Application Note

Genomics

Agilent

Trusted Answers

A Comparative Analysis of DNA Shearing Methods Upstream of the Avida DNA Workflow

Benefits of enzymatic fragmentation for the Agilent Avida DNA workflow

Author

Bilan Hsue, Keith Chen, Sergey Shiryaev, Carole Durand, and Bahram Arezi,
Agilent Technologies, Inc.

Abstract

The Agilent Avida DNA workflow is optimized to maximize DNA recovery and offer high sensitivity through innovative, high-performance target enrichment technology. Here, we describe the incorporation of the Agilent SureSelect enzymatic fragmentation kit into library preparation for the Avida DNA workflow using genomic DNA (gDNA) from a cell line, formalin-compromised samples, and formalin-fixed, paraffin embedded samples. Input ranged from 10 ng to 50 ng. Our results indicate that enzymatic fragmentation generates higher library yields, and so, an increased number of unique molecules compared to mechanical shearing. Both fragmentation methods show similar values for mapped reads, insert size, percent on-target, and coverage uniformity. Furthermore, we verified the detection of variants with allele frequencies as low as 1% using a multiplex reference formalin-compromised standard. For samples of low concentration and in the absence of vacuum centrifugation, an alternate protocol was developed to allow higher fragmentation volume. Taken together, these findings demonstrate the compatibility of enzymatic shearing with reliable, high-quality sequencing outcomes.

Introduction

The Agilent Avida DNA reagent kit for targeted DNA sequencing features innovative, high-performance target enrichment technology optimized for maximum DNA recovery and highly sensitive low-frequency variant detection.[1,2] Currently, the kit recommends only mechanical shearing for DNA fragmentation. While effective, mechanical shearing requires costly equipment and can lead to sample loss and potential DNA damage.[3] This study aims to demonstrate that enzymatic fragmentation is compatible with Avida library preparation and target enrichment, further streamlining the protocol.

Here, we describe the incorporation of enzymatic shearing into the workflow for the Avida DNA reagent kit using the Agilent SureSelect enzymatic fragmentation kit, a simple, rapid, and streamlined single-tube process (Figure 1). We evaluated intact genomic DNA (gDNA), formalin-compromised DNA (fcDNA), and formalin-fixed, paraffin-embedded (FFPE) DNA samples. The gDNA was enriched using the Agilent Avida DNA Focused Cancer panel as well as the Agilent Avida DNA Expanded Cancer panel. The fcDNA and FFPE samples were enriched with the Agilent Avida DNA Discovery Cancer panel.

The Avida DNA Focused Cancer panel (26 kb) is the smallest cancer panel currently offered, optimized for high recovery of a focused set of cancer targets, mainly hotspots and exons, from 14 key oncogenes and tumor suppressor genes. This makes the panel ideal for the detection of low-frequency variants with a modest sequencing budget. The Avida DNA Expanded Cancer panel (345 kb) encompasses over 100 genes, 80 of which have full exonic coverage, facilitating both mutation and copy number detection. Additionally, this panel includes key intronic regions to enable translocation detection. The Avida DNA Discovery Cancer panel (2.7 Mb) covers the full exons of 682 key genes and translocation hotspot introns, making it ideal for genome profiling and biomarker assessment applications.

This study positions enzymatic fragmentation as a cost-effective and efficient alternative to mechanical shearing for Avida DNA library preparation and target enrichment. However, it should be noted that Agilent Avida Methyl and Duo Methyl reagent kits are not compatible with the SureSelect enzymatic fragmentation kit, due to the loss of methylation markers.

Experimental

Samples and preparation

Haplotype map (HapMap) gDNA (intact gDNA, NA24385) was acquired from Coriell Life Sciences, fcDNA (HD799, DNA integrity number [DIN] 4.2), a quantitative multiplex reference standard, was obtained from Horizon Discovery, and two real FFPE DNA samples (lung, DIN 7.2 and breast, DIN 3.4) were extracted from FFPE tissue blocks (BioChain Institute Inc). The quality of the FFPE lung sample surpassed that of the FFPE breast sample based on the DIN assessed using the Agilent 4200 TapeStation system and Agilent gDNA ScreenTape assay. Three replicates of 10 ng intact gDNA, 20 ng fcDNA, and 2 replicates from each real FFPE DNA sample were tested.

