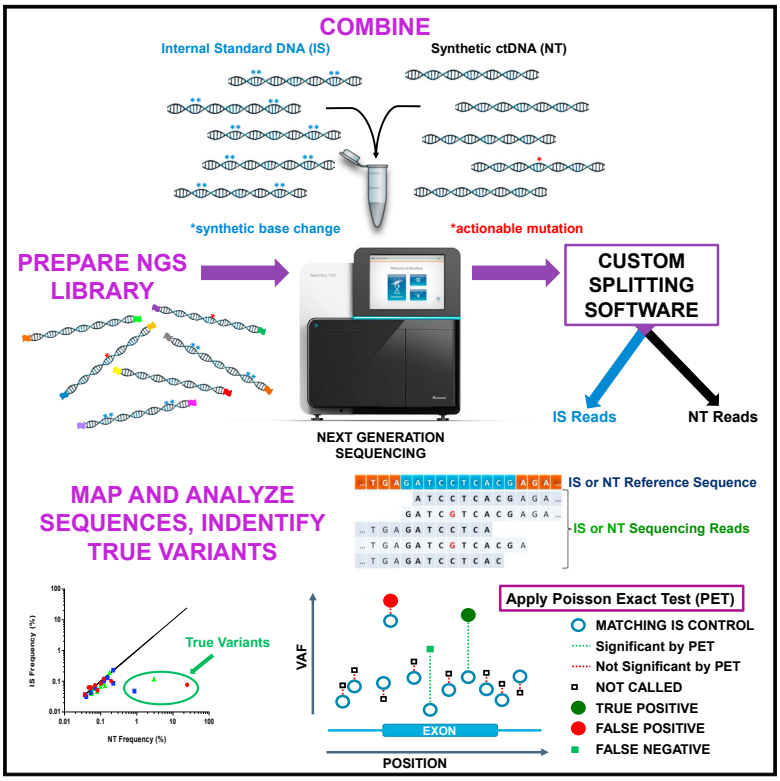
Cell Reports
Methods

Article

Advancing NGS quality control to enable measurement of actionable mutations in circulating tumor DNA

Graphical abstract



Highlights

- Synthetic spike-in IS control for technical errors in NGS
- IS enable measurement of true-positive mutations not detected by current practices
- IS included in hybrid capture NGS libraries do not interfere with existing workflows
- IS may be used as an orthogonal quality control for UMI or non-UMI library analysis

Authors

James C. Willey, Tom B. Morrison, Bradley Austermiller, ..., Timothy R. Mercer, Leihong Wu, Joshua Xu

Correspondence

james.willey2@utoledo.edu (J.C.W.), tmorrison@accugenomics.com (T.B.M.), joshua.xu@fda.hhs.gov (J.X.)

In brief

Willey et al. spiked synthetic internal standards (IS) into contrived circulating tumor DNA samples to control for technical error in NGS. IS enabled calculation of technical error rate and limit of detection for each actionable mutation in each sample, increasing the number of measurable true positives without loss of specificity.

Willey et al., 2021, Cell Reports Methods *1*, 100106
November 22, 2021
https://doi.org/10.1016/j.crmeth.2021.100106

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Advancing NGS quality control to enable measurement of actionable mutations in circulating tumor DNA

James C. Willey,[1,13,*] Tom B. Morrison,[2,*] Bradley Austermiller,[2] Erin L. Crawford,[1] Daniel J. Craig,[1] Thomas M. Blomquist,[1] Wendell D. Jones,[3] Aminah Wali,[3] Jennifer S. Lococo,[4] Nathan Haseley,[4] Todd A. Richmond,[5] Natalia Novoradovskaya,[6] Rebecca Kusko,[7] Guangchun Chen,[8] Quan-Zhen Li,[8] Donald J. Johann, Jr.,[9] Ira W. Deveson,[10,11] Timothy R. Mercer,[10,11] Leihong Wu,[12] and Joshua Xu[12,*]

[1]College of Medicine and Life Sciences, University of Toledo, Toledo, OH 43614, USA
[2]AccuGenomics Inc., The Atrium, Suite 105, 1410 Commonwealth Drive, Wilmington, NC 28403, USA
[3]Q[2] Solutions, EA Genomics, Morrisville, NC 27560, USA
[4]Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA
[5]Roche Sequencing Solutions Inc., Pleasanton, CA 94588, USA
[6]Agilent Technologies, La Jolla, CA 92037, USA
[7]Immuneering Corporation, Cambridge, MA 02142, USA
[8]University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
[9]Winthrop P Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, 4301 W Markham Street, Little Rock, AR 72205, USA
[10]Garvan Institute of Medical Research, Sydney, NSW 2010, Australia
[11]St. Vincent's Clinical School, University of New South Wales, Sydney, NSW 2010, Australia
[12]Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA
[13]Lead contact
*Correspondence: james.willey2@utoledo.edu (J.C.W.), tmorrison@accugenomics.com (T.B.M.), joshua.xu@fda.hhs.gov (J.X.)
https://doi.org/10.1016/j.crmeth.2021.100106

MOTIVATION Despite an increasing demand for precision medicine enabled by NGS measurement of actionable mutations in circulating tumor DNA (ctDNA) specimens, the ability to reliably measure and report low-frequency mutations using current NGS practices is limited. Challenges include low- or poor-quality specimens and technical errors that vary among samples and mutation sites. Here, we designed synthetic internal standards (IS) and methods for their use to better control for technical error in NGS in assessment of ctDNA specimens. The goal was to determine whether this would improve quality control, resulting in increased clinical sensitivity without loss of specificity.

SUMMARY

The primary objective of the FDA-led Sequencing and Quality Control Phase 2 (SEQC2) project is to develop standard analysis protocols and quality control metrics for use in DNA testing to enhance scientific research and precision medicine. This study reports a targeted next-generation sequencing (NGS) method that will enable more accurate detection of actionable mutations in circulating tumor DNA (ctDNA) clinical specimens. To accomplish this, a synthetic internal standard spike-in was designed for each actionable mutation target, suitable for use in NGS following hybrid capture enrichment and unique molecular index (UMI) or non-UMI library preparation. When mixed with contrived ctDNA reference samples, internal standards enabled calculation of technical error rate, limit of blank, and limit of detection for each variant at each nucleotide position in each sample. True-positive mutations with variant allele fraction too low for detection by current practice were detected with this method, thereby increasing sensitivity.

INTRODUCTION

Measurement of circulating tumor DNA (ctDNA) in blood samples with recently developed next-generation sequencing (NGS) provides a non-invasive "liquid biopsy" to detect and genetically target cancer as well as to monitor response to treatment (Beaubier et al., 2019; Chen et al., 2019; Schwaederle et al., 2016; Rodon et al., 2019; Zutter et al., 2015; Siravegna et al.,

2017; Vasan et al., 2019; Volckmar et al., 2018; Corcoran and Chabner, 2018; Oxnard et al., 2016). The US Food and Drug Administration (FDA) recently approved certain ctDNA assays for diagnostic applications. However, the approved assays are limited by current common practice. For example, a limit of detection (LOD) is estimated *a priori* for each type of actionable mutation, and then a constant predefined LOD value is applied to every sample. For a commercial platform recently approved by the FDA, the predefined variant allele fraction (VAF) LOD for measurement of an actionable BRAF mutation was 1.1% or 0.2% for DNA input of 5 ng or 30 ng, respectively (FDA, 2020). However, there is both variant-specific and inter-sample variation in LOD (Craig et al., 2019; Blomquist et al., 2015). Thus, application of a predefined VAF LOD to analysis of every actionable mutation in every sample is limiting if the LOD could be set lower in many ctDNA samples. Furthermore, even "conservative" predetermined LOD thresholds may not ensure reproducible results due to insufficient measurement of, and control for, sample-specific and site-specific technical error (Craig et al., 2019; Blomquist et al., 2015) and/or stochastic sampling associated with small and/or poor-quality ctDNA samples (Corcoran and Chabner, 2018; Blomquist et al., 2015; Fu et al., 2014; Merker et al., 2018; Kuderer et al., 2016; Stetson et al., 2019; Rossi and Ignatiadis, 2019; Torga and Pienta, 2018). In order to obtain maximum reliable information from limited ctDNA samples, there is a need for methods that provide sample- and variant-specific LOD, confidence limits for each test result, and inter-laboratory concordance (Blomquist et al., 2013, 2015; Kuderer et al., 2016; Squillace et al., 2015; Gargis et al., 2015). The method described here was designed to enable sample- and variant-specific LOD while ensuring reliable reporting of NGS VAF data. This study was conducted as a collaborative inter-laboratory effort through the FDA-led Sequencing Quality Control Consortium Phase 2 (SEQC2) project (FDA SEQC2, 2019).

Most false-positive variants are caused by NGS technical errors due to regional and site-specific chemical and physical factors during library preparation and sequencing (Craig et al., 2019; Blomquist et al., 2015). Multiple approaches have been developed to address this (Newman et al., 2014, 2016; Cibulskis et al., 2013; Kennedy et al., 2013; Schmitt et al., 2012; Sandmann et al., 2017; Davis et al., 2021; Craig et al., 2016; Ma and Zhang, 2019; Abelson et al., 2020; Salk et al., 2018; Cheng et al., 2015). One approach is to capture recurrent technical artifacts by sequencing many normal samples and then apply a variant-specific LOD value established *a priori* based on the rate of technical error that gives rise to each respective variant (Cibulskis et al., 2013). Refinements of this approach with reported increasing accuracy continue to be developed (Cheng et al., 2015; Davis et al., 2021; Ma and Zhang, 2019; Abelson et al., 2020). This approach typically requires sequencing data from more than 40 normal samples and is better suited to highly automated reference laboratories with sufficient testing bandwidth to regularly confirm that the technical error has not altered due to small drifts in reagents, instruments, and operators. Furthermore, this approach does not control for inter-measurement variation in sample-specific inhibitors, instrument performance, or operator proficiency. Another approach is to filter out sequencing artifacts by attachment of a unique molecular identifier (UMI) to each DNA molecule during library preparation, followed by bioinformatic identification of consensus among PCR replicates (Schmitt et al., 2012; Newman et al., 2014, 2016; Kennedy et al., 2013). This method facilitates identification and removal of technical errors that are inconsistent with the consensus, prior to variant calling. However, UMI methods may introduce new non-systematic errors, unrelated to chemical and physical technical errors, that must be identified and filtered empirically (Sandmann et al., 2017). It is reasonable to hypothesize that the combination of known chemical-physical sources and potential non-systematic bioinformatic sources of technical error define the limit of the blank (LOB) for measurement of each variant. Furthermore, it is reasonable to hypothesize that the primary determinants of LOD and confidence limits in each measurement of a variant are (1) variant-specific LOB and (2) the number of ctDNA sample variant copies captured in the library preparation.

