Genomic approaches to cancer and minimal residual disease detection using circulating tumor DNA

Nicholas P. Semenkovich
Jeffrey J. Szymanski
Noah Earland
Pradeep S. Chauhan
Bruna Pellini

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/oa_4

Part of the Medicine and Health Sciences Commons

Please let us know how this document benefits you.
Authors
Nicholas P. Semenkovich, Jeffrey J. Szymanski, Noah Earland, Pradeep S. Chauhan, Bruna Pellini, and Aadel A. Chaudhuri
Genomic approaches to cancer and minimal residual disease detection using circulating tumor DNA

Nicholas P Semenkovich, Jeffrey J Szymanski, Noah Earland, Pradeep S Chauhan, Bruna Pellini, Aadel A Chaudhuri

ABSTRACT
Liquid biopsies using cell-free circulating tumor DNA (ctDNA) are being used frequently in both research and clinical settings. ctDNA can be used to identify actionable mutations to personalize systemic therapy, detect post-treatment minimal residual disease (MRD), and predict responses to immunotherapy. ctDNA can also be isolated from a range of different biofluids, with the possibility of detecting locoregional MRD with increased sensitivity if sampling more proximally than blood plasma. However, ctDNA detection remains challenging in early-stage and post-treatment MRD settings where ctDNA levels are minuscule giving a high risk for false negative results, which is balanced with the risk of false positive results from clonal hematopoiesis. To address these challenges, researchers have developed ever-more elegant approaches to lower the limit of detection (LOD) of ctDNA assays toward the part-per-million range and boost assay sensitivity and specificity by reducing sources of low-level technical and biological noise, and by harnessing specific genomic and epigenomic features of ctDNA. In this review, we highlight a range of modern assays for ctDNA analysis, including advancements made to improve the signal-to-noise ratio. We further highlight the challenge of detecting ultra-rare tumor-associated variants, overcoming which will improve the sensitivity of post-treatment MRD detection and open a new frontier of personalized adjuvant treatment decision-making.

INTRODUCTION
After its initial discovery in 1948, plasma cell-free DNA (cfDNA) was noted to have associations with malignancy in 1977, and first entered clinical use through non-invasive prenatal testing in 2011 where it is now widely used in prenatal counseling to detect trisomies and other genetic syndromes. Within oncology, the first plasma cfDNA test approved by the United States Food and Drug Administration (FDA), the Roche cobas EGFR Mutation Test v2, was FDA-approved in 2016 to identify 42 mutations in the EGFR gene in patients with metastatic non-small cell lung cancer (NSCLC). Since then, numerous plasma cfDNA tests have entered clinical practice, focused on actionable mutations in solid tumors, including those of the colon, breast, prostate, ovary, and lung. More recently, liquid biopsy tests have emerged for the detection of minimal residual disease (MRD) after curative-intent treatment, as well as for early cancer detection, including in colorectal, breast, lung, and bladder cancers. In this review, we discuss the utility of cfDNA in identifying tumor-derived genomic alterations and describe the range of sequencing technologies for circulating tumor DNA (ctDNA) detection and key aspects of its analysis. We also highlight the role of ctDNA in selecting targeted therapies, detecting disease relapse and MRD, monitoring treatment response, and its emerging role in immuno-oncology.

SOURCES AND SCARCITY
In current clinical and research practice, peripheral blood plasma is the most common source for ctDNA, and collected volume and storage conditions can impact the sensitivity of ctDNA assays. Innovations in sample collection and storage have enabled plasma samples to be preserved at room temperature for up to 14 days without significant cfDNA degradation, though more rapid processing is needed if collecting blood in standard K₂EDTA tubes. Most commercial cfDNA assays target the collection of 8–20 mL of whole blood, which yields approximately 4–10 mL plasma.

Although assays may be performed on smaller plasma volumes, these reduced amounts can impact ctDNA detection sensitivity, especially in low disease burden settings such as post-treatment MRD and early cancer detection. As sequencing costs continue to fall, adjuvant systemic therapy options expand, and ctDNA technologies advance, there will be increased demand to detect MRD after curative-intent treatment.
with high clinical sensitivity to precisely inform adjuvant therapy decision-making.27 33 34

Limits of detection

The most fundamental challenge in the analysis of ctDNA is its scarcity—the majority of cfDNA (>90%–99.9%) within peripheral blood is derived from healthy host sources, predominantly PBMCs, though also from other healthy tissues including the endothelium.35–38 The concentration of ctDNA is highly variable and differs based on malignancy type and tumor burden, among other factors.39 Broadly speaking, ctDNA may comprise up to 10% of peripheral cfDNA in patients with advanced-stage cancer, 1% in locally advanced disease, and 0.1% of total cfDNA in early-stage disease or after curative-intent treatment.10 40

For patients with advanced-stage cancer, the higher levels of ctDNA have facilitated diagnostics for the detection of clinically actionable mutations at the time of diagnosis and at treatment resistance.9 12 14 41 42 However, for patients with early-stage disease or in the MRD setting after curative-intent treatment, the pool of ctDNA fragments is more limited, making technical analysis for detection more challenging.

