Options for Control Materials for NGS Analysis of Circulating Cell-Free DNA

Comparison of nucleosomal DNA and fragmented genomic DNA (gDNA).

Introduction

The analysis of circulating cell-free DNA (cfDNA) in cancer research holds promise for the development of noninvasive diagnostics and monitoring. To provide accurate information, reference materials are needed for benchmarking lab procedures and data quality. Traditionally, cfDNA analysis has employed shearing of gDNA to generate small fragments to use as reference materials. However, DNA sonication may cause DNA oxidation leading to artifactual base changes in sample DNA.[1] While oxidation-induced errors may be rare, false positives become more problematic as cancer researchers strive to lower the limit of detection for variant calling.

cfDNA is thought to be generated *in vivo* by endonuclease cleavage of chromatin, leading to a narrow fragment distribution around 170 bp.[2] Therefore, a method to digest nucleosomal DNA *in vitro* may be useful for making control materials that are biochemically similar to cfDNA. This technical note describes options for generating cfDNA control materials and compares the performance of cfDNA, nucleosomal DNA, and purified gDNA fragmented by either sonication or enzymatic digestion.

Methods

DNA preparation

cfDNA: Blood from healthy individuals (n = 7) was collected in Cell-Free DNA BCT blood collection tubes (Streck), and DNA was extracted with the QIAamp Circulating Nucleic Acid kit (QIAGEN).

Nucleosomal DNA: Nucleosomal DNA was generated from cultured human cells (n = 10), using the EZ Nucleosomal DNA Prep Kit (Zymo Research) following manufacturer instructions with the modification of extending enzyme incubation times from 20 to 30 minutes with Atlantis double-stranded DNase. Nuclei were isolated from one million cultured human cells and, following enzyme treatment, nucleosome-protected DNA was column purified. Approximately 1 µg of nucleosomal DNA was typically obtained from one million cells, although yield may vary between cell lines.

Digested gDNA: gDNA samples (n = 7) were digested using NEBNext dsDNA Fragmentase (New England Biolabs). One µl of fragmentase was used per 1 µg gDNA. Dual bead-based size selection was performed with the Agencourt AMPure XP system (Beckman Coulter) using a 1× and 0.4× bead:DNA ratio for the first and second selection, respectively. About 5% of input DNA was recovered.

Sheared gDNA: gDNA samples (n = 10) were mechanically sheared using the M220 Focused-ultrasonicator (Covaris), and the sheared DNA was size-selected to 130–210 bp with a targeted peak at 170 bp using the BluePippin system (Sage Science). DNA was then purified using the Agencourt AMPure XP system (Beckman Coulter) as previously described. About 10% of input DNA was recovered.

Quantification of cfDNA and cfDNA control materials

The concentration and size distribution of all DNA samples were evaluated on a Fragment Analyzer using the High Sensitivity Large Fragment Analysis Kit (Advanced Analytical). DNA concentration was measured using the 50–700 bp range to approximate the cfDNA-like fraction.

Library preparation and sequencing

For cfDNA, nucleosomal DNA, and digested gDNA, 30 ng DNA were used for library preparation. Due to an observed lower library conversion rate, 75 ng were used for sheared DNA. Libraries were prepared according to the reference guides for the TruSight Oncology UMI Reagents[3] and TruSight Tumor 170[4] following the DNA workflow only. Libraries were sequenced on the HiSeq™ 4000 System with eight multiplexed libraries per flow cell, resulting in ~40,000× raw coverage per sample.

Data analysis

Sequencing data were analyzed using an Illumina proprietary pipeline for error suppression. Using unique molecular identifiers (UMIs),[5] PCR duplicates from the same DNA template were identified as families. To prevent overcounting, median target coverage (MTC) is calculated based on number of fragments spanning the target region instead of number of reads. To calculate error rate, positions with nonreference allele frequency (< 5% or > 95%) were collected, excluding known 1000 Genome SNPs assumed to be potentially erroneous regions.[6] In the defined positions,error rate was calculated as the number of variant bases divided by total bases.

Results

All DNA types had similar size distributions (Figure 1) with peaks close to expected size of ~170 bp for cfDNA. Size selection with the BluePippin system resulted in a narrower size distribution for sheared gDNA. Nucleosomal DNA showed characteristic patterns of cfDNA, such as peaks spaced ~10 bp apart reflecting helical pitch of nucleosome-bound DNA, and a peak near 300 bp reflecting dinucleosomes. These chromatin-based properties were not present in the gDNA controls (Figure 1).

