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

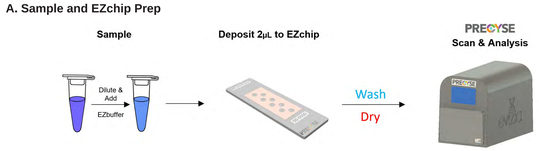
Introduction

Here we present data from a comprehensive benchmarking study for a novel benchtop biomolecule imaging and analysis platform, PRECYSE , from Evizia, that enables rapid visualization and analysis of DNA from 100 bp to >250,000 bp. Our study was conducted at six leading genomics core laboratories in the US and tested the performance in analyzing over 200 DNA and RNA samples of various types and qualities during library preparations for Next Generation Sequencing (NGS). Samples tested included long and short read DNA/RNA libraries, human gDNA from blood, saliva and FFPE, ultra-high molecular weight DNA for ONT, cfDNA size selections, polymerase and helicase bound libraries, and sheared gDNA.

Quality control (QC) in complex NGS workflows is a critical, established and clinically required step in the preparation of nucleic acid samples for analysis. However, existing QC methodologies such as gel electrophoresis have failed to keep pace with rapid advancements in sequencing technologies, impeding broader clinical adoption. Evizia aims to consolidate the complex instrument and kit offerings utilized today for NGS library QC with a single desktop instrument and consumable type, at lower cost, with no cold chain components, simplified workflow, and superior single molecule direct visualization and analysis.

Workflow

A. Sample and EZchip Prep



B. Image Acquisition and Analysis

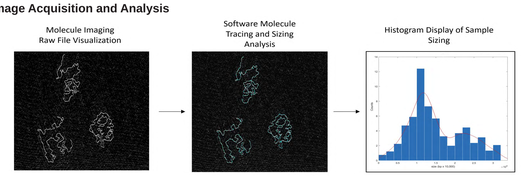


Figure 1. PRECYSE Workflow Overview.

A. Samples are diluted in water to 10 - 1600 pg/uL (range depends on approximate sizing required) and combined 1:1 with EZbuffer. 2 µL is added to each 'spot' of an EZchip followed by a short room temperature incubation. Washing, drying and heating is automated before loading the chip for scanning. Different sample types can be run together on a single multi-plexed EZchip, no ladders or standards required. B. PRECYSE invokes proprietary scanning based on HSAFM (High Speed Atomic Force Microscopy) technology for direct imaging from 100 bp to 250+ kbp with single molecule resolution. Raw image 'frames' are viewed in real time before onboard machine learning analysis software traces DNA molecules for exact molecule sizing and histogram and/or scatter plot generation.

Specifications

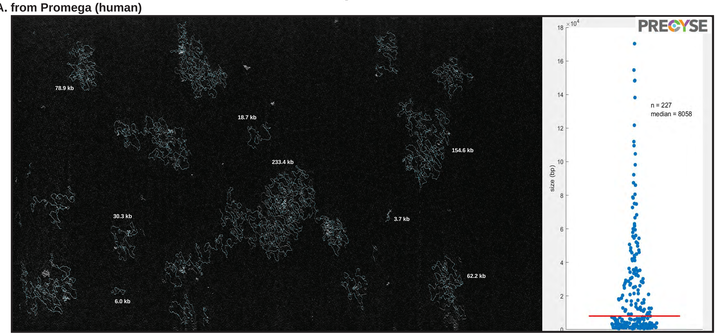
		PRECYSE	Automated Electrophoresis Instruments [3]	
	Technology	HSAFM	Vertical Mini-Gel	Pulse Field CE
TECHNICAL PERFORMANCE	Sequencer Type Supported	**Short & Long**	Short	Long
	Samples Per Chip [1]	**1-96**	2-16	1-11
	Max Sensitivity (pg) [1]	**Single molecule (0.1ag)** [2]	10	0.1
	Sizing Accuracy [1, 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

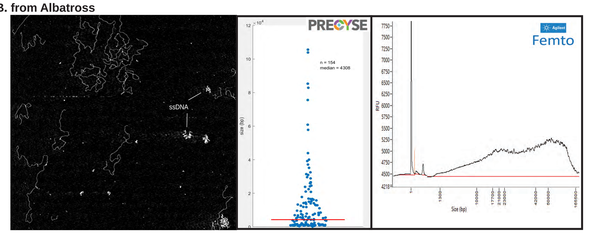
[1]Based on available published data or user feedback
[2]Based on 1x 100 bp DNA molecule
[3]Technical information sourced from vendor materials available through public sources

gDNA

A. from Promega (human)



B. from Albatross



C. from FFPE (human)

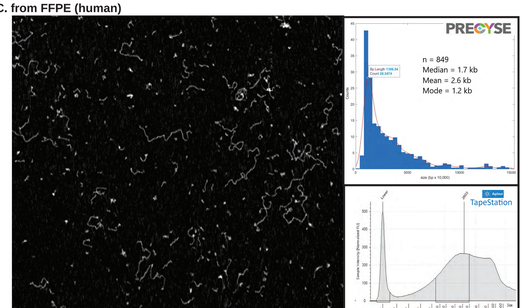


Figure 2. gDNA Analysis and Comparison to Gel Electrophoresis

A. Stock female Promega gDNA scanned by PRECYSE at VCU. One example frame shown. Blue outlines are the PRECYSE software tracing and sizing analysis. Scatter plot of sizing for 227 molecules shows a long distribution of larger molecules out to ~200 kb, indicative of intact high quality gDNA. B. gDNA extracted from albatross provided by Harvard University. PRECYSE imaging shows ssDNA and small particulates in the sample. Sizing analysis of 154 molecules indicates good quality gDNA extending out to ~111 kb. Corresponding Agilent Femto analysis shows a similar size distribution but severely underrepresents the percentage of smaller molecules due to mass-based fluorescence. The non-linear separation and data display of gel electrophoresis makes sizing of larger DNA difficult. C. Human gDNA extracted from FFPE and analyzed by PRECYSE at The Jackson Laboratory. PRECYSE sizing analysis of 849 molecules generated a median size of 1.7 kb with sizes ranging from 212 to 19,784 bp, indicative of degraded/sheared gDNA. Extremely small DNA or particulates, and cross-linked DNA (intense thicker white molecules), is shown. The Agilent TapeStation electropherogram of same FFPE gDNA sample produced sizing from 250 to >48,500 bp (max instrument resolution). This larger incorrect sizing is presumably due to the highly crosslinked nature of FFPE gDNA.

