



OPEN Molecular counting enables accurate and precise quantification of methylated ctDNA for tumor-naive cancer therapy response monitoring

Patrick Peiyong Ye¹✉, Robb Viens¹, Katherine E Shelburne¹, Sydne Scot Langpap¹, Xavier S Bower¹, Jonathan Jiahui Shi¹, Wen Zhou¹, Jan Christian Wignall¹, Joyce Jiawei Zhu¹, Brian D. Woodward², Hatim Husain², David S. Tsao¹ & Oguzhan Atay¹

Personalized cancer treatment can significantly extend survival and improve quality of life for many patients, but accurate and real-time therapy response monitoring remains challenging. To overcome logistical and technical challenges associated with therapy response monitoring via imaging scans or assays that track the variant allele fraction (VAF) of somatic mutations in circulating tumor DNA (ctDNA), we developed a tumor-naive liquid biopsy assay that leverages Quantitative Counting Template (QCT) technology to accurately and precisely quantify methylated ctDNA (Northstar Response™). The assay achieves <10% coefficient of variation at 1% tumor fraction, which is 2 × lower than tumor-naive, targeted-panel approaches using VAF. The assay accurately distinguishes 0.25% absolute changes in contrived tumor fraction (AUC > 0.94) and performs well in 12 solid tumor types. Finally, in a small cohort of patients with lung, colorectal, or pancreatic cancer, the assay detected changes in ctDNA methylation that correlate with clinical outcomes. With its precise quantification of ctDNA methylation, Northstar Response is a novel tool for therapy response monitoring with the potential to inform clinical decision making for cancer treatment.

While novel cancer treatments continue to be developed at an unprecedented pace, assessing whether a cancer treatment is effective for a particular patient remains cumbersome and relatively qualitative. Imaging is the current standard for measuring the state of a patient's cancer; however, therapy response monitoring using imaging faces limitations in accuracy and precision. Specifically, a patient's response to therapy may not be accurately reflected on imaging due to tumoral heterogeneity¹, pseudoprogression in patients receiving immunotherapy², scar tissue surrounding the tumor, or low contrast in certain tissues like bone and the peritoneum^{3–5}. These measurements may also be imprecise due to the subjectivity involved with measuring radiographic images¹. In addition, the relative infrequency of imaging, with scans often several months apart, lengthens the lead time for assessing response to treatment and identifying when a patient is progressing on a particular therapy. Increasing the frequency of imaging can be infeasible due to access, logistics of traveling to imaging centers, and cost. Protein biomarkers have some but limited value for monitoring, with varying utility across cancer types⁶. Thus, there is a need to more accurately, precisely, and frequently assess the extent to which a patient's cancer is responding to treatment to inform clinical decision making.

Non-invasive liquid biopsies that assay cell-free DNA (cfDNA) have been used to quantify the levels of circulating tumor DNA (ctDNA) from a blood sample⁷. Several clinical studies have found changes in ctDNA levels to be predictive of progression^{8–15}, indicating that ctDNA contains accurate and quantifiable information for longitudinally tracking tumor progression. These studies use commercially available liquid biopsies that quantify the abundance of ctDNA via measurements of the variant allele fraction (VAF) of somatic mutations, using either a tumor-naive, targeted-panel approach or a tumor-informed approach. Tumor-naive, targeted-panel assays face technical limitations for precise therapy response monitoring. First, these assays may not detect somatic mutations in a significant fraction of advanced cancer patients, around 10% to 20% averaged across cancer types^{15–18}, rendering the assay unusable for therapy response monitoring. If variants are detected, the

¹BillionToOne, Inc, Menlo Park, CA, USA. ²University of California San Diego Moores Cancer Center, La Jolla, CA, USA. ✉email: patrick@billiontoone.com

number of present variants can be limited (a mean between 3 and 4)^{13,16}, and the VAF is often small (< 0.5%)^{16,18,19}, reflecting a low abundance of total variant molecules. This results in a high coefficient of variation (CV) due to Poisson sampling alone. For example, a single variant at 0.5% VAF in a sample with 2,000 genomic equivalents (GE) would contain an average of 10 variant molecules, resulting in a minimum CV of 32% even with perfect assay quantification accuracy. A high CV means that each individual measurement has a high amount of noise and variability, making longitudinal measurements difficult to interpret, and potentially results in spurious apparent changes in ctDNA abundance in technical replicates simply due to random sampling²⁰.

To better ensure the detection of a sufficient number of somatic mutations and thus sufficient variant molecules, tumor-informed assays have been developed, in which a personalized set of somatic mutations are identified from a tumor biopsy and then assayed in a subsequent liquid biopsy^{12,21–24}. However, obtaining a biopsy is invasive, and sometimes impossible, and this two-step approach for assessing therapy response prolongs turnaround time. Moreover, tumor-informed assays track a fixed set of somatic mutations, whose abundance may not accurately reflect the tumor composition especially when the tumor clonal heterogeneity changes under pressure from a therapy, as is often true in the case of late-stage cancers²⁵. Sequencing the cfDNA more broadly with approaches like whole exome sequencing or whole genome sequencing would likely detect more somatic mutations without requiring a tumor sample. However, sampling more genomic space results in either increased false positive mutations or decreased sensitivity, while requiring more sequencing reads, and as such, these approaches have not been widely adopted for therapy response monitoring⁷.

Given that precision is often limited by the Poisson sampling noise associated with assaying somatic mutations in practice, methylation of ctDNA has been explored as an alternative biomarker to VAF measurements. Generally, methylation has been shown to be a strong, consistent, and genomically widespread biomarker for cancer that can also be detected in ctDNA^{26,27}. While methylation cancer signals are more abundant compared to somatic mutations, quantifying the amount of methylation accurately and precisely is a challenge. Methylation-specific qPCR has been shown to be correlated with clinical outcomes^{28–30}; however, because of its exponential nature, qPCR has high assay noise when calculating the absolute number of molecules, typically in the tens of percent³¹. Sampling hundreds of genomic locations for methylation, instead of just one or two, would significantly improve the assay performance, but interrogating multiple loci using qPCR is technically infeasible without obtaining additional cfDNA, i.e. additional tubes of blood for each locus, from the patient. Current DNA methylation sequencing methods also face challenges with precise quantification³². Therefore, a new technical approach is needed to precisely quantify ctDNA methylation while taking advantage of its abundance in the genome.

Herein, we describe Northstar Response, a novel assay that uses molecule counting to accurately and precisely quantify methylated ctDNA for tumor-naïve cancer therapy response monitoring. We built a multiplex assay that targets over 500 locations in the genome that are hypermethylated in cancer compared to normal tissue and leverages quantitative counting templates (QCTs)³³ to count the number of methylated molecules at each location. We analytically validated this assay, demonstrating accurate detection of 0.25% absolute changes in tumor fraction (AUC > 0.94) and high accuracy across 12 tumor tissue types. We show that this assay is not only correlated with but also up to two times more precise than tumor-naïve, targeted-panel methods measuring VAF, achieving CVs < 10% in 1% tumor fraction samples. Finally, we show high concordance between clinical outcomes from cancer patients and Northstar Response methylation measurements.

Results

Hypermethylation locus selection and assay design

We designed an amplicon-based, multiplex PCR assay that utilizes QCTs in conjunction with next-generation sequencing (NGS) to count the number of methylated molecules at more than 500 genomic locations known to be hypermethylated in cancer tissue compared to normal tissue (Fig. 1A). Given the limited number of clinically validated hypermethylated gene targets published in the literature, we analyzed The Cancer Genome Atlas (TCGA), a vast tumor and normal tissue methylation dataset obtained from cancer patients to identify hypermethylated locations³⁴. We downloaded methylation data for subject-matched tumor and normal tissue, and tumor hypermethylation was calculated at each CpG site by subtracting normal tissue beta values from tumor beta values, where beta represents the ratio of the methylated array intensity to the total array intensity. Tumor hypermethylation was averaged across all CpG sites in the same CpG island to remove spuriously methylated CpG sites and ranked in order of highest hypermethylation across 12 different tumor tissue types (Fig. 1B). Additionally, CpG islands with an average white blood cell beta > 0.2, measured from patients of similar age to cancer patients (mean = 63.9 years, sd = 13.3 years)³⁵, were filtered out to minimize background signal from buffy coat contributions to the cfDNA.

To determine the approximate number of loci required for the assay, we performed theoretical assay noise calculations based on the number of molecules sampled by the assay. We aimed for the assay to call 10% relative changes in tumor fraction, a percent change that would be more sensitive than established RECIST criteria for progressive disease and partial response³⁶. To call these changes with high accuracy, we aimed for the assay to achieve a 3% CV, which would call 10% relative changes in tumor fraction with at least three standard deviations of accuracy (e.g. a 99.7% confidence interval). Assuming only Poisson noise from molecule sampling and no additional background noise from non-cancer methylated molecules, a 3% CV would require sampling at least 1,000 methylated molecules. A typical 10 mL blood draw has approximately 5,000 cfDNA GE, and assuming a typical 1% tumor fraction for late-stage cancer patients, 50 GE of ctDNA are expected to be present at each locus. Accounting for an 80% cfDNA extraction yield, a 50% bisulfite conversion yield, and a 50% PCR capture efficiency, 10 ctDNA GE out of 1,000 total GE should be assayable at each methylated locus. Based on analysis of TCGA data and results from testing tumor tissue with prototype assays, we estimated that at least 20% of designed loci would be methylated in most cancers (Supplementary Fig. 1). This results in a final assay design requiring about 500 loci to achieve the desired CV (500 loci * 20% methylation ratio * 10 ctDNA GE per

