Biological Fragmentation of Circulating Cell-free DNA Alters Genetic Representation

Chen Hsiung Yeh*

Circulogene Theranostics, Birmingham, USA

***Corresponding author:** Chen Hsiung Yeh, Circulogene Theranostics, Birmingham, USA.

Short Communication

High-throughput sequencing of circulating cell-free DNA (cfDNA) as liquid biopsy has revolutionized tumor genome profiling by providing a more accurate, longitudinal, real-time and non-invasive mean for precision and personalized medicine. Current knowledge on cfDNA characteristics revealed that it exists mainly as double-stranded molecules, resulting from biological fragmentation into both short (<1 kb) and long segments (>10 kb) [1,2]. Short fractions are mostly derived from apoptosis via the activation of cellular endonucleases leading to the cleavage of chromatin DNA into inter-nucleosomal fragments [3], whereas necrosis generates relatively long fragments of DNA. It is now believed that circulating cfDNA pool of cancer patients is originated from a combination of apoptosis, necrosis, and active release [4]. Within the nucleosomal core, cfDNA is protected from blood nucleases by histones, whereas the linker is vulnerable to digestion. As a result, regions showing high or low frequency of fragmentation correspond to the sequences between or occupied by nucleosomes, respectively, with the majority of short-fragment length corresponds to single nucleosome size of 160-170 bp. These observations strongly support the notion that the patterns of cfDNA fragmentation are guided by chromatin structure particularly the interplay between nucleosome positioning, epigenetic regulation, and gene expression machinery [5,6].

Furthermore, cfDNA is highly heterogeneous since it represents numerous different tissues each of which has its own gene expression profiles. Indeed, biological fragmentation of nucleosome-bound DNA is never random leading to biased representation of cfDNA sequences and unbalanced read coverage and uniformity, especially near genomic regions of transcription start sites and exonic boundaries, where nucleosome positioning is highly phased [6,7]. Hence, it is believed that chromatin changes associated with loci overall expression level contribute

to the cfDNA fragmentation pattern, i.e., cfDNA patterning reflects a general picture of gene expression [8]. Accordingly, gene expression directed patterns of cfDNA fragmentation could have important impacts for next-generation sequencing (NGS) analysis. The required level of resolution for a NGS assay is achieved by providing sufficient coverage, which generally refers to the average number of reads that align to each base within the targeted gene regions. Theoretically, the coverage uniformity among different loci/alleles should be high and even to make calls with confidence. However, not all genes are born equally in terms of functionality and chromosome location, i.e., biological representation bias exists.

Information about the nature and mechanism of cfDNA fragmentation in their chromatin context is essential to understand not only the genetic representation but also the complex interactions that are responsible for tumor chromatin architecture. Changes in gene expression can alter dynamic chromatin states: gene activity is usually low with condensed and higher order packaged chromatin, whereas active gene expression always leads to relaxed and open chromatin where DNA fragmentation readily occurs (Figure 1).

Therefore, cfDNA fragmentation is not a random process evenly spread across the entire genome. As a result, it is expected that some genes in fragmented cfDNA are over-represented and others are under-represented. Most importantly, the real picture of uneven genetic representation of cfDNA caused by non-random fragmentation is greatly hindered by current DNA extraction methodology that depleted nucleosomes and other DNA-protein complexes in order to reach high sequencing uniformity.

The disparity in genetic representation in cfDNA fragments has been well documented [9-12]. Studies have shown that repetitive sequences Alu and certain satellite markers were found to be over-represented, while L1 and L2 repeats were under-represented in the cell-free apoptotic DNA. L1 elements are mainly located in

the transcriptionally inactive heterochromatin and Alu repeats are associated with gene-rich euchromatin regions that have high frequency of gene expression.

Examination of the fragment lengths present in cfDNA following silica extraction, we corroborated the findings in literature that fragmentation is primarily between nucleosomes with subsequent intra-nucleosomal cleavage along the DNA helical turn [13]. The fragment sizes corresponding to mono-, di- and tri-nucleosomal subunits appear to be prevalent (Figure 1). Only around 10-20 %

