FLEXIBLE, PRODUCTION-SCALE, HUMAN WHOLE GENOME SEQUENCING ON A BENCHTOP SEQUENCER

Benjamin Krajacich[1], Kevin Green[1], Kelly Wiseman[1], Peter Mains[1], Samantha Robertson[1], Sophie Billings[1], Mitch Sudkamp[1], Bryan R. Lajoie[1], Semyon Kruglyak[1], Junhua Zhao[1], Shawn Levy[1]

[1]Element Biosciences, San Diego, CA

Element Biosciences

Background

Whole-genome sequencing (WGS) is a pivotal tool in the research and study of rare diseases and can provide understanding for individuals with rare genetic disorders, offering molecular clarity crucial for research in targeted therapies and precision medicine.

Previous studies have shown that Avidite Base Chemistry (ABC) on the Element Biosciences AVITI™ benchtop sequencer is directly compatible with existing upstream and downstream workflows and can be used to study rare genetic diseases associated with inherited retinal degeneration (IRD) phenotype and other rare, undiagnosed neurological disorders (Biswas et al., 2024; Ramsey et al., 2023).

This study presents use-cases and applications over a range of experimental designs leveraging the AVITI system.

- Demonstrating throughput, we describe the generation of 800+ WGS samples to >30x depth with optimizations used to improve reliability and reduce re-queueing of samples, representing high-throughput and scalable use.
- Demonstrating large insert sequencing, we describe the generation and sequencing of >1kb+ libraries shown to provide more accurate variant calling and detection, with downstream beneficial effects on benchmarking and variant calling resolution and accuracy.
- Finally, demonstrating speed, we describe the use of AVITI to perform full coverage >30x depth of a human genome in ~12.5 hours for 2x100 sequencing or ~24 hours for 2x150.

These projects and applications demonstrate the flexibility in usage and ability to scale large throughput projects onto a benchtop system with consistent and reproducible results.

Methods

Library Preparation

To generate libraries ranging from 350bp insert to >1kb, gDNA (500ng-1ug) was mechanically sheared in 55uL of 10 mM Tris using the Covaris ME220 instrument (Woburn, MA, USA) with sonication protocols adjusted to enrich gDNA fragment distribution for the desired insert size range. Library conditions were optimized to produce aligned insert sizes of 350bp, 600bp, and >1kb. Table 1 outlines the shearing and SPRI ratios used during double sided size selection optimized for each library prep condition.

Table 1: Large insert library prep shearing and clean-up conditions.

Covaris Shearing							SPRI Ratio
Insert Size	Duration (s)	Peak Power	Duty % Factor	Cycles/Burst	Avg. Power	Iteration	Double Sided Size Selection
350 bp	10	50	20	1000	10	7	0.46x/0.62x
600 bp	30	50	20	200	10	N/A	0.4x/0.52x
>1kb	5	50	20	200	10	N/A	0.3x/0.42x

Higher throughput sample processing (intake QC, library preparation and QC, and circularization and QC) and sequencing by a single operator utilizing benchtop instruments is easily accomplished as outlined in Figure 1. Project 1 processed 807 unique patient samples using the KAPA HyperPrep Library Prep kit (Roche, USA). The first 77 samples were processed using the 350 bp insert size conditions outlined in Table 1 , then shifting to the 600 bp conditions for the remaining 730 samples. Project 2 processed 92 unique samples using the >1kb conditions using the Element Elevate Library Prep workflow (Element Biosciences, USA) with Element Full-lengths UDIs.

All projects utilized the Rapid Adept™ Workflow with NaOH Denaturation protocol (Element Biosciences, USA) to generate circular libraries unless otherwise stated. The final circularized libraries were quantified using the Applied Biosystems SYBR Green Universal PCR Master mix (Thermo Fisher Scientific, USA) and primers provided within the Rapid Adept kit. Circular libraries were diluted to 1 nM, then used as input into NaOH denaturation at the final library loading concentration along with PhiX libraries as a sequencing control.

Results

High-plexity initial pool runs improve reproducibility, reduce sample requeues, and allows for sample QC.

Table 2: Bulk Trio Pool design (48 samples, 16 trio pools per sequencing run.

Sample	1 nM Trio	Bulk Pool
Sample 1	Trio 1	Bulk Trio Pool
Sample 2		
Sample 3		
Sample 4	Trio 2	
Sample 5		
Sample 6		
Sample N	Trio N	
Sample N		
Sample N		

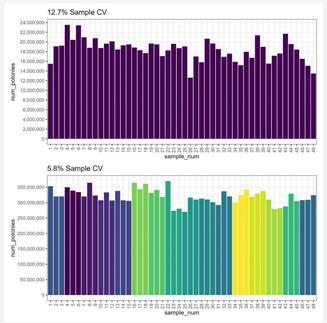

Figure 1: 48-plex topoff runs (top) improve loading cv for full density sequencing runs (bottom).

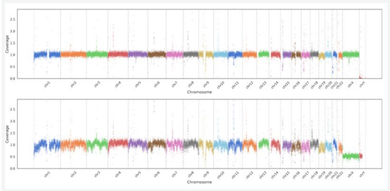

Figure 2: Top-off runs allow for 1x library QC showing where high quality material provides good uniformity (top) or "wavier" coverage (bottom).