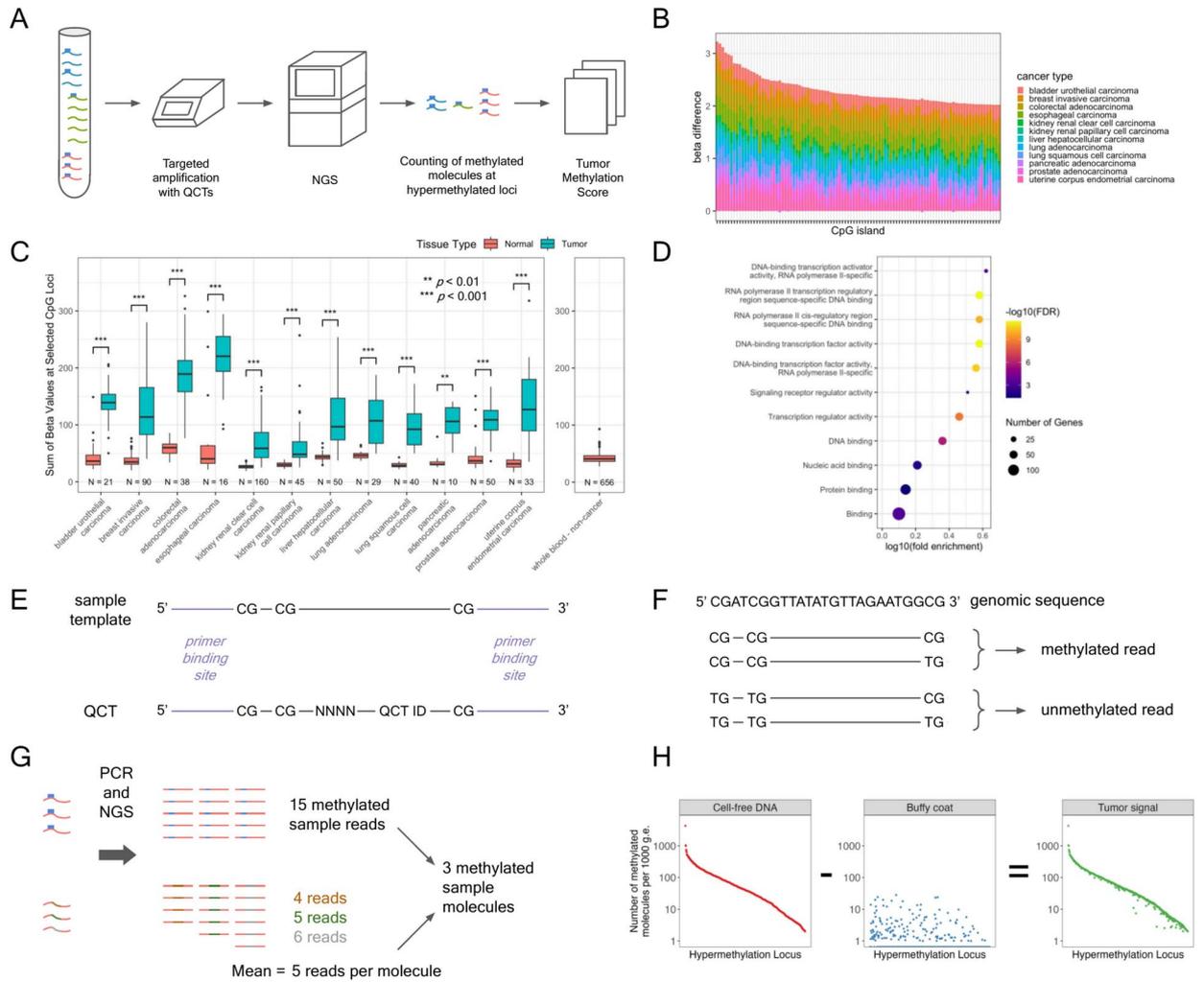


Fig. 1. Design and analytical methods for Northstar Response. **(A)** Workflow overview for Northstar Response. Cell-free DNA (cfDNA) molecules are hypermethylated (blue rectangles) in circulating tumor DNA (ctDNA) compared to normal tissue at multiple genomic locations (red, green, and blue wavy lines). These cfDNA molecules at targeted hypermethylated loci are co-amplified with Quantitative Counting Templates (QCTs) and sequenced using next-generation sequencing (NGS). This data is analyzed to calculate the number of methylated molecules at these targeted loci and aggregated across all loci to calculate the Tumor Methylation Score. **(B)** The top 100 CpG islands ranked by total differential methylation across 12 cancer types according to The Cancer Genome Atlas (TCGA) data. Beta is the fraction of DNA that is methylated, and the beta difference was calculated by subtracting the beta measured in normal tissue from the beta measured in tumor tissue and averaged across subjects. **(C)** Total methylation signal at CpG sites in the assay. Left: TCGA tumor samples have higher signal than paired normal tissue (Wilcoxon signed-rank test p value < 0.01 for all cancer types). Right: signal in whole blood samples from Hannum et al. 2013. **(D)** Functional enrichment analysis of molecular function Gene Ontology (GO) terms based on genes covered by the selected loci. All terms with false discovery rate (FDR) < 0.05 identified by STRING v12 are shown. **(E)** QCTs are designed for each targeted genomic location such that they have identical primer binding site sequences but have a number of random bases that determine an embedded molecular identifier (EMI) and a QCT identification sequence to distinguish QCTs from sample molecules. **(F)** Bisulfite converted sample reads are classified as methylated if the number of CpGs is equal to the maximum number of possible CpGs for that amplicon or the maximum number of possible CpGs minus one (in the example, 3 or 2 CpGs). Reads with fewer CpGs are classified as unmethylated. **(G)** Because QCTs are added at an abundance ensuring one molecule per EMI (brown, green, and gray), and because QCTs amplify at the same rate as sample molecules at the targeted genomic location, the average number of reads per EMI is equivalent to the number of reads per molecule at that genomic location. The number of methylated sample reads can then be divided by the average number of reads per molecule to calculate the number of methylated sample molecules at the start of PCR. **(H)** The numbers of methylated molecules measured in paired cfDNA and buffy coat are first normalized to the estimated input genomic equivalents (GE), then background methylation detected in buffy coat is subtracted from the methylation measured in plasma on a per-locus basis to extract tumor-associated signal.

locus = 1,000 molecules, resulting in a CV of $\sqrt{1,000/1,000} = 3\%$). We note that our experimental results validate the assumptions in this calculation, as we measured roughly 1,000 methylated molecules per 1,000 input GE across different cancer types at 1% tumor fraction (Supplementary Fig. 2A).

The top 300 CpG islands contained more than 2,000 individual CpG sites, which were input into a primer design pipeline to design 70–90 bp amplicons containing at least 3 CpG sites. We empirically removed amplicons that caused significant primer dimers and failed to amplify sufficiently, resulting in a multiplex assay targeting more than 500 total amplicons. Based on TCGA data, the targeted CpG sites are hypermethylated in multiple cancer types as opposed to just a few cancer types (Fig. 1B). Simulating measurements of individual patients' tumor and normal tissue based on the beta values from TCGA at CpG sites covered in the assay showed significantly increased methylation in cancer compared to corresponding normal tissue in all cancer types tested (Wilcoxon signed-rank test p value < 0.01 for all comparisons; Fig. 1C). Additionally, the total methylation signal in simulated measurements of whole blood samples from non-cancer subjects was comparable to the signal measured in other normal tissues (Fig. 1C).

The genomic locations covered in the assay have been implicated in diverse aspects of known cancer biology. Our literature-uninformed approach yielded several target locations in genes that have been clinically validated as cancer biomarkers; *IKZF1* is a hypermethylation target for colorectal cancer³⁷, and *ITGA4* and *PAX1* are hypermethylation targets for cervical cancer in separate assays³⁸. Furthermore, functional annotation of genes associated with the selected loci revealed significant enrichment for genes encoding proteins with DNA-binding transcription factor activity (Benjamini–Hochberg adjusted p value = 4.09×10^{-12} ; Fig. 1D), a molecular function known to be altered in cancer cells^{39–41}. In particular, the assay targets 20 of the 236 known homeobox genes, a significant enrichment of the homeobox transcription factor superfamily involved in many developmental processes and of which several members are involved in oncogenicity (adjusted p value = 5.11×10^{-11} ; Supplementary Fig. 3C)⁴². More broadly, the selected loci are significantly enriched for genes involved in biological processes related to development, morphogenesis, and cell differentiation, which may be dysregulated in cancer cells (Supplementary Fig. 3A)⁴³.

QCTs for precise quantification, background subtraction, and normalization

With an NGS approach, quantifying methylation based on the number of reads is technically challenging due to amplification variability across loci. Therefore, we utilized QCTs to calculate the number of input methylated molecules at each locus, effectively accounting for amplification bias without requiring any reference standards. For each target amplicon, a QCT was designed containing identical forward and reverse primer binding sites such that QCTs would amplify similarly to sample molecules in the PCR (Fig. 1E). Each QCT design also contains an embedded molecular identifier (EMI) composed of eleven random nucleotides. This results in 4^{11} possible EMI sequences for each QCT; adding on the order of 100 QCT molecules into the PCR probabilistically guarantees the uniqueness of each QCT molecule. One can then estimate the reads per sample molecule for the corresponding amplicon based on the average reads per QCT molecule. Reads that map to each amplicon are classified as methylated based on the number of CpGs contained in each read (Fig. 1F), and the total number of methylated reads can then be divided by the reads per molecule to calculate the number of methylated sample molecules at this amplicon prior to PCR (Fig. 1G).

As a significant portion of cfDNA is of white blood cell origin^{44,45}, we can account for non-cancer background methylation by subtracting the number of methylated molecules measured in the buffy coat from that measured in the paired cfDNA on a per-locus basis (Fig. 1H). This background subtraction step could also minimize spurious methylation signals resulting from effects that are non-specific to the cancer, such as the side effects of systemic therapy.

To perform this background subtraction accurately, the total number of molecules must be normalized to account for variable input amounts between cfDNA and buffy coat samples. This also allows for accurate comparisons across multiple time points for the same patient should cfDNA concentrations or plasma volumes vary significantly. To do this, we designed amplicons and corresponding QCTs that target CpG sites that are highly methylated in all tissues (e.g., buffy coat, tumor, and normal tissue) to estimate the total amount of input DNA. These normalization loci were selected using the same methylation data sources: CpG sites were ranked in order of highest white blood cell methylation signal and filtered for average beta > 0.9 across CpG islands for white blood cell, tumor tissue, and normal tissue for six tumor tissue types. The number of methylated molecules measured at these normalization loci were averaged to estimate the input amount of DNA.

After quantifying the number of methylated molecules, normalizing to an input of 1,000 GE of DNA, and subtracting background signal from the cfDNA on a per-locus basis, these normalized molecule counts are summed across loci to calculate the Tumor Methylation Score for a given sample.

