HHS Public Access

Author manuscript

Sci Transl Med. Author manuscript; available in PMC 2019 August 29.

Published in final edited form as:

Sci Transl Med. 2017 August 16; 9(403): . doi:10.1126/scitranslmed.aan2415.

Direct detection of early-stage cancers using circulating tumor DNA

Jillian Phallen[1], **Mark Sausen**[2], **Vilmos Adleff**[1], **Alessandro Leal**[1], **Carolyn Hruban**[1], **James White**[1], **Valsamo Anagnostou**[1], **Jacob Fiksel**[1], **Stephen Cristiano**[1], **Eniko Papp**[1,+], **Savannah Speir**[1], **Thomas Reinert**[3], **Mai-Britt Worm Orntoft**[3], **Brian D. Woodward**[4], **Derek Murphy**[2], **Sonya Parpart-Li**[2], **David Riley**[2], **Monica Nesselbush**[2], **Naomi Sengamalay**[2], **Andrew Georgiadis**[2], **Qing Kay Li**[1], **Mogens Rørbæk Madsen**[5], **Frank Viborg Mortensen**[6], **Joost Huiskens**[7,8], **Cornelis Punt**[8], **Nicole van Grieken**[9], **Remond Fijneman**[10], **Gerrit Meijer**[10], **Hatim Husain**[4], **Robert B. Scharpf**[1], **Luis A. Diaz Jr.**[1,+], **Siân Jones**[2], **Sam Angiuoli**[2], **Torben Ørntoft**[3], **Hans Jørgen Nielsen**[11], **Claus Lindbjerg Andersen**[3], **Victor E. Velculescu**[1,*]

[1]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA [2]Personal Genome Diagnostics, Baltimore, MD 21224, USA [3]Department of Molecular Medicine, Aarhus University Hospital, DK-8200 Aarhus, Denmark [4]Division of Hematology and Oncology, Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA [5]Department of Surgery, Herning Regional Hospital, DK-7400 Herning, Denmark [6]Department of Surgical Gastroenterology, Aarhus University Hospital, DK-8000 Aarhus, Denmark [7]Department of Surgery, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands [8]Department of Medical Oncology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands [9]Department of Pathology, VU University Medical Center, Amsterdam 1081HV, The Netherlands [10]Department of Pathology, The Netherlands Cancer Institute, Amsterdam 1066CX, The Netherlands [11]Department of Surgical Gastroenterology 360, Hvidovre Hospital, Hvidovre, Denmark

Abstract

Early detection and intervention are likely to be the most effective means for reducing morbidity and mortality of human cancer. However, development of methods for non-invasive detection of early stage tumors has remained a challenge. We have developed an approach called targeted error correction sequencing (TEC-Seq) that allows ultra-sensitive direct evaluation of sequence changes in circulating cell-free DNA using massively parallel sequencing. We have used this approach to examine 58 cancer-related genes encompassing 81 kb. Analysis of plasma from 44 healthy individuals identified genomic changes related to clonal hematopoiesis in 16% of asymptomatic

*To whom correspondence should be addressed: velculescu@jhmi.edu.

+Present Address: E.P., Personal Genome Diagnostics, Baltimore, MD 21224; L.A.D., Memorial Sloan Kettering Cancer Center, New York, NY 10065

Author Contributions: J.P. designed the study, performed experiments, performed analyses, provided funding, and wrote the paper. M.S. V. Adleff., and R.B.S. performed analyses and wrote the paper. A.L., C.H., J.W., J.F., E.P., S.S., T.R., M.W.O., B.D.W., D.M., S.P., D.R., M.N., N.S., A.G., Q.K.L., M.R.M., F.V.M., J.H., C.P. N.V., R.F., H.H., S.J., S.A., T.O., and H.J.N performed analyses. V. Anagnostou., S.C., G.M., and L.A.D. performed analyses and provided funding. C.L.A. performed analyses, provided funding, and wrote the paper. V.E.V designed the study, performed analyses, provided funding, and wrote the paper.

individuals but no alterations in driver genes related to solid cancers. Evaluation of 200 patients with colorectal, breast, lung, or ovarian cancer detected somatic mutations in the plasma of 71%, 59%, 59%, and 68%, respectively, of patients with stage I or II disease. Analyses of mutations in the circulation revealed high concordance with alterations in the tumors of these patients. In patients with resectable colorectal cancers, higher amounts of preoperative circulating tumor DNA were associated with disease recurrence and decreased overall survival. These analyses provide a broadly applicable approach for non-invasive detection of early stage tumors that may be useful for screening and management of patients with cancer.

INTRODUCTION

Over 14 million individuals are newly diagnosed with cancer world-wide each year, with the majority having invasive or metastatic disease (1). I t is well established that much of the morbidity and mortality in human cancer is related to the late diagnosis of this disease, where surgical and pharmacologic therapies are less effective (2). Unfortunately, clinically proven biomarkers that can be used to broadly diagnose and guide patient management early in the course of disease are not available. Serum-based protein biomarkers such as cancer antigen-125 (CA-125), carcinoembryonic antigen (CEA), prostate-specific antigen (PSA), and cancer antigen 19–9 (CA 19–9) are commonly used for monitoring cancer patients, but because these proteins are also found in the serum of individuals without cancer, they are typically not useful for disease diagnosis (3–7). Other approaches for early detection of cancer, such as stool-based molecular tests or colonoscopies, are limited to individual tumor types and have challenges in patient compliance (8, 9). Currently, no widely-applicable biomarkers have been developed for broad detection of human cancer.

The development of non-invasive liquid biopsy methods based on the analysis of cell free DNA (cfDNA) provides the opportunity for a new generation of diagnostic approaches. Although cfDNA in the circulation was first described over fifty years ago (10), abnormalities in cancer patients were observed only decades later (11, 12) and showed that such individuals have higher amounts of cfDNA. In patients with cancer, a fraction of cfDNA is tumor-derived and is termed circulating tumor DNA (ctDNA). In principle, analysis of ctDNA has the advantage of identifying alterations that are specific to the tumor. The application of next-generation sequencing (NGS) together with advanced computational methods has recently allowed ctDNA-based tumor genotyping in a variety of cancer types (13–22). However, these approaches have largely been applied in patients with late stage cancers or have used tumor tissue sequencing to guide mutational analyses in the blood.

In this study, we have developed an ultra-sensitive approach for direct analysis of sequence alterations in commonly altered cancer genes in cfDNA without prior knowledge of alterations in the tumor. The sensitivity and specificity of the methodology was evaluated in a clinically relevant cohort of healthy individuals as well as those with early stage disease in four common cancers. We identified sequence alterations in cell proliferation genes in individuals without cancer, established the sensitivity of the approach for detecting tumor-specific alterations in the plasma of cancer patients, evaluated concordance between plasma and tumor samples from the same patients, and showed that the amounts of ctDNA can serve

as a predictive marker of patient outcome. Overall, these analyses provide information on the potential utility and limitations of large-scale mutation-based measurements of ctDNA for early diagnosis in common cancers.

RESULTS

Targeted error correction sequencing

We developed a methodology for comprehensive analysis of sequence alterations in driver genes that are commonly mutated in colorectal, lung, ovarian, breast, and other cancers. Similar to targeted analyses of cancer tissues (23), we first selected genes that were frequently mutated in these tumors and focused our analyses on either the entire coding regions or the most highly mutated exons of these genes. An analysis of the frequency of these alterations in the COSMIC database of somatic mutations in cancer (24) revealed that over three quarters of patients would be expected to have at least one mutation in 55 genes among the intended cancers as well as other common tumor types (Table 1, table S1). We hypothesized that a larger panel of genes would increase the probability of detecting at least one gene alteration in the plasma from any given cancer patient. Because alterations in the blood have previously been reported in healthy individuals, we examined three additional genes as well as specific sequence positions in three genes of the 55 gene panel (table S1) that were known to be somatically altered in clonal hematopoietic expansion, myelodysplasia, or other hematological malignancies (25–27).

Detection of sequence alterations using conventional next generation sequencing is limited to a relatively high fraction of mutant to wild-type DNA (>1%) and as such is typically not useful for analyses of ctDNA which may be present in minute amounts in the blood. Although methods have been developed for analysis of ctDNA in late stage cancer patients (13–21), no method has been systematically applied for analysis of early stage disease. We developed a custom capture and sequencing approach called targeted error correction sequencing (TEC-Seq) to allow sensitive and specific detection of low abundance sequence alterations using next generation sequencing (Fig. 1). This methodology is based on targeted capture of multiple regions of the genome and deep sequencing (~30,000x) of DNA fragments. The 58 genes analyzed in this study comprised 80,930 captured bases. Specific steps were performed for analysis of rare tumor-specific alterations in DNA molecules and for elimination of potential amplification, sequencing, and contamination errors as well as other sources of alterations in the blood. These included (*1*) optimized library generation and capture for conversion of cfDNA for subsequent analyses, (*2*) maximizing representation of unique cfDNA molecules analyzed using mapping positions and a small number of pre-specified barcodes, (*3*) redundant sequencing, where multiple identical DNA molecules are generated, sequenced, and any sequence changes reconciled, (*4*) filtering of mapping and sequencing artifacts, and (*5*) identification and removal of germline and hematopoietic cell proliferation alterations.

