



# Utility of cfDNA Fragmentation Patterns in Designing the Liquid Biopsy Profiling Panels to Improve Their Sensitivity

Maxim Ivanov<sup>1\*</sup>, Polina Chernenko<sup>2</sup>, Valery Breder<sup>2</sup>, Konstantin Laktionov<sup>2</sup>, Ekaterina Rozhavskaia<sup>3,4</sup>, Sergey Musienko<sup>3</sup>, Ancha Baranova<sup>1,3,5,6</sup> and Vladislav Mileyko<sup>3</sup>

<sup>1</sup> Department of Biological and Medical Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Russia, <sup>2</sup> N.N. Blokhin Russian Cancer Research Center, Moscow, Russia, <sup>3</sup> Atlas Oncology Diagnostics, Ltd., Moscow, Russia, <sup>4</sup> Vavilov Institute of General Genetics, Moscow, Russia, <sup>5</sup> Research Centre for Medical Genetics, Moscow, Russia, <sup>6</sup> School of Systems Biology, George Mason University, Fairfax, VA, United States

## OPEN ACCESS

### Edited by:

Richard D. Emes,  
University of Nottingham,  
United Kingdom

### Reviewed by:

Vladimir B. Teif,  
University of Essex, United Kingdom  
Kuo-Ping Chiu,  
Academia Sinica, Taiwan  
Tatiana V. Tatarinova,  
University of La Verne, United States

### \*Correspondence:

Maxim Ivanov  
maksim.v.ivanov@phystech.edu

### Specialty section:

This article was submitted to  
Bioinformatics and Computational  
Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 19 October 2018

**Accepted:** 25 February 2019

**Published:** 12 March 2019

### Citation:

Ivanov M, Chernenko P, Breder V, Laktionov K, Rozhavskaia E, Musienko S, Baranova A and Mileyko V (2019) Utility of cfDNA Fragmentation Patterns in Designing the Liquid Biopsy Profiling Panels to Improve Their Sensitivity. *Front. Genet.* 10:194.  
doi: 10.3389/fgene.2019.00194

Genotyping of cell-free DNA (cfDNA) in plasma samples has the potential to allow for a noninvasive assessment of tumor biology, avoiding the inherent shortcomings of tissue biopsy. Next generation sequencing (NGS), a leading technology for liquid biopsy analysis, continues to be hurdled with several major issues with cfDNA samples, including low cfDNA concentration and high fragmentation. In this study, by employing Ion Torrent PGM semiconductor technology, we performed a comparison between two multi-biomarker amplicon-based NGS panels characterized by a substantial difference in average amplicon length. In course of the analysis of the peripheral blood from 13 diagnostic non-small cell lung cancer patients, equivalence of two panels, in terms of overall diagnostic sensitivity and specificity was shown. A pairwise comparison of the allele frequencies for the same somatic variants obtained from the pairs of panel-specific amplicons, demonstrated an identical analytical sensitivity in range of 140 to 170 bp amplicons in size. Further regression analysis between amplicon length and its coverage, illustrated that NGS sequencing of plasma cfDNA equally tolerates amplicons with lengths in the range of 120 to 170 bp. To increase the sensitivity of mutation detection in cfDNA, we performed a computational analysis of the features associated with genome-wide nucleosome maps, evident from the data on the prevalence of cfDNA fragments of certain sizes and their fragmentation patterns. By leveraging the support vector machine-based machine learning approach, we showed that a combination of nucleosome map associated features with GC content, results in the increased accuracy of prediction of high inter-sample sequencing coverage variation (areas under the receiver operating curve: 0.75, 95% CI: 0.750–0.752 vs. 0.65, 95% CI: 0.63–0.67). Thus, nucleosome-guided fragmentation should be utilized as a guide to design amplicon-based NGS panels for the genotyping of cfDNA samples.

**Keywords:** NGS, cfDNA, liquid biopsy, cancer, DNA fragmentation, nucleosome, amplicon, primer design

## INTRODUCTION

In an approach known as “liquid biopsy,” cell-free DNA (cfDNA) which circulates in the plasma may be used for a diagnostic detection of tumor-specific mutations (Dawson et al., 2013; Pupilli et al., 2013; Xi et al., 2016). In the frame of the Lab-Developed Tests (LDT) paradigm, analysis of cfDNA has already gained approval for a number of common indications, including the detection of the resistance mutation T790M in the EGFR encoding gene (Malapelle et al., 2016), which commonly emerges in lung adenocarcinomas treated with tyrosine kinase inhibitors.

At their inception, cfDNA-based LDTs commonly exploited one or another conventional DNA analysis technique, including real-time PCR, droplet digital PCR and beads, emulsions, amplification, and magnetics (BEAM)ing digital PCR (Dawson et al., 2013; Oxnard et al., 2014; Siravegna et al., 2015; Thress et al., 2015; Sacher et al., 2016). Many studies showed that the concordance of liquid biopsy and tissue-based analysis is relatively high; nevertheless, these approaches are not free of limitations. Typically, PCR-based and hybridization-based cfDNA profiling techniques are developed to detect particular DNA variants, which most commonly underlie one or another previously described pathophysiological process. These and other variant-specific techniques are not suitable for the exploratory analysis of cfDNA, which is necessary for acquisition of knowledge concerning non-conventional, emerging resistance pathways, for co-detection of the mismatch repair phenotype, and for off-label prescribing of anticancer medications commonly required for personalized treatment of metastatic tumors (Tafe et al., 2015; Wei et al., 2016; Zehir et al., 2017). These limitations are readily surmounted by an advent of sequencing-based technologies, including whole exome sequencing or, more applicable to cfDNA analysis, amplicon-based panels, which are limited to their target genes, but are still exploration-permissive.

With reported sensitivity and a specificity of more than 80%, and 98 to 100%, respectively (Krishnamurthy et al., 2017), a next generation sequencing (NGS) analysis of cfDNA has already inserted itself into the ranks of the commonly used LDTs. Nevertheless, further improvement of the sensitivity in liquid biopsy-based tests is warranted. The most common way to improve sensitivity of the mutation detection in liquid biopsy samples, is to increase the coverage, which in turn leads to a substantial increase in the cost of an assay. Deep or ultradeep coverage is necessary in order to account for low concentrations of total cfDNA in plasma samples that are compounded by the dilution of tumor-specific cfDNA fragments, by substantial amounts of non-tumoral cfDNA fragments (Hellwig et al., 2018).

Another physical characteristic of cfDNA, the distribution of the sizes of its fragments, is relevant to the detection of DNA variants both by sequencing and by PCR. Recent whole-genome sequencing (WGS) studies of cfDNA demonstrated that the distribution of the sizes of plasma derived DNA fragments is far from the typical lognormal distribution that reflects the patterning of DNA in formalin fixed-paraffin-embedded samples or snap-frozen tissues. In fact, cfDNA exhibits a predominant peak at a fragment length of ~167 bp accompanied by the

second, significantly less pronounced extremum at around 350 bp (Ma et al., 2017). These observations mean that the majority of these fragments are suitable to assess the technique that relies on conventional lengths of PCR amplicons. It is of note that tumor-derived cfDNA fragments are even shorter than those that originate from healthy cells of the same origin (Jiang et al., 2015). In the domain of conventional systems for the detection of DNA variants, these characteristic of cfDNA have prompted the development of ultra-short amplicon PCR, which allows for the substantial increase of analytical and, as a consequence, diagnostic sensitivity of these assays.

Moreover, recent studies have shown that fragmentation pattern of cfDNA is not random. As cfDNA degradation is guided by nucleosome patterns defined by epigenetic regulation within particular loci (Ivanov et al., 2015), recurrent underrepresentation of some regions in cfDNA introduces systematic bias in the PCR based enrichment of target amplicons and undermine the sensitivity at a local scale.

In this study, we investigated the effect of the amplicon length on the diagnostic and analytical sensitivity of mutation detection, using two amplicon-based NGS panels with diverse amplicon lengths. We also describe ways to utilize the knowledge of cfDNA fragmentation patterns to increase the sensitivity of mutation detection in a liquid biopsy setting.