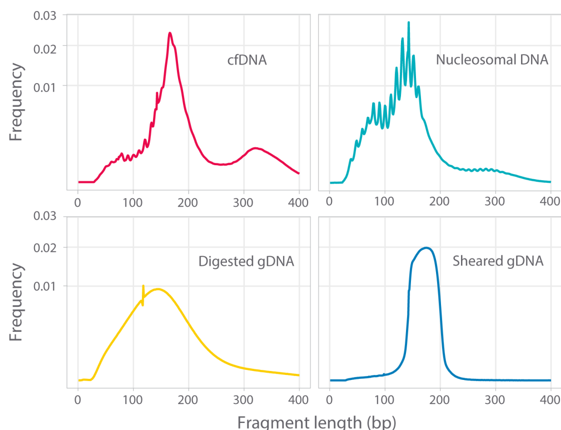

Figure 1: Fragment length distribution—DNA from four preparation methods were evaluated on a Fragment Analyzer.

MTC was comparable between sample types (Figure 2), and values were all above 1500×, which is recommended by Illumina for calling low-frequency variants. Recovery of both strands in the double-stranded DNA template (duplex rate) is valuable for error correction, enabling the filtering of variants that do not occur on both strands. In libraries from sheared DNA, the observed duplex rate was lower than with other control materials (Figure 3).

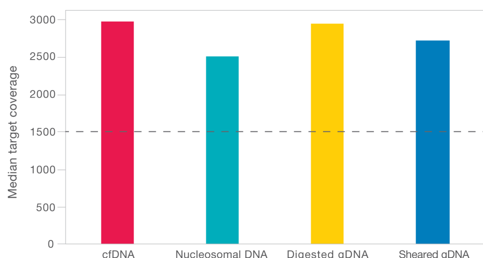

Figure 2: MTC comparison across different DNA materials—For variant calling as low as 0.4% variant allele frequency, a value of ≥ 1500× (indicated by dashed line) is recommended.

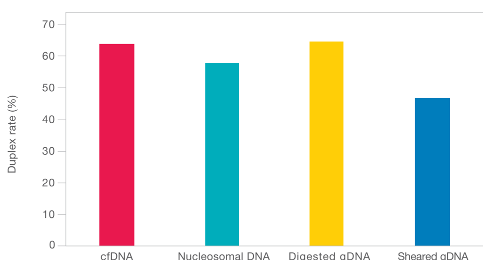

Figure 3: Duplex rate across different DNA materials

The overall error rate was also higher in sheared gDNA (Table 1). Although this may have partially resulted from lower duplex rate, the specific error rate for C to A base changes was significantly higher in sheared gDNA, which is a characteristic artifact from DNA oxidation that could have resulted from sonication (Table 1).

Table 1: Error rate for each error type

	Error Rate ($\times 10^{-5}$)			
Error Type	cfDNA	Nucleosomal DNA	Digested gDNA	Sheared gDNA
C to T	1.6	1.4	0.9	1.0
C to G	0.4	0.1	0.2	0.8
C to A	0.6	0.3	0.7	4.0
T to G	0.3	0.1	0.2	0.1
T to C	0.8	0.9	0.9	0.3
T to A	0.4	0.5	0.3	0.2
indel	0.2	0.2	0.3	0.1
Total Error Rate	2.2	1.8	1.9	3.3

Total error rate is a weighted sum of error rates by different nucleotide changes.

Summary

Nucleosomal DNA preparation is an efficient method to generate large amounts of control material that closely resembles natural cfDNA in its size distribution and biochemical properties. Nucleosomal DNA made in the laboratory also provides researchers with the flexibility of choosing cell lines with known variants. Size selection of gDNA provides a size distribution similar to cfDNA. However, sheared gDNA also produced a higher error rate than other DNA samples, which could have been a result of lower duplex rate formation and DNA oxidation during the sonication step. Both nucleosomal DNA and digested gDNA produced error rates that were lower and more similar to cfDNA than sheared gDNA.

Learn more

For more information about the TruSight Oncology UMI Reagents, visit www.illumina.com/UMI-Reagents

References

1. Chen L, Liu P, Evans TC, Ettwiller LM. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science*. 2017;355(6326):752-756.
2. 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(1-2):57-68.
3. Illumina (2017) TruSight Tumor 170 Reference Guide. (support.illumina.com/downloads/trusight-tumor-170-reference-guide-1000000024091.html).
4. Illumina (2018) TruSight Oncology UMI Reagents Workflow Integration Guide. (support.illumina.com/sequencing/sequencing_kits/trusight-oncology-umi-reagents.html).
5. Illumina (2018) TruSight Oncology UMI Reagents. (www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/trusight-oncology-umi-reagents-datasheet-1170-2017-007.pdf).
6. The International Genome Sample Resource. www.internationalgenome.org/category/dbsnp/. Accessed September 27, 2017.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

© 2018 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html. Pub. No. 1170-2017-020-A QB 5639

illumina®

For Research Use Only. Not for use in diagnostic procedures.