

LETTER TO THE EDITOR

Open Access

Deep cfDNA fragment end profiling enables cancer detection



Yulia V. Zhitnyuk¹, Anastasia P. Koval¹, Aleksandr A. Alferov², Yanina A. Shtykova³, Ilgar Z. Mamedov^{1,4,5}, Nikolay E. Kushlinskii², Dmitriy M. Chudakov^{1,4} and Dmitry S. Shcherbo^{1*} 

Keywords: cfDNA, Fragmentomics, cfDNA-FEP, Liquid biopsy, Early cancer detection, Renal cancer, Colorectal cancer

Background

The number of cancer cases is expected to increase by 40% in 20 years and reach nearly 30 million new cases per year in 2040 [1]. Therefore, it is of utmost importance to get a grip on cancer prevention and early diagnosis. Colorectal cancer (CRC) is the third most commonly diagnosed and the second most deadly cancer worldwide [1]. Because it begins insidiously, 20% of CRC cases are not discovered until cancer has already outgrown the colon [2]. Detecting tumors at an early stage represents a major opportunity to reduce CRC morbidity and mortality and improve patient prognosis. Renal cell carcinoma is the ninth most common cancer worldwide, with an increasing incidence due to growing obesity rates, smoking and alcohol consumption. In most cases, renal cell carcinoma is diagnosed incidentally on imaging, and rarely presents with either classic symptoms such as hematuria and flank mass or paraneoplastic syndromes or varicocele in men [3]. 35% of renal cell carcinoma cases are detected at the metastatic stage, and no current screening test is available.

Cell-free DNA (cfDNA) found in the bloodstream is primarily a byproduct of cell death in both normal and cancer cells [4]. Circulating DNA fragments are mainly short molecules with an average length of mononucleosome size that tend to be more fragmented in internucleosomal linkers and open chromatin regions. This leads to a biased, non-random fragmentation pattern [5]. Moreover,

tumor-derived DNA fragments (ctDNA) tend to be shorter than the non-tumor cell-derived fraction, and constantly accumulating evidence suggests that cfDNA fragmentation may serve as a cancer biomarker at the whole-genome level [6, 7]. Some studies argue the presence of specific genomic regions with preferential tissue-specific or tumor-specific cfDNA fragmentation [8]. Recently, several groups have thoughtfully characterized open chromatin landscapes in human cancer [9, 10], allowing further extrapolations to the cfDNA fragmentation footprints [11]. Here, we focus on targeted high-resolution profiling of cancer-specific open-chromatin regions in cfDNA from the blood of healthy individuals and patients with colorectal and renal cancers. We demonstrate that the proposed approach can facilitate cancer detection.

Results

Targeted fragment end profiling in cfDNA

To design an assay capable of capturing cfDNA fragment end profile shifts in cancer, we examined the available ATAC-seq datasets of 410 tumor samples from The Cancer Genome Atlas (TCGA). These data characterize chromatin accessibility in 23 cancer types, including colon adenocarcinoma (COAD) and renal cell carcinoma (RCC) with a peak resolution of 500 bp [10]. Of these, we selected 48 COAD-specific and 48 RCC-specific peaks based on their normalized scores and specificity for the corresponding cancer type (Supplementary Fig. 1). To accurately analyze cfDNA fragment end profiles (cfDNA-FEP) in genomic regions of interest, we used a modified anchored multiplex PCR approach followed by NGS [12]. Briefly, ligation of the universal adapter to