Quantification performance of Tumor Methylation Score in contrived samples

To assess the analytical performance of the Northstar Response assay, we tested technical replicates of contrived samples made from sheared tumor DNA spiked into sheared buffy coat DNA at 0%, 0.25%, 0.5%, 1%, and 2% tumor fraction by mass. Two sets of contrived samples were made from two tumors, one from a stage IV lung adenocarcinoma and another from a stage IV breast adenocarcinoma, and buffy coats from the same subject. By performing measurements on contrived samples with known relative amounts of tumor DNA, we determined the quantification accuracy of Tumor Methylation Score within each specific tumor.

For both tumors, the Tumor Methylation Score linearly increased with contrived tumor fraction, accurately quantifying the amount of methylated tumor DNA in contrived samples (breast $R^2 = 0.998$, lung $R^2 = 0.996$) (Fig. 2A). Notably, this relationship is directly proportional, with a doubling in contrived tumor fraction closely corresponding to a doubling in Tumor Methylation Score, and remains linear down to 0.25% tumor fraction. When the background subtraction used to calculate Tumor Methylation Score is not performed, this

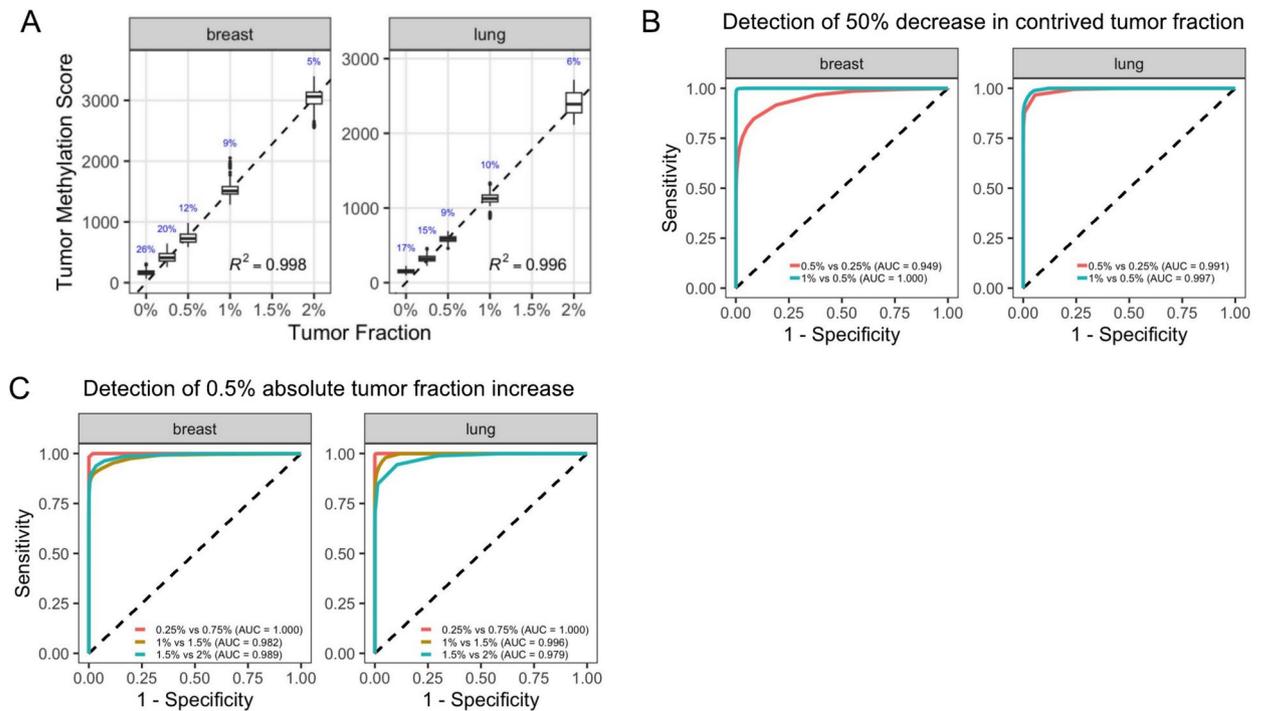


Fig. 2. Northstar Response accurately and precisely quantifies methylated tumor DNA. **(A)** Two tumors were used to generate technical replicates of contrived samples at different tumor fractions by mass, assuming that the tumor was 100% cancerous tissue (total input mass = 30 ng). The replicates were processed independently through the assay, and Tumor Methylation Score was calculated for every possible combination of contrived sample and buffy coat. The boxplots show the median and interquartile range in Tumor Methylation Score for each contrived tumor fraction. The dashed lines indicate a linear fit through the origin and the mean Tumor Methylation Score for each contrived tumor fraction. Percents printed in blue indicate coefficients of variation at each condition. **(B)** Receiver operating characteristic (ROC) curves for calling changes in tumor fraction for sample pairs with either a 50% decrease or no change in contrived tumor fraction based on the measured Tumor Methylation Scores. **(C)** ROC curves for calling changes in tumor fraction for sample pairs with either an increase of 0.5% contrived tumor fraction or no change in tumor fraction based on the measured Tumor Methylation Scores. Data for 0.75% and 1.5% tumor fraction were obtained by appropriately scaling down the signal from 1 and 2% replicates, respectively.

proportionality is lost due to background signal interference (Supplementary Fig. 4D). Additionally, the CV of the Tumor Methylation Score was <20% for all contrived tumor fractions tested, decreasing to about 6% at 2% tumor fraction (Fig. 2A). These results demonstrate that Northstar Response quantifies tumor DNA with high accuracy and precision.

Classification of changes in tumor fraction using Tumor Methylation Score

Because the assay is intended to monitor changes in methylated ctDNA abundance over time, we evaluated the assay's ability to distinguish between different contrived tumor fractions. Based on the measured Tumor Methylation Scores of a given pair of samples, a call was made as to whether there was a change in contrived tumor fraction from the first sample to the second sample. After making calls for the change between a pair of samples using a range of thresholds, we generated receiver operating characteristic (ROC) curves by assessing the concordance between the call and the ground truth. For example, a call of "decrease" for a change from 0.5% tumor fraction to 0.25% tumor fraction was classified as a true positive result, while a call of "no change" for a comparison between two samples at 0.5% tumor fraction was classified as a true negative result. Because the assay's ability to accurately call changes between samples is expected to depend on both the tumor fraction at each collection and the magnitude of the change in tumor fraction, ROC curves were generated for a variety of different comparisons. Overall, the assay was able to call changes in Tumor Methylation Score with high sensitivity and specificity (Fig. 2B, 2C).

Previous studies have shown that molecular response to immunotherapy, defined as a >50% decrease in mean VAF, is associated with favorable clinical outcomes^{8,13,15}. Starting at a contrived tumor fraction of 1%, the assay was able to distinguish whether there was a decrease to 0.5% tumor fraction or no change in tumor fraction with accuracy approaching 100% (Fig. 2B). For a 50% decrease from 0.5% to 0.25% contrived tumor fraction, the area under the ROC curve (AUC) was 0.949 for the breast tumor and 0.991 for the lung tumor. The assay performed similarly for a doubling in contrived tumor fraction, making correct calls for tumor fraction increases as small as from 0.25% to 0.5% tumor fraction (breast AUC = 0.918, lung AUC = 0.981) (Supplementary Fig. 4A).

While large reductions in measured ctDNA are most strongly associated with improved survival, capturing smaller changes in methylated ctDNA levels may provide additional clinical utility by enabling earlier detection of progression. The assay made accurate calls for step increases of 0.5% contrived tumor fraction, with $AUC > 0.97$ for all comparisons tested (Fig. 2C). Although distinguishing between samples with smaller step increases of 0.25% tumor fraction was more challenging, particularly as the relative change between the samples becomes small, the assay could still distinguish between 1.75% tumor and 2% contrived tumor fraction (breast $AUC = 0.818$, lung $AUC = 0.727$) (Supplementary Fig. 4B). As for the smallest contrived tumor fraction that could be reliably distinguished from 0% tumor samples, we looked at the rate of increase calls for non-zero tumor fraction samples. At 0.25% tumor fraction, the assay achieved 92.4% sensitivity for the breast tumor and 95.2% sensitivity for the lung tumor. Variability in the performance between the two tumors is likely due to differences in the level of background signal found in the paired buffy coat (Supplementary Fig. 4C). For the comparison between 0% and 0.5% tumor samples, the assay made calls with 100% sensitivity for both tumors ($AUC = 1.000$).

Analytical validation in clinical plasma samples

While the above results demonstrate that Northstar Response accurately quantifies the amount of tumor DNA in contrived samples, we also wanted to understand how Tumor Methylation Score measurements reflect the amount of ctDNA in clinical plasma samples. We leveraged a late-stage lung cancer patient cohort with paired plasma samples collected from the same blood draw prior to the start of a new treatment. We processed a plasma sample each with both Northstar Response and a tumor-naive, targeted-panel assay to obtain an orthogonal measurement of ctDNA abundance via VAF. Paired buffy coat samples from each patient were also processed with both Northstar Response and the tumor-naive, targeted-panel assay, enabling both subtraction of background methylation signal and removal of somatic mutations arising due to clonal hematopoiesis of indeterminate potential (CHIP), respectively. Additional demographic information for this cohort can be found in Supplementary Table 1.

Within this patient cohort of late-stage lung cancer patients ($N = 60$ samples from 51 patients), there is a positive correlation between Tumor Methylation Score and mean VAF ($R^2 = 0.647$, log-log regression) (Fig. 3A) as well as maximum VAF ($R^2 = 0.718$, log-log regression) (Supplementary Fig. 5A). Background subtraction removes a greater proportion of total methylation for samples with lower Tumor Methylation Scores (Supplementary Fig. 6).

We also investigated whether changes in ctDNA within patients were concordant between the methylation-based approach with Northstar Response and the VAF-based approach with the tumor-naive, targeted-panel assay. A total of eight patients underwent more than one treatment and had samples with non-CHIP-derived somatic mutations from more than one time point. We observed consistent trends in Tumor Methylation Score and mean and maximum VAF for 7 out of 8 patients with somatic mutations detected (Fig. 3B, Supplementary Fig. 5B).

