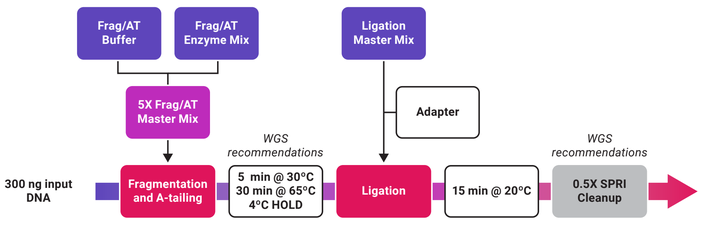
Benchmarking enzymatic library preparation workflows in a PCR-free whole genome sequencing context

Philip Benson, Josh Haimes, Zane Jafaar, Thomas Harrison, Skyler Mishkin, Andrew Gaines, Doug Wendel, Kristina Giorda, Martin Ranik, Brian Kudlow – Watchmaker Genomics, Boulder, CO

Introduction

PCR-free library preparation is considered the gold standard for whole-genome sequencing with minimal bias. These libraries should contain appropriate insert lengths to make efficient use of 2 x 150 bp read structure; however, size selection must not be so stringent as to discard a high fraction of yield, as clinical samples are often scarce. This study describes the ideal parameters for PCR-free WGS library preparation using eight commercially available library prep solutions, and assesses the sequencing data quality of each one.



Library prep time: 1 hr 30 min / Hands-on time: 45 min

Figure 1. PCR-free library prep recommended workflow. The Watchmaker DNA Library Prep Kit with Fragmentation offers a streamlined and automation-friendly workflow.

Library Size Tuning

Matching final library sizes was essential to make valid comparisons between the surveyed library preparation kits. Library size was first tuned by adjusting fragmentation temperature and time, with a target of 450 bp mode insert length. Post-ligation SPRI ratio was the second lever used to adjust library size, if minimal fragmentation parameters created libraries smaller than the desired length.

Table 1. Input mass and post-ligation SPRI ratios

Library Prep Kit	Input Mass	Post-ligation SPRI
Watchmaker	300 ng	0.5x
Watchmaker dsSPRI	300 ng	0.3 – 0.5x
Watchmaker 75 ng	75 ng	0.5x
KAPA HyperPrep	300 ng	0.3x
KAPA HyperPlus	300 ng	0.1 – 0.25x
KAPA EvoPlus	300 ng	0.4x
NEB Ultra II FS	300 ng	0.25 – 0.35x
IDT xGen EZ	300 ng	0.25x
QIAseq FX	300 ng	0.25x
Illumina	300 ng	0.8 – 1.8x

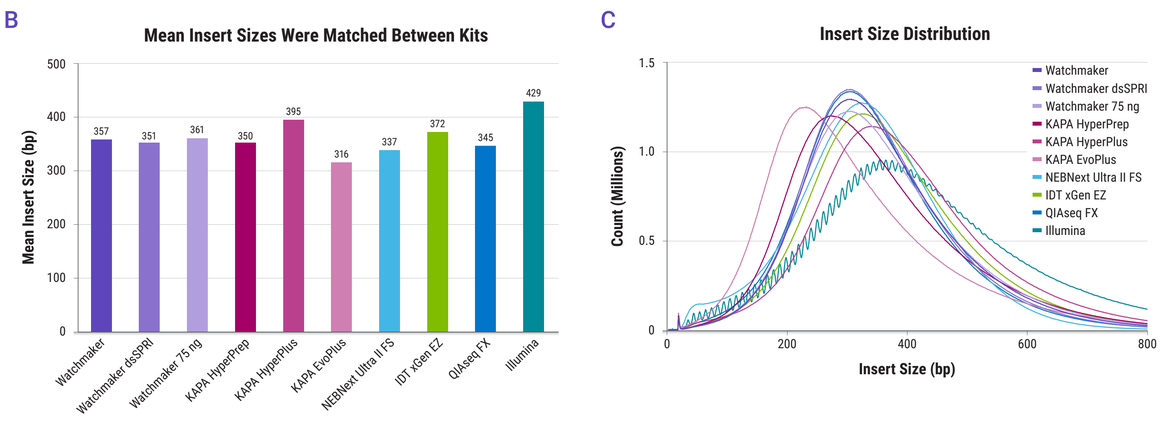


Figure 2. Fragmentation and SPRI cleanup parameters were tuned to match final library sizes. Libraries within 50 bp of the target 450 bp mode insert length were selected for sequencing. **(A)** TapeStation electropherogram traces of sequenced libraries. **(B)** Mean insert sizes were calculated and **(C)** insert size distributions were plotted using NovaSeq 6000 output data.