DNA fragmentation, targeting a mean fragment size of approximately 200 bp, was performed using either mechanical or enzymatic shearing. Mechanical shearing was conducted in 50 µL low-Tris-EDTA (TE) buffer with a Covaris E220 according to the Avida DNA reagent kit protocol (part number G9409-00000). Enzymatic shearing was performed using the SureSelect enzymatic fragmentation kit (part number 5191-4080) following the recommended protocol (part number G9702-90050). Briefly, the intact gDNA was sheared in a 10 µL reaction volume for 10 minutes at 37 °C, followed by 5 minutes at 65 °C. The fcDNA, FFPE lung, and FFPE breast samples were incubated at 37 °C for 15 minutes. After the enzymatic shearing step, 40 µL of water was added to the sheared DNA (Figure 1).

To enable a higher fragmentation volume of 50 µL, the enzyme volume was increased to 1.5 µL, the fragmentation buffer volume to 5 µL, and DNA volume to 43.5 µL. In this adapted protocol, no additional water was added after fragmentation. Accordingly, the incubation times at 37 °C were optimized to 25 minutes for intact and FFPE samples of medium to high quality (DIN ≥ 3) and 5–15 minutes for samples of low quality (DIN < 3). Please note that using higher amounts of enzyme and buffer per reaction will reduce the total number of reactions available based on the kit's label. All experiments in this study, except for those shown in Table 4 and Figure 6, were performed using standard 10 µL enzymatic fragmentation conditions.

The fragmented intact gDNA and fcDNA libraries were then prepared and target enriched using the off-the-shelf Avida DNA reagent kit, while the FFPE DNA libraries were prepared and enriched using the Agilent beta Avida DNA reagent kit. All library preparation was performed following the manufacturer's protocol for the Avida DNA kit (part number G9418A). For hybridization, the Avida DNA Focused Cancer panel (part number 5280-0050), Expanded Cancer panel (part number 5280-0047), or Discovery Cancer panel (part number 5280-0044) was used.

The captured DNA was washed and amplified to incorporate sample indexes. The final libraries were then pooled and sequenced at 2 x 150 on the Illumina NovaSeq6000. Sequencing depths were adjusted based on panel size to ensure that the maximum number of unique molecules was sequenced, as illustrated in each figure or table.

Data analysis

Alignment of FASTQ files was performed using the BWA (Burrows-Wheeler Aligner), followed by unique molecular identifier (UMI)-aware deduplication. Single-stranded UMI consensus reads were generated and used to calculate the following sequencing metrics: number of unique molecules per target region, insert size, uniformity, GC coverage, and variant detection. All other metrics presented here were calculated without using UMI information. Details on the analysis pipeline can be found in the Avida DNA Targeted Sequencing Analysis guide (part number G9409-90001).

