


Review

Bioinformatics Analysis for Circulating Cell-Free DNA in Cancer

Chiang-Ching Huang ^{1,*}, Meijun Du ² and Liang Wang ^{2,*} 

¹ Zilber School of Public Health, University of Wisconsin, Milwaukee, WI 53205, USA

² Department of Pathology and MCW Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226, USA; mdu@mcw.edu

* Correspondence: huangcc@uwm.edu (C.-C.H.); liwang@mcw.edu (L.W.);
Tel.: +1-414-227-5006 (C.-C.H.); +1-414-955-2574 (L.W.)

Received: 22 April 2019; Accepted: 6 June 2019; Published: 11 June 2019



Abstract: Molecular analysis of cell-free DNA (cfDNA) that circulates in plasma and other body fluids represents a “liquid biopsy” approach for non-invasive cancer screening or monitoring. The rapid development of sequencing technologies has made cfDNA a promising source to study cancer development and progression. Specific genetic and epigenetic alterations have been found in plasma, serum, and urine cfDNA and could potentially be used as diagnostic or prognostic biomarkers in various cancer types. In this review, we will discuss the molecular characteristics of cancer cfDNA and major bioinformatics approaches involved in the analysis of cfDNA sequencing data for detecting genetic mutation, copy number alteration, methylation change, and nucleosome positioning variation. We highlight specific challenges in sensitivity to detect genetic aberrations and robustness of statistical analysis. Finally, we provide perspectives regarding the standard and continuing development of bioinformatics analysis to move this promising screening tool into clinical practice.

Keywords: bioinformatics; copy number variation; cell-free DNA; methylation; mutation; next generation sequencing

1. Introduction

To date, tissue biopsy samples are widely used to characterize tumors. Although tissues allow the histological definition of the disease and can reveal details of the genetic profile of the tumor, enabling prediction of disease progression and response to therapies, the applications are limited on tissue availability, sampling frequency, and their genetic heterogeneity [1]. Therefore, attention is turning to liquid biopsies, which enable the analysis of tumor components, including circulating tumor cells (CTC) [2] and circulating tumor nucleic acids from various biological fluids, mostly blood but also other easily accessible fluids such as urine [3]. Compared to conventional tissue biopsy from a single tumor site, the main advantages of liquid biopsies include their non-invasive characteristics, multiple sampling capability, and comprehensive coverage to address issues of tumor heterogeneity [4,5].

Circulating cell-free DNA (cfDNA) is defined as extracellular DNA occurring in blood or other body fluids. It is usually released as small fragments (150–200 bp in length [6]) from normal or tumor cells by apoptosis and necrosis [7], or shed from viable cells [8]. Levels of cfDNA are higher in diseased than healthy individuals [9]. cfDNA can track the evolutionary dynamics and heterogeneity of tumors and detect the early emergence of therapy resistance, residual disease, and recurrence [10–12]. Therefore, analysis of cfDNA has been considered as a potential screening approach for tumor diagnosis and prognosis by detecting tumor-associated aberrations in peripheral blood [13,14].

Next generation sequencing (NGS) has emerged as a powerful tool for cfDNA analysis, which allows the detection of cancer-related genetic and epigenetic alterations such as mutations, copy number

variations (CNVs), and DNA methylation changes across wider genomic regions in many cancer types [15,16]. However, detection of cancer with high specificity and sensitivity is still challenging, especially in early-stage cancers, as there exist many barriers to the utilization of cfDNA in clinical applications, including lack of well-accepted sample collection protocol and sensitive detection approaches. Furthermore, analysis of cfDNA sequencing data requires specialized bioinformatics tools to identify robust biomarkers for clinical practice. In this review, we will discuss specific challenges in sensitivity to detect genetic aberrations and provide information on cfDNA bioinformatics approaches. We conclude with a perspective regarding future development in this rapidly evolving area. A simplified workflow of blood-based liquid biopsy is shown in Figure 1.

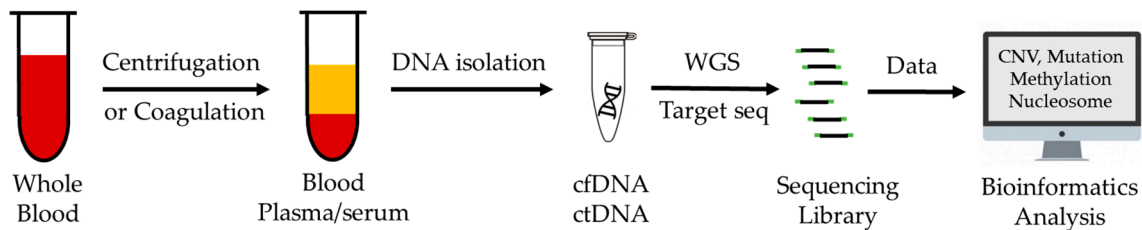


Figure 1. Workflow of blood-based liquid biopsy.

2. Characteristics of Circulating Tumor DNA (ctDNA)

The ctDNA is released from tumor cells only. The ctDNA can be derived from primary or metastatic tumors [17]. Most circulating ctDNA are 160–180 base pair fragments, roughly the size of a mononucleosomal unit [18,19]. However, recent studies have shown that ctDNA tends to be shorter than cfDNA from normal cells [20,21]. Therefore, ctDNA may be enriched by excising smaller DNA fragments from cfDNA on polyacrylamide gels [22]. Currently, cfDNA fragmentation patterns and their applications in liquid biopsy are an emerging research field. Although ctDNA can be used to detect the presence of cancer-related genetic and epigenetic changes, such changes usually vary from case to case, which makes the development of sensitive and generalizable approaches extremely challenging. One major challenge is low ctDNA fraction. In most cases, ctDNA accounts for a small fraction of total cfDNA since most cfDNA is derived from non-cancer cells, especially blood cells. In early-stage cancer patients, ctDNA fraction could be lower than 0.1%. To detect such a rare event with high specificity and sensitivity, a variety of approaches have been developed, which include droplet digital PCR (ddPCR) and molecular index-based next generation sequencing technologies [23,24].

3. Detection and Analysis of Somatic Mutations

Somatic mutations are involved in cancer development and progression. The presence or absence of a single genetic alteration in tumor DNA is currently employed to guide clinical decision making for a number of targeted agents [25–28]. Ever-increasing numbers of genomic alterations are being tested as putative predictive biomarkers in clinical trials of novel anticancer therapies [29]. To detect the cancer-associated alleles in the blood, real-time PCR (RT-PCR) and ddPCR “targeted” methods have been extensively adopted in most clinical trials [30]. Till now, clinical utility has been demonstrated for two FDA-approved cfDNA-based tests: the cobas epidermal growth factor receptor (EGFR) mutation test V2 (Roche Molecular Diagnostics), which detects EGFR mutation in plasma cfDNA from patients with lung cancer [31,32], and Epi proColon (Epigenomics AG), which reports on the methylation status of the Septin 9 promoter in plasma cfDNA from patients undergoing screening for colorectal cancer [33]. ddPCR is particularly useful to sensitively detect well-characterized mutations. The system can partition cfDNA into 20,000 nanoliter-sized droplets, where PCR amplification is carried out simultaneously. It is reported that the sensitivity of ddPCR can reach a limit of detection of 0.0005% BRAF V600E and V600K [34]. Another study reported that ddPCR can reliably detect AR-V7 expression from one spiked cell into 4000 lymphocytes (0.025%) [35]. Compared to the traditional NGS method, ddPCR is easier to use, has lower cost, and provides higher sensitivity and specificity. Although

molecular barcoding technology has significantly increased the sensitivity and specificity of NGS, the low cost and easy-to-use features will make ddPCR widely accepted in clinical practice.

