An evaluation of enzymatic fragmentation in the library preparation workflow: The Watchmaker Genomics kit

MARISSA GILDEA[1], SAM DELUCA[1], MICAH RICKLES-YOUNG[1], TOM HOWD[1], ZANE JAAFAR[2], MARTIN RANIK[2], NIALL LENNON[1], STACEY GABRIEL[1]

[1]Broad institute of MIT and Harvard, 320 Charles Street Cambridge, MA 02141
[2]Watchmaker Genomics, Boulder, CO 80301

Overview

As the demand for whole human genome sequencing (WGS) continues to grow, evaluations of both workflow changes and novel technologies will lead advances in scientific development, while ensuring competitive market pricing. Specifically, there is a need for the sample preparation process to evolve in order to match the dropping cost of sequencing. DNA fragmentation is a fundamental step in the library construction (LC) process, and mechanical shearing has historically been the frontrunner in workflow choice given its naturally unbiased and uniform performance. Recently, however, advances in enzymatic fragmentation workflows have been introduced as a truly competitive alternative to the mechanical shearing based library preparation workflows.

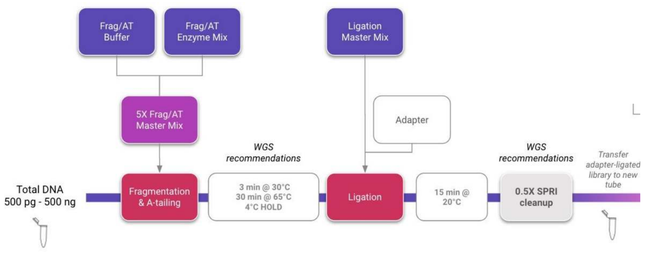

Figure 1. Watchmaker DNA Library Prep Kit with Fragmentation workflow chart

Our evaluation of the Watchmaker DNA Library Prep Kit with Fragmentation included the following steps:

1) Establish baseline relationships between fragmentation time, SPRI concentration, and DNA insert size
2) Evaluate GC bias plots to understand the advancement in non-biased enzymatic fragmentation
3) Compare standard Picard sequencing metrics (chimeric rates, coverage uniformity, etc) to current WGS workflows used in The Broad Institute's Genomics Platform
4) Perform variant calling analysis on PCR Free libraries at 30x mean coverage and assess SNP and InDel specificity and sensitivity.

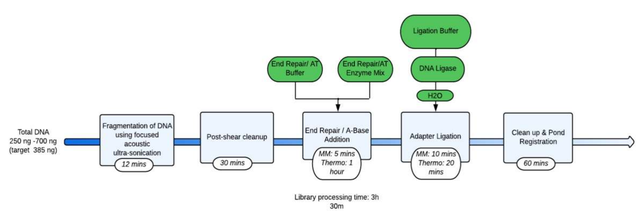

Figure 2. Acoustic shearing library preparation workflow chart

Key Points

- Watchmaker Genomics offers an extremely easy to use kit: two component Frag/AT mastermix and a ready-to-use ligation mastermix
- Ability to create high-quality libraries in less than 2 hours; ~ 1h 20 min time savings
- Achieved tight, reproducible replicates within the same fragmentation condition
- Comparable sequencing data achieved with enzymatic fragmentation and acoustic shearing
- 6 Samples were able to hit 30X mean coverage on 1 lane NovaSeq (0.167 lane fraction)

Evaluate Fragmentation: Time Titration

The lab performed a baseline time titration for fragmentation with the goal of producing an approximate 450 bp fragment size.