Conceptually, the number of genome equivalents analyzed provides a lower limit of detection for any genomic analysis. A high sensitivity approach would aim to maximize the number of unique molecules assessed while allowing for a broad and facile analysis in a range that is above the actual number of fragments present in a biologic sample. We

optimized methodologies for extraction and conversion of cfDNA to genomic libraries. Initially, we considered using the start and end genome mapping positions of paired-end sequenced fragments as “endogenous barcodes” to distinguish between individual molecules. However, Monte Carlo simulations suggested that the tight size distribution of cfDNA molecules observed in the plasma would result in a smaller number of possible end mapping combinations and therefore underestimate the true complexity of cfDNA in the circulation (fig. S1). To extend the complexity of endogenous barcodes, we introduced a limited set of sequence indices as “exogenous barcodes” in the initial steps of library generation. Kinde et al. reported use of a large number of random exogenous barcodes as unique identifiers for analysis of rare mutations in DNA populations (28). However, simulations with a relatively small number of long pre-specified exogenous barcodes (4–16) suggested that these in combination with endogenous barcodes would be sufficient to distinguish among different cfDNA molecules in the plasma from a typical blood draw (fig. S1). Extending the number of barcodes substantially beyond this number has the theoretical disadvantage of misassignment among barcodes through sequencing errors and of primer dimers that can be formed during library formation.

We first evaluated the characteristics of the TEC-Seq approach for detecting known tumor-specific alterations from a mixture of DNA from tumor cell lines at different dilutions (ranging from 100% to 0.1%) with unrelated wild-type DNA. Libraries with eight exogenous barcodes were sequenced with an average of ~32,224 sequence reads at each position among the 58 genes analyzed (table S2). We designed thresholds that were expected to identify >99% of alterations with a mutant allele fraction of 0.5% at the anticipated sequencing depth. Alterations were considered if they were present in all copies of multiple sequences of each DNA molecule with identical endogenous and exogenous barcodes and were not removed by additional error filtering steps. Hot-spot alterations at positions previously observed to be frequently altered in cancer patients were evaluated with more sensitive thresholds because the a priori probability that these alterations were tumor-derived is higher than that of other alterations. Alterations present in common germline variant databases or in 25% or greater of reads were considered germline and removed from further analysis unless the mutations were identical to known hot-spot alterations or represented truncating mutations in common tumor suppressor genes. Analysis of the altered positions in the dilution samples revealed high concordance to the expected fraction of mutant molecules ($r=0.93$, $P<0.0001$, Pearson correlation; fig. S2, table S2), as well as high sensitivity and specificity. The analytical sensitivity was 97.4% overall, and 100% and 89% for detecting mutations present at 0.2% and 0.1%, respectively, using minimum thresholds of 0.05% in hot-spot positions and 0.1% at all other locations. No false positives were detected over the 80,930 bases analyzed in 38 dilution analyses, resulting in less than one error in three million bases sequenced (error rate of $<3.3 \times 10^{-7}$ false positive mutation calls per base, specificity >99.9999%, table S2).

Evaluation of plasma from healthy individuals

We used TEC-Seq to examine plasma specimens from 44 healthy individuals (tables S3 and S4). These individuals were not known to have cancer and provided their blood samples as part of a routine cancer screening visit (colonoscopy or Papanicolaou test). Samples were

processed within two hours from collection and centrifuged twice at high speed to ensure that cells and cellular debris were removed and that only cfDNA was analyzed. From the ~4 ml of plasma obtained from each individual, we generated TEC-Seq libraries and sequenced these to ~30,000 fold coverage. Through these analyses, no mutations were observed in the cancer driver genes analyzed in our panel, consistent with the estimated specificity observed in our dilution analyses. Whereas conventional sequencing of these samples would have resulted in thousands of putative alterations among the regions analyzed, the TEC-Seq analyses significantly reduced the sequencing error rate to fewer than one false positive per three million bases sequenced ($<3\times10^{-7}$ false positive mutation calls per base, $P < 0.0001$, paired t test, Fig. 2). We compared the TEC-Seq error rate to those obtained through other liquid biopsy analyses. Reanalysis of our sequence data from 15 healthy individuals using the recently developed integrated digital error suppression (iDES) method (19, 21) resulted in multiple false positive alterations in the healthy cases, consistent with the reported error rate of this approach (21).

Analysis of six genes related to hematopoietic proliferation identified six individuals with a single mutation in their plasma samples, and a seventh had two detectable alterations (16% of patients analyzed, table S5). All of the alterations were identified in DNA Methyltransferase 3 Alpha *(DNMT3A)*, a gene that is clonally altered in pre-leukemic conditions and myelodysplasia (25–27). Three of the mutations were predicted to result in the R882C change previously observed in clonal hematopoiesis, but other alterations have not been previously reported. These mutations were identified at mutant allele fractions of 0.16% to 5.3%, substantially lower than previous observations in blood cells of healthy individuals (25–27). Our analyses suggest that a higher fraction of asymptomatic individuals may harbor such somatic alterations than had been previously reported through cellular analyses of these genes in the blood.

Analysis of plasma from patients with cancer

We next analyzed plasma samples from 194 patients with breast cancer (n=45), colorectal cancer (n=42), lung cancer (n=65), and ovarian cancer (n=42). The cohort consisted of untreated patients who had localized or metastatic disease, with the majority of patients diagnosed at stage I and II (table S3). We found that the concentration of cfDNA in plasma from cancer patients was ~29 ng/ml, significantly higher than that observed in healthy individuals (average of 7 ng/ml, P=0.001, unpaired t test, Fig. 3A). In the colorectal cancer cohort, where a larger number of later stage patients were analyzed, we found that samples from patients with metastatic disease had higher concentration of cfDNA than those from patients with earlier stages (average of 66 ng/ml for stage IV patients versus 21 ng/ml for stage I-III, P=0.006, unpaired t test; Fig. 3B).

We examined the cfDNA from these patients using the TEC-Seq approach. Of the 194 patients analyzed, over three quarters of colorectal cancer patients, two thirds of ovarian cancer patients, and the majority of lung and breast cancer patients had detectable alterations in driver genes (Table 2). These detection rates were higher in some cases than the theoretical estimates for these cancer types (Tables 1 and 2). Over three quarters of patients with advanced disease (stage III and IV) and 62% of patients with localized disease (stage I

and II) were detected among all tumor types (Table 2). The amounts of ctDNA varied among cancer types, with breast cancer having the lowest mutant allele fraction (P=0.028, unpaired t test, Fig. 3C). Similar to observations of cfDNA, the amounts of ctDNA were higher in metastatic disease compared to earlier stage disease among all cancer types (P<0.0001, unpaired t test, Fig. 3D and 4). Eighty of 128 detected cases had at least one alteration in a gene hot-spot position (Fig. 4). The affected genes and distribution of alterations for each tumor type were similar to common driver gene alterations that have previously been reported in these cancers (fig. S3). On average 2.1 alterations, including 0.9 changes at hot-spot positions, were observed in each patient with detectable ctDNA, with lung and colorectal cancers having a higher number of alterations per case (Fig. 4). By limiting analysis only to a specific set of hot-spot variants as others have reported (21), the fraction of cases detected was reduced to 56% of those identified by TEC-Seq. These observations highlight the benefit of analyzing a broader panel of driver gene regions to increase the possibility of detecting tumor-specific alterations in the plasma.

Comparison of mutations in plasma with those in matched tumor and blood cells

Of the 194 patients in our study, 152 cases had matched tumor and normal tissues that we analyzed using an independent targeted next-generation sequencing approach (tables S3 and S4). We examined these cases to determine whether the mutations identified in the plasma were tumor-specific or may have originated during blood cell expansion. The plasma analyses performed using TEC-Seq were performed separately and did not rely on any knowledge of alterations identified through these parallel tissue analyses.

We detected 87 changes in the circulation of 194 patients at allele fractions >25% and considered these to be likely germline variants. Analysis of 63 of these variants in the available corresponding blood cells identified all of these changes to be germline (table S6). These observations suggested that cfDNA can be used to accurately identify germline changes in the context of tumor-derived and blood cell proliferation alterations, and similarly that this approach can be used to distinguish these changes from somatic alterations.

Similar to our observations in healthy individuals, we identified alterations in *DNMT3A* and five other genes involved in blood cell proliferation in the plasma of cancer patients (table S5). The fraction of patients with detectable changes in these genes correlated with age, as previously observed (P=0.013, unpaired t test) (25–27). Unlike tumor-specific alterations, the allele fractions of blood cell proliferation alterations in cfDNA were similar among healthy individuals and patients with cancer, regardless of stage. Analysis of matched white blood cells from individuals with alterations in these genes identified the corresponding mutation in a majority of cases, consistent with the notion that the alterations in cfDNA originated from these cells (table S5).