Although PCR-based assays can detect known mutations, the assay requires previous knowledge of target genes. In addition, the assay does not cover whole spectrum mutations in specific genes. Restriction of multiplexing capacity limits the simultaneous analysis of a large number of gene targets. Therefore, it may fail to identify less common but clinically relevant mutations. On the other hand, NGS, based on massive parallel sequencing of millions of different DNA molecules, allows the detection of multiple mutations in multiple genes. By using focused gene panels on clinically relevant targets, each nucleotide of interest can be sequenced thousands of times, ensuring a high degree of sensitivity. However, the requirement for such a high degree of sensitivity can easily lead to false positive results due to potential errors of PCR amplification and sequencing. To address this challenge, new data analysis approaches have been developed, among which is a new unique molecular identifier (UMI) strategy [36]. Another challenge related to mutation detection in cfDNA is to differentiate tumor mutations from background somatic mutations. Somatic mutations are common in healthy individuals with a rate between 2–6 mutations per 1 Mb [37]. Given the fact that the majority of cfDNA is from blood cells and ctDNA fraction in cancer patients is generally low, it is likely that most of the mutations identified in cfDNA could be irrelevant to cancer development, thereby impeding their clinical application [38–40]. This challenge points to the need for a large experiment to systematically investigate the mutation spectrum from both cfDNA and white blood cells in healthy and cancer patients.

4. Unique Molecular Identifier (UMI)-Based Target Sequencing

Target enrichment is a critical component of targeted deep sequencing for cost-effective, accurate, and sensitive detection of mutations, CNVs, and methylations in cfDNA. Common bioinformatics workflows allow sensitive and specific variant identification down to 2–5% allele frequency. This provides a sound methodology for identifying somatic mutations from solid tumor biopsies [41]. However, low ctDNA content in the blood and sequencing artifacts currently limit analytical sensitivity. In analyzing cfDNA from healthy controls, background errors are increasingly evident below allele fractions of ~0.2%. It is reported that under an allele fraction of 0.02%, >50% of sequenced genomic positions had artifacts [42]. In addition, common NGS assays involve multiple steps, including end repair, ligation, PCR, and sequencing. These steps often introduce technical biases, limiting accurate quantification and, therefore, hindering the robust and clinically valid detection of biomarkers [43]. Furthermore, PCR-based target enrichment cannot distinguish PCR duplicates from copies of unique fragments generated by a pair of PCR primers.

To overcome these limitations, UMIs (also known as molecular barcodes) have been added into the adaptors to tag individual DNA molecules [44–47]. Such barcodes enable the precise tracking of individual molecules. UMIs can accurately distinguish PCR duplicates from copies of unique fragments generated by PCR amplification [36]. Moreover, UMIs can reduce quantitative bias during experimental processes to detect true ultra-rare variants by distinguishing authentic somatic mutations arising *in vivo* from artifacts introduced *ex vivo*. This is largely due to the fact that errors arising from artifacts during library construction and sequencing runs could be eliminated by comparing the sequences of PCR duplicates identified with a UMI sequence [42,48]. Figure 2 illustrates the basic principle of UMI application in the detection of true somatic mutations. Dedicated bioinformatics software packages (Table 1) have been developed for the UMI-tagged targeted resequencing data to improve ultra-rare variant calling by removing errors arising from the first cycle PCR [49,50].

Incorporation of molecular barcoding into a bioinformatics algorithm has significantly increased sensitivity of mutation detection in NGS data. The detection sensitivity can be down to 0.01% [57]. However, recent advances in statistical modeling has also increased sensitivity of variant detection without molecular barcoding. A method ERAS-Seq (Elimination of Recurrent Artifacts and Stochastic Errors) that utilizes technical replicates in conjunction with background error modelling has shown an

increased sensitivity of variant detection between 0.05% and 1% allele frequency [58]. By physically extracting and individually amplifying the DNA clones of erroneous reads, another barcoding-free method is reported to distinguish true variants of frequency >0.003% from the systematic NGS error. This method uses 10 times less sequencing reads compared to those from previous studies and achieved a PCR-induced error rate of 2.5×10^{-6} per base per doubling event [59].

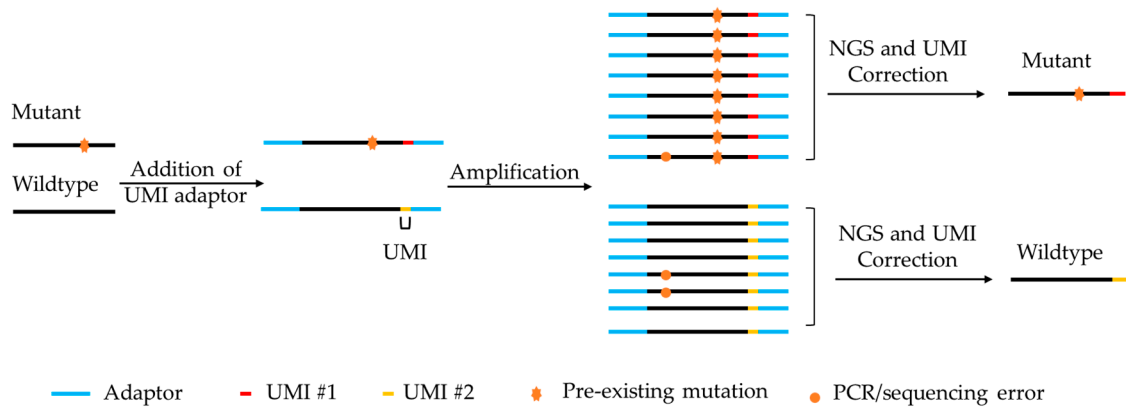


Figure 2. Principle of unique molecular identifiers (UMI) application in the detection of somatic mutations.

Table 1. Bioinformatics programs for detecting genetic and epigenetic changes in cancers.

Program	Website	Key Features	Reference
Mutation			
UMI-tools	https://github.com/CGATOxford/UMI-tools	identifies sequencing errors in the UMI sequence to improve quantification accuracy	[49]
MAGERI	https://github.com/mikessh/mageri	provides an efficient analysis pipeline for UMI-encoded data	[50]
Copy Number			
QDNA-seq	https://github.com/ccagc/QDNAseq	simultaneously corrects for GC and mappability bias	[51]
WisecondorX	https://github.com/CenterForMedicalGeneticsGhent/WisecondorX	optimizes segmentation by reducing noise from problematic bins	[52]
BIC-seq2	http://compbio.med.harvard.edu/BIC-seq/	Avoids high variability of reads in bins	[53]
CNVkit	https://github.com/etal/cnvkit	uses both the targeted reads and the nonspecifically captured off-target reads to infer copy number	[54]
Methylation			
CancerLocator	https://github.com/jasminezhoulab/CancerLocator	simultaneously infers the proportion and tissue of origin of ctDNA	[55]
CancerDetector	https://zhoulab.dgsom.ucla.edu/pages/CancerDetector	Improves ctDNA fraction estimation and identifies outlier markers	[56]

5. Detection of DNA Copy Number Alterations

Currently, most cfDNA applications in cancer screening have focused on somatic point mutations [23,24]. However, methods that interrogate other genomic aberrations should be incorporated to improve detection and characterization of early-stage cancers. One of such genomic abnormalities is CNVs that contribute significantly to genome instability [60,61]. Large-scale cancer genome studies have identified CNVs across various types of cancer and a majority of the CNVs are shared among several cancer types [62,63]. Recently, several lines of investigation have demonstrated the potential of CNVs from cfDNA as sensitive cancer biomarkers [64–66]. Both targeted and whole genome sequencing (WGS) have been employed to identify specific CNVs or genome-wide DNA copy number patterns in cancer patients. Extension of statistical and bioinformatics methods developed from microarray-based comparative genomic hybridization (aCGH) array or NGS are suitable for the detection of CNVs from cfDNA.