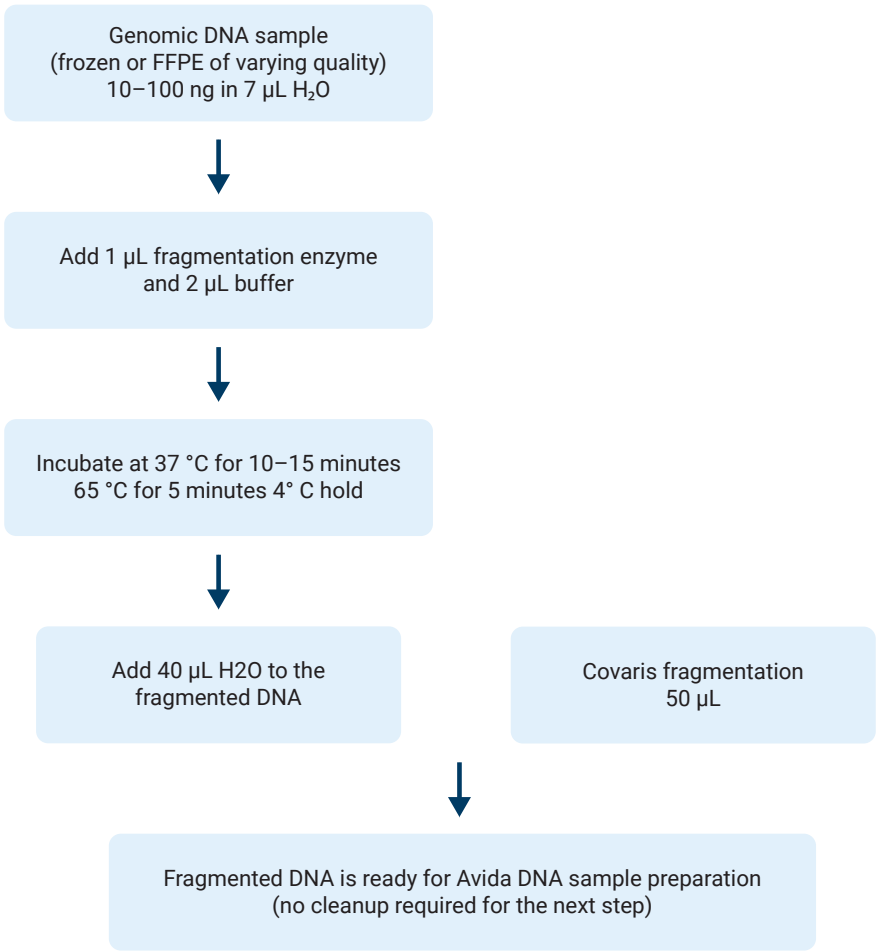

Figure 1. Overview of the fragmentation protocol using Covaris mechanical shearing or the Agilent SureSelect enzymatic fragmentation kit following the standard protocol (10 µL reaction volume).

Results and discussion

Library yields

Generating sufficient library yields from low-input amounts and low-quality DNA can be challenging, especially with mechanical shearing. This method can lead to DNA damage,[3] as well as sample loss during transfer steps. The results presented in Figure 2 show that enzymatic shearing produces greater post-capture library yields (from 1.51 to 2.97-fold higher) than Covaris shearing across different sample types (intact, formalin-compromised, FFPE lung, and FFPE breast), input amounts (10–50 ng), and probe panels (Focused, Expanded, and Discovery Cancer panels). The increased yields are likely due to the more streamlined workflow, where enzymatic shearing and end-repair/A-tailing are performed in a single tube, and/or enzymatic fragmentation generates more repairable ends compared to Covaris shearing.

Median insert size

Figure 3 shows that both shearing conditions (enzymatic and mechanical) produce similar insert sizes. The insert size for intact gDNA is larger across all sample types, ranging from 216 bp to 254 bp, while the insert sizes for formalin-compromised and FFPE samples are smaller, ranging from 170 bp to 204 bp. These results demonstrate that the enzymatic fragmentation kit is robust enough to produce libraries of insert sizes comparable to those generated by Covaris. Furthermore, this shows that the libraries are suitable for enrichment using Avida probes.

