

Published in final edited form as:

Sci Transl Med. 2012 November 28; 4(162): 162ra154. doi:10.1126/scitranslmed.3004742.

Detection of Chromosomal Alterations in the Circulation of Cancer Patients with Whole-Genome Sequencing

Rebecca J. Leary^{1,*}, Mark Sausen^{1,*}, Isaac Kinde^{1,*}, Nickolas Papadopoulos¹, John D. Carpten², David Craig², Joyce O'Shaughnessy³, Kenneth W. Kinzler¹, Giovanni Parmigiani^{4,5}, Bert Vogelstein¹, Luis A. Diaz Jr.^{1,†}, and Victor E. Velculescu^{1,†}

¹Ludwig Center for Cancer Genetics and Howard Hughes Medical Institutions, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21287, USA.

²Translational Genomics Research Institute, Phoenix, AZ 85044, USA.

³Baylor Sammons Cancer Center, Texas Oncology, US Oncology, Dallas, TX 75246, USA.

⁴Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.

⁵Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA.

Abstract

Clinical management of cancer patients could be improved through the development of noninvasive approaches for the detection of incipient, residual, and recurrent tumors. We describe an approach to directly identify tumor-derived chromosomal alterations through analysis of circulating cell-free DNA from cancer patients. Whole-genome analyses of DNA from the plasma of 10 colorectal and breast cancer patients and 10 healthy individuals with massively parallel sequencing identified, in all patients, structural alterations that were not present in plasma DNA from healthy subjects. Detected alterations comprised chromosomal copy number changes and rearrangements, including amplification of cancer driver genes such as *ERBB2* and *CDK6*. The level of circulating tumor DNA in the cancer patients ranged from 1.4 to 47.9%. The sensitivity and specificity of this approach are dependent on the amount of sequence data obtained and are derived from the fact that most cancers harbor multiple chromosomal alterations, each of which is unlikely to be present in normal cells. Given that chromosomal abnormalities are present in nearly all human cancers, this approach represents a useful method for the noninvasive detection of human tumors that is not dependent on the availability of tumor biopsies.

Copyright 2012 by the American Association for the Advancement of Science; all rights reserved.

[†]To whom correspondence should be addressed. velculescu@jhmi.edu (V.E.V.); ldiaz1@jhmi.edu (L.A.D.).

*These authors contributed equally to this work.

Author contributions: R.J.L., M.S., I.K., J.D.C., D.C., J.O., B.V., L.A.D., and V.E.V. designed the study; R.J.L., M.S., I.K., G.P., B.V., and V.E.V. generated and analyzed the data; and R.J.L., M.S., I.K., N.P., G.P., K.W.K., B.V., L.A.D., and V.E.V. wrote the manuscript.

Competing interests: K.W.K., B.V., L.A.D., and V.E.V. are co-founders of Inostics and Personal Genome Diagnostics and are members of their Scientific Advisory Boards. K.W.K., B.V., L.A.D., and V.E.V. own Inostics and Personal Genome Diagnostics stock, which is subject to certain restrictions under University policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies. G.P. is on the scientific advisory board of Counsyl.

SUPPLEMENTARY MATERIALS www.sciencetranslationalmedicine.org/cgi/content/full/4/162/162ra154/DC1

INTRODUCTION

Abnormal chromosome content, or aneuploidy, is a common characteristic of tumors, which manifests at the earliest stages of tumorigenesis and increases throughout subsequent tumor development (1–4). In addition to losses and gains of entire chromosomes, alterations of chromosome arms, focal amplifications and deletions, and rearrangements are observed in nearly all cancer genomes. Analysis of such alterations in cancer began with karyotyping but is now generally carried out with molecular methods that can more easily assess genomes in a comprehensive manner. For example, an approach based on sequencing and enumerating genomic DNA tags, called digital karyotyping (DK), was developed for the analysis of copy number alterations on a genome-wide scale (5). Similar tag-based approaches have been adapted to next-generation sequencing methods (6, 7). Likewise, the analysis of chromosomal rearrangements with large-scale DNA sequencing approaches allows for high-resolution mapping of rearrangement breakpoints (3).

Given the universal nature of chromosomal alterations in human cancer and improved methods for detecting such changes, we wondered whether we could directly identify chromosomal alterations in the circulation of cancer patients. Sequencing analyses of chromosome content in the maternal circulation are now being used for detection of fetal aneuploidy (8, 9), although such approaches have not been evaluated for detection of chromosomal alterations in cancer patients. Similarly, analysis of circulating tumor DNA in patients with hematopoietic malignancies has been useful for the detection of known recurrent chromosomal rearrangements, such as those that involve the *BCR-ABL* oncogene and genes that encode immunoglobulin chains, T cell receptor subunits, and the retinoic acid receptor (10–15). More recently, analysis of tumor rearrangements has allowed the development of patient-specific biomarkers that can be evaluated in plasma for the detection of residual disease or for tumor monitoring (6, 16). However, such monitoring approaches rely on analyses of known alterations identified in resected tumors from the same patients and cannot be directly applied to the detection of new alterations in the circulation of patients in whom biopsied material is unavailable. Recurrent mutations, including those identified in oncogenes such as *KRAS*, have also been readily identified in a fraction of patients with solid tumors (17–19).

An alternative to these approaches is the identification of de novo tumor-derived chromosomal alterations through massively parallel direct sequencing of DNA from the circulation of cancer patients. Such approaches would be applicable to more patients than those that rely on recurrent oncogene alterations and could theoretically permit noninvasive detection of nearly all cancer types. Herein, we compare whole-genome analyses of DNA from the plasma of late-stage cancer patients to healthy individuals with massively parallel sequencing and detect structural alterations specific to patients.

RESULTS

Overview

A schematic of our approach to examine chromosomal abnormalities directly in the plasma of cancer patients is illustrated in Fig. 1. As a proof-of-principle analysis, we obtained 4 to 18 ml of plasma from each of 10 healthy individuals (N1 to N10), 7 patients with colorectal cancer (CRC11 to CRC17), and 3 patients with breast cancer (BR1 to BR3) (table S1). Plasma DNA was purified and used to generate paired-end libraries for whole-genome sequencing, and each library was analyzed on two lanes of an Illumina HiSeq instrument (see Materials and Methods). An average of 249,378,422 distinct paired sequences [50 base pairs (bp) from each end] was obtained for each sample (Table 1). The resulting sequence

data from circulating DNA were analyzed for chromosome copy number changes and for intra- and interchromosomal rearrangements.

Analysis of chromosomal copy number changes

Losses or gains of specific chromosomal regions are a hallmark of many cancers and have been used historically to identify tumor suppressors or oncogenes targeted by the alterations (20–22). Such chromosomal imbalances could be useful as markers of tumorigenesis because they should, in principle, alter the chromosomal representation of circulating DNA. To adapt DK to detect tumor-specific (somatic) chromosomal alterations in the plasma, we used the equivalent of one lane of HiSeq single-read sequence data per sample (average of 144,543,191 distinct reads) and applied a number of filtering steps to remove sources of variation that were not tumor-specific (see Materials and Methods). For example, we removed sequences that are known to vary in the germlines of normal individuals, because these could confound identification of somatic copy number changes. In addition, we applied a weight to each sequence read based on local GC content. This weighting has been shown to remove bias introduced by next-generation sequencing and allows for a more accurate assessment of chromosomal representation of the original genomic DNA (see Materials and Methods) (23, 24). The resulting weighted reads were used to determine the proportion of reads that mapped to specific regions in the genome (fig. S1). We performed analyses of entire chromosomes, of chromosome arms, and of sequential regions of specified sizes (for example, 10 Mb) throughout the genome. Although each of these approaches has certain advantages, we chose to analyze chromosome arms because these were frequently altered in breast and colorectal cancer samples previously analyzed for copy number alterations and would be expected to be altered in most human cancers (see Materials and Methods).