Furthermore, we assessed whether the cfDNA methylated molecules quantified in the assay were of tumor origin by running subject-matched cfDNA, tumor, and buffy coat samples on the assay. In samples with Tumor Methylation Score greater than 1000, we found that the methylated loci were highly conserved between cfDNA and tumor samples in both their identity and their relative molecular abundance, and that the corresponding buffy coats did not contribute significant amounts of methylation (Supplementary Fig. 7A). In addition, the

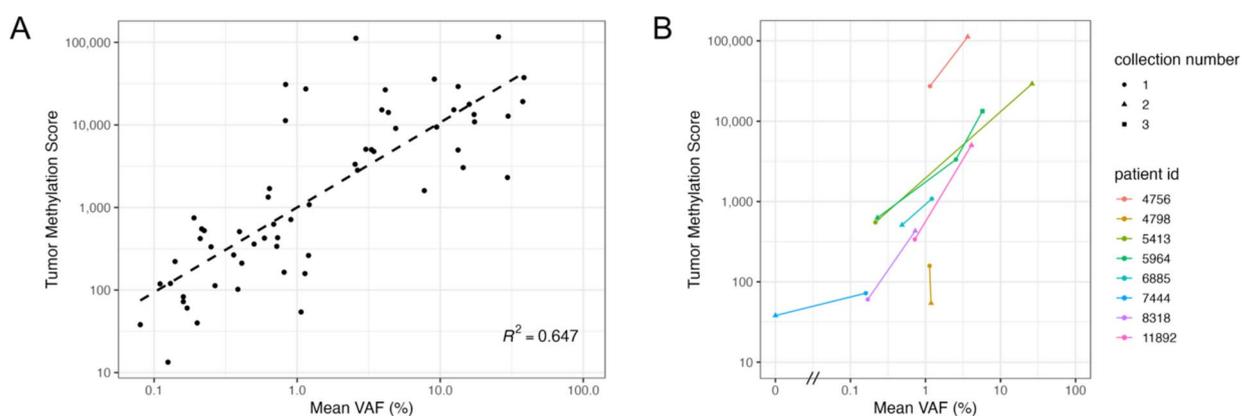


Fig. 3. Correlation of Tumor Methylation Score with mean variant allele fraction (VAF) in clinical samples from metastatic lung cancer patients. **(A)** Correlation of mean VAF among clinically significant variants as measured by a tumor-naive, targeted-panel assay with Tumor Methylation Score in clinical samples ($N = 60$ samples from 51 patients). Mutations of clonal hematopoiesis of indeterminate potential (CHIP) were excluded based on results from testing the matched buffy coat sample. A log-log fit and corresponding R^2 was calculated using the logarithm of mean VAF and the logarithm of Tumor Methylation Score. **(B)** Correlation of Tumor Methylation Score with mean VAF within patients for therapy response monitoring ($N = 8$ patients). CHIP mutations were similarly excluded. Only variants present in the first collection time point were included for calculating the mean VAF to avoid considering clonal evolution or the emergence of new variants.

cfDNA and tumor methylation profiles cluster closer within subjects than between subjects (Supplementary Fig. 7B).

Comparing precision with a tumor-naive, targeted-panel assay using VAF

To evaluate how the precision of the Northstar Response assay compares to typically performed tumor-naive VAF-based monitoring approaches, we compared the CV of Tumor Methylation Score to the CV of VAF measured using the same tumor-naive, targeted-panel assay. To generate replicate samples with many mutations at a wide range of VAFs, DNA from multiple tumors containing a total of 29 mutations was diluted into healthy background DNA at 10% and 20% tumor material by mass, with 20 replicates of each condition. Average VAF was calculated for each mutation across the 20 replicates and was used to filter and group the mutations into bins centered at 0.25%, 0.5%, 1%, and 2% tumor fraction. Separately considering the mutations by tumor fraction bin allows us to treat each bin as its own set of contrived samples representative of a different ctDNA level.

As measured by one commonly used tumor-naive, targeted-panel assay, more than half of samples in advanced cancer patients have four mutations or fewer^{13,16}. Therefore, we analyzed all possible combinations of one, two, three, and four mutations from each tumor fraction bin. For each combination of mutations, we calculated the average VAF across the mutations for each replicate, and then calculated the CV of these average VAFs (Fig. 4). In this way, we mimicked scenarios where either one, two, three, or four mutations were detected in a patient cfDNA sample and the VAFs for those mutations were then averaged for therapy response monitoring. To validate the results from these contrived replicate samples, a Poisson sampling-based simulation was performed assuming mutations at 0.25%, 0.5%, 1%, and 2% VAF to calculate the theoretical CV (Supplementary Fig. 8A). Finally, the results from both the technical replicates and the simulation were compared to the CV of Tumor Methylation Score in replicates of contrived samples at matching contrived tumor fractions of 0.25%, 0.5%, 1%, and 2% derived from the lung and breast tumor (Fig. 4, Supplementary Fig. 8A).

The CV of Tumor Methylation Score for the lung tumor contrived samples was lower than the median CV and simulated CV of mean VAF at all tested VAF tumor fraction bins (Fig. 4, Supplementary Fig. 8A). In fact, Northstar Response achieved CVs twice as low as that measured using average VAFs with the tumor-naive, targeted-panel assay, particularly when few total mutations are measured. Even if we assume no additional noise beyond Poisson sampling noise (e.g., no PCR or sequencing error, perfect molecular capture, and no clonal heterogeneity), an approach using average VAF would need to detect at least 10 mutations to match the precision of Northstar Response for a 2% tumor fraction sample (Supplementary Fig. 8B).

Analytical validation in several different tissues of tumor origin

While Northstar Response accurately and precisely quantified methylated tumor DNA from two tumors, we also tested its performance across a more expansive set of tumors. Using 54 solid tumors originating from 12 different tissues, we tested contrived samples that were made by combining sheared tumor DNA and sheared buffy coat DNA from the same subjects at 0%, 1%, and 2% tumor fraction. To account for the variable fraction of cancerous tissue in each specimen, we calculated a tumor purity score for each tumor and scaled tumor DNA inputs accordingly when preparing contrived samples (Supplementary Fig. 9).

Tumor Methylation Scores at 1% and 2% tumor fraction were calculated for each tumor by subtracting background methylation detected in 0% tumor fraction samples. The Tumor Methylation Scores for all 2%

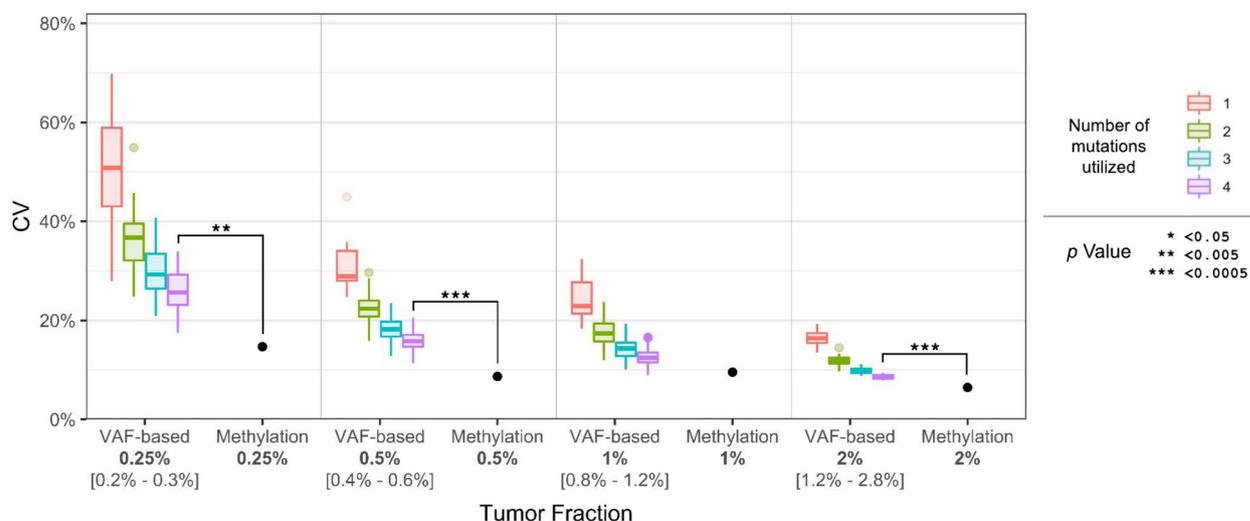


Fig. 4. Coefficients of variation (CV) for VAF-based and methylation-based therapy response monitoring. VAF-based CVs were calculated from average VAFs of different numbers of mutations from replicate contrived tumor samples (box plots). CV of Tumor Methylation Scores were calculated in contrived lung cancer samples at matching contrived tumor fractions (black points). *p* values for each tumor fraction were calculated comparing the Tumor Methylation Score CV to the distribution of VAF CVs for 4 mutations utilized using a 2-sided T test (black asterisks).

tumor fraction samples were $1.65\times$ – $2.35\times$ greater than those of the 1% tumor fraction samples (Fig. 5), except for one lung neuroendocrine tumor for which we measured little difference in signal likely due to low tumor purity (Supplementary Table 2). Quantifying the fold change in methylation without accounting for background signal detected in buffy coat resulted in less accurate and more variable measurements (Supplementary Fig. 10), which highlights the utility of subtracting background methylation when calculating Tumor Methylation Scores.

The 2% tumor samples were then called as having increased, unchanged, or decreased Tumor Methylation Scores relative to 1% tumor samples using the same calling method described earlier. We correctly made increase calls for 53/54 sample pairs, which corresponds to 98.1% [95% confidence interval: 90.11%, 99.95%] sensitivity across tissue types. While we observed variability in absolute Tumor Methylation Score between patients and across tissue types measured at a given contrived tumor fraction (Supplementary Fig. 2A), the assay was able to accurately call relative increases from 1 to 2% contrived tumor fraction within individual patients. The assay has similar performance when calling decreases between 2 and 1% tumor samples (Supplementary Fig. 11).

Comparing clinical results with outcomes

To investigate whether the Northstar Response assay could measure clinically meaningful signals, we tested cancer patient samples collected before and after treatment. Tumor Methylation Scores were determined at each sample collection time point, and calls were made by comparing the Tumor Methylation Score with the immediately preceding measurement. These longitudinal results were compared with clinic-reported clinical outcomes that were determined from imaging scans. A total of seven patients had at least three sample collection time points with at least one corresponding clinical outcome within 100 days of any sample collection time point. These patients all had advanced stage cancer, and their cancers spanned three different tissues of origin: lung, pancreas, and colorectal cancers. Additional demographic information can be found in Supplementary Table 3.

