Multi-Site Benchmarking Studies For PRECYSE, A Novel DNA/RNA Single Molecule Resolution Imaging And Analysis QC Platform For Long And Short Read NGS Libraries

Jason Reed[1], Katia Sol-Church[2], Kelly Cribari[3], ChewYee Ngan[4], Yongjun Huang[4], Sophie Low[5], Michelle Cipicchio[5], Evelyn Ng[6], Jack Wheeler[6], Guilin Wang[6], Karolina Aberg[7], and Brian Patterson[1]

[1]Evizia, Inc., Richmond, VA, [2]University of Virginia School of Medicine, Charlottesville, VA, [3]Harvard University, Cambridge, MA, [4]The Jackson Laboratory, Farmington, CT, [5]The Broad Institute of Harvard and MIT, Cambridge, MA, [6]Yale University School of Medicine KMSR, New Haven, CT, [7]Virginia Commonwealth University, Richmond, VA

Introduction

Quality control (QC) remains a fundamental component of complex DNA sequencing workflows, ensuring the integrity of nucleic acid samples prior to downstream analysis, and validating intermediate constructs during library construction. Despite its importance, current QC technologies employed have not kept pace with the rapid evolution of sequencing platforms and applications, presenting a barrier to broader research and clinical implementation.

In this study, we report findings from an extensive benchmarking evaluation of a novel biomolecular imaging and analysis system, PRECYSE, designed for rapid sizing and visualization of DNA fragments ranging from <100 to 500,000+ bp. Conducted across six prominent genomics core facilities in the United States, the study assessed over 200 DNA and RNA samples representing a broad spectrum of sample types and quality. These included short- and long-read sequencing libraries, gDNA derived from blood, saliva, and FFPE tissue, ultra-high molecular weight DNA for Oxford Nanopore sequencing, size-selected cfDNA, enzyme-bound libraries, and sheared genomic DNA.

