Clinical Chemistry 59:1
211–224 (2013)

Cancer Diagnostics

fast track

Cancer Genome Scanning in Plasma: Detection of Tumor-Associated Copy Number Aberrations, Single-Nucleotide Variants, and Tumoral Heterogeneity by Massively Parallel Sequencing

K.C. Allen Chan,[1,2,3] Peiyong Jiang,[1,2] Yama W.L. Zheng,[1,2] Gary J.W. Liao,[1,2] Hao Sun,[1,2] John Wong,[4] Shing Shun N. Siu,[5] Wing C. Chan,[6] Stephen L. Chan,[3,7] Anthony T.C. Chan,[3,7] Paul B.S. Lai,[4] Rossa W.K. Chiu,[1,2] and Y.M.D. Lo[1,2,3*]

BACKGROUND: Tumor-derived DNA can be found in the plasma of cancer patients. In this study, we explored the use of shotgun massively parallel sequencing (MPS) of plasma DNA from cancer patients to scan a cancer genome noninvasively.

METHODS: Four hepatocellular carcinoma patients and a patient with synchronous breast and ovarian cancers were recruited. DNA was extracted from the tumor tissues, and the preoperative and postoperative plasma samples of these patients were analyzed with shotgun MPS.

RESULTS: We achieved the genomewide profiling of copy number aberrations and point mutations in the plasma of the cancer patients. By detecting and quantifying the genomewide aggregated allelic loss and point mutations, we determined the fractional concentrations of tumor-derived DNA in plasma and correlated these values with tumor size and surgical treatment. We also demonstrated the potential utility of this approach for the analysis of complex oncologic scenarios by studying the patient with 2 synchronous cancers. Through the use of multiregional sequencing of tumoral tissues and shotgun sequencing of plasma DNA, we have shown that plasma DNA sequencing is a valuable approach for studying tumoral heterogeneity.

CONCLUSIONS: Shotgun DNA sequencing of plasma is a potentially powerful tool for cancer detection, monitoring, and research.

© 2012 American Association for Clinical Chemistry

The presence of tumor-derived DNA in the plasma of cancer patients offers exciting opportunities for the detection and monitoring of cancer *(1, 2)*. Indeed, cancer-associated microsatellite alterations *(3, 4)*, gene mutations *(5–9)*, DNA-methylation changes *(10, 11)*, and viral nucleic acids *(12)* have been found in the plasma of patients with different cancer types. Most of the previously published work on plasma DNA as a cancer marker has focused on the detection of specific and predetermined molecular targets known to be associated with cancer by means of such methods as the PCR *(3, 4, 12)*, digital PCR *(5–7)*, and digital ligation assays *(9)*. With the advent of massively parallel sequencing (MPS),[8] several groups have incorporated this approach for developing new plasma DNA–based cancer markers. One approach is to use MPS on tumor samples to first identify specific genomic rearrangements that can subsequently be detected in plasma *(13, 14)*. Another approach is based on the use of targeted amplicon sequencing to search for mutations of genes that are commonly found in cancer *(8)*.

[1] Li Ka Shing Institute of Health Sciences, [2] Department of Chemical Pathology, [3] State Key Laboratory in Oncology in South China, Sir Y.K. Pao Centre for Cancer, [4] Department of Surgery, and [5] Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China; [6] Department of Surgery, North District Hospital, Sheung Shui, New Territories, Hong Kong SAR, China; [7] Department of Clinical Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China.
* Address correspondence to this author at: Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, 30–32 Ngan Shing St., Shatin, New Territories, Hong Kong SAR. Fax +852-26365090; e-mail loym@cuhk.edu.hk.
Received September 6, 2012; accepted September 27, 2012.
Previously published online at DOI: 10.1373/clinchem.2012.196014
[8] Nonstandard abbreviations: MPS, massively parallel sequencing; HCC, hepatocellular carcinoma; SOAP2, Short Oligonucleotide Alignment Program 2; SNP, single-nucleotide polymorphism; LOH, loss of heterozygosity; LOESS, locally weighted scatterplot smoothing; SNV, single nucleotide variant; GAAL, genomewide aggregated allelic loss.

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.001

Owing to their targeted nature, the approaches outlined above can provide only a partial glimpse of the tumor genome in the plasma of cancer patients. For a genomewide view of the tumor genome in the circulation, a nontargeted random—or shotgun—sequencing approach would be desirable. In this regard, there has been much progress in the field of noninvasive prenatal diagnosis because of the results obtained with shotgun MPS of DNA from the plasma of pregnant women *(15)*. This approach has allowed the noninvasive detection of fetal chromosomal aneuploidies *(16–18)* and fetal genomic scanning *(19, 20)*.

In this article, we report the use of shotgun MPS to obtain a noninvasive, genomewide view of cancer-associated copy number variations and mutations in DNA in plasma. We have also sought to demonstrate the use of this approach for elucidating important tumoral characteristics, with tumoral heterogeneity used as an example.

Materials and Methods

SAMPLE COLLECTION

Hepatocellular carcinoma (HCC) patients and carriers of chronic hepatitis B were recruited from the Department of Surgery and the Department of Medicine and Therapeutics, respectively, of the Prince of Wales Hospital, Hong Kong, and informed consent and institutional review board approval were obtained. All HCC patients had Barcelona Clinic Liver Cancer stage A1 disease. Informed consent was obtained after the nature and possible consequences of the studies were explained. The patient with synchronous breast and ovarian cancers was recruited from the Department of Clinical Oncology, Prince of Wales Hospital. Peripheral blood samples from all participants were collected into EDTA-containing tubes. The tumor tissues of the HCC patients were obtained during their cancer-resection surgeries.

PROCESSING OF BLOOD

Peripheral blood samples were centrifuged at 1600*g* for 10 min at 4 °C. The plasma portion was recentrifuged at 16 000*g* for 10 min at 4 °C and then stored at −80 °C. Cell-free DNA molecules from 4.8 mL of plasma were extracted according to the blood and body fluid protocol of the QIAamp DSP DNA Blood Mini Kit (Qiagen). The plasma DNA was concentrated with a SpeedVac® Concentrator (Savant DNA120; Thermo Scientific) into a 40-μL final volume per case for subsequent preparation of the DNA-sequencing library.

GENOMIC DNA EXTRACTION

Genomic DNA was extracted from patients' buffy coat samples according to the blood and body fluid protocol of the QIAamp DSP DNA Blood Mini Kit. DNA was extracted from tumor tissues with the QIAamp DNA Mini Kit (Qiagen).

DNA SEQUENCING

Sequencing libraries of the genomic DNA samples were constructed with the Paired-End Sample Preparation Kit (Illumina) according to the manufacturer's instructions. In brief, 1–5 μg genomic DNA was first sheared with a Covaris S220 Focused-ultrasonicator to 200-bp fragments. Afterward, DNA molecules were end-repaired with T_4 DNA polymerase and Klenow polymerase; T_4 polynucleotide kinase was then used to phosphorylate the 5′ ends. A 3′ overhang was created with a 3′-to-5′ exonuclease–deficient Klenow fragment. Illumina adapter oligonucleotides were ligated to the sticky ends. The adapter-ligated DNA was enriched with a 12-cycle PCR. Because the plasma DNA molecules were short fragments *(21)* and the amounts of total DNA in the plasma samples were relatively small, we omitted the fragmentation steps and used a 15-cycle PCR when constructing the DNA libraries from the plasma samples.

