Review

Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies 🔄

Stanislav Volik[1], Miguel Alcaide[2], Ryan D. Morin[2,3], and Colin Collins[1,4]

Abstract

Precision oncology is predicated upon the ability to detect specific actionable genomic alterations and to monitor their adaptive evolution during treatment to counter resistance. Because of spatial and temporal heterogeneity and comorbidities associated with obtaining tumor tissues, especially in the case of metastatic disease, traditional methods for tumor sampling are impractical for this application. Known to be present in the blood of cancer patients for decades, cell-free DNA (cfDNA) is beginning to inform on tumor genetics, tumor burden, and mechanisms of progression and drug resistance. This substrate is amenable for inexpensive noninvasive testing and thus presents a viable approach to serial sampling for screening and monitoring tumor progression. The fragmentation, low yield, and variable admixture of normal DNA present formidable technical challenges for realization of this potential. This review summarizes the history of cfDNA discovery, its biological properties, and explores emerging technologies for clinically relevant sequence-based analysis of cfDNA in cancer patients. Molecular barcoding (or Unique Molecular Identifier, UMI)-based methods currently appear to offer an optimal balance between sensitivity, flexibility, and cost and constitute a promising approach for clinically relevant assays for near real-time monitoring of treatment-induced mutational adaptations to guide evidence-based precision oncology. *Mol Cancer Res; 14(10); 898–908. ©2016 AACR.*

Early Discovery and Applications of cfDNA

The presence of cell-free DNA (cfDNA) in blood plasma was discovered in 1948 by Mandel and Metais (1). Seventeen years later, in 1965, Bendich and colleagues hypothesized, that cancer-derived cfDNA could be involved in metastasis (2). However, it took another year to discover the first link to disease. In 1966, Tan and colleagues observed high levels of circulating cell-free DNA (cfDNA) in the blood of systemic lupus erythematosus patients (3). Eleven years later, in 1977, Leon and colleagues used radio-immunochemistry to demonstrate that for at least half of cancer patients the level of cfDNA in their blood was significantly higher than in normal control subjects (4). The authors noted that patients with metastatic cancer had significantly higher cfDNA levels in blood. Because of technological limitations, it took

another 12 years for the first experimental evidence to support that cfDNA in cancer patients does indeed contain tumor DNA based on temperature stability measurements (5).

The technological progress of the 1990s fuelled by the Human Genome Project allowed more direct demonstration for a tumor origin of at least some cancer patient cfDNA. In 1994, two groups reported the presence of tumor-specific mutations in cfDNA (6, 7). Both groups used mutation-specific primers to facilitate PCR amplification of tumor-specific (*N-RAS*) mutations in the plasma samples of patients with pancreatic adenocarcinoma and acute myelogenous leukemia (AML), respectively. This approach to detection of specific *a priori* known mutations in cfDNA was to become the preferred method of cfDNA studies until the advent of massively parallel sequencing (MPS). Circulating tumor DNA (ctDNA) is typically so diluted by normal DNA that existing sequencing methods (e.g., Sanger sequencing) were not sufficiently sensitive to detect mutant DNA molecules. As a result, mutation-specific PCR was the only available technology that could provide sufficient specificity for detection of the weak tumor signal. It was recognized in these pioneering studies that the detection of tumor DNA in circulation offers exciting implications for clinical translation for "diagnosis, determining response to treatment, and predicting prognosis" (7). Not surprisingly, soon after this initial breakthrough essentially all other types of tumor-specific DNA changes were discovered in cfDNA, such as changes in the status of microsatellite markers including loss of heterozygosity (LOH; refs. 8, 9); gene amplifications (10, 11); the presence of the oncogenic viral DNA (12–15); and hypermethylation of the promoter regions of tumor suppressor genes (16–18). While these early observations highlighted many possibilities for using ctDNA as a noninvasive approach to analyze tumor genomes, sufficiently sensitive and specific laboratory techniques to fully leverage this potential were not yet developed.

[1]Vancouver Prostate Centre, Vancouver, British Columbia, Canada. [2]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada. [3]Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. [4]Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada.

Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).

S. Volik and M. Alcaide contributed equally to this article.

Corresponding Authors: Ryan D. Morin, Department of Molecular Biology and Biochemistry, Simon Fraser University, South Sciences Building Room 8166, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. Phone: 778-782-9581; Fax: 778-782-5583; E-mail: rmorin@bcgsc.ca; and Colin Collins, Jack Bell Research Centre, 2660 Oak Street, Vancouver, British Columbia V6H 3Z6, Canada. Phone: 604-875-4818; Fax: 604-875-5654; E-mail: ccollins@prostatecentre.com
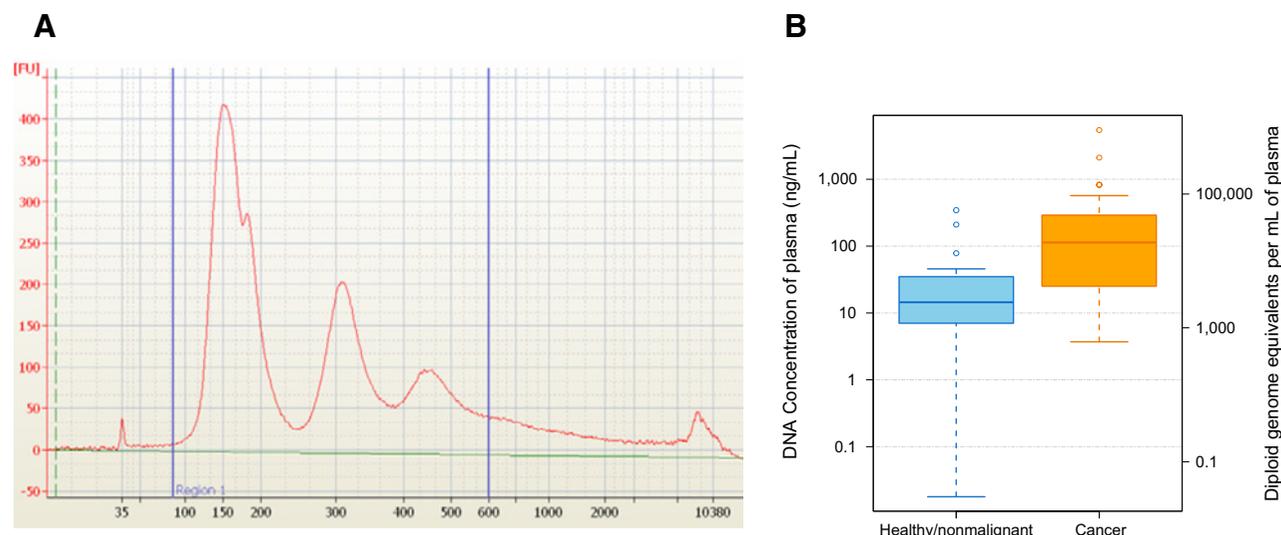
AACR

Figure 1.
A, a Bioanalyzer 2100 (Agilent Technologies) profile showing the length distribution of cfDNA isolated from a high-risk prostate cancer patient reported in ref. 31. The x-axis shows the size of DNA fragments (in base pairs); the y-axis, fluorescent intensity, proportional to DNA concentration. Note the unusual trimodal distribution of DNA sizes (150, 300, and 450 bp) most likely indicating a very high tumor burden. **B,** cfDNA concentration in blood of healthy subjects/nonmalignant patients (blue box) and in cancer patients (orange box) based on ref. 30. The left y-axis shows the cfDNA concentration on a logarithmic scale (in ng/mL of blood) and the right shows the number of diploid genome equivalents per mL of blood (assuming 6 pg of DNA per diploid genome).

