF material with no ERBB2 spike-in was included.

This material was analyzed using two distinct assays (ddPCR and NGS) to test for the abundance, coverage, and quality of the ERBB2 CNV reference standard. Three technical replicates of the native Twist Pan-cancer Reference Standard and ERBB2 Spike-in were used to determine total copy number amplification and sequence enrichment using a BioRad ddPCR™ Copy Number Variation Assay and the EIF2C1 locus as an internal reference. Additionally, these controls and test material were also assayed via target enrichment using the Twist Standard Capture v2 system (SKU: 105560) with the Twist Exome 2.0 TE panel (SKU: 104132). Sample concentrations were quantified and QC'd via Bioanalyzer HS, pooled, and sequenced on two NextSeq 550 High Output runs with 5% phiX spike-in. Sequence alignments were run using BWA mem resulting in a mean coverage of 102x.