An Agilent 2100 Bioanalyzer (Agilent Technologies) was used to check the quality and size of the adapter-ligated DNA libraries. DNA libraries were then measured by a KAPA Library Quantification Kit (Kapa Biosystems) according to the manufacturer's instructions.

The DNA library was diluted and hybridized to the paired-end sequencing flow cells. DNA clusters were generated on a cBot cluster generation system (Illumina) with the TruSeq PE Cluster Generation Kit v2 (Illumina), followed by 51 × 2 cycles or 76 × 2 cycles of sequencing on a HiSeq 2000 system (Illumina) with the TruSeq SBS Kit v2 (Illumina).

SEQUENCE ALIGNMENT AND FILTERING

The paired-end sequencing data were analyzed by means of the Short Oligonucleotide Alignment Program 2 (SOAP2) in the paired-end mode *(22)*. For each paired-end read, 50 bp or 75 bp from each end was aligned to the non–repeat-masked reference human genome (Hg18). Up to 2 nucleotide mismatches were allowed for the alignment of each end. The genomic coordinates of these potential alignments for the 2 ends were then analyzed to determine whether any combination would allow the 2 ends to be aligned to the same chromosome with the correct orientation, spanning an insert size ≤600 bp, and mapping to a single location in the reference human genome. Duplicated reads were defined as paired-end reads in which the insert DNA molecule showed identical start and end locations in the human genome; the duplicate reads were removed as previously described *(19)*.

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.002

MICROARRAY ANALYSIS

DNA extracted from the buffy coat and the tumor tissues of the HCC patients was genotyped with the Affymetrix Genome-Wide Human SNP Array 6.0 system, as previously described *(23)*. The microarray data were processed with the Affymetrix Genotyping Console version 4.1. Genotyping analysis and single-nucleotide polymorphism (SNP) calling were performed with the Birdseed v2 algorithm, as previously described *(24)*. The genotyping data for the buffy coat and the tumor tissues were used for identifying loss-of-heterozygosity (LOH) regions and for performing the copy number analysis. Copy number analysis was performed with the Genotyping Console with default parameters from Affymetrix and with a minimum genomic-segment size of 100 bp and a minimum of 5 genetic markers within the segment. Regions with LOH were identified as regions having 1 copy in the tumor tissue and 2 copies in the buffy coat, with the SNPs within these regions being heterozygous in the buffy coat but homozygous in the tumor tissue. For a genomic region exhibiting LOH in a tumor tissue, the SNP alleles that were present in the buffy coat but were absent from or of reduced intensity in the tumor tissues were considered to be the alleles on the deleted segment of the chromosomal region. The alleles that were present in both the buffy coat and the tumor tissue were deemed as having been derived from the nondeleted segment of the chromosomal region.

ARRAY COMPARATIVE GENOMIC HYBRIDIZATION ANALYSIS

DNA samples extracted from the buffy coat and the tumor tissues of the HCC patients were analyzed with the SurePrint G3 Human High Resolution Microarray Kit (Agilent) as previously described *(25)*. Array comparative genomic hybridization data for the HCC patients were analyzed for copy number variation with the Partek® Genomics Suite. In brief, the raw probe intensities were adjusted according to the GC content of the sequence. This adjustment was followed by probe-level normalization of signal intensity while simultaneously adjusting for fragment length and probe sequences across all samples. Copy number gains and losses were detected by applying the default parameters of the Genomic Segmentation algorithm available in Partek Genomics Suite version 6.5 to obtain the different partitions of the copy number state.

DETECTION OF COPY NUMBER ABERRATION IN TUMOR TISSUE SAMPLES BY SEQUENCING

To investigate genomic copy number aberrations (e.g., copy number gains and copy number losses), we divided the genome into equal-sized segments (1 Mb per window/bin), and tallied the numbers of sequence reads mapping to each bin. Owing to the presence of GC-dependent sequencing biases with high-throughput sequencing technologies *(26)*, we used a statistical correction method, locally weighted scatterplot smoothing (LOESS), to correct the GC-associated bias *(27)*. In this method, a correction factor is calculated for each bin according to the LOESS regression model, as previously described *(28)*. Then, the read counts of each bin were adjusted with the bin-specific correction factor and normalized with the median read counts of all bins. After GC correction, a ratio of the adjusted read counts of the tumor to those of the buffy coat was calculated with the following equation:

$$R = \frac{A_{\text{tumor}}}{A_{\text{BC}}},$$

where A_{tumor} is the normalized GC-adjusted read counts of the tumor tissue and A_{BC} is the normalized GC-adjusted read counts of the buffy coat.

We then constructed a frequency distribution of $\log_2(R)$ for all bins. This distribution plot was used to estimate the proportion of tumor cells (F) in the tumor tissue that showed a particular copy number distribution and, subsequently, the copy number change at each bin.

On the frequency-distribution plot, a central peak at which R is approximately equal to 1 [i.e., $\log_2(R) = 0$] was identified; this peak represents the genomic regions without copy number aberrations. Then, the peaks lying to the left and right of the central peak were identified. These peaks represented regions with a 1-copy loss and a 1-copy gain, respectively. The distances of the left and right peaks from the central peak were used to determine the proportion of tumor cells (F) in the tumor tissue, according to the following equation:

$$F = R_{\text{right}} - R_{\text{left}},$$

where R_{right} is the R value of the right peak and R_{left} is the R value of the left peak.

Then, the copy number change (CN) values for all 1-Mb bins were calculated with the following equation:

$$\text{CN} = \frac{R_{\text{bin}} - R_{\text{cen}}}{0.5 \times F},$$

where R_{bin} is the R value of the bin and R_{cen} is the R value of the central peak.

DETECTION OF COPY NUMBER ABERRATIONS IN PLASMA

We analyzed the genomic representation of plasma DNA for different genomic regions. First, the entire genome was divided into 1-Mb windows, similar to the analysis of copy number aberrations in the tumor tissues. The GC-corrected read count was then determined (as described above) for each 1-Mb window. A z

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.003

score statistic was used to determine if the plasma DNA representation in a 1-Mb window would be significantly increased or decreased when compared with the reference group. The reference group consisted of the plasma samples from 16 healthy control individuals. In the current study, the GC-corrected read counts of each 1-Mb bin were normalized to the median GC-corrected read counts of all bins in the sample. The normalized plasma DNA representation was then compared with the data from the controls. A z score was then calculated for each 1-Mb window by using the mean and SDs of the controls. Regions with z scores of <-3 and >3 were regarded as significantly under- and overrepresented, respectively.

NUMBER OF MOLECULES REQUIRED FOR IDENTIFYING COPY NUMBER ABERRATIONS IN PLASMA

For copy number aberration analysis, the sensitivity and specificity of detecting tumor-associated copy number aberrations in plasma were determined by the precision of measuring the representation of plasma DNA in a chromosomal region and the fractional concentration of the tumor-derived DNA in the plasma of the cancer patient. The precision of measuring the plasma DNA representation in turn was affected by the number of plasma DNA molecules analyzed. In this regard, we performed simulation analyses to determine the relationship between the number of plasma DNA molecules required for analysis and the fractional concentration of tumor-derived DNA in the plasma so we could achieve a sensitivity of 95% for the detection of tumor-associated copy number aberrations. Computer simulations were performed for scenarios in which the affected region had a copy number change of -1, $+1$, and $+2$ and for fractional concentrations of tumor-derived DNA ranging from 1% to 50%. In each simulation analysis, the entire genome was divided into 3000 bins. This number was similar to the one we used in the actual experimental analysis when a 1-Mb resolution was used.