The "chimeric" nature of cfDNA, the presence of both normal and tumor DNA in blood plasma, enabled development of applications in other fields. However, these are outside of the scope of the current review and will only be briefly mentioned here. Arguably the most successful application of cfDNA studies was the discovery of the high admixture of fetal-derived cfDNA in mother's blood stream by Lo and colleagues (19). Later the same group demonstrated that in 70% of women bearing male fetuses, fetal Y-chromosome sequences could be detected in just 10 μL of blood plasma (20). This discovery opened up a new avenue for development of fetal cfDNA-based prenatal genetic testing. For example, in one recent large study involving 1,914 women across 21 U.S. centres it was shown that cfDNA-based prenatal testing for fetal aneuploidy has a significantly lower false positive rate for detection of trisomies 21 and 18 compared with standard procedures (10 times lower for trisomy 21 and 3 times lower for trisomy 18) and significantly higher negative and positive predictive values (ref. 21; for detailed review of prenatal diagnostics application of cfDNA, see ref. 22). Another interesting application of cfDNA-based detection of "foreign" DNA is monitoring the status of solid organ transplants. As DNA is mostly released into the blood as a result of cell death, the level of the donor's DNA in recipients blood can be used a marker of rejection (for detailed review of this cfDNA application, see ref. 23).

Later studies linked cfDNA levels to outcomes in severe injury such as blunt trauma and burns (reviewed in ref. 24). cfDNA levels correlated with the length of hospital stay, burn surface area, the number of operations needed for scalds (though not for the flash/ flame burns). Plasma cfDNA levels also correlated with the need for patient ventilation in intensive care units (ICU). High (above 127 ng/mL) concentrations of cfDNA in blood were found to be a predictor of death for ICU patients (with sensitivity of 92% and specificity of 82%). In line with these findings, cfDNA levels in blood turned out to be higher and have certain predictive value for sepsis and septic shock, aseptic inflammation, myocardial infarction, stroke including patients with negative neuroimaging results, where cfDNA concentrations seem to predict poststroke morbidity and mortality in patients with negative neuroimaging, and sickle cell disease. In short, cfDNA concentration is elevated in conditions that involve increased rates of cell death and necrosis.

Properties of cfDNA

Multiple properties of cfDNA suggest cell death as its major origin. Importantly, cfDNA is double stranded and highly fragmented, with most molecules being approximately 150 bp in length (Fig. 1A). This matches the length of DNA occupied by a nucleosome, the primary unit for spatial organization of DNA in the nucleus (25). Moreover, the other fragment length peaks correspond well with linear progression of nucleosome units (two units for 300-bp band, three units for 450-bp band; Fig. 1A). Interestingly, there is still controversy on whether the higher or lower integrity of cfDNA is associated with cancer. In 2003, Wang and colleagues reported that the comparisons of the relative amounts of 100- and 400-bp PCR products of the β-actin gene demonstrated increased cfDNA integrity in 61 patients with breast and gynecologic cancers compared with 65 non-neoplastic patients (26). This observation is supported by the studies in numerous cancer types summarized in ref. 27. However, there are also conflicting reports, including by Madhavan and colleagues (27), who determined that decreased cfDNA integrity (defined as the ratio of concentrations of long, ∼260-bp Alu and LINE fragments to short, ∼100 bp, fragments determined by qPCR) correlates with worse outcome. The authors noted that the decreased cfDNA integrity would imply higher apoptotic rates, and that increased apoptosis correlates with higher tumor proliferation. This in turn would imply that the apoptotic, not necrotic cells, are the main source of cfDNA at least in cancer

patients (28). One corollary of this supposition is that tumors with higher proliferation may naturally yield higher proportion of cfDNA in line with well-established link between cell proliferation and apoptosis rates (for review, see ref. 29). In any case, all of the studies and our experience agree, that as much as 90% of the total cfDNA is contained in low molecular weight band (~150–180 bp).

The amount of cfDNA in cancer patients varies widely. A good summary can be found in 2008 Fleischhacker and Schmidt review (30) who assembled the results of 34 studies involving healthy subjects and patients with malignant and nonmalignant disease (summarized in Fig. 1B). While a trend toward the DNA concentration in the blood of cancer patients being much higher than in the blood of healthy controls and nonmalignant patients is clear, the cfDNA concentration varies considerably and is below 100 ng/mL for the majority of reported cancer patients. This is in line with our own data—in a group of 62 castrate-resistant metastatic prostate cancer (mCRPC) patients the mean cfDNA concentration was 53 ng/mL of blood (31). Another way to look at these numbers is to calculate how many genome equivalents can be identified in blood. Assuming 6 pg of DNA per diploid human genome, the majority of cancer patients have below 17,000 genome equivalents per mL of blood. Patients from our recent study (31) had on average ~9,000 genome equivalents per mL of blood.

Not unexpectedly, in addition to varying absolute levels of cfDNA, the fraction of DNA molecules in the circulation of cancer patients that can be recognized as being derived from tumor cells also varies. In 2001, Jahr and colleagues published one of the first attempts to estimate the proportion of circulating tumor DNA (ctDNA) to total cfDNA (32). The ctDNA/cfDNA ratio was determined by quantifying the amount of hypermethylated *CDKN2A* promoter that the authors assumed to be tumor-specific. Hypermethylation of the *CDKN2A* promoter was detected in 11 of 25 specimens, "in line with previous studies," and in all six cases where both tumor tissue and cfDNA was available, results of methylation-specific PCR were concordant. The proportion of tumor-specific hypermethylated *CDKN2A* sequences ranged from <10% to >90% of the total cfDNA. Four years later, Diehl and colleagues published a study (33) in which they reported that the tumor content in cfDNA of 33 colorectal cancer patients ranged from 0.01% to 1.7%. Interestingly, they also reported, that the percentage of mutant molecules of the APC gene increased 5- to 20-fold when the fragment size used for PCR decreased from 1,296 to 100 bp. Numerous studies have shown similar results. For example, amplicon sequencing of the *PIK3CA* and *TP53* genes and digital PCR of the identified structural variations and point mutations allowed Dawson and colleagues to determine that in metastatic breast cancer the ctDNA defined as a fraction of the somatic mutant allele comprised a median of 4% of total cfDNA (interquartile range 1–14; ref. 34). This is in line with our study (31), where we identified mutations in exon 8 of the AR gene in cfDNA of metastatic castrate resistant prostate cancer patients at a frequency of 0.1%–23% (median 1.5%). Given the total cfDNA content estimated above we expect the most cancer patients to have less than 3,500 tumor genomes per mL of blood. In our cohort of mCRPC patients (31), the median yield of tumor DNA was approximately 135 genome equivalents per mL of blood.

Finally, an important characteristic of cfDNA is its rapid turnover. The first report in the kinetics of foreign DNA clearance from animal's blood dates back to 1963, when Tsumita and Iwanaga used tritiated DNA injected into mice to show that 99% of the radioactivity is cleared from the bloodstream in 30 minutes (35). They have also reported the highest increase of radioactivity in kidneys, followed by liver and spleen, suggesting the importance of renal clearing of cfDNA. Later, in 1999, Lo and colleagues measured the half-life of the fetal cfDNA in mother's blood post-partum using real-time quantitative PCR of the *SRY* gene (36). They reported that the half-life of the male fetal DNA in women post-partum was 16.3 minutes, no detectable male-derived cfDNA were found in the mother's blood 2 hours after birth—a result that is very close to the original 1963 observation. Other studies confirmed the very short half-life of cfDNA in blood stream. For example, Fatouros and colleagues have measured the kinetics of cfDNA concentrations in athletes following vigorous exercise and reported that cfDNA increased to 15-fold postexercise, stabilized at 13-fold for 30 minutes after exercise and normalized 30 minutes later (37). In short, there is no controversy on the kinetics of cfDNA clearance from blood stream; however, the mechanisms of its clearance have not been studied in detail.

Collection and Processing of Blood for cfDNA

As discussed above, a large variability in the total plasma cfDNA levels has been reported among patients. Some of this variability may be explained by biological differences between patients, whereas some relates to different sensitivities of the analytic technologies employed by groups and even sources of contaminating DNA. Therefore, it is important to keep in mind that not all samples should be considered equivalent and that there are preanalytic considerations one should make when prospectively collecting samples specifically for cfDNA analyses. Much of our understanding of the biology of cfDNA and optimal methods for its collection and extraction comes from the study of fetal cfDNA for prenatal screening. Owing to the ability to readily distinguish fetal from maternal DNA (particularly with a male fetus), rigorous experiments have identified factors that affect yield and stability of cfDNA as well as sources of contamination (i.e., from maternal cells). For example, in 2001, Chiu and colleagues reported that different methods of isolation of cfDNA from the blood of pregnant women (such as filtration through 0.22-μm filters, centrifugation in Percoll gradients, and high-speed centrifugation) produced significantly different amounts of maternal but not fetal cfDNA (38). Not surprisingly, they have noted the importance of standardization at the level of blood collection, processing, and DNA extraction so that samples within individual studies remain comparable.