The analytical limits of ctDNA assays are frequently discussed in terms of the variant allele frequency (VAF)—also referred to as mutant allele fraction—which is the percentage of sequencing reads containing tumor-specific mutations among the total number of sequencing reads overlapping the same genomic loci. Practically, these limits can be better understood by thinking of individual molecules of tumor-derived cfDNA, and the total amount of genome equivalents (the amount of DNA in one whole copy of a genome) that exist within a sample of blood. From an idealized peripheral blood draw of one full blood collection tube (approximately 10 mL), one can isolate approximately 5 mL of plasma. In patients with cancer, this plasma is expected to contain roughly 50 ng of DNA (~10 ng/mL), which corresponds to approximately 15 000 haploid genome equivalents.43–46 At a VAF of 0.1%, consistent with localized malignancy or post-treatment MRD, this equates to only 15 molecules of tumor DNA (figure 1).

As it is rare to be able to sequence a cfDNA sample to exhaustion (in the aforementioned example, doing so would require a unique sequencing depth of 15,000×), it would be challenging to recover all 15 tumor DNA molecules with a specific mutation within the blood sample. Sequencing to deep coverage of 1000× would be expected to recover only 1 of the mutant molecules, with a high chance of false-negatives.45 This issue illustrates the challenge of early cancer and MRD detection at low VAFs, and was highlighted in a 2021 FDA evaluation of five commercial ctDNA assays.47 In their detailed validation including synthetic spike-in control DNA and cell line reference samples, the FDA noted that all commercial assays performed well at VAF levels above 0.5%, but were unreliable below that cut-off, demonstrating discordant results among vendors, labs and assay replicates.

Still, there are potential solutions to address these challenges of low VAF MRD detection through a combination of (1) enrichment of tumor variants (eg, through fragmentomic size selection), (2) personalized sequencing panels, (3) multimutation tracking, (4) molecular barcodes to distinguish tumor variants from PCR errors, and (5) background error correction to distinguish tumor variants from oxidative damage and non-biological alterations.23 24 33 43 48 49

CTDNA DETECTION APPROACHES

PCR

The most straightforward approach to ctDNA detection is through PCR. The first FDA-approved diagnostic for ctDNA was the Roche cobas EGFR Mutation Test V.2—an RT-PCR-based approach that targets 42 common mutations within the EGFR gene. Notably, this test was approved only for patients with known advanced NSCLC who did not have tissue available for EGFR sequencing, and the FDA recommended patients who were negative by this ctDNA test to undergo confirmatory tissue testing. The analytical LOD of this assay was modest, approximately 5% VAF (although this LOD is reported to vary between 1.4% and 13.4% depending on the specific mutation).6 7

One alternative that improves the LOD of PCR is digital droplet PCR (ddPCR)—a microfluidic technique.
which became broadly available in 2011 that performs individual PCR reactions within water-in-oil droplets. Both ddPCR and a closely related technique (BEAMing) enable a 10–100 × increase in sensitivity over traditional PCR, with studies attaining consistent detection of VAFs from 0.1% to 0.01%. However, these techniques are still limited by targeting single or a small number of known, predefined mutations, which make them somewhat inflexible and challenging to scale up for early detection and MRD settings where detecting ctDNA requires simultaneous tracking of several mutations.