After accounting for blood cell proliferation and germline alterations, we identified 313 candidate tumor-specific changes in the plasma samples from 128 of the 194 patients analyzed. We further evaluated 216 of these alterations in 100 patients where matched tumor tissue and blood cells were available. We found that 155 of the 216 (72%) alterations were identical in both plasma and tumor samples (Fig. 5). Among stage III and IV patients, 65 of

84 variants were concordant (77%), whereas for early stage patients, 90 of 132 alterations were concordant (68%). In line with these observations, we found that 70 of the 75 alterations (93%) with a mutant allele fraction > 1% in the plasma were detected in the tumor tissue of the same individual. Overall, 82 of the 100 patients (82%) had at least one alteration observed in the circulation that was identical to that in the tumor specimen.

To evaluate reproducibility of the approach between separate blood draws in the same patients, we assessed six late stage patients with lung cancer where blood was obtained early during the course of treatment. These patients were undergoing treatment but were observed to have progressive or stable disease. Despite the difference in time between the blood draws, we found that 90% of the alterations observed in the second blood draw were present at the time of the first blood draw (17 of 19 alterations), with one patient having no alterations at both time points (fig. S4). All alterations present with a mutant allele fraction > 1% were observed at both time points.

In a subset of colorectal cancer patients, we evaluated whether the observations we detected in the plasma could be independently confirmed using droplet digital PCR (ddPCR), a method that is known to be highly sensitive for detection of single base substitutions (29). We examined six driver alterations detected in the plasma, two that were also detected in matched tumors, and four that were absent. Five of the six driver alterations were detected in the plasma by ddPCR at levels similar to those observed by TEC-Seq (fig. S5A). Those not detected in tumors by targeted sequencing were similarly not identified through ddPCR approaches. We also evaluated 10 mutations that corresponded to the most common changes in *KRAS, PIK3CA*, and *BRAF* that we detected in these tumors but were not present in the plasma of these patients. Although we confirmed that these alterations were in the tumors of these patients, we found that those not detected by TEC-Seq analyses remained undetected by ddPCR in the plasma, presumably because the amounts of ctDNA corresponding to these alterations were extremely low in these patients (fig. S5B).

To assess the possibility that tumor heterogeneity may be responsible for the apparent lack of concordance between specific alterations in the plasma and those in the tumor, we analyzed multiple tumor sites from colorectal cancer patient CGCRC307 using ddPCR. We characterized ten different regions of the tumor as well as a subsequent metastatic site for a R201C alteration in the *GNAS* gene that we detected in the plasma but not in the tumor of this patient. Although we found a *BRAF* V600E alteration in all samples analyzed, the *GNAS* R201C substitution was not detected in the original tumor biopsy but was detected as a subclonal change in only a portion of the primary tumor, suggesting it developed later in tumorigenesis (fig. S6). The *GNAS* R201C change identified had been previously reported in colorectal cancers (30) and has been shown to promote intestinal tumorigenesis through activation of both Wnt and ERK pathways (31). Consistent with this notion, we found the *GNAS* alteration to be clonal in the metastatic lesion that was identified two years after the primary tumor in this patient (fig. S6). These results suggest that plasma alterations not detected in the matched tumor specimens may represent *bona fide* somatic mutations in ctDNA derived from heterogeneous primary or occult lesions.

ctDNA and disease progression

Tumor-specific markers may be useful for evaluating disease progression. In colorectal cancer, carcinoembryonic antigen (CEA) is commonly used to monitor patients after therapy to determine recurrence or progressive disease (7, 32). Of the 29 colorectal cancer patients for whom CEA values were available, all ten cases with CEA concentrations >5 ng/ml had detectable ctDNA (tables S3 and S6). However, among the 19 patients with negative or borderline CEA results, 13 had detectable ctDNA, including patients of all stages (tables S7 and S8). There was no significant correlation between ctDNA and CEA concentrations (Pearson correlation coefficient = −0.017, P=0.93).

We next examined whether preoperative ctDNA analyses may be related to disease recurrence and survival after surgical resection. We hypothesized that elevated amounts of ctDNA were more likely to be associated with large primary lesions that were incompletely resected or with occult metastases. A total of 31 colorectal cancer patients had potentially curative resections, including eight stage I, nine stage II, ten stage III, and four stage IV patients with liver-only metastases. For these patients, the median mutant allele fraction was 0.21%. However, several patients had mutant allele fractions greater than three median absolute deviations from the median mutant allele fraction, or >2%. As predicted, we found that high amounts of ctDNA correlated with poor prognosis (fig. S7). Patients with increased ctDNA had a shorter progression-free survival (PFS) and overall survival (OS) compared to patients with lower ctDNA amounts (P<0.0001 for PFS and OS, Log-rank test, Fig. 6A–B). The prognostic value for progression-free survival was statistically significant in multivariate models, adjusted for stage as a categorical covariate (Hazard ratio =36.3, 95% CI =2.8–471.1, P=0.006, Cox proportional hazards model). These same predictions were observed in patients with resectable stage I-III disease (P=0.0006 for PFS and P<0.0001 for OS, Log-rank, test, Fig. 6C–D). We also evaluated other thresholds of increased amounts of ctDNA and found that these were statistically significantly associated with worse outcome (P=0.008 for 0.5% mutant allele fraction, and P=0.0001 for 1% mutant allele fraction, Log-rank Test). In addition, we found that considering ctDNA amounts as a continuous variable correlated with outcome (Hazard ratio=1.13, 95% CI =1.03–1.24, P=0.01 for PFS and OS, Cox univariate test). Together, these results indicate that liquid biopsy analyses offer both a quantitative and qualitative assessment of disease progression. Although previous analyses have found a limited association between preoperative CEA concentrations and overall survival (7, 32), CEA concentrations among our patients were not associated with disease outcome (P=0.75 for PFS and P=0.73 for OS, Log-rank test, fig. S8). These analyses from a limited and heterogeneous cohort of patients suggest that pre-operative ctDNA amounts may provide a useful marker of disease outcome in operable colorectal cancer.

DISCUSSION

These analyses provide an approach for non-invasive direct detection of patients with early stage disease across common cancer types. A conceptual benefit of this approach is that detectable alterations in cfDNA are by definition clonal and therefore indicate an underlying population of cells with identical somatic mutations. This high degree of specificity is one of

the potential benefits of ctDNA detection compared to other blood-based biomarkers which may be increased in other normal tissues in patients without cancer.

Although ctDNA analyses have raised the possibility of direct detection of patients with early stage disease (13, 33), the de novo identification of somatic alterations has remained a major challenge for development of early detection approaches. The analytical performance characteristics of the TEC-Seq method suggest that it may be suitable for such analyses. Other methods have been used for analyses of cell-free DNA in late stage cancer patients (13–22), but the specificity and sensitivity of these methods may limit their applicability for detection of early stage disease. A variety of experimental and bioinformatic aspects may contribute to the high specificity of the TEC-Seq method compared to previous approaches, including deep sequencing (>30,000 fold coverage), use of a small number of adaptors with long pre-specified barcodes, and multiple bioinformatic filtering steps comprising error correction, removal of repetitive sequences and mapping artifacts, and identification and removal of germline and hematopoietic sequences.

Using the TEC-Seq approach, no tumor-derived alterations were identified in plasma of the healthy individuals in our study. Although the average age of the healthy cohort was younger than the cancer patients analyzed, this corresponds to an age at which cancer screening may be initiated. Likewise, the concordance between liquid and tumor biopsies was high and suggested that liquid biopsies may have advantages for detection of heterogeneous tumor-specific alterations that may be missed by tissue biopsies. In the colorectal cancer case analyzed through multiple tissue biopsies, we showed that heterogeneous alterations appeared to have lower amounts of ctDNA and may explain the wide range of mutant allele fractions in ctDNA in the same individuals. One concern is that clonal hematopoietic changes may be confounded with heterogeneous tumor-specific mutations (25, 27) and lead to over-diagnoses. Large-scale studies of cell-free alterations in healthy individuals will be important to catalog the frequency and spectrum of these changes in the circulation. The higher fraction of healthy individuals in whom we detected mutations in blood cell proliferation genes compared to previous studies (25–27) will require further investigation to see if these alterations become clinically relevant over time. Given the different tumors that could potentially be detected, imaging and other diagnostic studies will be needed to complement any positive ctDNA analysis to appropriately identify the tumor of origin. In the future, ctDNA mutations combined with other molecular characteristics (34) may be helpful to identify the source of occult lesions.

Achieving effective sensitivity in ctDNA analyses has similarly presented a major technical hurdle. The high rate of conversion of cfDNA molecules in TEC-Seq libraries, combined with the use of endogenous as well as a limited number of exogenous barcodes, has increased the number of molecules that can be evaluated through next generation sequencing approaches. The parallel analysis of 55 cancer driver genes in this approach has the advantage of detecting a high fraction of tumors without prior knowledge of the genetic make-up of these cancers. The ability to detect multiple alterations in each case can increase sensitivity even when an individual mutation may not be detected. The inclusion of additional genes in larger panels could increase sensitivity, although this would be associated with higher sequencing costs. In some cancer types we have surpassed the theoretical

estimate of cases that could be detected, potentially due to the limited number of cases analyzed or underestimates of mutation prevalence in existing databases. Overall sensitivity may be further improved by deeper sequencing, improved error correction methods, larger blood volumes, and repeated testing at regular intervals, but it is likely that biologic characteristics of ctDNA will ultimately determine the ability to detect very small tumors or pre-neoplastic lesions.

