Library preparation from circulating cell-free DNA

Summary

- Circulating cell-free DNA (cfDNA) is becoming an increasingly important marker for assessing many cancer types
- cfDNA extraction and library preparation for next-generation sequencing (NGS) is challenging, as only very limited quantities of material are present in blood samples available for research, and the DNA is degraded
- The Invitrogen™ Collibri™ PS DNA Library Prep Kit for Illumina™ Systems enables efficient cfDNA library generation for sequencing on Illumina systems, with a few modifications to the standard protocol

Introduction

The use of liquid biopsies for research has become increasingly prevalent to investigate potential biomarkers within genetic material from tumors circulating in peripheral blood. Elevated levels of cfDNA in peripheral blood are common with many cancers, due to the increased rate of apoptosis in many tumor cells.

Isolation of cfDNA and preparation of sequencing-ready libraries from cfDNA present a number of technical challenges. Despite cfDNA being elevated in some oncology research samples, the concentration of material available is still in the low ng/mL range [1]. Depending on the sample type, there is also considerable variation in the amount of cfDNA, and the difficulty of assessing this prior to extraction complicates workflow standardization. This issue is further compounded by the presence of genomic DNA (gDNA) from white blood cells, which can be present in plasma samples due to mechanical damage caused during sample processing and storage. The goal is therefore to recover the shorter, fragmented DNA while leaving the larger gDNA molecules, enriching the cfDNA for downstream analysis using NGS.

Another major challenge when preparing libraries from cfDNA is the low quality of the input material. During apoptosis, cleavage of DNA primarily occurs between nucleosomes, resulting in fragments of roughly 170 bp [2]. The DNA is then exposed, in peripheral circulation, to conditions that can lead to further degradation. The combination of low input material, gDNA contamination, and intrinsic degradation makes efficient and reproducible extraction of cfDNA essential to the performance of liquid biopsy samples in oncology research.

The Applied Biosystems™ MagMAX™ Cell-Free DNA and Cell-Free Total Nucleic Acid Isolation Kits use magnetic bead–based technology to purify enriched cfDNA or cell-free total nucleic acids from plasma, serum, or urine samples, without gDNA contamination. To help save time and increase reproducibility, MagMAX kits can be combined with Thermo Scientific™ KingFisher™ instruments for automated purification.

Researchers at the Central European Institute of Technology at Masaryk University in the Czech Republic—in collaboration with the EMBL Genomics Core Facility, Heidelberg, Germany—have recently developed a protocol for efficient cfDNA library preparation to investigate hemato-oncological disorders, mainly chronic lymphocytic leukemia (CLL). Combining this method with whole-genome sequencing (WGS), they have been able to observe genomic abnormalities that are specific to stages of disease, helping to guide their further research. This application note highlights the approach used in one of their studies.

Methods

cfDNA isolation and library preparation

Eleven human plasma samples from different research subjects representing various stages of CLL were studied. cfDNA was isolated from the plasma samples using the MagMAX Cell-Free DNA Isolation Kit, and libraries were prepared from the cfDNA samples using the **Collibri PS DNA Library Prep Kit for Illumina Systems**. Samples (1 µL) of the isolated cfDNA and prepared libraries were evaluated for quality on an Agilent™ Bioanalyzer™ system using the Agilent™ High Sensitivity DNA Kit (Figure 1). All steps of the isolation and library preparation workflow were performed according to the manufacturer's protocol, with the following modifications to the library preparation:

- To reflect the fragmented nature of the input material, the purified size range was adjusted from the 150–1,000 bp recommended in the product specifications to a desired insert size of 170 bp.
- Amounts tested as input varied from 1 ng to 10 ng of cfDNA (4 samples in duplicates). As 1 ng of cfDNA provided sufficient yields, this amount was used for further experiments.
- Due to the fragmented nature of the input material, it was not subjected to further fragmentation.
- A one-sided cleanup, option A in the user guide, was used for post-ligation purification. Post-ligation double-sided size selection, option B, was not used due to the short insert size.
- For post-ligation cleanup, a bead-to-sample volume ratio of 1.1:1 (77 µL of beads to 70 µL of reaction) was used to remove unligated adapters and to increase yield, instead of the ratio of 0.64:1 recommended for gDNA (45 µL of beads to 70 µL of reaction).
- The bead-to-sample ratio of 1.1:1 was also used for purification of the amplified DNA library (55 µL of beads to 50 µL of reaction).