## MATERIALS AND METHODS

### Sample Collection

The sequencing was performed on cfDNA fragments extracted from previously collected plasma samples of 13 non-small cell lung cancer (NSCLC) patients, treated at the Blokhin Russian Cancer Research Centre in 2014 to 2015. For each patient, tumor tissue-based EGFR mutation status was assessed using the thetascreen EGFR RGQ PCR Kit (Qiagen, Milan, Italy) according to the manufacturers protocol.

For nucleosome-guided cfDNA fragmentation pattern analysis we used publicly available, anonymized WGS data of cfDNA, described by Snyder et al. (2016) and included in dataset [PRJNA291063].

The present study was approved by the Atlas Biomed Internal Review Board. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### DNA Extraction and Sample Quality Control

For each NSCLC patient, a peripheral blood sample was collected into an EDTA-containing vacutainer tube (BD). Samples were fractionated into plasma and blood cells by centrifugation at 400 g for 15 min within 4 h after venipuncture, followed by a secondary spin at 1200 g for 20 min. Resultant plasma samples were frozen in aliquots and stored at  $-80^{\circ}\text{C}$  until DNA isolation. Circulating DNA was extracted from 4 ml of plasma using the Blood Plasma DNA Isolation Kit (BioSilica Ltd., Russia) according to the manufacturer's instructions, eluted by 120  $\mu\text{l}$  of nuclease-free water, mixed with 3  $\mu\text{l}$  of glycogen (20 mg/ml, Fermentas, Lithuania), 1/10 volume of 50 mM triethylamine

and then precipitated with 5 volumes of acetone (Bryzgunova et al., 2011). After reconstitution in 30–50  $\mu$ l of water, cfDNA concentrations were measured using the Qubit fluorometer.

## Library Preparation and Quality Control

Sequencing libraries were prepared according to the manufacturer's protocol for Ion AmpliSeq Cancer Hotspot Panel (ITCHP2), designed to amplify 207 target regions across 50 cancer-related genes. Additionally, a custom panel namely Atlas Clinical Panel (AODCP), was designed to cover the following genes: EGFR, IDH2, NRAS, KIT, BRAF, TP53, PDGFRA, PTEN, IDH1, KRAS, PIK3CA, ERBB2, CTNNB1 (AODCP, 55 target regions). The custom panel was designed using the Ion AmpliSeq Designer server (pipeline version 5.2). The two panels had several loci in common, allowing for their comparison.

## Sequencing and Data Analysis

Pooled libraries were sequenced employing Ion Torrent PGM, according to the manufacturer's protocol. As low frequency mutant alleles were expected, initial analysis was performed using Ion Torrent Suite software (version 5.2.0) on low stringency settings. In order to exclude false negative single nucleotide variant (SNV) calls, concomitant Bowtie2-Strelka pipeline analysis was carried out. After aligning all reads to the genome (GRCh37) (Bowtie2 parameters: `-rdg 5,2 -rfg 5,2 -N 1 -L 17`), further off-target reads were removed, while the remaining reads were realigned on target sequences. Primer sequences were excluded from reads employing in-house software (Ivanov et al., 2018). Somatic variant calling was performed employing Strelka (maxInputDepth set to `-1`; indelMaxRefRepeat set to `6`; indelMaxWindowFilteredBasecallFrac set to `0.4`; indelMaxIntHpoLength set to `6`; lower quality bound for SNV and indels set to `9` and `2`, respectively). Variants supported with less than 20 reads in total were discarded. If less than four reads supported alternative allele, the variant was omitted. Mutation hotspots were defined as nucleotide variations identified in ten or more COSMIC (Forbes et al., 2010) samples. Detected variants located within mutation hotspots were supposed to be confidently somatic. Variants outside mutation hotspots with minor allele frequency in the general population, as defined by 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015), of 5% and more were supposed to be confidently germline. Further analysis was limited to confidently somatic and confidently germline variants. Preprocessed fastq files were additionally screened for mutation hotspots by inputting wild type and expected mutant reads into the Poisson distribution statistical model with complexity-dependent variable expectation probability of SNVs and indels. Somatic variant calls were verified manually, in the Tablet (version 1.16.09.06) read alignment visualization tool (Milne et al., 2010). Variant allele frequencies were quantified within raw read sets as a ratio of reads confirming the mutation to the total count of qualified reads covering the mutation site. Normalization of mutation allele frequencies to amplicon coverage was performed by bootstrapping. The genome variation analysis was limited to the nucleotide changes affecting the protein sequence, unless otherwise

specified. Publicly available software and database versions used were Bowtie2 v. 2.1.0 (Langmead and Salzberg, 2012), Strelka v. 1.0.14 (Saunders et al., 2012), and SAMtools v. 0.1.19 (Li, 2011). COSMIC and dbSNP databases were assessed in December 2017.

GC content normalization for linear regression analysis was performed leveraging a simple adjustment according to the equation  $\tilde{r}_i = r_i^m / m_{GC}$ , where  $r_i$  stands for the read count of the  $i$ th amplicon,  $m_{GC}$  is the median read count of all windows with the same GC content as the  $i$ th amplicon, and  $m$  is the overall median of all the amplicons. Deviation of coverage from the mean was performed for 5% GC content bands rather than percentages of 0, 1, 2, 3, . . . , 100%. Linear regression analysis was performed employing simple least square fitting.

Nucleosome-guided cfDNA fragmentation patterns were analyzed in publicly available sequences obtained from plasma samples pooled from an unknown number of healthy individuals (GSM1833219). The details of the DNA extraction, library preparation and sequencing are provided in Snyder et al. (2016). Briefly, cfDNA libraries underwent paired-end sequencing with Illumina sequence-by-synthesis technology generating reads of 101 bp in size. Importantly, at the library preparation stage, plasma DNA samples did not undergo fragmentation by sonication and, thus, original cfDNA molecules were preserved, granting the opportunity to investigate its fragmentation patterns. The fastq read sequences were aligned to the human genome (aforementioned reference build) with BWA-mem v. 0.7.12 (Li and Durbin, 2009). cfDNA fragment length may exceed sequencing read length, however, paired-end sequencing allows to capture both start and end positions of the fragment. Paired reads, thus, continued to represent WGS fragments. Nucleosome position stringencies were calculated essentially as described in Valouev et al., using the NuMap software with standard parameters. NuMap performs the nucleosome mapping based on the kernel smoothed reads count calculation (Valouev et al., 2011).

For ITCHP2 and AODCP panel amplicons, fragment counts were generated *in silico* after matching both primers with the fragment amplified and sequenced experimentally. To understand the patterns of amplicon coverage by experimentally observed fragments, the fragments were generated using paired reads, then further filtered by length to include only fragments in the range of 80 to 250 bp. Dinucleosome fragments were therefore excluded. To improve resolution, resulting fragments were trimmed by 40 bp around dyads to generate a set of equal-length fragments. For each sequenced nucleotide position, counts of overlapping fragments were recorded. Generated data were subjected to a lowpass filter with the square pulse kernel with the width of 21 base pairs, then resulting coverage plots were mapped to amplicons genome positions.

Statistical analysis was performed using R, version 3.2.3. For machine learning, we used the open source library Orange (Demsar et al., 2013). Five machine learning algorithms were evaluated to find the best model, demonstrating the highest prediction accuracy based on all descriptors [support vector machine (SVM), neural network, multiple linear regression, naive Bayes, and random forest].

## RESULTS

### Sample Sequencing and Mutation Analysis

In this study, fourteen cfDNA samples collected from patients with NSCLC, were analyzed using the screening panels ITCHP2 and AODCP. The mean sequencing coverages across all experiments were set at 1150× for the AODCP panel and 802× for the ITCHP2 panel with corresponding medians of 1002× and 674×, respectively.

Variant detection results were completely concordant for two panels across 18 identified somatic mutations. Plasma variant detection results were concordant with baseline tissue analysis in 9 samples (69%). False negative samples were limited to the cases, characterized with low plasma DNA concentration (Figure 1). In addition to mutations identified by tissue analysis at baseline, namely, these in EGFR and RAS, the sequencing of 13 plasma cfDNA samples revealed five additional somatic missense mutations, including these in PIK3CA and TP53 genes (Figure 1).