We assumed that 10% of the bins would exhibit chromosomal aberrations in the tumor tissue. In the tumor tissue, the expected fraction (P) of total molecules falling into a bin within an affected region would be:

$$P = \frac{2 + \mathrm{CN}}{2 \times 3000 + 3000 \times 10\% \times \mathrm{CN}},$$

where CN is the copy number change. From this information, we calculated the expected change in the plasma.

In the plasma, the expected proportion of the total molecules (E) falling into a bin within an affected region can be calculated as:

$$E = P \times f + \frac{1 - f}{3000},$$

where f is the fractional concentration of tumor-derived DNA in plasma.

Simulations of 1000 normal cases and 1000 cancer cases were performed on the assumption of a binomial distribution of the plasma DNA molecules, with the expected plasma representations as calculated above and with an increasing number of molecules being analyzed until the 95% detection rate was reached. The simulation was conducted with the rbinom function in R (http://www.r-project.org/).

DETECTION OF TUMOR-ASSOCIATED SINGLE-NUCLEOTIDE VARIANTS

We sequenced the paired tumor and constitutional DNA samples to identify the tumor-associated single-nucleotide variants (SNVs). We focused on the SNVs occurring at homozygous sites in the constitutional DNA (i.e., buffy coat DNA). In principle, any nucleotide variation detected in the sequencing data of the tumor tissues but absent in the constitutional DNA could be a potential mutation (i.e., a SNV). Because of sequencing errors (0.1%–0.3% of sequenced nucleotides) *(29)*, however, millions of false positives would be identified in the genome if a single occurrence of any nucleotide change in the sequencing data of the tumor tissue were to be regarded as a tumor-associated SNV. One way to reduce the number of false positives would be to institute the criterion of observing multiple occurrences of the same nucleotide change in the sequencing data in the tumor tissue before a tumor-associated SNV would be called. Because the occurrence of sequencing errors is a stochastic process, the number of false positives due to sequencing errors would decrease exponentially with the increasing number of occurrences required for an observed SNV to be qualified as a tumor-associated SNV. On the other hand, the number of false positives would increase exponentially with increasing sequencing depth. These relationships could be predicted with Poisson and binomial distribution functions. In this regard, we have developed a mathematical algorithm to determine the dynamic threshold of occurrence for qualifying an observed SNV as tumor associated. This algorithm takes into account the actual coverage of the particular nucleotide in the tumor sequencing data, the sequencing error rate, the maximum false-positive rate allowed, and the desired sensitivity for mutation detection.

In this study, we set very stringent criteria to reduce false positives. We required a mutation to be completely absent in the constitutional DNA sequencing, and the sequencing depth for the particular nucleotide position had to be >20-fold. This threshold of occur-

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.004

rence was required to control the false-positive detection rate at $<1 \times 10^{-7}$. In this algorithm we also filtered out SNVs that were within centromeric, telomeric, and low-complexity regions to minimize false positives due to alignment artifacts. In addition, putative SNVs mapping to known SNPs in the dbSNP build 135 database were also removed.

Results

TUMOR-ASSOCIATED COPY NUMBER ABERRATIONS IN PLASMA

We investigated whether tumor-associated copy number aberrations could be detected in the plasma of cancer patients by shotgun MPS. Peripheral blood samples were obtained both before and 1 week after surgical resection with curative intent from 4 HCC patients. The blood samples were fractionated into plasma and blood cells. DNA was also obtained from each of the tumors. Copy number aberrations in the 4 tumor samples were analyzed with MPS and with 1 or 2 microarray platforms (Affymetrix and Agilent). Copy number aberrations were analyzed in 1-Mb windows across the genome in the tumor tissues and compared with the plasma samples from a group of 16 healthy control individuals. The data were consistent across the 3 platforms (see Fig. 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol59/issue1).

We then used MPS to analyze the pre- and post-resection plasma samples obtained from all 4 HCC patients. The mean sequencing depth was 17-fold coverage of the haploid human genome (range, 15.2-fold to 18.5-fold). Fig. 1 shows Circos plots *(30)* of the copy number aberrations across the genome in the tumor, the pre-resection plasma sample, and the post-resection plasma sample, for each patient. In each case, characteristic copy number aberrations seen in the tumor tissue sample were also observed in the pre-resection plasma sample (Fig. 1). A significant change in the regional representation of plasma DNA was defined as >3 SDs from the mean representation of the 16 healthy controls for the corresponding 1-Mb window.

For all cases, such copy number aberrations disappeared almost completely in the post-resection plasma sample (Fig. 1). The detectability of the different classes of tumor-associated genetic alterations in plasma is shown in Fig. 2. For comparison, we used the same approach to analyze plasma DNA samples from 4 hepatitis B carriers without HCC (Fig. 1E; see Fig. 2 in the online Data Supplement). These individuals were followed up for 1 additional year after blood sampling and had no evidence of HCC. For these individuals, 99% of the sequenced bins showed normal representations in plasma (see Table 1 in the online Data Supplement). Similarly, a mean of 98.9% of the sequenced bins in the 16 healthy controls showed normal representations in plasma (see Table 2 and Fig. 3 in the online Data Supplement). These results indicate that the analysis of copy number aberrations in plasma is specific for differentiating between cancer patients and individuals without a cancer; however, the specificity for plasma copy number analysis appeared to be reduced in the HCC patients. Hence, in the 4 HCC patients, a median of 15% (range, 2%–48%) of the regions at which no copy number aberrations occurred in the corresponding tumor tissue showed an aberrant plasma DNA representation (see Table 1 in the online Data Supplement). This issue will be discussed in more detail in the Discussion section.

FRACTIONAL CONCENTRATION OF TUMOR DNA IN PLASMA DETERMINED BY GENOMEWIDE AGGREGATED ALLELIC LOSS ANALYSIS

The fractional concentrations of tumor-derived DNA in plasma were determined by analyzing, in a genomewide manner, the allelic counts for SNPs exhibiting LOH in the plasma shotgun MPS data, which we term "genomewide aggregated allelic loss" (GAAL) analysis. For such an analysis, we chose SNPs that exhibited LOH in the tumors as demonstrated with the Affymetrix SNP 6.0 microarray. The alleles deleted in the tumors would have lower concentrations in the plasma than those that were not deleted. The difference in their concentrations was related to the concentration of tumor-derived DNA in the plasma sample. Thus, the plasma concentration of the tumor-derived DNA (C) can be deduced with the following equation:

$$C = \frac{N_{\text{nondel}} - N_{\text{del}}}{N_{\text{nondel}}},$$

where N_{nondel} represents the number of sequenced reads carrying the nondeleted alleles in the tumor tissues, and N_{del} represents the number of sequenced reads carrying the deleted alleles in the tumor tissues.

Table 1 lists the fractional concentrations of tumor DNA in the plasma samples for each of the 4 cases. The size of the tumor appears to be correlated with the estimated fractional concentration of tumor-derived DNA in plasma before surgical resection. For example, we estimated that tumor-derived DNA accounted for 52% of the total plasma DNA in the patient who had the largest tumor (13 cm) of the 4 HCC cases. For each of the 4 cases, we observed a reduction in the fractional concentration of tumor-derived DNA after surgical resection of the tumor (Table 1).