At about the same time, ruptured blood cells were identified as a main source of cfDNA contamination (i.e., ref. 39), which has largely shaped efforts on optimization of the cfDNA isolation protocols. Although there is no clear consensus on the best practices for sample handling, we refer the reader to a review covering preanalytic variables that affect cfDNA quality (40). One key observation reported therein is that while overall, serum may yield higher levels of cfDNA than plasma, the yield is more variable and the cfDNA quality may be severely impacted due to lysis of monocytes. Plasma is theoretically less likely to be contaminated with DNA from blood cells but, importantly, the time elapsed between blood collection and centrifugation can heavily influence this (41). Unfortunately, because the overall yield of DNA extracted from plasma or the amount of DNA

measured by qPCR is often used as a proxy for cfDNA, there is also conflicting evidence on the extent of contamination from the blood (42). As revisited later, contamination may be best assessed by determining the relative abundance of high molecular weight DNA fragments that are not consistent with the apoptotic signature characteristic of cfDNA. Methods to overcome such contamination or even accurately assess it remain under development. To better circumvent this issue, at the British Columbia Cancer Agency and Vancouver Prostate Centre, we have opted to collect blood in EDTA tubes and separate plasma by double centrifugation at 1,600 rpm within two hours after collection. It is important to note that heparin tubes are not typically compatible with ctDNA detection methods because in our experience the effect of heparin on polymerase activity can severely impact sensitivity. There are some options for analyzing cfDNA from heparin-containing tubes or tubes containing other PCR inhibitors including polymerases that are more robust to such inhibitors. However, in our opinion for prospective studies, these contaminants are best avoided (43). Understandably, so-called rapid processing of blood is not feasible in all settings (e.g., blood collection at sites lacking a centrifuge) and may not be cost-effective in smaller centers. Preservatives such as formaldehyde have also been proposed as a means to prevent cell lysis thereby obviating the necessity for rapid processing. However, owing to its' potential to damage DNA, other preservatives may be preferable, although we are not aware of any data showing a higher level of noise in cfDNA exposed to formaldehyde-based preservatives. An alternative are the cell-free DNA BCT tubes from Streck. These are advertised to prevent lysis at ambient temperature for up to two weeks but we and others have noted that processing within a week or less is more appropriate to minimize contamination (44). In summary, procedures for collecting plasma for cfDNA analysis should be standardized within centers (and studies) and, depending on the available infrastructure, one may opt to use rapid processing or cell-stabilizing tubes to minimize the risk of contamination from cellular fractions.

Biological Role of cfDNA

Despite almost 70 years of history, the biological function of cfDNA has only been unambiguously established for immune response and blood coagulation (28) and its function in other conditions, if any, remains nebulous. Neutrophils in human blood release so called neutrophil extracellular traps (NET) as one of responses to bacterial infection. cfDNA is an important component of these NETs, that allows them to bind and trap microbial pathogens. The release of DNA from neutrophils is thought to occur via an alternative mechanism of cell death termed NETosis (see refs. 45 and 46 for details). Briefly, two mechanisms are being considered, with the first involving dissolution of the plasma and nuclear membranes that is followed by release of the chromatin into extracellular space. Unlike apoptosis, NETosis does not result in display of the phagocyte-activating signals, so the neutrophils that undergo NETosis do not get cleared from the blood stream by phagocytes. An alternative theory postulates the existence of a DNA/serine nuclease extrusion mechanism from intact neutrophils and that autophagy contributes to NETosis. In any case, NETs appear to require intact chromatin lattices. Importantly, cfDNA in NETs triggers blood coagulation, a process that clearly needs to be tightly controlled. This control is affected by the DNaseI in the blood stream. It has

also been suggested, that another important function of the DNaseI-based mechanism of clearing cfDNA from human blood is the prevention of autoimmunity against DNA (47). Involvement of cfDNA was suggested for other biological processes as well, ranging from tumor dissemination (2, 48–50) to aging (51), but experimental evidence supporting such hypotheses remains tenuous, and considerable additional research is needed to clarify the involvement of cfDNA in processes other than immunologic response.

Genomic Analysis of cfDNA

The presence of tumor-derived DNA in cfDNA implies the entire spectrum of tumor genome aberrations are present and can thus be detected. However, the low amount, high degradation, and high admixture of normal DNA in cfDNA pose major challenges for the development of sensitive and robust detection pipelines. The fact that most (if not all) tumors are characterized by multiple subclonal populations with only a subset of somatic mutations shared among all cells (for review, see ref. 52) further complicates the issue. Broadly, current approaches for detection of tumor aberrations in cfDNA can be divided into two categories: methods targeting specific changes and methods allowing detection of all possible aberrations in DNA (including targeted and whole exome/genome sequencing). The latter next-generation sequencing (NGS)-based options offer numerous potential benefits for observing clonal differences in tumor cell populations, an advantage that until very recently was offset by more limited sensitivity and specificity.

Assessing specific DNA changes

A priori knowledge of specific DNA aberrations (mainly mutations and short insertions/deletions) and recurrent "hot spot" mutations allow implementation of a variety of sophisticated PCR-based methods for their detection in the cfDNA of cancer patients. These approaches historically had low DNA requirements and low noise levels, and proved quite efficient for cancer types where few genomic changes are important for the patient stratification.

Historically, one of the first applications for PCR-based analysis of ctDNA was the detection of high-level amplification of oncogenes. One of the earliest examples was the detection of increased levels of *MYCN* sequences in circulation in patients with neuroblastoma. *MYCN* amplification is a major prognostic factor in localized neuroblastoma. In 2002, Combaret and colleagues demonstrated that PCR and qPCR detect the presence of *MYCN* in blood of patients with *MYCN* amplification, but not in the blood of patients without amplification or healthy controls (11). However, this approach was not very efficient for patients with stage I and II disease (53), probably due to low cfDNA content and insufficient sensitivity of the employed methodology, as *MYCN* locus is present in a diploid state in all healthy cells and thus should be present in all cfDNA samples.

However, in general, detection of copy number aberrations (CNA) in cfDNA using aCGH or approaches that target single/limited loci targeting such as digital PCR has limitations that make it less appealing for ctDNA quantification. After all, unlike other types of cancer-specific mutations (i.e., point mutations or breakpoint detection), the assay must be capable of reliably distinguishing a small change in DNA copy number in a high background of diploid genomes. One of the first approaches to solve this

problem was measuring the number of copies of a locus of interest relative to one or more reference loci assumed to be diploid in all cells. For example, a digital PCR assay to detect ERBB2 amplifications compared the number of copies of this locus to another gene on the same chromosome and applied a threshold that could distinguish plasma of ERBB2 -amplified patients to unamplified patients at sensitivity of 64% and specificity of 94% (54). In general, such approaches are expected to be confounded by variability in copy number across patients. The sensitivity of CNA detection is also expected to suffer for patients with lower ctDNA abundance levels. Sequencing-based methods that broadly survey the genome have shown to be useful for cfDNA copy number profiling (i.e., ref. 55; discussed more below) when employed over large genomic regions, as is the case in prenatal screening for chromosome trisomies. Copy number status evaluation for smaller targeted panels remains challenging, especially for approaches that do not include the routine use of matched normal DNA.