We previously demonstrated that Standardized Nucleic Acid Quantification (SNAQ-SEQ) internal standard (IS) controls improved accuracy of variant calling with amplicon-based sequencing (Craig et al., 2019; Blomquist et al., 2015). Specifically, systematic chemical-physical technical errors were closely modeled by synthetic IS spike-in molecules added to each sample (Craig et al., 2019; Blomquist et al., 2015). IS controls were known to be the reference genome sequence except for the intended IS-specific variants. Thus, any new variant detected in IS reads was caused by technical errors accrued in the NGS testing process. Importantly, the technical error fraction for each detected IS variant closely reflected the error fraction for that same variant in the native template (NT) sequence of the sample tested. As such, addition of IS to a sample enabled calculation of within-sample variant-specific technical error rate at each genomic position covered by the IS and provided a basis for establishment of variant-specific LOD in that sample.

Targeted NGS libraries prepared with UMI following hybrid enrichment are the predominant NGS method used by the majority of oncopanels developed for ctDNA analysis (Deveson, 2021). There is an unment need for methods that provide an LOD for hybridge capture-targeted NGS measurement for each actionable mutation in each ctDNA sample. To address this need, we re-designed the SNAQ-SEQ IS so that they would be suitable for use in NGS following hybrid capture enrichment and library preparation with or without UMI incorporation.

The purpose of this study was to (1) evaluate whether synthetic spike-in IS control for technical errors associated with hybrid capture, (2) to assess utility of synthetic spike-in IS as an orthogonal quality control for UMI library analysis, and (3) to determine whether a mixture of IS with each sample interferes with existing hybrid capture workflow or causes additional errors. It was necessary to develop bioinformatic methods that efficiently separate the re-designed spike-in IS from sample NT sequence reads in FASTQ files prior to pipeline analysis and variant calling. Moreover, this approach required development of a statistical approach for variant calling based on comparative analysis of sequencing reads from each sample DNA NT and respective IS. To test the clinical utility of methods developed through these studies, we also prepared contrived ctDNA reference samples (Deveson, 2021) with ground truth true-positive (TP) variant sites and ground truth true-negative (TN) reference sites.

RESULTS

A ground truth dataset was established in SEQC2 reference *Sample A*

A ground truth variant dataset was identified for SEQC2-contrived tumor reference *Sample A* (a pool of ten diverse cancer cell lines) in a related SEQC2 project study (Deveson, 2021; Jones et al., 2021). For this study, it was important to establish a more comprehensive ground truth variant dataset for *Sample A* within the targeted regions spanned by the SNAQ-SEQ IS mixture. For this purpose, the Roche SeqCap EZ Choice custom PHC panel was used for hybrid capture enrichment followed by non-UMI library preparation. A 7.5 kb region common to the SNAQ-SEQ IS mixture and the Roche SeqCap EZ Choice custom PHC panel also spanned the 6.8 kb Illumina TST170 × SNAQ-SEQ IS consensus region (described in STAR Methods). In brief, as described in STAR Methods, ground truth TP variants were established by sequencing *Sample A* and samples from a dilution series, in which *Sample A* was diluted with a normal sample (SEQC2 reference *Sample B*). Each *Sample A*/*B* dilution sample was mixed with SNAQ-SEQ IS, enzymatically fragmented, subjected to non-UMI library preparation, then sequenced. As presented in Figure 1A, following sequencing, a FASTQ file was bioinformatically separated into IS and NT bins through alignment to respective IS-specific and NT-specific Hg19 reference genomes. Precision for separation of NT from IS reads was >99.9997% (i.e., <0.0003% IS base changes present in NT reads). Next, the respective IS FASTQ and NT FASTQ files were mapped with BWA MEM and positionally deduplicated with Picard Mark Duplicates. As an example of coverage covariation for IS and NT DNA fragments during enrichment, the IS and NT sequences mapped to *EGFR* gene are presented in Figure 1B. The histogram demonstrates that, although capture efficiency and subsequent sequence coverage varied considerably for each hybrid capture probe region, the IS (identified by presence of synthetic dinucleotide variants used to separate IS from NT) and NT were captured proportionately across all probes in the *EGFR* panel with peak capture efficiency in exonic regions.

TP were identified as variants that responded appropriately to titration in the *Sample A*/*B* serial dilution samples (Figure 2). Through this analysis, the set of TP in the 7.5 kb region consensus to the SNAQ-SEQ IS mixture and the Roche panel was expanded to 28. In the 6.8 kb region consensus to the Illumina TST170 panel and the SNAQ-SEQ IS mixture, the TP were increased to 20, an increase from 8 reported in a related SEQC2 study (Deveson, 2021). Furthermore, 3,500 positions in this consensus region had >1,000 coverage in the ground truth data, sufficient to establish them as TN. As such, any suspicious variant call in a test sample at one of these 3,500 positions could be confirmed as false-positive (FP) if there was not a TP variant at the same position in the ground truth data (i.e., demonstrated lack of titration response).

SNAQ-SEQ QC method development in Non-UMI libraries from *Sample A* undiluted with *Sample B*

Non-UMI libraries from *Sample A* (undiluted with *Sample B*) following Roche SeqCap EZ Choice custom PHC panel enrichment were used for development of the SNAQ-SEQ method. As detailed in the STAR Methods, SEQC2 Reference *Sample A* was mixed with SNAQ-SEQ IS mixture then subjected to hybrid capture target enrichment with the Roche SeqCap EZ Choice custom PHC panel, non-UMI library preparation, and sequencing.

VarDict (Lai et al., 2016) was used to call each IS and NT variant (or alternative [alt]) allele using the IS and NT reference sequences, respectively, then the pileup option was used to create IS and NT genome (g)VCF files. SEQC2 *Sample A* libraries yielded 16,310 NT alt bases at 8,571 positions within exons spanned by IS (other than IS dinucleotide sites). Of these *Sample A* NT alts, 5,608 had the same alt in the IS sequence, and this allowed bivariate plot of NT VAF (y axis) and IS VAF (x axis) (Figure 1C). Among the 16,310 NT alt bases, VarDict called 374 alts (i.e., PASS FILTER). Other than the expected 28 TP variants, the remaining VarDict PASS variants (n = 346) arose among the 3,500 TN positions and, by definition, were FP arising from technical errors.

SNAQ-SEQ QC used both the VarDict IS gVCF and NT gVCF files to determine the significance of each variant in the NT gVCF file. Possion exact test (PET) (see STAR Methods) was used to analyze the relative VAF values for each NT alt and corresponding IS alt to establish LOB and LOD for measurement of the NT alt, as schematically illustrated in Figure 1. Specifically, for each NT alt allele called by VarDict, statistical significance was determined by PET analysis of the number of alt observations (i.e., alt count and position coverage) for the respective NT and IS sequences. PET was an appropriate statistical test for this application because Poisson distribution best represents both (1) the low number of NT alts commonly observed in small and/or degraded DNA samples, such as ctDNA and cytologic specimens, and (2) the low number of NT and IS alts resulting from technical error (Craig et al., 2019). Frequencies of the 374 VarDict PASS NT alts, binned according to PET score, are presented in Figure 1D. VarDict NT PASS alts had PET scores less than zero (left side of plot), and VarDict IS PASS alts had PET scores greater than zero (right side of plot). Alts with PET score significantly different from NGS background error appear as green bars, and those not distinguishable from NGS background error appear as red bars. Because each IS was synthesized and sequence verified, any IS alts, including those with significant PET scores (green bars) relative to NT background, were technical errors (FP). Therefore, the highest PET score observed for an IS technical error variant was used to determine PET score significance threshold for NT variants.

The 28 ground truth TP variants were called by VarDict, confirmed by SNAQ-SEQ QC with PET analysis (green columns with negative PET score in Figure 1D), and appear as green circles in Figure 1C. In contrast, 346 alts called by VarDict (VarDict FP) were correctly classified as TN by SNAQ-SEQ QC (red columns in Figure 1D and red circles in Figure 1C). The remaining TN alts were correctly not called by VarDict or SNAQ-SEQ QC analysis and appear as gray circles in Figure 1C. As is evident in Figure 1C, genome positions with matched alts in NT and IS and NT alt confirmed as TN by SNAQ-SEQ QC (gray and red circles) clustered around the 1:1 correlation line. Furthermore, variation in NT and IS VAF among these technical error variants ranged over three logs (base10) consistent with amplicon library

