Enzymatic fragmentation enables scalable preparation of high-quality FFPE libraries with minimal artifacts

Giulia Corbet, Philip Benson, Kailee Reed, Skyler Mishkin, Thomas Harrison, Zane Jafaar, Kristina Giorda, Josh Haimes, Martin Ranik, Brian Kudlow | *Watchmaker Genomics, Boulder, CO USA*

Kristin Scott | *Colorado State University, Fort Collins, CO USA*

INTRODUCTION

The preparation of formalin-fixed paraffin-embedded (FFPE) samples from tissue biopsies is standard practice in clinical pathology, as it preserves cellular morphology for microscopy-assisted diagnosis and enables long-term archiving for retrospective studies. However, nucleic acids retrieved from such samples are chemically and physically damaged as a result of the fixation process, storage conditions, and extraction methodology.[1,2]

As a consequence, FFPE DNA is typically limited in quantity and highly variable in quality, which complicates next-generation sequencing (NGS) analysis. Low inputs and chemical modification (such as cross-linking) impact library preparation conversion rates, library complexity, and amplification efficiency, and induce molecular artifacts that impact variant calling accuracy and sequencing economy. Given these intrinsic challenges, NGS library preparation methods that limit further loss or distortion of biological information are critical in oncology and other high-sensitivity translational/clinical research applications.

Sonication is the most common method for shearing of FFPE DNA for short-read sequencing, traditionally due to lower fragmentation bias (compared to enzymatic methods, including tagmentation) and better control over library insert size and distribution. However, sonication requires a significant capital investment, is laborious and difficult to scale, results in the loss of already precious material, and introduces sequencing artifacts when working with FFPE samples.[2]

To address these pain points, Watchmaker Genomics has developed a robust enzymatic fragmentation solution for the preparation of high-quality libraries from FFPE DNA.

HIGHLIGHTS

- Improved chemistry and flexible parameters enable consistent fragmentation and control over FFPE library insert size.
- Single-tube protocol limits sample loss, improves library complexity and sequencing metrics, and enables full automation.
- Minimal artifacts (compared to other enzymatic fragmentation methods and sonication) facilitate data interpretation and improves sequencing economy.

Our primary goals were to improve sequencing data quality, workflow automatability, and operational efficiency for users. The Watchmaker DNA Library Prep Kit with Fragmentation (Figure 1) offers the convenience and scalability of enzymatic fragmentation-based library preparation kits, but with two important improvements:

- consistent, tunable insert sizes; independent of input amount or FFPE quality; and
- significant mitigation of molecular artifacts associated with the library construction process.

Here, we highlight key considerations when preparing FFPE DNA libraries for short-read sequencing, discuss mechanisms for controlling insert size and optimizing library quality, and demonstrate improvements — particularly with respect to sequencing artifacts — over libraries prepared with a workflow that relies on sonication.

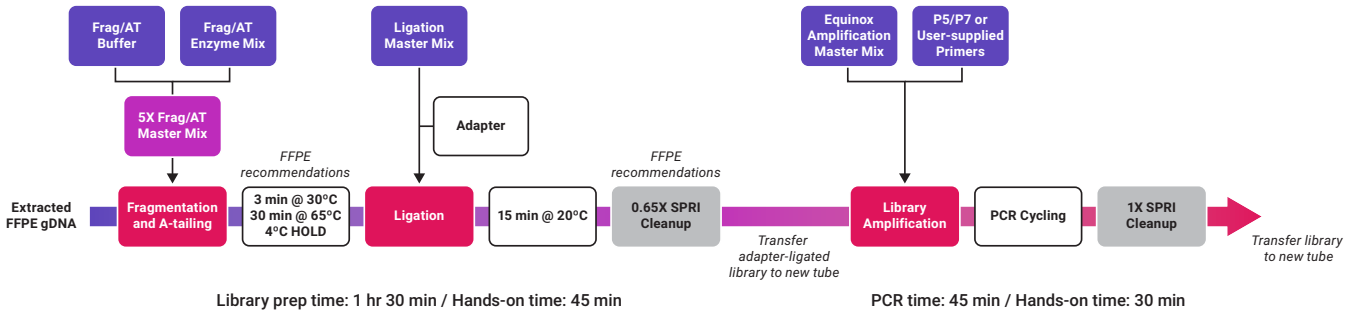

Figure 1. The Watchmaker DNA Library Prep Kit with Fragmentation offers a robust, scalable solution for high-quality library preparation from FFPE samples of variable quality. The simplified workflow combines integrated enzymatic fragmentation/A-tailing chemistry with a highly optimized Ligation Master Mix to produce adapter-ligated libraries in single tube with three reagent additions and two incubations. This minimizes sample loss due to tube transfers and facilitates automation. The protocol offers flexibility with respect to input amount and quality, fragmentation parameters, adapter design, cleanup ratios, and amplification parameters to achieve optimal results for different sample types and sequencing applications. Optimization strategies are outlined in Figure 4.