Still, PCR-based techniques are widely used for the detection of specific mutations, and can broadly be classified as either qualitative or quantitative in nature. Qualitative methods generally provide a yes/no readout for the presence of the target mutation and include amplification refractory mutation system (ARMS) PCR (56, 57); PNA clamping PCR (58) and ligation-based methods that use DNA ligase and wild-type and mutation-specific reporter probes to quantify mutant DNA (59). The major disadvantage of these approaches is, obviously, a lack of precise quantification of mutant DNA molecules. Digital methods include various implementations of digital PCR (dPCR) approaches (60, 61), usually in the form of emulsion PCR (62), including BEAMing [for Beads, Emulsions, Amplification, and Magnetics) technology (33, 63)]. This family of approaches incorporate a number of techniques used to improve specificity and sensitivity of mutation detection by PCR through separation of template molecules into individual reaction vessels with modern methods typically using either microfluidics (64), by separating sample and PCR reagents into droplets in an oil emulsion (65, 66) or by combining microfluidics and emulsion PCR to generate evenly sized droplets (67).

BEAMing is an interesting example, because it incorporates a number of techniques used to improve specificity and sensitivity of mutation detection by PCR. Briefly, this approach consists of the following steps: first, the cfDNA mixture is PCR amplified using primers that introduce sequence tags into the resulting amplicons. Then amplicons are combined with streptavidin-coated magnetic beads, coated with nested primers, and emulsified so that each drop in emulsion contains on average one bead and one DNA fragment and emulsion PCR is performed that results in clonal amplification of each template on the surface of the beads. The PCR emulsion is deemulsified and DNA-covered beads are magnetically purified and the DNA on beads is hybridized with oligos complementary to a sequence adjacent to the nucleotide of interest. Next, a single base extension is performed using fluorescently labeled bases that allow differential labeling of the wild-type and mutant alleles. Finally, fluorescently labeled beads are counted/purified on a flow cytometer (and optionally validated by Sanger sequencing).

BEAMing was the first method to allow quantitative sensitive interrogation of mutant cfDNA. In this study, the authors noted that the sensitivity of the detection of rare mutant fragments was mainly limited by two factors: the number of genome equivalents entering the assay (in other words DNA fragments spanning a given mutation and by the fidelity of the DNA polymerases employed in the two PCR steps). These inputs ranged from 1,350 to 230,000 per mL of blood in cancer patients and 1,150 to 8,280 fragments per mL of blood for control subjects, very close to our estimates above. However, most importantly, the errors introduced in the first PCR rounds cannot be eliminated because they would result in beads with homogeneous nonreference fragments, indistinguishable from the bona fide homozygous mutations (33). Interestingly, for this particular study the polymerase error rate was relatively less important than the limitations imposed by the low available amount of input cfDNA because of the possibility to use high-fidelity proofreading DNA polymerases and scoring of only specific base changes. The authors empirically determined that for the assessed targets and used polymerases the error rate after 30 PCR cycles was approximately 2×10^{-5}. The necessity of detecting multiple nonreference reads (three in this case) for identification of a mutation would limit sensitivity to the detection of nonreference base in $>1/1,333$ molecules or approximately 7.5×10^{-4} if 4,000 total cfDNA fragments were assayed (33). To conclude, since its publication, BEAMing has become commercialized and remains a research staple of some groups. For example, a recent application to plasma from patients with colon cancer allowed detection of circulating KRAS mutations, which are known to be acquired in response to EGFR blockade (68). It has also been extended for detection of methylated fragments (69).

In conclusion, even an early dPCR assay for quantifying mutant KRAS DNA suggested a sensitivity as high as one molecule in a background of 200,000 (70). This study also introduced the concept of combining multiple assays for different mutant alleles in a single experiment using different dye dilutions for individual alleles. Commercial kits that facilitate the multiplex screening of hot spots in KRAS now exist. Clearly, methods targeting specific sets of changes, including so-called "actionable" alterations, are of particular interest in clinical settings. By observing the presence of such mutations in a ctDNA sample, one can consider this as a "liquid biopsy" that could inform the clinician on a suitable treatment course. Currently, there are few drugs whose indication is associated with the presence or absence of specific genomic changes (Rubio-Perz and colleagues have listed 57 FDA-approved agents targeting 51 driver genes as of 2015 (71)). Therefore, a relatively small panel can test a patients' cfDNA for a large proportion of clinically relevant mutations with high specificity and sensitivity, which would be highly dependent on the level of ctDNA in the patient. However, as the number of targeted genes grows (Rubio-Perez and colleagues have counted a total of 96 targetable cancer driver genes if all current clinical trials are included; ref. 71) such assays become ever more unwieldy. This problem is exacerbated by the fact that many relevant mutations in cancer are not sufficiently recurrent to facilitate broad coverage using dPCR-based assays. Furthermore, copy number information can also be of clinical importance (see ref. 72 for review). As discussed above, the copy number status of a gene can theoretically be established by dPCR-based approaches with great precision in high-ctDNA scenarios (62); however, again at the expense of increasing number of assessed targets and a potentially high false negative rate when ctDNA is low. Finally, for some cancers, the absence of high frequency driver mutations such as prostate cancer, where the SPOP, mutated in 13% of patients is the most frequently

mutated gene (73), makes the development of such targeted panels even less practical.

Analyzing cfDNA using Massively Parallel Sequencing

The inherent limitations of targeted methods described above for determining more comprehensive mutational landscape of tumors prompted the development on more generalizable techniques. The necessity for development of such methods is further emphasized by the existence of intrapatient tumor heterogeneity—a well-documented phenomenon with relevance to treatment resistance and relapse (e.g., ref. 74). Improvements in read-length, sequence quality, and throughput allowed NGS methods to become a viable alternative for quantifying ctDNA. Limitations that remain in using NGS are the efficiency by which regions of interest can be captured/enriched from cfDNA and the higher error rate of sequencing relative to the accuracy of dPCR. A variety of strategies to drastically reduce the error rate of NGS for accurate ctDNA assessment have already been developed as will be discussed below. In general, current methods for using NGS in ctDNA quantification can be broadly divided into two groups. The first group relies on the amplification of the target regions using region-specific primers (often highly multiplexed), while the second relies upon hybridization-based capture of target regions using complementary oligonucleotides with subsequent amplification of the captured DNA (library). Both strategies are followed by highly redundant ("deep") sequencing to allow the relative amount of mutant and wild-type DNA molecules at each locus to be accurately counted.

PCR-based methods for cfDNA sequencing

A simple approach to amplify cfDNA for sequence-based characterization involves a PCR using site-specific primers with universal tails that facilitate library construction using nested PCR. An early example of a PCR-based strategy for analysis of mutations found in individual patients involved the design of a set of tailed site-specific primers followed by multiplex PCR (for preamplification) and subsequent uniplex PCR using each individual primer pair. The second PCR, in which locus-specific primers are applied individually, is accomplished using a Fluidigm AccessArray system and the entire procedure was named TAm-Seq (Tagged Amplicon Sequencing; ref. 75). This strategy can facilitate sequencing a panel of commonly mutated exons or may be guided by mutations identified through other means (e.g., genome or exome sequencing). This approach afforded the opportunity to quantify mutant DNA at many loci in each patient but had a relatively low sensitivity for mutations below approximately 1%–2%. Still, this pilot study showed clear evidence of a temporal correspondence between ctDNA levels and tumor burden when compared within individual patients. The multiplex nature of the assay also allows monitoring the level of many individual mutations in a single sample (or series) thereby allowing profiling cfDNA at each locus individually. Such distinct locus-specific profiles might be expected in patients with spatial heterogeneity of tumors or with ongoing clonal evolution in response to therapy. Anecdotal examples of this were first shown in a study by Dawson and colleagues, in which the level of mutant DNA corresponding to TP53 and PIK3CA showed strikingly different dynamics across series of plasma samples from patients with metastatic breast cancer (34).

Until recently PCR-based methods were preferred for cfDNA analysis because they allowed sequencing of much smaller input DNA amounts. For example, in our own study of mCRPC patients treated with abiraterone and enzalutamide (31), we chose to combine targeted sequencing with whole-genome copy number profiling using array comparative genomic hybridization. Whole-genome copy number profiling allowed us to determine that pre-existing amplification of AR in this cohort was a marker of adverse outcome for patients switched onto enzalutamide. We successfully sequenced exon 8 that encodes part of the ligand-binding domain (LBD) of the AR gene from as little as 1 ng of input cfDNA. Out of six detected nonsynonymous LBD mutations, three were not previously observed in prostate cancer. Importantly, we also identified cases where a patient had multiple (up to five) mutations in the AR, while no DNA read, spanning the sequenced region, had more than two mutations. The most parsimonious explanation for that phenomenon is the existence of multiple tumor subclones, each with unique version of AR protein.

