Optimized enzymatic fragmentation workflow for the rapid construction of high-quality libraries that is compatible with several DNA collection methodologies

Zane Jaafar[1], Aaron Johnson[2], Sean E. Hofherr[2], April Berlyoung[2], Josh Haimes[1], Thomas D. Harrison[1], Lindsay Peterkin[1], Kristina Giorda[1], Martin Ranik[1], Joseph M. Devaney[2], Brian Kudlow[1]

1 - Watchmaker Genomics, Boulder, CO | 2 - GeneDx/Sema4, Gaithersburg, MD

Introduction

Large-scale precision medicine and clinical research initiatives require robust and highly scalable workflows. The source tissue or body fluid and the DNA extraction methodology often have a significant impact on the effectiveness of enzymatic library preparation. We assessed the utility of the Watchmaker DNA Library Prep Kit with Fragmentation for the preparation of PCR-free WGS libraries from DNA extracted from a range of clinically relevant sample types.

Study Design and Methods

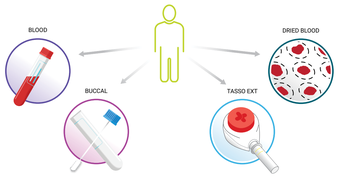

Figure 1. Evaluation methodology. Six patient-matched samples were extracted using the MagMAX DNA Multi-Sample Ultra 2.0 Kit on the KingFisher Flex from blood draws, buccal swabs, an at-home blood collection device (Tasso™) and dried blood spots. Libraries were constructed with the Watchmaker DNA Library Prep Kit with Fragmentation on the PerkinElmer Sciclone. Fragmentation was carried out for 4 minutes at 30°C. Samples were indexed using full-length unique dual index adapters for PCR-free sequencing. Post-ligation libraries were purified using 0.8X SPRI. Libraries were quantified by qPCR, pooled, and sequenced on a NovaSeq 6000 (Illumina).

Real World Sample Library Construction

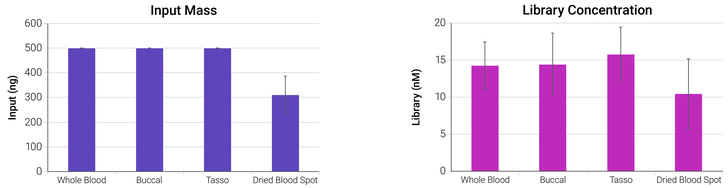

Figure 2. High library conversion across sample types. (A) PCR-free libraries were constructed using inputs, as indicated. **(B)** Final library concentrations were sufficient for downstream sequencing, consistent across samples, and correlated with the input mass used. N=6 for each sample source.

Sequencing Performance

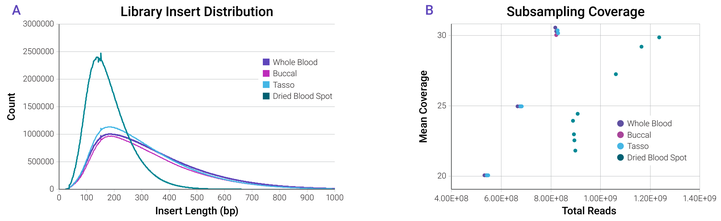

Figure 3. Library insert size strongly influences sequencing read requirements. (A) Library inserts were consistently longer across whole blood, buccal, and Tasso samples while dried blood spot samples produced shorter insert libraries. **(B)** Subsampling analysis indicated that read length was a primary driver of sequencing depth requirements. A larger sample set would be required to determine the input mass, mapping rate, and sample source contributions to sequencing requirements.