*Correspondence: dshcherbo@gmail.com

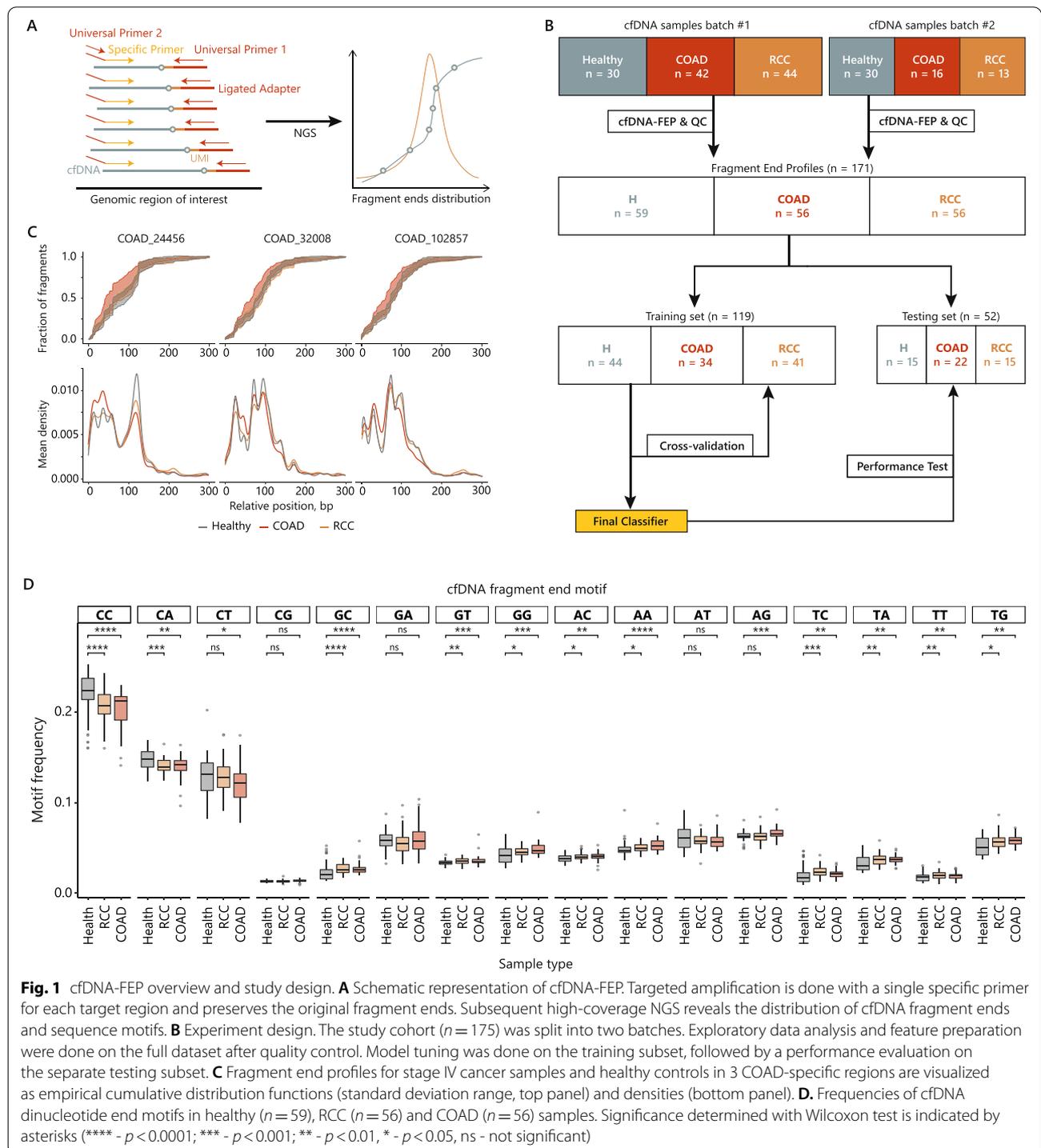
¹ Institute of Translational Medicine, Pirogov Russian National Research Medical University, 1 Ostrovityanova str, Moscow, Russia 117997
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

cfDNA is followed by primer extension from the target primer pool such that the resulting products contain universal adapter sequences at the 3'-ends. The ligated adapters contain unique molecular identifiers (UMIs) to effectively remove PCR duplicates during downstream analysis and converge read counts to the number of original cfDNA molecules. Subsequent amplification is

performed with universal primers to reduce PCR biases. This scheme allows for targeted amplification while preserving information about the original end coordinates of the cfDNA fragments (Fig. 1A). The distributions of relative end positions reflect cfDNA fragmentation profiles specific to each target region. We hypothesized that the cfDNA end profile pattern in open-chromatin



regions might differ between healthy individuals and cancer patients.