For the WGS-based CNV analysis, depth of coverage (DOC) methods (Table 1) are the most used techniques to estimate copy number from the sequence depth in the genome [51–54]. Other methods such as assembly-based, split-read, and read-pair methods [67] can be used to infer copy number changes and chromosomal rearrangement. However, these methods may require high sequence coverage or specific molecular size and thus may not be practical in diagnostic application. The DOC methods can be divided into two major categories depending on whether a reference signal is required. In general, the pseudo-autosomal region on the Y chromosome and genomic regions with low mappability should be removed before the sequencing alignment procedure. This step is especially critical for reference free methods to ensure that the short reads can be mapped to a unique genomic location instead of multiple possible locations. The GEM (GENome Multitool) mappability algorithm [68] is an efficient program that provides mappability information for multiple genomes. In addition, it is important to filter genomic regions that tend to show artificially high signal (i.e., excessive unstructured anomalous reads mapping). These blacklisted regions in the human genome are often found in highly variable regions (e.g., alternative haplotypes overrepresented on chromosome 19) or at specific types of problematic repeats such as centromeres, telomeres, and satellite repeats. The ENCODE and modENCODE consortia have identified these regions and made them available online [69] at <https://sites.google.com/site/anshulkundaje/projects/blacklists>. However, empirical data analysis indicates that the ENCODE blacklist may not be sufficient to remove all problematic regions. As such, the QDNaseq algorithm [51] provides a data-driven approach to identify additional regions that should be removed before downstream analysis.

Due to the high cost of WGS assay, current cfDNA-based approaches to CNVs detection normally have low-sequence coverage (e.g., $0.1\times\sim 0.5\times$ coverage depth) [64,70,71]. As such, the binning procedure is generally required to aggregate reads mapped to a genomic window. After removing the low mappability reads and blacklisted regions, reads in different genomic windows are counted and normalized by the total number of reads. Depending on the read depth, a fixed bin size is normally chosen such that sufficient detection resolution can be achieved while excessive variation of read counts between adjacent windows can be reduced, thereby enhancing the detection sensitivity for CNVs. Although simple, using a fixed bin size may lead to high variability of read counts among bins with a substantially different number of mappable positions. To overcome this problem, the BIC-seq2 algorithm [53] normalizes read counts at a nucleotide level rather than at the bin level. It calculates the expected number of mapped reads for every position in the mappability map. The ratio of the observed read number and expected number of mappable reads is thus used to infer copy number for a specific genomic region. The normalized read counts can be further subject to GC content correction using smoothing techniques such as LOWESS [72]. The GC-corrected read counts are then normalized to the GC-corrected read counts of cfDNA from a group of reference samples (e.g., healthy controls or patient's own germline DNA) and expressed as \log_2 ratio values. For reference-free methods, median normalization can be used to obtain \log_2 ratio values.

Segmentation on the \log_2 ratio values is generally performed to identify the genomic areas with potential CNVs. The purpose of segmentation is to merge adjacent data points with the same copy number into one segment and divide regions with different copy numbers into different segments. Several statistical techniques and tools have been developed. Two of the most popular methods are circular binary segmentation (CBS) [73,74] and the hidden Markov model (HMM) [75,76]. Thorough review and systematic evaluation of CNV detection methods and software resources have been documented previously [52,77–79]. Researchers may use the information therein to choose appropriate algorithms for their projects. After the segmentation, aberration calling will be made to infer DNA regions with abnormal copy number (e.g., >2 or <2 DNA copies for gain or loss). A commonly used method for determining CNVs from the cfDNA of cancer patients using high throughput sequencing is the Z-score based approach [64,80–82]. These methods identify CNV segments by determining regions in the cfDNA that are significantly different from the reference panel (e.g., Z-score distribution from normal control). Other methods that make formal statistical inference for copy number are available [83,84]. For example, CGHcall [83] uses a two-level hierarchical mixture model to infer for each segment the likelihood of being one of six states of copy number: double deletion, single deletion, normal, gain, double gain, and amplification. This method uses \log_2 ratio data to estimate the proportion of different copy number states at the chromosome arm level. Therefore, it may require a large number of samples for robust inference, especially for chromosomes in which abnormal DNA copy numbers are rare. A summary of the bioinformatics procedure for WGS-based CNV analysis in cfDNA is shown in Figure 3.

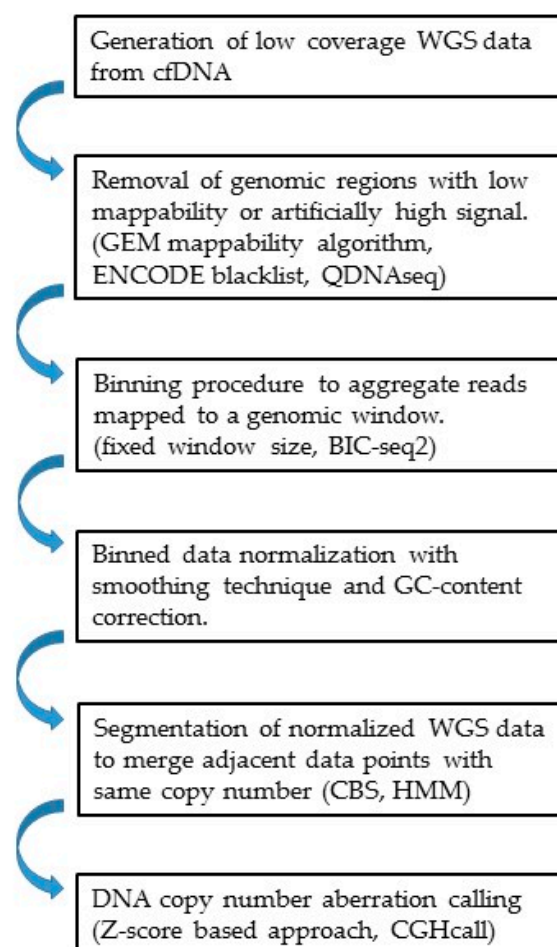


Figure 3. Bioinformatics procedure and techniques/resources used to detect copy number variations (CNVs) from low coverage whole genome sequencing (WGS) data.

One of the challenges to infer CNVs from the cfDNA sequencing data is attributable to the ctDNA content and tumor heterogeneity. In a large portion of cfDNA samples with low ctDNA content (i.e., <2%), especially in the early stages of cancer, sequencing reads are dominated by the DNA from non-cancer cells. Therefore, the signals of CNVs from cancer cells are almost entirely masked, leading to very little statistical power for any segmentation algorithms to detect CNVs, especially for focal amplifications or deletions. In addition, multiple clones of cancer cells could coexist in a cfDNA sample. This will make it even more difficult to detect CNVs due to genetic heterogeneity. To overcome this obstacle, Kirkizar et al. [85] developed a method that employs single-nucleotide polymorphism (SNP)-targeted massively multiplexed PCR (mmPCR) followed by NGS (mmPCR-NGS). Haplotype information is then obtained from the experiment to identify both single nucleotide variants (SNVs) and CNVs with high sensitivity and an average allelic imbalance as low as 0.5%. This method can also detect both clonal and subclonal CNVs in ctDNA.

6. Identification of DNA Methylation Changes from cfDNA

DNA methylation is essential for normal development and plays an important role in epigenetic control of gene activity. Changes in DNA methylation have been recognized as one of the most common molecular alterations in tumorigenesis [86,87]. It is well known that each tissue possesses unique methylation signatures and a genome-wide methylation pattern is distinguished between cancer and normal cells [16,88,89]. Therefore, whole genome methylation profiling from cfDNA could be a potentially powerful tool to detect the presence of specific cancer. Lehmann-Werman et al. [90] first demonstrated the feasibility to identify tissue origin using cfDNA. By leveraging whole genome methylation data sets from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) repositories, they identified individual CpG dinucleotides that were unmethylated in the tissue of interest but methylated in other tissues. By comparing genome-wide methylation data from 35 human tissues generated using the Illumina Infinium HumanMethylation450k BeadChip, tissue-specific DNA methylation markers were selected. Subsequently, Moss et al. [91] generated a reference methylation atlas of 25 human tissues including major organs and cells involved in common diseases. For each tissue or cell type, both uniquely hypermethylated and uniquely hypomethylated CpG sites were identified. Additional CpG sites were further identified to differentiate any two cell types that were found to be most similar in the atlas.