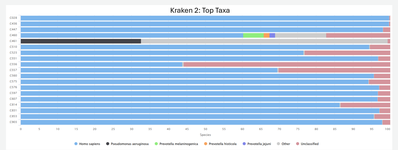

Figure 3: Top-off runs can catch contaminated sample material missed through standard library QC in a cost-effective manner.

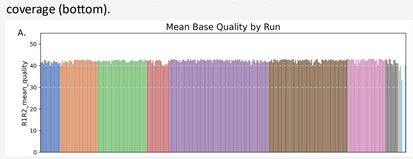

Figure 4: AVITI sequencing has shows consistent quality across human trio sequencing runs with 313/315 completing successfully (2 failed due to operator error).

Large insert library preps increase benchmarking performance while maintaining high quality.

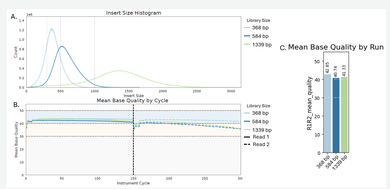

Figure 5: Large insert library preparations (aligned insert sizes up to 2kb, panel A), can be sequenced with high quality (panel B and C) through recipe optimization on AVITI.

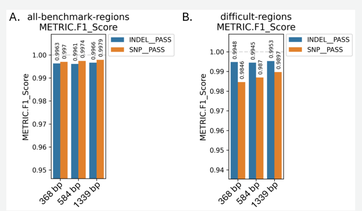

Figure 6: Large insert library preparations increase F1 score in difficult to map regions.

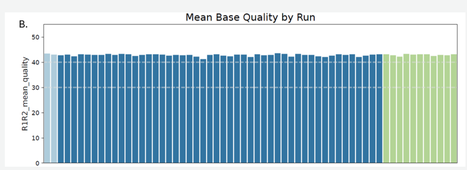

Figure 7: Large insert library preparation and sequencing methodologies are robust enough for higher-throughput projects.

Results

Rapid mode sequencing gives fast time-to-answer with minimal quality trade-offs.

Table 3: Run time options across rapid whole genome sequencing (rWGS) options on AVITI.

	Total run time (hr)		
Read Length	Standard AVITI - 1 FC - CB MO	rWGS Strategy 1 - 2 FC	rWGS Strategy 2 - 1 FC
2x75	18.9	10.1	16.0
2x100	22.9	12.5	18.6
2x150	31.0	17.3	24.0

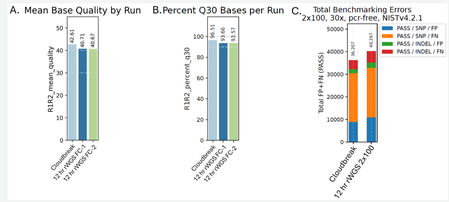

Figure 8: 12-hour 2x100 sequencing (2 FC strategy) shows slightly reduced overall quality and benchmarking errors (HG001) to Cloudbreak Chemistry (24 hour) with reduced output (587M total polonies for 2 FCs with rWGS vs/ 1.1B with in 1FC in these runs).

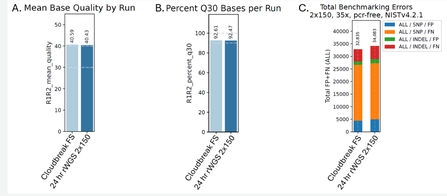

Figure 9: 24-hour 2x150 sequencing shows similar overall quality and benchmarking errors (HG001) to Cloudbreak Chemistry (38 hour) with reduced output (549M in 1FC vs 1.09B)

Conclusions

- Flexibility in sequencing applications, library times, run times, and plexity enables a greater diversity of workflows without needing to batch samples on a factory-scale sequencer.
- Early "top-off" QC runs enables a greater degree of quality control than standard, wet-lab library QC approaches and provides low-pass data suitable for evaluation for contamination, uniformity, library size, and imputation.
- Large insert libraries allow for greater mapping of difficult regions, leading to an increase in benchmarking performance on reference material that is further improved through graph alignment approaches.
- Rapid Whole Genome Sequencing, enabled through recipe modification and/or mixture of flowcells, produces slightly reduced or comparable data quality at a reduced output in up to half the time.

Biswas, P., Villanueva, A., Krajacich, B. J., Moreno, J., Zhao, J., Berry, A. M., Lazaro, D., Lajoie, B. R., Kruglyak, S., & Ayyagari, R. (2024). Avidity sequencing of whole genomes from retinal degeneration pedigrees identifies causal variants. *PLOS ONE*, *19*(10), e0307266. https://doi.org/10.1371/journal.pone.0307266

Ramsey, K., Kruglyak, S., Naymik, M., Lajoie, B. R., Wiseman, K. N., Sanchez-Castillo, M., Billings, S., Jepsen, W., Huentelman, M., & Narayanan, V. (2023). *An efficient design for whole genome trio sequencing identifies key variants in rare neurological disorder cases* (p. 2023.10.13.23296768). medRxiv. https://doi.org/10.1101/2023.10.13.23296768

Visit **elementbiosciences.com** for more information.