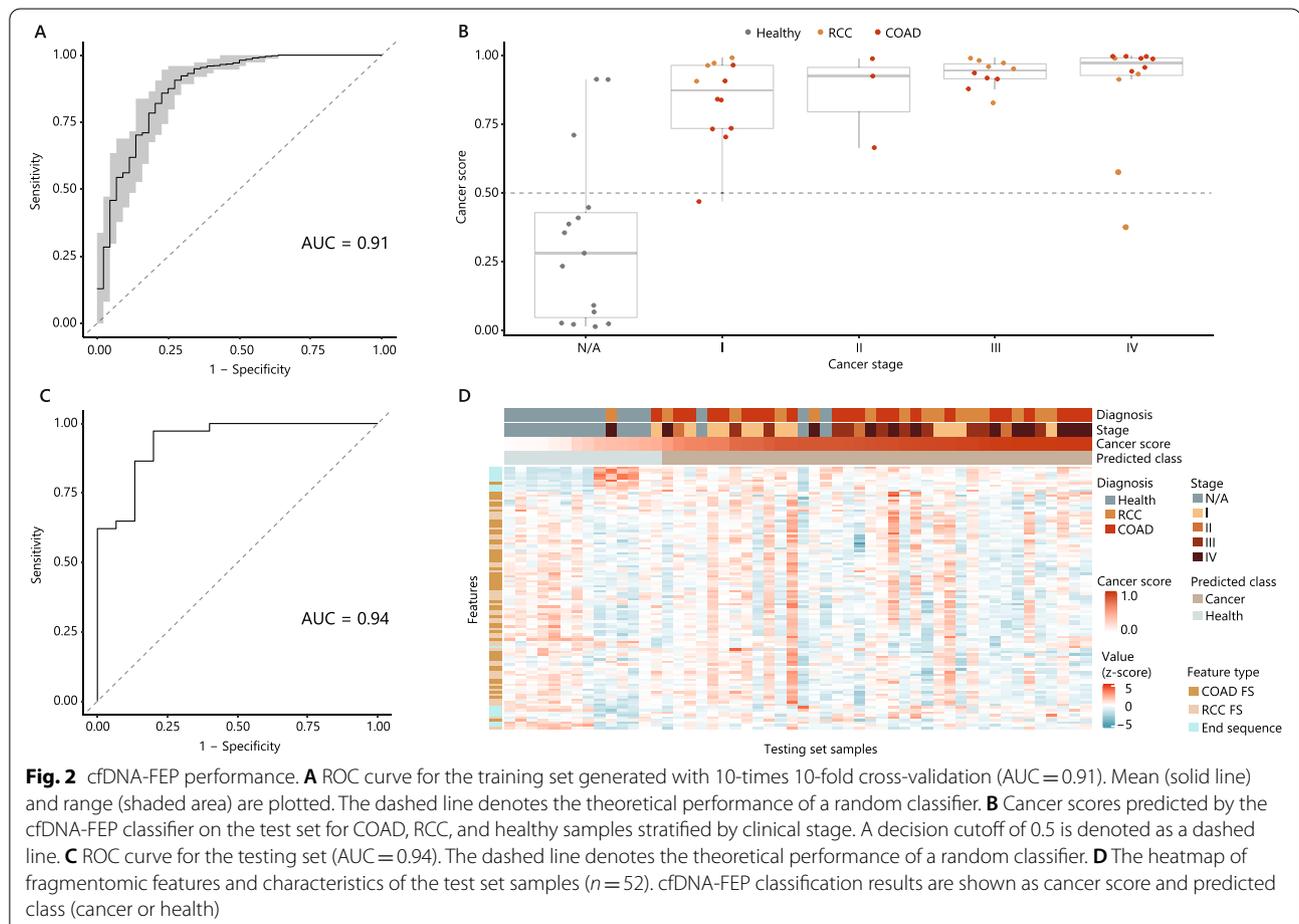
cfDNA-FEP on a clinical cohort

A cohort of 175 individuals with histologically confirmed CRC ($n=58$) and RCC ($n=57$), as well as age-matched healthy individuals ($n=60$), was divided into two batches ($n=116$ and $n=59$) that were processed individually to account for potential batch effects (Fig. 1B). After library preparation and paired-end next-generation sequencing, we performed UMI-clustering to remove PCR duplicates and then aligned clustered reads to the reference genome. To generate end profiles, we retrieved only proper pairs where the second reads overlapped the target primer binding sites. The relative start positions of the corresponding first reads represent the end profile of cfDNA molecules for each region (Fig. 1C, top). We examined the densities of the distributions in each target region and found that, in at least some regions, the average density at the peaks differed between cancer and control groups (Fig. 1C, bottom). These peaks in density denote sites

where cfDNA is predominantly cleaved and may vary due to nucleosome repositioning, change in chromatin state, or aberrant nuclease activity associated with pathology [13]. To build a classifier, we selected the most prominent peaks in the range of 0–99 bp (Peak1) and 100–300 bp (Peak2) within each target region (Supplementary Fig. 2). The normalized log₂ ratio of the densities in Peak1 and Peak2 served as a single-value metric of the fragmentation profile for each target region, or the fragmentation score (FS). We further noted that dinucleotides at the ends of the cfDNA fragments were not evenly distributed, with CC being the most frequent motif (Fig. 1D). This is consistent with the previous reports on hepatocellular carcinoma and may be a consequence of aberrant nuclease activity in cancer [14]. Therefore, we used the frequencies of sequence motifs along with FS values as predictors for the cfDNA-FEP model.