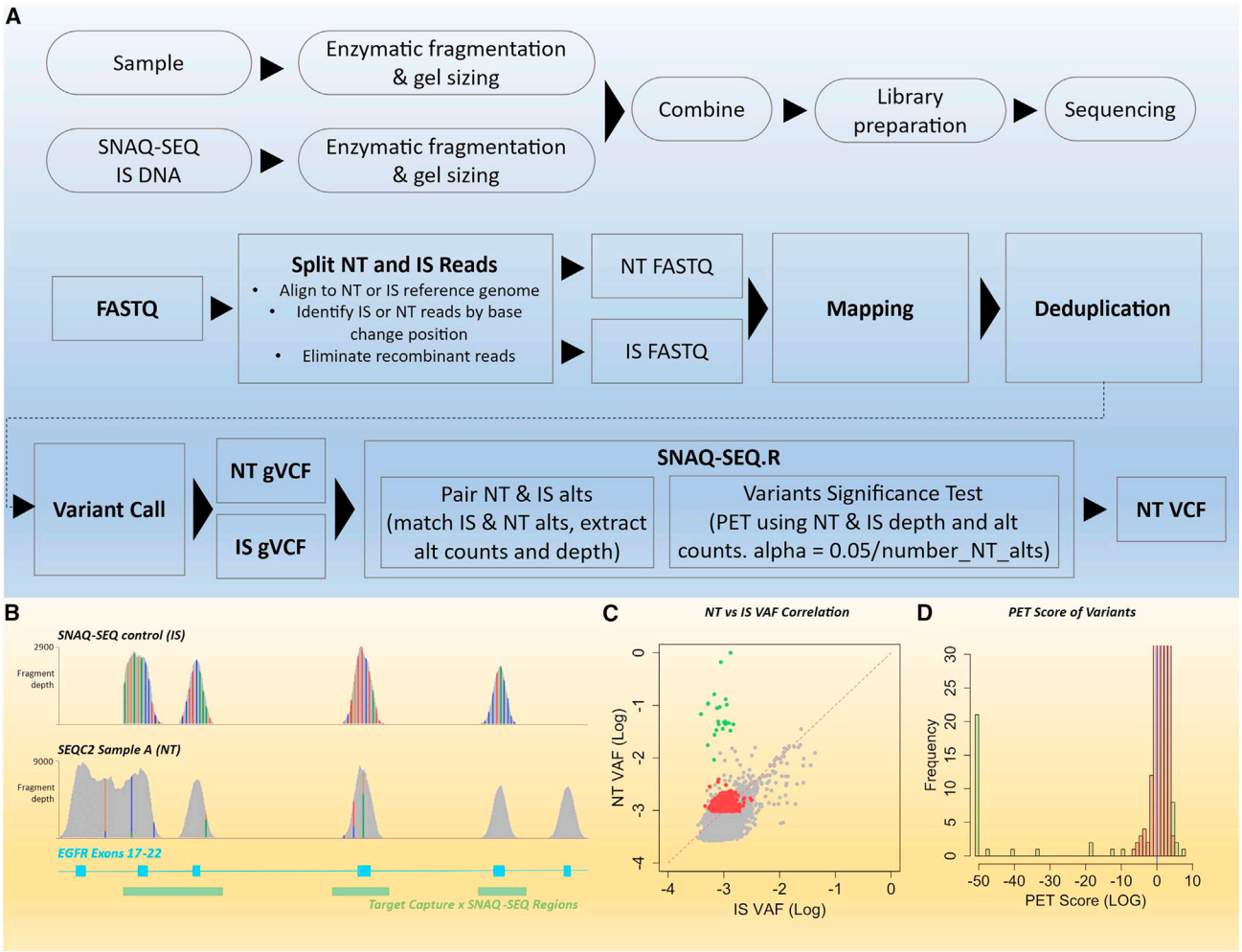


Figure 1. SNAQ-SEQ method

(A) Integration of SNAQ-SEQ into the NGS pipeline (see details in STAR Methods).

(B) Analysis of *EGFR* region in SEQC2 reference *Sample A* mixed with SNAQ-SEQ IS followed by enrichment on the Roche SeqCap EZ Choice custom PHC panel. Histogram of *EGFR* region (x axis) versus fragment depth (coverage) (y axis) for IS (top) and native template (NT) (bottom) sequence reads. Colored columns indicate bases alternate (alt) to hg19 reference with >5% VAF. IS alts were engineered dinucleotides; NT alts were SNPs derived from one or more cancer cell lines in *Sample A*. The three sequence regions spanned by IS (green bars) overlap four of the six depicted *EGFR* exons (blue bars) bounded by the Roche panel target regions.

(C) The 5,608 *Sample A* NT alts with matched alts in the IS sequence were plotted: NT VAF (y axis) and IS VAF (x axis). NT alts called by VarDict are indicated as significant (green filled) or not significant (red filled) by SNAQ-SEQ QC PET analysis, and alts that were not called by VarDict or SNAQ-SEQ QC PET analysis are indicated as gray circles. Red dashed line is 1:1 correlation in VAF value.

(D) Frequency (y axis) of alts binned according to PET score (x axis) for the 374 NT alts called by VarDict (left side of plot, scores < 0) and 895 IS alts (right side of plot, scores > 0), with significant alts as green columns and non-significant as red columns.

data in previous reports (Blomquist et al., 2015; Craig et al., 2019). This high correlation between SNAQ-SEQ IS sequences and corresponding human DNA sample NT sequences at TN sites enabled use of IS as an internal negative control to estimate NGS background.

In analysis of *Sample A*, the single-nucleotide variant (SNV) LOD (i.e., SNV detected in three of four replicates) for the Roche library non-UMI library preparation with SNAQ-SEQ QC was between 0.5% and 1% (data not shown), limited by the high IS NGS background evident in Figure 1C. This is an improvement on the LOD of 2% observed for Roche non-UMI analysis of the same variants without SNAQ-SEQ QC in a parallel SEQC2 study (Deveson, 2021).

Demonstration of SNAQ-SEQ QC utility: SNAQ-SEQ QC improved variant calling in SEQC2-contrived reference ctDNA

To evaluate the utility of SNAQ-SEQ QC to improve NGS analysis of ctDNA samples, we used the Illumina TST170 enrichment platform, followed by Illumina UMI library preparation and variant calling. This UMI dataset represented one of the best commercially available error-corrected NGS methods and therefore

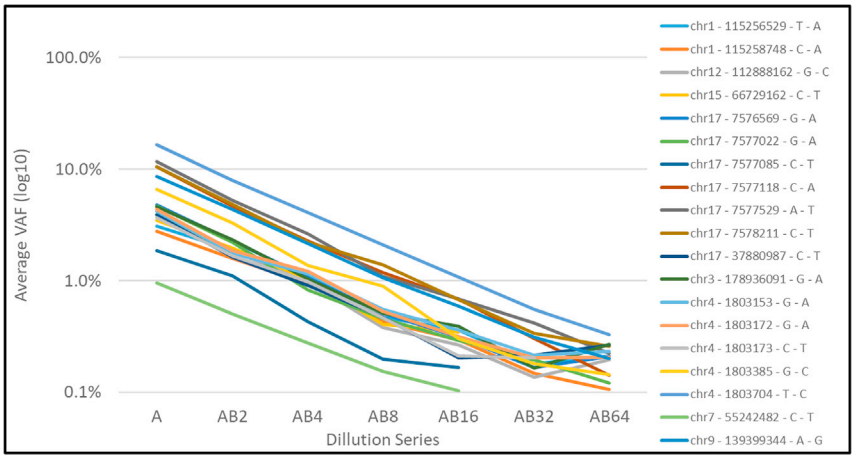


Figure 2. Identification of *Sample A* ground truth TP variants

Ground truth TP variants were identified within the consensus region for the Roche SeqCap EZ Choice custom PHC, panel TST170 panel, and SNAQ-SEQ IS panel, based on serial dilution of *Sample A* with *Sample B* from 2- to 64-fold (AB2 through AB64). Plotted are the average replicate VAF values for 19 variants that demonstrated a dilution response, with information for each variant in the legend. One additional *Sample A* TP variant restricted to the consensus region between TST170 and SNAQ-SEQ IS (i.e., not in Roche panel) was identified through serial dilution (data not shown), resulting in a total of 20 ground truth TP variants (see Table 2).

was a suitable challenge for SNAQ-SEQ to demonstrate additional QC utility. The SEQC2-contrived ctDNA reference samples used in this study closely resemble those used in a related SEQC2 study (Deveson, 2021) (STAR Methods). *Sample A* was a mixture of ten cancer cell lines. *Samples D* and *E* were created through 5- or 25-fold dilution of *Sample A* DNA with normal cell line *Sample B* DNA. Therefore, the set of variants comprised of *Samples D* and *E* represent the intersection of *Samples A* and *B* but at different variant allele frequencies (see details in STAR Methods) (Deveson, 2021; Jones et al., 2021). As described in the related manuscript, SEQC2 *Samples B*, *D*, and *E* were subjected to enzymatic fragmentation followed by size selection to create *Samples Bf*, *Df*, and *Ef* (Deveson, 2021). The samples named *Df* and *Ef* in this study were derived from the same batch of material as the SEQC2 reference ctDNA samples (*ct-high* and *ct-low*, respectively) used in the related study from the SEQC2 project, and VAF values were expected to be equivalent (Deveson, 2021; Jones et al., 2021).

Samples Bf, *Df*, and *Ef* were each mixed with aliquots of the fragmented IS spike-in mixture to make *Samples BfIS*, *DfIS*, and *EfIS*, respectively (see STAR Methods). The fragment size distribution of SNAQ-SEQ IS controls closely paralleled that of the SEQC2 reference samples (Figure S1).

Based on the ground truth data, the expected VAF range in *Sample A* for the 20 TP variants covered by the TST170/SNAQ-SEQ IS overlap was 2.5%–5%, and following 5-fold dilution of *Sample A* in *Sample B* to create *Sample DfIS*, the expected VAF range was 0.5%–1%. For this study, the two *Sample B* heterozygote variants and a single homozygote variant were filtered out in analysis of *Samples DfIS* and *EfIS*. Assuming good TST170 library preparation capture efficiency, the four *Sample DfIS* replicates at 50 ng input each (14,500 haploid genomes) would provide sufficient read counts for detection of each variant in the 0.5%–1% range. In contrast, in *Sample EfIS*, these same variants would range from below 0.1%–0.5% VAF, and those at the lower end of this VAF range would challenge the detection limit.

The optimal ratio of the IS:NT read count is a balance to minimize IS flow cell reads while still providing adequate power to estimate NGS background. According to the Illumina-optimized UMI pipeline, detection of a variant requires at least three alt counts. Therefore, the goal was to achieve an IS level that would yield a 95% confidence that the NT variant was above background LOB when no IS alt was detected. For the PET analysis, this requirement was met at a depth ratio of 2.5:1 IS:NT.

Thus, the SNAQ-SEQ IS mixture was mixed with *Samples DfIS* and *EfIS* to achieve at least a 2.5:1 IS:NT depth ratio. The TST170 UMI pipeline called variants with a UMI deduplicated read count of at least three. Assuming no variant was detected in IS, a variant with three NT alts would have a significant PET value ($p < 0.05$) (i.e., there were sufficient observations of zero alts in IS to reach confidence that the signal in NT was above noise) when the IS depth was 2.5-fold more than NT. The average IS:NT ratio across each position in the targeted regions was measured in each of the sequencing files for four replicate libraries of *Samples BfIS*, *DfIS*, and *EfIS*. This average IS:NT ratio was 4.8, and only 3% of NT alts had <2.5-fold IS coverage relative to NT (Table 3). Furthermore, there was high inter-library and inter-sample correlation for the IS/NT ratio measured for each of the targeted regions (mean CV, 7.6%), as was observed for analysis of EGFR in *Sample A*, presented in Figure 1B. Inter-target variation in average IS/NT ratio ranged from 2.9- to 7.3-fold. This approximately 2.5-fold range in IS/NT ratio among the targeted regions arose from variation in concentration among the IS when the IS mixture was created.