Despite these limitations, the ability to detect half to three quarters of patients with early stage colorectal, ovarian, lung, or breast cancer provides opportunities for early detection and intervention. The survival difference between late stage and early stage disease in these cancers accounts for over a million lives world-wide each year (1). Circulating tumor DNA-based cancer detection followed by appropriate intervention at earlier stages in even a fraction of individuals would likely dwarf the current health impact of most late-stage cancer therapies. Additionally, as we observed in colorectal cancer, the amount and type of ctDNA at the time of diagnosis may provide additional insight related to patient prognosis that could inform further clinical intervention. Although screening for ctDNA will require larger validation studies, the success of cancer screening efforts based on other molecular tests (35) suggests that these approaches could in principle be implemented on a broad scale.

MATERIALS AND METHODS

Study design

This study presents a retrospective analysis of cfDNA using an ultrasensitive sequencing and analysis platform to detect somatic sequence alterations in early stage cancers. We analyzed 250 plasma samples from 244 individuals including 44 healthy individuals as well as 200 patients with colorectal (n=42), lung (n=71), ovarian (n=42), or breast (n=45) cancer over a range of stages, with most patients exhibiting localized disease. We estimated that analysis of at least 42 patients for each tumor type would provide a 96% power to detect 50% of cases with a 95% confidence interval of 35% to 65%. We evaluated the sensitivity and specificity of the TEC-Seq method to detect ctDNA in early stage patients without prior knowledge of alterations in their tumors. We detected sequence alterations in hematopoietic expansion genes in healthy individuals, established the sensitivity of the approach for detecting tumor-specific alterations in the blood of cancer patients, evaluated concordance between alterations identified in cfDNA and tumor samples from the same patients, and assessed whether pre-operative ctDNA can serve as a marker of patient outcome.

Patient and sample characteristics

Plasma samples from healthy individuals as well as plasma and tissue samples from patients with breast, lung, ovarian, or colorectal cancers were obtained from ILSBio/Bioreclamation, Aarhus University, the Academic Medical Center of the University of Amsterdam, and UCSD. All samples were obtained under Institutional Review Board (IRB) approved protocols with informed consent for research use at participating institutions.

Plasma samples from healthy individuals were obtained at the time of routine screening, including for colonoscopies or pap smears. Individuals were considered healthy if they had

no prior history of cancer and negative screening results. Plasma samples from individuals with colorectal, lung, ovarian, and breast cancer were obtained at the time of diagnosis, before tumor resection. Serially collected plasma samples from lung cancer patients were collected over a course of treatment during which the patients experienced stable or progressive disease.

Matched FFPE or frozen tumor tissue and buffy coat (as a source of germline DNA) were obtained from patients whenever available. Tumor specimens were obtained from primary resection, with the exception of stage IV colorectal cancer patients with liver-only metastases, for whom the samples were obtained from the liver metastases. All tumor samples had ≥10% viable tumor cell content by histopathologic assessment. Clinical data for all patients included and sample data for the tissue types assayed in this study are listed in table S3.

Sample preparation and next-generation sequencing of cfDNA

Whole blood was collected in EDTA tubes and processed immediately or within 2 hours after storage at 4°C to separate plasma and cellular components by centrifugation at 800 g for 10 minutes at 4°C. Plasma was centrifuged a second time at 18,000 g at room temperature to remove any remaining cellular debris and stored at −80°C until the time of DNA extraction. DNA was isolated from plasma using the Qiagen Circulating Nucleic Acids Kit (Qiagen GmbH) and eluted in LoBind tubes (Eppendorf AG). Concentration and quality of cfDNA were assessed using the Bioanalyzer 2100 (Agilent Technologies).

TEC-Seq next-generation sequencing cell-free DNA libraries were prepared from 5 to 250 ng of cfDNA. Genomic libraries were prepared using the NEBNext DNA Library Prep Kit for Illumina (New England Biolabs (NEB)) with four main modifications to the manufacturer's guidelines: 1) The library purification steps used the on-bead Ampure XP approach to minimize sample loss during elution and tube transfer steps (36), 2) NEBNext end-repair, A-tailing, and adapter ligation enzyme and buffer volumes were adjusted as appropriate to accommodate the on-bead Ampure XP purification strategy, 3) a pool of 8 unique lllumina dual index adapters with 8 bp barcodes was used in the ligation reaction instead of the standard lllumina single or dual index adapters with 6 bp or 8 bp barcodes, respectively, and 4) cell-free DNA libraries were amplified with Hotstart Phusion Polymerase. Incorporation of these modifications improved conversion efficiency from 13.4% before modifications to 34.1% in validation analyses of 38 cases incorporating these changes. Analysis of plasma samples from healthy individuals and cancer patients revealed a conversion efficiency of 40%, with a significant correlation between input DNA amount and the number of distinct molecules analyzed (Pearson correlation: $r=0.55$, 95% CI=0.46–0.64, $p<0.0001$, fig. S9).

Briefly, cell-free DNA was combined with End-Repair Reaction Buffer (NEB) and End-Repair Enzyme Mix (NEB) and incubated for 30 minutes at 20°C. The end-repair reaction was purified with Agencourt AMPure XP Beads (Beckman Coulter). A-tailing was performed by adding 6 μl of dA Tailing Reaction Buffer (NEB) and 3.6 μl of Klenow (NEB) to the end-repaired cfDNA and incubating for 30 minutes at 37°C. A-tailed cfDNA was purified using Agencourt AMPure XP Buffer (Beckman Coulter). Adaptor oligonucleotides

containing the TEC-Seq dual index pools and Quick T4 DNA Ligase (NEB) were mixed with A-tailed, on-bead cfDNA and incubated for 15 min at 20°C. Ligated cfDNA was purified with two rounds of Agencourt AMPure XP Buffer. The cfDNA library was amplified using Phusion Hot Start DNA polymerase (Thermo Fisher Scientific) and PCR primers published for the Nextera DNA library prep kit: 5'-AATGATACGGCGACCACCGA and 5'-CAAGCAGAAGACGGCATACGA (Illumina Inc.). For each genomic library, PCR reactions contained 2 μl of cfDNA library, 15.5 μl of H2O, 1.25 pl of dimethyl sulfoxide, 5.0 pl of 5X Phusion HF Buffer, 0.5 pl of dNTP mix containing 10 mM of each dNTP (Life Technologies), 0.5 pl of each primer, and 0.25 pl of Hotstart Phusion Polymerase. The following PCR conditions were used: 98°C for 30 seconds; 12 cycles of 98°C for 10 seconds, 60°C for 30 seconds, and 72°C for 30 seconds; and 72°C for 5 minutes. Purification of the amplified cfDNA library was performed using Agencourt AMPure XP Beads. Concentration and quality of cfDNA libraries was assessed using the Bioanalyzer 2100 (Agilent Technologies).

Targeted capture was performed using the Agilent SureSelect reagents and a custom set of hybridization probes targeting 58 genes (table S1) per the manufacturer's guidelines. The captured library was amplified with HotStart Phusion Polymerase (New England Biolabs). The concentration and quality of captured cfDNA libraries was assessed on the Bioanalyzer 2100 using the DNA 1000 Kit (Agilent Technologies). TEC-Seq libraries were sequenced using 100 bp paired end runs on the Illumina HiSeq 2000/2500 (Illumina).

Sample preparation and next-generation sequencing of tumor-normal pairs

Sample preparation, library construction, targeted capture, next-generation sequencing, and bioinformatic analyses of tumor and normal samples were performed as previously described (23, 37). Briefly, DNA was extracted from matched FFPE or frozen tumor tissue and buffy coat samples using the Qiagen DNA FFPE Tissue Kit or Qiagen DNA Blood Mini Kit (Qiagen GmbH). Genomic DNA from tumor and normal samples was fragmented and used for Illumina TruSeq library construction (Illumina) as previously described (23, 37). Targeted regions of interest were captured using Agilent SureSelect in-solution capture reagents and a custom targeted panel for genes of interest according to the manufacturer's instructions (Agilent). Paired-end sequencing, resulting in 150 bases from each end of the fragment for targeted libraries, was performed using the Illumina MiSeq (Illumina).

Analyses of Next-Generation Sequencing Data from Cell-Free DNA

Primary processing of next-generation sequence data for cell-free DNA samples was performed using Illumina CASAVA software (v1.8), including demultiplexing and masking of dual index adapter sequences. Sequence reads were aligned against the human reference genome (version hg18 or hg19) using Novoalign with additional realignment of select regions using the Needleman-Wunsch method (23). The positions of the alterations we have identified have not been affected by the different genome builds.

Next, candidate somatic mutations, consisting of point mutations, small insertions, and deletions were identified using VariantDx (23) across the targeted regions of interest. VariantDx examined sequence alignments of cell-free DNA plasma samples while applying

filters to exclude alignment and sequencing artifacts. Specifically, an alignment filter was applied to exclude quality failed reads, unpaired reads, and poorly mapped reads in the plasma. A base quality filter was applied to only include bases with reported phred quality score > 30.