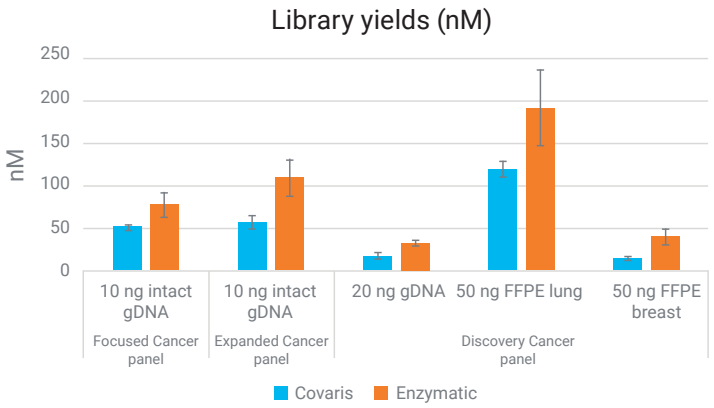

Figure 2. Enhanced library yields with enzymatic shearing compared to Covaris, quantified by the Agilent 4200 TapeStation system and Agilent D1000 ScreenTape . The post-capture library yields are from various DNA samples subjected to mechanical and enzymatic shearing. Samples include 10 ng intact gDNA enriched by either the Avida Focused or Avida Expanded Cancer panel, as well as 20 ng fcDNA, 50 ng FFPE lung, and 50 ng FFPE breast, all enriched by the Avida Discovery Cancer panel. Each data point represents the average yield from two or three technical replicates with error bars indicating standard deviations.

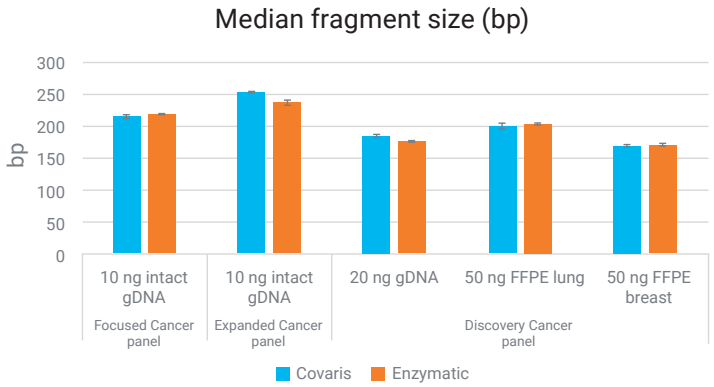

Figure 3. Similar insert sizes obtained across sample types and panels. The bar graphs represent the median insert sizes derived from Picard software. Avida Focused Cancer panel samples were downsampled to 1.74 M read pairs, Avida Expanded Cancer panel samples were downsampled to 25 M read pairs, and Avida Discovery Cancer panel samples were downsampled to 33 M read pairs. Each data point is the average of two or three technical replicates.

Coverage and complexity: number of unique molecules per target region

Figure 4A shows the median count of unique molecules per target region from 10 ng intact gDNA enriched using the Focused and Expanded Cancer panels. The number of unique molecules per target region refers to the number of unique, single-stranded DNA molecules that include the center base of each target region obtained after UMI deduplication. It is a measurement of capture efficacy for each target and reflects the overall recovery of the method. The results indicate approximately 1.8-fold more unique molecules with enzymatic shearing compared to mechanical fragmentation. Figure 4B illustrates an increase of 1.24- to 2.11-fold in the number of unique molecules for fcDNA and two FFPE DNA samples when enzymatically fragmented and enriched with the Discovery Cancer panel. This suggests greater library complexity for enzymatic shearing across all sample types and inputs tested.

Percentage of mapped reads, on-target rates, and uniformity

To provide a more general overview of the performance of enzymatic shearing compared to mechanical fragmentation, additional metrics, including the percentage of mapped reads, on-target rates, and uniformity, are shown in Table 1. The percentage of mapped reads is 99.5% or higher for both methods, indicating minimal amounts of adaptor dimer or other sequencing artifacts. Both shearing methods generated similar on-target rates across all samples with different panels. On-target is defined as the number of reads that map to within the 100 bases flanking the target region. The slightly lower on-target read percentage of the Avida DNA Expanded Cancer panel, compared to the other two panels, is likely due to the inclusion of intronic regions for fusion detection. Intronic sequences are less complex and repetitive, making probes targeting these regions prone to cross-hybridization, thus lowering the on-target rate.

Uniformity is measured by the fraction of targeted regions with coverage greater than 50% of the mean coverage across all regions. Both methods exhibit similar uniformity, except for the low-quality FFPE breast sample (DIN 3.4). In this case, Covaris shearing achieved an average uniformity of 67.02%, whereas enzymatic shearing resulted in a higher uniformity of 72.66%. Further studies are needed to investigate whether this improvement in base coverage uniformity with enzymatic shearing is reproducible across lower-quality FFPE samples.