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.005

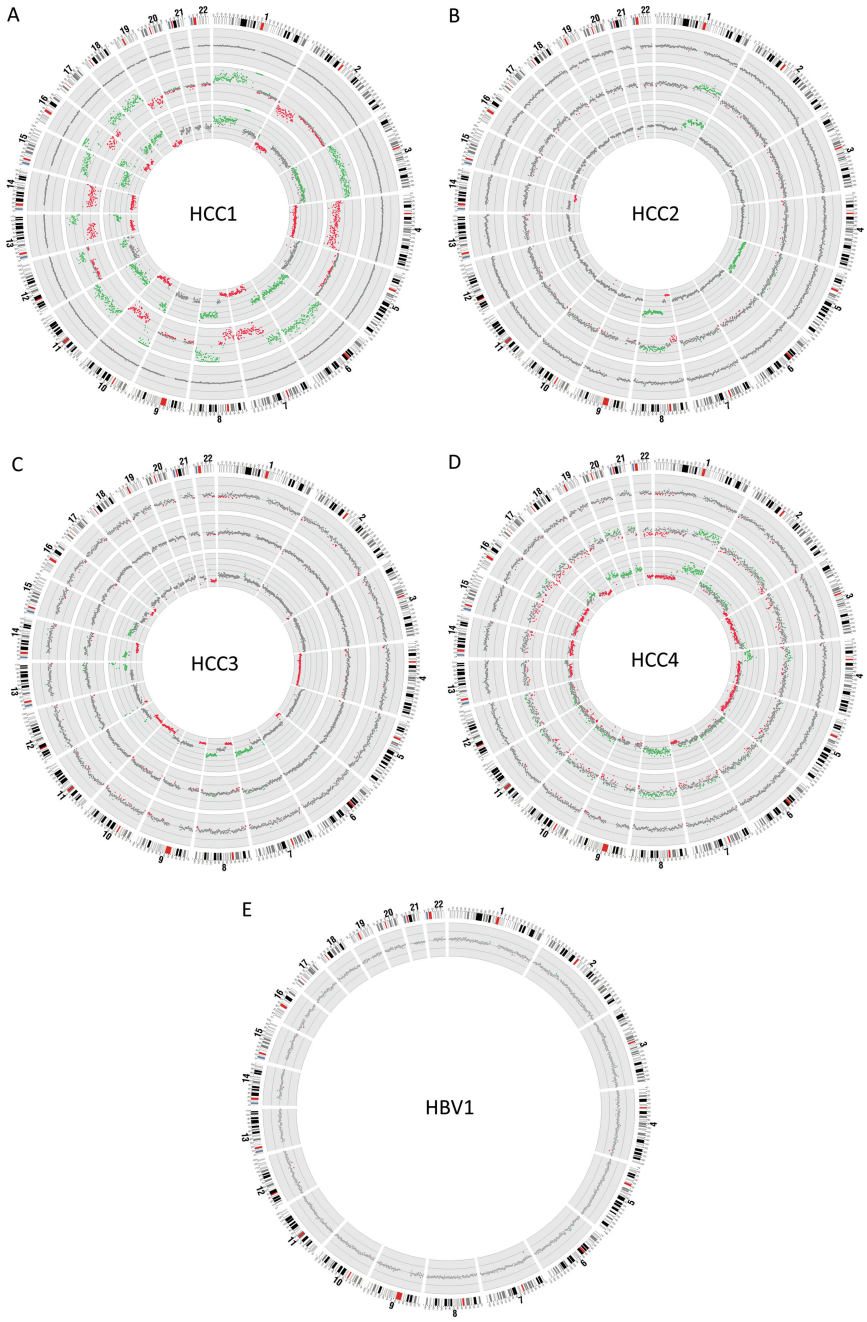

Fig. 1. Copy number aberrations detected in the tumor tissue sample (inner ring), presurgery plasma sample (middle ring), and postsurgery plasma sample (outer ring) for the 4 HCC cases (A–D) and genomic representation of plasma DNA samples for a hepatitis B virus carrier (HBV1) without HCC (E).

Chromosome ideograms (outside the plots) are oriented pter to qter in a clockwise direction. For the tumor tissues, the numbers of copies gained (green) or lost (red) are presented. For the plasma samples, regions with increased and reduced representation in plasma are shown in green and red, respectively. Regions with *z* scores between −3 and +3 are represented by gray dots. For HCC1, the distance between 2 consecutive horizontal lines represents a *z* score of 30. For the other cases, the distance represents a *z* score of 10.

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.006

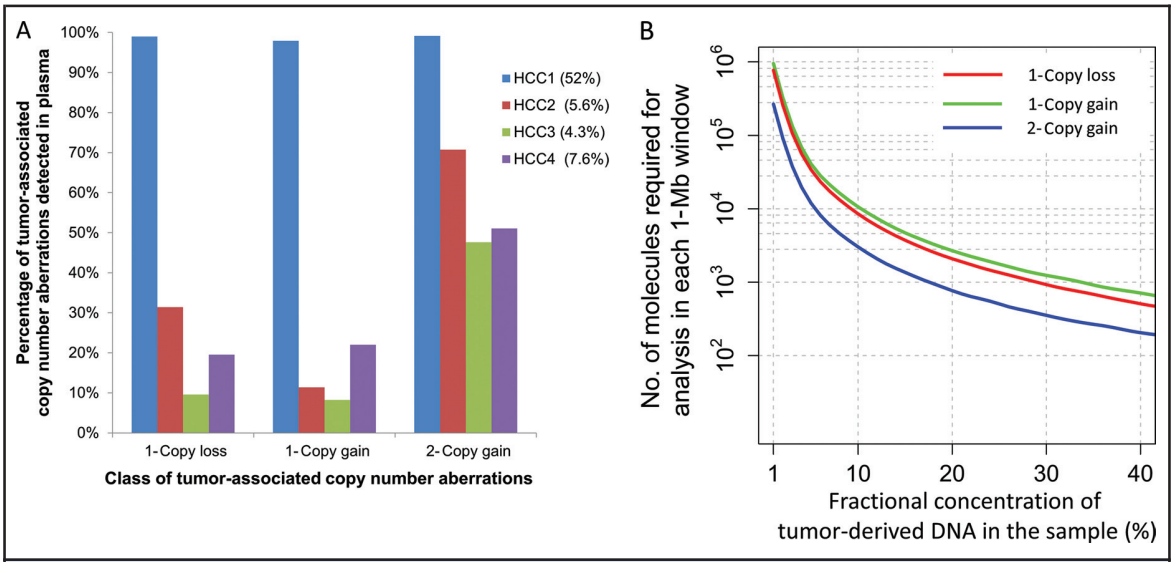

Fig. 2. (A), Detection rates for different classes of tumor-associated copy number aberrations in the plasma of the HCC patients.

The fractional concentration of tumor-derived DNA in plasma (in parentheses) determined for each case by GAAL analysis. (B), Number of molecules required for analysis in each region of interest (1-Mb window) for detecting 95% of the cancer-associated copy number aberrations in plasma.

FACTORS INFLUENCING THE DETECTION OF COPY NUMBER ABERRATIONS IN PLASMA

The fractional concentration of tumor DNA in plasma and the class of copy number aberrations strongly influenced the detectability of such alterations in plasma. Fig. 1 and Fig. 2A show that the proportions of tumor-associated genetic aberrations that could be seen in the plasma DNA–sequencing results were correlated with the fractional concentrations of tumor DNA in plasma. For example, case HCC1, which had the largest tumor and the highest fractional concentration of tumor DNA in plasma, also had the largest proportion of tumor-associated copy number aberrations detected in plasma (Fig. 1A).