Table 2. Average sequencing and alignment coverage metrics for each sample source

Sample source	Total reads	PF reads aligned	Chimera	Softclip	Mean coverage	Median coverage	Adapter	Exc MAPQ	Exc duplicates
Whole blood	8.26E+08	95.5%	1.604%	0.969%	30.20	30.60	0.0081%	3.05%	4.24%
Buccal	9.24E+08	87.3%	1.821%	1.727%	28.75	28.80	0.0321%	3.35%	4.08%
Tasso	8.33E+08	95.5%	1.559%	0.920%	30.09	30.75	0.0065%	3.08%	3.78%
Dried blood spot	1.09E+09	95.3%	1.797%	1.317%	27.72	28.25	0.0099%	3.27%	4.87%

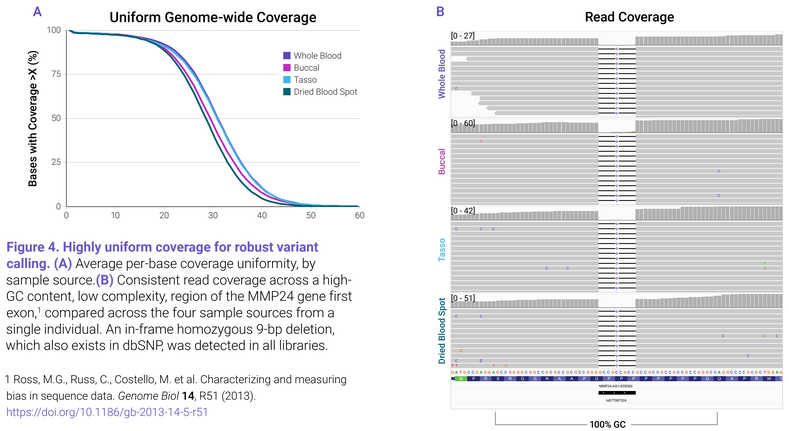

Figure 4. Highly uniform coverage for robust variant calling. (A) Average per-base coverage uniformity, by sample source.**(B)** Consistent read coverage across a high-GC content, low complexity, region of the MMP24 gene first exon,[1] compared across the four sample sources from a single individual. An in-frame homozygous 9-bp deletion, which also exists in dbSNP, was detected in all libraries.

1 Ross, M.G., Russ, C., Costello, M. et al. Characterizing and measuring bias in sequence data. *Genome Biol* **14**, R51 (2013). https://doi.org/10.1186/gb-2013-14-5-r51

Variant Calling Concordance

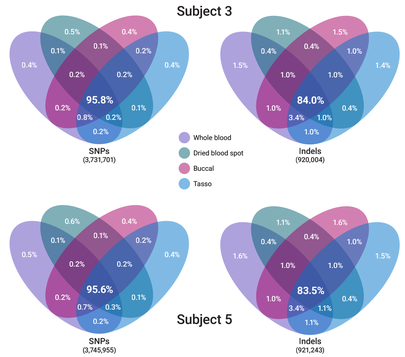

Figure 5. SNP and Indel concordance across sample sources. SNP and Indel variants between four extraction methods were highly consistent for the cohort. The total number of SNPs and Indels for each sample set are provided below each Venn diagram. Representative results from two subjects are shown.

Size Selection for Larger Insert Libraries

Table 3. Library sizing by electrophoresis vs. sequencing

		TapeStation	MiSeq	MiSeq
	Post Ligation SPRI	Mode Library Size	Mean Insert Size	Median Insert Size
Lot 1	0.4X	854	541	534
	0.5X	706	456	435
	0.6X	628	398	365
Lot 2	0.4X	852	543	536
	0.5X	694	452	427
	0.6X	688	394	363
Lot 3	0.4X	909	552	550
	0.5X	763	469	439
	0.6X	715	421	387

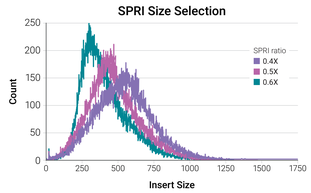

Figure 6. Tuning post-ligation size selection for larger insert libraries. Libraries were constructed from genomic DNA fragmented at 30°C for 3 min. Post-ligation SPRI cleanups were adjusted, as indicated. Library sizes were compared to sequencing insert sizes using TapeStation analysis and MiSeq sequencing, respectively (Table 3).

Conclusions

Watchmaker DNA Library Prep Kits with Fragmentation delivers a scalable and accurate library construction suitable for real-world samples:

- Broad sample compatibility with whole blood, buccal, Tasso, and dried blood spot samples
- Excellent sequencing performance and high variant calling concordance
- Adjusting post-ligation SPRI cleanup may be implemented to further improve performance