### Significance of Amplicon Length for Mutation Detection Sensitivity and Specificity

The average length of amplicons in panel AODCP was much shorter than that in panel ITCHP2 (Figure 2A), with median amplicon lengths to include primer sequences at 137 and 156 bp, respectively. Despite the difference in amplicon sizes, variant calling results obtained for each panel were completely concordant, with a total of 51 either somatic or germline variants detected. Therefore, diagnostic sensitivity and specificity of these two detection systems were the same at the study power.

In order to explore possible influences of the amplicon length on the limits of detection and, therefore, analytical sensitivity to the presence of the mutations in liquid biopsy, we performed a pairwise comparison of the frequencies for same mutated allele in reads obtained from pairs of panel-specific amplicons. For the synonymous germline variant, namely, EGFR p.Gln787= with the total of 15 alleles identified (1000 Genomes MAF 0.43), allele frequencies extracted from analysis of AODCP and ITCHP2 amplicons were equivalent (Wilcoxon signed rank test  $p$ -value = 0.88). On the other hand, analysis of somatic mutations, which are typically present in a relatively small fraction of the reads, showed Pearson's correlation coefficients of 0.88 ( $p$ -value = 0.02; Wilcoxon signed rank test  $p$ -value = 0.44) for point mutations in genes *EGFR*, *TP53*, and *PIK3CA*, and 0.95 for the deletions of the *EGFR* exon 19 ( $p$ -value = 0.001; Wilcoxon signed rank test  $p$ -value = 0.53) (Figure 3). Since *EGFR* deletions further reduce the length of amplified fragments by 15 or more bp, their presence should, at least in theory, increase analytical sensitivity of the detection system (Figure 2B). Notably, the geometric mean ratio of the allele frequency of the *EGFR* exon 19 deletions, detected by two panels, was 1.16 (95% CI, 0.72–1.88;  $p$ -value > 0.1). This indicates that the analytical sensitivity of this assay is unlikely to change even if the difference in the average sizes of amplicons would increase further.

Finally, we performed a regression analysis to estimate the relationship between amplicon length and its average coverage across samples for the ITCHP2 panel, representing a wider spectrum of amplicon lengths. After normalization on GC-content and overall sample read count, linear regression analysis employing the least squares fitting approach, demonstrated a negative slope with a Student  $t$ -test  $p$ -value of 0.0063. However, regression analysis across the set of amplicons with a length of 170 bp or less yielded a non-significant slope coefficient ( $p$ -value 0.69) (Figures 4A,B). Regression analysis between amplicon length and its coverage covariance demonstrated no significant correlation in any amplicon length range (data not shown). Considering that amplicons with a length of 120 or less comprises of only 5% of that set, this indicates that the NGS sequencing of plasma cfDNA equally tolerates amplicons with a length in the 120–170 bp range.

### Nucleosome-Guided Pattern May Facilitate Primer Panel Design

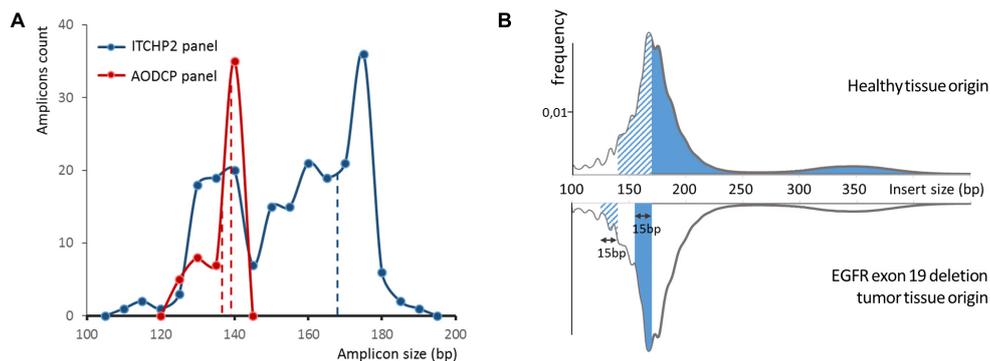
According to the most commonly cited hypothesis, plasma cfDNA originates from apoptotic cells where genomic DNA is digested by a set of nucleases (Ma et al., 2017). Wrapping around nucleosomes protects some of the DNA fragments from digestion; that is why cfDNA fragments correspond primarily to the mononucleosome bound regions. Originally supported only by a unimodal distribution of cfDNA fragments sizes (Fan et al., 2008; Lo et al., 2010), this hypothesis has been recently validated in several studies (Chandrananda et al., 2015; Snyder et al., 2016; Ulz et al., 2016). In particular, employing whole exome sequencing of cfDNA fragments to infer the read depth coverage allowed the construction of 'plasma genome-wide nucleosome maps. Mapping the fragments covered by the ITCHP2 panel, to these nucleosome maps, showed that the positions of the ITCHP2 primers were selected in a non-optimal way with respect to the nucleosome positioning ( $p$ -value for nucleosome peaks and amplicons interception 0.36). An amplicon covering *KRAS* exon 4 serves as a good illustration for non-optimal selection of primers which fall in between two peaks (Figure 5A). Because of that, amounts of spanning cfDNA fragments are much lower than for the primers selected to amplify the fragment located within the same peak. A similar situation may be observed for the *EGFR* exon 21; shifting positions of the primers by the order of 100 nucleotides may result in an increase of the depth and the uniformity of the coverage, without compromising amplification of the clinically relevant, mutation-harboring locus.

At the next stage of analysis, we inquired whether efficiency of targeted resequencing of cfDNA samples depends on the pattern of DNA fragmentation. To perform this analysis, for all amplicons represented in the ITCHP2 panel, the fragmentation patterns were extracted from the repository of reads obtained after a shotgun sequencing of cfDNA fragments purified from the pool of plasma samples, of healthy individuals and from five individual patients with solid tumors (Figure 5B).

It is known that both the nucleosome positioning (Struhl and Segal, 2013), which, in turn, guides the fragmentation of cDNA (Ma et al., 2017), and the depth and the uniformity of

Patient Id	Plasma DNA concentration (ng/ $\mu$ l)	Tumor alteration	Plasma EGFR status			Additional mutations identified		
			overall	L858R	ex19del	Gene	mutation	allele frequency
PI10	NA	EGFR	negative					
PI84	15,6	EGFR	negative					
PI87	16,7	EGFR	positive		2,6%	PIK3CA	p.His1060Gln	23,60%
PI90	32,8	ALK	positive		10,4%			
PI92	38,5	EGFR	positive	7,7%	2,8%			
PI94	46,3	EGFR	positive	5,9%	0,8%	TP53	p.Arg248Trp	3,2%
PI89	59,1	EGFR	positive		1,2%			
PI93	62,4	EGFR	positive		13,7%			
PI91	64,9	EGFR	positive		90,6%	PIK3CA	p.Glu542Gln	20,00%
						TP53	p.Arg213Leu	34,20%
PI88	84,2	EGFR	positive		0,4%			
PI19	86,1	EGFR	positive		0,6%			
PI29	92,4	EGFR	positive	7,9%		TP53	p.Ser241Phe	3,30%
PI13	169,8	EGFR	positive		4,1%			

**FIGURE 1** | Samples used for data analysis as well as mutations identified during NGS sequencing and allele frequencies thereof (plasma EGFR status). Mutations identified employing a conventional sequencing method indicated in the tumor alteration column while its match (green) or mismatch (red) with NGS results specified in plasma EGFR status column.



**FIGURE 2** | (A) Ion torrent cancer hotspot 2 (ITCHP2) and custom AODCP primer panels amplicon length distribution. A constant window of 5 bp was used to discretize amplicon length. Dotted lines demonstrate length of amplicons, covering exon 19 of the EGFR. (B) cfDNA fragment length distribution influence available for the amplification DNA molecules in plasma and, thus, amplification effectiveness. Solid fill at the top panel demonstrates the spectrum of cfDNA fragments involved in EGFR exon 19 PCR amplification employing the ITCHP2 panel. Dashed fill demonstrates the extension of that spectrum in case the AODCP panel is used. Fills in the bottom panel demonstrate the spectrum extension for two panels, respectively, in case of the 15 bp exon 19 deletion mutation.