Case HCC1 has a fractional concentration of tumor-derived DNA of 52%. Before treatment, most of the tumoral copy number aberrations could also be seen in the plasma. At most of the chromosomal re-

Table 1. Fractional concentration of tumor-derived DNA in plasma by GAAL analysis.

Case	Maximal dimension of the tumor, cm	SNPs with LOH in tumor, n	Plasma			
			Collection time	N_{nondel}[a]	N_{del}	Fractional concentration of tumor-derived DNA
HCC1	13.0	11 310	Before surgery	88 499	42 829	52%
			After surgery	94 442	93 584	0.9%
HCC2	6.2	6040	Before surgery	56 611	53 428	5.6%
			After surgery	63 694	63 083	0.9%
HCC3	4.2	24 783	Before surgery	198 688	190 163	4.3%
			After surgery	236 655	233 280	1.4%
HCC4	6.2	498	Before surgery	2465	2277	7.6%
			After surgery	2773	2699	2.7%

[a] N_{nondel}, number of sequenced reads carrying the nondeleted alleles in tumor tissues; N_{del}, number of sequenced reads carrying the deleted alleles in the tumor tissues.

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.007

Table 2. Fractional concentrations of tumor-derived DNA in plasma determined by SNV analysis.

Case	No. of SNVs detected in tumor tissue	Time point	No. of tumor-associated SNVs sequenced from plasma (percentage of SNVs seen in tumor tissue)	No. of plasma DNA sequence reads showing SNVs (p)	No. of plasma DNA sequence reads showing wild-type sequence (q)	Deduced fractional concentration of tumor-derived DNA by SNV analysis $\left(\frac{2p}{p+q}\right)$	Deduced fractional concentration of tumor-derived DNA by GAAL analysis
HCC1	2840	Before surgery	2569 (94%)	11 389	31 602	53%	52%
		After surgery	44 (1.5%)	91	46 898	0.4%	0.9%
HCC2	3105	Before surgery	1097 (35%)	1490	57 865	5.0%	5.6%
		After surgery	72 (2.3%)	206	66 692	0.6%	0.9%
HCC3	3171	Before surgery	461 (15%)	525	48 886	2.1%	4.3%
		After surgery	31 (1%)	67	58 862	0.2%	1.4%
HCC4	1334	Before surgery	201 (15%)	248	18 527	2.6%	7.6%
		After surgery	74 (5.5%)	149	22 144	1.3%	2.7%

gions with a single-copy gain in the tumor (e.g., chromosomes 1p, 3, and 6), the z scores were >20, indicating that the plasma representation was 20 SDs above the mean representation of the healthy control individuals for these regions. On the other hand, case HCC3 had a fractional tumor DNA concentration in plasma of 4.3%, and a smaller proportion of the cancer-associated aberrations could be observed in the plasma. None of the regions with a single-copy gain (7q, 8q, 13q, and 14p) had a z score of >10 in the plasma.

The classes of copy number aberrations that were studied included 1-copy losses, 1-copy gains, and 2-copy gains. The percentages of such changes that were detected in plasma in each of the 4 HCC cases are plotted in Fig. 2A and listed in Table 3 in the online Data Supplement. For each case, 2-copy gains could be detected with higher sensitivity in plasma than 1-copy changes. For all 4 HCC patients, most of these tumor-associated chromosomal aberrations disappeared after surgical resection of the tumor (Fig. 1; see Table 3 in the online Data Supplement).

Previous work on noninvasive prenatal diagnosis has revealed that a greater sequencing depth would enable the detection in plasma of an aneuploid fetus at a lower fractional fetal DNA concentration *(31)*. Using computer simulation, we explored the relationship between the depths of sequencing that would be needed to detect different classes of tumor-associated copy number aberrations in plasma at different fractional concentrations of tumor DNA in plasma (Fig. 2B). For illustration purposes we fixed the detection rate at 95% and explored 3 classes of genetic aberrations—1-copy loss, 1-copy gain, and 2-copy gain. When the fractional concentration of tumor-derived DNA was 40%, the detection of aberrations with 2-copy gains and 1-copy gains would require the analysis of approximately 180 and 800 molecules, respectively, per 1-Mb window. When the fractional concentration of tumor-derived DNA drops to 10%, the analysis of approximately 2500 and 12 000 molecules per 1-Mb window is necessary to detect these respective changes. The requirement of an exponential increase in the number of molecules with a decreasing fractional concentration of tumor-derived DNA was consistent with the requirement for noninvasive prenatal diagnosis of fetal chromosomal aneuploidies via analysis of maternal plasma DNA *(32)*.

TUMOR-DERIVED SNVs IN PLASMA

We next explored the genomewide detection of tumor-derived SNVs in the plasma of the 4 HCC patients. We sequenced tumor DNA and buffy coat DNA to mean depths of 29.5-fold (range, 27-fold to 33-fold) and 43-fold (range, 39-fold to 46-fold) haploid genome coverage, respectively. The MPS data from the tumor DNA and the buffy coat DNA from each of the 4 HCC patients were compared, and SNVs present in the tumor DNA but not in the buffy coat DNA were mined with a stringent bioinformatics algorithm. This algorithm required a putative SNV to be present in at least a threshold number of sequenced tumor DNA fragments before it would be classified as a true SNV. The threshold number was determined by taking into account the sequencing depth of a particular nucleotide and the sequencing error rate.

The number of tumor-associated SNVs ranged from 1334 to 3171 in the 4 HCC cases. The proportions of such SNVs that were detectable in plasma are listed in Table 2. Before treatment, 15%–94% of the tumor-associated SNVs were detected in plasma. The fractional concentrations of tumor-derived DNA in plasma were

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.008

determined by the fractional counts of the mutant with respect to the total (i.e., mutant plus wild type) sequences (Table 2). These fractional concentrations were well correlated with those determined with GAAL analysis and were reduced after surgery (Table 2).

To estimate the specificity of the SNV analysis approach, we analyzed the plasma of the healthy controls for the tumor-associated SNVs (see Table 4 in the online Data Supplement). The presence of a small number of these putative tumor-associated mutations in the plasma of the 16 healthy controls represented the "stochastic noise" of this method and was likely due to sequencing errors. The mean fractional concentration estimated from such noise was 0.38%.

PLASMA DNA ANALYSIS IN A PATIENT WITH MULTIPLE SYNCHRONOUS CANCERS

To illustrate the use of shotgun MPS of plasma DNA to monitor a complex oncologic scenario, we studied a 58-year-old female patient with a *BRCA1* (breast cancer 1, early onset) mutation (p.Cys1697*) presenting with synchronous breast and ovarian cancers (Fig. 3). The breast cancer was a 3-cm infiltrating ductal carcinoma located in the left breast. The patient had serous adenocarcinomas of both ovaries. The tumor in the left ovary measured 6 cm in the longest dimension, and the one on the right side measured 12 cm. There were also multiple intra-abdominal tumor deposits involving the omentum and the colon. Surgical resections of the breast tumor and the ovarian tumors, together with the omentum and the sigmoid colon, were performed on the same day. The breast tumor and the ovarian tumor tissues from the left and right sides were collected for this study. Plasma samples were also collected at diagnosis and 1 day after the operation. The breast tumor DNA and 4 regions of the ovarian tumors (2 from the left ovary and 2 from the right ovary) were analyzed by MPS. The regions sampled for the ovarian tumor on the same side were separated by 4 cm.