A mutation identified in cell-free DNA was considered a candidate somatic mutation only when: (i) Three distinct paired reads contained the mutation in the plasma (each redundantly sequenced at least three times) with a distribution of start and cycle positions when compared to the reference genome, and the number of distinct paired reads containing a particular mutation in the plasma was at least 0.1% of the total distinct read pairs; or (ii) Four distinct paired reads contained the mutation in the plasma (each redundantly sequenced at least four times) with a distribution of start and cycle positions when compared to the reference genome, and the number of distinct paired reads containing a particular mutation in the plasma was at least 0.05% and less than 0.1% of the total distinct read pairs; and (iii) the mismatched base was not present in >1% of the reads in a panel of unmatched normal samples as well as not present in a custom database of common germline variants derived from dbSNP.

Mutations arising from misplaced genome alignments, including paralogous sequences, were identified and excluded by searching the reference genome. Candidate somatic mutations were further filtered based on gene annotation to identify those occurring in protein coding regions. Functional consequences were predicted using snpEff and a custom database of CCDS, RefSeq, and Ensembl annotations using the latest transcript versions available on hg18 and hg19 from UCSC (https://genome.ucsc.edu/). Predictions were ordered to prefer transcripts with canonical start and stop codons and CCDS or Refseq transcripts over Ensembl when available. Finally, mutations were filtered to exclude intronic and silent changes, while retaining mutations resulting in missense mutations, nonsense mutations, frameshifts, or splice site alterations.

Candidate alterations were defined as somatic hot-spots if the nucleotide change and amino acid change were identical to an alteration observed in ≥ 20 cancer cases reported in the COSMIC database. Alterations that were not hot-spots were retained only if either (i) seven or more distinct paired reads contained the mutation in the plasma and the number of distinct paired reads containing a particular mutation in the plasma was at least 0.1% and less than 0.2% of the total distinct read pairs, or (ii) six or more distinct paired reads contained the mutation in the plasma and the number of distinct paired reads containing a particular mutation in the plasma was at least 0.2% of the total distinct read pairs.

Candidate mutations were further limited through identification and removal of common germline variants present in ≥ 25% of reads or < 25% of reads if the variant was recurrent and the majority of alterations at that position had a mutant allele fraction ≥ 25% (table S6). Variants known to be at a somatic hot-spot position or producing a truncating mutation in a tumor suppressor gene were not excluded as germline changes. Because of the high frequency of mutations in specific genes and the possible confounding between somatic and germline changes, we limited analyses in the *APC* gene to frameshift or nonsense mutations, and in *KRAS, HRAS*, and *NRAS* to positions to 12, 13, 61, and 146. Finally, we excluded

hematopoietic expansion-related variants that have been previously described, including those in *DNMT3A, IDH1*, and *IDH2* and specific alterations within *ATM* (residue 3008), *GNAS* (residue 202), or *JAK2* (residue 617) (table S1) (25–27).

To evaluate the sensitivity of TEC-Seq approach using dilutions of cell lines with known mutations, we used a mixture of cell lines obtained from ATCC and combined in ratios to reflect the mutant allele frequency. The cell lines in the mutant pool included CCL-237, CRL-2158, CRL-2547, CRL-7585, CRL-9068, CRL-2177, CCL-231, CRL-2871, CRL-5908, CRL-5908, CCL-224, and CRL-5894. To evaluate sensitivity and specificity, we used dilutions of a cell line (CGBR4C, CRL-2338) which had been previously sequenced to examine both mutant and wild-type bases in the 58 genes in our panel (30). For analyses at all dilutions, we considered those alterations where the mutant allele fraction was expected to be at 0.1% or higher. To calculate the per base error rate for conventional sequencing in samples from healthy individuals, we summed the number of false positive calls at each genomic position and divided this by the total coverage at that base for the 44 healthy individuals. The upper limit of the per base error rate of TEC-Seq was determined by assuming one alteration per base if no error was identified and dividing by the total coverage at each base for the 44 healthy individuals analyzed.

To compare the TEC-Seq bioinformatic approach to iDES enhanced CAPP-Seq, we used the bioinformatic components of iDES combined with the requirement of multiple distinct read families based on endogenous and exogenous barcodes (19, 21) (https://cappseq.stanford.edu/ides/).

Analyses of Next-Generation Sequencing Data from tumor-normal pairs

Primary processing of next-generation sequencing data from tumor-normal pairs and identification of putative somatic mutations was completed using Illumina CASAVA (Consensus Assessment of Sequence and Variation) software V1.8 and VariantDx custom software, respectively, as previously described (23).

Statistical Analyses

We used a variety of methods for determining significance. To test the linear association between expected and observed mutant allele fractions (fig. S2), we used Pearson's product moment correlation coefficient. To quantify the difference in mean error rate by genomic position for conventional sequencing and TEC-Seq, we used a paired (by genomic position) t-test assuming equal variances. Differences in means of unpaired (independent) samples were tested using a two sample t-test assuming equal variances (such as for comparisons involving the concentration of cfDNA in plasma between healthy and cancer populations). To assess whether high mutant allele fractions are associated with patient outcomes, we defined patients with high mutant allele fractions as those with values more than three median absolute deviations from the median mutant allele fraction observed in 31 CRC patients analyzed. We used a median absolute deviation rather than a standard deviation because the mutant allele fractions were skewed and the median absolute deviation provides a more robust-to-outlier measure of the standard deviation. We compared progression-free survival and overall survival between patients with low and high mutant allele fraction using

the log-rank test in univariate analyses and the Cox proportional hazards in multivariate analyses (38, 39).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank members of our laboratories for critical review of the manuscript.

Funding: This work was supported in part by US National Institutes of Health grants CA121113, CA006973, CA180950, the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, the Stand Up to Cancer-Dutch Cancer Society International Translational Cancer Research Dream Team Grant (SU2C-AACRDT1415), the Commonwealth Foundation, The Cigarette Restitution Fund, the Burroughs-Wellcome Fund and the Maryland-Genetic, Epidemiology and Medicine Training Program (MD-GEM), IASLC/Prevent Cancer Foundation, the Danish Council for Independent Research (11–105240), the Danish Council for Strategic Research (1309–00006B), the Novo Nordisk Foundation (NNF14OC0012747), and the Danish Cancer Society (R133-A8520 and R40-A1965–11-S2). Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

Competing interests:

J.P. and V.E.V. are inventors on patent applications (62/501,686 and 62/516,009) submitted by Johns Hopkins University related to detection of cancer. M.S., L.A.D., and V.E.V. are inventors on a patent application (no. 62/422,355) submitted by Personal Genome Diagnostics related to non-unique barcodes for genotyping. V.E.V. is a founder of Personal Genome Diagnostics, is a member of its Scientific Advisory Board and Board of Directors, and owns Personal Genome Diagnostics stock, which is subject to certain restrictions under university policy. V.E.V. is also on the Scientific Advisory Board for Ignyta. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict of interest policies. L.A.D. is a founder of Personal Genome Diagnostics and Papgene and stock owner for both entities, a member of the Personal Genome Diagnostics Board of Directors, and a consultant for Personal Genome Diagnostics, Merck, and Cell Design Laboratories. V. Adleff is a consultant for Personal Genome Diagnostics. Data and materials availability: Data have been deposited at the European Genome-phenome Archive, which is hosted at the European Bioinformatics Institute, under study accession EGAS00001002577.

REFERENCES AND NOTES

1. Torre LA et al., Global cancer statistics, 2012. CA Cancer J Clin 65, 87 (2015). [PubMed: 25651787]
2. W. H. Organization, Guide to Cancer Early Diagnosis. Guide to Cancer Early Diagnosis, (2017).
3. Mazzucchelli R, Colanzi P, Pomante R, Muzzonigro G, Montironi R, Prostate tissue and serum markers. Advances in clinical pathology : the official journal of Adriatic Society of Pathology 4, 111 (2000). [PubMed: 11080790]
4. Ruibal Morell A, CEA serum levels in non-neoplastic disease. The International journal of biological markers 7, 160 (1992). [PubMed: 1431339]
5. Galli C, Basso D, Plebani M, CA 19–9: handle with care. Clinical chemistry and laboratory medicine 51, 1369 (2013). [PubMed: 23370912]
6. Sikaris KA, CA125--a test with a change of heart. Heart, lung & circulation 20, 634 (2011).