With the data for tissue-specific and cancer methylation signatures, deconvolution algorithms [92], a commonly used algorithm to recover the original signal from a mixture of signal sources, can be used to map tumor tissue of origin from cfDNA. Sun et al. [93] used optimization programming to calculate the methylation densities of 5820 methylation markers in cfDNA from bisulfite sequencing data for 14 human tissues. To improve the selection of informative methylation markers, Guo et al. [94] identified 147,888 blocks of tightly coupled CpG sites, called methylation haplotype blocks, after a comprehensive analysis of a large amount of whole-genome bisulfite sequencing data, reduced-representation bisulfite sequencing data, and methylation array data. The deconvolution algorithm was then applied for tissue-specific methylation analysis at the block level. This method was successfully applied to estimate ctDNA content and differentiate among clinical plasma samples from normal individuals and patients of lung cancer and colorectal cancer.

Recently, probabilistic models have been formulated to identify specific cancer types from cfDNA. Kang et al. developed a method, termed CancerLocator [55], to simultaneously infer the proportion and tissue of origin of ctDNA using whole-genome DNA methylation data. By using TCGA Infinium HumanMethylation450 microarray data from both normal and tumor samples, CancerLocator identified as feature input a large number of CpG clusters that have high inter-individual methylation variation across all normal and cancer types. Since cfDNA from the peripheral blood is a mixture of normal and tumor DNA if a cancer cell is present, the methylation level for each CpG cluster, one for normal and the other one for a cancer type, can be estimated and the ctDNA fraction and the likelihood of the presence of a specific cancer type can be inferred based on the methylation data of informative

CpG clusters. CancerLocator demonstrated a superior prediction performance over popular machine learning algorithms (i.e., random forest and support vector machine) on low-coverage sequencing data, especially for samples with low to moderate ctDNA fraction. However, a challenge facing this method is that the classification accuracy depends substantially on the estimated ctDNA fraction of a specific tumor type.

A variation of CancerLocator was developed later by Li et al. [56]. This method, called CancerDetector, differs slightly from CancerLocator in genomic marker selection and estimation. To identify sensitive genomic markers, CpG clusters were identified such that the level of methylation in a specific cancer tissue differs from matched normal tissue as well as normal plasma samples. This procedure ensures that selected markers are not tissue specific and the methylation signal can be detected in the blood. With selected CpG clusters, a similar probabilistic model to CancerLocator was implemented to predict cancer types and ctDNA fraction. To improve the estimation of ctDNA fraction, an iteration procedure was developed to remove outlier markers whose estimated ctDNA fraction are far from the estimated ctDNA fraction when all markers were used. CancerDetector demonstrated substantial improvement over CancerLocator with high sensitivity and specificity in detecting tumor cfDNAs on real plasma data. Figure 4 illustrates the major principle of the bioinformatics approach for tumor tissue-specific methylation analysis.

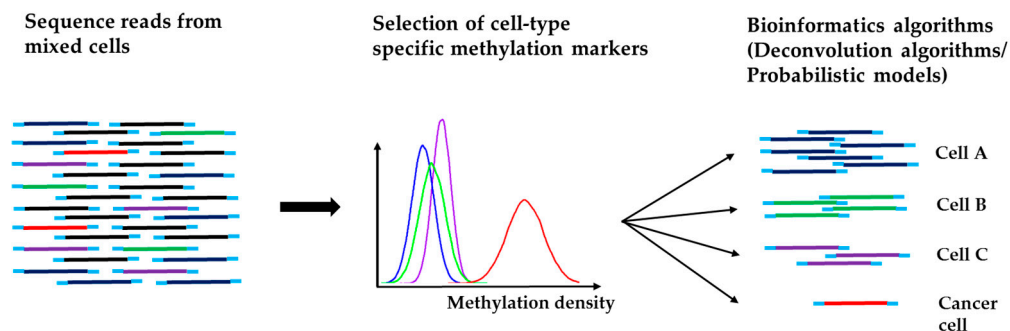


Figure 4. Schematic approach to map cancer tissue of origin from WGS methylation analysis.

7. Association of Nucleosome and Fragmentation Pattern with Tissue of Origin in cfDNA

In addition to DNA methylation, cfDNA fragmentation and/or nucleosome occupancy patterns are another epigenetic feature to trace gene activity and tissue origin [95]. Compaction of nucleosomal structures creates a barrier for DNA-binding transcription factors to access their cognate *cis*-regulatory elements. Usually, active promoters lack nucleosomes, while inactive promoters have densely packed nucleosomes. Nucleosome positioning through genome-wide mapping is shown to be associated with gene activation and expression in a development-dependent and tissue-specific manner [95,96]. Therefore, investigation of nucleosome positioning in a patient's cfDNA may reveal the existence of a specific cancer type.

As cfDNA is preferentially released from apoptotic cells, the size distribution of cfDNA fragments (160–180 bp) can resemble the size of mononucleosome-protected DNA. Specifically, peak sizes correspond to nucleosomes (~147 bp) and chromatosomes (nucleosome + linker histone; ~167 bp), suggesting they could bear the information of the cell type of origin [97]. Based on the expectation that fragment endpoints should cluster next to nucleosome boundaries and should be depleted at sites of nucleosome occupancy, Snyder et al. showed that nucleosome spacing patterns can inform the cell type of origin from cfDNA [98]. The study showed that nucleosome spacing inferred from cfDNA in healthy individuals correlated strongly with epigenetic features of lymphoid and myeloid cells, consistent with hematopoietic cell death as a major source of cfDNA, while the patterns of nucleosome spacing in late-stage cancer patients match the anatomical origin of the patient's cancer. Therefore, different nucleosome footprints between the tumor and the normal source of cfDNA may enable the noninvasive monitoring of a much broader set of clinical conditions than currently possible [98].

8. Conclusions and Future Direction

cfDNA molecules have emerged as promising biomarkers for cancer detection and monitoring due to the easy access to clinical samples from blood or urine. The advent of NGS technology provides an unprecedented opportunity to systematically examine the characteristics of cfDNA for tumor-specific changes. However, the massive amount of sequencing data requires sophisticated bioinformatics analysis to accurately identify genomic abnormalities in cancer. This review discussed major bioinformatics applications of cfDNA in oncological research to identify point mutations, copy number abnormalities, DNA methylation changes, and nucleosome positioning patterns. Using sophisticated bioinformatics analysis, advances have been made to better understand the property of cfDNA through fragmentation and nucleosome spacing patterns. Analysis by leveraging large-scale cancer genomic databases in conjunction with state-of-the-art statistical algorithms demonstrates the great potential of using methylation biomarkers for identification of cancer cell origin. Moreover, patterns of CNV through the WGS analysis can further reveal the extent of tumor heterogeneity. Nevertheless, to move cfDNA into routine clinical practices for better patient management, future studies will need to address several issues. First, studies need to focus more on detection sensitivity in early-stage cancer because there are many barriers to utilizing cfDNA for such applications. For example, most studies that demonstrated the feasibility of cfDNA in cancer detection used samples from late-stage cancer patients. However, the fraction of ctDNA in the plasma from early-stage cancer patients is generally very low. Although a range of NGS-based approaches have been used to characterize tumor genomes in detail and new bioinformatics techniques and analysis tools are rapidly evolving, current technologies and bioinformatics algorithms are not sensitive enough to detect such low level of genetic or epigenetic abnormalities. How to develop advanced technologies to detect mutations, CNVs, and epigenetic changes at the low ctDNA level is likely to be one of the most challenging issues to resolve. Another issue is related to cfDNA contaminations by the lysed blood cells and significant variation into cfDNA due to DNA isolation protocols and choice of instrument. Therefore, a standard protocol for quality control and bioinformatics analysis procedures need to be developed before these technologies can be successfully and reliably used in clinical practice and regulatory decision-making. A joint effort from the scientific community for the MicroArray Quality Control (MAQC) project [99] is an excellent example to follow to attain this goal. Finally, other biomarkers should be further explored for liquid biopsy in addition to genetic and epigenetic markers and nucleosome spacing patterns discussed in this review. For example, recent studies have shown that circulating cell-free RNA (cfRNA), which encompasses miRNAs, lncRNAs, and mRNAs, could also serve as valuable biomarkers for liquid biopsy [100,101]. Given the finding that transcriptome profiling alone from tissue biopsies can robustly determine cancerous status and tissue origin [102], the multiparameter analyses incorporating the molecular profiles at cfDNA, cfRNA, and protein will result in an improved understanding of molecular aberrations and their functional roles across tumor types, as well as facilitate the identification of novel tumor subtypes [103]. As most of the cfDNA interrogations to date are proof-of-principle studies, large-scale, multi-site cohort studies that systematically investigate all these aspects of molecular profiles are needed to evaluate the complementary nature of their screening power so that liquid biopsy signatures can be refined, validated, and utilized in clinical practice. Eventually, these efforts will lead to the identification of new oncological biomarkers for early detection and outcome prediction, which is a prerequisite for realizing the promise of precision medicine.