of the cfDNA population are 170 bp increments with a laddering pattern, which is related to DNA wrapped around multiple nucleosomes and protected from nuclease cleavage. The higher abundance of longer fragment cfDNA population (>1 kb) could be due to the regions being generally inaccessible by enzymes because of the dense packaging. In addition, utilizing NGS analysis with each purified fraction, we are able to show that longer fragments exhibited much higher coverage uniformity, on-target rate and mean depth than shorter fragments (Table 1).
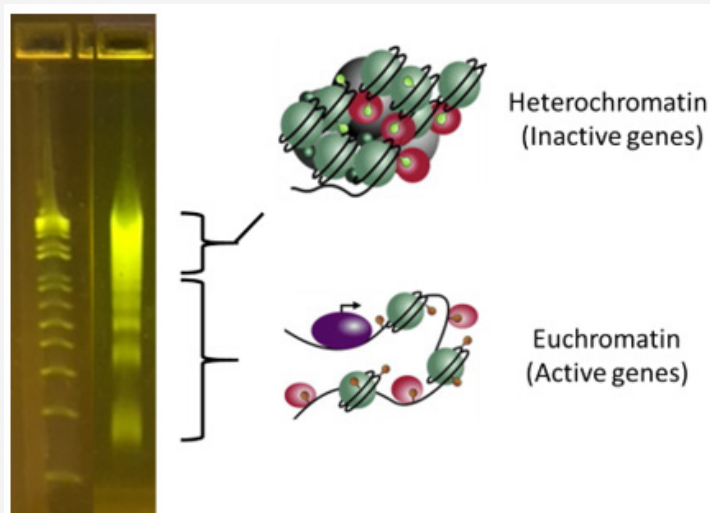

Figure 1: Fractionation of Qiagen silica prepared cfDNA in a 2% agarose gel. The longer fragments (>10kb) are enriched by heterochromatin which contains mainly inactive genes. The shorter fragments (<1kb) represent euchromatin with predominantly active genes. The fragment sizes of DNA laddering pattern correspond to mono-, di-, tri-, and multi-nucleosomal subunits were evident. DNA fragments of >10 kb, 300-500 bp and 170 bp were individually excised, purified and then analyzed by next-generation sequencing.

Table 1: Next-generation sequencing characteristics of size-selected cfDNA by Qiagen silica extraction.

Qiagen Prep cfDNA	Mapped Reads	ON-target	Mean Depth	Uniformity
No size-selectin	2,589,412	97.67%	11,684	95.76%
Long fragments (>10 kb)	2, 251, 411	98.87%	11,545	93.10%
Medium fragments (300-500 bp)	1,217,436	11.19%	890	70.86%
Short fragments (170) bp	2,123,795	10.01%	472	61.90%

These inconsistencies and chromosome position effect in cfDNA sequencing data when averaging across the genome and read coverage imbalances between different regions need to be taken into account. Our findings here illustrated the artificial bias incurred by standard silica-based extraction methodology which captures both long and short cfDNA fragments, although achieving overall high sequencing uniformity but losing the real in vivo picture and representation complexity. Uneven coverage in cfDNA sequencing data does not vary simply due to technical biases such as GC-content and read mappability along the genome but largely due to non-random DNA fragmentation. The imbalance in sequence representation in NGS data could be the end-product from a complex mixture of cfDNA in circulation due to various factors i.e., different proportions of apoptotic/necrotic input, endo- and exo-nuclease activity and different tissue origins, all of which warrant further investigation. The extraction bias also impacts copy-number variation (CNV) analysis where CNVs in certain regions would be harder to detect simply due to lack of cfDNA fragments

originating from these regions despite the high uniformity of sequencing. It is thus of great benefit to develop alternative cfDNA preparation approaches to preserve this fine scale unbalanced genetic representation for higher sensitivity and more accurate mutation detection in liquid biopsy.

Here we deployed an extraction-free in situ cfDNA sample preparation technology [14-16] to eliminate extraction bias and systematically analyzed NGS sequencing coverage patterns across multiple patients and gene regions. The NGS gene panel consists of 207 amplicons (y-axis), representing portions of 50 genes. The global heatmap pattern showed strikingly similar sequencing coverage profiles across various patient samples (x-axis) in 2 independent runs (Figure 2). Sequencing coverage statistics was assessed and categorized into >5,000X (green), 100X-5,000X (yellow), and <100X (red) for each of the 207 amplicons. Our analysis highlighted 7 genes (VHL, FGFR3, CDKN2A, NOTCH1, HRAS, AKT1 and STK11) contained amplicons having consistently

lower-than-average coverage, demonstrating unbalanced genetic representation. In contrast, two silica-purified control DNA samples in each run (Figure 2, the farthest right 2 columns in each heatmap) lost the genetic imbalance and reached high uniformity. Our study showed that there is considerable extraction bias in cfDNA sequencing data that could be harnessed to improve existing NGS analysis toward consistently high sensitivity and improve detection