7. Wanebo HJ et al., Preoperative carcinoembryonic antigen level as a prognostic indicator in colorectal cancer. N Engl J Med 299, 448 (1978). [PubMed: 683276]
8. Lin JS et al., in Screening for Colorectal Cancer: A Systematic Review for the U.S. Preventive Services Task Force. (Rockville (MD), 2016).
9. Zauber AG, The impact of screening on colorectal cancer mortality and incidence: has it really made a difference? Dig Dis Sci 60, 681 (2015). [PubMed: 25740556]
10. Mandel P, Metais P, [Not Available]. Comptes rendus des seances de la Societe de biologie et de ses filiales 142, 241 (1948). [PubMed: 18875018]
11. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA, Isolation and characterization of DNA from the plasma of cancer patients. European journal of cancer & clinical oncology 23, 707 (1987). [PubMed: 3653190]
12. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ, Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 37, 646 (1977). [PubMed: 837366]
13. Bettegowda C et al., Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6, 224ra24 (2014).
14. Leary RJ et al., Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 4, 162ra154 (2012).
15. Sausen M et al., Clinical implications of genomic alterations in the tumour and circulation of pancreatic cancer patients. Nat Commun 6, 7686 (2015). [PubMed: 26154128]
16. Dawson SJ et al., Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 368, 1199 (2013). [PubMed: 23484797]
17. Forshew T et al., Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 4, 136ra68 (2012).
18. Murtaza M et al., Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 497, 108 (2013). [PubMed: 23563269]
19. Newman AM et al., An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 20, 548 (2014). [PubMed: 24705333]
20. Kim ST et al., Prospective blinded study of somatic mutation detection in cell-free DNA utilizing a targeted 54-gene next generation sequencing panel in metastatic solid tumor patients. Oncotarget 6, 40360 (2015). [PubMed: 26452027]
21. Newman AM et al., Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 34, 547 (2016). [PubMed: 27018799]
22. Lanman RB 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). [PubMed: 26474073]
23. Jones S et al., Personalized genomic analyses for cancer mutation discovery and interpretation. Sci Transl Med 7, 283ra53 (2015).
24. Forbes SA et al., COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res 38, D652 (2010). [PubMed: 19906727]
25. Xie M et al., Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 20, 1472 (2014). [PubMed: 25326804]
26. McKerrell T et al., Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep 10, 1239 (2015). [PubMed: 25732814]
27. Genovese G et al., Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 371, 2477 (2014). [PubMed: 25426838]
28. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B, Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 108, 9530 (2011). [PubMed: 21586637]
29. Hindson BJ et al., High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical chemistry 83, 8604 (2011). [PubMed: 22035192]
30. Sjoblom T et al., The consensus coding sequences of human breast and colorectal cancers. Science 314, 268 (2006). [PubMed: 16959974]

31. Wilson CH, McIntyre RE, Arends MJ, Adams DJ, The activating mutation R201C in GNAS promotes intestinal tumourigenesis in Apc(Min/+) mice through activation of Wnt and ERK1/2 MAPK pathways. Oncogene 29, 4567 (2010). [PubMed: 20531296]
32. Locker GY et al., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. J Clin Oncol 24, 5313 (2006). [PubMed: 17060676]
33. Haber DA, Velculescu VE, Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. Cancer discovery 4, 650 (2014). [PubMed: 24801577]
34. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J, Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell 164, 57 (2016). [PubMed: 26771485]
35. Toes-Zoutendijk E et al., Real-Time Monitoring of Results During First Year of Dutch Colorectal Cancer Screening Program and Optimization by Altering Fecal Immunochemical Test Cut-Off Levels. Gastroenterology 152, 767 (2017). [PubMed: 27890769]
36. Fisher S et al., A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12, R1 (2011). [PubMed: 21205303]
37. Sausen M et al., Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. Nat Genet 45, 12 (2013). [PubMed: 23202128]
38. Harrington DP and Fleming TR A class of rank test procedures for censored survival data. Biometrika 69, 553–566 (1982).
39. Andersen P and Gill R Cox's regression model for counting processes, a large sample study. Annals of Statistics 10, 1100–1120 (1982).

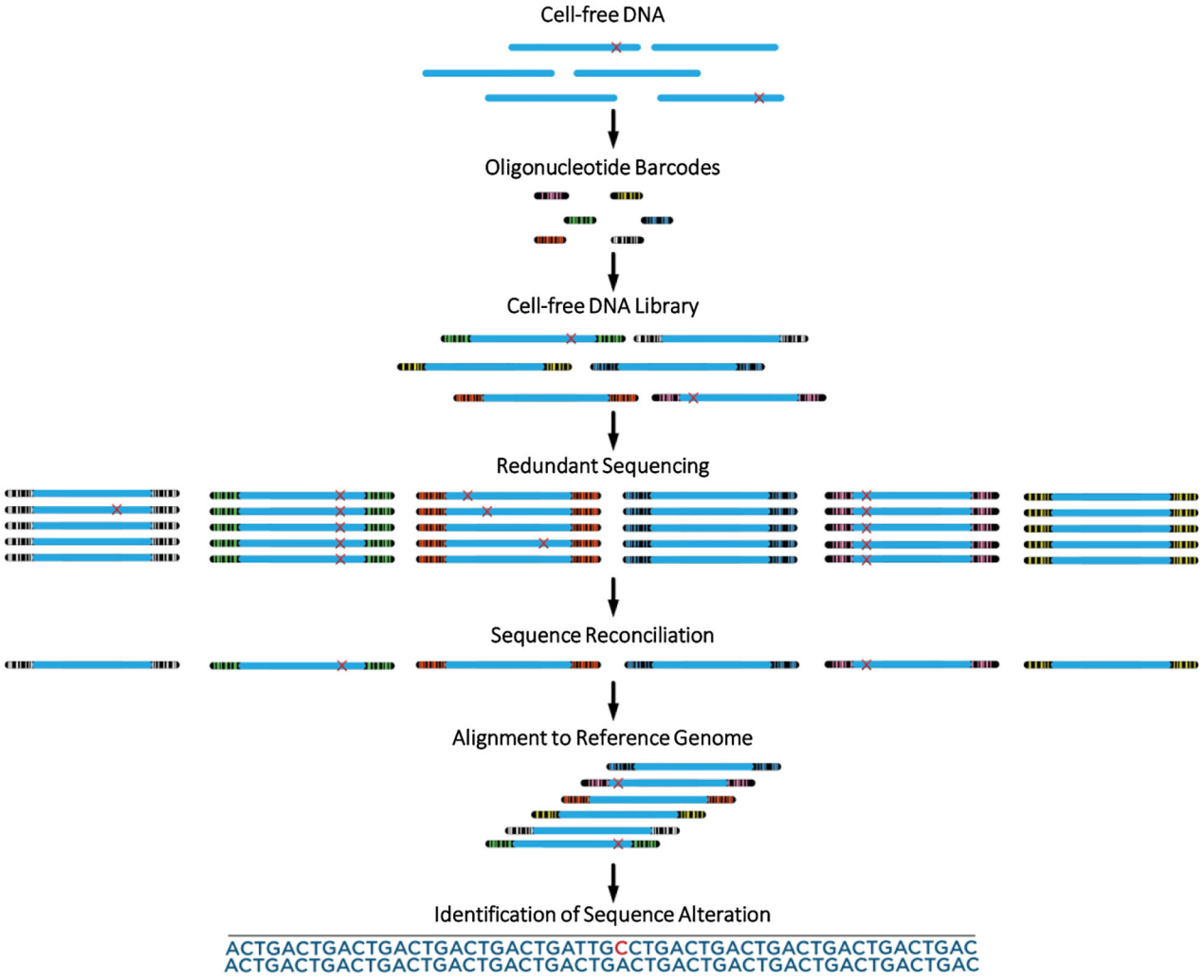

Figure 1. Schematic of TEC-Seq method.

Cell-free DNA is extracted from blood and converted to a genomic library through ligation of a pool containing a small number of dual index barcode adapters. The resulting cfDNA library is captured and redundantly sequenced to produce multiple duplicates of each DNA fragment. Sequence reconciliation among duplicate fragments identifies alterations present in identical DNA molecules with the same start and end position and exogenous barcodes. Alignment to the reference genome of multiple distinct molecules containing identical redundant changes is used to identify *bona fide* alterations.

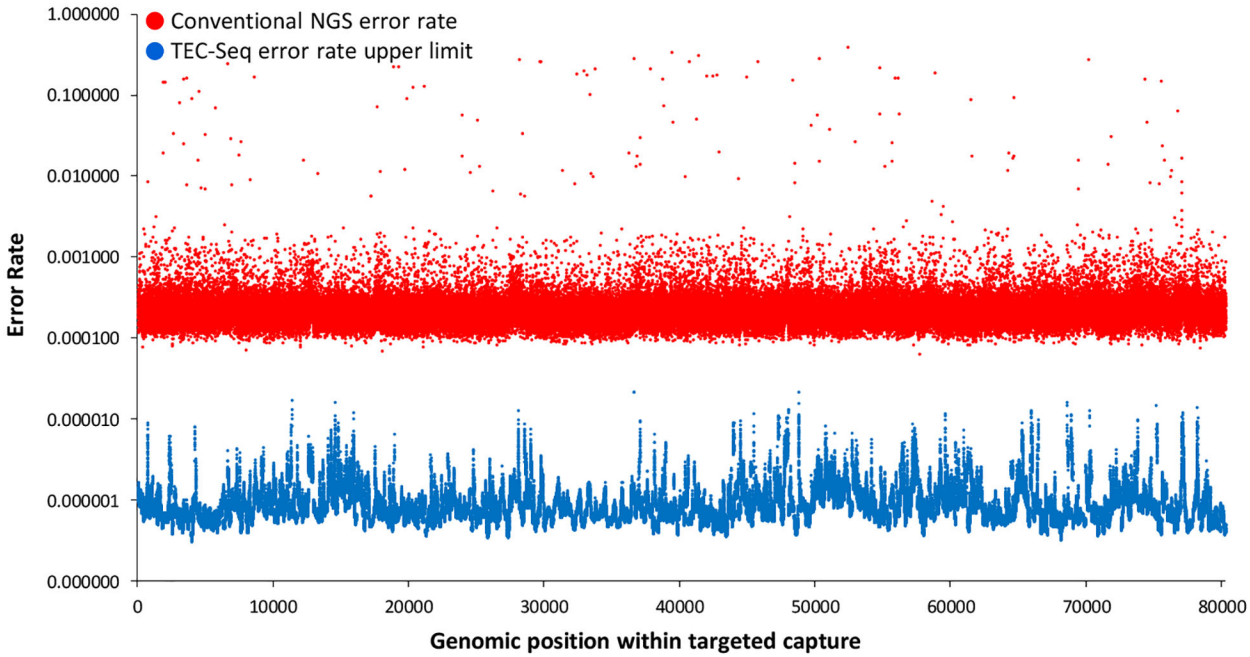

Figure 2. TEC-Seq error correction.