Workflow

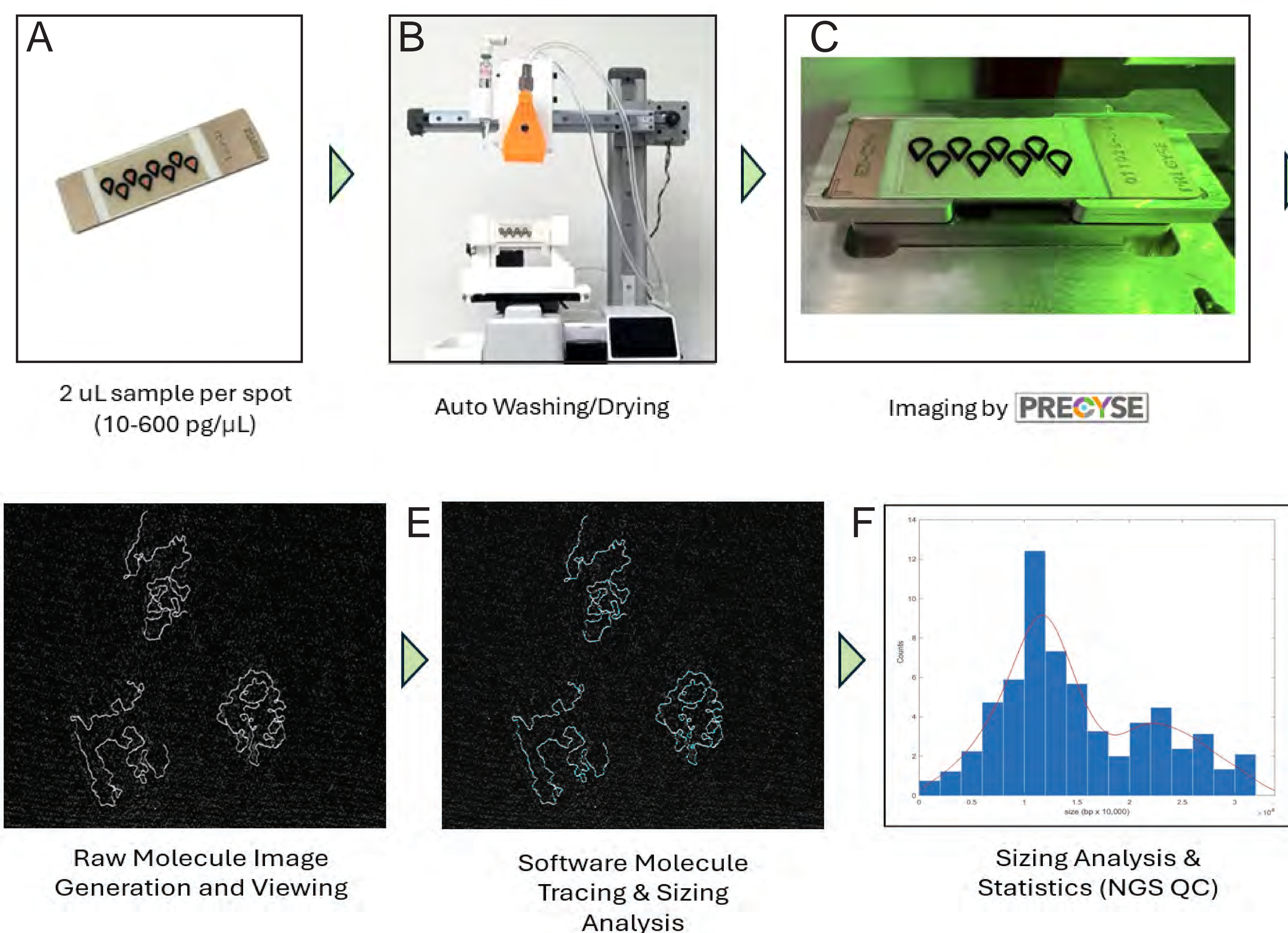

Figure 1. PRECYSE Automated Workflow Overview.

PRECYSE invokes proprietary scanning based on HSAFM (High Speed Atomic Force Microscopy) technology for direct imaging from 70 bp to 500+ kbp DNA samples with single molecule resolution. Exact molecule sizing and distribution is displayed as a histogram or swarm (scatter) plot. Sample to data analysis in as little as 20 min. (A) Samples of varying DNA sizes are added to EZchip, 2 µL per spot (well) at 10 - 600 pg/µL after a 1:1 addition with EZbuffer. (B) EZchip washing and drying system, ~2 min per chip. (C) EZchip shown in PRECYSE instrument ready for scanning. Each sample is scanned, and data analyzed in 2 - 20 minutes depending on scan resolution, which is dictated by approximate size of DNA in sample (gDNA, PacBio vs Illumina libraries). (D) Raw image output is viewed in real-time during scanning as partial 'frames'. Shown here is a zoomed-in section of 1 frame showing 3 PacBio library molecules. (E) Molecules are analyzed and sized by Molecular Explorer software in real-time during scanning. (F) All molecules per sample are pooled together to generate a sizing

Specifications

		PRECYSE	Automated Electrophoresis Instruments [3]	
	Technology	HSAFM	Vertical Mini-Gel	Pulse Field CE
TECHNICAL PERFORMANCE	Sequencer Type Supported	**Short & Long**	Short	Long
	Samples Per Chip [1]	**1-96**	2-16	1-11
	Max Sensitivity (pg) [1]	**Single molecule (0.1ag)** [2]	10	0.1
	Sizing Accuracy [1, 3]	**1-5%**	10-15%	15%
	Sizing Precision (%CV) [1]	**2-7**	5-15	20
	Upper Limit of Sizing (kbp) [1]	**250+**	20-40	165
SIMPLIFIED WORKFLOW	Consumable Prep Time (min) [1]	**15**	30+	40+
	Minimum Run Time (min) [1]	**2**	5	90
	Cold Chain & Perishable	**No**	Yes	Yes
	Multiple Kits	**No**	Yes	Yes
	Standards Required to Size	**No**	Yes	Yes