The proportion of sequences that represented each chromosome arm (excluding short arms of acrocentric chromosomes) was calculated, for each sample, by dividing the sum of the weighted reads mapping to that arm by the total number of weighted reads mapping to the reference genome. For the normal samples, N1 to N10, the proportion of chromosomal arm sequences ranged from 0.46 to 6.19%, closely corresponding to the expected fraction based on genomic size and the applied mapping criteria (table S2) ($R^2 = 0.95$; $P < 0.0001$, Pearson correlation). The variation among the normalized proportions of each chromosomal arm in the plasma from normal individuals was very low (average, $2.56 \pm 0.0065\%$; range of SD, $\pm 0.0025\%$ to $\pm 0.014\%$). These results are consistent with similar measurements of circulating DNA from the plasma of pregnant women carrying euploid fetuses (8, 9). In contrast, the normalized proportions of chromosomal arm sequences in the plasma of cancer patients were much more variable, ranging from 0.61- to 1.97-fold of the average found in the plasma of normal individuals (table S2).

To determine whether sequenced reads for an individual patient sample deviate from patterns in normal samples, we used the fraction of reads that mapped to each arm to calculate a z score. For each arm, the z score was calculated as the number of SDs from the mean of the reference plasma samples (N1 to N10). After Bonferroni correction for multiple comparisons of the 39 chromosomal arms, an absolute z score of ≥ 1.88 was determined to represent a statistically significant gain or loss of a chromosomal arm ($P < 0.05$, Student's t test). All chromosome arms of the 10 normal plasma samples had absolute z scores of less than 2.62. In contrast, plasma samples from all 10 of the cancer patients showed evidence of copy number gains or losses, with the highest absolute z score in each sample ranging from 13.3 to 434.4 (Fig. 2A).

Although such analyses could be used to evaluate specific chromosomal arms, a statistical approach that uses a combination of the most markedly altered chromosome arms in each

sample would be expected to provide a more sensitive measure of circulating tumor DNA. We analyzed previously obtained genome-wide copy number alterations detected from single-nucleotide polymorphism (SNP) arrays of 36 colorectal cancer samples (25) to determine how frequently tumors lost multiple chromosome arms. As shown in fig. S2, we found that the mean number of chromosome arms altered in these colorectal cancers was 21 and ranged from 5 to 35. Accordingly, we constructed a log-scale plasma aneuploidy score (PA score) based on the five chromosomes whose arms had the highest absolute *z* scores (see Materials and Methods). The PA score from the plasma of healthy individuals ranged from 0.1 to 2.4, and we calculated that a threshold PA score of 5.84 would provide a specificity of >99% (Student's *t* distribution) for indicating aneuploidy (Fig. 2B). All plasma samples from the colorectal and breast cancer patients had PA scores above this threshold, ranging from 11.9 to 41.5 (Fig. 2B and tables S1 and S2). The two plasma samples with the lowest PA scores represented those with the lowest amounts of circulating tumor DNA, and the PA score generally correlated with tumor burden ($R^2 = 0.53$; $P = 0.017$, Pearson correlation) (Fig. 2B, table S2, and Materials and Methods).

Analysis of rearrangements

The chromosomal instability that underlies large chromosomal gains and losses in tumorigenesis is associated with genomic rearrangements. Such somatic rearrangements are not present in normal cells in a clonal fashion and would therefore be expected to provide a highly sensitive and specific marker for the presence of clonal tumor-specific genetic alterations. We previously developed a technique, personalized analysis of rearranged ends (PARE), to identify rearranged breakpoints from tumor DNA for individual patients (see Materials and Methods). A challenge in adapting PARE to detection of rearrangements directly from plasma DNA is distinguishing the relatively few somatic rearrangements present in circulating tumor DNA from the much larger number of structural variants resulting from copy number variations in the germline of all individuals. To overcome this obstacle, we used bioinformatic filters that enriched for high-confidence somatic structural alterations while removing germline and artifactual changes. These filters included selecting paired-end reads that (i) mapped to different chromosomes or to the same chromosome but at large distances (≥ 30 kb) apart, (ii) spanned rearrangement junctions that were observed in multiple reads, (iii) contained sequenced rearrangement breakpoints, and (iv) mapped to genomic regions that did not contain known germline copy number variants or repeated sequences (26, 27) (fig. S1).

Paired-end Illumina sequence data for DNA in plasma samples from the 10 cancer patients and 10 healthy individuals (table S1) revealed a total of 65,402,563 aberrantly mapped paired-end reads, most of which were expected to result from either germline changes or mapping artifacts (26, 27). Application of the criteria described above identified 14 candidate rearrangements in 9 of the 10 plasma samples from cancer patients but none in the plasma samples from healthy individuals (Fig. 3). These rearranged sequences were evaluated further by polymerase chain reaction (PCR) amplifications across the rearrangement junctions in tumor and normal DNA from the same nine cancer patients, and all were confirmed to be present in the tumor samples but not in the matched normal DNA. Independent sequencing of the rearranged regions identified the expected rearrangement junctions in all nine cases analyzed. We further evaluated the specificity of the approach by analyzing more than 5.6 billion paired-end Illumina reads of normal lymphocyte DNA from 28 individuals (see table S3 and Materials and Methods). These analyses did not identify any candidate rearrangements, providing further evidence that the approach is highly specific to tumor-derived structural alterations.

Two of the rearrangement regions were associated with amplified genes known to be drivers of cancer development (Fig. 3). The chromosomal rearrangement identified in the CRC16

plasma sample corresponded to a breakpoint resulting from the amplification of the genetic locus that contains the *ERBB2* gene, which encodes HER2/neu, the target of trastuzumab (Herceptin) (Fig. 3 and fig. S3). The level of amplification in the plasma was estimated using DK to be 10.5-fold higher than that of plasma from a healthy individual. In addition, two of the four candidate rearrangements detected in plasma sample BR1 were associated with amplification of the cell cycle regulatory gene *cyclin-dependent kinase 6* (*CDK6*) (Fig. 3), where the level of amplification in the plasma was estimated using DK to be 6.5-fold. Inhibition of the CDK6 protein is currently being evaluated in clinical trials for breast and other cancer types ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01320592) identifier NCT01320592). These analyses show that amplified genes can be identified through detection of amplification-associated rearrangements by direct sequencing in the plasma.

Rearrangements detected directly in patient plasma may be used to develop PCR-based breakpoint-specific biomarkers for the analysis of circulating tumor DNA levels in plasma samples. Such breakpoint biomarkers as identified by PARE could be useful for providing a measure of circulating tumor DNA at the time of detection or for quantitative monitoring during therapy. Analyses of the nine patient plasma samples with plasma rearrangements were found to have concentrations of circulating mutant DNA ranging from 4.7 to 47.9% as determined by digital PCR with a PARE biomarker or as estimated using chromosomal representation (see Materials and Methods). The absence of detected rearrangements in the plasma of CRC12 could be a result either of the absence of structural alterations in the tumor DNA or of the failure to detect such rearrangements in the plasma DNA. To distinguish between these possibilities, we searched for rearrangements in the DNA of this patient's tumor using whole-genome sequencing (Table 1). We identified three rearrangements and showed that rearranged sequences were indeed present in the plasma DNA, the matching tumor, and the CRC12 plasma DNA library with PCR using primers that spanned the rearranged sequences. Using digital PCR, we estimated that the fraction of circulating mutant DNA in the plasma of this patient was 1.4%. These analyses suggested that obtaining additional sequence data would likely have identified the tumor rearrangements in this plasma sample.