Watchmaker Prep Kits Maximize PCR-free Yields

The necessary input mass for PCR-free library preparation to produce sufficient yield for NovaSeq 6000 sequencing was determined. A minimum input of 300 ng resulted in approximately 4 nM final libraries across various chemistries. To demonstrate Watchmaker's efficiency, libraries were also generated with just 75 ng of input, delivering superior yield and similar library insert size distribution to other methods (Figure 2C), as measured by qPCR.

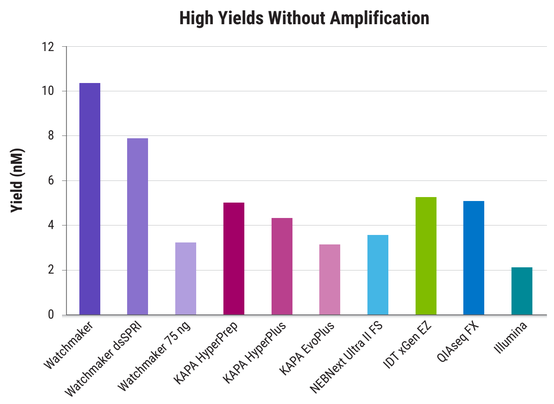


Figure 3. Efficient conversion in high-stringency applications. Double-sided size selection was necessary to achieve matching final library insert sizes for some of the methods assessed. However, it's important to note that this additional step could potentially have a detrimental effect on final library yields, while also extending the total workflow time and cost. Post-ligation library yields quantified by qPCR.

Sequencing Results

Table 2. Sequencing coverage

Library Prep Kit	Mean Coverage	Median Coverage
Watchmaker	28.6	30
Watchmaker dsSPRI	28.8	31
Watchmaker 75 ng	28.2	30
KAPA HyperPrep	27.9	30
KAPA HyperPlus	28.9	31
KAPA EvoPlus	25.9	27
NEB Ultra II FS	28.1	30
IDT xGen EZ	29.0	31
QIAseq FX	28.6	30
Illumina	29.2	31

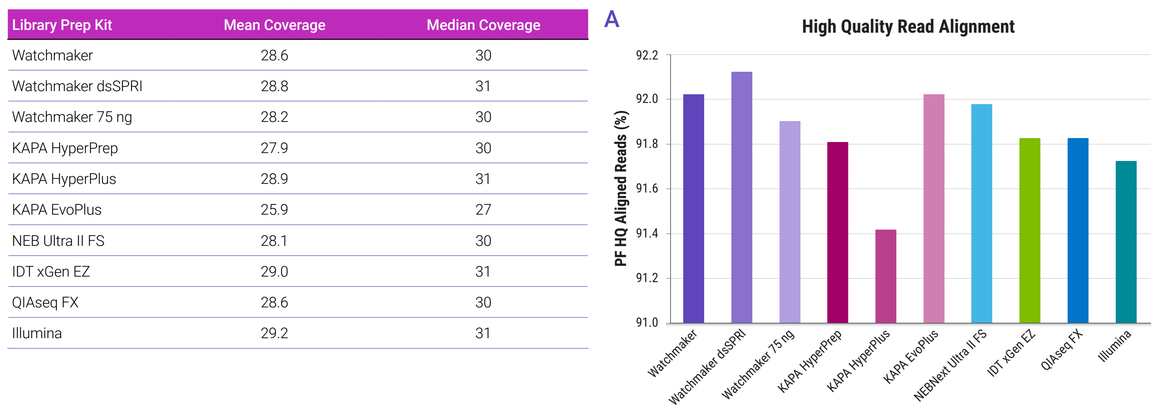


Figure 4. Maximize coverage while minimizing biases. At whole genome sequencing depths **(Table 2)**, Watchmaker libraries achieve **(A)** efficient alignment of high-quality reads, **(B)** unbiased-GC coverage and **(C)** exceptional coverage at GC-rich promoter regions.[1]