Illumina UMI deduplication and variant calling was conducted to create VCF files for each replicate library, test for the presence of each expected TP, and assign a FILTER value for each detected TP: PASS or LowSupport. The Illumina pipeline reported only PASS variants. However, for purposes of this study, SNAQ-SEQ QC with PET analysis was applied to each alt with at least three alt counts and with either a PASS or LowSupport FILTER value, but not blacklisted. As presented for analysis of TP in *Sample A* non-UMI libraries following Roche panel hybrid capture (Figures 1C and 1D), SNAQ-SEQ QC with PET analysis was used to estimate the technical error and calculate the LOB at each nucleotide position in the IS sequence in the Illumina TST170 UMI libraries. In analysis of data from *Sample BfIS*, none of the 20 *Sample A* TP variants appeared and no other FP variants were reported (see data code and availability,

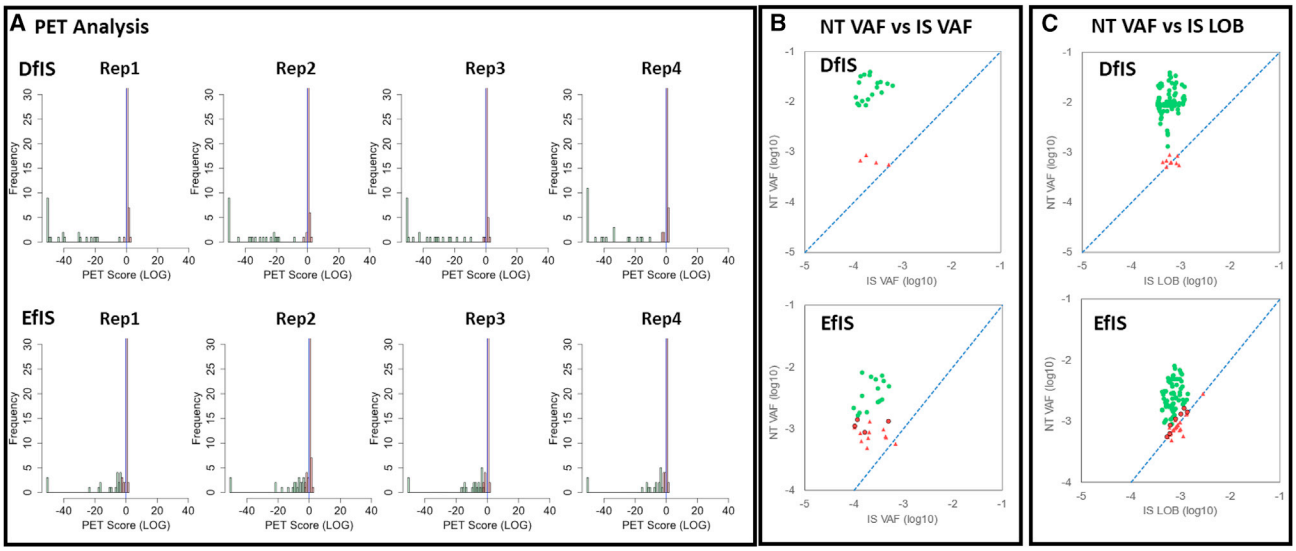


Figure 3. Effect of SNAQ-SEQ QC on ILM UMI VCF file variants from *Samples DfIS* and *EfIS*

(A) PET analysis of ILM alts for each replicate library. PET score distribution of Illumina VCF file alts (alts with PMEAN > 25 [see Figure S2] and PASS or LowSupport FILTER) for *Sample DfIS* (upper panels) and *Sample EfIS* (lower panels) replicate libraries; x axis, PET score for NT alts (scores < 0) and IS alts (scores > 0); y axis, alt frequency. SNAQ-SEQ QC PET analysis indicated alts significantly above background NGS technical error (green bars) or not significant (red bars).

(B) Bivariate analysis of paired NT and IS alt VAF values. NT alt VAF (y axis) versus matched IS alt VAF in (x axis) plotted for each NT alt. Circles, ground TP variants; green filled, PET significant; red filled, non-significant; and triangles, ILM FILTER LowSupport VCF that were not PET significant. The dashed reference line is the 1:1 correlation between NT and IS VAF values.

(C) Bivariate analysis of paired NT VAF versus IS-determined LOB values. For each alt displayed, the PET score from (A) was used to estimate the IS-determined LOB (x axis) which was plotted relative to VAF (y axis); symbols as described for (B). Dashed line is the 1:1 correlation in IS-derived LOB and corresponding NT VAF values.

STAR Methods, and key resources table). Graphs that present the distribution of PET scores for each variant in each replicate library of *Samples DfIS* and *EfIS* are depicted in Figure 3A. As is evident, in *Sample DfIS* (upper panel) the PET scores corresponding to the TP variants were highly significant (green bars) and well separated from the background alts. Importantly, due to effective UMI suppression of technical error, no PET significant IS variants were observed in any of the *Sample DfIS* or *EfIS* replicate libraries, in contrast to the presence of PET significant IS variants in the non-UMI *Sample A* libraries (Figure 1D). An alternative view of the difference between TP and NGS sequencing error is depicted in the bivariate plots, Figures 3B and 3C. In Figure 3B, the VAF for each NT alt (y axis) with a matching IS alt (x axis) was presented. For *Sample DfIS*, there were two well-separated clusters. On the one hand, the VAF for each TP variant was two to three logs higher than for the matching IS alt (green-filled circles). In contrast, the cluster associated with the 1:1 line comprises LowSupport alts that were not validated by PET analysis and therefore were not separable from but ratherconsistent with technical error (red-filled triangles). In Figure 3C, the IS VAF value was replaced with the LOB value estimated based on PET analysis. As with Figure 3B, the *Sample DfIS* TP variants (green-filled circles) cluster significantly above the LOB (red-filled triangles). In Figure 3C, some *Sample DfIS* NT alts appear on the plot that are not present on Figure 3B because, using PET analysis, a LOB can be calculated even for NT alts for which zero alts were observed in IS. In contrast to *Sample DfIS* (Figures 3B and 3C, upper panels), for *Sample EfIS* (Figures 3B and 3C, lower panels), separation between TP (circles) and NGS error (triangles) was reduced due to the 5-fold further dilution of each *Sample A* TP variant in *Sample B* (see STAR Methods). Correspondingly, the distribution of PET values for TP variants in *Sample EfIS* (Figure 3A, lower panel) was closer to that for technical error.

SNAQ-SEQ pipeline analysis improved clinical sensitivity without loss of specificity

In *Sample EfIS*, *Sample A* was 25-fold diluted with *Sample B* (5-fold greater dilution with *Sample B* than in *Sample DfIS*). Among the 4 replicate *Sample EfIS* libraries, 80 ground truth TP variant measurements were expected. However, the mean TP VAF in *Sample EfIS* value was expected to be less than 0.5% based on ground truth experiments in this study and prior analysis of *Sample Ef* (Deveson, 2021) (termed *ct-low* in Deveson). Therefore, this sample presented a stringent test for clinical sensitivity of NGS measurement platforms to measure the ground truth TP variants. The effect of SNAQ-SEQ QC with PET analysis on ILM UMI NGS analysis of ground truth TP variants in *Sample EfIS* replicate libraries is presented in Table 1. Some TP variants were not detected by ILM (i.e., no alts with either VCF PASS or LowSupport FILTER) and are represented as blank cells. The TP that were called significant by SNAQ-SEQ QC with PET analysis and detected but not called by ILM UMI NGS alone (i.e., FILTER LowSupport but not PASS) are indicated as VAF values with bold text. One *Sample EfIS* variant called by ILM (i.e., VCF PASS FILTER) was not supported by

Table 1. Effect of SNAQ-SEQ QC on ILM UMI NGS analysis of ground truth true-positive variants in *Sample EfIS* replicate libraries

CHROM	Position	Ref.	Alt	Gene	VAF 1 (%)	VAF 2 (%)	VAF 3 (%)	VAF 4 (%)
chr1	115,256,529	T	A	NRAS	0.4	0.2	0.2	0.2
chr1	115,258,748	C	A	NRAS	0.1	(0.06)	0.2	–
chr12	112,888,162	G	C	PTPN11	0.1**	0.3	0.2	–
chr15	66,729,162	C	T	MAP2K1	–	0.1**	0.4	0.2
chr17	7,576,569	G	A	TP53	(0.16)	0.3	0.2**	0.2**
chr17	7,577,022	G	A	TP53	0.3	0.2	0.2	0.1**
chr17	7,577,085	C	T	TP53	0.1	0.2	0.1	0.1
chr17	7,577,118	C	A	TP53	0.5	0.5	0.3	0.5
chr17	7,577,529	A	T	TP53	0.6	0.4	0.5	0.2
chr17	7,578,211	C	T	TP53	0.4	0.3	0.4	0.4
chr17	7,578,671	C	T	TP53	0.2	0.2	0.5	(0.11)
chr17	7,578,679	A	G	TP53	0.3	0.5	0.7	0.5
chr17	37,880,987	C	T	ERBB2	(0.09)	0.2	0.2	0.2
chr3	178,936,091	G	A	PIK3CA	0.2	0.2	0.3	0.1*
chr4	1,803,153	G	A	FGFR3	0.2	0.2	0.2	0.2
chr4	1,803,172	G	A	FGFR3	0.2**	0.2	0.2	0.2
chr4	1,803,173	C	T	FGFR3	0.3	0.1**	0.1**	0.2
chr4	1,803,385	G	C	FGFR3	0.2	0.2	0.2	0.3
chr4	1,803,704	T	C	FGFR3	0.8	0.7	0.6	0.6
chr9	139,399,344	A	G	NOTCH1	0.3	0.3	0.5	(0.13)

True-positive (TP) variant annotation—CHROM, POS, REF, ALT, GENE—indicate hg19 chromosome location, nucleotide position, reference base, alternate base, and HUGO Gene Nomenclature Committee symbol; TP variant allele fraction (VAF) (%) in replicates 1, 2, 3, and 4; cells with an – indicate variant *not detected* by Illumina pipeline (i.e., no alts with either VCF PASS or LowSupport FILTER); cells with two asterisks (**) indicates variants *detected but not called* by Illumina (i.e., VCF alts with LowSupport FILTER only) that were called by SNAQ-SEQ QC PET analysis; cells with asterisk (*) indicate an Illumina VCF alt called by Illumina (i.e., VCF PASS FILTER) but not significant by SNAQ-SEQ QC PET analysis; cells with VAF enclosed in parentheses indicate detected alts determined to be not significant either by Illumina (i.e., LowSupport FILTER only) or SNAQ-SEQ; and remaining variant cells contain VAF values for variants that were called by both methods (i.e., Illumina VCF PASS FILTER and significant by SNAQ-SEQ QC PET analysis).