- 240 ng input (6ng/uL @ 40uL)
- Fragmentation conditions - 4 min, 5 min, 6 min thermocycler (30℃)
- 0.60X SPRI adlig clean up
- LC was performed manually
- Samples were sequenced on 1 lane MiSeq 2x151
- Important to consider MiSeq vs NovaSeq(~40bp larger on NovaSeq)
- % Chimera < 0.33%

Condition	Frag Size (bp)	Yield (nM)*
4 min	438	11.30
4 min	438	9.28
4 min	445	12.61
4 min	438	10.09
5 min	418	15.77
5 min	419	15.31
5 min	412	16.78
5 min	402	19.14
6 min	411	12.78
6 min	414	18.92
6 min	417	15.32
6 min	420	15.62

Figure 3. qPCR yield based on media insert size value *Minimum yield of 0.9nM in 20uL needed for successful sequencing attempt

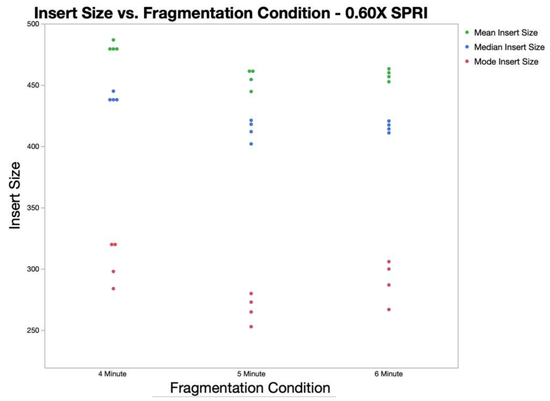

Figure 4. Mean, median, and mode insert size data for the 4, 5, and 6 min fragmentation conditions. Tight, reproducible data points achieved between conditions.

Evaluate Fragmentation: Time Titration + SPRI Titration

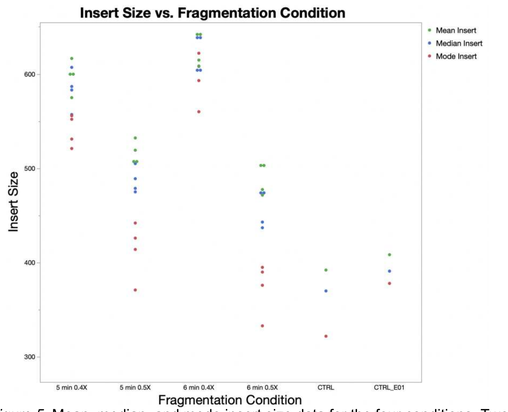

Figure 5. Mean, median, and mode insert size data for the four conditions. Two controls that were processed through acoustic shearing and sequenced on a MiSeq were added as comparisons (CTRL and CTRL_E01).

Condition	Frag Size (bp)	Yield (nM)
6 min 0.4X	639	6.21
6 min 0.4X	638	6.25
6 min 0.4X	603	7.28
6 min 0.4X	605	8.65
6 min 0.5X	443	17.00
6 min 0.5X	475	15.83
6 min 0.5X	437	15.29
6 min 0.5X	473	15.82
5 min 0.4X	583	6.53
5 min 0.4X	607	5.86
5 min 0.4X	557	7.97
5 min 0.4X	584	9.63
5 min 0.5X	489	15.31
5 min 0.5X	475	15.89
5 min 0.5X	478	14.22
5 min 0.5X	505	12.76

Figure 6. qPCR yield based on condition. Calculated with median insert size.

Following the baseline fragmentation time titration, the lab proceeded with adding in two different SPRI conditions for the adapter ligation clean up.

- Uniform 240 ng input (6ng/uL @ 40uL)
- 4 samples per condition (30℃)
  - 5 minute fragmentation @ 0.4X SPRI
  - 5 minute fragmentation @ 0.5X SPRI
  - 6 minute fragmentation @ 0.4X SPRI
  - 6 minute fragmentation @ 0.5X SPRI
- Library preparation was completed by hand
- Samples were sequenced on 1 lane MiSeq 2x151
- % Chimera ranged from 0.44-0.63%

Incorporation of Automated Agilent Bravo Scripts

One large consideration was to remove any human error that possibly could have led to run-to-run variation.