Table 1. PRECYSE Comparison to Automated Gel Electrophoresis Systems

[1]Based on available published data or user feedback
[2]Based on 1x 100 bp DNA molecule
[3]Technical information sourced from vendor materials available through public sources

Short Read Libraries

PCR Free

cfDNA

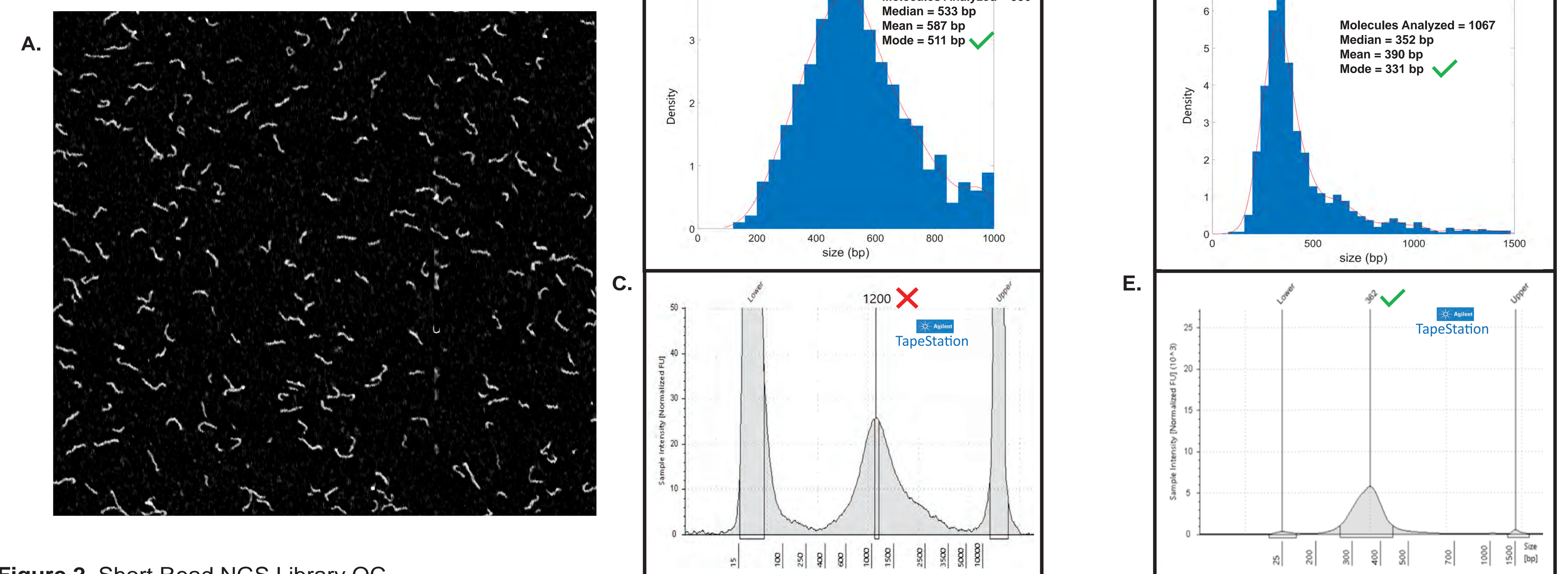

Figure 2. Short Read NGS Library QC

(A,B) Example partial frame image from PRECYSE and corresponding sizing analysis of a short read Mouse PCR free Illumina library generated by The Jackson Laboratory. 2 µL of 200 pg/µL sample, 90 sec total analysis time, generated 936 molecules with median of 533 bp, mean of 587 bp and mode at 511 bp. (C) Agilent TapeStation electropherogram of same PCR Free Illumina library showed a mode of 1200 bp. Slow migration of molecules on the Tapestaion due to Y-adaptors overestimated true sizing. PRECYSE sizing is not impacted by adaptors and secondary structure, and accurately sizes the PCR Free Illumina library. (D) Human cfDNA sample analyzed at Yale University. Sample is post Pippin Prep (Sage Science) size selected targeting 250-450 bp. PRECYSE sizing analysis of 1067 molecules produced a mode of 331 bp with a tail of larger molecules and a second mode at 620bp that increased the overall mean to 390 bp. (E) Corresponding TapeStation electropherogram of same cfDNA sample produced a mode of 362 bp and no visible larger molecule tail. TapeStation and PRECYSE primary sizing modes align well. Sizing of small DNA < 2 kb with PRECYSE requires < 50 pg of material. Sub-pg analysis is possible with extended scanning times that could benefit workflows with limited DNA input such as cfDNA.