We observed changes in Tumor Methylation Scores that reflected the dynamics of cancer treatment (Fig. 6). All seven patients had an initial decrease in Tumor Methylation Score after starting treatment (Fig. 6H). Some of these decreases were large, with one patient's Tumor Methylation Score decreasing by almost 80% within 7 days (Fig. 6F) and several patients' Tumor Methylation Scores decreasing by more than 90% compared to pre-treatment levels (Fig. 6A–E, 6G). We also observed large increases in Tumor Methylation Score later during treatment, with some patients increasing between fourfold and 160-fold (Fig. 6A–E). Changes as small as twofold were also observed (Fig. 6A, 6F), and sometimes, the Tumor Methylation Score remained below the calling threshold in consecutive time points (Fig. 6E, 6G).

Tumor Methylation Scores showed high concordance with clinical outcomes. Of the six patients that displayed an increase in Tumor Methylation Score, all six had progression noted as a clinical outcome shortly before or shortly after the sample was collected (Fig. 6H). In three of those six patients, an increase in Tumor Methylation Score was measured before clinical progression was noted (Fig. 6D, 6E, 6F). In addition, when a partial response clinical outcome was noted, the Tumor Methylation Score had decreased compared to the pre-treatment time point (Fig. 6A, Fig. 6G). In one of these patients, Tumor Methylation Scores had decreased by over tenfold 81 days prior to a reported partial response clinical outcome (Fig. 6A).

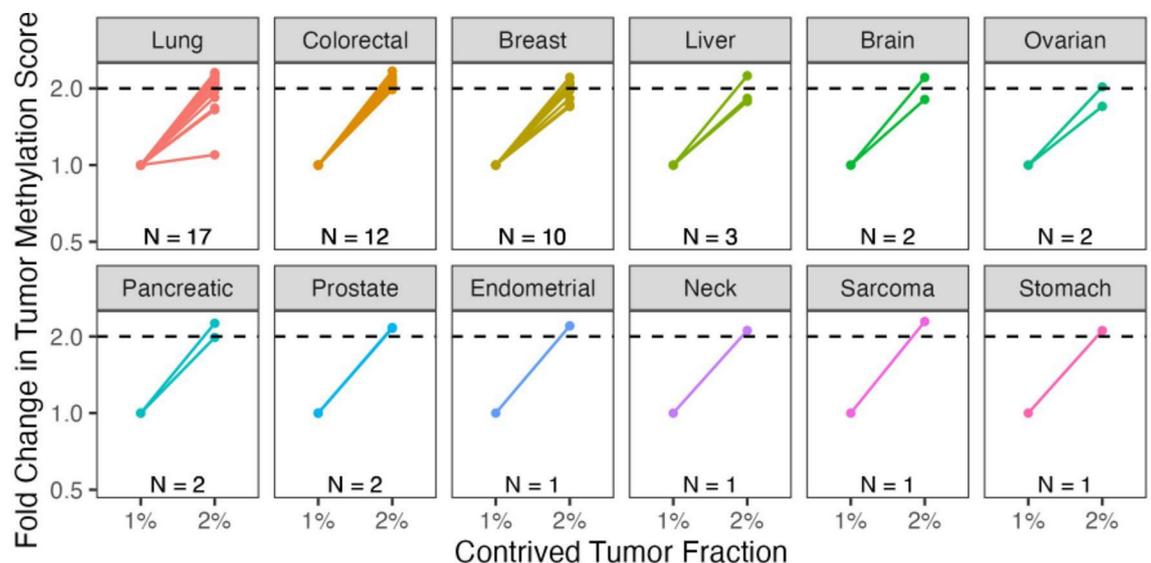


Fig. 5. Northstar Response accurately measures increases in methylation for many cancer types in contrived samples. Tumor Methylation Scores were significantly increased at 2% tumor fraction relative to 1% tumor fraction for 53/54 different tumor specimens across 12 different tissues of origin. The dashed line indicates the expected fold change of 2x.

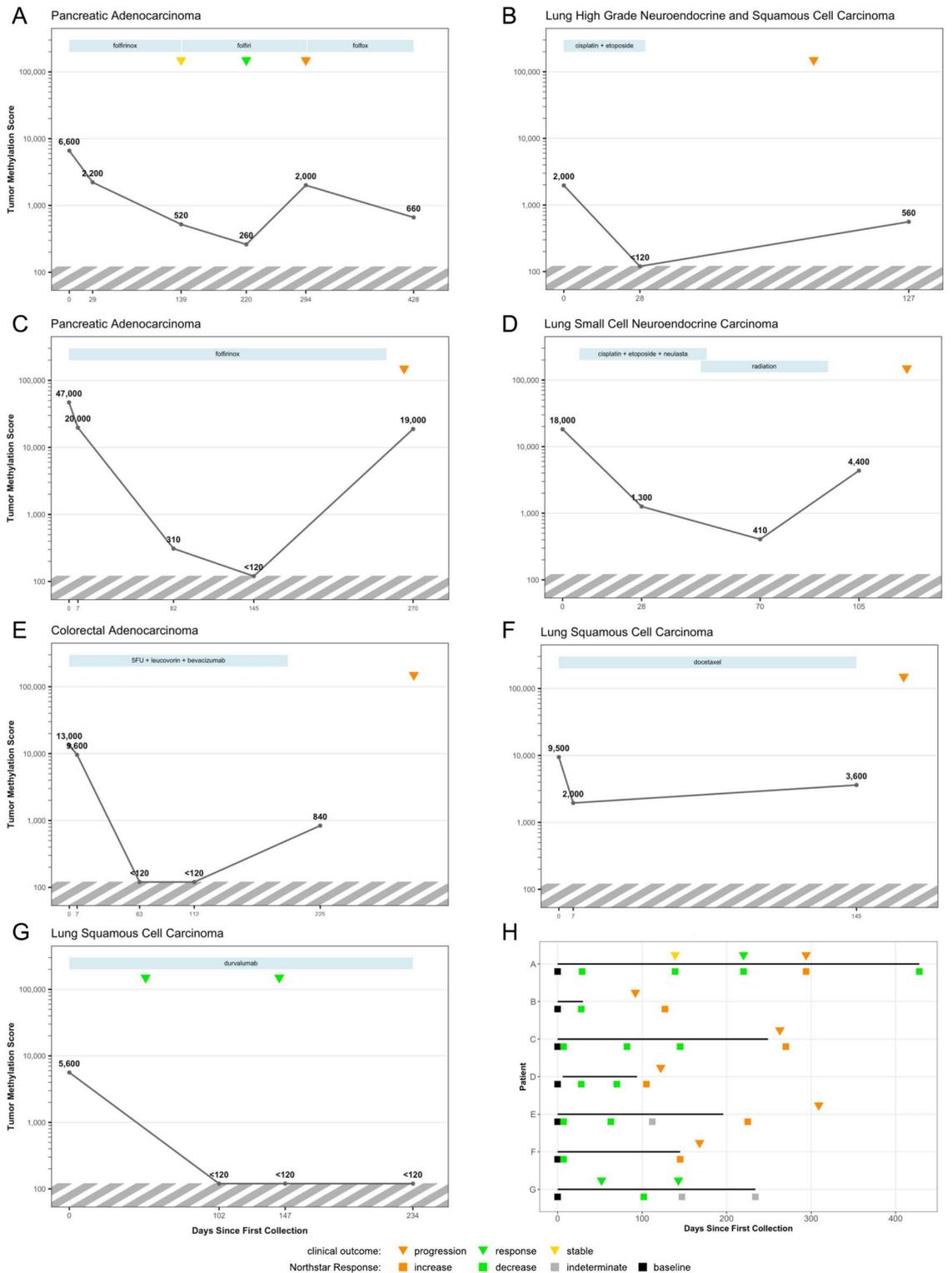


Fig. 6. Longitudinal clinical results from Northstar Response. (A–G) Tumor Methylation Score measurements from seven cancer patients, annotated with clinical outcomes indicated by triangles (orange = progression, green = response, yellow = stable) as reported by the clinic. Blue rectangles plotted at the top indicate the duration of each treatment regimen. The striped region on each plot indicates the calling threshold; any Tumor Methylation Scores below 120 are reported as <120. (H) Swimmer plot summarizing Tumor Methylation Score and clinical outcomes for all seven cancer patients. Squares indicate calls made by Northstar Response based on changes in Tumor Methylation Score compared to the previous sample point (orange = increase, green = decrease, gray = indeterminate, black = baseline measurement). Segments indicate the duration of treatment.

Discussion

Therapy response monitoring through liquid biopsy has been increasingly shown to be clinically useful for late-stage cancer patients. However, current liquid biopsy approaches for therapy response monitoring face several technical challenges that limit both the accuracy and precision of those measurements. In this study, we describe Northstar Response, a novel, tumor-naïve, methylation-based assay that overcomes these challenges.

Because liquid biopsy requires quantification of ctDNA within a background of healthy cfDNA, the assay was designed to maximize tumor-specific hypermethylation signal while minimizing background from normal tissues. Our analysis of aggregate methylation metrics yielded a set of over 500 target loci, which collectively measure significantly higher signal in tumor than normal tissue on a per-sample basis. Although we did not select loci based on their biological significance, our approach to assay design resulted in significant enrichment for genes involved in transcriptional regulation, cell differentiation, and developmental processes, all of which are known to be dysregulated in cancer. Because the selected loci were not restricted to promoter regions, it is possible that the assay captures hypermethylation sites with roles in cancer biology beyond gene silencing via promoter methylation. Recent work suggests that gene-body hypermethylation can play a role in activation of certain genes, including oncogenic homeobox genes, in cancer cells⁴⁶. Interestingly, roughly 26% of the selected loci were not associated with particular genes, suggesting that methylation at these locations is associated with cancer or causal of cancer through indirect regulatory mechanisms. Future research may reveal additional relationships between cancer biology and tumor-specific hypermethylation at the sites included in this assay.