- 3 automated Bravo scripts were created
  - Fragmentation mastermix addition
  - Adapter ligation protocol
  - 0.55X SPRI adapter ligation clean up
- Repeated both 5 minute and 6 minute fragmentation (30℃)
  - 4 samples per condition - 240 ng input
- Samples were sequenced on 1 lane MiSeq 2X151
- % Chimera for all 8 samples <0.358%

Condition	Frag Size (bp)	Yield (nM)
5 min	436	21.88
5 min	440	19.52
5 min	444	20.19
5 min	429	25.35
6 min	406	22.68
6 min	411	20.35
6 min	365	34.60
6 min	414	31.76

Figure 7. qPCR yield for Automated run 1 based on condition. Yield was calculated with median insert size.

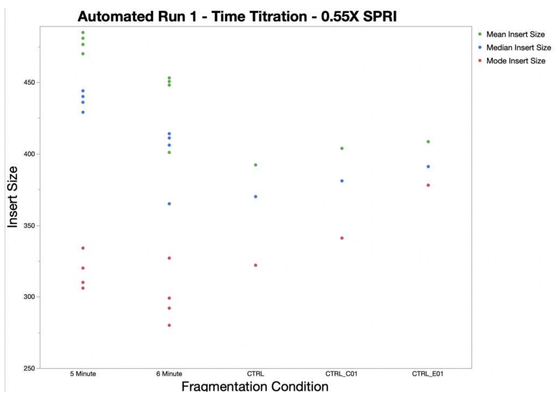

Figure 8. Mean, median, and mode insert size data for the two fragmentation conditions. Three controls that were processed through acoustic shearing and sequenced on MiSeq were added (CTRL, CTRL_C01, CTRL_E01)

Automated Run 2 + NA12878 trio + Ashkenazi trio

Sample Name	HapMap Sample	qPCR Yield (nM) based on 540 bp frag	Mean Coverage (Raw)	Median Coverage (Raw)
A02_91_255ng	NA12891	14.28	35.348288	34
B01_92_253ng	NA12892	14.67	46.667506	45
B06_78_292ng	NA12878	15.54	30.485302	29
C03_43_328ng	NA24143	14.82	36.583202	35
C04_49_344ng	NA24149	15.55	44.520669	43
C05_85_308ng	NA24385	17.02	36.302809	35

Figure 9. Coverage data for the 6 samples that were sequenced on 1 lane NovaSeq S4 flowcell

Metric	Automated Run 2 Data Avg	Current Production 60 Day Avg (n= 8,981)	Current Production 120 Day Avg (n= 14,482)
Estimated Lib Size	4,390,449,032.67	3,182,934,753.85	3,307,355,465.68
% Duplication	11.47	17.4	16.75
% Chimera	0.73	0.78	0.80
% Adapter	0.01	0.01	0.01
Mean Insert Size	559.6	474.75	464.25
Median Insert Size	517.67	446.82	436.79
Mode Insert Size	369	382.36	376
Contamination	0.02	0.19	0.12

Figure 10 . Picard sequencing metrics for the 6 samples (average) in comparison to current production 60 day and 120 day average.

The second automated run involved processing 6 different HapMap cell lines and ultimately sequence on 1 lane of a NovaSeq S4 flowcell. The same automated Bravo protocols were used from the previous library construction run.

- 6 HapMap samples were processed in triplicate through LC
- Input ranged from 250 ng - 344 ng
- 5 minute fragmentation (30℃) @ 0.55X SPRI
- The goal was to sequence samples to ~30x mean coverage and compare benchmarking quality metrics with libraries made with current production workflow.