FFPE LIBRARY PREPARATION: KEY CONSIDERATIONS

FFPE quality assessment

Unlike high-quality genomic DNA preparations, FFPE DNA samples are extremely varied. The degree of physical degradation and chemical modification differs between samples and depends on the original fixation method, how long and under which conditions blocks were stored, and the protocol and reagents used to extract DNA from the archival material. Even within a batch of samples handled and prepared in a similar manner, the final FFPE library fragment size distribution and yield may vary significantly.

Quality assessment of input DNA is an invaluable tool in the establishment and optimization of an FFPE library preparation workflow. Electrophoretic methods used routinely for library QC offer a good indication of DNA degradation (degree of fragmentation), but do not provide any insight into chemical damage such as crosslinking, deamination, or other base modifications that impede the conversion of FFPE DNA into sequencing libraries. Instead, a quantitative PCR (qPCR)-based method — such as the one described by Saelee, et al. (2022) — is recommended to determine the amount of amplifiable (utilizable) DNA in a sample.[3] "Quality scores" determined with such assays are typically good predictors of FFPE library prep outcomes.

Fragmentation parameters

The Watchmaker DNA Library Prep Kit with Fragmentation contains a new generation of enzyme cocktails specifically formulated to enable consistent fragmentation across DNA input amounts (Figure 2). This obviates the need to fine-tune fragmentation parameters (time and in some cases, temperature) for different sample cohorts or subsets of samples, thereby facilitating adoption in clinical/translational settings, and high-throughput/automated pipelines. As indicated in Figure 1, mild fragmentation parameters (3 min at 30°C) are recommended as a starting point for FFPE DNA and support efficient library construction.

Post-ligation cleanup ratio

Mean insert sizes for FFPE libraries typically decrease with decreasing FFPE quality, irrespective of the fragmentation method used (Figure 2). This is attributed to chemical damage that renders DNA resistant to fragmentation and amplification. Sequencing read length is commonly tailored to expected library insert size to minimize read overlap and maximize sequencing economy, but the extent to which this can be achieved depends on operational preferences and constraints (such as having to pool libraries from low-quality FFPE samples with other samples for sequencing on a production-scale sequencer).

With the Watchmaker solution, mean library insert size may be tailored to the preferred sequencing read length by adjusting the post-ligation SPRI ratio. Reducing the ratio (from the standard 0.8X to 0.65X, or as low as 0.5X) favors the retention of longer fragments (see Figure 3A), but comes at a cost to final library yield.

Input into library prep

Lower library yields resulting from selecting for longer insert sizes are effectively offset by the fact that enzymatic fragmentation preserves more input DNA in comparison to mechanical shearing, where up to 44% of template

DNA may be lost during sonication (see Figure A1 in the Appendix). Additionally, DNA loss during sonication is not uniform which necessitates a normalization step after sonication and prior to end-repair. With the Watchmaker solution, the entire amount of input DNA is available for library preparation. This supports higher library complexity at the end of ligation, irrespective of whether the post-ligation SPRI ratio is modified to optimize fragment size. Library yields can be further improved by increasing the amount of input DNA if feasible (Figure 3B). Final library yields may, of course, be boosted by performing more library amplification cycles — but this only addresses output mass requirements, not library diversity.