Long Read Libraries

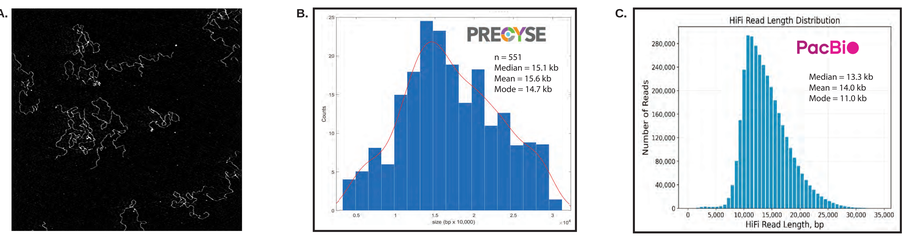


Figure 3. Long Read NGS Library QC

PacBio library provided by The Jackson Laboratory. A. Example raw PRECYSE image of several molecules within one frame. B. PRECYSE sizing analysis of PacBio Library and 551 molecules spanning multiple image frames. Median size of 15.1 kb and mode at 14.7 kb measured. C. HiFi read length distribution from the PacBio Revio showed a median of 13.3 kb and mode of 11 kb. Femto analysis (data not shown) produced a median size of 17.4 kb and Mode of 15.9 kb. PRECYSE analysis more closely aligns with the HiFi read distribution. Note that the non-HiFi reads are not shown.

Short Read Libraries

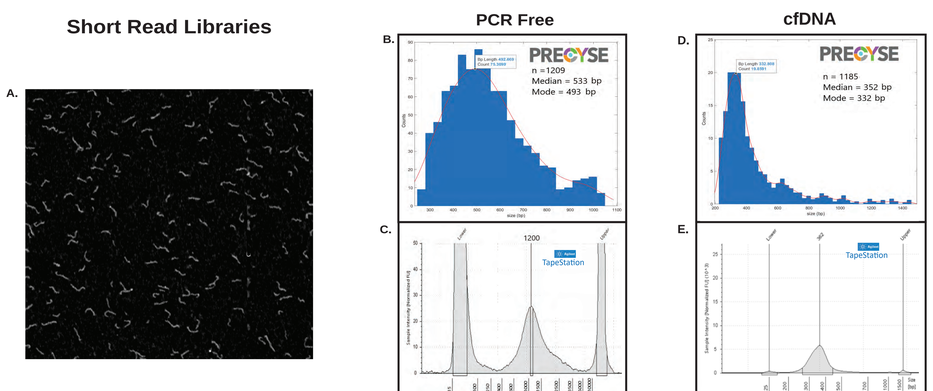


Figure 4. Short Read NGS Library QC

A. Example partial frame image from PRECYSE of a short read Illumina library. B. PCR-free mouse Illumina library provided by The Jackson Laboratory. Sizing analysis of 1209 molecules generated a median of 533 bp and mode at 493 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 overestimates 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 size selected targeting 250-450 bp (Sage Science). E. Corresponding TapeStation electropherogram of same cfDNA sample produced a median of 362 bp. TapeStation and PRECYSE primary sizing modes align well. PRECYSE was able to show improved sizing granularity with a second mode at ~620 bp, and larger molecules extending out to >1400 bp. Sizing of small DNA < 2 kb with PRECYSE requires < 50 pg of material. Sub pg analysis is possible with extended scanning times.

Conclusion

Evizia successfully completed Alpha instrument testing onsite with early access collaborators. PRECYSE performed well in standard laboratory settings, requiring only a sturdy lab bench. Hundreds of samples spanning numerous types of sample inputs and library preps for NGS were analyzed and compared by collaborators to Agilent Femto or TapeStation automated electrophoresis systems. We found that the sizing accuracy of short-read libraries was comparable with the exception of PCR free adapted samples, while being faster from loading to data acquisition. For the first time, collaborators directly observed sample molecules, secondary structure, and presence of sample contaminants and proteins. Including consumable preparation and analysis time, the instrument proved to be both faster (<17 min per sample vs. 30+ min) and more accurate (absolute error <15% vs 15%+) than electrophoresis QC in the analysis of DNA longer than 10,000 bp, including gDNA, sheared DNA, and long-read libraries. Importantly, we observed that electrophoresis QC consistently overestimates the true sizing of samples (as validated by the NGS read length output), as well as severely underrepresented smaller DNA populations. We were able to directly observe polymerase and helicase binding to sequencing library molecules by comparing fully ligated libraries with and without enzyme addition (data not shown). RNA integrity analysis was demonstrated showing the reduction of 18S and 28S populations related to quality and sample degradation with comparable results to electrophoresis (data not shown). Further, DNA from FFPE samples could be directly observed and sized. Across all sample types, the instrument was able to accurately size DNA and detect in samples which passed electrophoresis QC, various sample contaminants and nucleic acid secondary structure that could inhibit downstream applications including sequencing performance.

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.