SNAQ-SEQ QC analysis and is indicated as a VAF value with an asterisk (*). The remaining cells contain VAF values of variants that were called by both methods (i.e., Illumina VCF PASS FILTER and significant by SNAQ-SEQ QC PET analysis). SNAQ-SEQ QC analysis of ILM VCF file variants with LowSupport FILTER annotation did not result in any FP calls.

As indicated in the summary table (Table 2), the Illumina gVCF file used for SNAQ analysis *detected* (i.e., PASS or LowSupport FILTER) 77/80 TP in the four replicate libraries from *Sample EfIS* but only *called* (i.e., PASS filter) 64 of the 77. Thus, overall 64/80 TP were called, for 80% sensitivity. Due to the Illumina LowSupport FILTER label, the remaining 13 TP in the VCF file were not reportable. In comparison, SNAQ-SEQ QC analysis called 71 of the 80 TP (89% sensitivity), including 8 of the 13 LowSupport variants. In Table 2, sensitivity/specificity results for ground truth known variants are presented according to different expected VAF bins, including 0.1%–0.3% VAF, >0.3% VAF, and all VAF. Notably, all of the increase in sensitivity occurred among the very low VAF TP variants in *Sample EfIS* ($0.1\% \leq \text{VAF} \leq 0.3\%$). Among the known TP variants in this VAF range, the Illumina pipeline called (i.e., PASS FILTER) 73%, whereas SNAQ-SEQ QC analysis of the same ILM VCF file with either PASS or LowSupport FILTER increased TP detection sensitivity to 86%, for a 13% increase in sensitivity.

DISCUSSION

Use of SNAQ-SEQ QC to report variants based on calculation of an LOB, LOD, and significance score for each variant at each targeted nucleotide position in each ctDNA sample is a significant advance compared to current practice in which a predefined VAF LOD is imposed or a threshold coverage across the targeted regions is required (Deveson, 2021; FDA, 2020). Mixture of each sample with a synthetic spike-in IS for each targeted region enables direct measurement of NGS technical background errors that are experiment-specific, sample-specific, and/or sequence-context specific. More accurate technical error measurements enable more accurate mutation detection. No other method provides within-run measurement of background, as is presented here. For comparison, the closest method is perhaps the approach to modeling sequence context-specific background errors after analysis of many cfDNA samples from healthy subjects (Davis et al., 2021; Abelson et al., 2020; Ma and Zhang, 2019; Cheng et al., 2015; Cibulskis et al., 2013). However, stochastic errors during library preparation and sequencing will vary from NGS experiment to experiment and in the absence of SNAQ-SEQ QC, we observed poor accuracy for calling variants with VAF <0.5% in ctDNA (Deveson, 2021). Causes of low accuracy for variants with VAF <0.5% may be

Table 2. Summary: SNAQ effect on detection sensitivity in *Sample EfIS*

Variant VAF range	Variant caller	LOW VAF	HIGH VAF	ALL VAF
Expected TP calls	–	56	24	80
Observed TP calls	ILM	41	23	64
Observed TP calls	SNAQ	48	23	71
Sensitivity	ILM	73%	96%	80%
Sensitivity	SNAQ	86%	96%	89%
SNAQ sensitivity increase	–	13%	0%	9%

Detection rate of ground truth TP variants by ILM variant caller (i.e., variants with ILM VCF PASS FILTER) or ILM caller plus SNAQ-SEQ QC (SNAQ) caller for variants with ILM VCF PASS + LowSupport FILTERS. ILM and SNAQ variant callers were compared among variants with mean inter-replicate VAF values in LOW range (VAF $\geq$ 0.10% and $\leq$ 0.3%), HIGH range (VAF > 0.3%), or ALL. Sensitivity was calculated by dividing the number of measured variants by expected variants, with the increase in sensitivity indicated in the SNAQ sensitivity increase row.

observed due to (1) variable low yield from library preparation, (2) variable loss of signal associated with stochastic sampling, and (3) low VAF signal-to-noise (i.e., position- and nucleotide-change-specific NGS background technical error). Variant callers (e.g., VarDict) based on statistical analysis alone are tuned to a narrow range of coverage and assumption of NGS error rates based on panels of "normal" samples. In contrast, the SNAQ-SEQ QC significance scores based on PET analysis adjust to sample- and target-specific coverage and directly tested error rate. This is evident in analysis of variants in *Sample EfIS* that were 5-fold diluted relative to *Sample DfIS* (Figure 3; Table 1) or in down-sampled datasets that simulated loading of 25 ng or 10 ng of sample (data not shown). Thus, the SNAQ-SEQ method will inform the user when any value cannot be reported with confidence due to low coverage. Importantly, we are able to minimize stochastic sampling effect on IS error detection by ensuring an IS/NT ratio of 2.5 or greater. Furthermore, under the specified conditions, in which a minimum IS:NT ratio of 2.5:1 is used, the IS will always be able to estimate the NGS background error for the given sample as long as the minimum NT alt count is observed. Therefore, data presented here support the expectation that any NT variants called with this approach will be reliable, even in clinical samples with sub-optimal sequence data resulting from low genomic input or low insert yields due to sample quality or unforeseen variation in reagents or equipment function.

In this and previous studies, there was no evidence that inclusion of SNAQ-SEQ IS in samples interfered with testing sensitivity or specificity. The SNAQ-SEQ IS mixture used in this study was designed to span each known actionable mutation. This IS mixture spanned 4% of the TST170 panel, thus there was no significant impact on sample coverage. Furthermore, mixture of SNAQ-SEQ IS with samples had no effect on workflow or performance of established targeted NGS platforms in an accompanying SEQC2 study (Deveson, 2021). Specifically, 50 ng of *Sample Ef,* termed *ct-low* in the Deveson SEQC2 study, was mixed with 10-fold higher SNAQ-SEQ IS than was used in that study before analysis by TST170. Despite the 10-fold higher IS in this study, sensitivity using the standard Illumina pipeline was unaffected. For example, presence of 10-fold higher IS had no adverse effect on specificity in that no false positives were detected in either study. These observations provide strong evidence that the SNAQ-SEQ IS did not interfere with vendor-approved variant calling performance. Based on these criteria, inclusion of SNAQ-SEQ IS with SEQC2-contrived ctDNA reference samples also did not interfere with the other hybrid capture library preparation platforms used in the parallel SEQC2 study (Deveson, 2021).

Adaptability of SNAQ-SEQ to other NGS platforms

The applicability and utility of SNAQ-SEQ in both UMI and non-UMI library hybrid capture NGS platforms is supported by data presented in this study. That said, there is significant variation among other NGS platforms not used in this study with respect to library preparation design, bioinformatics pipeline analysis, and variant-calling algorithms that might affect performance of SNAQ-SEQ. To implement SNAQ-SEQ with other NGS methods, pipelines, and variant callers, the process would need to be adjusted to account for performance differences. For example, it would be simple to adjust to a two alt NT cutoff for PASS instead of a three alt cutoff used by the Illumina TST170 pipeline or an NGS background not estimated by a Bonferroni corrected alpha. As mentioned in the STAR Methods, alignment to hg38 would simply involve appending the IS sequences to an hg38 FASTA.

Application of SNAQ-SEQ QC to non-UMI libraries

As reported in Results, the 0.5%–1.0% VAF LOD observed with application of SNAQ-SEQ QC to the Roche platform with non-UMI libraries was better than the 2% VAF cutoff for the same sample run in a non-UMI Roche platform without SNAQ-SEQ QC (Gong et al., 2021). In addition, when we analyzed the TST170 UMI data without application of UMI information to suppress technical error, we observed a VAF LOD of 0.5%–1%, similar to that observed for the Roche non-UMI library preparation. Thus, SNAQ-SEQ QC is applicable to, and improves, QC when applied to both UMI and non-UMI NGS datasets following hybrid capture enrichment.

Taken together, these findings demonstrate that application of synthetic spike-in IS will provide critical test- and sample-specific quality control in NGS diagnosis of actionable mutations, as they do in other key molecular diagnostic testing methods, including liquid and gas chromatography, mass spectroscopy (Dolan, 2012; Takats et al., 2005), and the FDA-approved Roche Cobas qPCR tests (Diagnostics, 2015).

Importantly, we assume that the IS initially is a reference sequence, without detectable error prevalence. We then assume that during library preparation the IS will accumulate the same amount of chemical and physical damage as the actual sample. We have determined that most technical error in samples is well-modeled by the IS in that there is very high correlation between IS and NT samples for site-specific error (Blomquist et al., 2015). That said, actual clinical circulating free DNA (cfDNA) samples may accumulate damage during sample processing and storage prior to the spiking in of the IS, and this could introduce detectable variants not mimicked by IS. In this study, it was not our aim to

Table 3. Average IS/NT ratio for each IS in the SNAQ-SEQ IS mixture

IS plasmid ID	Mean IS:NT ratio	SD	CV (%)
JAK3	2.9	0.14	5
MAP2K1_3	3.4	0.50	14
NRAS_3	3.6	0.18	5
PDGFRA_12	3.7	0.11	3
ERBB2_RST	3.8	0.18	5
NRAS_2	3.9	0.35	9
EGFR19_RST	4.0	0.13	3
PTEN	4.2	0.30	7
EGFR21_RST	4.4	0.25	6
KRAS_3	4.4	0.17	4
JAK2_14	4.5	0.31	7
ATM	4.5	0.21	5
FGFR2_7	4.6	0.12	3
EGFR18_RST	4.7	0.32	7
PIK3CA_10	4.8	1.00	21
JAK2_12	4.8	0.67	14
EGFR20_RST	4.8	0.26	5
NOTCH1	5.0	0.76	15
TP53	5.1	0.74	14
MET	5.2	0.37	7
PTPN11	5.6	0.12	2
FBXW7_16	6.0	0.18	3
CTNNB1	6.0	0.47	8
FGFR3	6.1	0.58	9
KIT_17	6.1	0.71	11
KIT_9	6.8	0.54	8
FBXW7_13	7.3	0.28	4
All plasmid mean	4.8	0.37	7.6

The position-specific IS:NT ratio for each targeted region was averaged across each of the targeted regions within a plasmid for four replicate libraries of *Samples BfIS*, *DfIS*, and *EfIS* (mean). Standard deviation (SD) and coefficient of variation (CV) for each plasmid also are presented as well as the mean, SD, and CV for the entire mixture of IS plasmids (all plasmid mean).

differentiate variants introduced after sample collection but prior to library preparation from the true biological mutations. However, it is clear from this study that SNAQ-SEQ will increase the ability to accurately distinguish technical error introduced during library preparation and sequencing from damage accumulated during sample processing and storage. As such, SNAQ-SEQ analysis will enable more systemic analysis and control for this type of pre-analytical error. How well our fragmented IS will mimic clinical plasma cfDNA will be resolved with future studies.