Sequencing error rates of conventional next generation sequencing and theoretical upper limit for TEC-Seq are indicated at each base in the captured regions of interest ($P < 0.0001$, paired t test). Error rates are determined by identifying the number of alterations at each base (or assuming one alteration per base if no error was identified) divided by the total coverage at each base among the 44 healthy individuals analyzed.

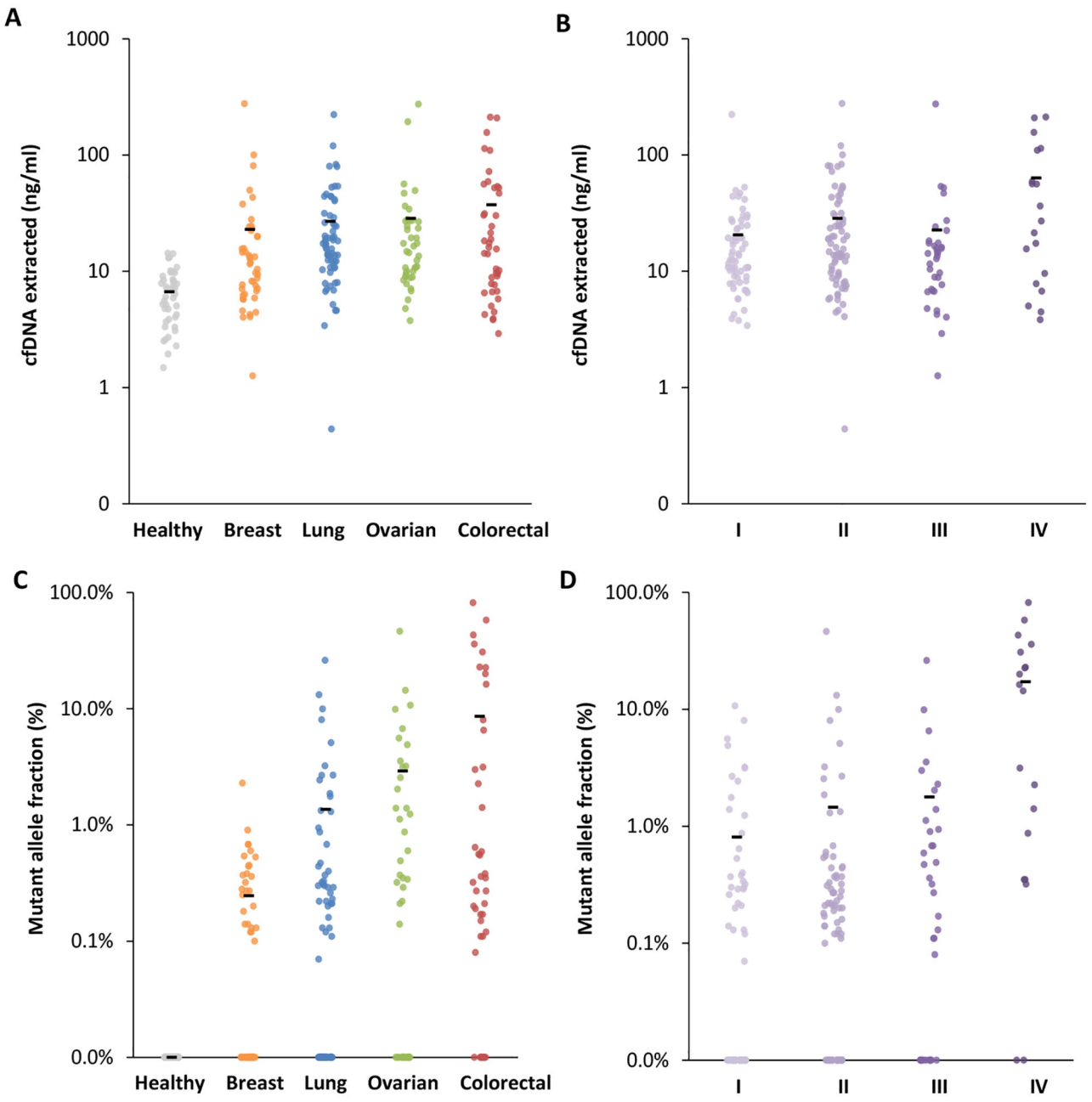

Figure 3. cfDNA and ctDNA in healthy individuals and patients with cancer.

Amount of cfDNA extracted from all heathy individuals and patients with different cancer types (ng/ml) (**A**) and from cancer patients of different stages (**B**). Mutant allele fraction (%) of ctDNA detected in heathy individuals and patients with different cancer types **(C)** and in cancer patients of different stages **(D)**. Means for each group are represented by the black bars in the columns analyzed. In patients for whom multiple alterations were detected, the highest value is indicated. Clinical characteristics of patients and stages are indicated in table S3.

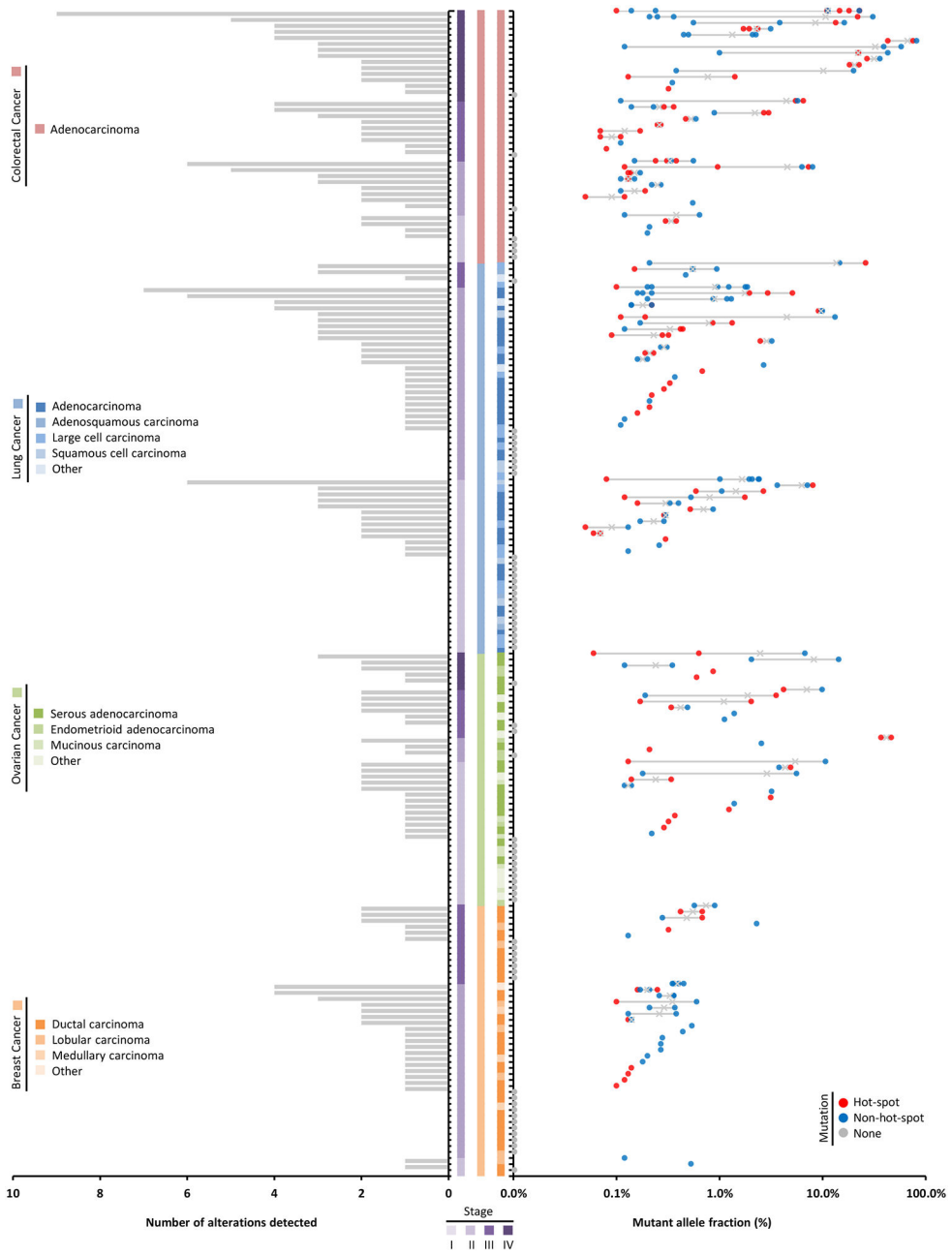

Figure 4. ctDNA in patients with breast, colorectal, lung, and ovarian cancer.