We have also observed changes in patients' AR LBD mutation landscape during the course of treatment. To understand the functional significance of these changes and to enable rational design of novel antiandrogens, we have established a resource for the functional characterization of all identified AR mutants. Therefore we characterized the effects of various steroids (DHT, estradiol, progesterone, and hydrocortisone) and different antiandrogens including enzalutamide and a novel agent developed at the Vancouver Prostate Centre (VPC) on transcriptional activity of the receptor. We established, in vitro, that all mutations detected in the plasma samples of mCRPC patients were resistant to at least one specific antiandrogen treatment and allowed us to explain some of the observed treatment-induced AR mutation landscape shifts. We also demonstrated that a novel AR inhibitor VPC-13566 under development at the VPC was able to efficiently target all tested AR mutants (76). To summarize, we have prototyped an analytic pipeline for evidence-based selection of optimal treatment strategies for mCRPC patients that may eventually enable rational and rapid selection of specific AR inhibitors to combat resistance.

In a broader application of this type of approach, Carreira and colleagues reported on the temporal sequential analysis of approximately 38-kb region using a custom Ampliseq panel (Thermo Fisher Scientific) from as little as 6 ng of cfDNA in a cohort of 16 TMPRSS2-ERG–positive prostate cancer patients (77). This work demonstrates the advantages of targeted sequencing approaches as a single assay enabled the detection of both CNAs and point mutations. Importantly, the authors also developed an approach for assessment of the abundance of ctDNA in total cfDNA, based either on the analysis of the allelic frequencies in monoallelic deletion regions, or in absence of those, on the comparison of read depths of the autosomal and nonautosomal regions in tumor and matched normal samples. This work also illustrated the main disadvantage of an amplicon-based targeted sequencing—varying PCR efficiency of the panel primer pairs. For example, for the copy number estimates the authors chose to retain only amplicons with read coverage falling in the range of mean $\pm$ SD for more, than 10 samples. This resulted in discarding 35% (120/337) of autosomal amplicons (ref. 77; Supplementary Data). It also should be noted, that 6 ng of input DNA is arguably the lower practical limit of cfDNA input for sequencing as it corresponds to approximately 1,000 diploid genome equivalents. Assuming 1% ctDNA content, this amount would contain just 10

Volik et al.

copies of tumor genomes—a level, likely to result in high sampling variance. Another key consideration when designing PCR primers for such experiments is that the distance between primers should correspond to the size of DNA fragments expected in ctDNA. In our experience, we have observed a bias towards PCR under-representing the level of ctDNA proportionally relating to the size of amplicons in accordance with observations made in the Diehl study (33). While keeping amplicon size to a minimum must be a priority (e.g., 60–80 bp) it should be also noted that this can be challenging for assays relying on hydrolysis probes, as little space is available in such a small region for two primers and a probe. Even in spite of such accommodations, any PCR-based method is expected to under-sample ctDNA owing to fragments in which both priming sites are not represented and this may be exacerbated in samples with contaminating DNA that can be less fragmented.

Ligation and hybridization-capture methods

A recently developed method for performing targeted sequencing is based on oligonucleotide DNA capture and has given rise to affordable approaches to sequence gene panels and even the entire human exome (78). This approach is based upon sequencing library construction followed by hybridization of the library to a pool of DNA or RNA oligonucleotides complementary to regions of interest. The hybrid molecules are then isolated (typically via immobilization on streptavidin beads) and amplified using universal primer pairs complementary to the library adaptors. Importantly, a convenient feature of such strategies that owes to the natural size distribution of ctDNA is that libraries can be generated using ligation-based chemistry directly without the need for shearing or transposon-based library construction. It is notable that in absence of shearing step much of the contaminating DNA from nonapoptotic processes is expected to be naturally excluded from libraries thereby naturally biasing the library contents in favor of true ctDNA fragments. Assuming sufficiently high ligation efficiency, this would result in a higher fraction of mutant ctDNA molecules making it through sequencing pipeline than PCR-based applications. This assumption has been experimentally validated in libraries prepared from cfDNA of hepatocellular carcinoma patients (79). The authors observed a shift towards the characteristic cfDNA fragment length in read pairs mapping in regions associated with copy number gains, where we would expect greater proportion of ctDNA-derived fragments. Despite this potential benefit, this does not preclude the need for standardization of blood collection and handling methods. Until recently, the major disadvantages of this type of approach to cfDNA analysis were relatively high requirements for the quality and quantity of input DNA. However, development of the solution-based hybridization workflows (80) and refinements in sequencing library construction protocols such as improvements of ligation efficiency allowed drastic reduction of the quality and quantity of input DNA requirements, resulting in the development of the capability to analyze DNA samples as scarce and highly fragmented as cfDNA.

One of the best examples of successful application of this strategy to development of a clinically relevant cfDNA-based analysis is the work of Newman and colleagues (81). In this seminal study, the authors first analyzed available sequence data to determine a set of genomic regions (a "selector") comprising a set of mutations present in majority of patients with stage II–IV

non–small cell lung cancer (NSCLC). Custom oligonucleotides covering these loci were purchased to allow hybridization capture of all DNA from these regions. They then combined a modified library construction strategy with sophisticated bioinformatics methods to sequence these loci in patient constitutional DNA, tumor, and cfDNA samples. Importantly, mutations identified using this method may include indels, single base substitutions, and breakpoints that underlie structural alterations such as those affecting *ALK* or large deletions. The latter types afford virtually perfect specificity for tumor DNA. Using the level of mutant DNA detected across the mutations found in the tumor, they demonstrated that cfDNA-based analysis allowed for earlier assessment of response to treatment than standard-of-care radiographic approaches and could distinguish between residual disease and treatment-related imaging changes, such as postradiotherapy inflammation. This so-called "CAPP-seq" method allowed analysis of as little as 7 ng of input cfDNA (∼1,100 genome equivalents), essentially at the level of the best existing PCR-based target sequencing methods. This was a significant result, because it demonstrates the potential for hybridization-based workflows to compete with PCR-based strategies while allowing for a much broader and more uniform representation of the analyzed DNA than PCR-based techniques. We note that very inexpensive options now exist for obtaining individual capture oligonucleotides or pools that target the exons of a small number of genes such as biotinylated DNA LockDown oligonucleotides offered by Integrated DNA Technologies. This approach allows for flexibility in designing and modifying such selectors for individual patients or cohorts. For some tumor types, a smaller selector focusing on a small panel of exons may be suitable, whereas other may be better covered using a larger selector that includes genomic regions commonly affected by structural alterations. An important consideration when designing a selector is that sequencing cost is proportional to the size of the region being sequenced. It should be noted, that targeting very small regions is associated with its own hurdles because the enrichment efficiency of a single capture are typically on the order of 10^4-fold, so for the targets smaller, than approximately 100 kb additional steps, such as two rounds of capture may be necessary to ensure sufficiently high on-target mapping rate (82). For patient-specific applications or cancers with a high mutation rate (or recurrence in a small number of genes), this may allow deep sequencing to be achieved at a low cost using "bench-top" NGS devices.

A logical extension of this approach that requires no prior knowledge of the tumor but assumes the presence of somatic point mutations affecting exons is the use of whole-exome sequencing. Early exploration of this strategy was demonstrated in six patients with a mixture of advanced breast, ovarian, and lung cancers by Murtaza and colleagues (83). Overall, a strong correspondence between the mutations was detected between the plasma and the matched tumor and the variant allelic frequency (VAF) was largely reflective of the level of ctDNA in each plasma sample. In almost all patients, at least one mutation was identified that showed an increase in VAF over the time of the study, and these included mutations in genes thought to be associated with treatment resistance in their respective disease such as *PIK3CA* in a breast cancer patient treated with paclitaxel. In another example, Butler and colleagues (84) compared tumor and plasma exomes from two patients with metastatic sarcoma and metastatic breast tumor. For sarcoma patients, 47 of 48 mutations identified in the tumor sample were also found by exome sequencing of the

cfDNA. However, for the patient with metastatic breast tumor the authors observed discordance for the H1047R *PIK3CA* mutation status. This mutation was detected in primary tumor but not in matched metastatic and cfDNA samples. *ESR1* mutation (D538G), on the contrary, was observed in both metastatic and cfDNA samples but not in tumor, and could possibly explain patient's resistance to estrogen deprivation therapy. The potential for observing discordance between tumor tissue and liquid biopsies via exome sequencing offers many avenues of research for studying the clonal complexity and patterns of evolution in cancer not previously readily accessible due to the invasive nature of tissue biopsies.