In summary, the SNAQ-SEQ orthogonal biochemical/statistical approach to NGS QC improved clinical sensitivity for measurement of ground truth-positive variants in SEQC2-contrived ctDNA reference samples without loss of clinical specificity. This advance was possible due to increased analytical specificity. In particular, use of SNAQ-SEQ unique within-run QC enabled determination of technical error, LOB, confidence limits, and calculation of lower LOD for each variant at each nucleotide position in each sample. This represents a major advance because, due to the limited number of genome copies in ctDNA specimens, it is critical to maximize the ability to reliably report low VAF actionable mutations. This approach controls for systematic chemical-physical NGS errors, pipeline-specific non-systematic errors, and coverage. Our work promises to provide a reliable method to increase the yield of reportable variant calls in low input clinical ctDNA samples.

Limitations of the study

A limitation of this study is that contrived ctDNA samples were used due to challenges in acquiring sufficient amounts of clinical sample material to use in the multiple collaborative studies conducted by the SEQC2 group. Although efforts were made to mimic an actual clinical ctDNA sample, the performance characteristics of the mixture of fragment IS and associated methodology will need to be validated in future studies using clinical samples from multiple sources.

In addition, this study revealed analytical performance characteristics of the SNAQ-SEQ QC method that may be improved. First, there was systematic inter-platform variation in hybrid capture of IS relative to NT (data not shown). That said, once the appropriate IS input to achieve the desired IS:NT ratio (i.e., >2.5 IS:NT) was empirically identified for a hybrid capture platform, that platform yielded an inter-experimentally consistent result (Table 3). Therefore, until the basis for this systematic bias is better understood, the protocol will be to experimentally determine for each hybrid capture panel platform the IS DNA input required to achieve the desired IS:NT ratio following library preparation. Second, higher IS coverage increases power to establish LOB based on PET, due to the larger number of *potential* observations for technical error. We empirically determined that an IS:NT ratio of at least 2.5:1 provided optimal power to measure LOB by PET across all positions/nucleotide exchanges represented in this study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Design and synthesis of reference DNA internal standard [IS] spike-ins for use in hybrid capture NGS
  - SEQC2 reference samples
  - Establishment of ground truth dataset
  - Illumina TST170 hybrid capture NGS library preparations and analysis
  - Illumina UMI pipeline and variant calling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - SNAQ-SEQ QC variant calling of Illumina UMI deduplicated BAM files
  - Performance metrics

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2021.100106.

ACKNOWLEDGMENTS

All SEQC2 participants freely donated their time, samples, reagents, and computing resources for the completion and analysis of this project. We thank SEQC2 Oncopanel Sequencing Working Group participants for providing useful feedback during manuscript preparation. This work was supported by The George Isaac Chair for Cancer Research fund and grants from the National Cancer Institute (5U24CA086368; 1U01CA243483), the FDA (BAA grant HHSF223201510172C), the National Health and Medical Research Council (NHMRC) of Australia (grants APP1108254, APP1114016, APP1173594, and Cancer Institute NSW Early Career Fellowship 2018/ECF013). None of these funding bodies had a role in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript.

AUTHOR CONTRIBUTIONS

J.C.W. contributed to the conception and design of the study, analysis and interpretation of data, and drafting and revision of the manuscript. T.B.M. contributed to the conception and design of the study; the acquisition, analysis, and interpretation of data; the creation of software used in the work; and the drafting and revision of the manuscript. B.A. contributed to the acquisition and analysis of the data and the creation of software used in the work. E.L.C. contributed to the conception and design of the study, acquisition and interpretation of data, and revision of the manuscript. D.J.C. contributed to the conception and design of the study; the acquisition, analysis, and interpretation of data; the creation of software used in the work; and the revision of the manuscript. T.M.B. contributed to the conception and design of the study, the creation of software used in the work, and the revision of the manuscript. W.D.J. contributed to the acquisition of data and revision of manuscript. A.W. contributed to the acquisition of data. J.S.L. contributed to the acquisition, analysis, and interpretation of data and revision of the manuscript. N.H. contributed to the acquisition, analysis, and interpretation of data. T.A.R. contributed to the acquisition, analysis, and interpretation of data and revision of the manuscript. N.N. contributed to the acquisition, analysis, and interpretation of data and revision of the manuscript. R.K. contributed to the interpretation of data and revision of the manuscript. G.C. contributed to the acquisition, analysis, and interpretation of data and revision of the manuscript. Q.-Z.L. contributed to the acquisition, analysis, and interpretation of data and revision of the manuscript. D.J.J. contributed to the analysis and interpretation of data and revision of the manuscript. I.W.D. contributed to the interpretation of data and revision of the manuscript. T.R.M. contributed to the interpretation of data and revision of the manuscript. L.W. contributed to the analysis and interpretation of data, the creation of software used in the work, and the revision of the manuscript. J.X. contributed to the conception and design of the study, the analysis and interpretation of data, the creation of software used in the work, and the revision of the manuscript.

DECLARATION OF INTERESTS

J.C.W. has 5%–10% equity interest in and serves as a consultant to AccuGenomics, Inc. Technology relevant to this manuscript was developed and patented by J.C.W., E.L.C., and T.B.M. and is licensed to AccuGenomics, Inc. These relationships do not alter our adherence to all policies on sharing data and materials. The views presented in this article do not necessarily reflect the current or future opinion or policy of the US FDA. Any mention of commercial products is for clarification and not intended as an endorsement.

Received: April 12, 2021
Revised: May 31, 2021
Accepted: October 11, 2021
Published: November 3, 2021

REFERENCES

Abelson, S., Zeng, A.G.X., Nofech-Mozes, I., Wang, T.T., Ng, S.W.K., Minden, M.D., Pugh, T.J., Awadalla, P., Shlush, L.I., Murphy, T., et al. (2020). Integration of intra-sample contextual error modeling for improved detection of somatic mutations from deep sequencing. Sci. Adv. *6*, eabe3722.

Beaubier, N., Tell, R., Lau, D., Parsons, J.R., Bush, S., Perera, J., Sorrells, S., Baker, T., Chang, A., Michuda, J., et al. (2019). Clinical validation of the tempus xT next-generation targeted oncology sequencing assay. Oncotarget *10*, 2384–2396.

Blomquist, T.M., Crawford, E.L., Lovett, J.L., Yeo, J., Stanoszek, L.M., Levin, A., Li, J., Lu, M., Shi, L., Muldrew, K., and Willey, J.C. (2013). Targeted RNA-sequencing with competitive multiplex-PCR amplicon libraries. PLoS One *8*, e79120.

Blomquist, T., Crawford, E.L., Yeo, J., Zhang, X., and Willey, J.C. (2015). Control for stochastic sampling variation and qualitative sequencing error in next generation sequencing. Biomol. Detect Quantif. *5*, 30–37.

Chen, H.Z., Bonneville, R., and Roychowdhury, S. (2019). Implementing precision cancer medicine in the genomic era. Semin. Cancer Biol. *55*, 16–27.

Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z.Y., Won, H.H., Scott, S.N., and Brannon, A.R. (2015). Memorial Sloan Kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT) a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J. Mol. Diagn. *17*, 251–264.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Corcoran, R.B., and Chabner, B.A. (2018). Application of cell-free DNA analysis to cancer treatment. N. Engl. J. Med. *379*, 1754–1765.

Craig, D.W., Nasser, S., Corbett, R., Chan, S.K., Murray, L., Legendre, C., Tembe, W., Adkins, J., Kim, N., Wong, S., et al. (2016). A somatic reference standard for cancer genome sequencing. Sci. Rep. *6*, 24607.

Craig, D.J., Morrison, T., Khuder, S.A., Crawford, E.L., Wu, L., Xu, J., Blomquist, T.M., and Willey, J.C. (2019). Technical advance in targeted NGS analysis enables identification of lung cancer risk-associated low frequency TP53, PIK3CA, and BRAF mutations in airway epithelial cells. BMC Cancer *19*, 1081.

Davis, E.M., Sun, Y., Liu, Y.L., Kolekar, P., Shao, Y., Szlachta, K., Mulder, H.L., Ren, D.R., Rice, S.V., Wang, Z.M., et al. (2021). SequencErr: measuring and suppressing sequencer errors in next-generation sequencing data. Genome Biol. *22*, 1–18.

Deveson, I. (2021). Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. Nat. Biotechnol. *39*, 1–14.

Diagnostics, R. (2015). Roche molecular systems: cobas EGFR mutation test v2. https://diagnostics.roche.com/us/en/products/params/cobas-egfr-mutation-test-v2.html.

Dolan, J.W. (2012). When should an internal standard be used? LCGC N. Am. *30*, 316–322.

FDA (2020). PMA approval letter for Guardant360 CDx. https://www.accessdata.fda.gov/cdrh_docs/pdf20/P200010A.pdf.

FDA SEQC2 (2019). MicroArray/Sequencing Quality Control (MAQC/SEQC). https://www.fda.gov/science-research/bioinformatics-tools/microarraysequencing-quality-control-maqcseqc#maqc_iv.

Fu, G.K., Xu, W., Wilhelmy, J., Mindrinos, M.N., Davis, R.W., and Xiao, W. (2014). Molecular indexing enables quantitative targeted RNA sequencing and reveals poor efficiencies in standard library preparations. Proc. Natl. Acad. Sci. *111*, 1891–1896.

Gargis, A.S., Kalman, L., Bick, D.P., da Silva, C., Dimmock, D.P., Funke, B.H., Gowrisankar, S., Hegde, M.R., Kulkarni, S., Mason, C.E., et al. (2015). Good laboratory practice for clinical next-generation sequencing informatics pipelines. Nat. Biotechnol. *33*, 689.

Gong, B.S., Li, D., Kusko, R., Novoradovskaya, N., Zhang, Y.F., Wang, S.Z., Pabon-Pena, C., Zhang, Z.H., Lai, K., Cai, W.S., et al. (2021). Cross-oncopanel study reveals high sensitivity and accuracy with overall analytical performance depending on genomic regions. Genome Biol. *22*, 1–23.

Illumina (2020). BaseSpace Sequence Hub. https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub.html.

Jones, W., Gong, B., Novoradovskaya, N., Li, D., Kusko, R., Richmond, T.A., Johann, D.J., Bisgin, H., Sahraeian, S.M.E., Bushel, P.R., et al. (2021). A verified genomic reference sample for assessing performance of cancer panels detecting small variants of low allele frequency. Genome Biol. *22*, 111.