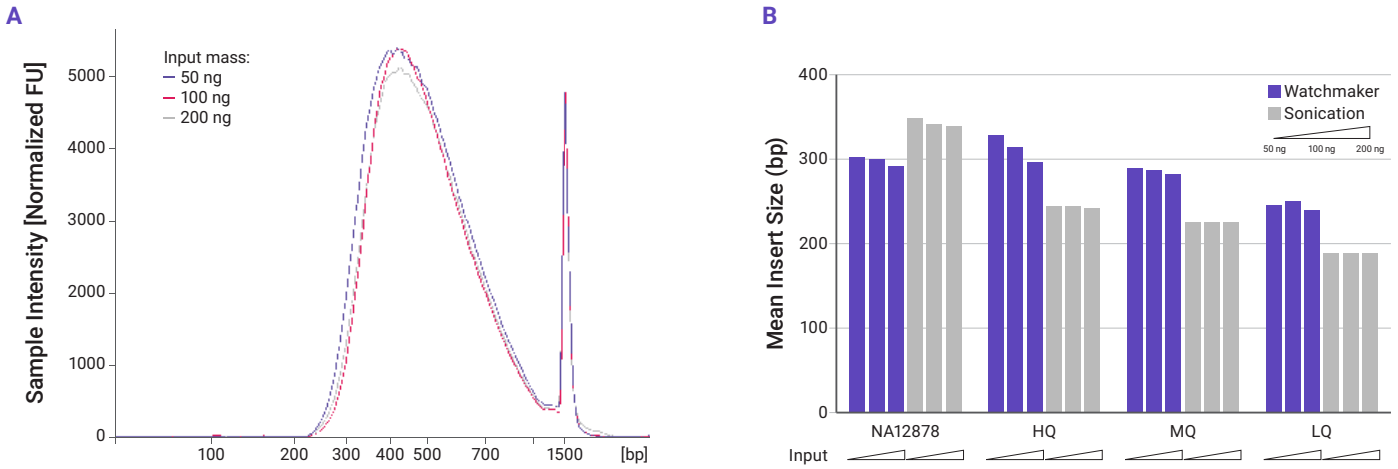

Figure 2. Consistent fragmentation across a range of input amounts. (A) Representative electropherograms for FFPE libraries produced from 50 – 200 ng of input DNA with fragmentation for 3 min at 30°C and a 0.65X post-ligation SPRI ratio. **(B)** Mean insert sizes (determined from Illumina® sequencing data after adapter trimming) for targeted sequencing libraries prepared from 50, 100, or 200 ng of FFPE DNA of variable quality or high-quality NA12878 genomic DNA, using the Watchmaker kit (purple) or a library preparation workflow with sonication (gray). Sonication parameters were set to obtain a mean library insert size of 450 bp, but insert sizes became progressively shorter with decreasing DNA quality. HQ: high quality, MQ: medium quality, LQ: low quality.

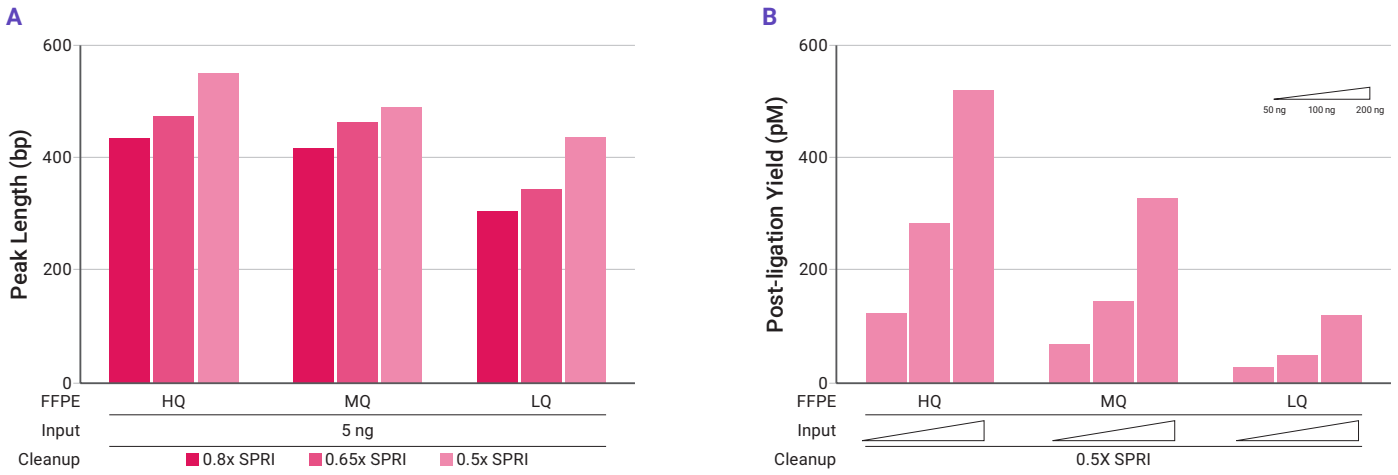

Figure 3. Optimization of key parameters to control library quality. (A) Optimization of final, mean fragment size (determined using the Agilent® 4200 TapeStation system and D5000 TapeStation assay) by adjusting the SPRI ratio used in the post-ligation cleanup. A ratio of 0.5X increased the peak fragment size for libraries produced from 5 ng of low-quality FFPE DNA to that obtained from a high-quality FFPE sample using the standard 0.8X ratio. **(B)** Increasing the input into library construction compensates for lower post-ligation yields (resulting from lower input DNA quality or a 0.5X post-ligation SPRI ratio). In this experiment, increasing the input to 200 ng restored the post-ligation yield for a low-quality FFPE sample to the level obtained using 50 ng of a high-quality FFPE DNA. Post-ligation yields were determined using an in-house developed qPCR-based library quantification assay. HQ: high quality, MQ: medium quality, LQ: low quality.