accuracy. Since the prevalence of cfDNA fragments may depend on the chromatin positioning at given DNA locus, PCR primer sets, or target-capturing probes may need be tuned specifically to those less-fragmented hotspot mutation regions. Unfortunately, this avenue for cfDNA-based fragmentation remains unexplored. The biological fragmentation-driven expedition will certainly open a novel field in liquid biopsy.
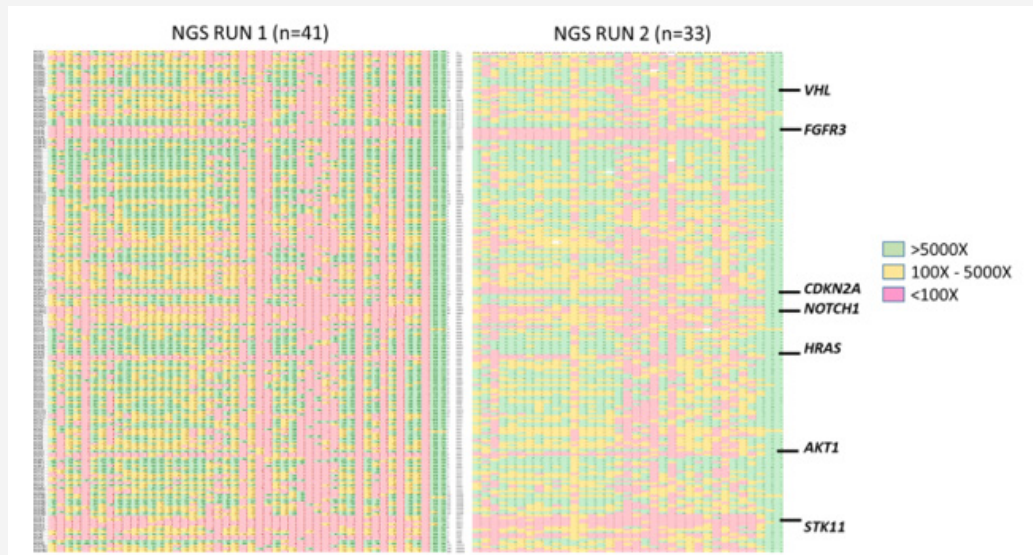

Figure 2: Next-generation sequencing coverage heatmap of cfDNA samples prepared by a proprietary extraction-free in situ technology highlighted unbalanced representation. The NGS gene panel consists of 207 amplicons (y-axis) which encompass hotspot mutation regions within 50 cancer driver genes. The global heatmap pattern showed strikingly similar sequencing coverage profiles across various patient samples (x-axis) in 2 separate runs. Sequencing coverage statistics was classified into >5,000X (green; over-represented), 100X-5,000X (yellow; average-represented), and <100X (red; under-represented) for each of the 207 amplicons.

Acknowledgment

Conflict of Interest

No conflict of interest.

References

1. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, et al. (2001) DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 61(4): 1659-1665.

2. Li Y, Zimmermann B, Rusterholz C, Kang A, Holzgreve W, et al. (2004) Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. Clin Chem 50(6): 1002-1011.

3. Zhivotosky B, Orrenius S (2001) Assessment of apoptosis and necrosis by DNA fragmentation and morphological criteria. Biochem Pharmacol 66: 1527-1535.

4. Delgado PO, Alves BC, Gehrke Fde S, Kuniyoshi RK, Wroclavski ML, et al. (2013) Characterization of cell-free circulating DNA in plasma in patients with prostate cancer. Tumor Biol 34(2): 983-986.

5. Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, et al. (2016) MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nat Commun 7: 11485.

6. Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V (2015) Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. BMC Genomics 16: S1

7. Ma X, Zhu L, Wu X, Bao H, Wang X, et al. (2017) Cell-Free DNA provides a good representation of the tumor genome despite its biased fragmentation patterns. PLoS ONE 12: e0169231.

8. Ford A, Brown C, Yeh CH (2017) When liquid biopsy cell-free DNA meets with tissue biopsy RNA. Biomed J Sci Tech Res 1: 256-257.

9. Beck J, Urnovitz HB, Riggert J, Clerici M, Schütz E (2009) Profile of the circulating DNA in apparently healthy individuals. Clin Chem 55(4): 730-738.

10. Morozkin ES, Loseva EM, Morozov IV, Kurilshikov AM, Bondar AA, et al (2012) A comparative study of cell-free apoptotic and genomic DNA using FISH and massive parallel sequencing. Expert Opin Biol Ther 12: S11-7.

11. Stroun M, Lyautey J, Lederrey C, Mulcahy HE, Anker P (2001) Alu repeat sequences are present in increased proportions compared to a unique gene in plasma/serum DNA: evidence for a preferential release from viable cells? Ann NY Acad Sci 945: 258-264.

12. Van der Vaart M, Pretorius PJ (2008) A method for characterization of total circulating DNA. Ann NY Acad Sci 1137: 92-97.

13. Tsui NBY, Jiang P, Chow KCK, Su X, Leung TY, et al. (2012) High resolution size analysis of fetal DNA in the urine of pregnant women by paired end massively parallel sequencing. PLoS One 7(10): e48319.

14. Ford A, Brown C, Yeh CH (2017) Sample preparation of circulating cell-free DNA by direct-on-specimen and silica-based methods. J Biol Sci 3: 23-35.

15. Ford A, Brown C, Yeh CH (2018) Pre-analytical assessment of circulating cell-free DNA prepared by an isolation-free enrichment technology. ACTA Sci Cancer Biol 2: 2-6.

16. Ford A, Brown C, Caver E, Yeh CH (2018) The evolution of in situ genetic technology. J Clin Res Oncol 1: 1-5.