Kennedy, S.R., Salk, J.J., Schmitt, M.W., and Loeb, L.A. (2013). Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. Plos Genet. *9*, e1003794.

Kuderer, N.M., Burton, K.A., Blau, S., Rose, A.L., Parker, S., Lyman, G.H., and Blau, C.A. (2016). Comparison of 2 commercially available next-generation sequencing platforms in oncology. JAMA Oncol. *3*, 996–998.

Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., Mcewen, R., Johnson, J., Dougherty, B., Barrett, J.C., and Dry, J.R. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. *44*, e108.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11.

Ma, X.T., and Zhang, J.H. (2019). Analysis of error profiles in deep next-generation sequencing data. Cancer Res. *79*, 50.

Merker, J.D., Oxnard, G.R., Compton, C., Diehn, M., Hurley, P., Lazar, A.J., Lindeman, N., Lockwood, C.M., Rai, A.J., Schilsky, R.L., et al. (2018). Circulating tumor DNA analysis in patients with cancer: American Society of Clinical Oncology and College of American Pathologists Joint Review. J. Clin. Oncol. *36*, 1631–1641.

Newman, A.M., Bratman, S.V., To, J., Wynne, J.F., Eclov, N.C., Modlin, L.A., Liu, C.L., Neal, J.W., Wakelee, H.A., Merritt, R.E., et al. (2014). An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat. Med. *20*, 548–554.

Newman, A.M., Lovejoy, A.F., Klass, D.M., Kurtz, D.M., Chabon, J.J., Scherer, F., Stehr, H., Liu, C.L., Bratman, S.V., Say, C., et al. (2016). Integrated digital error suppression for improved detection of circulating tumor DNA. Nat. Biotechnol. *34*, 547–555.

Oxnard, G.R., Thress, K.S., Alden, R.S., Lawrance, R., Paweletz, C.P., Cantarini, M., Yang, J.C., Barrett, J.C., and Janne, P.A. (2016). Association between plasma genotyping and outcomes of treatment with Osimertinib (AZD9291) in advanced non-small-cell lung cancer. J. Clin. Oncol. *34*, 3375–3382.

Rodon, J., Soria, J.C., Berger, R., Miller, W.H., Rubin, E., Kugel, A., Tsimberidou, A., Saintigny, P., Ackerstein, A., Brana, I., et al. (2019). Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. Nat. Med. *25*, 751–758.

Rossi, G., and Ignatiadis, M. (2019). Promises and pitfalls of using liquid biopsy for precision medicine. Cancer Res. *79*, 2798–2804.

Salk, J.J., Schmitt, M.W., and Loeb, L.A. (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat. Rev. Genet. *19*, 269–285.

Sandmann, S., de Graaf, A.O., Karimi, M., van der Reijden, B.A., Hellstrom-Lindberg, E., Jansen, J.H., and Dugas, M. (2017). Evaluating variant calling tools for non-matched next-generation sequencing data. Sci. Rep. *7*, 43169.

Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B., and Loeb, L.A. (2012). Detection of ultra-rare mutations by next-generation sequencing. Proc. Natl. Acad. Sci. U. S. A. *109*, 14508–14513.

Schwaederle, M., Zhao, M., Lee, J.J., Lazar, V., Leyland-Jones, B., Schilsky, R.L., Mendelsohn, J., and Kurzrock, R. (2016). Association of biomarker-based treatment strategies with response rates and progression-free survival in refractory malignant neoplasms: a meta-analysis. JAMA Oncol. *2*, 1452–1459.

Sherry, S.T., Ward, M., and Sirotkin, K. (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. *9*, 677–679.

Siravegna, G., Marsoni, S., Siena, S., and Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. Nat. Rev. Clin. Oncol. *14*, 531–548.

Squillace, R.M., Frampton, G.M., Stephens, P.J., Ross, J.S., and Miller, V.A. (2015). Comparing two assays for clinical genomic profiling: the devil is in the data. OncoTargets Ther. *8*, 2237–2242.

Stetson, D., Ahmed, A., Xu, X., Nuttall, B.R.B., Lubinski, T.J., Johnson, J.H., Barrett, J.C., and Dougherty, B.A. (2019). Orthogonal comparison of four plasma NGS tests with tumor suggests technical factors are a major source of assay discordance. JCO Precis. Oncol., 1–9.

Takats, Z., Wiseman, J.M., and Cooks, R.G. (2005). Ambient mass spectrometry using desorption electrospray ionization (DESI): instrumentation, mechanisms and applications in forensics, chemistry, and biology. J. Mass Spectrom. *40*, 1261–1275.

Thermo Fisher Scientific (2020). Oncomine Pan-Cancer Cell-Free Assay User Guide. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017064_PanCancer_UG.pdf.

Torga, G., and Pienta, K.J. (2018). Patient-paired sample congruence between 2 commercial liquid biopsy tests. JAMA Oncol. *4*, 868–870.

Vasan, N., Baselga, J., and Hyman, D.M. (2019). A view on drug resistance in cancer. Nature *575*, 299–309.

Volckmar, A.L., Sultmann, H., Riediger, A., Fioretos, T., Schirmacher, P., Endris, V., Stenzinger, A., and Dietz, S. (2018). A field guide for cancer diagnostics using cell-free DNA: from principles to practice and clinical applications. Genes Chromosomes Cancer *57*, 123–139.

Zutter, M.M., Bloom, K.J., Cheng, L., Hagemann, I.S., Kaufman, J.H., Krasinskas, A.M., Lazar, A.J., Leonard, D.G., Lindeman, N.I., Moyer, A.M., et al. (2015). The cancer genomics resource list 2014. Arch. Pathol. Lab Med. *139*, 989–1008.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
SEQC2 contrived ctDNA reference materials, Samples B, D and E	Jones, et al., 2021 SEQC2 Oncopanel Sequencing Working Group	N/A
Critical commercial assays		
SNAQ-SEQ IS mixture	This paper, Acccugenomics, Inc.	N/A
Roche SeqCap EZ Choice custom PHC panel	Roche	06266312001
Illumina TST170 hybrid capture panel	Illumina, Inc.	OP-101-1004
Deposited data		
Raw and analyzed data, dataset1	This paper	http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/
Raw and analyzed data, dataset2	This paper	http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/
Human reference genome NCBI build 37, GRCh37	Broad Institute	https://registry.opendata.aws/broad-references/
IS-specific reference genome	This paper, Acccugenomics, Inc.	http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/SIST_REFS.zip
Software and algorithms		
SIST3.1.4.pl Split Script	Accugenomics, Inc.	http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/
The Genome Analysis Toolkit (GATK) v4.0.11.0	Broad Institute	https://gatk.broadinstitute.org/hc/en-us
VarDict 1.5.7	Lai et al. (2016)	https://github.com/AstraZeneca-NGS/VarDict
R library package exactci	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/exactci/index.html
Samtools	Li et al. (2009)	http://samtools.sourceforge.net/

RESOURCE AVAILABILITY

Lead contact

Requests for information, reagents and resources will be coordinated by Dr. James C. Willey (james.willey2@utoledo.edu).

Materials availability

The SNAQ-SEQ IS mixture used in this work is available through AccuGenomics, Inc. Availability of the SEQC2 contrived ctDNA reference materials used in this study is limited due to extensive use by members of the Sequencing Quality Control Consortium Phase 2 (SEQC2) group led by the FDA.

Data and code availability

- The full sequencing datasets representing data presented in this manuscript along with the custom PERL script (SIST3.1.4.pl Split Script) used to split reads into NT and IS FASTQ files are accessible at this site: http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/. The IS-specific reference genome is accessible at: http://accugenomicspublic.s3-website-us-east-1.amazonaws.com/LBMS/SIST_REFS.zip.
- This paper does not report original code
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

CellPress
OPEN ACCESS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

SEQC2 contrived ctDNA reference materials used in this study were obtained from the SEQC2 Oncopanel Sequencing Working Group (Jones et al., 2021) as intact and fragmented genomic DNA from mixtures of human cell lines (see SEQC2 Reference Samples). No cell lines, subjects or animal models were used for this study.

METHOD DETAILS

The aims of this study included the following. First, design methods for incorporation of synthetic human genome reference sequence internal standard spike-ins into hybrid-capture targeted-NGS UMI library analysis. Second, validate these methods as orthogonal QC in SEQC2 *Sample A* in non-UMI libraries following Roche SeqCap EZ Choice custom PHC panel enrichment, positional deduplication, and Poisson Exact Test (PET) analysis of NT variant reads relative to IS variant reads. Third, evaluate clinical utility of this method in SEQC2 contrived ctDNA *Samples DfIS* and *EfIS* following Illumina TST170 hybrid capture panel enrichment, UMI deduplication, and PET analysis of NT variant and IS variant reads.

Design and synthesis of reference DNA internal standard [IS] spike-ins for use in hybrid capture NGS

We prepared a mixture of synthetic Standardized Nucleic Acid Quantification for Sequencing (SNAQ-SEQ) IS DNA controls that span 38 kb and correspond to 32 exons and 61 actionable mutations (Craig et al., 2019; Thermo Fisher Scientific, 2020; Liu et al., 2018). This IS mixture was used successfully in previous work to provide quality control in amplicon libraries prepared with DNA extracted from tissue samples (Craig et al., 2019). To provide quality-control for targeted hybrid capture library analysis of SEQC2 contrived ctDNA reference materials used in this study, the IS mixture was enzymatically sheared to approximate the modal distribution of typical ctDNA samples (Deveson, 2021). Of this 38kb region, 6.8 kb overlap with the Roche SeqCap EZ Choice custom PHC panel used to establish ground truth in *Sample A* and the Illumina TST170 hybrid capture panel used to analyze the SEQC2 contrived ctDNA *Samples DfIS* and *EfIS*.

Each IS sequence matched the respective GRCh37 (hg19) human reference genome sequence (sourced from the Broad Institute) except for unique dinucleotides (DN) every 40–60 bp to allow for bioinformatic separation of IS from native template (NT) genomic reads. Each IS control spanned sequence that comprised one or more full exons containing the actionable mutation(s) plus up to 500 bp of DNA flanking each side of each mutation. Flanking DNA was included in an effort to ensure that enzymatic or mechanical fragmentation methods would yield a size spectrum around each targeted mutation that closely approximated that of sample DNA. The DN positions were chosen to avoid actionable mutation sites, known primer binding sites used by commercial PCR amplicon library vendors, and known genomic variants with VAF $\geq$ 0.1% based on HapMap database (Sherry et al., 1999). Each IS control was cloned into a pUC plasmid and the sequence was verified. Other than the DN sites, it was expected that every IS position matched reference genome due to the low replication error rate of *E. coli* DNA, which is $<1 \times 10^{-8}$. Thus, spanning 38,000 bp the expected error rate would be < 0.004%. Plasmids were linearized, quantified and mixed at an equimolar concentration with <20% CV for inter-IS variation in mixture. For this study, a large batch of IS mixture was then enzymatically fragmented with goal to achieve a 130–170 bp modal size distribution (Figure S1), closely approximating that observed in typical ctDNA clinical specimens as well as the contrived SEQC2 ctDNA reference materials (Deveson, 2021). Frozen aliquots of IS mixture were then distributed to participating SEQC2 sites for use in this and related SEQC2 studies (Deveson, 2021).