Workflow optimization

A general strategy for establishing an FFPE library prep workflow with the Watchmaker DNA Library Prep Kit with Fragmentation is illustrated in Figure 4. Initial assessment of FFPE DNA quality enables the selection of a small, diverse, and representative set of samples for workflow optimization. A fixed input and fragmentation time is used to determine the best post-ligation SPRI ratio. Fragmentation time, input amount, and/or the number of PCR cycles may then be adjusted to achieve the desired fragment size distribution, yields, and sequencing performance metrics during routine processing of future sample cohorts. Once the workflow has been optimized and qualified, it should not be necessary to perform quality assessment of input DNA or adjust parameters from run to run or between sample batches.

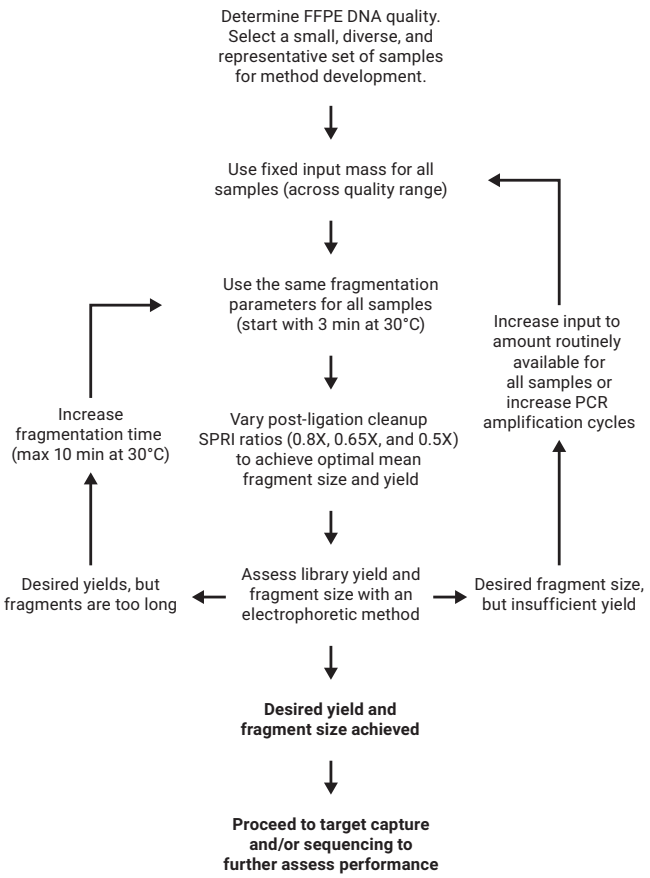

Figure 4. Strategy for establishing and optimizing library prep for FFPE samples of variable quality. Optimization may be performed with a relatively small set of samples, provided that they represent the quality range that will be processed routinely. Compile a set of library and sequencing QC metrics with suitable acceptance criteria and use this to optimize the workflow. Adjust the post-ligation SPRI ratio and other parameters systematically until acceptance criteria are consistently met, and then qualify the workflow with a broader sample set. Quality assessment of FFPE samples is recommended to select the appropriate sample set for method development, but is not required for routine sample processing.

RESULTS: IMPROVED LIBRARY QUALITY AND SEQUENCING ECONOMY

An optimized Watchmaker workflow for variable quality FFPE DNA samples generates consistent, high quality libraries. Compared to protocols that rely on sonication, the Watchmaker solution enables higher conversion rates (Figure A2 in the Appendix) and library quality as a result of:

- **not** losing unique molecules during tube transfers;
- highly optimized chemistry, designed to make every enzymatic step (fragmentation, A-tailing, and adapter ligation) as efficient as possible; and
- highly efficient, ultra-high-fidelity amplification with the Equinox Library Amplification Kit, which reduces length, GC, and unique molecular identifier (UMI) bias.

Libraries prepared from FFPE samples are often subjected to target enrichment using a hybridization capture approach to achieve higher sensitivity and sequencing economy. Pre-capture library quality directly impacts sequencing data quality and enrichment metrics. In this experiment, libraries prepared from FFPE samples of variable quality with the Watchmaker solution returned comparable or better sequencing performance compared to libraries prepared from the same samples with a kit that employs sonication — despite the fact that input masses for the latter were quantified and normalized after sonication to compensate for losses incurred during the shearing process (Figure 5 and Table A1 in the Appendix).