Copy number aberrations for the breast and ovarian cancers are plotted in Fig. 4A. The 4 ovarian tumoral regions showed highly similar patterns of copy number aberrations. On the other hand, the breast cancer exhibited a different pattern of copy number aberrations. The pattern of aberrations in the presurgery plasma sample was a composite of the patterns of aberrations in the breast and ovarian cancers (Fig. 4B). Examples of genetic aberrations that were specific to the breast cancer included a deleted segment on chromosome 6p and amplifications on chromosomes 1q, 7p, and 15q (Fig. 4A). On the other hand, examples of genetic aberrations that were specific to the ovarian tumors included deletions on chromosomes 2, 4p, 11p, 12q, 18q, and 22q

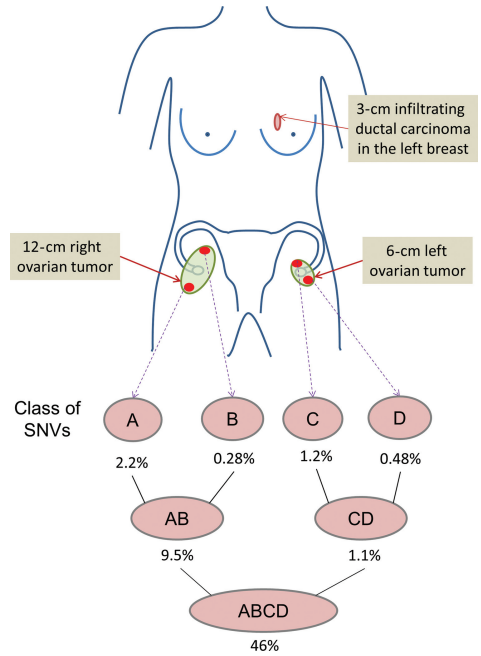

Fig. 3. Illustration showing the locations of the tumors in a patient with synchronous breast and ovarian cancers.

Four regions (A, B, C, and D) from the ovarian tumors were sampled. Shown are the fractional concentrations of tumor-derived plasma DNA estimated with SNVs associated with each of the ovarian tumoral regions. The calculated fractional concentration of tumor DNA in plasma increases when the estimate is based on SNVs associated with multiple tumor regions.

and amplifications on chromosomes 3q, 5p, and 21q (Fig. 4B). These genetic aberrations specific to the breast cancer or to the ovarian cancers were all observed in the presurgery plasma sample but were cleared in the postsurgery plasma sample (Fig. 4A). An expanded view of 2 genomic regions exhibiting copy number aberrations that were present in the breast cancer but absent in the ovarian tumors is shown in Fig. 4C.

To examine the relative contributions made by the breast and ovarian cancers to the plasma DNA of the patient, we conducted GAAL analyses for genomic regions that exhibited deletions specific to either the breast tumor or the ovarian tumors. These results indicated that the breast cancer and the ovarian cancers contributed 2.1% and 46%, respectively, of the DNA in plasma before surgery (see Table 5 in the online Data Supplement). The fractional concentrations of DNA contributed by each of these tumors dropped to 1.3%

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.009

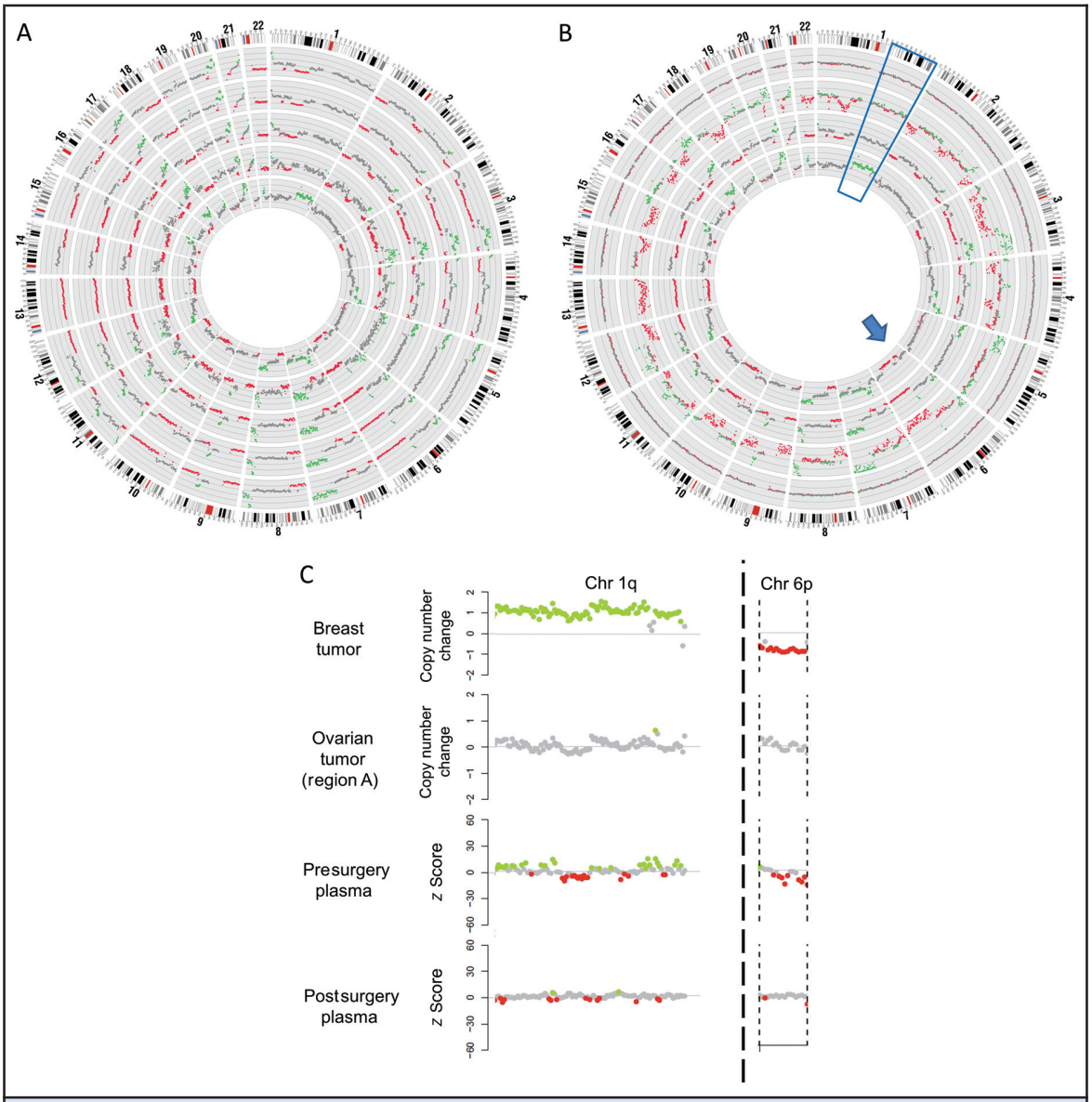

Fig. 4. (A), Copy number aberrations detected in the breast cancer (innermost ring) and the ovarian tumors (second innermost ring outwards: regions A, B, C, and D).

(B), Copy number aberrations in the breast tumor, ovarian tumor (region A), presurgery plasma, and postsurgery plasma (from inside to outside). Two genomic regions exhibiting copy number aberrations specific to the breast cancer are marked with a box (chromosome 1q) and an arrow (chromosome 6p). (C), Expanded view of the chromosome 1q and 6p regions in tumor and plasma samples.

and 0.66%, respectively, after surgery (see Table 5 in the online Data Supplement).