Author Contributions: Conceptualization, C.-C.H. and L.W.; Writing—original draft preparation, C.-C.H. and M.D.; Writing—review and editing, L.W. and C.-C.H.; Supervision, L.W.; funding acquisition, C.-C.H. and L.W.

Funding: This research was supported by National Institute of Health (R01CA212097) to L.W. and by National Institutes of Health (NIH) CTSA award (UL1TR001436) to C.-C.H.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gerlinger, M.; Rowan, A.J.; Horswell, S.; Math, M.; Larkin, J.; Endesfelder, D.; Gronroos, E.; Martinez, P.; Matthews, N.; Stewart, A.; et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **2012**, *366*, 883–892. [[CrossRef](#)] [[PubMed](#)]
2. Millner, L.M.; Linder, M.W.; Valdes, R., Jr. Circulating tumor cells: A review of present methods and the need to identify heterogeneous phenotypes. *Ann. Clin. Lab. Sci.* **2013**, *43*, 295–304.
3. Xia, Y.; Huang, C.C.; Dittmar, R.; Du, M.; Wang, Y.; Liu, H.; Shenoy, N.; Wang, L.; Kohli, M. Copy number variations in urine cell free DNA as biomarkers in advanced prostate cancer. *Oncotarget* **2016**, *7*, 35818–35831. [[CrossRef](#)] [[PubMed](#)]
4. Ilie, M.; Hofman, P. Pros: Can tissue biopsy be replaced by liquid biopsy? *Transl. Lung Cancer Res.* **2016**, *5*, 420–423. [[CrossRef](#)] [[PubMed](#)]
5. Gonzalez-Billalabeitia, E.; Conteduca, V.; Wetterskog, D.; Jayaram, A.; Attard, G. Circulating tumor DNA in advanced prostate cancer: Transitioning from discovery to a clinically implemented test. *Prostate Cancer Prostatic Dis.* **2019**, *22*, 195–205. [[CrossRef](#)] [[PubMed](#)]
6. Fleischhacker, M.; Schmidt, B. Circulating nucleic acids (CNAs) and cancer—A survey. *Biochim. Biophys. Acta* **2007**, *1775*, 181–232. [[CrossRef](#)] [[PubMed](#)]
7. Jahr, S.; Hentze, H.; Englisch, S.; Hardt, D.; Fackelmayer, F.O.; Hesch, R.D.; Knippers, R. DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **2001**, *61*, 1659–1665.
8. Alix-Panabieres, C.; Pantel, K. Challenges in circulating tumour cell research. *Nat. Rev. Cancer* **2014**, *14*, 623–631. [[CrossRef](#)] [[PubMed](#)]
9. Koffler, D.; Agnello, V.; Winchester, R.; Kunkel, H.G. The occurrence of single-stranded DNA in the serum of patients with systemic lupus erythematosus and other diseases. *J. Clin. Investig.* **1973**, *52*, 198–204. [[CrossRef](#)]
10. Abbosh, C.; Birkbak, N.J.; Wilson, G.A.; Jamal-Hanjani, M.; Constantin, T.; Salari, R.; Le Quesne, J.; Moore, D.A.; Veeriah, S.; Rosenthal, R.; et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **2017**, *545*, 446–451. [[CrossRef](#)]
11. Qin, Z.; Ljubimov, V.A.; Zhou, C.; Tong, Y.; Liang, J. Cell-free circulating tumor DNA in cancer. *Chin. J. Cancer* **2016**, *35*, 36. [[CrossRef](#)] [[PubMed](#)]
12. Tie, J.; Wang, Y.; Tomasetti, C.; Li, L.; Springer, S.; Kinde, I.; Silliman, N.; Tacey, M.; Wong, H.L.; Christie, M.; et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **2016**, *8*, 346ra92. [[CrossRef](#)] [[PubMed](#)]
13. Crowley, E.; Di Nicolantonio, F.; Loupakis, F.; Bardelli, A. Liquid biopsy: Monitoring cancer-genetics in the blood. *Nat. Rev. Clin. Oncol.* **2013**, *10*, 472–484. [[CrossRef](#)] [[PubMed](#)]
14. Diaz, L.A., Jr.; Bardelli, A. Liquid biopsies: Genotyping circulating tumor DNA. *J. Clin. Oncol.* **2014**, *32*, 579–586. [[CrossRef](#)] [[PubMed](#)]
15. Zehir, A.; Benayed, R.; Shah, R.H.; Syed, A.; Middha, S.; Kim, H.R.; Srinivasan, P.; Gao, J.; Chakravarty, D.; Devlin, S.M.; et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **2017**, *23*, 703–713. [[CrossRef](#)] [[PubMed](#)]
16. Saghafinia, S.; Mina, M.; Riggi, N.; Hanahan, D.; Ciriello, G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* **2018**, *25*, 1066–1080. [[CrossRef](#)] [[PubMed](#)]
17. Heitzer, E.; Auer, M.; Hoffmann, E.M.; Pichler, M.; Gasch, C.; Ulz, P.; Lax, S.; Waldispuehl-Geigl, J.; Mauermann, O.; Mohan, S.; et al. Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int. J. Cancer* **2013**, *133*, 346–356. [[CrossRef](#)] [[PubMed](#)]
18. Jung, K.; Fleischhacker, M.; Rabien, A. Cell-free DNA in the blood as a solid tumor biomarker—A critical appraisal of the literature. *Clin. Chim. Acta* **2010**, *411*, 1611–1624. [[CrossRef](#)] [[PubMed](#)]
19. Jiang, P.; Chan, C.W.; Chan, K.C.; Cheng, S.H.; Wong, J.; Wong, V.W.; Wong, G.L.; Chan, S.L.; Mok, T.S.; Chan, H.L.; et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1317–E1325. [[CrossRef](#)]
20. Jiang, P.; Lo, Y.M.D. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet.* **2016**, *32*, 360–371. [[CrossRef](#)]