Next-generation sequencing-based approaches
The limitations of PCR-only approaches prompted the development of next-generation sequencing (NGS) technologies to improve sensitivity, lower the limit of detection (LOD), and add flexibility. An extensive range of techniques has been demonstrated for ctDNA detection, though modern studies are generally dominated by multiplex PCR-based NGS and hybrid capture-based NGS. Both hybrid capture and multiplex PCR-based NGS represent a significant improvement over more traditional PCR and enable a much wider range of analysis of genomic variants. In multiplex PCR-based NGS, which was popularized for ctDNA detection in part by Safe-SeqS (and now SaferSeqS), unique molecular identifiers (UMIs) are incorporated into ctDNA fragments before further PCR amplification and sequencing. Related techniques power Natera’s Signatera, Invitae’s RaDaR, and Invitae’s PCM ctDNA detection assays. Hybrid capture-based NGS was popularized for ctDNA detection by approaches such as Cancer Personalized Profiling by deep Sequencing (CAPP-Seq), targeted error correction sequencing (TEC-seq), and tagged-amplicon deep sequencing (TAD-seq). Modern hybrid capture-based approaches incorporate UMIs and form the basis for multiple clinical ctDNA assays (such as Guardant360, FoundationOne Liquid, and Tempus xF) and research use only assays (such as Roche AVENIO).

Hybrid capture-based NGS and CAPP-Seq
Hybrid capture-based NGS was initially developed for whole exome sequencing of cellular DNA before being adapted to ctDNA. In hybrid capture, genomic regions of interest are identified and then biotin-linked complementary probes are designed to cover and flank these regions. ctDNA molecules are then ligated to barcoded adapters, amplified by PCR, and a biotinylated probe set is used to ‘capture’ the targeted regions. These probes are then isolated by binding to streptavidin-coated beads and the captured fragments are sequenced.

CAPP-Seq, an early hybrid capture NGS technology, was developed initially for the analysis of ctDNA in NSCLC patients. In the initial CAPP-Seq study, a custom hybrid capture panel was designed against recurrently mutated genomic loci from population-level tumor sequencing data (via The Cancer Genome Atlas and the Catalogue of Somatic Mutations in Cancer) along with fusion and breakpoint regions in the ALK, ROS1 and RET genes. The resulting NSCLC-specific panel was approximately 125 kB in size and was validated both computationally and in human samples and cell lines, demonstrating 96% sensitivity for detecting VAFs down to ~0.02%.

Custom CAPP-Seq panels have subsequently been applied to a number of malignancies including diffuse large B cell lymphoma, esophageal cancer, bladder cancer, prostate cancer, colorectal cancer, pediatric sarcoma, and pancreatic cancer. The approach of hybrid capture NGS now backs two of the FDA-approved ctDNA panels for solid malignancies: Guardant360 (targeting 74 genes encompassing SNVs, indels, amplifications, and fusions) and FoundationOne Liquid CDx (targeting 311 genes, including 309 with whole exon coverage). We also first showed that hybrid-capture NGS of ctDNA can be used to infer exome-wide tumor mutational burden (TMB), a finding that was extended further by others using the FoundationOne Liquid CDx assay. Currently, both commercial hybrid capture-based NGS tests (FoundationOne Liquid CDx and Guardant360) enable clinicians to noninvasively infer TMB and detect microsatellite instability (MSI), which can have important roles in immunotherapy response prediction.

In clinical practice, both Guardant360 and FoundationOne Liquid CDx are approved as companion diagnostics to help match patients with an established diagnosis of solid tumor malignancy with potential therapies. Both tests show variable sensitivity, performing best at detecting SNVs and indels, with less sensitivity for rearrangements. For example, FoundationOne Liquid CDx shows a median LOD of approximately 0.3% for actionable EGFR mutations (both L858R and exon 19 deletions), but only ~0.9% for the NPM1-ALK fusion. As a result, FDA approvals for both of these liquid biopsy assays emphasize that negative ctDNA results should be reflexed to tissue mutation testing if feasible. However, a liquid first strategy is recommended as an alternative option to tissue genotyping when time to results is clinically important or tissue biopsy is unavailable.

Lowering the LOD for more sensitive MRD detection
The original version of CAPP-Seq had an LOD of ~0.02%. The subsequent iteration of CAPP-Seq lowered this LOD by ~10 fold to ~0.002% by including two key innovations known collectively as integrated digital error suppression (iDES): molecular barcoding to distinguish true mutations from PCR errors, and background polishing to suppress errors arising from oxidative damage during the library preparation process. We used this iDES-enhanced version of CAPP-Seq to detect MRD in localized patients with lung cancer after curative-intent treatment with 94% sensitivity and 100% specificity at levels as low as ~0.003%.