INVESTIGATION OF TUMORAL HETEROGENEITY

We further explored the phenomenon of tumoral heterogeneity *(33)* by studying the 4 regions sampled from the ovarian tumors. There were no observable differences among these 4 regions in terms of copy number aberrations (Fig. 4A). We compared the SNV profile from each of these regions with the buffy coat DNA of the patient. We classified the SNVs into 7 groups: 4 groups containing mutations that were unique to each

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.010

Table 3. SNV analysis of multiple regions of the ovarian cancer in the patient with synchronous cancers.

		Presurgery plasma			Postsurgery plasma		
Group of SNVs	No. of SNVs in tumor	No. of sequence reads showing SNVs in plasma (p)	No. of sequence reads showing wild-type sequence in plasma (q)	Deduced fractional concentration of tumor-derived DNA $\left(\frac{2p}{p+q}\right)$	No. of sequence reads showing SNVs in plasma (p)	No. of sequence reads showing wild-type sequence in plasma (q)	Deduced fractional concentration of tumor-derived DNA $\left(\frac{2p}{p+q}\right)$
A	71	37	3321	2.2%	5	2149	0.46%
B	122	8	5633	0.28%	1	3516	0.06%
C	168	51	8507	1.2%	2	5438	0.07%
D	248	26	10 880	0.48%	2	7060	0.06%
AB	10	21	423	9.5%	0	297	0%
CD	12	3	569	1.1%	0	333	0%
ABCD	2216	21 344	70 940	46%	54	61 417	0.18%

region (i.e., groups A, B, C, and D), 2 groups that contained mutations shared by the 2 regions on each side of the body (i.e., groups AB and CD), and a final group containing mutations shared by all 4 regions (i.e., group ABCD) (Table 3 and Fig. 3). We randomly selected approximately 10 SNVs for each of the 7 SNV groups for validation with a mass spectrometry–based single-nucleotide extension method (iPLEX analysis; Sequenom) (see Table 6 in the online Data Supplement). A total of 67 mutations were subjected to validation. More than 95% of the SNV results determined by sequencing were validated in the iPLEX analyses.

SNVs from each of these groups were then sought out in the plasma shotgun MPS data. The fractional concentrations of circulating tumor DNA were determined with each SNV group (Table 3). The fractional concentrations of tumor DNA in plasma before surgery and after surgery, as determined by SNVs shared by all 4 regions (i.e., group ABCD), were 46% and 0.18%, respectively. These latter percentages correlated well with those obtained in GAAL analyses (see Table 5 in the online Data Supplement). The fractional concentrations of tumor-derived DNA in preoperative plasma determined with SNVs from groups AB and CD were 9.5% and 1.1%, respectively (Table 3). These concentrations were consistent with the relative sizes of the right and left ovarian tumors (Fig. 3). The fractional concentrations of tumor-derived DNA determined with the region-unique SNVs (i.e., those in groups A, B, C, and D) were generally low. The trend of a reduction in the observed fractional tumor-derived DNA in plasma as measured with SNVs of increasing "regional specificity" will be explored in the Discussion.

Discussion

Our data indicate that shotgun sequencing of plasma samples from cancer patients would allow cancer-associated copy number aberrations and mutations to be analyzed noninvasively and in a genomewide fashion (Figs. 1 and 4). This finding is an important step forward from previous work, which has generally been focused on detecting a small number of tumor-associated genetic changes in plasma *(1)*. This approach would allow genomic aberrations harbored by tumor cells within a patient to be scanned at different levels of resolution, ranging from a global view of cancer-associated copy number aberrations to point mutations carried by different clones of tumor cells. This approach would allow qualitative and quantitative aspects of such aberrations to be followed serially, as evidenced by the changes we observed in the pre- and postsurgery plasma samples. The method would also allow tumor load to be assessed, as indicated by the correlation between the fractional tumor DNA concentration measured and the size of the tumor. In this study, these general characteristics were observed in the plasma DNA profiles derived from HCC, breast cancer, and ovarian cancer.

Both GAAL analysis and SNV analysis allow the fractional concentrations of tumor-derived DNA in plasma to be measured for any cancer. These approaches would allow one to compare DNA released into the plasma from different types of tumors. Through the pooling of the sequence counts from SNPs involved in LOH and point mutations across the genome, the measurements made by these analyses are expected to be much more precise than those made on

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.011

the basis of individual tumor-associated genetic changes.

The control group for establishing the baseline for plasma-based copy number analysis consisted of healthy individuals without cancer. The specificity of this type of analysis for individuals without cancer can been seen from bootstrap analysis of this control group, as well as in chronic hepatitis B carriers without HCC. For these 2 groups, only a small fraction of regions had aberrant plasma DNA representation with a z score <-3 or >3. These results indicate that this approach is specific for differentiating between cancer patients and individuals without cancer. When this approach was used for analyzing the plasma of HCC patients, however, the specificity appeared to be lower than for study participants without cancer. The explanation for this observation is not entirely clear at the moment; however, we can think of at least 2 possible explanations. First, the presence of tumor DNA with regions exhibiting copy number aberrations in the plasma of cancer patients would affect the observed relative contribution of DNA from regions not exhibiting copy number aberrations. For example, for a tumor exhibiting many copy number losses across the genome, one would observe a relatively increased contribution of plasma DNA originating from genomic regions with a normal copy number. This effect would be greatest for plasma samples containing high concentrations of tumor DNA. Indeed, the percentages of misclassification of normal regions are the highest for case HCC1 (see Table 1 in the online Data Supplement), which has the highest fractional tumor DNA concentration in plasma (52%). Second, the phenomenon of tumoral heterogeneity could explain a proportion of such apparent false-positive results. In other words, the single tumoral region sampled from each HCC tumor might not contain a particular copy number aberration that has been released into the plasma from another tumor clone not contained in the sampled region. Although the phenomenon mentioned in this paragraph requires more study, it does not appear at present to adversely affect the ability to distinguish between individuals with cancer and those without cancer.

The patient with synchronous breast and ovarian cancers is particularly noteworthy, because the case illustrates a number of important concepts. First, it shows that careful analysis of the sequencing data permits one to dissect out the presence of DNA in the plasma originating from individual tumor cell populations. Targeting genomic regions exhibiting copy number aberrations specific to the breast and ovarian cancers allowed the elucidation of the relative contributions of these tumors to the circulating DNA pool (Fig. 4B; see Table 5 in the online Data Supplement).

To study the effect of tumoral heterogeneity on the plasma DNA profile, we analyzed the mutational profile of the 4 regions of the bilateral ovarian cancer via comparisons with the patient's constitutional DNA. We were able to group these mutations into 7 categories according to the degree of sharing of these mutations between the 4 regions. It is interesting that mutations that were shared by all 4 regions (i.e., group ABCD) contributed the highest fractional contribution of tumor-derived DNA to the plasma. On the other hand, mutations that were more region specific had a reduced contribution to plasma. It was therefore not surprising that the GAAL analysis, which is based on aggregating the amount of allelic loss across the genome, produced a fractional tumor DNA concentration similar to that of the analysis based on SNVs from the ABCD category. These data suggest that for an accurate measurement of the total tumor load in a cancer patient, the use of a genomewide shotgun approach might provide a more representative picture, compared with the more traditional approach of targeting specific tumor-associated mutations. For the latter approach, if only a subset of the tumor cells possesses the targeted mutations, one might miss important information regarding imminent relapse or disease progression caused by tumor cells not possessing the targeted mutations, or one might miss the emergence of a treatment-resistant clone.