Variant Calling Analysis Results

Benchmarking analysis was performed on the 4 samples that had sufficient NIST truth data. The libraries were benchmarked and compared against other HapMaps that were processed through acoustic shearing

- All samples were downsampled to match in coverage
- HaplotypeCaller WDL - uses GATK 4.2.0.0
- GenotypeGvcfs WDS to convert to VCF
- BenchmarkVCFs to get comparison data

	Watchmaker Enzymatic Fragmentation					Acoustic Shearing					
NIST Truth Sample	NA12878	NA24143	NA24149	NA24385	AVERAGE (n=4)	NA12878	NA12878	NA12878	NA24149	NA24385	AVERAGE (n=5)
indelF1Score	0.983	0.987	0.988	0.990	0.987	0.988	0.988	0.988	0.985	0.986	0.987
indelPrecision	0.985	0.989	0.990	0.990	0.988	0.989	0.989	0.989	0.988	0.989	0.989
indelRecall	0.981	0.985	0.987	0.989	0.985	0.987	0.987	0.987	0.982	0.983	0.985
snpF1Score	0.996	0.997	0.996	0.996	0.996	0.997	0.997	0.997	0.995	0.995	0.996
snpPrecision	0.995	0.995	0.994	0.994	0.994	0.995	0.995	0.995	0.993	0.993	0.994
snpRecall	0.998	0.998	0.999	0.998	0.998	0.999	0.999	0.999	0.998	0.997	0.998

Figure 11. Benchmarking data post-downsampling of WGS metrics between fragmentation methods. Averaged benchmarking values are almost identical between the two conditions.

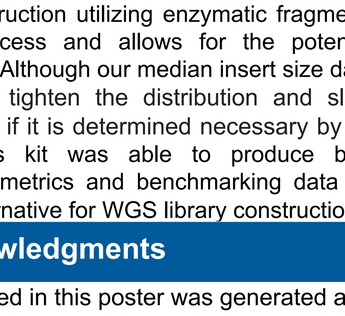

Figure 12. Average IndelF1score and snpF1score of HapMaps between Fragmentation Methods across HapMaps for WM (n=4) and acoustic shearing (n=5) are almost identical.

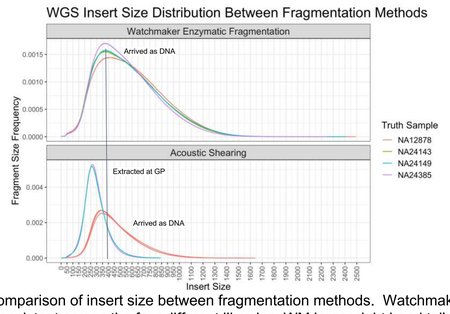

Figure 13. A comparison of insert size between fragmentation methods. Watchmaker fragment size remains consistent across the four different libraries. WM larger right hand tail increases MIS.

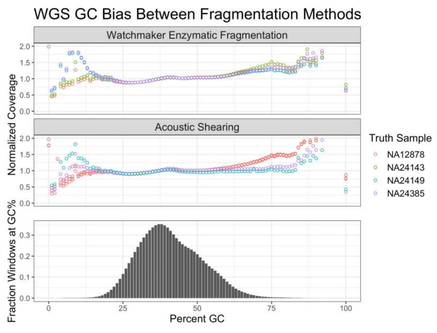

Figure 14. WGS GC Bias between WM enzymatic fragmentation and acoustic shearing. % GC bias is very similar across fragmentation methods.

Conclusion

The Watchmaker DNA Library Prep Kit with Fragmentation is very user-friendly and it promotes a streamlined and cost effective laboratory workflow. As demonstrated with this kit, library construction utilizing enzymatic fragmentation is a time efficient process and allows for the potential to scale in throughput. Although our median insert size data was > 450bp, we plan to tighten the distribution and slightly lower the median size if it is determined necessary by further analysis. Overall, this kit was able to produce both comparable sequencing metrics and benchmarking data proving to be a feasible alternative for WGS library construction.

Acknowledgments

Data used in this poster was generated at the Broad Institute, for more information please visit: http://genomics.broadinstitute.org/