Other groups, however, have reported lower sensitivity at ~40% for MRD detection using modern ctDNA assays. Indeed, while ctDNA detection using these
approaches has been shown to be highly specific for MRD detection, the sensitivity remains modest with a false negative rate that may be too high for robust clinical implementation using current technologies. These challenges with false-negative ctDNA detection for MRD have led to the development of ultra-sensitive platforms such as MRDetect and PhasED-seq which employ novel strategies to lower the analytical LOD of ctDNA down even further, into the part-per-million range.86 87

Although the LOD of iDES-enhanced CAPP-Seq improved nearly 10-fold when requiring duplex variants (variants identified on both strands of a DNA molecule by deep NGS and molecular barcode-matching),43 at ultra-low mutant allele frequencies, CAPP-Seq can still struggle to detect ctDNA. PhasED-seq builds on this concept by focusing on phased variants (PVs)—that is, two SNVs that occur in cis (on the same strand of DNA). PVs may have a higher practical recovery rate than duplex variants in cancer types that have a high mutational burden.88 By identifying PVs, PhasED-seq can call true mutations in these mutationally rich cancers with high confidence, as the probability that two (or more) mutations occurring due to chance on the same strand is extremely low. This technique has demonstrated a remarkably low limit of detection, in the parts-per-million (ppm) range, with reproducible linearity down to 1 part per 2 million molecules in ctDNA serial dilution experiments.86

At a technical level, PhasED-seq uses hybrid capture to select regions to sequence. In the original paper, the authors identified tumors with high PV burden, analyzing public sequencing databases for variants that occur within ~170 bp of each other (the average size of a cfDNA fragment). They noted that certain malignancies have considerable rates of PVs (≥3% of total SNVs), notably B-cell lymphomas (which have hypermutation driven by AID), melanoma, and NSCLC, while PV rates were lower in other cancer types.86

Given the PV enrichment in B-cell lymphomas, the PhasED-seq authors focused their study on diffuse large B-cell lymphoma (DLBCL). They compared PhasED-seq to CAPP-Seq in a cohort of 107 DLBCL patients receiving standard immuno-chemotherapy. Among 88 patients with samples available after two cycles of treatment (a time point used to assess major molecular response (MMR)86), 59% (52/88) had undetectable ctDNA by CAPP-Seq, while PhasED-seq detected PVs in 25% of those samples (13/52) at levels as low as ~3 parts per million. The authors additionally showed that detection of even these ultra-low-levels of ctDNA PVs was prognostic for event-free survival, and that DLBCL patients with undetectable ctDNA by PhasED-seq after treatment had favorable outcomes compared with their PV-positive counterparts.86

The challenge with PhasED-seq’s broader application is the lower rates of PVs in solid tumor malignancies compared with lymphoid cancers. While some solid tumors show APOBEC3B-associated kataegis hypermutation,89 90 and the PhasED-seq authors also noted that PVs in multiple tumor types were associated with SBS mutations (a signature of tobacco use)—these do not approach the high density of PVs in DLBCL with hypermutation phenotypes. To extend PhasED-seq beyond B cell malignancies, the PhasED-seq authors proposed the development of personalized PV-enriched panels for solid tumors that are informed by up-front whole genome sequencing (WGS) of tumor-normal pairs. The authors demonstrated the feasibility of this approach in 24 plasma samples from five patients with lung cancer and one with breast cancer, showing that their technique achieves ctDNA detection at levels as low as 0.94 parts per million and at multiple timepoints deemed negative by CAPP-Seq.86

Clonal hematopoiesis

Although the requirement to develop individually personalized tumor and normal sequencing panels for solid tumor malignancies can be time and resource-consuming, this approach is already being adopted by a number of groups to address a key source of biological noise: clonal hematopoiesis of indeterminate potential (CHIP). CHIP refers to the age-associated accumulation of somatic mutations in hematopoietic cells, and is a risk factor for future hematologic malignancy and cardiovascular disease.91–95 CHIP can significantly confound ctDNA detection96 97 and result in false positives if not carefully addressed.