In addition, the Watchmaker solution minimizes sequencing artifacts that complicate NGS data interpretation and impact sequencing economy:

- Improved control of FFPE library insert length enables less read overlap, thereby increasing the amount of non-redundant data for downstream analysis (Figure 6A).
- Reduced soft clipping (masking of unaligned bases at the 5'- and 3'-ends of reads; Figure 6B) minimizes data loss and improves the confidence of variant calls close to the ends of library fragments (Figure 6C).
- "Hairpin" artifacts (a type of chimeric read; Figure 6D) are elevated in sequencing data generated from FFPE DNA libraries.[2,5] In this experiment, a significant level of hairpin artifacts was observed in sonication libraries, whereas they were virtually absent from libraries prepared with the Watchmaker solution (Figure 6E).

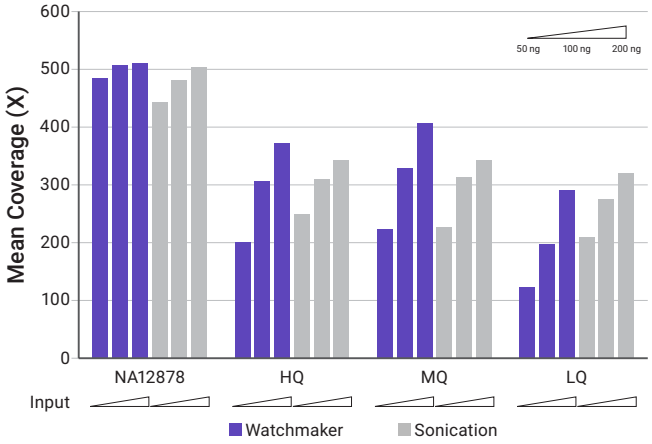

Figure 5. High coverage in targeted sequencing workflows. Libraries were prepared from 50, 100, or 200 ng of variable quality FFPE DNA and a high-quality reference sample. DNA mass into fragmentation and A-tailing (Watchmaker DNA Library Prep Kit with Fragmentation) and end repair/A-tailing (sonication control) was kept equivalent through the implementation of a normalization step after sonication. This effectively masked the impact of any DNA template loss during sonication (see Figure A1 in the Appendix for more information).

CONCLUSIONS

DNA library preparation kits with enzymatic fragmentation were first introduced almost a decade ago. Despite obvious operational advantages and early demonstration of higher library prep efficiency,[5] adoption of these kits for FFPE samples has been relatively slow, especially in higher throughput settings. Key reasons for this trend include poor control over insert size when working with large cohorts of real-life samples, and the discovery of sequencing artifacts attributable to fragmentation enzyme cocktails — which can have significant impacts on variant calling in translational/clinical research.[4,6]

The Watchmaker DNA Library Prep Kit with Fragmentation was designed to address these issues. Novel enzyme chemistry, a highly streamlined, single-tube protocol, mild fragmentation, and flexible library prep parameters offer the control, scalability, and reliable performance needed to process FFPE samples of highly variable quality in targeted sequencing pipelines. Hairpin and other sequencing artifacts that may stochastically occur during library preparation are also minimized with the Watchmaker chemistry.

Library prep parameters outlined here were specifically optimized for FFPE samples and may likewise be tailored for other applications, including whole-genome sequencing (WGS).

MATERIALS AND METHODS

DNA samples and QC. Genome in a Bottle NA12878 genomic DNA was obtained from the Coriell Institute for Medical Research. FFPE blocks were purchased from the BioChain Institute. DNA was extracted using the ReliaPrep™ FFPE gDNA Miniprep System (Promega Corporation). DNA was quantified using a QFX fluorometer and Broad Range dsDNA Assay (Denovix DS-11 FX+).

FFPE DNA quality was assessed on the basis of (i) DNA Integrity Number (DIN) obtained with a 4200 TapeStation system and Genomic DNA ScreenTape assay (Agilent® Technologies) or (ii) ΔCq scores determined using an in-house developed qPCR-based method similar to one described elsewhere.[3]

Library preparation. Libraries were constructed using the Watchmaker DNA Library Prep Kit with Fragmentation (7K0019-024 or 7K0019-096) or the KAPA HyperPrep Kit (Roche). Watchmaker libraries were prepared according to the standard protocol using xGen UDI-UMI Adapters (Integrated DNA Technologies).[7] Unless specified otherwise, (i) FFPE DNA was fragmented for 3 min at 30°C and NA12878 DNA for 10 min at 30°C, (ii) the post-ligation cleanup was performed with a SPRI ratio of 0.65X for FFPE libraries and 0.8X for NA12878 libraries, and (iii) all libraries were amplified with the Equinox Library Amplification Master Mix (2X) and P5/P7 Primer Mix (10X). The number of amplification cycles (FFPE: 11 – 15; NA12878: 6 – 8) was based on post-ligation quantification using an in-house developed qPCR assay. KAPA HyperPrep ("sonication") libraries were prepared according to the manufacturer's instructions, from DNA sheared with a Covaris® E220 ultrasonicator.[8] Shearing parameters were based on a desired DNA fragment size of 450 bp. All final (pre-capture) libraries were quantified and fragment size distributions confirmed using the 4200 TapeStation system and D5000 ScreenTape assay (Agilent® Technologies).