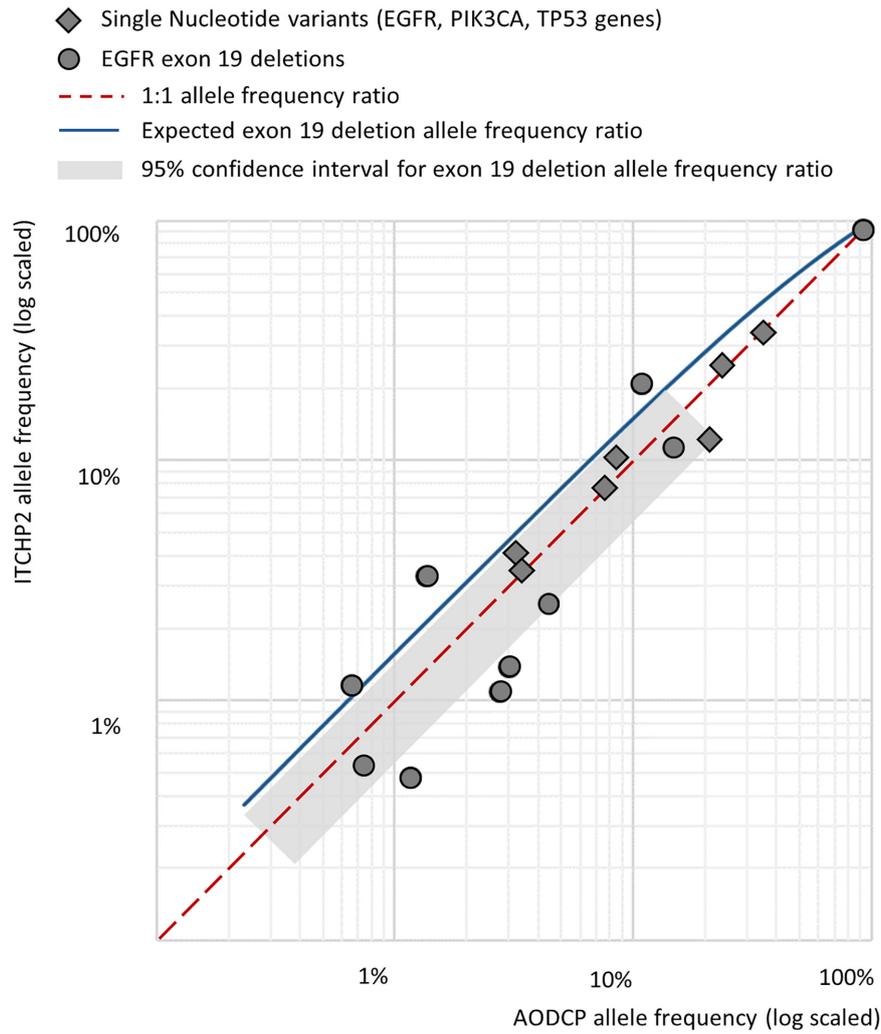
the coverage by sequencing reads (Benjamini and Speed, 2012), are influenced by the GC content. In the following analysis, we aimed at finding out whether any characteristic related to the fragmentation pattern of cfDNA within the locus of interest may influence the depth and the uniformity of coverage with amplification based sequencing reads.

For the ITCHP2 panel, each amplicon was matched to an individual nucleosome map and evaluated according to four features: (i) absolute count of experimentally observed continuous cfDNA fragments spanning the whole amplicon (Feature A), (ii) read signal amplitude within the amplicon (Feature B), (iii) read signal change at the boundaries of amplicon (Feature C), and (iv) read signal shape defined as the area between its linear approximation and itself (Feature D) (Figure 5C). Uniformity of the coverage was defined as a coefficient of inter-individual variation in read coverage between all cfDNA samples. To calculate the robustness of the nucleosome mapping, we assessed the inter-sample variance of the defined

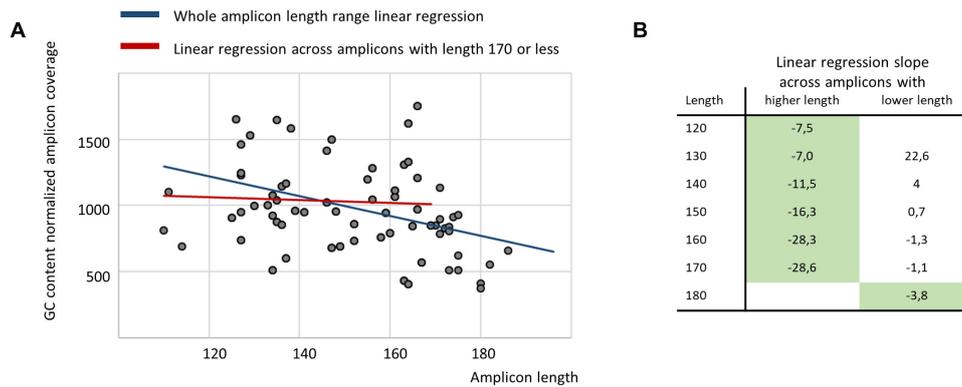
features calculated for each amplicon. Averaged coefficients of the variation of features D, B and C were at 390, 68, and 38%, respectively, pointing at significant inter-sample variation.

Further, we estimated the feature quality, employing the RReliefF method (Robnik-Sikonja and Kononenko, 2003) estimating how well their values distinguish between target variables that are near to each other. Despite previously demonstrated low robustness of the nucleosome associated features, the count of spanning fragments (Feature A) was ranked even higher than the GC content, while the other three features, B, C, and D, closely followed feature A and the GC content (Figure 6A). This finding indicates that uniformity of the locus coverage, with amplified sequencing reads, may depend on the underlying pattern of cfDNA fragmentation.

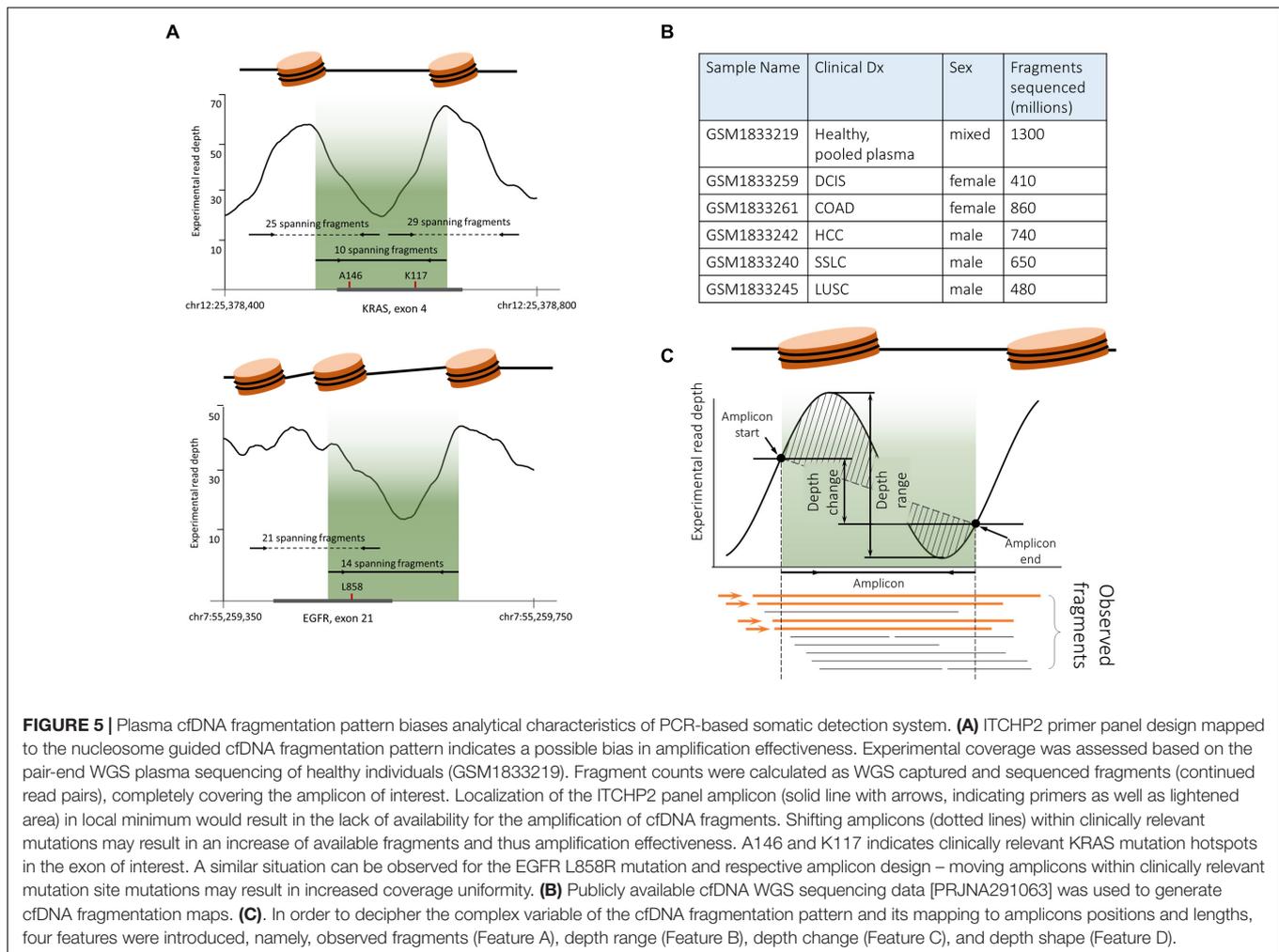
Univariate polynomial regression of the sequencing coverage depth and its coefficient of variation based on the GC content with second degree polynomial yielded coefficients of determination of 0.29 and 0.19, respectively. Furthermore,