When defined as a VAF>2% in peripheral blood, CHIP mutations have been found in 40% of individuals over 60 years of age,98 with rates increasing with age.99 100 When using lower VAF thresholds and modern sequencing approaches, CHIP mutations as low as 0.03% VAF may encompass 95% of the 50–70 years old population,99 although the clinical implications of this high prevalence of low-VAF CHIP remain unclear.100 While CHIP mutations commonly occur in hematologic malignancy-associated genes—classically DNMT3A, TET2, and ASXL1—they also occur in TP53, APC, KRAS, BRCA1 and a range of other genes relevant to solid tumors, and may also result from chemotherapy or radiation therapy.101 102 103 Strikingly, 10% of all CHIP mutations in the Circulating Cancer Genome Atlas (CCGA) study involved the TP53 gene, highlighting the challenge they pose when trying to detect solid tumor malignancy using mutation-based approaches to ctDNA analysis.86

The noise introduced by CHIP was evident even in the initial CAPP-Seq study of NSCLC, where a specific TP53 mutant allele was noted at a median frequency of 0.18% across all samples (including healthy controls) and had to be manually excluded from the analysis.48 This manual curation may be less feasible for larger-scale deployment of ctDNA-based diagnostics, although variants that remain at similar frequencies across serial plasma samples have been shown to be more characteristic of CHIP.104

One of the most comprehensive ctDNA versus CHIP studies compared ctDNA derived from 124 patients with metastatic malignancies to 47 healthy controls.105 Patients with malignancies had paired tumor samples available,
and all subjects underwent targeted ultra-deep sequencing (at 60,000× raw depth) of 508 genes from cfDNA, along with paired PBMCs. Remarkably, most cfDNA mutations (81.6% in controls and 53.2% in patients with cancer) were also found in paired PBMC samples, consistent with CHIP. Similarly, the CCGA study of 836 patients with solid tumor malignancy and 576 controls sequenced matched plasma and PBMCs, and noted that nearly all individuals had somatic mutations due to CHIP. CHIP rates increased at lower VAFs, with 7% of individuals harboring CHIP with VAF>10%, 39% with CHIP at VAFs>1%, and 92% with CHIP at VAFs>0.1%. Given these findings, the most conservative approach to addressing CHIP is through paired deep sequencing of matched PBMCs to filter results from plasma, which is especially critical for mutation-based tumor-naïve assays querying ctDNA at low levels (figure 2).

Groups are also developing bioinformatic approaches to try to more reliably distinguish CHIP from ctDNA. Another approach is to use a tumor-informed assay—that is, to track variants in plasma that are first identified within a patient’s tumor tissue biopsy. As tumor biopsies are commonly obtained at the time of cancer diagnosis and as the question of MRD is particularly important after surgical tumor resection, this type of personalized tumor-informed liquid biopsy approach is clinically feasible for many patients and is used in commercial ctDNA MRD assays such as PCM (Invitae), RaDaR (Inivata), Signatera (Natera), and NeXT Personal (Personalis).

However, this tumor-informed approach may have some limitations too, as the biopsy specimen used to design the assay may miss subclonal variants that were not present within the sampled tissue. Tumor-informed assays can also be limited by the type of biopsy obtained at the time of diagnosis (e.g., core vs fine-needle aspiration), the tumor purity of the sample, and the quality of the surgical resection specimen (e.g., effects of neoadjuvant treatment). There are also clinical scenarios where curative-intent treatment may be rendered without any prior tissue biopsy, such as via stereotactic radiotherapy for select lung and liver cancers, where a tumor-naïve ctDNA MRD detection approach would be more practical.

**Emerging techniques**

Panel-based mutation detection in cfDNA is ultimately limited by the number of ctDNA fragments possessing each on-panel variant. Several groups including ours have now shown that broader sequencing to survey beyond focal recurrent mutations can improve the ctDNA limit of detection, with important implications for early cancer diagnostics and emerging applications to MRD detection. Targets of these broader approaches include methylated DNA, genome-wide copy number alterations (CNAs), and fragment-level sequencing features (fragmentomics).

For example, Zviran et al developed and published MRDetect in 2020, a WGS-based ctDNA assay for global SNV detection with read-centric noise suppression of features known to correlate with sequencing errors (such as variant position within a read and variant base quality). The authors also measured genome-wide CNAs in patient plasma and compared this to background signal in healthy controls. Integrating genome-wide SNV and CNA signals from plasma sequenced to 35× coverage using WGS, they reported the ability to detect ctDNA VAFs as low as 10 parts per million. This approach demonstrated sensitive postoperative ctDNA MRD detection and identified patients who would go on to develop disease recurrence in a small cohort of patients with colorectal cancer (n=19) and another cohort of lung adenocarcinoma patients (n=22).