21. Lo, Y.M.; Chan, K.C.; Sun, H.; Chen, E.Z.; Jiang, P.; Lun, F.M.; Zheng, Y.W.; Leung, T.Y.; Lau, T.K.; Cantor, C.R.; et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2010**, *2*, 61ra91. [[CrossRef](#)] [[PubMed](#)]
22. Underhill, H.R.; Kitzman, J.O.; Hellwig, S.; Welker, N.C.; Daza, R.; Baker, D.N.; Gligorich, K.M.; Rostomily, R.C.; Bronner, M.P.; Shendure, J. Fragment Length of Circulating Tumor DNA. *PLoS Genet.* **2016**, *12*, e1006162. [[CrossRef](#)] [[PubMed](#)]
23. Volik, S.; Alcaide, M.; Morin, R.D.; Collins, C. Cell-free DNA (cfDNA): Clinical Significance and Utility in Cancer Shaped by Emerging Technologies. *Mol. Cancer Res.* **2016**, *14*, 898–908. [[CrossRef](#)] [[PubMed](#)]
24. Wood-Bouwens, C.; Lau, B.T.; Handy, C.M.; Lee, H.; Ji, H.P. Single-Color Digital PCR Provides High-Performance Detection of Cancer Mutations from Circulating DNA. *J. Mol. Diagn.* **2017**, *19*, 697–710. [[CrossRef](#)] [[PubMed](#)]
25. Allegra, C.J.; Jessup, J.M.; Somerfield, M.R.; Hamilton, S.R.; Hammond, E.H.; Hayes, D.F.; McAllister, P.K.; Morton, R.F.; Schilsky, R.L. American Society of Clinical Oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J. Clin. Oncol.* **2009**, *27*, 2091–2096. [[CrossRef](#)]
26. Shaw, A.T.; Engelman, J.A. ALK in lung cancer: Past, present, and future. *J. Clin. Oncol.* **2013**, *31*, 1105–1111. [[CrossRef](#)] [[PubMed](#)]
27. Gonzalez, D.; Fearfield, L.; Nathan, P.; Taniere, P.; Wallace, A.; Brown, E.; Harwood, C.; Marsden, J.; Whittaker, S. BRAF mutation testing algorithm for vemurafenib treatment in melanoma: Recommendations from an expert panel. *Br. J. Derm.* **2013**, *168*, 700–707. [[CrossRef](#)]
28. Marchetti, A.; Palma, J.F.; Felicioni, L.; De Pas, T.M.; Chiari, R.; Del Grammastro, M.; Filice, G.; Ludovini, V.; Brandes, A.A.; Chella, A.; et al. Early Prediction of Response to Tyrosine Kinase Inhibitors by Quantification of EGFR Mutations in Plasma of NSCLC Patients. *J. Thorac. Oncol.* **2015**, *10*, 1437–1443. [[CrossRef](#)] [[PubMed](#)]
29. Simon, R.; Roychowdhury, S. Implementing personalized cancer genomics in clinical trials. *Nat. Rev. Drug Discov.* **2013**, *12*, 358–369. [[CrossRef](#)]
30. Gevensleben, H.; Garcia-Murillas, I.; Graeser, M.K.; Schiavon, G.; Osin, P.; Parton, M.; Smith, I.E.; Ashworth, A.; Turner, N.C. Noninvasive detection of HER2 amplification with plasma DNA digital PCR. *Clin. Cancer Res.* **2013**, *19*, 3276–3284. [[CrossRef](#)]
31. Sacher, A.G.; Paweletz, C.; Dahlberg, S.E.; Alden, R.S.; O’Connell, A.; Feeney, N.; Mach, S.L.; Janne, P.A.; Oxnard, G.R. Prospective Validation of Rapid Plasma Genotyping for the Detection of EGFR and KRAS Mutations in Advanced Lung Cancer. *JAMA Oncol.* **2016**, *2*, 1014–1022. [[CrossRef](#)]
32. Leighl, N.B.; Rekhtman, N.; Biermann, W.A.; Huang, J.; Mino-Kenudson, M.; Ramalingam, S.S.; West, H.; Whitlock, S.; Somerfield, M.R. Molecular testing for selection of patients with lung cancer for epidermal growth factor receptor and anaplastic lymphoma kinase tyrosine kinase inhibitors: American Society of Clinical Oncology endorsement of the College of American Pathologists/International Association for the study of lung cancer/association for molecular pathology guideline. *J. Clin. Oncol.* **2014**, *32*, 3673–3679. [[PubMed](#)]
33. Warren, J.D.; Xiong, W.; Bunker, A.M.; Vaughn, C.P.; Furtado, L.V.; Roberts, W.L.; Fang, J.C.; Samowitz, W.S.; Heichman, K.A. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med.* **2011**, *9*, 133. [[CrossRef](#)] [[PubMed](#)]
34. Reid, A.L.; Freeman, J.B.; Millward, M.; Ziman, M.; Gray, E.S. Detection of BRAF-V600E and V600K in melanoma circulating tumour cells by droplet digital PCR. *Clin. Biochem.* **2015**, *48*, 999–1002. [[CrossRef](#)] [[PubMed](#)]
35. Ma, Y.; Luk, A.; Young, F.P.; Lynch, D.; Chua, W.; Balakrishnar, B.; de Souza, P.; Becker, T.M. Droplet Digital PCR Based Androgen Receptor Variant 7 (AR-V7) Detection from Prostate Cancer Patient Blood Biopsies. *Int. J. Mol. Sci.* **2016**, *17*, 1264. [[CrossRef](#)] [[PubMed](#)]
36. Kivioja, T.; Vaharautio, A.; Karlsson, K.; Bonke, M.; Enge, M.; Linnarsson, S.; Taipale, J. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **2011**, *9*, 72–74. [[CrossRef](#)] [[PubMed](#)]
37. Martincorena, I.; Roshan, A.; Gerstung, M.; Ellis, P.; Van Loo, P.; McLaren, S.; Wedge, D.C.; Fullam, A.; Alexandrov, L.B.; Tubio, J.M.; et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **2015**, *348*, 880–886. [[CrossRef](#)] [[PubMed](#)]

38. Liu, J.; Chen, X.; Wang, J.; Zhou, S.; Wang, C.L.; Ye, M.Z.; Wang, X.Y.; Song, Y.; Wang, Y.Q.; Zhang, L.T.; et al. Biological background of the genomic variations of cf-DNA in healthy individuals. *Ann. Oncol.* **2019**, *30*, 464–470. [[CrossRef](#)] [[PubMed](#)]
39. Bauml, J.; Levy, B. Clonal Hematopoiesis: A New Layer in the Liquid Biopsy Story in Lung Cancer. *Clin. Cancer Res.* **2018**, *24*, 4352–4354. [[CrossRef](#)] [[PubMed](#)]
40. Chin, R.L.; Chen, K.; Usmani, A.; Chua, C.; Harris, P.K.; Binkley, M.S.; Azad, T.D.; Dudley, J.C.; Chaudhuri, A.A. Detection of Solid Tumor Molecular Residual Disease (MRD) Using Circulating Tumor DNA (ctDNA). *Mol. Diagn. Ther.* **2019**, *23*, 311–331. [[CrossRef](#)] [[PubMed](#)]
41. Frampton, G.M.; Fichtenholtz, A.; Otto, G.A.; Wang, K.; Downing, S.R.; He, J.; Schnall-Levin, M.; White, J.; Sanford, E.M.; An, P.; et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **2013**, *31*, 1023–1031. [[CrossRef](#)] [[PubMed](#)]
42. Newman, A.M.; Lovejoy, A.F.; Klass, D.M.; Kurtz, D.M.; Chabon, J.J.; Scherer, F.; Stehr, H.; Liu, C.L.; Bratman, S.V.; Say, C.; et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **2016**, *34*, 547–555. [[CrossRef](#)] [[PubMed](#)]
43. Wan, J.C.M.; Massie, C.; Garcia-Corbacho, J.; Mouliere, F.; Brenton, J.D.; Caldas, C.; Pacey, S.; Baird, R.; Rosenfeld, N. Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **2017**, *17*, 223–238. [[CrossRef](#)] [[PubMed](#)]
44. Kennedy, S.R.; Schmitt, M.W.; Fox, E.J.; Kohrn, B.F.; Salk, J.J.; Ahn, E.H.; Prindle, M.J.; Kuong, K.J.; Shen, J.C.; Risques, R.A.; et al. Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat. Protoc.* **2014**, *9*, 2586–2606. [[CrossRef](#)] [[PubMed](#)]
45. Schmitt, M.W.; Fox, E.J.; Prindle, M.J.; Reid-Bayliss, K.S.; True, L.D.; Radich, J.P.; Loeb, L.A. Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat. Methods* **2015**, *12*, 423–425. [[CrossRef](#)] [[PubMed](#)]
46. Chung, J.; Lee, K.W.; Lee, C.; Shin, S.H.; Kyung, S.; Jeon, H.J.; Kim, S.Y.; Cho, E.; Yoo, C.E.; Son, D.S.; et al. Performance evaluation of commercial library construction kits for PCR-based targeted sequencing using a unique molecular identifier. *BMC Genom.* **2019**, *20*, 216. [[CrossRef](#)] [[PubMed](#)]
47. Teder, H.; Koel, M.; Paluoja, P.; Jatsenko, T.; Rekker, K.; Laisk-Podar, T.; Kukuskina, V.; Velthut-Meikas, A.; Fjodorova, O.; Peters, M.; et al. TAC-seq: Targeted DNA and RNA sequencing for precise biomarker molecule counting. *NPJ Genom. Med.* **2018**, *3*, 34. [[CrossRef](#)]
48. Phallen, J.; Sausen, M.; Adleff, V.; Leal, A.; Hruban, C.; White, J.; Anagnostou, V.; Fiksel, J.; Cristiano, S.; Papp, E.; et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **2017**, *9*, eaan2415. [[CrossRef](#)]
49. Smith, T.; Heger, A.; Sudbery, I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **2017**, *27*, 491–499. [[CrossRef](#)]
50. Shugay, M.; Zaretsky, A.R.; Shagin, D.A.; Shagina, I.A.; Volchenkov, I.A.; Shelenkov, A.A.; Lebedin, M.Y.; Bagaev, D.V.; Lukyanov, S.; Chudakov, D.M. MAGERI: Computational pipeline for molecular-barcoded targeted resequencing. *PLoS Comput. Biol.* **2017**, *13*, e1005480. [[CrossRef](#)]
51. Scheinin, I.; Sie, D.; Bengtsson, H.; van de Wiel, M.A.; Olshen, A.B.; van Thuijl, H.F.; van Essen, H.F.; Eijk, P.P.; Rustenburg, F.; Meijer, G.A.; et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **2014**, *24*, 2022–2032. [[CrossRef](#)] [[PubMed](#)]
52. Raman, L.; Dheedene, A.; De Smet, M.; Van Dorpe, J.; Menten, B. WisecondorX: Improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Res.* **2019**, *47*, 1605–1614. [[CrossRef](#)] [[PubMed](#)]
53. Xi, R.; Lee, S.; Xia, Y.; Kim, T.M.; Park, P.J. Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants. *Nucleic Acids Res.* **2016**, *44*, 6274–6286. [[CrossRef](#)] [[PubMed](#)]
54. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2016**, *12*, e1004873. [[CrossRef](#)] [[PubMed](#)]
55. Kang, S.; Li, Q.; Chen, Q.; Zhou, Y.; Park, S.; Lee, G.; Grimes, B.; Krysan, K.; Yu, M.; Wang, W.; et al. CancerLocator: Non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* **2017**, *18*, 53. [[CrossRef](#)]