Sensitivity of detection

As shown above, the sequence information obtained from circulating plasma DNA can be analyzed in an integrated fashion to obtain a comprehensive analysis of chromosome content and rearrangements of the same sample. For cases CRC14, CRC15, and CRC16, multiple samples, including plasma and the primary tumor, were available and could be directly examined for chromosomal abnormalities (6). These analyses allowed us to evaluate similarities between copy number alterations in plasma and the primary tumor and the sensitivity of detecting alterations in plasma during disease progression. For CRC15 and CRC16, we analyzed primary tumor samples and plasma from 33 and 50 months after surgery, respectively. For CRC14, the analyzed samples included plasma at the time of initial evaluation (CRC14-0), tumor tissue obtained from surgical resection 1 week later (CRC14-PT), plasma from 4 months after diagnosis subsequent to chemotherapy and resection of a metastatic lesion (CRC14-4), and plasma from 62 months after diagnosis at which time the tumor had recurred (CRC14). The analyses were normalized to the amount of sequence data obtained, and chromosomal representation analyses using DK of tumor and plasma DNA samples are shown in Fig. 4 and fig. S3. The copy number patterns observed for plasma samples at the time of initial evaluation (CRC14-0) and recurrence (CRC14) were markedly similar to those of the resected tumor (CRC14-PT), with significant losses of chromosomes 1p, 4q, 14q, and 22q and gains of chromosomes 13q and 20q (for each alteration, $P < 0.05$, Student's t test) (Fig. 4). Likewise, similar patterns of chromosomal alterations between plasma samples and primary tumors were observed for CRC15 and

CRC16 (table S2 and fig. S3). Overall, for samples for which tumors were also analyzed using our stringent bioinformatic criteria (CRC12, CRC14, CRC15, and CRC16), most of the detected structural alterations in the tumors were independently identified in the plasma (67 of 125, tables S2 and S4). For plasma sample CRC14-4, although the copy number graphs appeared similar to those derived from normal DNA, there was still significant alteration in chromosomal arm content (PA score = 6.6) (table S4 and Fig. 4). The fraction of mutant tumor DNA in CRC14-4 was previously measured using a PARE biomarker to be 0.3% (6), consistent with the predicted sensitivity of the approach using the available sequencing data (see Materials and Methods).

To evaluate the potential sensitivity and specificity of applying the DK approach to cell-free DNA for discriminating individuals with colorectal and breast cancer from healthy individuals, we performed receiver operating characteristic (ROC) analyses of simulated next-generation sequencing data from 81 cancer patients and 10,000 simulated normal controls based on data from 10 healthy individuals. For the 81 tumor cases, chromosomal arm alterations were determined using previously obtained genome-wide copy number information from SNP arrays of 36 colorectal cancers and 45 breast cancers (25). These analyses simulated mixtures of different concentrations of each tumor DNA with normal DNA (as would be expected in the circulation of cancer patients) using the experimentally observed means and SDs for each chromosome arm proportion for unaffected individuals (N1 to N10) (table S2, see Materials and Methods). Using the equivalent of one lane of HiSeq reads, these analyses suggested that tumor DNA concentrations at levels $\geq 0.75\%$ could be detected in the circulation of patients with breast and colorectal cancers with a sensitivity of $>90\%$ and specificity of $>99\%$ when the five chromosome arms with the largest absolute z scores were evaluated (Fig. 5). Analyses of the single most altered chromosome arm (17p) were much less sensitive, and a specificity of $>99\%$ could only be achieved with circulating tumor DNA concentrations of 5% or more. This single chromosome arm sensitivity is in accord with the results of previously described approaches for the detection of fetal trisomy 21 in maternal DNA (8, 9) (Fig. 5).

To determine the potential of this approach for detecting circulating tumor DNA at levels below 0.75%, we used simulations to predict the amount of sequencing required to achieve $>90\%$ sensitivity and $>99\%$ specificity using both copy number and rearrangement analyses (fig. S4). These simulations showed that the ability to detect chromosomal arm gains or losses increased proportionately as one over the square root of the number of reads obtained, similar to that expected from previous analyses of circulating fetal DNA (8, 9) (see Materials and Methods). On the other hand, the lower limit of detection of rearranged sequences decreased proportionately as one over the total number of reads obtained. This suggests that the sensitivity of PARE for detecting very low levels of circulating tumor DNA is higher than DK when assessing a large number of reads ($>10^9$). One advantage of an integrated approach using both methods is that the overall detection limit of the combined approaches would be expected to be the greater of the two at any given sequence depth.

DISCUSSION

In this study, we have demonstrated the feasibility of directly detecting chromosomal alterations in the plasma of cancer patients. Like many large-scale genomic analyses, our approach has limitations. First, sensitivity is largely dependent on the amount of sequence data obtained. Previous studies of circulating tumor DNA have shown that a sensitivity of $<0.10\%$ is often needed to detect patients with potentially curative tumors (17, 18). Currently, the cost of the sequencing necessary for detection of rearrangements at this level is prohibitive for routine clinical implementation. Detection of chromosomal copy number changes requires less sequencing and has been shown to be feasible at levels of 0.75% in

this study. Next-generation methods aimed at detection of somatic alterations in known driver genes require substantially less sequencing but are limited to patients with alterations in the analyzed regions (19). If sequencing technologies continue to improve at their current pace (28), the amount of sequencing needed for detection of whole-genome structural alterations will soon become affordable. Although substantial clinical studies will be needed to determine the use for early-stage disease as well as for direct genotyping of specific structural alterations, detection of medium- to late-stage tumors, such as those analyzed in this study, may provide clinical benefit for certain tumor types (29).

Second, the stringent approaches used to enrich for bona fide somatic alterations may fail to detect certain structural alterations (for example, small rearrangements or copy number changes), thereby underestimating the number of total alterations present in a sample. Analysis of a larger number of normal DNA samples combined with deeper sequencing may allow for a more comprehensive detection of somatic alterations in the plasma. Additional development of methods for detection of point mutations in cell-free DNA may provide a complementary approach for detecting disease in a subset of patients (17–19). Third, previously undetected constitutional germline or mosaic structural alterations along with mapping or sequence artifacts could lead to false positives in individual patients (30–32). Performing a comparative sequence analysis of plasma to another normal tissue (for example, buccal, skin, or lymphocyte DNA) from the same individual could help minimize this issue by removal of such alterations. Fourth, the information obtained through these analyses does not directly indicate the source of circulating tumor DNA. Further clinical evaluation combined with imaging studies will be helpful to determine the tumor location and subsequent interventions.

Despite these limitations, the combination of these methods has the potential to detect cancers in a noninvasive, specific, and unbiased manner. Direct identification of amplified genes in patient plasma may provide information for potential therapeutic targets without the need for tumor biopsies. Given the major contribution to morbidity and mortality caused by delayed diagnosis of primary or recurrent tumors, the approach described here, combined with further advances in sequencing technologies, has the potential to improve patient management and outcomes.

MATERIALS AND METHODS

Sample collection and preparation

Plasma samples were collected from cancer patients CRC11 to CRC17, BR1 to BR3 (4 to 17 ml each), and unrelated normal controls N1 to N10 (5 to 18 ml each). Matching tumor DNA samples were obtained from patients' formalin-fixed paraffin-embedded (FFPE) surgically resected tumor. Normal DNA samples were obtained from either matched lymphocytes or matched normal FFPE tissue obtained at the time of surgery. Whole-genome sequence data from normal lymphocytes of 28 individuals with neuroblastoma or pancreatic cancer were obtained from previous studies (33). Genotype and focal amplification data from 36 colorectal and 45 breast cancer samples, representing early-passage cell lines or xenografts established from patients with late-stage disease, were obtained from previous studies (25). All samples were obtained in accordance with the Health Insurance Portability and Accountability Act.