We also performed WGS in a study we published in *PLOS Medicine* in 2021, with the goal of detecting neurofibromatosis type 1 (NF1) patients who harbor malignant peripheral nerve sheath tumor (MPNST) versus the non-malignant plexiform neurofibroma precursor lesion. We used a highly economical and scalable approach called ultra-low-pass WGS (ULP-WGS), sequencing the full genome to only ~0.6× depth of coverage. We then measured CNAs and inferred the liquid biopsy tumor fraction using the ichorCNA platform. While the ULP-WGS-derived tumor fraction was not able to accurately discriminate MPNST from non-malignant plexiform neurofibroma on its own, combining it with fragment size information enabled us to discriminate MPNST from its plexiform neurofibroma precursor with 89% accuracy. Additionally, our work also established that cfDNA fragments from patients with cancer appear to be shorter than fragments from patients with the corresponding precancerous lesion, extending on prior findings showing that patients with cancer have overall shorter cfDNA fragments than healthy donors.
To explore cfDNA fragment size distributions and CNA integration in greater detail, Cristiano et al. used low-pass WGS to track fragment sizes across the genome in five megabase bins, and measured bin-wise ratios of short (100–150 bp) to long (151–220 bp) cfDNA fragments. These features were then included in a machine learning model (after GC content and library size normalization), along with mitochondrial copy number and chromosomal arm copy number features. Using this model, called ‘DNA evaluation of fragments for early interception’ (DELFI), with 10-fold cross-validation, the authors were able to distinguish 208 patients with cancer from 215 healthy individuals with an area under the receiver operating characteristic curve (AUROC) of 0.94 across patients with stages I–IV cancer across seven different malignancies. Sensitivity of the approach remained high for stage I cancer at 71% with 98% specificity within this internal cross-validation framework. The research group further showed that DELFI scores could be used to stratify cancer-specific survival in patients with lung cancer, and could outperform serum alpha fetoprotein for liver cancer detection.

cfDNA fragments are cleaved into canonically sized ~170 base pair fragments by nuclease and are protected from further cleavage by the nucleosomes that these fragments are wrapped around. Emerging evidence suggests that the DNases responsible for cleaving genomic DNA into cfDNA fragments show preferences for specific sequence motifs, methylation states, and epigenetic modifications, all features that can be used to identify tumor-derived fragments. Jiang et al. found that specific 4-mer end motifs were enriched in cfDNA fragments from hepatocellular carcinoma (HCC) patients compared with either healthy controls or patients with hepatitis B, a finding that may be related to the downregulation of DNASE1L3 in HCC and other tumor types. A machine learning model applied to plasma cfDNA using all possible 256 4-megend motifs had an AUROC of 0.89 to distinguish HCC from healthy donors. Tumor-derived plasma cfDNA fragments from HCC patients were also observed to have a higher rate of single-stranded ends (‘jagged ends’) compared with non-tumoral DNA. Jagged ends were also observed in urine cfDNA and were found to be present at a lower rate in patients with bladder cancer than in healthy controls. Interestingly, there was a relationship between jagged ends in plasma and urine cfDNA and nucleosome occupancy, suggesting that jaggedness could be used to infer epigenomic structure.

Epigenomic state can also be inferred from cfDNA fragment populations sequenced with basic WGS. This was first demonstrated by Snyder et al. who showed that a ‘windowed protection score’ of fragments spanning a window of genomic positions versus fragments with endpoints within the window could be used to infer nucleosome positioning at DNase I hypersensitivity and transcription factor binding sites. This enabled the prediction of tissue types contributing to cfDNA fragments, which correlated with the tumor tissue or origin in five patients with late-stage cancer. Ulz et al. also used WGS and showed that relative coverage of cfDNA at the transcription start site (TSS) could be used to infer nucleosome positioning at the TSS, which in turn could predict gene expression status. Both of these earlier approaches used high-coverage WGS, which may not be economically practical or scalable. More recently, however, Doebley et al. developed a framework for profiling nucleosome protection and accessibility from cfDNA sequenced with ultra-low-pass WGS (as low as 0.1× coverage) and employed GC correction tailored to variable cfDNA fragment sizes. These optimizations facilitated better cancer detection and tumor subtype classification accuracy in a more economical and scalable format.