Patient samples classification

We trained support vector machine classifiers on a dedicated training dataset to select the best-performing model based on the area under the ROC curve. The final classifier



was able to distinguish cancer and healthy samples on the training dataset (10 times 10-fold cross-validation) with a mean AUC = 0.91 (sd = 0.09, $n = 100$) (Fig. 2A) and on the unseen test dataset with an AUC = 0.94 and an accuracy of 0.9 (Fig. 2C). The cfDNA-FEP classifier generates a cancer score for each cfDNA sample. This metric reflects the probability that the cfDNA sample is from a patient with cancer (Fig. 2B). For samples from the test dataset, we observed only a slight decrease in classifier performance for early-stage (I, II) cancer (AUC = 0.91, accuracy 0.87) compared with late-stage (III, IV) disease (AUC = 0.96, accuracy 0.89). The median cancer score for healthy and stage I cancer samples was 0.275 (sd = 0.294, $n = 15$), and 0.831 (sd = 0.162, $n = 12$), respectively, making a classification of even early-stage cancers feasible with the decision cut-off of 0.5 (Fig. 2B). Stage IV cancer samples ($n = 12$) were labeled with a median cancer score of 0.955, sd = 0.205. The cfDNA-FEP was able to detect both cancer types in the test set with similar performance: AUC for RCC and COAD test set samples was 0.94.

Discussion

Fragmentomic cfDNA features can be considered as independent analytes or additional biomarker layers in liquid biopsies. Several studies have demonstrated the utility of fragmentomic markers for cancer detection using whole-genome sequencing [8, 15, 16]. However, there are few reports of targeted assays [17] that are potentially more effective in the clinical setting due to their lower sequencing requirements and ability to complement existing approaches. Detection of cfDNA end profiles does not require additional treatments and can be combined with other NGS assays (the detection of cytosine methylation changes or somatic mutations). Moreover, current experimental evidence of cfDNA fragmentomic-based tumor detection is enriched with cancer types that are believed to shed more ctDNA into the bloodstream (e.g., liver, colorectal, lung, and breast), while fewer reports of successful detection of low-shedding cancers, including renal, are available [16]. In this report, we show that deep targeted profiling of cfDNA ends distributions and sequence motifs can reveal the presence of early-stage colorectal and renal cancers with an AUC = 0.94. The limitation of our study design is the lack of a group with benign pathological lesions in the colon and kidneys, so our results do not demonstrate the ability of the approach to distinguish cancer from other forms of tissue damage. Another direction for improvement would be a wider screening for additional targets derived from sources other than ATAC-seq data, such as regions of stable nucleosome repositioning specific to cancer cells or tumor-specific transcription factor binding sites.

Conclusion

Our results show that deep profiling of cfDNA fragment ends can facilitate the detection of colorectal and renal cancers. We believe that cfDNA-FEP can be further extended to non-invasively detect more cancer types with higher precision.

Abbreviations

ATAC-seq: Assay for transposase-accessible chromatin using sequencing; AUC: Area under the curve; cfDNA: Cell-free DNA; cfDNA-FEP: Cell-free DNA fragment end profiling; COAD: Colon adenocarcinoma; CRC: Colorectal cancer; ctDNA: Circulating tumor DNA; FIT: Fecal immunohistochemical test; FS: Fragmentation score; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma; NGS: Next-generation sequencing; RCC: Renal cell carcinoma; ROC: Receiver operating characteristic; UMI: Unique molecular identifier.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12943-021-01491-8>.

Additional file 1. Supplementary Methods.

Additional file 2: Supplementary Fig 1. ATAC-seq signal in RCC (KIRC and KIRP) and COAD-specific open-chromatin regions (rows) analyzed in this study shown for the samples (columns) from the TCGA cohort. Data from [10].

Additional file 3: Supplementary Fig 2. Densities of fragment end distributions in all target regions analyzed in this study plotted for COAD, RCC, and healthy cfDNA samples. Black vertical lines represent positions of Peak1 and Peak2.