Sequencing libraries from corresponding normal and primary tumor gDNA were prepared following the standard protocol with double-sided size selection for post-ligation purification.

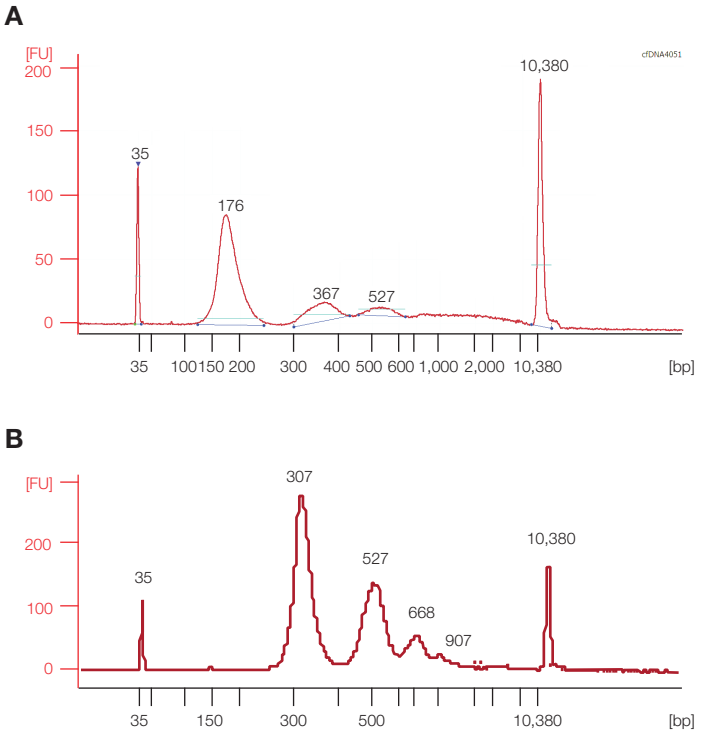

Figure 1. Size distribution of isolated cfDNA or prepared library. The peaks at 35 and 10,380 bp are size markers. **(A)** Isolated cfDNA (0.6 ng/µL) showing the expected prominent peak at 176 bp. **(B)** Prepared library (13.6 ng/µL) showing the correct size distribution, including a prominent peak at 307 bp that reflects the cfDNA plus adapters.

NGS and data analysis

The prepared libraries were sequenced on an Illumina™ NextSeq™ 550 System using the NextSeq™ 500/550 High Output Kit v2.5 (150 cycles). Four cfDNA sequencing libraries were pooled at equimolar ratios for each run on the NextSeq 550 System and sequenced in paired-end mode (2 x 75 bp). Base calling was carried out using Illumina™ bcl2fastq software. Sequenced reads were aligned to the hg19 human genome build using Burrows-Wheeler Aligner (BWA) software [3]. Alignment files were sorted and indexed using SAMtools software [4]. Quality control on alignments and computation of genome-wide read-depth profiles were performed using Alfred software [5].

Results

Median coverage of libraries prepared from cfDNA of 6x was reached after deduplication; the median PCR duplicate rate was 11%. Somatic copy-number alterations (SCNAs) in the primary tumor samples were compared to the cfDNA samples, using custom scripts. Specifically, analyses were focused on clonal SCNAs in the primary tumor gDNA to investigate if the same abnormalities were identified in the cfDNA (Figure 2). Specific SCNAs in cfDNA could be correlated to disease stages. In the majority of cases with active disease, tumor-associated SCNAs were detected despite low white blood cell counts. In samples of inactive disease, tumor-associated SCNAs were undetectable in cfDNA.