56. Li, W.; Li, Q.; Kang, S.; Same, M.; Zhou, Y.; Sun, C.; Liu, C.C.; Matsuoka, L.; Sher, L.; Wong, W.H.; et al. CancerDetector: Ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res.* **2018**, *46*, e89. [[CrossRef](#)] [[PubMed](#)]
57. Schmitt, M.W.; Kennedy, S.R.; Salk, J.J.; Fox, E.J.; Hiatt, J.B.; Loeb, L.A. Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14508–14513. [[CrossRef](#)]
58. Kamps-Hughes, N.; McUsic, A.; Kurihara, L.; Harkins, T.T.; Pal, P.; Ray, C.; Ionescu-Zanetti, C. ERASE-Seq: Leveraging replicate measurements to enhance ultralow frequency variant detection in NGS data. *PLoS ONE* **2018**, *13*, e0195272. [[CrossRef](#)]
59. Yeom, H.; Lee, Y.; Ryu, T.; Noh, J.; Lee, A.C.; Lee, H.B.; Kang, E.; Song, S.W.; Kwon, S. Barcode-free next-generation sequencing error validation for ultra-rare variant detection. *Nat. Commun.* **2019**, *10*, 977. [[CrossRef](#)]
60. Andor, N.; Maley, C.C.; Ji, H.P. Genomic Instability in Cancer: Teetering on the Limit of Tolerance. *Cancer Res.* **2017**, *77*, 2179–2185. [[CrossRef](#)]
61. Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: The next generation. *Cell* **2011**, *144*, 646–674. [[CrossRef](#)]
62. Beroukhi, R.; Mermel, C.H.; Porter, D.; Wei, G.; Raychaudhuri, S.; Donovan, J.; Barretina, J.; Boehm, J.S.; Dobson, J.; Urashima, M.; et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **2010**, *463*, 899–905. [[CrossRef](#)] [[PubMed](#)]
63. Zack, T.I.; Schumacher, S.E.; Carter, S.L.; Cherniack, A.D.; Saksena, G.; Tabak, B.; Lawrence, M.S.; Zhsng, C.Z.; Wala, J.; Mermel, C.H.; et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **2013**, *45*, 1134–1140. [[CrossRef](#)] [[PubMed](#)]
64. Heitzer, E.; Ulz, P.; Belic, J.; Gutsch, S.; Quehenberger, F.; Fischereder, K.; Benezeder, T.; Auer, M.; Pischler, C.; Mannweiler, S.; et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **2013**, *5*, 30. [[CrossRef](#)]
65. Dawson, S.J.; Tsui, D.W.; Murtaza, M.; Biggs, H.; Rueda, O.M.; Chin, S.F.; Dunning, M.J.; Gale, D.; Forshew, T.; Mahler-Araujo, B.; et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **2013**, *368*, 1199–1209. [[CrossRef](#)] [[PubMed](#)]
66. Leary, R.J.; Sausen, M.; Kinde, I.; Papadopoulos, N.; Carpten, J.D.; Craig, D.; O’Shaughnessy, J.; Kinzler, K.W.; Parmigiani, G.; Vogelstein, B.; et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci. Transl. Med.* **2012**, *4*, 162ra154. [[CrossRef](#)]
67. Pirooznia, M.; Goes, F.S.; Zandi, P.P. Whole-genome CNV analysis: Advances in computational approaches. *Front. Genet.* **2015**, *6*, 138. [[CrossRef](#)]
68. Derrien, T.; Estelle, J.; Marco Sola, S.; Knowles, D.G.; Raineri, E.; Guigo, R.; Ribeca, P. Fast computation and applications of genome mappability. *PLoS ONE* **2012**, *7*, e30377. [[CrossRef](#)] [[PubMed](#)]
69. Kundaje, A. A Comprehensive Collection of Signal Artifact Blacklist Regions in the Human Genome. Available online: <https://personal.broadinstitute.org/anshul/projects/encode/rawdata/blacklists/hg19-blacklist-README.pdf> (accessed on 3 June 2019).
70. Xia, S.; Huang, C.C.; Le, M.; Dittmar, R.; Du, M.; Yuan, T.; Guo, Y.; Wang, Y.; Wang, X.; Tsai, S.; et al. Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. *Lung Cancer* **2015**, *90*, 78–84. [[CrossRef](#)]
71. Hovelson, D.H.; Liu, C.J.; Wang, Y.; Kang, Q.; Henderson, J.; Gursky, A.; Brockman, S.; Ramnath, N.; Krauss, J.C.; Talpaz, M.; et al. Rapid, ultra low coverage copy number profiling of cell-free DNA as a precision oncology screening strategy. *Oncotarget* **2017**, *8*, 89848–89866. [[CrossRef](#)]
72. Benjamini, Y.; Speed, T.P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **2012**, *40*, e72. [[CrossRef](#)] [[PubMed](#)]
73. Olshen, A.B.; Venkatraman, E.S.; Lucito, R.; Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **2004**, *5*, 557–572. [[CrossRef](#)] [[PubMed](#)]
74. Venkatraman, E.S.; Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **2007**, *23*, 657–663. [[CrossRef](#)] [[PubMed](#)]
75. Shah, S.P.; Xuan, X.; DeLeeuw, R.J.; Khojasteh, M.; Lam, W.L.; Ng, R.; Murphy, K.P. Integrating copy number polymorphisms into array CGH analysis using a robust HMM. *Bioinformatics* **2006**, *22*, e431–e439. [[CrossRef](#)] [[PubMed](#)]