Preparation of Illumina fragment sequencing libraries

Plasma DNA libraries were prepared following Illumina's suggested protocol with the following modifications: (i) circulating DNA isolated from plasma was mixed with 10 μ l of End Repair Reaction Buffer and 5 μ l of End Repair Enzyme Mix (catalog no. E6050, New

England Biolabs) in a final volume of 100 μ l. The end-repair mixture was incubated at 20°C for 30 min, purified by a PCR purification kit (catalog no. 28104, Qiagen), and eluted with 45 μ l of elution buffer (EB) prewarmed to 70°C. If starting DNA volumes were >85 μ l, multiple 100 μ l end-repair reactions were used per sample and purified over the same Qiagen column as indicated above. (ii) To A-tail, all 42 μ l of end-repaired DNA was mixed with 5 μ l of 10 \times dA Tailing Reaction Buffer and 3 μ l of Klenow (exo-) (catalog no. E6053, New England Biolabs). The 50- μ l mixture was incubated at 37°C for 30 min before DNA was purified with a MinElute PCR purification kit (catalog no. 28004, Qiagen). Purified DNA was eluted with 27 μ l of EB. (iii) For adaptor ligation, 25 μ l of A-tailed DNA was mixed with 10 μ l of PE-adaptor (Illumina), 10 μ l of 5 \times Ligation buffer, and 5 μ l of Quick T4 DNA ligase (catalog no. E6056, New England Biolabs). The ligation mixture was incubated at 20°C for 15 min. (iv) To purify adaptor-ligated DNA, 50 μ l of ligation mixture from step (iv) was mixed with 200 μ l of NT buffer and cleaned up by NucleoSpin column (catalog no. 636972, Clontech). DNA was eluted in 50 μ l of EB. (v) To obtain an amplified library, 10 PCRs of 25 μ l each were set up, each including 12 μ l of H₂O, 5 μ l of 5 \times Phusion HF buffer, 0.5 μ l of a deoxynucleotide triphosphate (dNTP) mix containing 10 mM of each dNTP, 1.25 μ l of dimethyl sulfoxide (DMSO), 0.5 μ l of Illumina PE primer 1, 0.5 μ l of Illumina PE primer 2, 0.25 μ l of Hot Start Phusion polymerase (M0530L, New England Biolabs), and 5 μ l of the DNA from step (v). The following PCR program was used: 98°C for 1 min; 10 cycles of 98°C for 20 s, 65°C for 30 s, 72°C for 30 s; and 72°C for 5 min. The pooled 250- μ l PCR was purified by two sequential AMPure XP bead (catalog no. A63881, Beckman Genomics) purifications with a 1:1 PCR product/AMPure bead mix. Library DNA was eluted with 40 μ l of EB, and the DNA concentration and library quality were determined with both NanoDrop and BioAnalyzer.

Genomic DNA libraries from FFPE tissue were prepared as described above with the following modifications: (i) Before end repair, 0.5 to 1 μ g of genomic tumor DNA in 100 μ l of TE was fragmented in a Covaris sonicator to a size of 100 to 500 bp. To remove fragments shorter than 150 bp, we mixed DNA with 25 μ l of 5 \times Phusion HF buffer, 416 μ l of ddH₂O, and 84 μ l of NT binding buffer and loaded it into NucleoSpin column. The column was centrifuged at 14,000g in a desktop centrifuge for 1 min, washed once with 600 μ l of wash buffer NT3 (Clontech), and centrifuged again for 2 min to dry completely. DNA was eluted in 45 μ l of EB included in the kit. (ii) The PCR program used was as follows: 98°C for 1 min; 10 cycles of 98°C for 20 s, 65°C for 30 s, 72°C for 30 s; and 72°C for 5 min.

Sequencing and analyses of Illumina fragment libraries

All samples were sequenced on Illumina HiSeq and Genome Analyzer II instruments. Sequence reads were analyzed and aligned to human genome hg18 with the ELAND algorithm using CASAVA 1.7 software (Illumina). Reads were mapped with the default seed-and-extend algorithm, allowing a maximum of two mismatched bases in the first 32 bp of sequence. Paired reads with duplicate start positions or reads with mapping scores of <50 were removed from further analysis.

Identification of somatic rearrangements

Somatic rearrangements were identified by querying aberrantly mapping paired-end sequences identified through ELAND. The discordantly mapping pairs were grouped into 1-kb bins and further considered when more than two distinct read pairs (with distinct start sites) spanned the same two 1-kb bins and suggested either an inter- or an intrachromosomal rearrangement \geq 30 kb in size. To identify all high-confidence genomic rearrangements, we required candidate rearrangements to have at least one read sequenced across the rearrangement breakpoint. Breakpoints were identified at base-pair resolution with BLAT

(34) to compare the putative rearrangement to the human genome sequence (March 2006, NCBI36/hg18) with zero mismatches and <5 bp overlapping between the two rearranged loci. To reduce the number of false-positive somatic rearrangements identified, we applied a final set of filters. Candidate somatic rearrangements were required to (i) not contain sequences that are entirely repeat-masked in the 200 bp surrounding the breakpoint, (ii) contain paired-end reads spanning the breakpoint where >66% of the aberrantly mapped paired reads for each loci were involved in the candidate rearrangement, (iii) uniquely map to a region of the reference genome using a 36-bp sequence (two or less mismatches), and (iv) occur in a region not implicated in known human germline variation (35).

These first three criteria were designed to eliminate those rearrangements identified by alignment artifacts to the reference genome, whereas the fourth criteria removed common germline polymorphisms. Candidate rearrangements were independently confirmed to be somatic when a 10- μ l PCR-based reaction [containing 5.9 μ l of H₂O, 1 μ l of 10 \times PCR buffer, 1 μ l of 10 mM dNTPs, 0.6 μ l of DMSO, 0.4 μ l of 25 μ M primers, 0.1 μ l of Platinum Taq, and 1 μ l of DNA (3 ng/ μ l)] resulted in the amplification of a product of the expected size using tumor DNA but not DNA from the matched normal tissue or lymphocytes.

Detection of copy number alterations

For each sample, distinct reads with mapping scores of ≥ 30 were mapped uniquely to chromosome arms. The total number of reads (one read was analyzed per cluster generated on the flow cell) mapping to each chromosome arm was compared to the distinct reads mapping to all 39 autosomal chromosome arms (only one chromosome arm from each of five acrocentric chromosomes was evaluated). Given the known GC bias of both next-generation sequencing library preparation and next-generation sequencing, we corrected for GC content across the genome (23, 24). The average number of distinct reads mapping to each 5-kb region of GC content (in 0.1% GC intervals) was determined. A weight was calculated for each GC interval equal to the global average number of distinct reads mapping to any interval of GC content divided by a specific interval of GC content (excluding intervals with more or less than 1.96 SDs of the global average number of distinct reads). After regions of known copy number polymorphism were removed (35), the remaining GC-corrected reads were summed for each chromosome arm. The number of reads mapping to a given chromosome arm was compared to the total number of reads mapping to all 39 chromosome arms to calculate the percentage of genomic representation (GR).

To determine whether the observed percentage of reads mapping to a given chromosome arm in the plasma derived from patients with cancer was different from the observed percentage of reads mapping to a given chromosome arm in the plasma derived from normal patients, we calculated the *z* score.

$$z \text{ score} = \frac{\text{observed GR [case]} - \text{mean GR [normal cases]}}{\text{SD GR [normal cases]}}$$

A cutoff of ≥ 1.88 or ≤ -1.88 was applied ($P < 0.05$ after Bonferroni correction for 39 chromosome arms, Student's *t* test with three degrees of freedom) to indicate either a gain or a loss of the chromosome arm.

To combine the significance of multiple chromosome arm alterations in plasma cancer samples, we used the five chromosomes containing the arms with the largest absolute *z* scores in each case to calculate a PA score. The top five *z* scores for each case were

converted to P values (Student's t distribution, three degrees of freedom), and the negative of the sum of the logarithms of the P values was calculated for each tumor and normal case. The PA score was calculated as follows:

$$\text{PA score} = \frac{|\log P [\text{case}] - \text{mean } \log P [\text{normal cases}]|}{\text{SD } \log P [\text{normal cases}]}$$

where

$$\text{Log } P = \sum_{n=1}^5 [-\log(P \text{ value})_n]$$

PA scores higher than the threshold score of 5.84 provide a specificity greater than 99% (Student t distribution, three degrees of freedom) for the presence of aneuploid circulating tumor DNA.