An orthogonal approach to cancer detection and tumor tissue of origin inference uses methylation sequencing of cfDNA. The Circulating Cell-free Genome Atlas (CCGA) study reported that whole-genome bisulfite sequencing of cfDNA at 30× coverage had higher sensitivity than targeted sequencing for SNVs across several cancer types. This group went on to generate a 17.2 Mb panel covering over 100,000 informative regions which they applied using targeted bisulfite sequencing to a 6089-participant cohort of patients with cancer and healthy individuals. Using this approach, they achieved 18% sensitivity for stage I cancer detection from cfDNA with >99% specificity across >50 cancer types. Updated results from this group in the prospective PATHFINDER study in a screening population of adults 50 years of age or older revealed a detected cancer signal in 1.4% (95% confidence interval 1.1%–1.8%) of participants, with cancer confirmed in 35 of these 92 participants. Test specificity was 99.1%, the negative predictive value was 98.6%, and positive predictive value was 38.0%. This study revealed that multicancer early detection was feasible in the outpatient screening setting. An alternative methodological approach is cell-free methylated DNA immunoprecipitation and high-throughput sequencing (cfMeDIP-seq), which also showed promising results for detecting early-stage lung and pancreatic cancers. Methyl-based cfDNA assays are also able identify cancer tissue of origin based on known tissue-specific methylation signatures, with the CCGA group showing >75% accuracy of tissue of origin prediction in stage I cancers, which rose to >90% in more advanced malignancies.

Recent studies further suggest that the biological underpinnings of cfDNA methylation and fragmentation profiles are deeply intertwined, such that it may be possible to infer one set of features from the other. Ultimately, both cfDNA methylation and fragmentomics are measures of epigenomic phenomena. In addition to paradigm-shifting clinical potential in early cancer and MRD detection, these technologies have promise in more basic biological research to facilitate a greater understanding of the cancer epigenome with the ability to track its evolution noninvasively via liquid biopsy.
Alternative sources of cfDNA

One strategy for improving the clinical sensitivity of ctDNA assays in low disease burden settings may be through sampling non-plasma biological compartments that are more proximal to the tumor site.34 69 70 124 125 Although cfDNA and ctDNA frequently refer to peripheral blood plasma-derived DNA, numerous studies have analyzed cfDNA and identified ctDNA across different biological fluids, including urine,69 70 124 tears,126 saliva,127 CSF,125 pleural, and peritoneal fluid among others128–132 (Figure 3). These distinct biological compartments each pose both opportunities and challenges for analyzing genomic alterations in cfDNA. Indeed, fluid isolated from these alternative sources can be enriched for fragments with genomic alterations from locoregionally present malignancies, a finding that has been shown in stool (for colorectal cancer),73 urine (for urothelial cancers),68–70 134–136 CSF (for CNS malignancies),125 137 138 and pleural and bronchoalveolar lavage fluids (for lung cancer)131 139 among others.129 Our data further suggest that if an alternative biofluid (such as urine) is distal to plasma for a tumor type, the reverse relationship is also true with significant de-enrichment of ctDNA in the alternative biofluid compared with plasma.34 One challenge in studying these non-plasma biofluids is that they are generally more difficult to serially collect and process, although biofluids such as saliva and urine could be collected even more readily than plasma without requiring phlebotomy.

A number of groups including ours are actively exploring the utility of these alternative biofluid sources of cfDNA to more sensitively detect MRD after curative-intent treatment—especially in patients at high risk for locoregional relapse—and to provide greater insights into geographical tumor heterogeneity.140 Importantly, there are several mechanisms underlying cfDNA release into the plasma extracellular space including apoptosis, necrosis, and active secretion via extracellular vesicles (EVs),141–143 with recently published data indicating that tumor-derived cfDNA fragments are primarily free-floating within plasma, while ctDNA encapsulated within exosomes is mostly normal.144 145 It will be important to study the topology of ctDNA versus normal cfDNA in alternative biofluids too, where these mechanisms are less well understood. Additionally, the tumor microenvironment may provide different environmental pressures that alter the presence and composition of ctDNA molecules,140 and sampling of these alternative biological fluids may be critical to decipher complex tumor ecosystems.

PREDICTION AND MONITORING OF IMMUNOTHERAPY RESPONSE

An emerging goal of modern cfDNA liquid biopsy technology is to predict immunotherapy response and personalize the administration of immune checkpoint blockade.146 TMB, which is the quantification of tumor-specific non-synonymous mutations measured from sequencing data of tumor tissue, is a precision biomarker for immunotherapy response prediction.147 148 We showed in our 2017 Cancer Discovery paper23 that TMB...
can also be estimated from hybrid-capture cfDNA technology applied to blood plasma, by interpolating the number of non-synonymous mutations in the whole exome from a targeted panel. We applied similar methodology to infer TMB from plasma from patients with colorectal cancer, and from urine in patients with bladder cancer.