Patients (n = 194) are each represented by a tick mark. **Left:** Bar chart shows the number of alterations detected for each case. **Center:** Stage, cancer type, and histopathological subtype are represented by colored vertical bars. **Right:** Mutant allele fractions for each alteration detected per patient are indicated with an 'x' at the mean. Alterations are colored based on hot-spot status and whether any alterations were detected in the case.

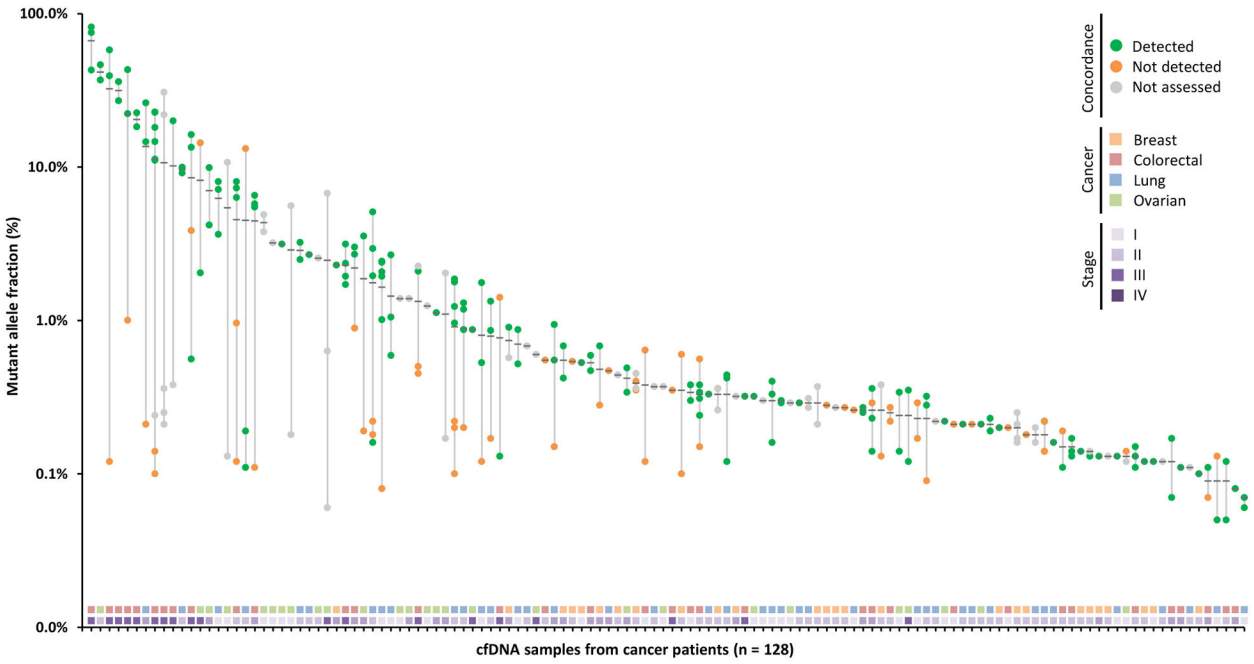

Figure 5. Concordance between alterations in plasma and tissue.

Mutant allele fractions observed in the plasma are indicated for each alteration identified with a black bar at the mean. Presence of alterations in matched tumor specimens is indicated with green dots, whereas non-concordant alterations are indicated in orange and those not assessed in gray. Stage and cancer type for each patient are plotted in the two horizontal tracks at the bottom of the figure.

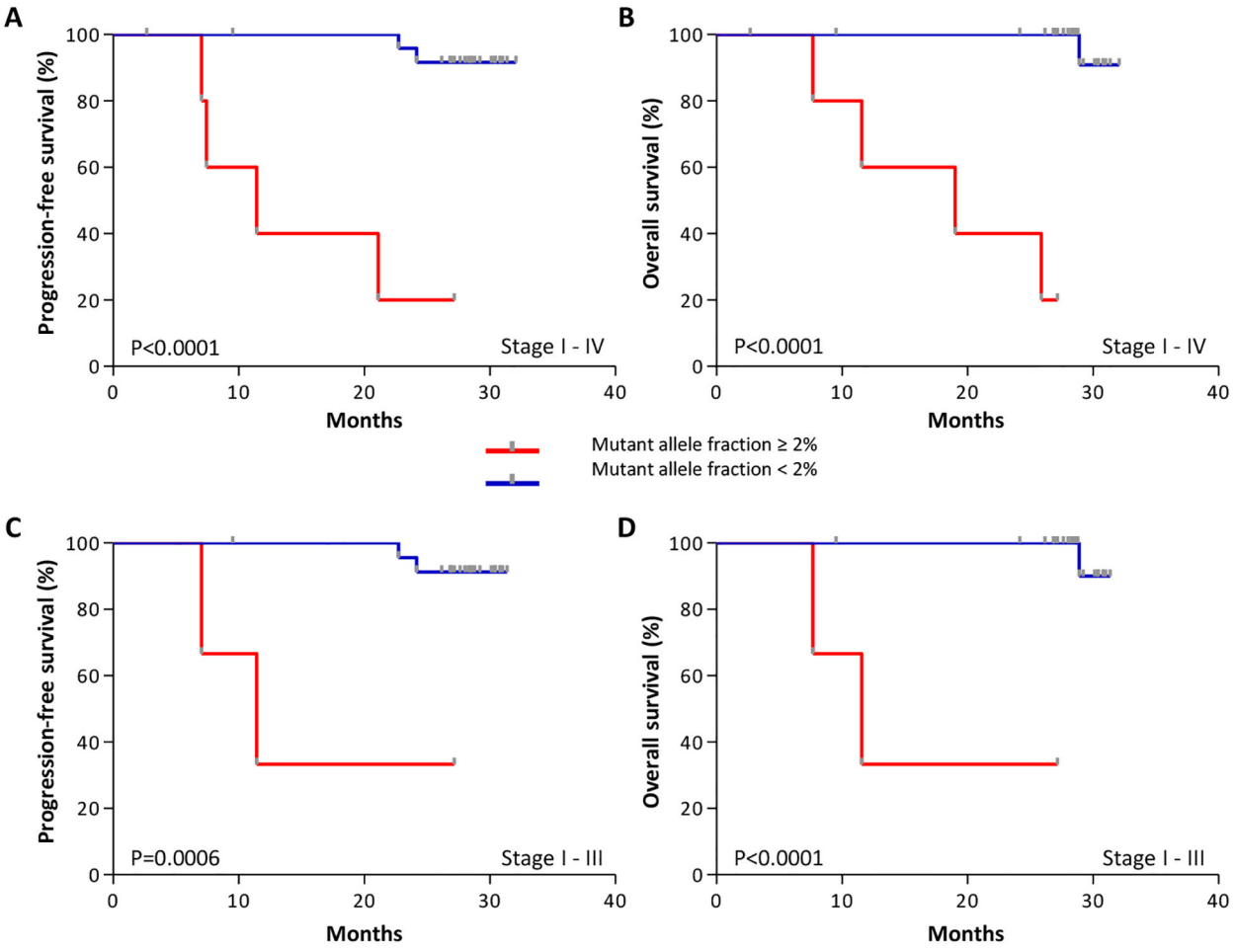

Figure 6. Pre-operative ctDNA amounts and outcome in colorectal cancer patients.

Kaplan-Meier curves depict progression-free survival (**A**, Log-rank test $p < 0.0001$) and overall survival (**B**, Log-rank test $p < 0.0001$) of 31 colorectal cancer patients, stage I - IV, stratified based on a ctDNA mutant allele fraction threshold of 2%. Kaplan-Meier analyses of the 27 patients with stage I - III disease for progression-free survival (**C**, Log-rank test $p = 0.0006$) and overall survival (**D**, Log-rank test $p < 0.0001$) were performed using the same threshold to examine the association of ctDNA with outcome in patients without stage IV disease.

Table 1.

Cancer cases containing alterations in driver genes.

Tissue Type	Cases in COSMIC	Detectable Cases*	Detectable Fraction
Breast	1,002	719	72%
Colorectal	1,248	1,071	86%
Lung	1,198	932	78%
Ovarian	647	524	81%

* Detectable cases indicate those with at least one alteration in the cancer driver genes analyzed (table S1).

Table 2.

Cancer patients detected using TEC-Seq

Cancer Type	Patients (n)	Patients with ctDNA Alterations (n)	Fraction of patients with ctDNA Alterations (%)
Colorectal			
I	8	4	50%
II	9	8	89%
III	10	9	90%
IV	15	14	93%
I-IV	42	35	83%
Lung			
I	29	13	45%
II	32	23	72%
III	4	3	75%
IV	6	5	83%
I-IV	71	44	62%
Ovarian			
I	24	16	67%
II	4	3	75%
III	8	6	75%
IV	6	5	83%
I-IV	42	30	71%
Breast			
I	3	2	67%
II	29	17	59%
III	13	6	46%
IV	0	NA	NA
I-IV	45	25	56%
All			
I, II	138	86	62%
III, IV	62	48	77%
I-IV	200	134	67%