**FIGURE 3** | Pairwise comparison of the frequencies for same mutated allele in reads obtained from pairs of panel-specific amplicons across detected somatic variants.



**FIGURE 4** | The regression analysis demonstrates no relationship between amplicon length and its average coverage when restricting the analysis to amplicons with a length of 170 or less (**A**), while, broken stick regression indicates that the system suffers a shock after 170 bp (**B**) (green fill indicates statistically significant slope coefficient with  $p$ -value of 0.05 or less).



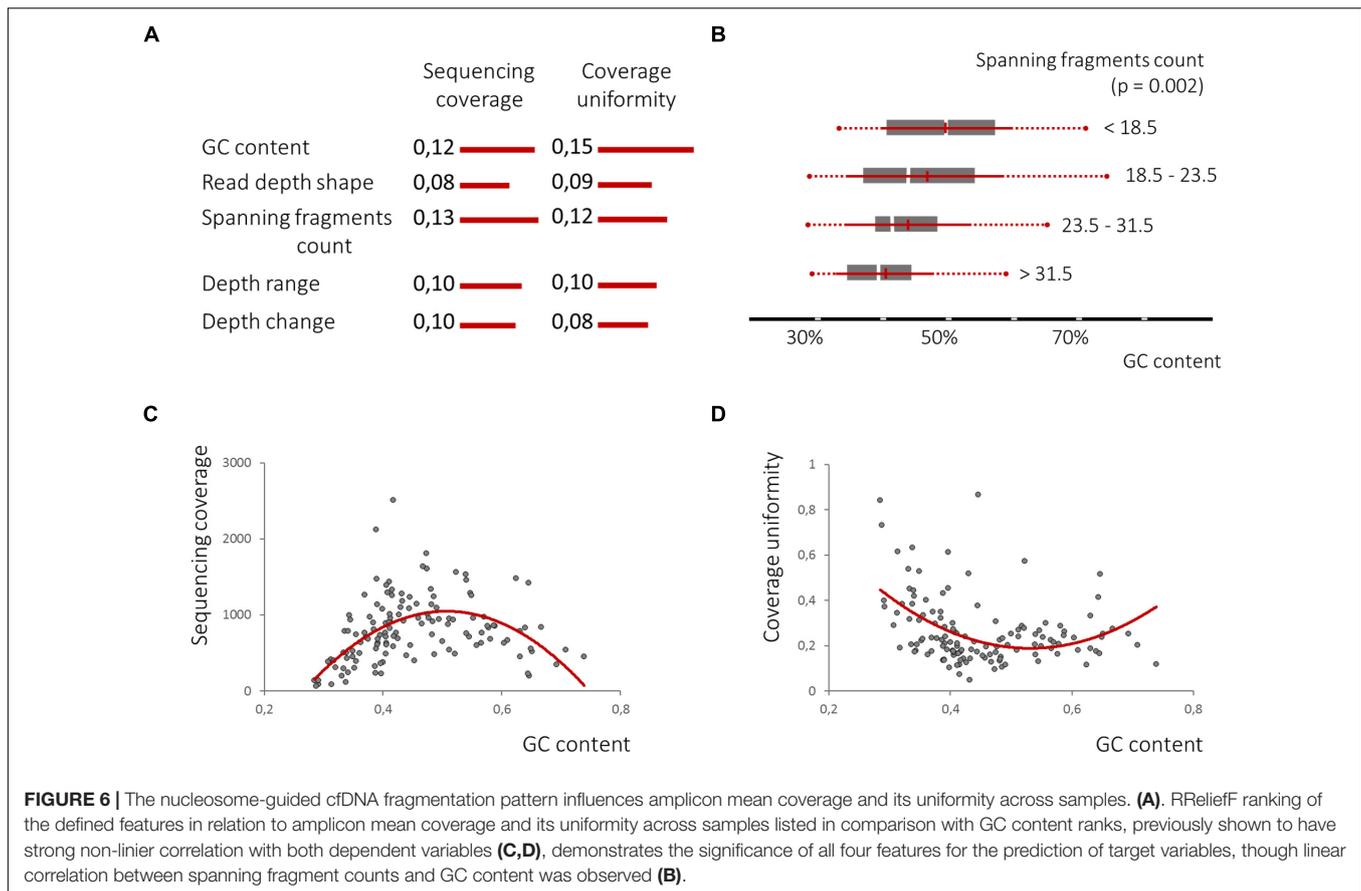
GC content equal-frequency discretization (four groups) and analysis of variance of both dependent variables between groups, yielded a  $p$ -value of less than  $1e-6$ . Thus, a strong non-linear correlation between the GC content, a sequencing coverage and its uniformity (Figures 6C,D) was detected. Despite significant linear correlation between counts of spanning fragments and the GC contents (Figure 6B), no similar relationship between this feature and sequencing coverage was seen (Figures 7A,B). In contrast, as for coverage uniformity, both spanning fragments, count and read depth coverage, shape the demonstrated correlation in relation to it (ANOVA test  $p$ -value of 0.037 and 0.013, respectively) (Figures 7C,D). No correlation was seen for depth change or depth range (data not shown).

Finally, we tested the performance of the SVM classifier for its prediction of coverage depth and coverage uniformity by either employing the GC content as a single feature or in a combination with all the other features analyzed above. Following 3-groups equal-frequency discretization, the target classes were defined as coverage depth in the lowest third tertile and coverage uniformity in the highest third tertile. For predicting the depth of coverage, GC content in combination with depth change (Feature C) were selected as features. To predict the uniformity of

coverage, GC content in combination with the spanning fragment counts (Feature A) and read depth shape (Feature D) were selected as features. A radial basis function (RBF)-kernel utilizing SVM classifier was then applied, using threefold cross-validation. Performance of the SVM classifiers, built upon several features for predicting coverage uniformity, was better than that of the GC-content only classifiers (areas under the receiver operating curve (AUROCs) of 0.75, 95% CI: 0.750–0.752 vs. 0.65, 95% CI: 0.63–0.67; precision – 0.74 vs. 0.68). This indicates that non-GC content features may aid in the prediction of the amplicons with a high coverage variation across samples. For coverage depths, however, applying a similar strategy has not resulted in a significant improvement (AUROCs of 0.69, 95% CI: 0.68–0.70 vs. 0.70, 95% CI: 0.70–0.71; precision – 0.69 vs. 0.69) (Figures 7E,F).