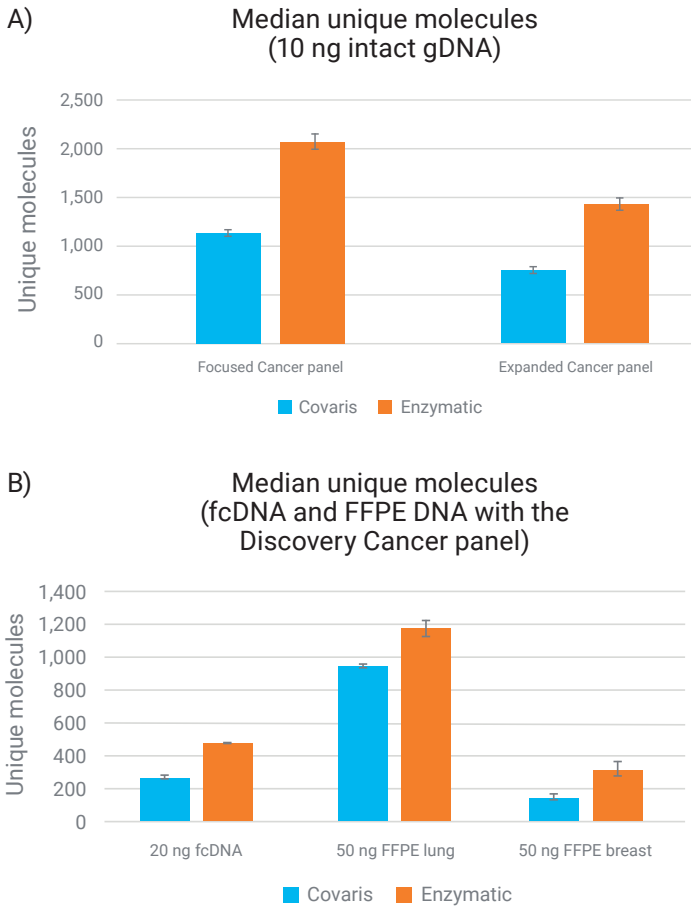

Figure 4. Enzymatic shearing generates more unique molecules across sample types and panels. (A) Unique molecules obtained from 10 ng intact gDNA and captured with the Avida DNA Focused Cancer panel (1.74 M read pairs) or the Avida DNA Expanded Cancer panel (25 M read pairs). (B) Unique molecules obtained from 20 ng fcDNA and two 50 ng FFPE DNA samples and captured with the Avida DNA Discovery Cancer panel (33 M read pairs). Each data point is the average of two or three technical replicates with error bars indicating standard deviations.

Table 1. Summary of the percentage of mapped reads, on-target rate, and uniformity. Avida Focused Cancer panel samples were downsampled to 1.74 M read pairs, Avida Expanded Cancer panel samples were downsampled to 25 M read pairs, and Avida Discovery Cancer panel samples were downsampled to 33 M read pairs. Each data point is the average of two or three technical replicates.

		% Mapped		% On-Target		% Uniformity	
Sample Type	Probe Panel	Covaris	Enzymatic	Covaris	Enzymatic	Covaris	Enzymatic
10 ng Intact gDNA	Focused Cancer	97.14	99.66	88.92	88.75	99.91	100.00
	Expanded Cancer	99.50	99.66	73.43	72.99	96.21	95.74
20 ng fcDNA, DIN 4.2	Discovery Cancer	99.61	99.74	90.73	89.57	85.14	85.64
50 ng FFPE lung, DIN 7.2		99.77	99.74	89.26	87.45	91.81	89.85
50 ng FFPE breast, DIN 3.4		99.77	99.72	88.21	87.56	67.02	72.66