Target enrichment. Library pools (6 x 187.5 ng for NA12878 or 9 x 187.5 ng for FFPE samples) were prepared for multiplexed capture using the Twist Target Enrichment Standard Hybridization v1 Protocol and a 37 kb custom panel (Twist Biosciences).[9]

Sequencing and data analysis. Paired-end (2 x 150 bp) sequencing was performed on the Illumina® platform. Read subsampling was performed with seqtk, adapter trimming with cutadapt, and alignment (to the GRCh38/hg38 reference genome) with bwa mem. Deduplication based on standard sequencing indices was performed with Picard MarkDuplicates, using the REMOVE_DUPLICATE flag set. For deduplication based on UMIs, the fgbio consensus generation workflow with GroupReadsByUmi and CallMolecularConsensusReads was used. Alignment and hybrid selection (HS) metrics were generated using the appropriate Picard tools. Sequencing artifacts were identified using custom, in-house developed python scripts based on the FADE software.[4]

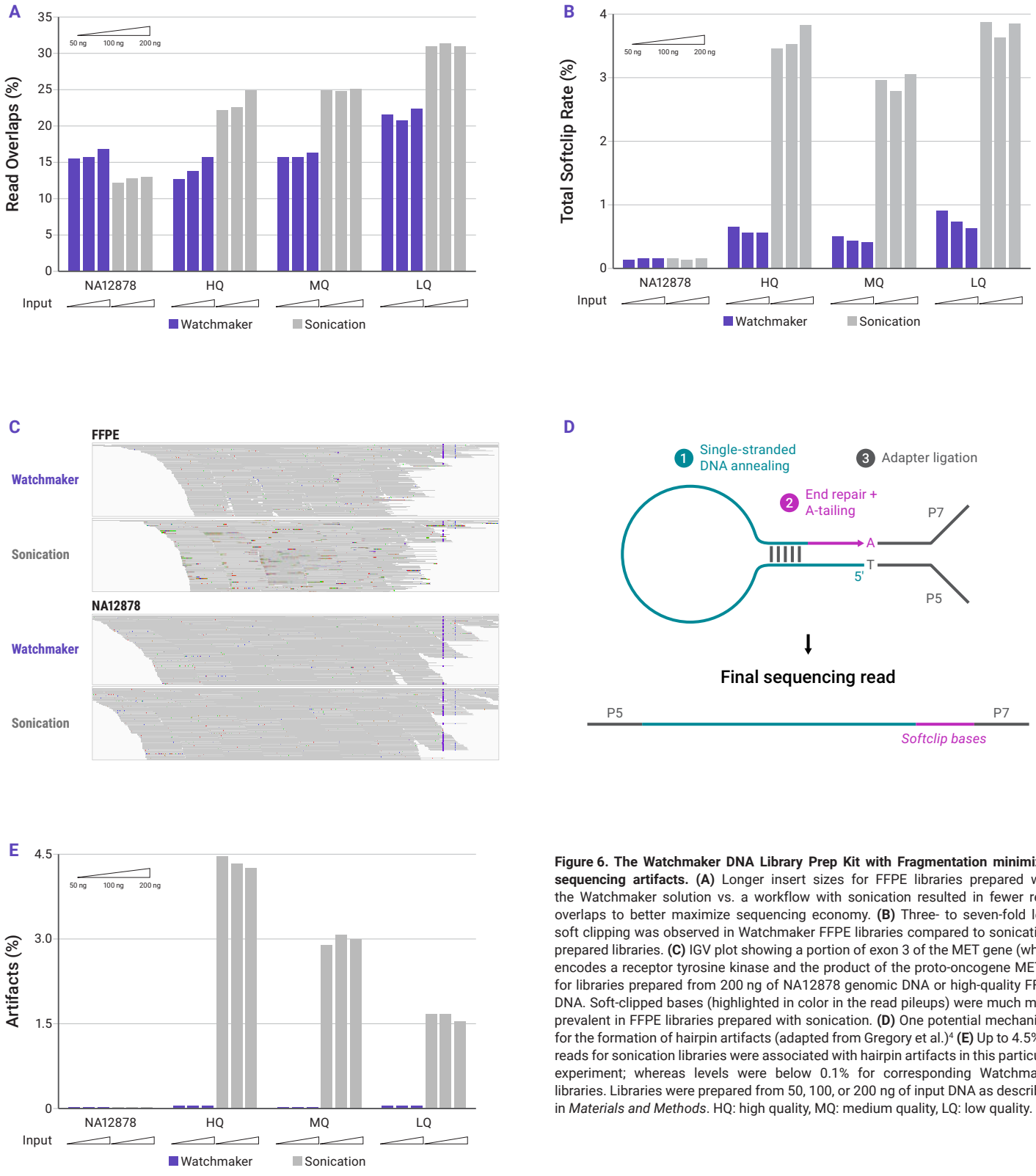

Figure 6. The Watchmaker DNA Library Prep Kit with Fragmentation minimizes sequencing artifacts. (A) Longer insert sizes for FFPE libraries prepared with the Watchmaker solution vs. a workflow with sonication resulted in fewer read overlaps to better maximize sequencing economy. **(B)** Three- to seven-fold less soft clipping was observed in Watchmaker FFPE libraries compared to sonication-prepared libraries. **(C)** IGV plot showing a portion of exon 3 of the MET gene (which encodes a receptor tyrosine kinase and the product of the proto-oncogene MET)[10] for libraries prepared from 200 ng of NA12878 genomic DNA or high-quality FFPE DNA. Soft-clipped bases (highlighted in color in the read pileups) were much more prevalent in FFPE libraries prepared with sonication. **(D)** One potential mechanism for the formation of hairpin artifacts (adapted from Gregory et al.)