The assay is highly accurate and precise based on extensive testing of technical replicates of contrived samples. Tumor Methylation Scores are linear with contrived tumor fraction down to 0.25% tumor fraction, and that linearity is enabled by background subtraction of subject-matched buffy coat signal. At 2% contrived tumor fraction, the CV of the measured Tumor Methylation Score is less than 6.5%. This level of precision enables the detection of 20% changes in Tumor Methylation Score with 99.7% confidence (an interval of 3 standard deviations), meaning that the assay could accurately call an increase in ctDNA burden from 2% to 2.4% tumor fraction or a decrease from 2% to 1.6% tumor fraction. It is worth noting that contrived samples composed of sheared genomic DNA do not completely reflect the complex biology of cfDNA (e.g. non-random fragmentation pattern due to open chromatin regions and active promoters), so these results from contrived samples can be taken as the upper end of assay performance. While the clinical significance of small changes in ctDNA burden remains to be elucidated within the field of liquid biopsy, the ability to detect these changes may provide earlier insight into a patient's response to therapy compared to traditional imaging methods.

To further characterize Northstar Response's ability to quantify ctDNA in clinical samples, we processed patient samples using both Northstar Response and an orthogonal tumor-naïve, targeted-panel assay measuring VAF. We found that the Tumor Methylation Score was correlated with VAF across patients and measured concordant trends in ctDNA dynamics across more than one time point from the same patient. The absolute level of Tumor Methylation Score may have clinical significance, given that we found Tumor Methylation Score to be correlated with VAF, and that several studies have shown the absolute level of the VAF to be prognostic of treatment outcomes^{15,47}. Further studies in additional cancer types are needed to establish the robustness of this correlation. However, while Tumor Methylation Score has a good correlation with VAF, the correlation is not robust enough to convert Tumor Methylation Score into a tumor fraction metric. One reason could be that while somatic mutations and hypermethylation are both cancer-specific, they do not originate from the same cancerous biological mechanisms. In addition, not all cancer cells from a given patient will contain a specific somatic mutation or be equally hypermethylated at a given locus. Nevertheless, the relatively strong concordance between Tumor Methylation Score and VAF, coupled with the high analytical sensitivity and specificity on contrived samples, leads us to believe that Northstar Response can accurately quantify ctDNA in clinical samples for therapy response monitoring.

Today, the field commonly uses VAF measurements from liquid biopsy assays to track the relative abundance of somatic mutations for therapy response monitoring. However, the precision of VAF metrics is often limited by Poisson counting noise due to the small number of tumor molecules, making these assays poorly suited for therapy response monitoring in those situations. In one analytical validation study using a tumor-naïve, targeted-panel approach performed on synthetic DNA, two-fold changes could not be reliably distinguished below 0.8% VAF²⁰. In another analytical validation study using a tumor-informed approach, the assay had quantitative precision greater than 15% at an intended 1.60% VAF, in line with our own analysis of the precision of a VAF-based approach⁴⁸. This lack of precision in VAF measurements when counting noise is significant and could misclassify patients as molecular responders / non-responders, such as when predicting clinical outcomes for immunotherapy. We found that Northstar Response consistently achieved lower CV and thus greater precision compared to these VAF measurements when only several mutations are present. One reason for this improved performance is that the abundance of informative methylated loci is much greater than the typical number of somatic mutations detected in a tumor-naïve assay (Supplementary Fig. 2B), which significantly reduces the Poisson sampling noise of the assay. While we did not test contrived tumor fractions less than 0.25% in this study, Northstar Response should theoretically have superior CV at even smaller tumor fractions, which is particularly relevant because a significant percent of somatic mutations in late-stage cancer patients have VAF below 0.2%¹⁹. The improved precision allows for the detection of smaller changes in the abundance of ctDNA, which could translate to measuring signs of tumor progression, or response, earlier.

Northstar Response maintained accuracy and sensitivity across a diversity of tumors and patients, suggesting that methylation-based therapy response monitoring has the potential to benefit a broad patient population. The assay correctly detected increases from 1 to 2% tumor fraction with 98.1% sensitivity for 54 solid tumors from 12 distinct tissue types. Despite the heterogeneity often observed in cancer biology, this implies that sampling a panel of hypermethylated loci enables tumor-naïve tracking of tumor progression in a variety of tissue types. In

the future, additional tumor samples from untested tissues of origin should be tested to further validate the assay as an effective therapy response monitoring test for all solid tumors.

To further assess how Northstar Response performs in clinical samples, we performed the assay on longitudinally collected samples from cancer patients, and we found that Tumor Methylation Scores correlated strongly with clinical outcomes in both increasing and decreasing directions. If the clinical accuracy of the assay holds beyond this small initial cohort of patients with known clinical outcomes, we foresee potential clinical utility for detecting molecular progression and molecular response. It is possible that clinicians could take action based on these findings, for example, accelerating the timing of the next imaging scan and adjusting treatment regimens in light of molecular progression, or determining the efficacy of therapies like immunotherapy^{8,11,13,15,24,49,50} and targeted therapy^{9,14} at an earlier stage of treatment. Aside from clinical accuracy, it remains undetermined whether the unprecedented precision of the assay could enable the accurate detection of smaller yet still meaningful molecular progression or response. Further clinical utility studies with larger cohort size are needed to confirm the clinical accuracy of the assay as well as establish Tumor Methylation Score as a clinically actionable biomarker^{51,52}. Overall, Northstar Response is a significant technical advance that may improve therapy response monitoring and enhance cancer patient care.

Methods

Target selection

To choose targets that are hypermethylated in tumors, we queried The Cancer Genome Atlas (TCGA) for subjects with methylation data from matched tumor and normal tissue of the following cancer types: bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), pancreatic adenocarcinoma (PAAD), prostate adenocarcinoma (PRAD), and uterine corpus endometrial carcinoma (UCEC)³⁴. To choose targets for normalization loci that are highly methylated in buffy coat, tumor, and normal tissue, methylation data was obtained and analyzed from six tumor tissue types: breast invasive carcinoma, colon adenocarcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, and pancreatic adenocarcinoma.

To verify that the selected targets collectively captured more methylation signal in tumor tissues than normal tissues, we simulated measurements from individual samples by taking the sum of beta values at the selected targets. This analysis was performed with paired samples of tumor and normal tissue from the TCGA database and individual samples of whole blood from Hannum et al. 2013, both of which were processed with the Illumina Infinium HumanMethylation450 BeadChip^{34,35}. Additionally, functional enrichment analysis of the gene set corresponding to selected CpG sites was performed using STRING v12⁵³.

Oligo design and preparation

Primer3 (<https://primer3.org/>) was used to design primer pairs targeting the genomic locations of chosen CpG locations and using a reference genome that was bisulfite converted in silico. Each amplicon was required to contain at least 3 CpG sites, and amplicon length was constrained between 70–90 bp (average length = 82.2 bp). This process was performed assuming full methylation of the reference hg19 human genome, converting all Cs to Ts except for CpGs. The ideal melting temperature was set to 60C. Primers were synthesized by Eurofins Genomics (Louisville, KY).

Primer mixes were created by pooling all primer pairs and iteratively removing and/or rebalancing the concentrations of each primer pair to optimize for balanced read depth across target amplicons. Two iterations of the primer mix were prepared, version v0 and version v1.1. These two versions share a core set of 494 target hypermethylation amplicons. Between the initial version and the second version, 60 hypermethylation amplicons were removed due to poor amplification or high background signal, and 26 new hypermethylation amplicons were added. This resulted in an initial version with 590 total amplicons (554 hypermethylation), and a second version with 551 total amplicons (520 hypermethylation). The v0 version of the assay was used for all clinical samples presented in the main figures, and the v1.1 version of the assay was used for all other results. The two versions of the assay achieve comparable CV in the lung contrived samples (Supplementary Fig. 12) and have consistent Tumor Methylation Score trends in clinical samples run on both assays (Supplementary Fig. 13).

Single stranded Quantitative Counting Templates (QCTs)³³ were designed for each amplicon. The predicted amplicon sequence for each primer pair was determined using Bowtie and the converted human genome, and 17 bases of flanking genomic sequence were added to both 5' and 3' ends. Eleven bases in the insert region of the QCT were replaced with Ns, allowing for a small number of unique QCTs to be added to each PCR reaction. QCTs were synthesized by either Eurofins Genomics (Louisville, KY) or IDT (Coralville, IA). QCTs were diluted to 200 molecules per PCR reaction at each amplicon.

Specimen sourcing and processing

To obtain tumor specimens spanning a variety of tissue types, banked flash frozen tumors and buffy coats from the same subjects were obtained from Corewell Health SHARE Biorepository (Grand Rapids, MI) on behalf of Accio Biobank Online (UK). This study was approved by the Spectrum Health Institutional Review Board, and all the methods were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from patients for all specimens according to the IRB.

To compare the performance of Northstar Response with an orthogonal tumor-naive, targeted-panel assay for measuring VAF, banked plasma and matched buffy coat from late-stage lung cancer patients were processed through both assays. Whole blood was collected in K2 EDTA tubes and processed immediately or within 2 h after storage at 4C. Plasma and cellular components were separated by centrifugation at 800 g for 10 min at 4C.

Plasma was centrifuged a second time at 18,000 g at room temperature to remove any remaining cellular debris. Plasma and buffy coat samples were stored at -80°C until the time of DNA extraction. This study was approved by the University of California San Diego Institutional Review Board, and all the methods were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from patients for all specimens according to the IRB.

To test whether the assay could detect changes in methylation that were concordant with clinical outcomes, samples were prospectively collected from cancer patients at Ochsner Health (New Orleans, LA) and Christus Health (Irving, TX) on behalf of Accio Biobank Online (UK) and iSpecimen (Lexington, MA). Patients that were diagnosed with cancer and had not started treatment were enrolled. Up to 25 mL of blood was collected in Streck tubes (La Vista, NE) pre-treatment and at subsequent time points post-treatment. Blood tubes were then shipped from the clinic to BillionToOne's laboratory. Clinical outcomes were provided when available. This study was approved by the Ochsner Clinic Foundation Institutional Review Board and Christus Health Institutional Review Board, and all the methods were performed in accordance with relevant guidelines and regulations. Informed consent was obtained from patients for all specimens according to the IRBs.

Plasma and buffy coat were isolated from Streck tubes within 3 days of collection. Plasma and cellular components were separated by centrifugation at 1,600 g for 10 min at 22°C . Plasma was centrifuged a second time at 16,000 g for 10 min without temperature control to remove any remaining cellular debris. cfDNA was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany), and genomic DNA was extracted from tumor samples and buffy coat samples using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany).