AT/GC coverage

The data shown in Figure 5 represent the coverage of the target regions as a function of their %GC, using 10 ng of intact gDNA enriched with the Avida Focused Cancer panel. This panel targets the exons and hotspots of 13 oncogenes (AKT1, ALK, BRAF, EGFR, ERBB2, ESR1, KRAS, MAP2K1, MET, NRAS, PIK3CA, RET, and ROS1) and one tumor suppressor gene (TP53). Both shearing methods exhibit even coverage profiles across the entire GC content range of 30% to 75%, with enzymatic shearing generating approximately 2x higher read counts. While enzymatic shearing appears to result in greater variability across the gene targets, this is an artifact of the higher mean coverage for the enzymatic condition. To account for this, the coverage variability across the targets was assessed using the coefficient of variation (CV), which normalizes the data by average coverage. This analysis shows that the CV for the two shearing methods is nearly identical (Covaris: 9.6% versus Enzymatic: 9.7%), indicating that the overall variability between the two data sets is similar.

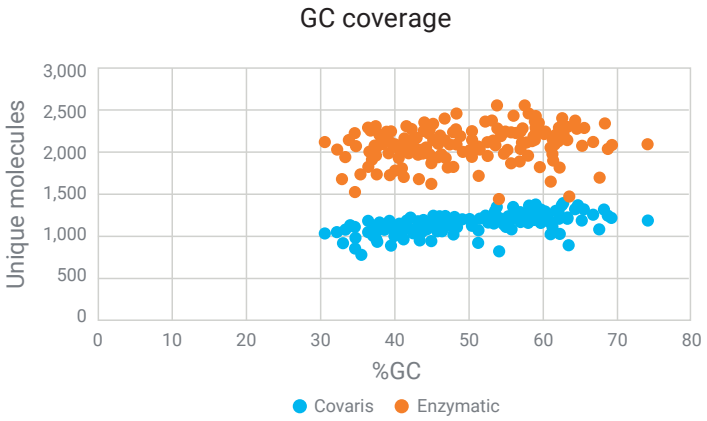

Figure 5. AT/GC coverage for samples sheared with either Covaris or enzymatic methods. Samples with 10 ng gDNA input were enriched with the Avida DNA Focused Cancer panel. Data were downsampled to 1.74 M read pairs.

Variant detection

The Avida DNA library prep and target enrichment is optimized to detect low-frequency alleles in low input samples. Therefore, we used formalin-compromised reference standards, such as HD799, as well as some previously characterized FFPE samples (lung and breast), to examine the detection rate of variants at different allele frequencies (VAFs), ranging from 32% to as low as 1%. The Avida Discovery Cancer panel was used for enrichment in this analysis.

Table 2 shows the VAF data from HD799, which contains 11 variants with allele frequencies ranging from 1% to 24.5%. All 11 variants in HD799 were detected at similar rates in samples sheared by either Covaris or enzymatic methods. Table 3 shows the VAFs obtained from the two real FFPE DNA samples at 50 ng input. Based on previous analysis using Covaris fragmentation and Avida library prep, the FFPE breast DNA (DIN 7.2) contains two variants with VAFs of 9.80% and 13.00%, while the lung FFPE lung DNA (DIN 3.4) has one variant with a VAF of 32.70%. All variants were also detected in these FFPE samples, regardless of the shearing method used.

Additionally, enzymatic shearing generated higher total read depth as well as a higher number of reads containing variants across all samples compared to Covaris fragmentation (Tables 2 and 3). This is not surprising given the higher number of unique molecules (complexity) for enzymatic shearing (Figure 4). These results suggest that enzymatic shearing may enable variant detection at a lower sequencing depth compared to Covaris fragmentation, potentially lowering sequencing costs. Further studies are needed to determine whether enzymatic shearing, in combination with Avida DNA library prep, can provide higher sensitivity compared to Covaris fragmentation in detecting lower-frequency variants.

Table 2. Variant allele frequency (VAF) detection using 20 ng HD799 input. Data shown are the averages across three replicates. Samples were downsampled to 33 M read pairs.