76. Lai, D.; Ha, G.; Shah, S. HMMcopy: Copy number prediction with correction for GC and mappability bias for HTS data. R Package Version 1.26.0. 2019. Available online: <http://bioconductor.org/packages/release/bioc/html/HMMcopy.html> (accessed on 3 June 2019).
77. Liu, B.; Morrison, C.D.; Johnson, C.S.; Trump, D.L.; Qin, M.; Conroy, J.C.; Wang, J.; Liu, S. Computational methods for detecting copy number variations in cancer genome using next generation sequencing: Principles and challenges. *Oncotarget* **2013**, *4*, 1868–1881. [[CrossRef](#)] [[PubMed](#)]
78. Eckel-Passow, J.E.; Atkinson, E.J.; Maharjan, S.; Kardia, S.L.; de Andrade, M. Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. *BMC Bioinform.* **2011**, *12*, 220. [[CrossRef](#)] [[PubMed](#)]
79. Zhang, X.; Du, R.; Li, S.; Zhang, F.; Jin, L.; Wang, H. Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinform.* **2014**, *15*, 50. [[CrossRef](#)] [[PubMed](#)]
80. Mohan, S.; Heitzer, E.; Ulz, P.; Lafer, I.; Lax, S.; Auer, M.; Pichler, M.; Gerger, A.; Eisner, F.; Hoefler, G.; et al. Changes in colorectal carcinoma genomes under anti-EGFR therapy identified by whole-genome plasma DNA sequencing. *PLoS Genet.* **2014**, *10*, e1004271. [[CrossRef](#)]
81. Xu, H.; Zhu, X.; Xu, Z.; Hu, Y.; Bo, S.; Xing, T.; Zhu, K. Non-invasive Analysis of Genomic Copy Number Variation in Patients with Hepatocellular Carcinoma by Next Generation DNA Sequencing. *J. Cancer* **2015**, *6*, 247–253. [[CrossRef](#)] [[PubMed](#)]
82. Ulz, P.; Belic, J.; Graf, R.; Auer, M.; Lafer, I.; Fischereder, K.; Webersinke, G.; Pummer, K.; Augustin, H.; Pichler, M.; et al. Whole-genome plasma sequencing reveals focal amplifications as a driving force in metastatic prostate cancer. *Nat. Commun.* **2016**, *7*, 12008. [[CrossRef](#)]
83. Van de Wiel, M.A.; Kim, K.I.; Vosse, S.J.; van Wieringen, W.N.; Wilting, S.M.; Ylstra, B. CGHcall: Calling aberrations for array CGH tumor profiles. *Bioinformatics* **2007**, *23*, 892–894. [[CrossRef](#)] [[PubMed](#)]
84. Engler, D.A.; Mohapatra, G.; Louis, D.N.; Betensky, R.A. A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics* **2006**, *7*, 399–421. [[CrossRef](#)] [[PubMed](#)]
85. Kirkizlar, E.; Zimmermann, B.; Constantin, T.; Swenerton, R.; Hoang, B.; Wayham, N.; Babiarz, J.E.; Demko, Z.; Pelham, R.J.; Kareht, S.; et al. Detection of Clonal and Subclonal Copy-Number Variants in Cell-Free DNA from Patients with Breast Cancer Using a Massively Multiplexed PCR Methodology. *Transl. Oncol.* **2015**, *8*, 407–416. [[CrossRef](#)] [[PubMed](#)]
86. Baylin, S.B.; Herman, J.G. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet.* **2000**, *16*, 168–174. [[CrossRef](#)]
87. Jones, P.A.; Laird, P.W. Cancer epigenetics comes of age. *Nat. Genet.* **1999**, *21*, 163–167. [[CrossRef](#)] [[PubMed](#)]
88. Chen, Y.; Breeze, C.E.; Zhen, S.; Beck, S.; Teschendorff, A.E. Tissue-independent and tissue-specific patterns of DNA methylation alteration in cancer. *Epigenet. Chromatin* **2016**, *9*, 10. [[CrossRef](#)] [[PubMed](#)]
89. Zhang, B.; Zhou, Y.; Lin, N.; Lowdon, R.F.; Hong, C.; Nagarajan, R.P.; Cheng, J.B.; Li, D.; Stevens, M.; Lee, H.J.; et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res.* **2013**, *23*, 1522–1540. [[PubMed](#)]
90. Lehmann-Werman, R.; Neiman, D.; Zemmour, H.; Moss, J.; Magenheimer, J.; Vaknin-Dembinsky, A.; Rubertsson, S.; Nellgard, B.; Blennow, K.; Zetterberg, H.; et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1826–E1834. [[CrossRef](#)]
91. Moss, J.; Magenheimer, J.; Neiman, D.; Zemmour, H.; Loyfer, N.; Korach, A.; Samet, Y.; Maoz, M.; Druid, H.; Arner, P.; et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **2018**, *9*, 5068. [[CrossRef](#)] [[PubMed](#)]
92. Teschendorff, A.E.; Breeze, C.E.; Zheng, S.C.; Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinform.* **2017**, *18*, 105. [[CrossRef](#)]
93. Sun, K.; Jiang, P.; Chan, K.C.; Wong, J.; Cheng, Y.K.; Liang, R.H.; Chan, W.K.; Ma, E.S.; Chan, S.L.; Cheng, S.H.; et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5503–E5512. [[CrossRef](#)] [[PubMed](#)]
94. Guo, S.; Diep, D.; Plongthongkum, N.; Fung, H.L.; Zhang, K.; Zhang, K. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **2017**, *49*, 635–642. [[CrossRef](#)] [[PubMed](#)]

95. Kelly, T.K.; Liu, Y.; Lay, F.D.; Liang, G.; Berman, B.P.; Jones, P.A. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **2012**, *22*, 2497–2506. [[CrossRef](#)] [[PubMed](#)]
96. Ye, Z.; Chen, Z.; Sunkel, B.; Fietze, S.; Huang, T.H.; Wang, Q.; Jin, V.X. Genome-wide analysis reveals positional-nucleosome-oriented binding pattern of pioneer factor FOXA1. *Nucleic Acids Res.* **2016**, *44*, 7540–7554. [[CrossRef](#)] [[PubMed](#)]
97. Fan, H.C.; Blumenfeld, Y.J.; Chitkara, U.; Hudgins, L.; Quake, S.R. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16266–16271. [[CrossRef](#)] [[PubMed](#)]
98. Snyder, M.W.; Kircher, M.; Hill, A.J.; Daza, R.M.; Shendure, J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **2016**, *164*, 57–68. [[CrossRef](#)] [[PubMed](#)]
99. Consortium, M.; Shi, L.; Reid, L.H.; Jones, W.D.; Shippy, R.; Warrington, J.A.; Baker, S.C.; Collins, P.J.; de Longueville, F.; Kawasaki, E.S.; et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* **2006**, *24*, 1151–1161. [[CrossRef](#)] [[PubMed](#)]
100. Heitzer, E.; Perakis, S.; Geigl, J.B.; Speicher, M.R. The potential of liquid biopsies for the early detection of cancer. *NPJ Precis. Oncol.* **2017**, *1*, 36. [[CrossRef](#)]
101. Koh, W.; Pan, W.; Gawad, C.; Fan, H.C.; Kerchner, G.A.; Wyss-Coray, T.; Blumenfeld, Y.J.; El-Sayed, Y.Y.; Quake, S.R. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7361–7366. [[CrossRef](#)] [[PubMed](#)]
102. Sun, K.; Wang, J.; Wang, H.; Sun, H. GeneCT: A generalizable cancerous status and tissue origin classifier for pan-cancer biopsies. *Bioinformatics* **2018**, *34*, 4129–4130. [[CrossRef](#)] [[PubMed](#)]
103. Hodara, E.; Morrison, G.; Cunha, A.; Zainfeld, D.; Xu, T.; Xu, Y.; Dempsey, P.W.; Pagano, P.C.; Bischoff, F.; Khurana, A.; et al. Multiparametric liquid biopsy analysis in metastatic prostate cancer. *JCI Insight* **2019**, *4*. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).