Contrived samples resembling cfDNA from cancer patients were created by mixing 30 ng total of tumor DNA and buffy coat DNA at various tumor fractions. To mimic the fragment length of cfDNA, all tumor and buffy coat DNA samples were sheared with a Covaris E220 Sonicator (Covaris, Woburn, MA) to a fragment length of ~ 175 bp prior to mixing. Replicates of contrived tumor samples from a stage IV lung adenocarcinoma and a stage IV breast adenocarcinoma were prepared by combining sheared tumor DNA and sheared buffy coat DNA samples at 0%, 0.25%, 0.5%, 1%, and 2% tumor fraction by mass, assuming that the tumor was 100% cancerous tissue ($N = 12$ 0% breast replicates; $N = 13$ 0% lung replicates; $N = 15$ 0.5% breast replicates; $N = 16$ for all other conditions). To prepare samples at 1% and 2% tumor fraction for tumors from a variety of tissue types, sheared tumor DNA and sheared buffy coat DNA samples were combined in different ratios to account for tumor purity. For instance, if we estimated a tumor-extracted DNA sample to be 20% pure, the 1% tumor fraction sample would have contained 5% tumor-extracted DNA to account for purity.

Analytical validation samples were bisulfite converted using the EZ-96 DNA Methylation-Lightning MagPrep kit (Zymo, Irvine, CA) for all analytical experiments, and clinical samples were bisulfite converted with the Premium Bisulfite kit (Diagenode, Denville, NJ) which uses identical incubation times and temperatures. When either cfDNA or genomic DNA sample volumes were larger than the recommended input volume, samples were split in half, converted separately, and then re-combined during purification steps before elution of the converted DNA.

Multiplex PCR was performed on bisulfite converted specimens using Q5U polymerase (NEB, Ipswich, MA). Subsequently, indexing PCR was performed using Q5 polymerase (NEB, Ipswich, MA) in order to sequence multiple samples on the same sequencing run with dual indexes. Pooled libraries were bead cleaned and loaded on NextSeq 2000 (Illumina, San Diego, CA) sequencing instruments using P3 100 cycle reagents for single-directional sequencing.

Bioinformatic processing

Fastq files were adapter trimmed on the 3' end using BBDuk and then mapped using BWA-MEM to a custom genome composed of the target hypermethylated and highly methylated amplicon and QCT sequences. Amplicon sequences came from the expected amplified human genome sequence assuming all cytosines were converted and all CpGs were methylated.

For each amplicon, reads that mapped to the target amplicon were classified as either methylated or unmethylated based on the number of CpGs contained in each read (Fig. 1F). Each read was classified as methylated if the number of CpGs was equal to the maximum number of possible CpGs for that amplicon or the maximum number of possible CpGs minus one. Allowing for the maximum number of CpGs minus one accounts for possible bisulfite conversion, amplification, or sequencing error in order to maximize the number of methylated molecules without significantly compromising specificity of the assay. Reads mapping to the QCTs corresponding to each amplicon were separately processed and grouped based on the sequence of random bases of the QCT. The number of methylated molecules for each amplicon is calculated using the grouped QCT reads as described in the main text. Background subtraction with the buffy coat methylation signal was performed as described in the main text, with any calculated negative numbers of molecules after subtraction capped at zero.

Based on testing healthy subjects, certain hypermethylation loci have been found to have significant methylation signal even after buffy coat subtraction. To minimize the amount of false cancer signal, certain loci were filtered out from analysis. Any loci found to either consistently contribute a moderate amount of methylation or occasionally contribute a high amount of methylation in healthy subjects were added to a list of loci to ignore. These loci often contain high amounts of methylation in buffy coat. Remaining loci with more than 2 methylated molecules are identified as informative loci for calculating Tumor Methylation Score.

After filtering hypermethylation loci, the results from healthy subjects are used to establish a calling threshold, below which the methylation signal is not interpretable because it is comparable to the amount seen in healthy subjects.

Calling changes in Tumor Methylation Score

When two time points of data are available, a call can be made as to whether there has been an increase, decrease, or no change in the Tumor Methylation Score. The call is made by modeling each time point's measurement as a normally distributed variable with a mean equal to the Tumor Methylation Score, and a standard deviation based on the CV measurements from contrived tumor samples at different tumor fractions for two tumors (Fig. 2A). We fit those CV measurements using a logarithmic regression model as a function of the number of methylated molecules after background subtraction and additionally accounting for expected Poisson sampling noise, and then adding cfDNA extraction noise in quadrature (Supplementary Fig. 14). The difference between a first and second time point can then be modeled as the difference between two normally distributed variables with a mean corresponding to the difference between the two means and a standard deviation corresponding to the square root of the sum of the two variances.

An Increase call is made if the mean of this distribution is $\geq 15\%$ and the log₂ likelihood ratio that the difference is $\geq 15\%$ compared to $< 15\%$ is greater than 3. Similarly, a Decrease call is made if the mean of this distribution is $\leq -15\%$ and the log₂ likelihood ratio that the difference is $\leq -15\%$ compared to $> -15\%$ is greater than 3. If the relative difference is not of sufficiently large magnitude or the statistical likelihood is not sufficiently strong, a No Change call is made. If either time point is below the calling threshold, the mean for that time point is set to the calling threshold, the standard deviation is still calculated based on the total number of methylated molecules, and a call can still be made. If the methylation signal from both time points is below the calling threshold, an Below Quantifiable Range call is made.

In silico scaling of Tumor Methylation Scores at untested tumor fractions

To analyze the assay's performance for step increases of 0.25% and 0.5% tumor fraction, we scaled Tumor Methylation Scores to tumor fractions that were not experimentally tested. For tumor fractions between 0 and 1%, the measured Tumor Methylation Scores of 1% tumor fraction replicates were multiplied by the ratio of the desired tumor fraction and the original tumor fraction to yield simulated Tumor Methylation Scores. Similarly, Tumor Methylation Scores for tumor fractions between 1 and 2% were calculated by appropriately scaling down the Tumor Methylation Scores of 2% tumor fraction replicates.

Methylation-based estimation of tumor purity

One challenge that complicates comparisons of assay performance across many different tumors is the unknown fraction of cancerous tissue in each specimen. Histological analyses (e.g. hematoxylin and eosin) to determine tumor purity can be subjective, and a two-dimensional slide may not accurately represent the entire tumor. We tested matched tumor and buffy gDNA for each subject and calculated a tumor purity score for each tumor. Tumor purity scores were calculated as follows: 1) The number of methylated molecules detected at each hypermethylation locus was normalized to 1,000 GE, as measured by highly methylated control loci; 2) The percent hypermethylation was calculated for each locus as the difference between the number of methylated molecules measured in tumor gDNA and buffy coat gDNA divided by 1,000 minus the number of methylated molecules measured in buffy coat; 3) The tumor purity of a given tumor was defined as the 75th percentile of percent hypermethylation across all informative loci detected in a tumor sample. The tumor purity scores ranged from 0.7% to 104% across tumors, and we capped the estimated tumor purity at 100% when accounting for tumor purity to make contrived samples at 1% and 2% tumor fraction (Supplementary Fig. 9A).

To validate the computed tumor purity score as a meaningful metric, we compared tumor purity scores for a variety of tumors to orthogonal tumor purity estimates that were based on somatic mutations. We used a least squares approach to identify a percentile that maximizes concordance between the two orthogonal measurements of tumor purity, which is how we decided to define the tumor purity as the 75th percentile of percent hypermethylation across loci (Supplementary Fig. 9B, 9C).

Orthogonal tumor-naive, targeted-panel assay

A research-use, tumor-naive, targeted-panel assay was used to measure VAF of somatic mutations across a panel of 80 genes (Supplemental Table 4). Library prep for the assay consists of a custom hybrid capture panel spanning 217 kb (IDT, Coralville, IA). The target enriched libraries were run on a NextSeq 2000 sequencer (Illumina, San Diego, CA) at approximately 100 million reads per sample. Mutations were called using a custom-built bioinformatics pipeline, which involves adapter trimming (Trimmomatic), mapping to the human genome (bwa mem), consensus calling (fgbio), custom filtering and thresholding of data to account for background and erroneous signals, and calling of somatic mutations with VAF determined based on the ratio of mutant to total molecule counts (VarDict). The clinical significance of any detected mutations was assessed using Qiagen Clinical Insight (Qiagen, Hilden, Germany). Only clinically significant mutations were included for VAF analysis.

To determine the performance of using VAF measurements for therapy response monitoring, samples were prepared by extracting genomic DNA from 10 different tumor samples with known mutations, and one healthy tissue sample. These genomic DNA samples were sheared via sonication with a Covaris E220 Sonicator (Covaris, Woburn, MA) to an average length of ~ 175 bp to mimic the size of native cfDNA fragments. To generate a wide range of VAFs in the contrived sample, the tumor samples were combined at equal proportions to target various allele frequencies, and the tumor mixture was sequenced in 6 replicates to determine the VAF in the mix for each mutation. The combined tumor mix sheared DNA sample was added into a healthy background (buffy coat sheared DNA) sample at 10% and 20% by mass. 20 replicates of 30 ng of each of these mixes were run on the in-house hybrid capture-based assay. Mutations were filtered by VAF to create groups of mutations centered around each VAF of interest: 0.25% (8 mutations ranging from 0.2% to 0.3%, with mean 0.25%), 0.5% (11 mutations ranging from 0.4% to 0.6%, with mean 0.53%), 1% (10 mutations ranging from 0.8% to 1.2%, with mean 0.97%), and 2% (5 mutations ranging from 1.2% to 2.8%, with mean 1.68%).

In order to validate the precision of VAF from contrived replicate samples, a simulation was performed to determine whether the noise in the VAF measurement agrees with statistical theory. Data was generated assuming Poisson sampling of mutant molecules around a mean of the number of mutant molecules (central VAF [0.25%, 0.5%, 1%, 2%] \times 2,000 GE), with 10,000 mutations simulated for each VAF. 2,000 GE was chosen because it was the average detected across all the mutations in the experimental dataset. The data was split into subgroups of different numbers of mutations and the VAF was summed across them, then the noise was calculated by taking the CV for each individual or summed VAF for each group size. The simulation was rerun 100 times to determine the variability of this measurement. The median across the 100 replicates of the simulation is shown in Fig. 4, for 1, 2, 3, and 4 mutations utilized. The full results are shown in Supplementary Fig. 8.