This inference of TMB from liquid biopsy-targeted NGS data was corroborated by Gandara et al. using the FoundationOne Liquid CDx assay, applied to plasma cfDNA samples from NSCLC patients from the POPULAR and OAK studies, with elevated blood-derived TMB (bTMB) patients demonstrating a response to immunotherapy versus chemotherapy, and with bTMB levels correlating with progression-free and overall survival in a dose-dependent fashion. More recently, the bTMB cutpoint of 16 was tested prospectively in stage IIB–IVB NSCLC patients. Among 119 analyzable patients, there was no significant progression-free survival or overall survival benefit in the bTMB≥16 arm with a median follow-up of 20.9 months, however, survival benefits were seen at longer follow-up (median 36.5 months). In addition to inferring TMB, hybrid-capture NGS liquid biopsy assays from Foundation Medicine and Guardant also include microsatellite loci consisting of short-variant tandem repeats, which can measure MSI, another important biomarker for immunotherapy response.

Another strategy for predicting immunotherapy response is via ctDNA dynamics. Specifically, ctDNA levels changing from pre-immunotherapy to on-immunotherapy have been shown to correlate strongly with immunotherapy response, yielding results earlier than standard-of-care imaging. There is also potential for ctDNA MRD detection to serve as a predictive biomarker for adjuvant immunotherapy response after curative-intent surgery or radiotherapy. In this regard, Powles et al. analyzed plasma samples from 581 muscle-invasive urothelial carcinoma patients enrolled onto the IMvigor010 study using the Signatara tumor-informed PCR-based NGS assay. Strikingly, patients with detectable ctDNA MRD after surgery but prior to immunotherapy achieved both a disease-free survival and overall survival benefit with immunotherapy, while patients with undetectable ctDNA MRD after surgery did not. Additionally, patients whose ctDNA was detectable before immunotherapy but became undetectable during immunotherapy had superior disease-free survival compared with those whose ctDNA remained detectable. Moding et al. demonstrated similar results in a retrospective analysis of locally advanced NSCLC treated with definitive-intent chemoradiation, showing that patients with ctDNA MRD detectable after radiotherapy appeared to selectively benefit from consolidation immunotherapy, and that decreasing ctDNA levels during consolidation immunotherapy were associated with longer freedom from progression.

**Future directions**

The promise of ctDNA to robustly detect MRD and predict treatment response in clinical practice is captivating, and efforts continue toward enhancing the sensitivity and specificity of ctDNA assays to identify and track ultra-rare variants while accounting for sources of background noise. The landscape of ctDNA analyses is continuously changing, with many of the techniques highlighted here developed within the past few years. Ultimately, ctDNA studies may evolve to include integrative machine learning models, such as those advanced by MRDetect to overcome sequencing noise and detect ultra-low frequency mutations. Outside the scope of this review, there are further liquid analytes such as circulating RNA, circulating tumor cells, tumor-educated platelets, and EVs that merit further discussion. Additionally, within the ctDNA space, there are many other active topics of investigation such as delineation of subclonal architecture and tumor evolution, molecular response in advanced disease, sensitivity of single time point MRD detection versus serial monitoring, differences in ctDNA shedding by cancer type, how tumor-specific genomic characteristics can influence assay and analytical strategy, and ctDNA plus other liquid biopsy analyte multisonic approaches to early cancer detection.

While there is excitement regarding ctDNA as an oncogenic biomarker that could supplant standard imaging, pathology, and laboratory tests in the future, it will be important in the nearer term to be able to precisely fit ctDNA testing within the context of standard-of-care diagnostic modalities. This is being done already with ctDNA tests for actionable mutations in metastatic cancer patients, where guidelines recommend that a negative test be reflexed to tissue mutation testing if possible. Similarly, for MRD and surveillance testing in localized cancer patients, the results from ctDNA assays will need to be integrated seamlessly into standard clinical practice to guide clinical decision-making while minimizing confusion and patient anxiety. For early cancer detection in the screening setting, both assay sensitivity and specificity will need to be superb, and there will need to be clear clinical guidelines for addressing positive results, which should also include psychosocial considerations given the anxiety associated with false positive results in otherwise healthy individuals. These near-term challenges are similar to those faced by other game-changing diagnostic technologies in oncology such as PCR, NGS, mammography, and functional imaging. Like these other technologies, ctDNA has the potential to be the next frontier of personalized medicine.

**Author affiliations**

1 Division of Endocrinology, Metabolism, and Lipid Research, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA
2 Division of Cancer Biology, Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri, USA
3 Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, Missouri, USA
Open access