Simulated sensitivity of PARE approaches

Estimates of the number of sequence reads needed for detection of circulating cell-free tumor DNA were determined for PARE with paired sequence reads of 100 bp in length and the following assumptions:

- i. Each tumor was conservatively assumed to have at least two separate structural alterations per genome. This is consistent with previous reports of most human cancer samples analyzed to date using whole-genome sequencing having at least these many alterations per genome (36).
- ii. Circulating cell-free DNA was assumed to be a mixture derived from normal and tumor DNA.
- iii. Rearrangements were assumed to be either heterozygous or amplified at 50 copies per nucleus. These conditions illustrate the boundaries of rearrangement copy number changes in tumor samples and show the increased sensitivity of detecting such alterations in tumors with focal amplifications. The upper level of 50 copies per nucleus is based on reported examples of *EGFR* and *MYCN* amplifications in human cancers (37, 38).
- iv. The sequence data obtained were sufficient to allow sequencing at least one of the rearrangement breakpoints at least twice to pass bioinformatic filters.
- v. Rearrangements that passed our bioinformatic filters were assumed to only be present in tumor cells.

Using this approach, the minimum fraction of circulating tumor DNA detectable was estimated for given levels of sequencing requiring a detection sensitivity of 90% for detection of at least one rearrangement.

Simulated sensitivity of DK

Estimates of the number of sequence reads needed for detection of circulating cell-free tumor DNA were determined for DK with paired sequence reads of 100 bp in length and the following assumptions:

- i. The simulated next-generation sequence data for 10,000 simulated normal cases obeyed a normal distribution centered on the observed experimental means and SD

for each chromosome arm proportion observed in the 10 experimental normal cases analyzed (N1 to N10).

- ii. The simulated next-generation sequence data for 36 colorectal cancer cases and 45 breast cancer cases obeyed a normal distribution with SD for each chromosome arm proportion observed in the 10 experimental normal cases analyzed (N1 to N10).
- iii. The karyotypes of SNP data previously obtained for 36 colorectal cancers and 45 breast cancers (25) were inferred with SNP allele frequencies (39). Copy number was inferred from the genotypes of alleles A and B with the following criteria: B allele frequency of 0.1 to 0.3 (tetraploid, AAAB genotype), 0.3 to 0.4 (triploid, AAB genotype), 0.4 to 0.6 (diploid, AB genotype), 0.6 to 0.7 (triploid BBA), and 0.7 to 0.9 (tetraploid BBBA). The above criteria were used to determine copy number of each chromosome arm based on informative SNPs. If multiple genotypes were present on an arm, the most frequently observed genotype was used to determine copy number. If >60% of informative SNPs did not meet the above criteria, the arm was considered to be present at a single copy. At least five chromosome arm alterations were observed in every sample analyzed in contrast to other copy number alterations (for example, focal amplifications more than fivefold and whole chromosome changes), which were observed in only a fraction of cases.
- iv. As the amount of sequencing increased, the SD decreased proportionately to the square root of the number of reads obtained, similar to that expected from previous analyses of circulating fetal DNA (8, 9).

To estimate the sensitivity and specificity of detecting copy number alterations in the plasma using one HiSeq lane of sequencing, we calculated the chromosome arm proportions in R (version 2.3.1) (40), adjusted for the aneuploid karyotype by multiplying the simulated genomic proportion for each arm by the number of copies of the respective arm divided by two for each of the 81 human cancers based on the karyotypes indicated above in (iii) and then converted to *z* scores. PA scores were calculated from *z* scores as described above with data from 10,000 simulated unaffected cases as normal controls. ROC analyses were performed with the PA scores at different cutoffs for a given tumor fraction to assess sensitivity and specificity. Similar analyses were performed at sequencing levels higher and lower than one Illumina HiSeq lane with the observed mean proportions and fold-adjusted SD. In each case, the theoretical SD (calculated as the square root of the ratio comprising the genomic proportion multiplied by one minus the genomic proportion divided by the total tags sequenced) was based on the amount of simulated sequence data and multiplied by the fold difference between the observed and theoretical SDs from the observed data of unaffected individuals (N1 to N10) obtained with one Illumina HiSeq lane (table S2). The minimum fraction of circulating tumor DNA detected for a given level of sequencing was then interpolated with the observed mean and SD for each chromosome arm at 90% sensitivity and 99% specificity.

Quantification of tumor burden in plasma

Digital PCR with PARE biomarker-specific primers was used to quantify circulating tumor DNA in patient plasma as described (6, 41). Briefly, circulating tumor DNA was PCR-amplified with patient rearrangement-specific primers (at a final concentration of 0.5 μ M each) and 2 \times Phusion Flash PCR Master Mix. Control primers to the unrearranged allele of each locus were used to establish the amount of unrearranged DNA in each sample. The total DNA amount in each sample was considered to be the sum of rearranged and unrearranged DNA. Quantification through copy number analysis of chromosome arms was performed as previously described (24) for cases CRC13 and CRC16, where the average

tumor fraction of chromosome arms with absolute z scores of ≥ 1.88 were used to determine circulating tumor burden. A Pearson correlation between the tumor burden in plasma samples and PA scores was calculated for all cases, where circulating tumor DNA levels could be independently measured with digital PCR with PARE biomarker-specific primers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank T. Mosbrugger, C. Blair, L. Dobbyn, M. Kadan, J. Ptak, K. Romans, J. Schaefer, N. Silliman, and D. Singh for technical assistance.

Funding: The project was supported by NIH grants CA121113, CA057345, CA043460; The European Community's Seventh Framework Programme; the Virginia & D. K. Ludwig Fund for Cancer Research; the American Association for Cancer Research Stand Up To Cancer–Dream Team Translational Cancer Research Grant; National Colorectal Cancer Research Alliance; United Negro College Fund–Merck Fellowship; and Swim Across America.

REFERENCES AND NOTES

1. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*. 2007; 7:233–245. [PubMed: 17361217]
2. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, Stebbings LA, Leroy C, Edkins S, Hardy C, Teague JW, Menzies A, Goodhead I, Turner DJ, Clee CM, Quail MA, Cox A, Brown C, Durbin R, Hurler ME, Edwards PA, Bignell GR, Stratton MR, Futreal PA. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet*. 2008; 40:722–729. [PubMed: 18438408]
3. Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT, Stebbings LA, Leroy C, Edkins S, Mudie LJ, Greenman CD, Jia M, Latimer C, Teague JW, Lau KW, Burton J, Quail MA, Swerdlow H, Churcher C, Natrajan R, Sieuwerts AM, Martens JW, Silver DP, Langerød A, Russnes HE, Foekens JA, Reis-Filho JS, van 't Veer L, Richardson AL, Børresen-Dale AL, Campbell PJ, Futreal PA, Stratton MR. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*. 2009; 462:1005–1010. [PubMed: 20033038]
4. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998; 396:643–649. [PubMed: 9872311]
5. Wang TL, Maierhofer C, Speicher MR, Lengauer C, Vogelstein B, Kinzler KW, Velculescu VE. Digital karyotyping. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:16156–16161. [PubMed: 12461184]
6. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, Antipova A, Lee C, McKernan K, De La Vega FM, Kinzler KW, Vogelstein B, Diaz LA Jr. Velculescu VE. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl. Med.* 2010; 2:20ra14.
7. Beroukhi R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010; 463:899–905. [PubMed: 20164920]
8. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:16266–16271. [PubMed: 18838674]
9. Chiu RW, Chan KC, Gao Y, Lau VY, Zheng W, Leung TY, Foo CH, Xie B, Tsui NB, Lun FM, Zee BC, Lau TK, Cantor CR, Lo YM. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy

- by massively parallel genomic sequencing of DNA in maternal plasma. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:20458–20463. [PubMed: 19073917]
10. Grimwade D, Jovanovic JV, Hills RK, Nugent EA, Patel Y, Flora R, Diverio D, Jones K, Aslett H, Batson E, Rennie K, Angell R, Clark RE, Solomon E, Lo-Coco F, Wheatley K, Burnett AK. Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy. *J. Clin. Oncol.* 2009; 27:3650–3658. [PubMed: 19506161]
 11. Bregni M, Siena S, Neri A, Bassan R, Barbui T, Delia D, Bonadonna G, Dalla Favera R, Gianni AM. Minimal residual disease in acute lymphoblastic leukemia detected by immune selection and gene rearrangement analysis. *J. Clin. Oncol.* 1989; 7:338–343. [PubMed: 2918330]
 12. van der Velden VH, Panzer-Grümayer ER, Cazzaniga G, Flohr T, Sutton R, Schrauder A, Basso G, Schrappe M, Wijkhuijs JM, Konrad M, Bartram CR, Masera G, Biondi A, van Dongen JJ. Optimization of PCR-based minimal residual disease diagnostics for childhood acute lymphoblastic leukemia in a multi-center setting. *Leukemia.* 2007; 21:706–713. [PubMed: 17287857]
 13. Lion T. Minimal residual disease. *Curr. Opin. Hematol.* 1999; 6:406–411. [PubMed: 10546795]
 14. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J, Baccarani M, Cortes J, Cross NC, Druker BJ, Gabert J, Grimwade D, Hehlmann R, Kamel-Reid S, Lipton JH, Longtine J, Martinelli G, Saglio G, Soverini S, Stock W, Goldman JM. Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: Review and recommendations for harmonizing current methodology for detecting *BCR-ABL* transcripts and kinase domain mutations and for expressing results. *Blood.* 2006; 108:28–37. [PubMed: 16522812]
 15. He J, Wu J, Jiao Y, Wagner-Johnston N, Ambinder RF, Diaz LA Jr. Kinzler KW, Vogelstein B, Papadopoulos N. IgH gene rearrangements as plasma biomarkers in non- Hodgkin's lymphoma patients. *Oncotarget.* 2011; 2:178–185. [PubMed: 21399237]
 16. McBride DJ, Orpana AK, Sotiriou C, Joensuu H, Stephens PJ, Mudie LJ, Hämäläinen E, Stebbings LA, Andersson LC, Flanagan AM, Durbecq V, Ignatiadis M, Kallioniemi O, Heckman CA, Alitalo K, Edgren H, Futreal PA, Stratton MR, Campbell PJ. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer.* 2010; 49:1062–1069. [PubMed: 20725990]
 17. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, Diaz LA Jr. Goodman SN, David KA, Juhl H, Kinzler KW, Vogelstein B. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* 2005; 102:16368–16373. [PubMed: 16258065]
 18. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA Jr. Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* 2008; 14:985–990. [PubMed: 18670422]
 19. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* 2012; 4:136ra68.
 20. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM, vanTuinen P, Ledbetter DH, Barker DF, Nakamura Y, White R, Vogelstein B. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science.* 1989; 244:217–221. [PubMed: 2649981]
 21. Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D, Finniear R, Markham A, Groffen J, Boguski MS, Altschul SF, Horii A, Ando H, Miyoshi Y, Miki Y, Nishisho I, Nakamura Y. Identification of FAP locus genes from chromosome 5q21. *Science.* 1991; 253:661–665. [PubMed: 1651562]
 22. Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science.* 1991; 253:665–669. [PubMed: 1651563]
 23. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat. Genet.* 2009; 41:1061–1067. [PubMed: 19718026]

24. Fan HC, Quake SR. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. *PLoS One*. 2010; 5:e10439. [PubMed: 20454671]
25. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, Jones S, Sjöblom T, Park BH, Parsons R, Willis J, Dawson D, Willson JK, Nikolskaya T, Nikolsky Y, Kopelovich L, Papadopoulos N, Pennacchio LA, Wang TL, Markowitz SD, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proc. Natl. Acad. Sci. U.S.A.* 2008; 105:16224–16229. [PubMed: 18852474]
26. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
27. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stütz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Ye K, Eichler EE, Gerstein MB, Hurles ME, Lee C, McCarroll SA, Korbelt JO, 1000 Genomes Project. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011; 470:59–65. [PubMed: 21293372]
28. Stein LD. The case for cloud computing in genome informatics. *Genome Biol*. 2010; 11:207. [PubMed: 20441614]
29. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J. Clin.* 2012; 62:10–29. [PubMed: 22237781]
30. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, Hutchinson A, Deng X, Liu C, Horner MJ, Cullen M, Epstein CG, Burdett L, Dean MC, Chatterjee N, Sampson J, Chung CC, Kovaks J, Gapstur SM, Stevens VL, Teras LT, Gaudet MM, Albanes D, Weinstein SJ, Virtamo J, Taylor PR, Freedman ND, Abnet CC, Goldstein AM, Hu N, Yu K, Yuan JM, Liao L, Ding T, Qiao YL, Gao YT, Koh WP, Xiang YB, Tang ZZ, Fan JH, Aldrich MC, Amos C, Blot WJ, Bock CH, Gillanders EM, Harris CC, Haiman CA, Henderson BE, Kolonel LN, Le Marchand L, McNeill LH, Rybicki BA, Schwartz AG, Signorello LB, Spitz MR, Wiencke JK, Wrensch M, Wu X, Zanetti KA, Ziegler RG, Figueroa JD, Garcia-Closas M, Malats N, Marenne G, Prokunina-Olsson L, Baris D, Schwenn M, Johnson A, Landi MT, Goldin L, Consonni D, Bertazzi PA, Rotunno M, Rajaraman P, Andersson U, Freeman LE, Berg CD, Buring JE, Butler MA, Carreon T, Feychting M, Ahlbom A, Gaziano JM, Giles GG, Hallmans G, Hankinson SE, Hartge P, Henriksson R, Inskip PD, Johansen C, Landgren A, McKean-Cowdin R, Michaud DS, Melin BS, Peters U, Ruder AM, Sesso HD, Severi G, Shu XO, Visvanathan K, White E, Wolk A, Zeleniuch-Jacquotte A, Zheng W, Silverman DT, Kogevinas M, Gonzalez JR, Villa O, Li D, Duell EJ, Risch HA, Olson SH, Kooperberg C, Wolpin BM, Jiao L, Hassan M, Wheeler W, Arslan AA, Bueno-de-Mesquita HB, Fuchs CS, Gallinger S, Gross MD, Holly EA, Klein AP, LaCroix A, Mandelsohn MT, Petersen G, Boutron-Ruault MC, Bracci PM, Canzian F, Chang K, Cotterchio M, Giovannucci EL, Goggins M, Bolton JA, Jenab M, Khaw KT, Krogh V, Kurtz RC, McWilliams RR, Mendelsohn JB, Rabe KG, Riboli E, Tjønneland A, Tobias GS, Trichopoulos D, Elena JW, Yu H, Amundadottir L, Stolzenberg-Solomon RZ, Kraft P, Schumacher F, Stram D, Savage SA, Mirabello L, Andrulis IL, Wunder JS, Patiño García AP, Sierrasesúmaga L, Barkauskas DA, Gorlick RG, Purdue M, Chow WH, Moore LE, Schwartz KL, Davis FG, Hsing AW, Berndt SI, Black A, Wentzensen N, Brinton LA, Lissowska J, Peplonska B, McGlynn KA, Cook MB, Graubard BI, Kratz CP, Greene MH, Erickson RL, Hunter DJ, Thomas G, Hoover RN, Real FX, Fraumeni JF Jr, Caporaso NE, Tucker M, Rothman N, Pérez-Jurado LA, Chanock SJ. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* 2012; 44:651–658. [PubMed: 22561519]
31. Laurie CC, Laurie CA, Rice K, Doherty KF, Zelnick LR, McHugh CP, Ling H, Hetrick KN, Pugh EW, Amos C, Wei Q, Wang LE, Lee JE, Barnes KC, Hansel NN, Mathias R, Daley D, Beaty TH, Scott AF, Ruczinski I, Scharpf RB, Bierut LJ, Hartz SM, Landi MT, Freedman ND, Goldin LR, Ginsburg D, Li J, Desch KC, Strom SS, Blot WJ, Signorello LB, Ingles SA, Chanock SJ, Berndt SI, Le Marchand L, Henderson BE, Monroe KR, Heit JA, de Andrade M, Armasu SM, Regnier C, Lowe WL, Hayes MG, Marazita ML, Feingold E, Murray JC, Melbye M, Feenstra B, Kang JH,