			Covaris			Enzymatic		
Gene	Variant	Expected %AF	No. Reads Containing Variant	No. Total Reads	%VAF	No. Reads Containing Variant	No. Total Reads	%VAF
EGFR	T790M	1.00	9	663	1.31	15	1070	1.43
EGFR	ΔE746 - A750	2.00	8	538	1.55	9	973	0.96
EGFR	L858R	3.00	20	680	2.89	40	1147	3.46
KRAS	G12D	6.25	20	433	4.54	31	682	4.55
PIK3CA	E545K	9.00	9	224	4.17	28	461	6.14
KIT	D816V	10.0	17	262	6.48	37	451	8.28
BRAF	V600E	10.50	61	598	10.15	110	1017	10.79
NRAS	Q61K	12.5	26	355	7.42	44	481	9.15
KRAS	G13D	15.0	64	411	15.64	103	699	14.73
PIK3CA	H1047R	17.5	107	648	16.56	193	987	19.59
EGFR	G719S	24.5	99	468	21.23	211	945	22.30

Table 3. Variant allele frequency (VAF) detection for two 50 ng FFPE DNA samples (lung and breast). Data shown are the averages across two technical replicates. Samples were downsampled to 33 M read pairs.

				Covaris			Enzymatic		
Sample	Gene	Variant	Expected %AF	No. Reads Containing Variant	No. Total Reads	%VAF	No. Reads Containing Variant	No. Total Reads	%VAF
50 ng FFPE lung	TP53	H179L	13	243	1146	21.21%	294	1398	21.04%
	PIK3CA	E542K	9.8	206	1735	11.84%	236	2049	11.50%
50 ng FFPE breast	TP53	G245S	32.7	119	436	27.18%	206	751	27.38%

Enzymatic shearing results using increased reaction volume (50 µL)

To further streamline the enzymatic fragmentation protocol, particularly for samples of low concentration and for researchers with no access to vacuum centrifugation, we developed an alternate protocol to allow higher fragmentation volume. Table 4 compares key sequencing metrics for 10 ng intact and 50 ng FFPE pancreas (DIN 2.6) samples, sheared both mechanically (control) and enzymatically (in 50 µL per the Experimental section). Libraries were prepped using the Avida DNA workflow and enriched using the Avida DNA Expanded and Focused Cancer panels. Similar to results from the standard fragmentation reaction using 10 µL, higher library yield (1.59- to 3.03- fold) and more unique molecules (1.51- to 1.85- fold) were observed in the enzymatically sheared samples compared to Covaris controls for both sample types. However, the median insert size, percentage of mapped reads, on-target rate, and uniformity are similar for Covaris and enzymatic methods.

Furthermore, a side-by-side comparison of the main sequencing metrics, obtained using the 10 µL and 50 µL enzymatic fragmentation volumes, demonstrates that the median insert size, number of unique molecules, on-target rate, and uniformity are comparable between the two reaction volumes (Figures 6A and 6B). This confirms that increasing the fragmentation volume to 50 µL does not adversely affect the sequencing results.