## DISCUSSION

The share of cfDNA fragments originating from tumor rather than normal tissues, may vary greatly among patients. In early-stage disease, the share could be as low as 0.01% of the total cfDNA (Thierry et al., 2017). Because of that, the issue of the



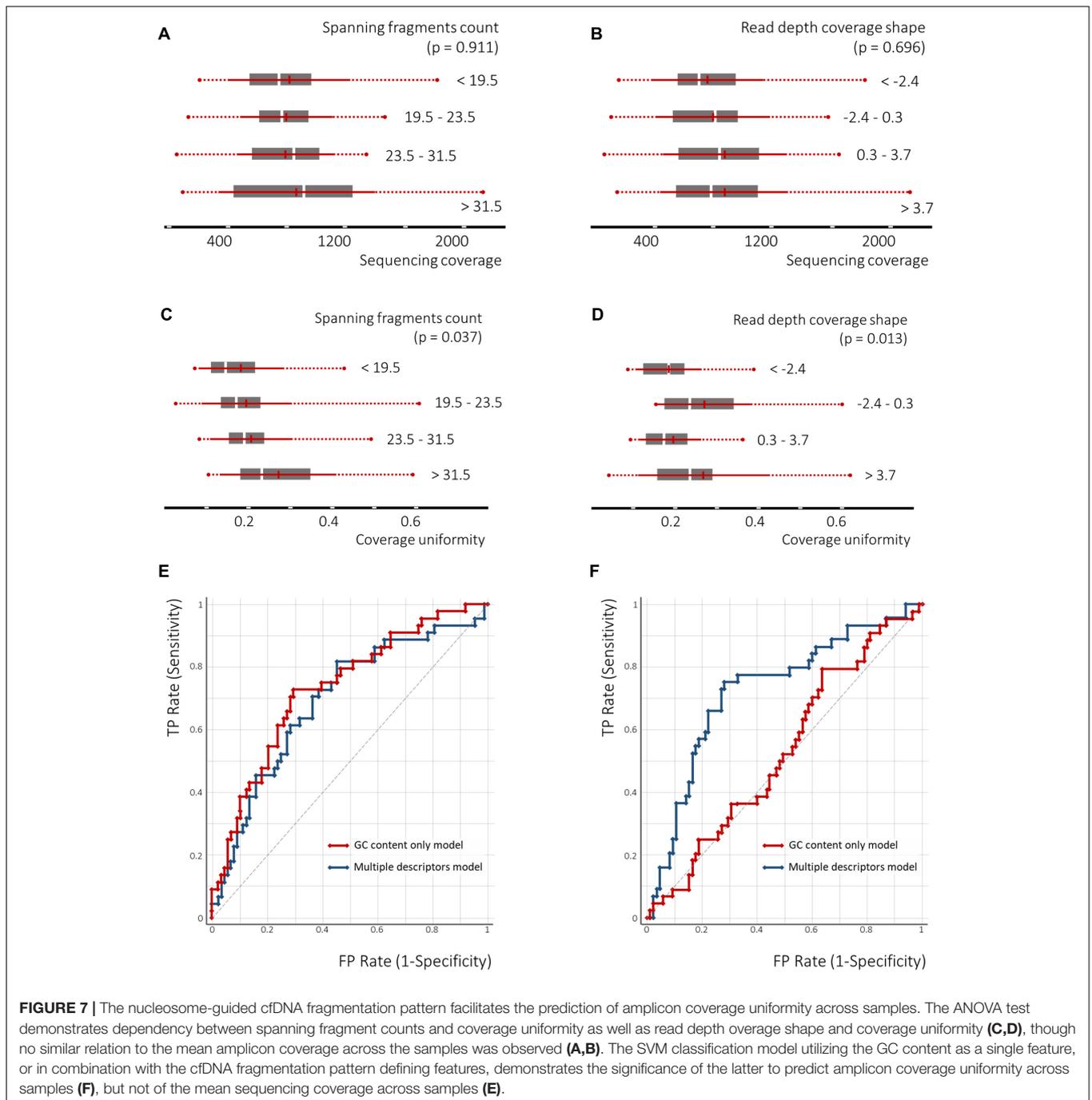
detection of low frequency mutant alleles, represents one of the biggest technical challenges to the development of diagnostic and prognostic assays involving the sequencing of cfDNA. In this study we examined various approaches to increase diagnostic and analytical sensitivity of the detection of somatic mutations in liquid biopsy samples.

In a heterogeneous cohort of patients, the liquid biopsy was performed at baseline, at disease progression and/or within the framework of disease monitoring. Overall diagnostic sensitivity of NGS to detect EGFR mutations in cfDNA was at 83%. Of note, when we limited the sample set to the plasma specimens with DNA concentration of 20 ng/ml and higher, the false negative rate was reduced from 17 to 0%. This observation points at low concentrations of cfDNA samples as a primary contributor to imperfect sensitivity of the liquid biopsy assays and at a necessity to either improve the recovery of tumor DNA fragments, or to require cfDNA profiling labs to introduce more stringent QC metrics, which may render many samples ineligible for downstream processing.

Sensitivity of cfDNA based mutation detection assays may be aided by an improvement of amplification efficiency. Plasma cfDNA is known to be highly fragmented (Fleischhacker et al., 2011; Klevebring et al., 2014; **Figure 2B**). Therefore, it is commonly recognized that an increase in length of PCR amplicons may result in the elimination of a majority of the extracted DNA fragments as possible templates. In this

study we sought to dissect how much of the amplicon length influences the sensitivity of subsequent mutation detection. For this we performed, to the best of our knowledge, the first comparison of two amplicon based NGS panels characterized by a substantial difference in average amplicon length (**Figure 2A**). The comparison was performed in relation to the panels' diagnostic and analytical sensitivity. Surprisingly, the yield of both the germline and somatic mutations between two panels were completely concordant, pointing at an irrelevance of amplicon size of the specified short range to diagnostic sensitivity of resultant assays.

As a particular example defying “the shorter amplicon, the better amplification efficiency” logic, we dissected the detection of EGFR exon 19 deletion alleles by amplicons of 138 and 168 nt in length. Based on the area under the fragment length distribution curves (**Figure 2B**), mutant alleles should be amplified 1.45 times more efficiently than wild-type ones by the panel with larger amplicons, while the panel with shorter amplicons would be 1.04 times more efficient for mutant cfDNA fragments. Considering that tumor-derived cfDNA fragments are even shorter than normal tissue-derived ones (Jiang et al., 2018), these rates would increase to 1.84 and 1.16, respectively (**Figure 3**). This should result in approximately and increase of 1.6 times of the mutant allele frequencies detected with a larger-amplicon panel as compared to a smaller-amplicon panel. In our experiment, no statistically significant difference in mutant allele frequencies



was noted, with the observed trend being the opposite to what was expected, indicating that the size of the amplicons does not contribute to the analytical sensitivity of cfDNA assays.

Notably, our observations contradict some previous work (Chan et al., 2004; Koide et al., 2005), which show a length-dependent decrease in efficiency of amplification of cfDNA templates in up to a 250 nt fragment range, which corresponds to the mononucleosome fraction representing approximately 85% of all cfDNA fragments (Figure 2B). In these previous studies, the yield of DNA dropped by almost 30 and 60% when using

amplicons with a size of 145 nt instead of 105 and 201 nt instead of 145 nt, while for amplicons with larger sizes no pronounced effect was observed. Furthermore, another study demonstrated that increases in the DNA yield may be observed at a lower amplicon size range: a direct digital PCR comparison of the 50 bp to the 84 bp amplicon resulted in significant favoring of the shorter amplicon (Koide et al., 2005; Sikora et al., 2010). It is important, however, to note that reported observations were obtained in course of analysis if cfDNA samples collected either from healthy individuals or in setting of prenatal diagnostics

aimed at amplifying fetal cfDNA and, therefore, cannot be directly projected onto the templates of tumor-derived cfDNA which is known for the shorter sizes of its fragments (Pinzani et al., 2011; Mouliere and Rosenfeld, 2015) and lower integrity (Underhill et al., 2016). The studies of cfDNA specimens collected from patients with tumors show that 60 bp fragments are almost five times more abundant than 150 bp ones, thus pointing at the necessity to use amplicons with sizes of 100 bp or lower (Mouliere et al., 2011).