SEQC2 reference samples

The SEQC2 Oncopanel Sequencing Working Group developed five reference samples (*Samples A*, *B*, *C*, *D*, and *E*), described in detail in related studies from the SEQC2 project (Jones et al., 2021). SEQC2 *Sample A* comprised a mixture of ten cancer cell lines, with an expected VAF for most variants ranging from 2% to 15%. Lower VAF samples were created by mixing *Samples A* and *B*. Briefly, *Sample A* was combined with DNA extracted from a non-cancer background cell-line (*Sample B*) to create three additional reference samples, *Sample C* (1:2 dilution; 50% *Sample A*/50% *Sample B*), *Sample D* (1:5 dilution; 20% *Sample A*/80% *Sample B*), and *E* (1:25 dilution; 4% *Sample A*/96% *Sample B*). We used four of these samples in this study; *Samples A*, *B*, *D*, and *E*. Samples A and B were mixed together in a 2-fold serial dilution series to develop enriched ground truth data set within our 6.8 kg region of interest (see below section). *Samples B, D,* and *E* were enzymatically fragmented to create *Samples Bf, Df*, and *Ef* with modal 130–170 bp fragment size (Deveson, 2021). *Samples Bf*, *Df*, and *Ef* were mixed with the fragmented IS spike-in mixture to make SEQC2 test *Samples BfIS, DfIS* and *EfIS*, respectively. Each of these test samples then were distributed to each test site that participated in this study. With this design, it was intended that following mixture with a sample, each IS DNA sequence would closely parallel its corresponding sample sequence during hybrid capture library preparation and sequencing yet be easily separated bioinformatically from sample sequence during pipeline analysis.

SEQC2 reference Samples A, B, D, and E were included in this study. Analysis of *Samples A* and *B* was necessary for establishment of ground truth (see below section). Analysis of *Sample Df* was a used to develop confidence with the analysis pipeline, including PET analysis. Analysis of *Sample Ef* was used to test clinical utility. Preparation and analysis of a *Sample CfIS* would have added cost and would not have added additional worthwhile information given that the SNAQ-SEQ method was capable of calling all known positives in *Sample Df* in which the average *Sample A* known variant VAF was lower than what would have been in a *Sample CfIS*.

Establishment of ground truth dataset

There was limited ground truth data available for *Samples A* and *B* within the 6.8 kb TST170 × SNAQ-SEQ IS overlap region based on the related SEQC studies (Jones et al., 2021). Because this would limit the assessment of SNAQ-SEQ controls as an independent NGS quality control, a more complete ground truth was created. Specifically, SEQC2 *Sample A* was serially two-fold diluted with *Sample B* (normal subject DNA) (Jones et al., 2021) to create 1, 2, 4, 8, 16, 32 and 64-fold dilutions. Then, each serial dilution sample was mixed with SNAQ-SEQ synthetic spike-in controls to achieve IS:NT ratio of >2.5 for each target (Jones et al., 2021). Following enzymatic fragmentation, triplicate aliquots of each sample were enriched on the Roche SeqCap EZ Choice custom PHC panel. The Roche panel x SNAQ-SEQ IS overlap was 7.5 kb and comprised all of the 6.8 kb TST170 × SNAQ-SEQ IS overlap region. Following enrichment, non-UMI libraries were prepared then sequenced on an Illumina Novaseq S4 at the University of Texas Southwestern. As schematically depicted in Figure 1A, the FASTQ files were shipped on hard drives to AccuGenomics, Inc., custom PERL script was used to split reads into NT and IS FASTQ files using IS-specific reference genome in which the IS overlapping regions in the NT-specific hg19 FASTA were replaced with IS sequence, reads were aligned with BWA mem 0.7.17-r1194-dirty (default parameters), removing duplicate reads using GATK markDuplicates v4.0.11.0 (default parameters) and a genomic (g)VCF file generated with VarDict 1.5.7 accepting variants with >0.1% for each sample. Ground truth positive (TP) variants in *Sample A* were defined as those for which VAF responded to serial dilution. The serial dilution VAF results are presented in Figure 2. The positionally deduplicated coverage for each target was above 1,100 for each replicate, and this coverage was sufficient to confirm that a position was ground truth negative (TN) down to at least 0.5% VAF. The goal was to both establish ground truth positive (TP) variants and confirm that any other variant was a false positive (FP) by confirming lack of titration response. Data from this study also were used to develop the SEQC-SEQ QC method, as presented in Figures 1C and 1D and presented in Results section.

Illumina TST170 hybrid capture NGS library preparations and analysis

50 ng aliquots of *Samples BfIS*, *DfIS* and *EfIS* were each included in four replicate TST170 UMI libraries prepared at Q2 Solutions, Inc. according to the ILM2 method previously described (Deveson, 2021), and sequenced on NextSeq. Following target enrichment, library preparation, and sequencing, NT and IS reads from *Samples BfIS, DfIS*, or *EfIS*, were bioinformatically isolated into their respective bins. Specifically, resulting FASTQ files were separated into NT, IS and suspicious using a custom designed SIST3.1.4.pl PERL script prior to running each sequence pool into the analysis pipeline (Figure 1A). Suspicious reads had both NT and IS base changes in the same read, made up 0.1% of reads, and likely arose by sequencing error or recombination between NT and IS during library preparation.

Illumina UMI pipeline and variant calling

After bioinformatic isolation of NT and IS reads into different FASTQ files at AccuGenomics the NT and IS FASTQ files were uploaded to BaseSpace (Illumina, 2020). For analysis of SEQC2 contrived samples DfIS and EfIS, Illumina (co-authors NH and JL) provided to AccuGenomics (co-author TM) the GRCh37 (hg19) reference sequence that Illumina used in their standard pipeline, AccuGenomics replaced the IS overlapping regions with IS sequence, then AccuGenomics returned the IS reference genome to Illumina. Each NT FASTQ file was then processed through the standard Illumina protocol using the respective NT and IS reference sequence information, which yielded a UMI deduplicated BAM file, and VCF file. Each IS FASTQ file was processed through the same protocol except for alignment to the hg19 human reference genome modified to include the IS-specific dinucleotide changes, which yielded UMI deduplicated BAM files and VCF files. Alignment to hg38 would simply involve appending the IS sequences to an hg38 FASTA. Because each IS was known to be reference sequence, any variant alleles were, by definition, technical error. Analysis of the IS VCF file indicated certain artifactual variants that were not detected by the Illumina variant caller. These variants likely were created during fragment end repair from 3-prime self- priming events and it was determined empirically that a VarDict PMEAN filter effectively removed them. Specifically, the VarDict variant caller PMEAN score indicated the average distance a variant occurred from a fragment end. To eliminate these end artifacts, VarDict analysis generated a PMEAN score for each variant in UMI deduplicated BAM files from each of 12 libraries (four replicate libraries each for *Samples BfIS*, *DfIS* and *EfIS*), then an R-script was used to calculate the median PMEAN score for each variant across all 12 samples. Analysis of the Illumina VCF file variants indicated that a median PMEAN of 25 would separate the fragment end repair artifact variants from TP variants (Figure S2). The <25 PMEAN positions were blacklisted from further analysis, reducing the total IS and NT variant sequences by 10%. Following application of the PMEAN 25 filter, Illumina UMI deduplicated BAM files were used to generate VCF files that reported all variants and variant positions. Within each Illumina VCF file, called variants were annotated as FILTER PASS, and statistically less likely non-called variants were annotated with FILTER LowSupport, Blacklist, or LowVar.

QUANTIFICATION AND STATISTICAL ANALYSIS

SNAQ-SEQ QC variant calling of Illumina UMI deduplicated BAM files

At AccuGenomics, SNAQ-SEQ QC analysis was combined with Poisson Exact Test (PET) analysis to calculate the statistical difference between each sample library NT variant VAF and respective IS variant VAF based on NT variant count and position coverage and IS count and position coverage. PET performs an exact test of a simple null hypothesis about the ratio between two rate parameters in Poisson distribution. SNAQ-SEQ QC with PET analysis was performed on any variant meeting the following criteria: a) not in

Illumina's blacklist, b) not in the PMEAN blacklist, c) $\geq$ three deduplicated reads, and 4) Illumina VCF FILTER containing PASS or LowSupport.

Because the IS was reference sequence prior to library preparation, it was assumed that any IS variant observed following library preparation and sequencing resulted from technical error. An NT variant with VAF PET score less than the cutoff indicated the genomic variant passed SNAQ-SEQ QC and was called. An NT variant VAF PET score greater than the significance cutoff was not distinguishable from NGS error and was not called. SNAQ-SEQ analysis of IS, which treats the IS as the ''sample'' and the NT sequence as ''background'' was used to examine the test pipeline performance, as a working pipeline should yield no statistically significant IS variants.

Performance metrics

Traditionally, a false negative (FN) is a ground truth positive (TP) variant not detected by sequence analysis, and a false positive (FP) is a variant detected at a ground truth negative (TN) site by sequence analysis. With SNAQ-SEQ QC, a FN could arise when a TP variant did not have a significant PET score, and a FP may arise if the observed NT variant derives from technical error in the sample that does not covary to the same extent in the IS. In addition, a true positive may be missed if it is not initially called by the variant calling software as PASS or LowSupport, or if removed in a blacklist.

Sequencing datasets from libraries prepared with loading 50 ng of sample were down-sampled using Samtools (Li et al., 2009) view to simulate loading of 25 ng or 10 ng of sample (Deveson, 2021). Statistical differences between each NT and respective IS variant VAF were calculated by Poisson Exact Test (PET) using R library package exactci (https://cran.r-project.org/package=exactci). The 5% alpha used as the cutoff for NT variant VAF significance was adjusted using a Bonferroni correction, 0.05 divided by the number of the sample NT variants being tested (n = 20 for each replicate library of *Sample EfIS*).