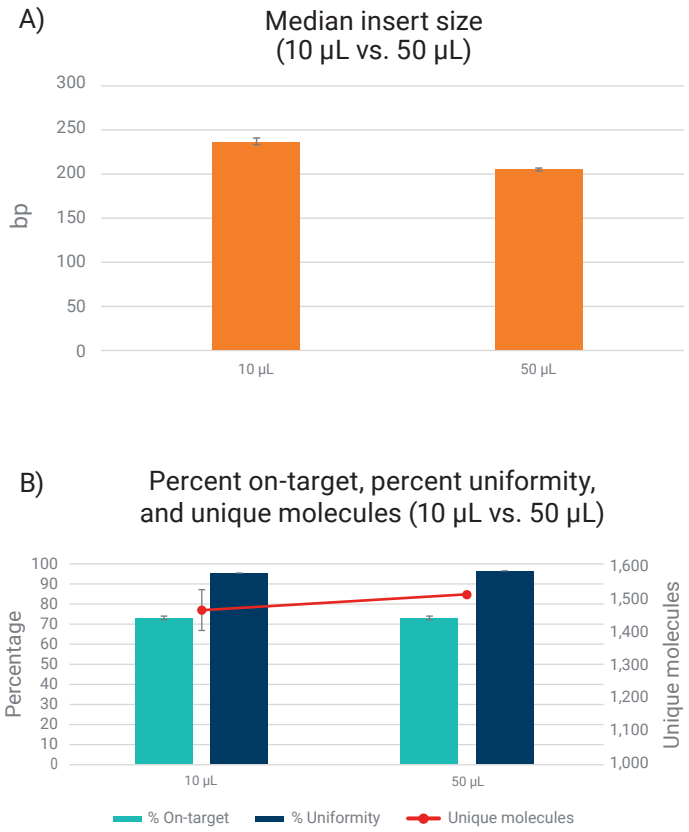

Figure 6. Comparable sequencing performance between samples sheared in 10 µL and 50 µL reaction volumes. The samples were enriched using the Avida Expanded Cancer panel and downsampled to 25 M read pairs. (A) Median insert size. (B) On-target rate, uniformity, and number of unique molecules.

Table 4. Key performance metrics for 10 ng intact and 50 ng FFPE pancreas gDNA samples. The samples were fragmented enzymatically in a 50 µL volume and compared to Covaris controls. Avida Expanded Cancer panel samples were downsampled to 25 M read pairs, and Avida Focused Cancer panel samples were downsampled to 3.5 M read pairs. Each data point is the average of two technical replicates.

Shearing Method	Yield (nM)	Med. Insert Size	Unique Molecules	% Mapped	% On-Target	% Uniformity
10 ng intact gDNA (Expanded Cancer panel)						
Covaris	30.10	227	815	99.46	74.48	96.24
Enzymatic	91.25	205	1508	99.65	73.09	96.49
50 ng FFPE pancreas (DIN 2.6) (Focused Cancer panel)						
Covaris	28.75	118	113	99.64	36.72	79.47
Enzymatic	45.75	124	171	99.68	41.67	85.93

Conclusion

The Agilent Avida DNA workflow is optimized for maximal DNA recovery, offering high sensitivity through innovative, high-performance target enrichment technology. This study demonstrates that the Avida DNA workflow can be further supported by the Agilent SureSelect enzymatic fragmentation kit, which provides a robust and efficient alternative to mechanical DNA shearing. Enzymatic fragmentation generates higher library yields and an increased number of unique molecules compared to mechanical shearing, while both methods show similar median fragment size, on-target rates, uniformity, VAF detection rates, and GC coverage. The fast and simple protocol of the SureSelect enzymatic fragmentation kit is not only cost-effective but also versatile, accommodating a wide range of genomic DNA types and input amounts with a single optimized procedure. Therefore, the SureSelect enzymatic fragmentation kit can effectively enhance the Avida workflow with more options for labs, putting users in control of the fragmentation method they use.

References

1. Corioni, M.; Chen, X.; Wang, H.; Zhao, G. Evaluating cfDNA NGS Performance with the Agilent Avida DNA Workflow and SeraSeq cfDNA Mutation Mix v4; Agilent Technologies application note, publication number 5594-7953EN, **2024**.
2. Ruvolo, M. Enhancing Targeted Next-Generation Sequencing with Cooperative Hybrid-Capture for DNA and Methyl Sequencing; Agilent Technologies Application note, publication number 5594-7815EN, **2024**.
3. Costello, M.; Pugh, T. J.; Fennell, T. J.; Stewart, C.; Lichtenstein, L.; Meldrim, J. C.; Fostel, J. L.; Friedrich, D. C.; Perrin, D.; Dionne, D.; et al. Discovery and Characterization of Artifactual Mutations in Deep Coverage Targeted Capture Sequencing Data Due to Oxidative DNA Damage During Sample Preparation. *Nucleic Acids Res*. **2013**, 41 (6), e67. DOI: 10.1093/nar/gks1443.

www.agilent.com/chem/avidadna

PR7001-4088

Published in the USA, March 31, 2025
5994-8275EN

Agilent
Trusted Answers