Importantly, in many cases, reaping the benefit of shorter amplicon size may not be possible due to complications arising from the necessity of the precise positioning of the primers restricting optimization of their GC content, matching melting temperatures and preventing oligonucleotide dimerization. While designing PCR systems for select loci may be still possible, with EGFR analysis being the common example (Reckamp et al., 2016), the introduction of ultra-short amplicons into highly multiplexed systems aiming at a broader molecular profiling of human tumors, may not be feasible. Particular concerns about this multiplexing precluding approach to the amplicon design are owed to the recent observations of a wide mutational spectrum in the liquid biopsies of metastatic cancer patients and its relevance to possible inclusion in clinical trials (Rothé et al., 2014; Frenel et al., 2015). In light of an obvious necessity for multiplexing, the finding that varying amplicon sizes in a range from 140 up to 170 nt does not influence analytical sensitivity is significant, as it shifts the attention of panel designers from minimizing the length of the amplicons to optimizing compatibility of oligonucleotides.

Additionally, cfDNA as a template for a designed PCR-based assay may introduce a set of additional restraints. Both the prevalence of cfDNA fragments of certain sizes and the fragmentation patterns depend on the positioning of the nucleosomes within its tissue of origin. To describe this novel complex variable depicting nucleosome positioning, we introduced four features namely, a spanning fragment count, a read depth change, a read depth range and a read depth shape (Figure 5C), which collectively portray the coverage of select amplicon by experimentally obtained WGS reads. When read coverage maps of WGS-sequenced cfDNA fragments from pooled plasma of healthy patients were aligned to the amplicons employed for liquid biopsy analysis of patients with NSCLC, these four features were utilized to determine the extent of the influence of nucleosome positioning on two dependent variables: sequencing coverage and coverage uniformity. A SVM-based classifier demonstrated that combining the GC content with spanning fragment counts and read depth shape, results in an increased accuracy of prediction of both dependent variables. Therefore, this variable should be taken in consideration when designing PCR primer systems.

Nevertheless, the overall robustness of nucleosome positioning remains unclear. It is known that several regulatory events defining the gene expression require the strict positioning of nucleosomes; these events are typically associated with promoter regions (Hesson et al., 2014; Lövkvist et al., 2018). However, nucleosome positioning is not absolute, and even with major shifts in gene expression, some cells fail to change

nucleosome configuration (Small et al., 2014), thus, indicating an underlying complexity of nucleosome positioning. Importantly, the majority of clinically relevant mutations are located within exons, which, according to the current view of cfDNA nucleosome maps, do not retain a strict pattern of cfDNA fragmentation. Therefore, nucleosome arranging within such exons may be variable, either between molecular subtypes of the same disease or even between normal tissue specimens. Nevertheless, despite a potential for low robustness, a substantial correlation observed between nucleosome maps revealed by unbiased read coverage in cfDNA from healthy patients, and the sequencing coverage and its uniformity in amplicons obtained in cfDNA of patients with NSCLC, indicates that the efficiency of amplification may be improved if the unbiased read coverages are taken into account.

In conclusion, low plasma cfDNA concentration remains the major factor that limits the sensitivity of liquid biopsy assays. Above we showed that the design of a highly multiplexed system equally tolerates amplicons in the range of 140–170 bp in size, thus allowing the shift of attention toward the melting temperature, GC clamps, cross homology and other controllable variables. We have also provided evidence that the nucleosome placement in the tissue of origin and the resultant genome-wide cfDNA fragmentation pattern, may be used as a guide for primer positioning to improve both the sequencing coverage and its uniformity.

## DATA AVAILABILITY

The datasets generated for this study can be found in the Sequence Read Archive under accession number SRP167082 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP167082>). The additional datasets analyzed for this study can be found in the Sequence Read Archive under accession number SRP061633.

## AUTHOR CONTRIBUTIONS

VM, AB, MI, and SM designed the work. PC, KL, and VB collected the samples. ER performed the experiments. All authors participated in the interpretation of the results and in writing the article.

## FUNDING

This study was supported by the Ministry of Science and Education, Russia (Project No. RFMEFI60714X0098).

## ACKNOWLEDGMENTS

The authors wish to gratefully acknowledge technical support from laboratory of epigenetics, Medical Genetic Science Center RAMS. Special thanks to Drs. Strelnikov and Tanas for their technical assistance and helpful discussions of presented here results.

## REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Benjamini, Y., and Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* 40:e72. doi: 10.1093/nar/gks001
- Bryzgunova, O., Bondar, A., Morozkin, E., Mileyko, V., Vlassov, V., and Laktionov, P. (2011). A reliable method to concentrate circulating DNA. *Anal. Biochem.* 408, 354–356. doi: 10.1016/j.ab.2010.09.005
- Chan, K. C., Zhang, J., Hui, A. B., Wong, N., Lau, T. K., Leung, T. N., et al. (2004). Size distributions of maternal and fetal DNA in maternal plasma. *Clin. Chem.* 50, 88–92. doi: 10.1373/clinchem.2003.024893
- Chandrananda, D., Thorne, N. P., and Bahlo, M. (2015). High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genomics* 8:29. doi: 10.1186/s12920-015-0107-z
- Dawson, S. J., Tsui, D. W., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S. F., et al. (2013). Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* 368, 1199–1209. doi: 10.1056/NEJMoa1213261
- Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., et al. (2013). Orange: data mining toolbox in python. *J. Mach. Learn. Res.* 14, 2349–2353.
- Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., and Quake, S. R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16266–16271. doi: 10.1073/pnas.0808319105
- Fleischhacker, M., Schmidt, B., Weickmann, S., Fersching, D. M., Leszinski, G. S., Siegle, B., et al. (2011). Methods for isolation of cell-free plasma DNA strongly affect DNA yield. *Clin. Chim. Acta* 412, 2085–2088. doi: 10.1016/j.cca.2011.07.011
- Forbes, S. A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., et al. (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucleic Acids Res.* 38, D652–D657. doi: 10.1093/nar/gkp995
- Frenel, J. S., Carreira, S., Goodall, J., Roda, D., Perez-Lopez, R., Tunariu, N., et al. (2015). Serial next-generation sequencing of circulating cell-free DNA evaluating tumor clone response to molecularly targeted drug administration. *Clin. Cancer Res.* 21, 4586–4596. doi: 10.1158/1078-0432.CCR-15-0584
- Hellwig, S., Nix, D. A., Gligorich, K. M., O’Shea, J. M., Thomas, A., Fuertes, C. L., et al. (2018). Automated size selection for short cell-free DNA fragments enriches for circulating tumor DNA and improves error correction during next generation sequencing. *PLoS One* 13:e0197333. doi: 10.1371/journal.pone.0197333
- Hesson, L. B., Sloane, M. A., Wong, J. W., Nunez, A. C., Srivastava, S., Ng, B., et al. (2014). Altered promoter nucleosome positioning is an early event in gene silencing. *Epigenetics* 9, 1422–1430. doi: 10.4161/15592294.2014.970077
- Ivanov, M., Baranova, A., Butler, T., Spellman, P., and Mileyko, V. (2015). Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genomics* 16(Suppl. 13):S1. doi: 10.1186/1471-2164-16-S13-S1
- Ivanov, M., Matsvay, A., Glazova, O., Krasovskiy, S., Usacheva, M., Amelina, E., et al. (2018). Targeted sequencing reveals complex, phenotype-correlated genotypes in cystic fibrosis. *BMC Med. Genomics* 11(Suppl. 1):13. doi: 10.1186/s12920-018-0328-z
- Jiang, P., Chan, C. W., Chan, K. C., Cheng, S. H., Wong, J., Wong, V. W., et al. (2015). Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1317–E1325. doi: 10.1073/pnas.1500076112
- Jiang, P., Sun, K., Tong, Y. K., Cheng, S. H., Cheng, T. H. T., Heung, M. M. S., et al. (2018). Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 115, E10925–E10933. doi: 10.1073/pnas.1814616115
- Klevebring, D., Neiman, M., Sundling, S., Eriksson, L., Darai Ramqvist, E., Celebioglu, F., et al. (2014). Evaluation of exome sequencing to estimate tumor burden in plasma. *PLoS One* 9:e104417. doi: 10.1371/journal.pone.0104417
- Koide, K., Sekizawa, A., Iwasaki, M., Matsuoka, R., Honma, S., Farina, A., et al. (2005). Fragmentation of cell-free fetal DNA in plasma and urine of pregnant women. *Prenat. Diagn.* 25, 604–607. doi: 10.1002/pd.1213
- Krishnamurthy, N., Spencer, E., Torkamani, A., and Nicholson, L. (2017). Liquid biopsies for cancer: coming to a patient near you. *J. Clin. Med.* 6:E3. doi: 10.3390/jcm6010003
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lo, Y. M., Chan, K. C., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M., et al. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* 2:61ra91. doi: 10.1126/scitranslmed.3001720
- Lövkvist, C., Sneppen, K., and Haerter, J. O. (2018). Exploring the link between nucleosome occupancy and DNA methylation. *Front. Genet.* 8:232. doi: 10.3389/fgene.2017.00232
- Ma, X., Zhu, L., Wu, X., Bao, H., Wang, X., Chang, Z., et al. (2017). Cell-Free DNA provides a good representation of the tumor genome despite its biased fragmentation patterns. *PLoS One* 12:e0169231. doi: 10.1371/journal.pone.0169231
- Malapelle, U., Pisapia, P., Rocco, D., Smeraglio, R., di Spirito, M., Bellocicene, C., et al. (2016). Next generation sequencing techniques in liquid biopsy: focus on non-small cell lung cancer patients. *Transl. Lung Cancer Res.* 5, 505–510. doi: 10.21037/tlcr.2016.10.08
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., et al. (2010). Tablet–next generation sequence assembly visualization. *Bioinformatics* 26, 401–402. doi: 10.1093/bioinformatics/btp666
- Mouliere, F., Robert, B., Arnau Peyrotte, E., Del Rio, M., Ychou, M., Molina, F., et al. (2011). High fragmentation characterizes tumour-derived circulating DNA. *PLoS One* 6:e23418. doi: 10.1371/journal.pone.0023418
- Mouliere, F., and Rosenfeld, N. (2015). Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3178–3179. doi: 10.1073/pnas.1501321112
- Oxnard, G. R., Paweletz, C. P., Kuang, Y., Mach, S. L., O’Connell, A., Messineo, M. M., et al. (2014). Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin. Cancer Res.* 20, 1698–1705. doi: 10.1158/1078-0432.CCR-13-2482
- Pinzani, P., Salvianti, F., Zaccara, S., Massi, D., De Giorgi, V., Pazzagli, M., et al. (2011). Circulating cell-free DNA in plasma of melanoma patients: qualitative and quantitative considerations. *Clin. Chim. Acta* 412, 2141–2145. doi: 10.1016/j.cca.2011.07.027
- Pupilli, C., Pinzani, P., Salvianti, F., Fibbi, B., Rossi, M., Petrone, L., et al. (2013). Circulating BRAFV600E in the diagnosis and follow-up of differentiated papillary thyroid carcinoma. *J. Clin. Endocrinol. Metab.* 98, 3359–3365. doi: 10.1210/jc.2013-1072
- Reckamp, K. L., Melnikova, V. O., Karlovich, C., Sequist, L. V., Camidge, D. R., Wakelee, H., et al. (2016). A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. *J. Thorac. Oncol.* 11, 1690–1700. doi: 10.1016/j.jtho.2016.05.035
- Robnik-Sikonja, M., and Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* 53, 23–69. doi: 10.1023/A:1025667309714
- Rothé, F., Laes, J. F., Lambrechts, D., Smeets, D., Vincent, D., Maetens, M., et al. (2014). Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Ann. Oncol.* 25, 1959–1965. doi: 10.1093/annonc/mdu288
- Sacher, A. G., Paweletz, C., Dahlberg, S. E., Alden, R. S., O’Connell, A., Feeney, N., et al. (2016). Prospective validation of rapid plasma genotyping for the detection of EGFR and KRAS mutations in advanced lung cancer. *JAMA Oncol.* 2, 1014–1022. doi: 10.1001/jamaoncol.2016.0173

- Saunders, C. T., Wong, W. S., Swamy, S., Becq, J., Murray, L. J., and Cheetham, R. K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817. doi: 10.1093/bioinformatics/bts271
- Sikora, A., Zimmermann, B. G., Rusterholz, C., Birri, D., Kolla, V., Lapaire, O., et al. (2010). Detection of increased amounts of cell-free fetal DNA with short PCR amplicons. *Clin. Chem.* 56, 136–138. doi: 10.1373/clinchem.2009.132951
- Siravegna, G., Mussolin, B., Buscarino, M., Corti, G., Cassingena, A., Crisafulli, G., et al. (2015). Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat. Med.* 21, 795–801. doi: 10.1038/nm.3870
- Small, E. C., Xi, L., Wang, J. P., Widom, J., and Licht, J. D. (2014). Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2462–2471. doi: 10.1073/pnas.1400517111
- Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M., and Shendure, J. (2016). Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 164, 57–68. doi: 10.1016/j.cell.2015.11.050
- Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273. doi: 10.1038/nsmb.2506
- Tafe, L. J., Gorlov, I. P., de Abreu, F. B., Lefferts, J. A., Liu, X., Pettus, J. R., et al. (2015). Implementation of a molecular tumor board: the impact on treatment decisions for 35 patients evaluated at dartmouth-hitchcock medical center. *Oncologist* 20, 1011–1018. doi: 10.1634/theoncologist.2015-0097
- Thierry, A. R., Pastor, B., Jiang, Z. Q., Katsiampoura, A. D., Parseghian, C., Loree, J. M., et al. (2017). Circulating DNA demonstrates convergent evolution and common resistance mechanisms during treatment of colorectal cancer. *Clin. Cancer Res.* 23, 4578–4591. doi: 10.1158/1078-0432.CCR-17-0232
- Thress, K. S., Brant, R., Carr, T. H., Dearden, S., Jenkins, S., Brown, H., et al. (2015). EGFR mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer* 90, 509–515. doi: 10.1016/j.lungcan.2015.10.004
- Ulz, P., Thallinger, G. G., Auer, M., Graf, R., Kashofer, K., Jahn, S. W., et al. (2016). Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* 48, 1273–1278. doi: 10.1038/ng.3648
- Underhill, H. R., Kitzman, J. O., Hellwig, S., Welker, N. C., Daza, R., Baker, D. N., et al. (2016). Fragment length of circulating tumor DNA. *PLoS Genet.* 12:e1006162. doi: 10.1371/journal.pgen.1006162
- Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520. doi: 10.1038/nature10002
- Wei, Z., Shah, N., Deng, C., Xiao, X., Zhong, T., and Li, X. (2016). Circulating DNA addresses cancer monitoring in non small cell lung cancer patients for detection and capturing the dynamic changes of the disease. *Springerplus* 5:531. doi: 10.1186/s40064-016-2141-5
- Xi, L., Pham, T. H., Payabyab, E. C., Sherry, R. M., Rosenberg, S. A., and Raffeld, M. (2016). Circulating tumor DNA as an early indicator of response to T-cell transfer immunotherapy in metastatic melanoma. *Clin. Cancer Res.* 22, 5480–5486. doi: 10.1158/1078-0432.CCR-16-0613
- Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., et al. (2017). Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* 23, 703–713. doi: 10.1038/nm.4333

**Conflict of Interest Statement:** ER, SM, VM, and AB were employed by company Atlas Oncology Diagnostics, Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer, TT declared a past co-authorship with one of the authors AB to the handling Editor.

Copyright © 2019 Ivanov, Chernenko, Breder, Laktionov, Rozhavskaia, Musienko, Baranova and Mileyko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.