Long Read Libraries

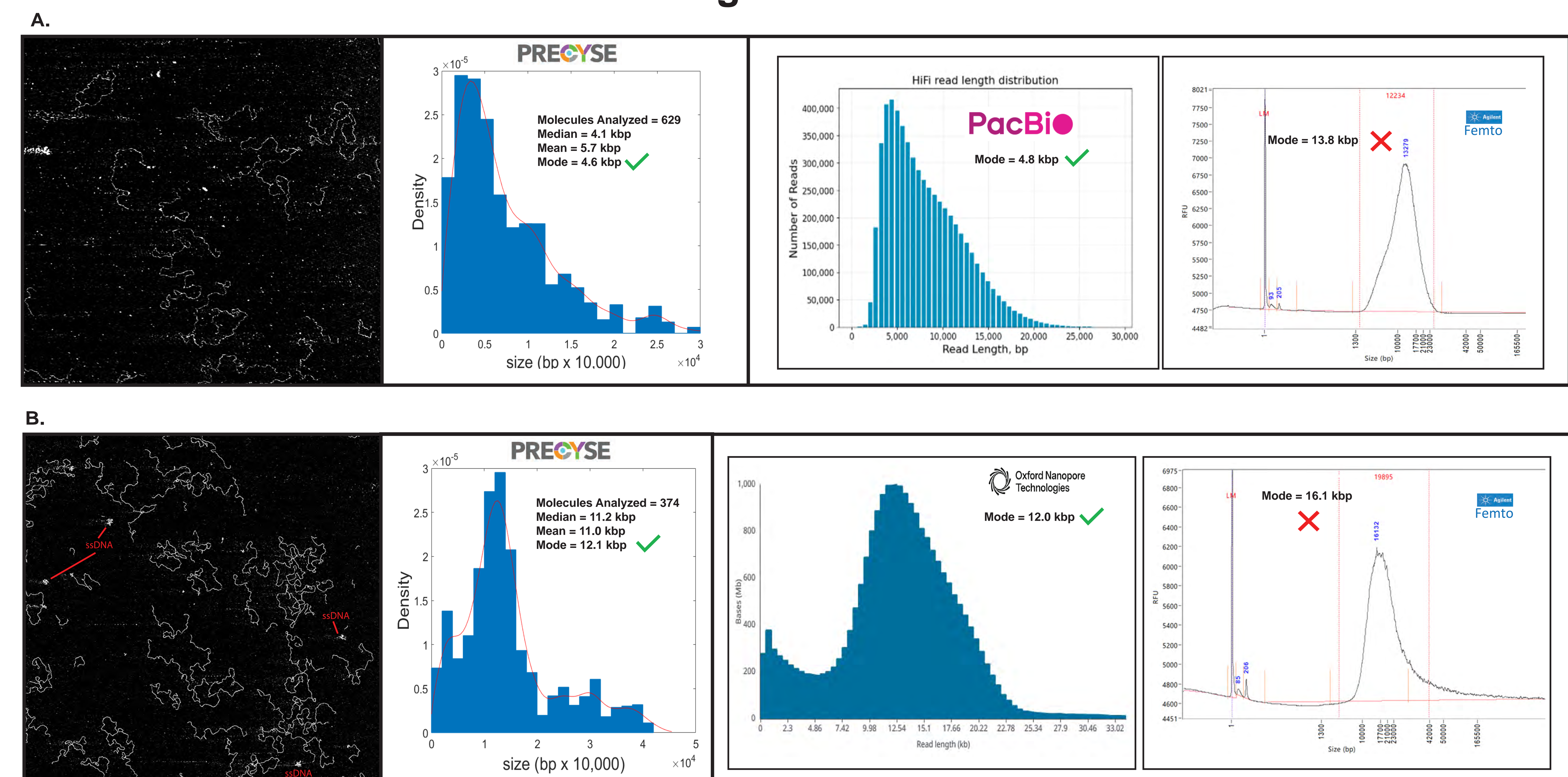

Figure 3. Long Read NGS Library QC

PacBio and Oxford Nanopore (ONT) library QC comparison between PRECYSE and Agilent Femto. All QC and sequencing performed by Harvard Bauer Center. (A) PacBio final library QC of mouse gDNA prior to polymerase addition. PRECYSE generated a mode of 4.6 kbp, Agilent Femto of 13.8 kbp. Final PacBio read length distribution produced a mode of 4.8 kbp. Strong alignment of sizing between PRECYSE and PacBio, while the Femto overestimated sizing by 188%. Additional visual qualitative QC offered only by PRECYSE showed the PacBio library to contain a high concentration of particulates, observed by the white specs. Example image shown is ~15% of a single image 'frame' generated by PRECYSE. Four total frames were imaged from 2 µL of 200 pg/µL sample producing 629 molecules for sample sizing analysis). Total PRECYSE scanning and analysis time of ~10 min. (B) Oxford Nanopore library QC of Pycnogonida (Sea Spider) gDNA after shearing and cleanup just prior to sequencing. PRECYSE generated a mode of 12.1 kbp, Agilent Femto of 16.1 kbp. Final ONT read length distribution produced a mode of 12.0 kbp. Strong sizing alignment between PRECYSE and ONT, while the Femto overestimated sizing by 34%. Additional visual qualitative QC offered only by PRECYSE showed the ONT library contained less particulate than the PacBio mouse sample but had some amount of ssDNA. Four total frames were imaged from 2µL of 200 pg/µL sample producing 374 molecule for analysis with a total sample analysis time of ~10 min.

Conclusion