[4] **(E)** Up to 4.5% of reads for sonication libraries were associated with hairpin artifacts in this particular experiment; whereas levels were below 0.1% for corresponding Watchmaker libraries. Libraries were prepared from 50, 100, or 200 ng of input DNA as described in *Materials and Methods*. HQ: high quality, MQ: medium quality, LQ: low quality.

APPENDIX: SUPPLEMENTARY DATA

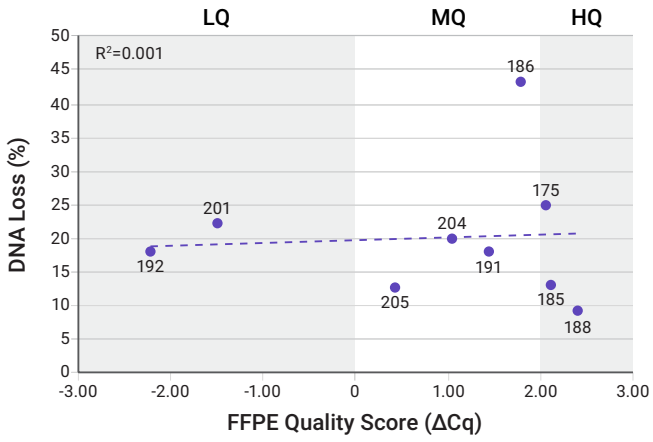

Figure A1. FFPE DNA loss during sonication. DNA loss during sonication was independent of template quality and ranged from 9% to 44%, with a majority having between 15% and 25% loss.

DNA quality was determined using an internally developed qPCR assay. A 54 or 297 bp amplicon (corresponding to a highly abundant and conserved sequence) was amplified from FFPE DNA. For each sample, the Cq value for the 297 bp amplicon was subtracted from the Cq value for the 54 bp amplicon to obtain the ΔCq score, with larger scores corresponding to higher DNA quality. In this assay, the ΔCq for high-quality genomic DNA (e.g., commercial preparations of NA12878 DNA) is typically around 3.25. FFPE samples with a ΔCq>1.5 are regarded as high quality (HQ), whereas 1.5≥ΔCq≥0.0 for medium-quality (MQ) and ΔCq<0 for low-quality (LQ) FFPE samples.

DNA was quantified both pre- and post-sonication using a fluorometric method as described in *Materials and Methods*. Mode fragment sizes were determined using a 4200 TapeStation system and Genomic DNA ScreenTape assay (Agilent® Technologies).

Data labels indicate the mode fragment size for each sample (in bp).

