COMMENTARY

Circulating tumor-derived DNA is shorter than somatic DNA in plasma

Florent Mouliere and Nitzan Rosenfeld[1]

Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, United Kingdom

Circulating tumor DNA (ctDNA) is now widely investigated as a biomarker in translational and clinical research (1). However, despite the growing field of clinical applications, the biology of ctDNA remains unclear. In trying to learn about the origins of ctDNA, nature provides us with very few clues. One of the important accessible parameters is the size of those DNA fragments. In addition, a well-informed model of these sizes and biases can help design more efficient and accurate diagnostic methods. In PNAS, Jiang et al. take an important step in that direction (2).

Previous efforts to characterize the size distribution of ctDNA were conducted with a variety of methods, and in different cancer types and stages, yielding contradictory evidence (3, 4). Such observations were hindered by technological limitations that only enabled assessment of limited fragment sizes and loci, or by methods that could not effectively differentiate germ-line DNA from DNA of tumor origin. Jiang et al.'s (2) work, directed by Dennis Lo, focuses on those limitations and proposes a novel design that makes use of next-generation sequencing and recurrent chromosomal aneuploidies frequently found in hepatocellular carcinoma (HCC) (2).

Sizing Up Circulating Tumor DNA at a Chromosomal Scale

One of the challenges in the analysis of plasma DNA is to differentiate circulating DNA derived from the tumor from nontumor circulating DNA. Animal models of xenografted human cancer cells have been used to differentiate tumor DNA (through its human sequences) from the nontumor, nonhuman genome (3, 5). However, investigating specifically tumor-derived DNA in human blood samples relies on genetic alterations in the DNA released by the tumor. Point mutations are infrequent throughout the genome [~4.2 somatic point mutations per megabase for HCC (6)], and not regularly distributed or easily assayed at such scale. Leveraging copy number aberrations is a pragmatic means to assess ctDNA on a genomic scale (7, 8), without explicit identification of tumor-derived fragments. Copy number aberrations frequently occur in HCC and often involve large portions of chromosome arms (9). Jiang et al. (2) leverage the fact that the fraction of tumor DNA in plasma would be higher for genomic regions that are amplified in the tumor tissue compared with nonamplified regions, whereas this fraction would be lower for regions that are deleted. A *z*-score statistic was used to determine if the plasma DNA representation of a chromosomal arm was significantly increased or decreased compared with a reference group.

Jiang et al. extract DNA from plasma using a method that filters and captures DNA of a particular size range, and measure DNA fragment sizes by paired-end sequencing.

Using massively parallel sequencing, the size profile of plasma DNA was characterized at single-base resolution for 225 samples, including 90 HCC patients, healthy subjects, and importantly, also individuals with hepatitis B virus infections, with and without liver cirrhosis. The distribution of cell-free DNA (cfDNA) fragment lengths was found to have a mode near 166 bp in all samples, in accordance with previous reports (10). In HCC patients, differences in the size distribution were observed: plasma DNA of patients with high fractions of tumor DNA in plasma appeared more fragmented than healthy individuals. Paradoxically, patients with low fractions of tumor DNA in plasma had an increased representation of larger-size fragments. Focusing on differences in the distributions of fragment sizes of sequences originating from chromosomal arms 8q (frequently amplified in HCC) and 8p (frequently deleted in those cancers), the authors found a consistent trend toward smaller sizes of DNA fragments from the chromosomal arms amplified in HCC. This elegant analysis provides evidence that ctDNA fragments in plasma with high tumor content have a bias toward shorter fragments compared with DNA originating from somatic tissues.

Origins and Significance of Circulating DNA Fragmentation

"Liquid biopsy" applications of ctDNA that aim to measure the fractional level of tumor alleles (1), or that rely on it for identification of cancer mutations (11), may suffer or benefit from such biases. How strong is this effect? The observable extent of the differences in fragment sizes varied strongly among the cases presented (2). Because cfDNA size profiles were derived by Jiang et al. through aggregate analysis and were dependent on undetermined amplification copy numbers, it is still difficult to extrapolate as to the effects this may have across patients with various cancers and with other measurement techniques. However, in extreme cases, such as those showcased by Jiang et al. (see figure 6 in ref. 2), one would expect to obtain a higher fraction of mutant alleles from an assay that uses short PCR amplicons compared with an estimate based on longer amplicons, hybrid-capture, or genome-wide methods. For specific loci, possible coverage biases of different analysis methods may be affected by GC content, alignment accuracy, or motif-specific cleavage (12). It would be important to bear in mind such potential biases as future applications of ctDNA become more refined.

The biology behind cfDNA fragmentation is still unclear. Fragment sizes of cfDNA are centered around 166 bp, approximately the length of DNA wrapped around a nucleosome plus its linker (10). Fragment lengths of 166 bp may result from the action of a caspase-dependant endonuclease that cleaves

Author contributions: F.M. and N.R. wrote the paper.

Conflict of interest statement: N.R. is cofounder of and consultant to Inivata Limited, and coinventor of patent applications describing methods for analysis of rare DNA fragments.

See companion article on page E1317.

[1]To whom correspondence should be addressed. Email: Nitzan.Rosenfeld@cruk.cam.ac.uk.