Evizia successfully completed on-site beta instrument testing with early access collaborators. PRECYSE performed reliably in standard laboratory environments, requiring only a sturdy lab bench and house air for operation. Hundreds of challenging samples were analyzed and results benchmarked against Agilent Femto and TapeStation systems, and PacBio/ONT sequencer read-length histograms. For short-read libraries, sizing accuracy was comparable to electrophoresis, except for PCR-free samples due to adaptors and secondary structure negatively impacting electrophoresis accuracy. Collaborators were able to directly observe individual sample molecules, secondary structures, presence of contaminants, and binding of proteins. Combining consumable preparation and analysis time, PRECYSE delivered faster turnaround (<17 minutes per sample vs. 30+ minutes) and greater accuracy (absolute error <15% vs. ≥15%). We observed that electrophoresis overestimated true fragment sizing, increasing with DNA size, and significantly underrepresented smaller DNA populations (validated by sequencer read length outputs and sizing standards). RNA integrity analysis demonstrated expected reductions in 18S and 28S rRNA populations corresponding to sample degradation, with results comparable to electrophoresis (data not shown). PRECYSE was able to accurately size DNA from FFPE samples despite significant secondary structure and cross-linking, and polymerase and helicase DNA binding was observed. Across all sample types tested, PRECYSE accurately sized DNA from 70 bp to 300 kbp+ and introduced new quantitative and qualitative stringencies into NGS QC that highlight its potential to improve overall sequencing workflow reliability and success.

Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number R44HG014079. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.