- Wiggs JL, Jarvik GP, McDavid AN, Seshan VE, Mirel DB, Crenshaw A, Sharopova N, Wise A, Shen J, Crosslin DR, Levine DM, Zheng X, Udren JI, Bennett S, Nelson SC, Gogarten SM, Conomos MP, Heagerty P, Manolio T, Pasquale LR, Haiman CA, Caporaso N, Weir BS. Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* 2012; 44:642–650. [PubMed: 22561516]
32. Bäscke J, Griesinger F, Trümper L, Brittinger G. Leukemia- and lymphoma-associated genetic aberrations in healthy individuals. *Ann. Hematol.* 2002; 81:64–75. [PubMed: 11907785]
33. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, Axilbund J, Schulick R, Ali SZ, Eshleman JR, Velculescu VE, Goggins M, Vogelstein B, Papadopoulos N, Hruban RH, Kinzler KW, Klein AP. ATM mutations in patients with hereditary pancreatic cancer. *Cancer Discov.* 2012; 2:41–46. [PubMed: 22585167]
34. Kent WJ. BLAT—The BLAST-like alignment tool. *Genome Res.* 2002; 12:656–664. [PubMed: 11932250]
35. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat. Genet.* 2004; 36:949–951. [PubMed: 15286789]
36. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M, Shlien A, Cooke SL, Hinton J, Menzies A, Stebbings LA, Leroy C, Jia M, Rance R, Mudie LJ, Gamble SJ, Stephens PJ, McLaren S, Tarpey PS, Papaemmanuil E, Davies HR, Varela I, McBride DJ, Bignell GR, Leung K, Butler AP, Teague JW, Martin S, Jönsson G, Mariani O, Boyault S, Miron P, Fatima A, Langerød A, Aparicio SA, Tutt A, Sieuwerts AM, Borg Å, Thomas G, Salomon AV, Richardson AL, Børresen-Dale AL, Futreal PA, Stratton MR, Campbell PJ, Breast Cancer Working Group of the International Cancer Genome Consortium. The life history of 21 breast cancers. *Cell.* 2012; 149:994–1007. [PubMed: 22608083]
37. Brodeur GM, Seeger RC. Gene amplification in human neuroblastomas: Basic mechanisms and clinical implications. *Cancer Genet. Cytogenet.* 1986; 19:101–111. [PubMed: 3940169]
38. Sauter G, Maeda T, Waldman FM, Davis RL, Feuerstein BG. Patterns of epidermal growth factor receptor amplification in malignant gliomas. *Am. J. Pathol.* 1996; 148:1047–1053. [PubMed: 8644846]
39. Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw CA, Belmont J, Cheung SW, Shen RM, Barker DL, Gunderson KL. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res.* 2006; 16:1136–1148. [PubMed: 16899659]
40. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. <http://www.R-project.org>
41. Vogelstein B, Kinzler KW. Digital PCR. *Proc. Natl. Acad. Sci. U.S.A.* 1999; 96:9236–9241. [PubMed: 10430926]

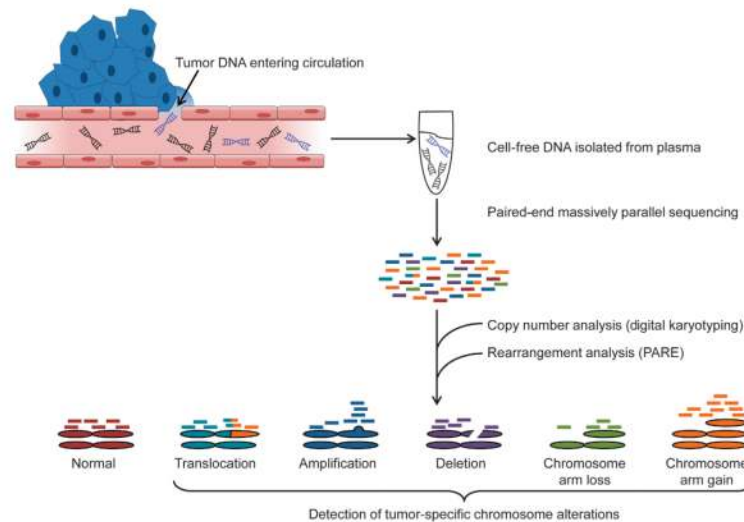
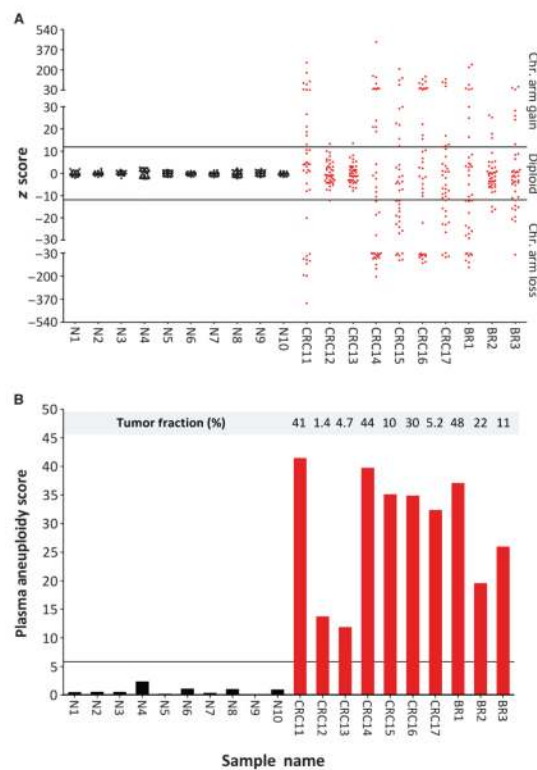
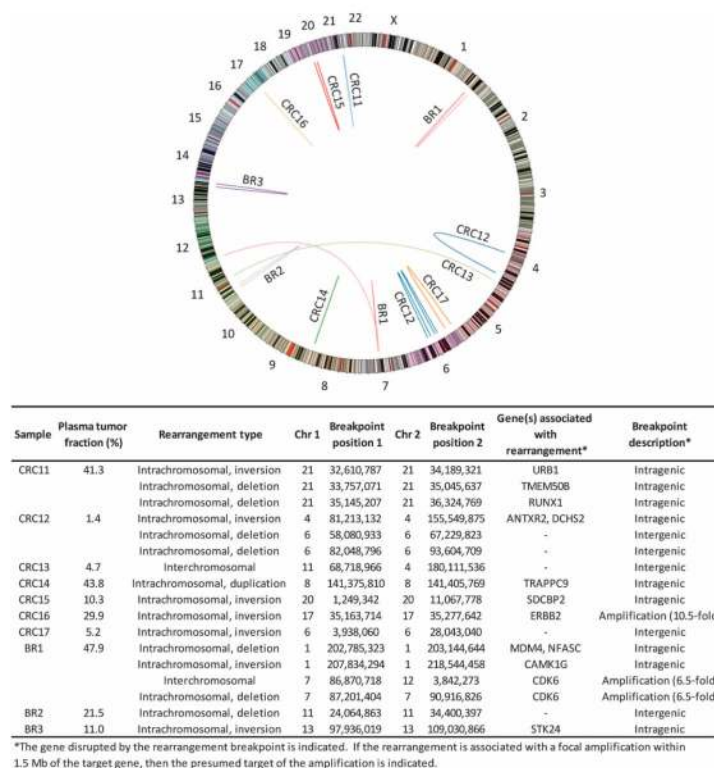


Fig. 1. Schematic of analyses for direct detection of chromosomal alterations in plasma. The method uses next-generation paired-end sequencing of cell-free DNA isolated from plasma to identify chromosomal alterations characteristic of tumor DNA. Such alterations include copy number changes (gains and losses of chromosome arms) as well as rearrangements resulting from translocations, amplifications, or deletions.