Importantly, despite some potential to observe tumor evolution and identify variants outside of gene panels or selectors in more targeted capture-based assays, exome sequencing remains a niche application. Large target size makes achieving the same level of coverage as for targeted gene sets impractical. Currently, that and the higher input DNA amount requirements (∼100 ng for Butler and colleagues study; ref. 84), limits exome sequencing to the analysis of samples with relatively high ctDNA levels that would allow robust detection of mutations at the relatively shallow coverage of individual loci. In such scenarios, it is advisable to assess ctDNA first using a targeted approach to identify suitable samples such as those with VAFs of at least 5% at known mutant sites and high amounts of cfDNA in blood.

Error suppression methods

As we have demonstrated earlier, the mutant VAF in patients' cfDNA can go as low as 0.01%. Clearly, any strategy to suppress sources of errors thereby increasing accuracy in detecting mutant DNA molecules is important for cfDNA analysis. This problem is equivalent to an extreme case of identifying low-abundance mutations in tissue samples, which has also proven difficult for NGS methods. Guntry and Vijg noted that "the single most important limitation of current MPS approaches from mutation analysis is the inability to address low-abundance mutations that turn somatic tissues into mosaics of cells" (85). This argument was based on the comparison of the natural mutation rates in mammalian cells (e.g., 0.05×10^{-9} per base per cell division for human cells) versus well-established error rates of existing sequencing platforms, which was estimated in 2012 to be approximately 0.05%–1% and has not significantly changed since.

Besides simple strategies such as using proofreading-enabled polymerase and a low number of PCR cycles, the most promising strategy for error suppression is based on the fact that because DNA is naturally double-stranded, true mutations should be present in identical positions on both DNA strands from a single duplex. However, because all widely employed sequencing strategies now include a PCR amplification step, the real problem becomes distinguishing reads that originate from the same original duplex fragment (and thus should be identical) from reads that originate from multiple DNA fragments covering the same locus (that were produced from the other allele(s) or other cells in the analyzed mixture). One of the first solutions to this problem was proposed in 2011 by Kinde and colleagues. Their "Safe-Seqs" approach involves usage of pools of primers containing degenerate molecular tags that would uniquely barcode all DNA fragments resulting from the initial PCR cycle. The procedure is followed by a nested PCR to amplify each "family" of fragments that are later collapsed into consensus sequences. Interestingly, the authors provide hard data to assess the efficiency of their

approach. In one of the experiments, the authors used DNA from 1,750 individual cells to assess mutation frequency in one gene. Using endogenous UID approach, 1,057 molecules were assessed or approximately 30% of the total available amount (86). Comparable strategies for ligation-based library preparation with modified adaptors allow individual ligation events to be distinguished and even facilitate recognition of the two strands of the original duplex. Conventional methods to recognize reads deriving from the same template DNA molecule (i.e., duplicate pairs) are not suitable for ctDNA because the distribution of fragment ends is non-random and is likely dictated, in part, by nucleosome occupancy (87). As the read lengths and the robustness of paired-end sequencing of the massively parallel sequencing platforms improved, it became feasible to sequence approximately 200-bp fragments from both ends. Coupled with the molecular tagging, this opened the way to drastically decrease the sequencing pipeline error rate. For example, the authors of one of such approach, Schmitt and colleagues estimated the error rate of 3.8×10^{-10} in a model experiment (88). The same group demonstrated the robustness of their approach by detecting a single *ABL1* imatinib conferring mutation in a sample from the chronic myeloid leukemia patient (82). The E279K mutation was unambiguously detected at 1% rate, whereas the error rate of raw sequencing data would have either completely or partially obscured this mutation. In our own data, (Assouline and colleagues, in review) we have directly inferred mutations from ctDNA libraries using a similar error suppression strategy at levels as low as 1% across a broad gene panel. An interesting combination approach termed "integrated digital error suppression" (iDES) was described in 2016 by Newman and colleagues (89). The authors redesigned their previously published Capp-Seq panel (81) to integrate molecular barcoding into their workflow. However, analysis of the cfDNA sequencing results revealed the presence of particular sequencing artifacts in their cfDNA samples that were not suppressed by bar coding. This prompted the authors to rigorously examine the patterns of sequencing artifacts and note that G→T changes were much more prevalent than C→A changes. Moreover, this imbalance increased proportionally to an increase of the hybridization time. This allowed the authors to develop an *in silico* filter to remove this bias. The combination of *in silico* filtering and barcoding approaches allowed development of the workflow that had 15 times lower error rate than original CAPP-Seq approach.

An interesting alternative to barcoding techniques capitalized on the increased read length possible with newer Illumina chemistry and suggested the use of rolling circle amplification to ensure redundant sequencing of individual DNA fragments (90). Briefly, DNA is fragmented to approximately 130 bp, denatured, circularized, and amplified using Phi29 polymerase. Strand displacement activity of this polymerase ensures rolling circle amplification of individual DNA fragments. These can be sequenced on Illumina MiSeq machines using 500-cycle chemistry. As a result, each individual fragment of the original DNA mixture will be read on average three times. Interestingly, the authors demonstrated that redundant sequencing alone results only in approximately 2-fold reduction of error rate. More detailed analysis of the sequencing data revealed that most of the noise resulted from deaminated nucleotides (C→T and G→A transitions due to deamination of cytosine into uracyl and of guanine into xanthine, respectively). Treating the circularized DNA fragments with uracil-DNA glycosylase and (UDG) and formamidopyrimidine-DNA-glycosylase (Fpg) resulted in excision of deaminated bases, and,

Volik et al.

thus, in linearization of the fragments with such bases. Clearly, linearized fragments are excluded from subsequent rolling circle amplification and sequencing. This allowed the authors to achieve approximately 100-fold reduction in sequencing noise thereby improving specificity, but presumably at a cost to overall sensitivity due to the loss of some molecules prior to sequencing (90).

To summarize, all of the aforementioned approaches rely on redundant sequencing of individual DNA fragments to achieve drastic noise reduction and thus come at a higher cost than standard applications of DNA sequencing. Circle sequencing stands apart from other approaches due to lowest sequencing redundancy, but requires UDG/Fpg enzyme treatment step to ensure significant noise reduction. Another common problem for the barcoding approaches is their sensitivity to efficiency of library construction protocol, starting at ligation of barcoded adaptors and ending with final PCR amplification of selected DNA fragments. This number can be measured by calculating the ratio of the detected individual molecules after sequencing to the total approximate number of DNA molecules that enter the assay. Over last five years, this metric improved from around 30% in 2011 (86) to more than 60% in 2016 (89).

Commercialization efforts

The practical advantages of tumor monitoring from blood have spurred a number of commercialization efforts. To date, one of the most successful enterprises in this space is Guardant Health Inc. Its proprietary approach appears to be a variant of the barcoding method termed "digital sequencing" (91) based on the "nonunique" heptamer barcodes that tag both 5′ and 3′ ends of cfDNA fragments. The company claims that this process is 5–10 times more efficient than other existing workflows and that minor alleles at approximately 0.1% prevalence can be detected with extremely high specificity. The analysis workflow also allows for identification of the copy number gains, as long as the CN gain exceeds 2.2-fold in the cfDNA. Guardant Health offers a 54-gene panel for cfDNA-based screening of melanoma, lung, and breast cancer patients that allows both SNV detection and gene copy number evaluation. Other important players in this field including Personal Genome Diagnostics Inc. led by John Hopkins team that includes Drs. L. Diaz and V. Velculescu and Cambridge, MA based Foundation Medicine. The latter recently announced its own version of a barcoding assay that is also aimed at detecting low-level contamination in cfDNA samples.