Indeed, the phenomenon of tumoral heterogeneity *(33)* has created challenges for developing truly representative tumor markers for monitoring purposes. Because plasma receives DNA from the various heterogeneous tumor clones in the body, shotgun sequencing of plasma DNA might be a readily available and noninvasive method for studying and monitoring tumoral heterogeneity and the total tumor burden in the body.

In this work, we used tumor-associated SNVs as tumor markers and did not focus on elucidating the biology of the mutational landscape of the studied cancers. Thus, the present work has not carried out a detailed characterization of the mutated sequences. We expect, however, that most of the tumor-associated SNVs we detected were "passenger" rather than "driver" mutations. It was the aim of this study to develop a noninvasive approach to assess the tumor burden of the body. We therefore believe that the detection of mutations—passengers and drivers—in a genomewide manner provides a numerical advantage that has the potential for more sensitive, precise, and representative detection of tumor DNA in plasma. On the other hand, selective analyses of driver mutations may allow more readily the identification of medically actionable targets *(8)*. Consequently, the nonselective genomewide approach can be combined in a synergistic man-

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.012

ner with the more conventional detection strategies for selective targets to achieve better cancer management.

Given that plasma DNA in cancer patients consists of a mixture of tumor and nontumor DNA, with the former being present as a minority population in many patients, one needs to perform relatively deep sequencing to obtain the necessary analytical power. Thus, as currently implemented, the shotgun sequencing approach we have described is relatively expensive compared with many targeted approaches *(8)*. With the continuing and, thus far, rapid reduction in sequencing costs, cost is not expected to be as much of an obstacle in the near future. One can also envision the possibility of using shotgun sequencing of a plasma sample upon presentation of a cancer patient, mining the shotgun sequencing data of plasma DNA for copy number aberrations and SNVs, and then specifically targeting such sequences for serial monitoring purposes. In this regard, solution-phase hybridization-based capture approaches have already been successfully used to enrich plasma DNA for noninvasive prenatal diagnosis *(34, 35)*. Hence, we see that the targeted and shotgun approaches can be used in combination for cancer detection and monitoring.

In summary, we have demonstrated the performance of cancer genome scanning via shotgun sequencing of plasma DNA. This development has numerous clinical and research applications. It would therefore be a research priority to evaluate this approach on a large cohort of cancers of multiple types.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: Y.M.D. Lo, *Clinical Chemistry*, AACC.
Consultant or Advisory Role: R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom.
Stock Ownership: R.W.K. Chiu, Sequenom; Y.M.D. Lo, Sequenom.
Honoraria: R.W.K. Chiu, Life Technologies (travel grants) and Illumina (travel grants); Y.M.D. Lo, Illumina and Life Technologies.
Research Funding: K.C.A. Chan, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10); H. Sun, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10); A.T.C. Chan, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10), and the Innovation and Technology Fund under the State Key Laboratory Programme; R.W.K. Chiu, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10) and S.K. Yee Foundation; Y.M.D. Lo, Hong Kong Research Grants Council Theme-based Research Scheme (T12-CUHK05/10), S.K. Yee Foundation, Innovation and Technology Fund under the State Key Laboratory Programme, and an endowed chair from the Li Ka Shing Foundation.
Expert Testimony: None declared.
Patents: K.C.A. Chan, P. Jiang, R.W.K. Chiu, and Y.M.D. Lo have all declared the same set of 5 US patent applications filed on this work: 13/308473, 61/662878, 61/682725, 61/695795, and 61/711172.
Data and Materials Availability: Sequence data and genotype data have been deposited at the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the European Bioinformatics Institute (EBI), under accession number EGAS00001000370.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank L. Chan, Y. Jin, C. Lee, K. Chow, S.-W. Yeung, X. Su, and C. Chan for technical assistance.

References

1. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 2011;11:426–37.
2. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. N Engl J Med 2011;364:340–50.
3. Nawroz H, Koch W, Anker P, Stroun M, Sidransky D. Microsatellite alterations in serum DNA of head and neck cancer patients. Nat Med 1996;2:1035–7.
4. Chen XQ, Stroun M, Magnenat JL, Nicod LP, Kurt AM, Lyautey J, et al. Microsatellite alterations in plasma DNA of small cell lung cancer patients. Nat Med 1996;2:1033–5.
5. Yung TK, Chan KCA, Mok TS, Tong J, To KF, Lo YMD. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. Clin Cancer Res 2009;15:2076–84.
6. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A 2005;102:16368–73.
7. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2008;14:985–90.
8. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012;4:136–68.
9. Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 2012;486:537–40.
10. Wong IH, Lo YMD, Zhang J, Liew CT, Ng MH, Wong N, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res 1999;59:71–3.
11. Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin Cancer Res 2011;17:4494–503.
12. Lo YMD, Chan LYS, Lo KW, Leung SF, Zhang J, Chan AT, et al. Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. Cancer Res 1999;59:1188–91.
13. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarker using massively parallel sequencing. Sci Transl Med 2010;2:20–14.
14. McBride DJ, Orpana AK, Sotiriou C, Joensuu H, Stephens PJ, Mudie LJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosomes Cancer 2010;49:1062–9.
15. Lo YMD, Chiu RWK. Genomic analysis of fetal nucleic acids in maternal blood. Annu Rev Genomics Hum Genet 2012;13:285–306.
16. Chiu RWK, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, et al. Noninvasive prenatal diagnosis

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.013

of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc Natl Acad Sci U S A 2008;105: 20458–63.
17. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, et al. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet Med 2011;13:913–20.
18. Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet Gynecol 2012;119:890–901.
19. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61–91.
20. Kitzman JO, Snyder MW, Ventura M, Lewis AP, Qiu R, Simmons LE, et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med 2012;4:137–76.
21. Chan KCA, Zhang J, Hui AB, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 2004;50:88–92.
22. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 2009;25: 1966–7.
23. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet 2009;41:334–41.
24. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 2008;40:1253–60.
25. De Keersmaecker K, Real PJ, Gatta GD, Palomero T, Sulis ML, Tosello V, et al. The TLX1 oncogene drives aneuploidy in T cell transformation. Nat Med 2010;16:1321–7.
26. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 2008;36:e105.
27. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 2009;41:1061–7.
28. Chen EZ, Chiu RW, Sun H, Akolekar R, Chan KC, Leung TY, et al. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. PLoS One 2011;6: e21791.
29. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol 2011;12:R112.
30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639–45.
31. Chiu RW, Akolekar R, Zheng YW, Leung TY, Sun H, Chan KC, et al. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 2011;342:c7401.
32. Lo YMD, Lun FMF, Chan KCA, Tsui NBY, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci U S A 2007;104:13116–21.
33. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366: 883–92.
34. Liao GJ, Lun FM, Zheng YW, Chan KC, Leung TY, Lau TK, et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. Clin Chem 2011;57:92–101.
35. Lam KW, Jiang P, Liao GJ, Chan KCA, Leung TY, Chiu RW, Lo YMD. Noninvasive prenatal diagnosis of monogenic diseases by targeted massively parallel sequencing of maternal plasma: application to β-thalassemia. Clin Chem 2012;58:1467–75.

Downloaded from https://academic.oup.com/clinchem/article/59/1/211/5622098 by Weil Gotshal and Manges LLP (Global) user on 11 October 2023

NEOGENOMICS EX.1059.014