Data availability

The datasets generated and analyzed during the current study are available in the European Genome-Phenome Archive under study accession number: EGAS5000000734 (<https://ega-archive.org/studies/EGAS5000000734>).

Received: 27 December 2024; Accepted: 10 February 2025

Published online: 18 February 2025

References

- Nishino, M. Tumor response assessment for precision cancer therapy: response evaluation criteria in solid tumors and beyond. *Am. Soc. Clin. Oncol. Edu. Book*. https://doi.org/10.1200/edbk_201441 (2018).
- Jia, W., Gao, Q., Han, A., Zhu, H. & Yu, J. The potential mechanism, recognition and clinical significance of tumor pseudoprogression after immunotherapy. *Cancer Biol. Med.* **16**, 655–670 (2019).
- Kyriazi, S., Kaye, S. B. & deSouza, N. M. Imaging ovarian cancer and peritoneal metastases—current and emerging techniques. *Nat. Rev. Clin. Oncol.* **7**, 381–393 (2010).
- Wolf, D. K., Padhani, A. R. & Makris, A. Assessing response to treatment of bone metastases from breast cancer: what should be the standard of care?. *Ann. Oncol.* **26**, 1048–1057 (2015).
- Hamaoka, T., Madewell, J. E., Podoloff, D. A., Hortobagyi, G. N. & Ueno, N. T. Bone imaging in metastatic breast cancer. *J. Clin. Oncol.* **22**, 2942–2953 (2004).
- Landegren, U. & Hammond, M. Cancer diagnostics based on plasma protein biomarkers: hard times but great expectations. *Mol. Oncol.* **15**, 1715–1726 (2021).
- Sanz-Garcia, E., Zhao, E., Bratman, S. V. & Siu, L. L. Monitoring and adapting cancer treatment using circulating tumor DNA kinetics: Current research, opportunities, and challenges. *Sci. Adv.* **8**, eabi8618 (2022).
- Goldberg, S. B. et al. Early assessment of lung cancer immunotherapy response via circulating tumor DNA. *Clin. Cancer Res.* **24**, 1872–1880 (2018).
- Mack, P. C. et al. Circulating tumor DNA kinetics predict progression-free and overall survival in EGFR TKI-treated patients with EGFR-mutant NSCLC (SWOG S1403). *Clin. Cancer Res.* **28**, 3752–3760 (2022).
- Kurtz, D. M. et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-021-00981-w> (2021).
- Vega, D. M. et al. Changes in circulating tumor DNA reflect clinical benefit across multiple studies of patients with non-small-cell lung cancer treated with immune checkpoint inhibitors. *JCO Precis Oncol.* **6**, e2100372 (2022).
- Hanna, G. J. et al. Personalized circulating tumor DNA for monitoring disease status in head and neck squamous cell carcinoma. *Clin. Cancer Res.* <https://doi.org/10.1158/1078-0432.ccr-24-0590> (2024).
- Thompson, J. C. et al. Serial monitoring of circulating Tumor DNA by Next-Generation gene sequencing as a biomarker of response and survival in patients with advanced NSCLC receiving pembrolizumab-based therapy. *JCO Precis Oncol.* **5**, 510–524 (2021).
- Phallen, J. et al. Early noninvasive detection of response to targeted therapy in non-small cell lung cancer. *Cancer Res.* **79**, 1204–1213 (2019).
- Zhang, Q. et al. Prognostic and predictive impact of circulating tumor dna in patients with advanced cancers treated with immune checkpoint blockade. *Cancer Discov.* **10**, 1842–1853 (2020).
- Zill, O. A. et al. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin. Cancer Res.* **24**, 3528–3538 (2018).
- Pairawan, S. et al. Cell-free circulating tumor DNA variant allele frequency associates with survival in metastatic cancer. *Clin. Cancer Res.* **26**, 1924–1931 (2020).
- Mack, P. C. et al. Spectrum of driver mutations and clinical impact of circulating tumor DNA analysis in non-small cell lung cancer: Analysis of over 8000 cases. *Cancer* **126**, 3219–3228 (2020).
- Lanman, R. B. et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS One* **10**, e0140712 (2015).
- Deveson, I. W. et al. Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology. *Nat. Biotechnol.* **39**, 1115–1128 (2021).
- Zhang, S. et al. Utility of tumor-informed circulating tumor DNA in the clinical management of gastrointestinal malignancies. *J. Gastrointest Oncol.* **12**, 2643–2652 (2021).
- Azzi, G. et al. Using tumor-informed circulating tumor DNA (ctDNA)-based testing for patients with anal squamous cell carcinoma. *Oncologist* **28**, 220–229 (2023).
- van der Leest, P. et al. Circulating tumor DNA as a biomarker for monitoring early treatment responses of patients with advanced lung adenocarcinoma receiving immune checkpoint inhibitors. *Mol. Oncol.* **15**, 2910–2922 (2021).
- Nabet, B. Y. et al. Noninvasive early identification of therapeutic benefit from immune checkpoint inhibition. *Cell* **183**, 363–376.e13 (2020).
- Janku, F. Tumor heterogeneity in the clinic: is it a real problem?. *Ther. Adv. Med. Oncol.* **6**, 43–51 (2014).
- Xu, R.-H. et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16**, 1155–1161 (2017).
- Vrba, L. et al. DNA methylation biomarkers discovered in silico detect cancer in liquid biopsies from non-small cell lung cancer patients. *Epigenetics* **15**, 419–430 (2020).
- de Vos, L. et al. Treatment response monitoring in patients with advanced malignancies using cell-free SHOX2 and SEPT9 DNA methylation in blood: an observational prospective study. *J. Mol. Diagn.* **22**, 920–933 (2020).
- Schmidt, B. et al. Quantification of cell-free mSHOX2 plasma DNA for therapy monitoring in advanced stage non-small cell (NSCLC) and small-cell lung cancer (SCLC) patients. *PLoS One* **10**, 1–10 (2015).
- Fackler, M. J. et al. Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. *Cancer Res.* **74**, 2160–2170 (2014).

31. Forootan, A. et al. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol. Detect Quantif* **12**, 1–6 (2017).
32. Lau, B. T. et al. Single-molecule methylation profiles of cell-free DNA in cancer with nanopore sequencing. *Genome Med.* **15**, 33 (2023).
33. Tsao, D. S. et al. A novel high-throughput molecular counting method with single base-pair resolution enables accurate single-gene NIPT. *Sci. Rep.* **9**, 14382 (2019).
34. Cancer Genome Atlas Research Network et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113–20 (2013).
35. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
36. Eisenhauer, E. A. et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* **45**, 228–247 (2009).
37. Pedersen, S. K. et al. A Two-Gene blood test for methylated dna sensitive for colorectal cancer. *PLoS One* **10**, e0125041 (2015).
38. Taryma-Leśniak, O., Sokolowska, K. E. & Wojdacz, T. K. Current status of development of methylation biomarkers for in vitro diagnostic IVD applications. *Clin. Epigenetics* **12**, 100 (2020).
39. Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell*. **168** 629–643 Preprint at <https://doi.org/10.1016/j.cell.2016.12.013> (2017).
40. Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell*. **152** 1237–1251 Preprint at <https://doi.org/10.1016/j.cell.2013.02.014> (2013).
41. Vishnoi, K., Viswakarma, N., Rana, A. & Rana, B. Transcription factors in cancer development and therapy. *Cancers*. **12** 1–32 Preprint at <https://doi.org/10.3390/cancers12082296> (2020).
42. Mark, M., Rijli, F. M. & Chambon, P. Homeobox Genes in embryogenesis and pathogenesis. *Pediatr. Res.* **42**, 421–429 (1997).
43. Ma, Y. et al. The relationship between early embryo development and tumorigenesis. *J. Cell Mol. Med.* **14**, 2697–2701 (2010).
44. Moss, J. et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).
45. Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA comprises an in vivo nucleosome footprint that informs Its Tissues-Of-Origin. *Cell* **164**, 57–68 (2016).
46. Su, J. et al. Homeobox oncogene activation by pan-cancer DNA hypermethylation. *Genome. Biol.* **19**, 108 (2018).
47. Bratman, S. V. et al. Personalized circulating tumor DNA analysis as a predictive biomarker in solid tumor patients treated with pembrolizumab. *Nat. Cancer* **1**, 873–881 (2020).
48. Zollinger, D. R. et al. Analytical validation of a novel comprehensive genomic profiling informed circulating tumor DNA monitoring assay for solid tumors. *PLoS One* **19**, e0302129 (2024).
49. Ricciuti, B. et al. Early plasma circulating tumor DNA (ctDNA) changes predict response to first-line pembrolizumab-based therapy in non-small cell lung cancer (NSCLC). *J. Immunother. Cancer* **9**, e001504 (2021).
50. Raja, R. et al. Early reduction in ctDNA predicts survival in patients with lung and bladder cancer treated with durvalumab. *Clin. Cancer Res.* **24**, 6212–6222 (2018).
51. Angeli-Pahim, I. et al. Methylated ctDNA quantification: noninvasive approach to monitoring hepatocellular carcinoma burden. *J. Am. Coll Surg.* **238**, 770–778 (2024).
52. Hsiao, A. et al. Brief report: Methylation-based ctdna serial monitoring correlates with immunotherapy response in NSCLC. *Clin. Lung Cancer* <https://doi.org/10.1016/j.clcc.2024.10.013> (2024).
53. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic. Acids Res.* **51**, D638–D646 (2023).

Acknowledgements

We would like to acknowledge Naomi Searle for her expertise using QCI for variant classification. We would like to acknowledge Matthew Varga, John ten Bosch, and Shan Riku for providing editorial support.

Author contributions

PY, RV, KS, XB, WZ, HH, DT, and OA conceived and designed experiments. RV, KS, SL, XB, JW, JZ, and BW conducted experiments. PY, RV, KS, XB, JS, JW, and WZ analyzed the results. PY, RV, KS, and XB wrote the manuscript. All authors reviewed the manuscript.

Declarations

Competing interests

PY, RV, KS, SL, XB, JS, WZ, JW, JZ, DT, and OA are employees of BillionToOne and/or hold stock or options to hold stock in the company. Patent disclosures about this work have been filed with BillionToOne, Inc. BillionToOne, Inc. funded portions of this study. BW and HH do not have competing interests to disclose.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-90013-3>.

Correspondence and requests for materials should be addressed to P.P.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025