**Fig. 2.**

Copy number analysis of plasma samples. **(A)** The z scores for each chromosome arm indicate the number of SDs from the mean of the mapped read fraction of the plasma DNA from unaffected individuals (N1 to N10). Positive z scores indicate chromosome gains, whereas negative z scores indicate chromosome losses. Significant chromosome arm gains and losses were observed only in plasma samples from patients with cancer (CRC11 to CRC17 and BR1 to BR3). **(B)** The PA score was calculated as the number of SDs from the mean of the sum of the $-\log$ of the P values for the top five chromosome z scores of the 10 reference samples (N1 to N10). A PA score of 5.84 (horizontal line) was estimated to indicate aneuploidy in the plasma sample at a specificity greater than 99% (Student's t distribution) (see Materials and Methods).

**Fig. 3.**

Detection of tumor-specific rearrangements in plasma samples. The Circos plot at the top indicates the rearrangements identified in plasma samples from cancer patients (CRC11 to CRC17 and BR1 to BR3). The type and individual boundaries of the rearrangements are indicated in the lower table. No rearrangements were identified in plasma samples from unaffected individuals (N1 to N10). Rearrangements listed for sample CRC12 were identified in tumor DNA and confirmed in patient plasma, whereas those listed for all other samples were identified directly from patient plasma.

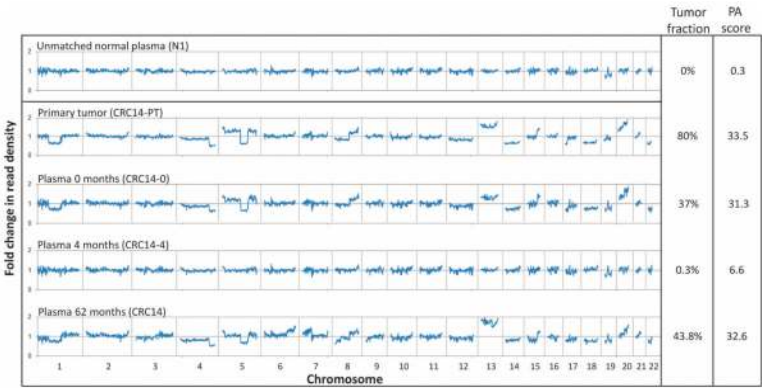


Fig. 4. Copy number analyses of tumor and serial plasma samples from patient CRC14. CRC14 primary tumor and plasma samples taken at various time points over 62 months of multimodality treatment were analyzed using DK in nonoverlapping 1-Mb windows and compared with unmatched normal plasma using the same methodology. The plasma samples were obtained at the time of initial evaluation (0 months), after extensive chemotherapy and surgical intervention (4 months), and at the time of cancer recurrence (62 months).

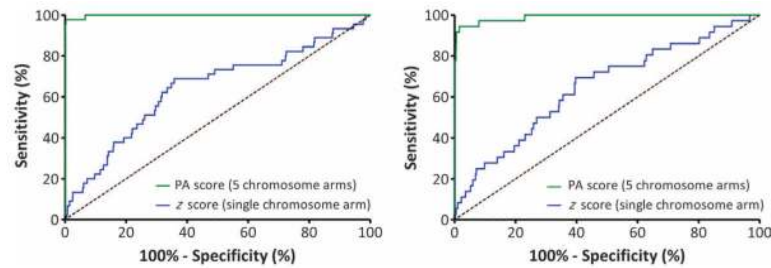


Fig. 5.

Detection of circulating tumor DNA in breast and colon cancers using simulated copy number analyses. ROC analyses of simulated mixtures of breast cancer DNA (left) or colorectal cancer DNA (right) with normal plasma DNA using the PA score derived from the five chromosomal arm copy number alterations with the highest absolute z scores in each sample. Detection of 0.75% circulating tumor DNA could be achieved with a sensitivity of >90% and specificity of >99% using the equivalent of one HiSeq lane of sequencing and a fixed PA score threshold in both tumor types (see Materials and Methods). ROC analyses of a z score from a single chromosome arm, 17p, were similar to chance alone at this simulated tumor DNA concentration in the plasma.

Table 1

Summary of next-generation sequencing analyses performed. Data were obtained using next-generation sequencing analyses performed on Illumina HiSeq instruments using 50-bp PE reads. Distinct paired reads correspond to read pairs having unique genomic start sites. Sequence coverage indicates average number of reads per base per haploid genome. Physical coverage indicates average number of paired reads spanning any base in a haploid genome assuming a 165-bp fragment size.

Sample name	Patient diagnosis	Sample origin	Total bases sequenced	Total distinct paired reads	Sequence coverage	Physical coverage
CRC11	Colorectal cancer	Plasma	24,728,144,682	216,092,204	8.2	11.9
CRC12	Colorectal cancer	Plasma	25,707,029,400	221,881,499	8.6	12.2
CRC12-PT	Colorectal cancer	Tumor	14,984,097,228	113,414,859	5.0	6.2
CRC13	Colorectal cancer	Plasma	24,033,905,652	206,926,843	8.0	11.4
CRC14	Colorectal cancer	Plasma	23,774,411,124	201,571,426	7.9	11.1
CRC14-0	Colorectal cancer	Plasma	3,113,755,960	17,297,631	1.0	1.0
CRC14-4	Colorectal cancer	Plasma	3,755,921,750	21,193,176	1.3	1.2
CRC14-PT	Colorectal cancer	Tumor	7,156,542,105	55,882,387	2.4	3.1
CRC15	Colorectal cancer	Plasma	34,216,224,513	299,895,779	11.4	16.5
CRC15-PT	Colorectal cancer	Tumor	12,466,375,200	102,193,630	4.2	5.6
CRC16	Colorectal cancer	Plasma	32,670,584,037	284,131,684	10.9	15.6
CRC16-PT	Colorectal cancer	Tumor	6,199,959,150	41,700,493	2.1	2.3
CRC17	Colorectal cancer	Plasma	33,060,006,522	286,524,067	11.0	15.8
BR1	Breast cancer	Plasma	34,918,959,429	304,798,840	11.6	16.8
BR2	Breast cancer	Plasma	30,171,911,865	240,263,505	10.1	13.2
BR3	Breast cancer	Plasma	31,294,294,671	259,443,659	10.4	14.3
N1	Normal	Plasma	26,918,560,359	231,520,314	9.0	12.7
N2	Normal	Plasma	25,928,759,499	224,708,017	8.6	12.4
N3	Normal	Plasma	21,331,401,576	161,874,934	7.1	8.9
N4	Normal	Plasma	25,735,971,696	223,342,309	8.6	12.3
N5	Normal	Plasma	33,535,967,796	288,318,721	11.2	15.9
N6	Normal	Plasma	32,892,667,872	285,785,107	11.0	15.7
N7	Normal	Plasma	27,558,615,816	226,653,076	9.2	12.5
N8	Normal	Plasma	32,472,392,886	279,224,469	10.8	15.4
N9	Normal	Plasma	30,058,183,548	257,431,312	10.0	14.2
N10	Normal	Plasma	33,068,060,850	287,180,675	11.0	15.8