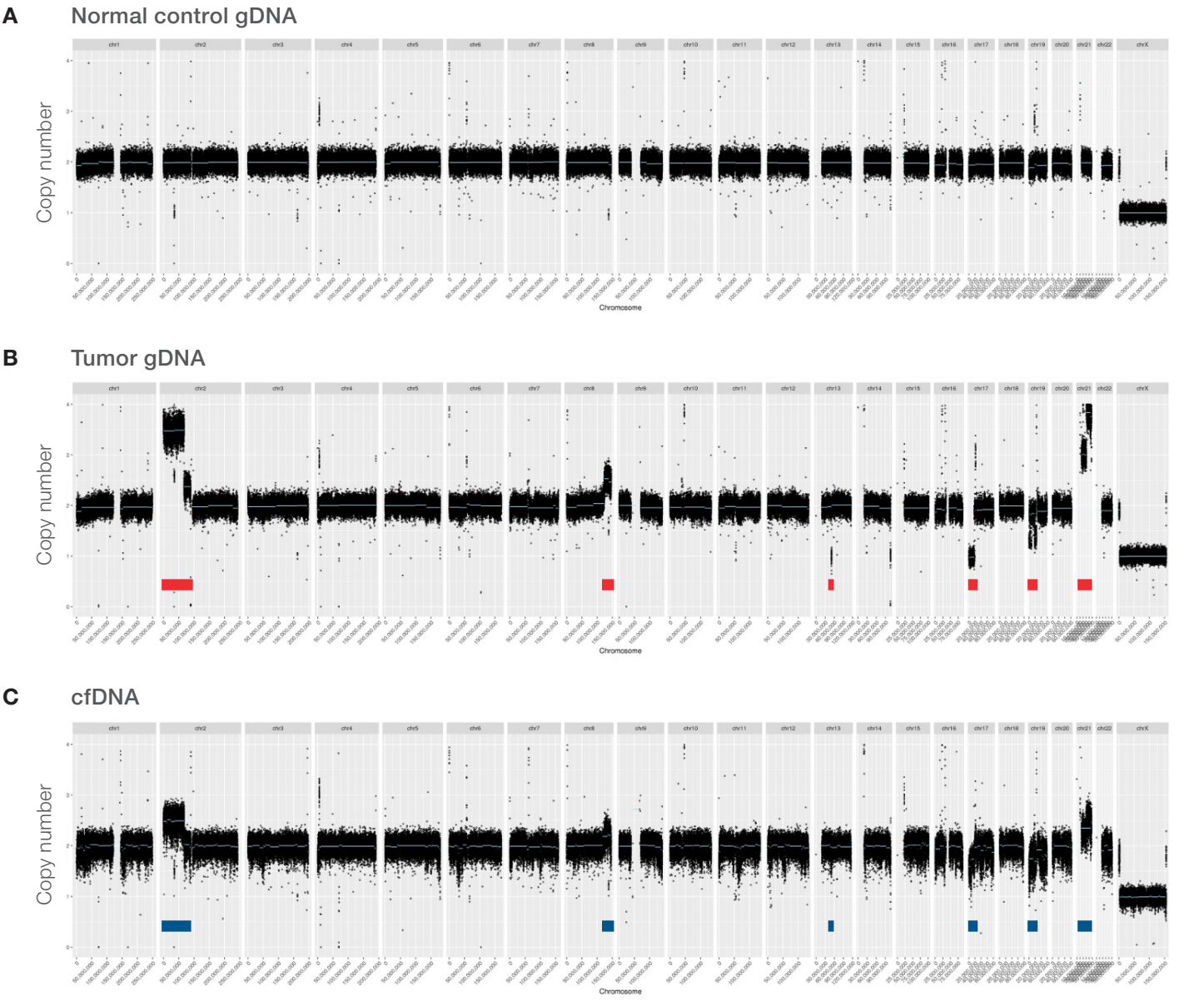

Figure 2. Detection of SCNAs in corresponding donor DNA samples. (A) Normal gDNA of a single donor, with no SCNAs detected. **(B)** Tumor gDNA from sorted CLL cells of the same donor, showing SCNAs (red bands). **(C)** cfDNA of the same donor, showing the corresponding SCNAs (blue bands).

invitrogen

Conclusions

- cfDNA analysis has emerged as a potential method for studying hemato-oncological disorders at various stages of assessment.
- WGS is suitable for the study of disease-specific genomic abnormalities in cfDNA.
- The Collibri PS DNA Library Prep Kit for Illumina Systems provides consistent and reproducible cfDNA library preparation from human plasma samples.

References

1. Bronkhorst AJ, Vida Ungerer V, Holdenrieder S (2019) The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif* 17:100087.
2. Snyder MW, Kircher M, Hill AJ et al. (2016) Cell-free DNA comprises an *in vivo* nucleosome footprint that informs its tissues-of-origin. *Cell* 164:57–68.
3. Durbin R, Li H (2009) Fast and accurate short-read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
4. Li H, Handsaker B, Wysoker A et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
5. Rausch T, Hsi-Yang Fritz M, Korbel JO et al. (2019) Alfred: interactive multi-sample BAM alignment statistics, feature counting, and feature annotation for long- and short-read sequencing. *Bioinformatics* 35:2489–2491.

Find out more at **thermofisher.com/collibriDNA**

ThermoFisher SCIENTIFIC