One potential application of cfDNA that has been long envisioned is its use in identifying cancers prior to any symptoms, but owing to the difficulty and likely cost of tackling this problem, there are no studies yet demonstrating the feasibility of finding a solution. In reference to this "holy grail" application, in January 2016, Illumina Inc announced launch of GRAIL, a new company dedicated to developing and implementing cfDNA-based assay for the early detection of cancer in asymptomatic individuals. No details on the underlying technology have been released, beyond the statement that ultra-deep sequencing (20,000× or more) will be one of the cornerstones of the approach. A recent presentation for investors released by Illumina outlines two scenarios: one "best case" which assumes that GRAIL will develop a test suitable for high-risk individuals in whom cancers at higher stages (stage II and above) would be sought. Ultimately, they also plan to target patients in the general population using an approach suitable for early-stage cancer. In both cases, the analysis cost is expected to fall

in $500—1,000 range. While the investor's presentation mentions "error-corrected reads", no details on the GRAIL's approach to error correction are available yet. To summarize, the liquid biopsy clearly represents an appealing area for commercialization. Thus far, the commercial success appears to be favoring companies that employ some form of error correction in their technologies, a trend we expect to continue.

Conclusion

The concept of precision oncology has as its foundation the ability to detect clinically relevant and actionable tumor-specific changes in a timely fashion. This may be achieved using temporal cfDNA assays to monitor adaptation to therapy and identify actionable mutations. cfDNA-based profiling of cancer patients offers a number of critical advantages for essentially real-time monitoring of a tumor response to therapy in cancer patients. These include integral representations of tumor heterogeneity, ease of sampling, minimal invasiveness and morbidity, and low cost. However, tumor-derived DNA usually constitutes only a small percentage of total cfDNA so the ability to detect rare genome aberrations is an essential requirement for cfDNA analysis pipelines. Another important parameter is the spectrum of genomic changes the technology is capable of detecting. While targeted assays can be fruitful in the clinical setting, sequence-based approaches offer clear advantages in terms of flexibility of coverage and the ability to detect a wide range of aberrations in tumor genomes. This flexibility will be especially important for managing metastasis and resistance to therapy; widely recognized to be among the most important problems in cancer management. Resistance to therapy can be driven by a wide range of genomic aberrations such as point mutations and copy number aberrations. Moreover, resistant subclones can constitute a very small proportion of the tumors total clonal population until the selective pressure of therapy leads to their rapid expansion. Clearly, the early detection of resistant clones requires sensitivity to detect such events. However, this requires minimizing noise in the ctDNA analyses and pushing the sensitivity of detection to the theoretical limits imposed by the plasma levels of ctDNA. Recent evidence suggests that the most promising technology for this is based on molecular tagging-based workflows that suppress errors introduced by PCR and sequencing. This approach is limited mainly by the ctDNA sampling efficiency and is straightforward to scale and thus offer enormous potential for monitoring ctDNA in cancer patients and possibly for screening healthy asymptomatic individuals.

References

1. Mandel P, Metais P. [Les acides nucléiques du plasma sanguin chez l'homme]. C R Seances Soc Biol Fil 1948;142:241–3.
2. Bendich A, Wilczok T, Borenfreund E. Circulating DNA as a possible factor in oncogenesis. Science 1965;148:374–6.
3. Tan EM, Schur PH, Carr RI, Kunkel HG. Deoxybonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. J Clin Invest 1966;45:1732–40.
4. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977;37:646–50.
5. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology 1989;46:318–22.
6. Vasioukhin V, Anker P, Maurice P, Lyautey J, Lederrey C, Stroun M. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. Br J Haematol 1994;86:774–9.
7. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev 1994;3:67–71.
8. 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.
9. 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.
10. Chiang P-W, Beer DG, Wei W-L, Orringer MB, Kurnit DM. Detection of erbB-2 amplifications in tumors and sera from esophageal carcinoma patients. Clin Cancer Res 1999;5:1381–6.
11. Combaret V, Audoynaud C, Iacono I, Favrot MC, Schell M, Bergeron C, et al. Circulating MYCN DNA as a tumor-specific marker in neuroblastoma patients. Cancer Res 2002;62:3646–8.
12. Mutirangura A, Pornthanakasem W, Theamboonlers A, Sriuranpong V, Lertsanguansinchi P, Yenrudi S, et al. Epstein-Barr viral DNA in serum of patients with nasopharyngeal carcinoma. Clin Cancer Res 1998;4:665–9.
13. Lo YMD, Chan LYS, Lo K-W, Leung S-F, Zhang J, Chan ATC, et al. Quantitative analysis of cell-free epstein-barr virus DNA in plasma of patients with nasopharyngeal carcinoma. Cancer Res 1999;59:1188–91.
14. Capone RB, Pai SI, Koch WM, Gillison ML, Danish HN, Westra WH, et al. Detection and quantitation of human papillomavirus (HPV) DNA in the sera of patients with HPV-associated head and neck squamous cell carcinoma. Clin Cancer Res 2000;6:4171–5.
15. Pornthanakasem W, Shotelersuk K, Termrungruanglert W, Voravud N, Niruthisard S, Mutirangura A. Human papillomavirus DNA in plasma of patients with cervical cancer. BMC Cancer 2001;1:2.
16. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Res 1999;59:67–70.
17. Wong IHN, Dennis Lo YM, Zhang J, Liew C-T, Ng MHL, Wong N, et al. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. Cancer Res 1999;59:71–3.
18. Silva JM, Dominguez G, Villanueva MJ, Gonzalez R, Garcia JM, Corbacho C, et al. Aberrant DNA methylation of the p16(INK4a) gene in plasma DNA of breast cancer patients. Br J Cancer 1999;80:1262–4.
19. Lo YMD, Wainscoat JS, Gillmer MDG, Patel P, Sampietro M, Fleming KA. Prenatal sex determination by DNA amplification from maternal peripheral blood. Lancet 1989;334:1363–5.
20. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CWG, et al. Presence of fetal DNA in maternal plasma and serum. Lancet 1997;350:485–7.
21. Bianchi DW, Parker RL, Wentworth J, Madankumar R, Saffer C, Das AF, et al. DNA sequencing versus standard prenatal aneuploidy screening. N Engl J Med 2014;370:799–808.
22. Greene MF, Phimister EG. Screening for trisomies in circulating DNA. N Engl J Med 2014;370:874–5.
23. Beck J, Oellerich M, Schulz U, Schauerte V, Reinhard L, Fuchs U, et al. Donor-derived cell-free DNA is a novel universal biomarker for allograft rejection in solid organ transplantation. Transplant Proc 2015;47:2400–3.
24. Butt AN, Swaminathan R. Overview of circulating nucleic acids in plasma/serum. Ann N Y Acad Sci 2008;1137:236–42.
25. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 1999;98:285–94.
26. Wang BG, Huang H-Y, Chen Y-C, Bristow RE, Kassauei K, Cheng C-C, et al. Increased plasma DNA integrity in cancer patients. Cancer Res 2003;63:3966–8.
27. Madhavan D, Wallwiener M, Bents K, Zucknick M, Nees J, Schott S, et al. Plasma DNA integrity as a biomarker for primary and metastatic breast cancer and potential marker for early diagnosis. Breast Cancer Res Treat 2014;146:163–74.
28. Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. Science 2004;303:1532–5.
29. Evan GI, Vousden KH. Proliferation, cell cycle and apoptosis in cancer. Nature 2001;411:342–8.
30. Fleischhacker M, Schmidt B. Circulating nucleic acids (CNAs) and cancer—A survey. Biochim Biophys Acta 2007;1775:181–232.
31. Azad AA, Volik SV, Wyatt AW, Haegert A, Le Bihan S, Bell RH, et al. Androgen receptor gene aberrations in circulating cell-free DNA: biomarkers of therapeutic resistance in castration-resistant prostate cancer. Clin Cancer Res 2015;21:2315–24.
32. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch R-D, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659–65.
33. 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.
34. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013;368:1199–209.
35. Tsumita T, Iwanaga M. Fate of injected deoxyribonucleic acid in mice. Nature 1963;198:1088–9.
36. Lo YMD, Zhang J, Leung TN, Lau TK, Chang AMZ, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 1999;64:218–24.
37. Fatouros IG, Jamurtas AZ, Nikolaidis MG, Destouni A, Michailidis Y, Vrettou C, et al. Time of sampling is crucial for measurement of cell-free plasma DNA following acute aseptic inflammation induced by exercise. Clin Biochem 2010;43:1368–70.
38. Chiu RWK, Poon LLM, Lau TK, Leung TN, Wong EMC, Lo YMD. Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem 2001;47:1607–13.
39. Lui YYN, Chik K-W, Chiu RWK, Ho C-Y, Lam CWK, Lo YMD. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem 2002;48:421–7.
40. Bronkhorst AJ, Aucamp J, Pretorius PJ. Cell-free DNA: preanalytical variables. Clin Chim Acta 2015;450:243–53.
41. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: preanalytical considerations. Clin Chim Acta 2013;424:222–30.
42. Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, et al. Influence of plasma processing on recovery and analysis of circulating nucleic acids. PLoS ONE 2013;8:e77963.
43. Zhang Z, Kermekchiev MB, Barnes WM. Direct DNA amplification from crude clinical samples using a PCR enhancer cocktail and novel mutants of Taq. J Mol Diagn 2010;12:152–61.
44. Wang P, Bahreini A, Gyanchandani R, Lucas PC, Hartmaier RJ, Watters RJ, et al. Sensitive detection of mono- and polyclonal ESR1 mutations in primary tumors, metastatic lesions and cell free DNA of breast cancer patients. Clin Cancer Res 2016;22:1130–7.
45. Kaplan MJ, Radic M. Neutrophil extracellular traps: double-edged swords of innate immunity. J Immunol 2012;189:2689–95.
46. Gould TJ, Lysov Z, Liaw PC. Extracellular DNA and histones: double-edged swords in immunothrombosis. J Thromb Haemost 2015;13:S82–S91.
47. Mai SH, Khan M, Dwivedi DJ, Ross CA, Zhou J, Gould TJ, et al. Delayed but not early treatment with DNase reduces organ damage and improves outcome in a murine model of sepsis. Shock 2015;44:166–72.
48. Garcia-Olmo D, Garcia-Olmo DC, Ontanon J, Martinez E, Vallejo M. Tumor DNA circulating in the plasma might play a role in metastasis. The hypothesis of the genometastasis. Histol Histopathol 1999;14:1159–64.