Additional file 4: Supplementary Fig 3. Demographic and clinical characteristics of the cohort. **A-C.** Age and sex distribution across clinical groups (A), batches (B), and train/test split (C). **D.** The cfDNA yields across clinical groups and cancer stages. **E, F.** Stage and diagnosis composition of the batches (E), training and test sets (F).

Additional file 5: Supplementary Table S1. Cohort Characteristics.

Additional file 6: Supplementary Table S2. List of Used Oligonucleotides.

Authors' contributions

YVZ and APK performed cfDNA and NGS experiments. AAA and YAS collected and described clinical samples. DSS analyzed the data. DSS, YVZ, and APK wrote the manuscript. DSS, DMC, NEK, and IZM planned, conceptualized, and supervised the study. DMC critically revised the manuscript. The author(s) read and approved the final manuscript.

Funding

This study was supported by a grant from the Russian Science Foundation (project #20-75-10008).

Availability of data and materials

Code and cfDNA fragment end positions in target regions for all samples are available from the GitHub repository <https://github.com/dshcherbo/cfDNA-FEP>. Sensitive patient-derived cfDNA sequencing data is available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was endorsed by the Local Ethics Committee of the Russian National Research Medical University (Record # 200). Written informed consent was obtained from each participant.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Translational Medicine, Pirogov Russian National Research Medical University, 1 Ostrovityanova str, Moscow, Russia 117997. ²Laboratory of Clinical Biochemistry, N.N. Blokhin National Medical Research Center of Oncology, 23 Kashirskoye Highway, Moscow, Russia 115478. ³Federal Center for Brain and Neurotechnology, 1/10 Ostrovityanova str, Moscow, Russia 117513. ⁴Department of Genomics of Adaptive Immunity, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, 16/10 Miklukho-Maklaya str, Moscow, Russia 117997. ⁵Dmitry Rogachev National Medical and Research Center of Pediatric Hematology, Oncology and Immunology, 1 Samory Mashela str, Moscow, Russia 117997.

Received: 9 November 2021 Accepted: 26 December 2021

Published online: 21 January 2022

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality world-wide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71:209–49.
- Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal Cancer: a review. *JAMA.* 2021;325:669–85.
- Decastro GJ, McKiernan JM. Epidemiology, clinical staging, and presentation of renal cell carcinoma. *Urol Clin North Am.* 2008;35(581–92):vi.
- Heitzer E, Auinger L, Speicher MR. Cell-free DNA and apoptosis: how dead cells inform about the living. *Trends Mol Med.* 2020;26:519–28.
- Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell.* 2016;164:57–68.
- Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature.* 2019. <https://doi.org/10.1038/s41586-019-1272-6>.
- Mouliere F, Robert B, Peyrotte E, Del Rio M, Ychou M, Molina F, et al. High fragmentation characterizes tumour-derived circulating DNA. *PLoS One.* 2011;6. <https://doi.org/10.1371/journal.pone.0023418>.
- Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proc Natl Acad Sci U S A.* 2018;115:E10925–33.
- Wang Z, Tu K, Xia L, Luo K, Luo W, Tang J, et al. The open chromatin landscape of non-small cell lung carcinoma. *Cancer Res.* 2019. <https://doi.org/10.1158/0008-5472.CAN-18-3663>.
- Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science.* 2018;362. <https://doi.org/10.1126/science.aav1898>.
- Ulz P, Perakis S, Zhou Q, Moser T, Belic J, Lazzeri I, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat Commun.* 2019;10:4666.
- Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, et al. Anchored multiplex PCR for targeted next-generation sequencing. *Nat Med.* 2014;20:1479–84.
- Han DSC, Ni M, Chan RWY, Chan VWH, Lui KO, Chiu RWK, et al. The biology of cell-free DNA fragmentation and the roles of DNASE1, DNASE1L3, and DFFB. *Am J Hum Genet.* 2020;106:202–14.
- Jiang P, Sun K, Peng W, Cheng SH, Ni M, Yeung PC, et al. Plasma DNA end-motif profiling as a Fragmentomic marker in Cancer, pregnancy, and transplantation. *Cancer Discov.* 2020;10:664–73.
- Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat Commun.* 2021;12:5060.
- Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med.* 2018;10:eaat4921.
- Zhu G, Guo YA, Ho D, Poon P, Poh ZW, Wong PM, et al. Tissue-specific cell-free DNA degradation quantifies circulating tumor DNA burden. *Nat Commun.* 2021;12:2229.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