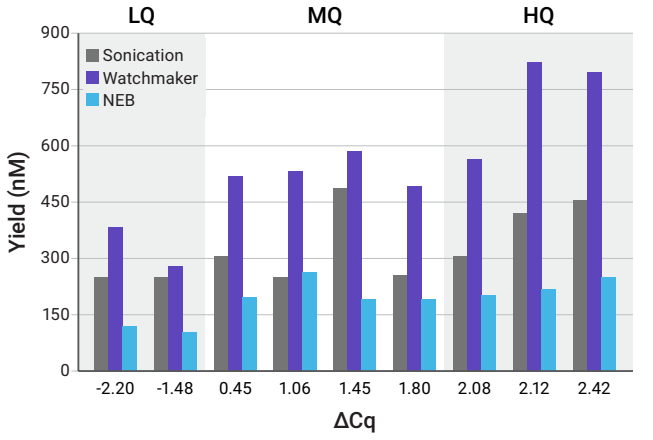

Figure A2. The Watchmaker DNA Library Prep Kit with Fragmentation enables higher conversion rates. Libraries were prepared from 100 ng inputs of FFPE samples of variable quality using the Watchmaker Library Prep Kit with Fragmentation, KAPA HyperPrep Kit with sonication, or enzymatic fragmentation-based NEBNext® Ultra™ II FS DNA Library Prep Kit. Watchmaker and KAPA HyperPrep libraries were prepared as described in *Materials and Methods* (with a 0.8X SPRI ratio for the post-ligation cleanup), and NEBNext libraries according to per manufacturer's recommendations.[11] Library yields were measured using an internally developed qPCR-based library quantification assay.

Table A1. Select quality and sequencing metrics for samples used in this study

	Quality score		PF reads aligned (%)		Improper read pairs (%)		Duplicate reads (%)		Bases on + near target (%)	
DNA Sample	ΔCq	DIN	**WMG**	Sonication	**WMG**	Sonication	**WMG**	Sonication	**WMG**	Sonication
NA12878	3.25	8.9	**100.00**	100.00	**0.09**	0.02	**10.3**	6.8	**80.2**	78.7
HQ FFPE	1.58	3.4	**99.99**	99.98	**0.22**	0.27	**38.0**	28.0	**77.6**	75.9
MQ FFPE	0.06	2.8	**99.99**	99.98	**0.21**	0.29	**23.1**	26.7	**79.4**	75.8
LQ FFPE	-1.64	2.5	**99.98**	99.98	**0.39**	0.35	**22.0**	27.2	**76.2**	75.8
Average	N/A	N/A	**99.99**	99.99	**0.23**	0.23	**23.4**	22.2	**78.4**	76.6

WMG: Watchmaker DNA Library Prep Kit with Fragmentation, Sonication: KAPA HyperPrep Kit with Covaris® shearing, HQ: high quality, MQ: medium quality, LQ: low quality. ΔCq: quality score determined with qPCR-based method, DIN: DNA Integrity Number determined with TapeStation assay. PF: passed filter. For this analysis, data were randomly subsampled to 211,000 read pairs per library.

REFERENCES

1. Hosein A.N., Song S., McCart Reed A.E., et al. Evaluating the repair of DNA derived from formalin-fixed paraffin-embedded tissues prior to genomic profiling by SNP–CGH analysis *Lab. Invest.* 93: 701-710 (2013). doi: 10.1038/labinvest.2013.54.

2. Haile S., Corbett, R.D., Bilobram, S. et al. Sources of erroneous sequences and artifact chimeric reads in next generation sequencing of genomic DNA from formalin-fixed paraffin-embedded samples *Nucleic Acids Res.* 47(2): e12 (2018). doi: 10.1093/nar/gky1142.

3. Saelee S.L., Lovejoy A.F., Hinzmann B., et al. Quantitative PCR–based method to assess cell-free DNA quality, adjust input mass, and improve next-generation sequencing assay performance. *J. Mol. Diagn.* 24(6): 566–575 (2022). doi: 10.1016/j.jmoldx.2022.02.005.

4. Gregory T., Ngankeu, A., Orwick, S., et al. Characterization and mitigation of fragmentation enzyme-induced dual stranded artifacts. *NAR Genom. Bioinform.* 2(4): lqaa070 (2020). doi: 10.1093/nargab/lqaa070.

5. Aigrain, L, Gu, Y., and Quail, M.A. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays—a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics* 17: 458 (2016). doi: 10.1186/s12864-016-2757-4.

6. Tanaka, N., Takahara, A., Hagio, T., et al. Sequencing artifacts derived from a library preparation method using enzymatic fragmentation. *PLoS ONE* 15(1): e0227427 (2020). doi: 10.1371/journal.pone.0227427.

7. Watchmaker Genomics (2022). Watchmaker DNA Library Prep Kit with Fragmentation User Guide, v2.0.1022.

8. Roche Diagnostics (2023). KAPA HyperPrep Kit Technical Data Sheet, v10.23.

9. Twist Biosciences (2022). Twist Target Enrichment Standard Hybridization v1 Protocol. Rev 4.0.

10. https://www.ncbi.nlm.nih.gov/gene/4233.

11. New England Biolabs (2022). NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® Instruction Manual, v3.0_9/22.

WATCHMAKER GENOMICS

5744 Central Avenue, Suite 100
Boulder, CO 80301
www.watchmakergenomics.com

For Technical Support, please contact support@watchmakergenomics.com.

WMAN001