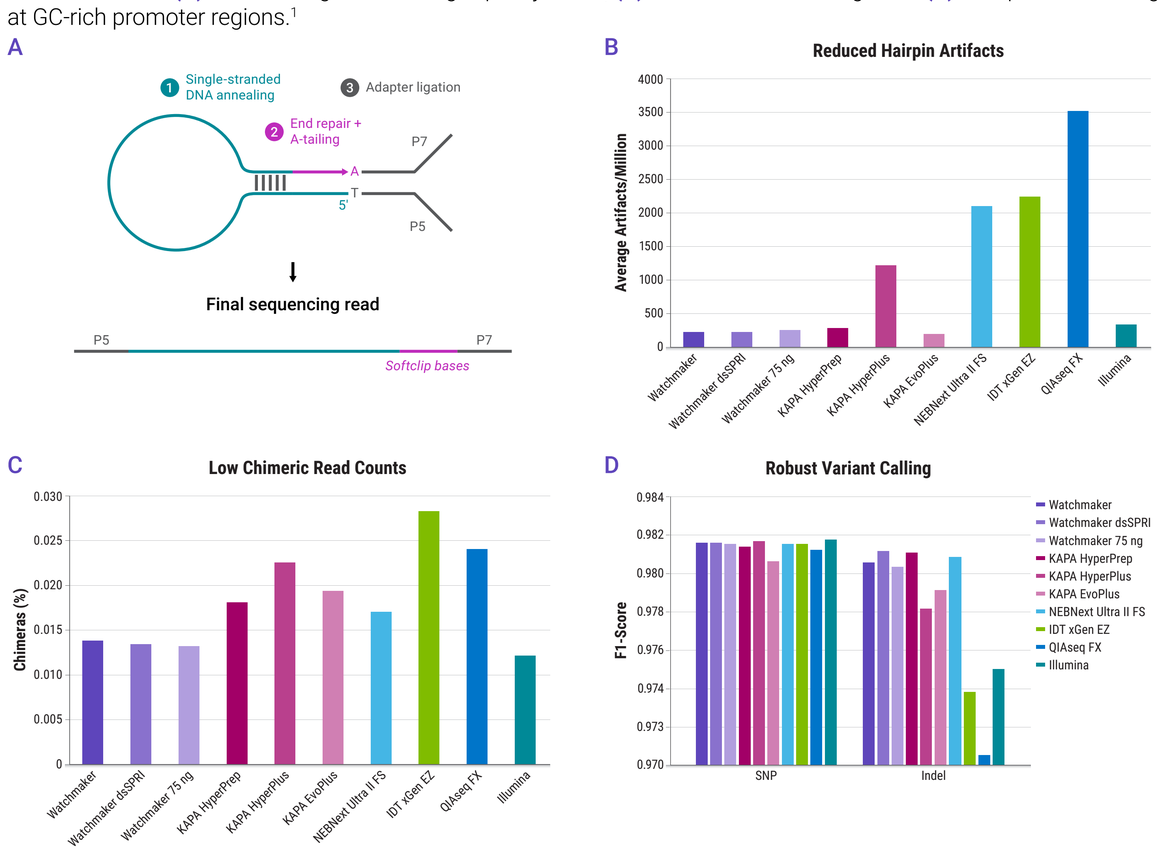


Figure 5. Prevent sequencing artifacts to increase the accuracy of variant calls. **(A)** A hypothesized mechanism of hairpin artifact formation, which interferes with variant calling accuracy. The rate of **(B)** hairpin artifacts and **(C)** chimeric reads in Watchmaker libraries were consistently some of the lowest observed. **(D)** The combination of low artifactual representation and high depth of coverage in Watchmaker libraries resulted in high precision and recall when calling SNP and indel variants.

Methods

PCR-free WGS. Enzymatic fragmentation libraries were constructed from 300 ng gDNA (NA12878, Coriell Institute) using the Watchmaker DNA Library Prep Kit with Fragmentation, KAPA HyperPlus, KAPA EvoPlus, NEBNext Ultra II FS DNA Library Preparation Kit, IDT xGen DNA Library Prep EZ Kit, QIAseq FX DNA Library Kit, or Illumina DNA PCR-free Prep Kit, per manufacturer's recommendations. The sonication control library was constructed from 300 ng of Covaris sheared gDNA (NA12878, Coriell Institute) using the KAPA HyperPrep Kit. Watchmaker's DNA Library Prep Kit with Fragmentation utilized a 5 minute at 30°C fragmentation followed by a 0.5X post-ligation SPRI for longer fragments. The post-ligation SPRI ratio for all other library prep kits was optimized to make 450 bp fragments as measured by HSD5000 ScreenTapes.

Conclusions

Watchmaker DNA Library Prep Kit with Fragmentation delivers:

- High conversion of gDNA inputs without the need for PCR amplification
- High quality, uniform and unbiased whole genome sequencing results
- Best-in-class coverage across challenging GC-rich regions

[1]Michael G Ross, Carsten Russ, Maura Costello, Andrew Hollinger, Niall J Lennon, Ryan Hegarty, Chad Nusbaum, David B Jaffe, Characterizing and measuring bias in sequence data, *Genome Biology*, 14(5):R51. doi: 10.1186/gb-2013-14-5-r51