DNA after a core histone and its linker. The 10-bp periodicity observed for fragments smaller than 166 bp (2, 10) corresponds to a turn of the DNA helix wrapped around the core histone, which may protect one part of the DNA sequence from the action of other blood-borne nucleases. These features of the fragmentation pattern support the view that apoptosis may be a major source of cfDNA release, and that histones are probably the key protein complex associated with DNA in the blood circulation. Nucleosomes are not distributed regularly along the genome (13, 14). For example, the nucleosome density seems to be higher in exonic regions in comparison with intronic area (14), and this may be reflected in different abundance or fragment sizes of exonic compared with intronic regions. A better characterization of ctDNA associated-complexes will also be important to clarify the potential impact of ctDNA on cancer biology (15).

ctDNA levels vary widely across multiple types and stages of cancer (16), and this will be reflected in the overall size profile of plasma DNA. Short DNA fragments were found to be more frequent in the plasma of patients with metastatic compared with earlier-stage breast cancer (17), and shorter fragments were enriched in HCC patients with higher levels of ctDNA (2). Diversity in fragment-size profiles is also to be expected because of biological differences as well as metabolic effects. For example, some cancers may be more prone to release of ctDNA through necrosis, whereas plasma ctDNA from glioblastomas would reflect filtration effects of the blood–brain barrier. Observations in HCC patients may be confounded by impaired liver function, and confirmation in other cancer types will be informative.

Mitochondrial cell-free DNA (mcfDNA) released by cancer cells is also being investigated as a potential biomarker for cancer. Higher concentrations of mcfDNA compared with healthy individuals were previously observed for patients with other cancer types (18). Jiang et al. find that mcfDNA concentrations were higher in HCC patients, and had a good discriminatory power compared with healthy individuals, subjects with hepatitis B virus, and also patients with liver cirrhosis but no cancer (2). As previously reported by the same team for maternal plasma (10), mcfDNA is much more fragmented compared with nuclear cfDNA, with no peak at 166 bp and no evidence of protective effects of nucleosomes (2).

Jiang et al. extract DNA from plasma using a method that filters and captures DNA of a particular size range, and measure DNA fragment sizes by paired-end sequencing (2). Although these methods are highly relevant for emerging applications of ctDNA (1), it is important to bear in mind that they do not span the full diversity of circulating DNA. DNA fragments that lie outside the studied range, in particular short fragments that are difficult to extract or assay, may yet play an important role in our understanding of cfDNA biology and in future applications. Long fragments may reach circulation through nonapoptotic mechanisms and may prove to be informative or biologically active. Size profiles of cfDNA in plasma, without hybridization-based extraction, showed very different patterns of fragmentation (19). Direct visualization of DNA fragments, PCR-based approaches, and animal model systems, can be used as a complementary tool to study smaller DNA fragments (3, 20). It would be useful to perform comparative studies that will apply a range of methods to the same samples and to study serial samples from the same patient (or model system) during tumor progression or treatment.

A better understanding of the biology of circulating DNA may lead to the optimization or the development of new approaches. If nucleosomes play a central role in the biology of cfDNA, they can be useful in extracting or enriching for ctDNA, and nucleosome occupancy or positioning along the genome could be informative in assay design. Identifying loci of preferential fragmentation could help avoid sequences that would be inefficient for primer binding for PCR or targeted sequencing. Recently, Lo and colleagues developed a size-selection assay for prenatal diagnostics (21), using their prior finding that fetal cfDNA is more fragmented than maternal cfDNA in maternal plasma (10). Similarly, if ctDNA of patients with high tumor fraction content has a shorter size profile, a size selection of short DNA fragment may improve the sensitivity for detection of ctDNA alterations.

ACKNOWLEDGMENTS. The authors thank the University of Cambridge, Cancer Research UK, and Hutchison Whampoa Limited for support.

1 Diaz LA, Jr, Bardelli A (2014) Liquid biopsies: Genotyping circulating tumor DNA. *J Clin Oncol* 32(6):579–586.
2 Jiang P, et al. (2015) Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci USA* 112:E1317–E1325.
3 Mouliere F, et al. (2011) High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* 6(9):e23418.
4 Umetani N, et al. (2006) Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol* 24(26):4270–4276.
5 Thierry AR, et al. (2010) Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res* 38(18):6159–6175.
6 Fujimoto A, et al. (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 44(7):760–764.
7 Leary RJ, et al. (2012) Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* 4(162):ra154.
8 Heitzer E, et al. (2013) Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int J Cancer* 133(2):346–356.
9 Kan Z, et al. (2013) Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res* 23(9): 1422–1433.
10 Lo YM, et al. (2010) Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med* 2(61):61ra91.
11 Forshew T, et al. (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 4(136):36ra68.
12 Chandrananda D, et al. (2014) Investigating and correcting plasma DNA sequencing coverage bias to enhance aneuploidy discovery. *PLoS ONE* 9(1):e86993.
13 Struhl K, Segal E (2013) Determinants of nucleosome positioning. *Nat Struct Mol Biol* 20(3):267–273.
14 Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16(9):990–995.
15 Dvořáková M, et al. (2013) DNA released by leukemic cells contributes to the disruption of the bone marrow microenvironment. *Oncogene* 32(44):5201–5209.
16 Bettegowda C, et al. (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 6(224): 24ra24.
17 Madhavan D, et al. (2014) Plasma DNA integrity as a biomarker for primary and metastatic breast cancer and potential marker for early diagnosis. *Breast Cancer Res Treat* 146(1):163–174.
18 Ellinger J, et al. (2012) Circulating mitochondrial DNA in serum: A universal diagnostic biomarker for patients with urological malignancies. *Urol Oncol* 30(4):509–515.
19 Breitbach S, et al. (2014) Direct quantification of cell-free, circulating DNA from unpurified plasma. *PLoS ONE* 9(3):e87838.
20 Mouliere F, El Messaoudi S, Pang D, Dritschilo A, Thierry AR (2014) Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol* 8(5): 927–941.
21 Yu SC, et al. (2014) Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci USA* 111(23):8583–8588.