49. García-Olmo DC, Domínguez C, García-Arranz M, Anker P, Stroun M, García-Verdugo JM, et al. Cell-free nucleic acids circulating in the plasma of colorectal cancer patients induce the oncogenic transformation of susceptible cultured cells. Cancer Res 2010;70:560–7.

50. Trejo-Becerril C, Pérez-Cárdenas E, Taja-Chayeb L, Anker P, Herrera-Goepfert R, Medina-Velázquez LA, et al. Cancer progression mediated by horizontal gene transfer in an in vivo model. PLoS ONE 2012;7:e52754.

51. Mittra I, Nair N, Mishra P. Nucleic acids in circulation: Are they harmful to the host? J Biosci 2012;37:301–12.

52. Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. Nat Med 2016;22:105–13.

53. Combaret V, Hogarty MD, London WB, McGrady P, Iacono I, Brejon S, et al. Influence of neuroblastoma stage on serum-based detection of MYCN amplification. Pediatr Blood Cancer 2009;53:329–31.

54. Gevensleben H, Garcia-Murillas I, Graeser MK, Schiavon G, Osin P, Parton M, et al. Noninvasive detection of HER2 amplification with plasma DNA digital PCR. Clin Cancer Res 2013;19:3276–84.

55. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 2012;4:162ra54–ra54.

56. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotech 1999;17:804–7.

57. Kimura H, Kasahara K, Kawaishi M, Kunitoh H, Tamura T, Holloway B, et al. Detection of epidermal growth factor receptor mutations in serum as a predictor of the response to gefitinib in patients with non–small-cell lung cancer. Clin Cancer Res 2006;12:3915–21.

58. Won JK, Keam B, Koh J, Cho HJ, Jeon YK, Kim TM, et al. Concomitant ALK translocation and EGFR mutation in lung cancer: a comparison of direct sequencing and sensitive assays and the impact on responsiveness to tyrosine kinase inhibitor. Ann Oncol 2015;26:348–54.

59. Wu J, Matthaei H, Maitra A, Molin MD, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med 2011;3:92ra66–92ra66.

60. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci 1999;96:9236–41.

61. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. Biotechniques 1992;13:444–9.

62. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 2011;83:8604–10.

63. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci 2003;100:8817–22.

64. Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. 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.

65. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. Nat Biotech 1998;16:652–6.

66. Wetmur JG, Kumar M, Zhang L, Palomeque C, Wallenstein S, Chen J. Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes. Nucleic Acids Res 2005;33:2615–9.

67. Janku F, Vibat CRT, Kosco K, Holley VR, Cabrilo G, Meric-Bernstam F, et al. BRAF V600E mutations in urine and plasma cell-free DNA from patients with Erdheim-Chester disease. Oncotarget 2014;5:3607–10.

68. 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.

69. Li M, Chen W-D, Papadopoulos N, Goodman SN, Bjerregaard NC, Laurberg S, et al. Sensitive digital quantification of DNA methylation in clinical samples. Nat Biotech 2009;27:858–63.

70. Pekin D, Skhiri Y, Baret J-C, Le Corre D, Mazutis L, Salem CB, et al. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab Chip 2011;11:2156–66.

71. Rubio-Perez C, Tamborero D, Schroeder Michael P, Antolín Albert A, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell 2015;27:382–96.

72. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463:899–905.

73. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat J-P, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685–9.

74. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JM, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. Nature 2015;520:353–7.

75. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012;4:136ra68–ra68.

76. Lallous N, Volik SV, Awrey S, Leblanc E, Tse R, Murillo J, et al. Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. Genome Biol 2016;17:10.

77. Carreira S, Romanel A, Goodall J, Grist E, Ferraldeschi R, Miranda S, et al. Tumor clone dynamics in lethal prostate cancer. Sci Transl Med 2014;6:254ra125–254ra125.

78. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 2009;461:272–6.

79. Jiang P, Chan CW, Chan KC, Cheng SH, Wong J, Wong VW, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci U S A 2015;112:E1317–25.

80. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotech 2009;27:182–9.

81. Newman AM, Bratman SV, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 2014;20:548–54.

82. Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, Radich JP, et al. Sequencing small genomic targets with high efficiency and extreme accuracy. Nat Methods 2015;12:423–5.

83. Murtaza M, Dawson S-J, Tsui DWY, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature 2013;497:108–12.

84. Butler TM, Johnson-Camacho K, Peto M, Wang NJ, Macey TA, Korkola JE, et al. Exome sequencing of cell-free DNA from metastatic cancer patients identifies clinically actionable mutations distinct from primary disease. PLoS One 2015;10:e0136407.

85. Gundry M, Vijg J. Direct mutation analysis by high-throughput sequencing: From germline to low-abundant, somatic variants. Mutat Res 2012;729:1–15.

86. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci 2011;108:9530–5.

87. Snyder Matthew W, Kircher M, Hill Andrew J, Daza Riza M, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its Tissues-Of-Origin. Cell 2016;164:57–68.

88. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012;109:14508–13.

89. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotech 2016;34:547–55.

90. Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, et al. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. Proc Natl Acad Sci 2013;110:19872–7.

91. Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. PLoS ONE 2015;10:e0140712.

AAGR American Association for Cancer Research

Molecular Cancer Research

Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped By Emerging Technologies

Stanislav Volik, Miguel Alcaide, Ryan D. Morin, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/1541-7786.MCR-16-0044

Cited articles	This article cites 91 articles, 32 of which you can access for free at: http://mcr.aacrjournals.org/content/14/10/898.full#ref-list-1
Citing articles	This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://mcr.aacrjournals.org/content/14/10/898.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://mcr.aacrjournals.org/content/14/10/898. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.