APP NOTE | PURIGEN BIOSYSTEMS
VERSION 1.0 | JUNE 2020

Increased DNA Yield from FFPE Samples Purified Using the Ionic® System Reduces PCR Duplicates in NGS

Doug Fadrosh, Cameron Pospisil, Lewis Marshall, Randy Dyer, and **Klint Rose**
Purigen Biosystems | Pleasanton, CA

OVERVIEW

FFPE tissue samples are known to produce low yields of nucleic acid that is often of a lower quality by comparison to nucleic acid from fresh-frozen tissue, particularly samples that have been stored for several years or longer. When FFPE tissue is limited, conventional extraction and purification technologies may not produce enough high-quality DNA to meet the recommended library preparation input requirements for next-generation sequencing (NGS). To compensate for this, most library preparation methods use additional PCR cycles to create enough copies of the original DNA. This results in a higher number of PCR duplicates that reduce sequencing efficiency. Improving the quality and quantity of DNA extracted and purified from FFPE samples can greatly improve sequencing results and interpretation by reducing the number of PCR cycles needed during library preparation yielding more usable high-quality sequencing reads.

To compare the efficiency of DNA extraction and purification from FFPE samples, DNA was purified from a single 10 µm thick section from eight FFPE tissue blocks (18 years of age) containing breast, colon, or lung tissue using the Purigen Ionic® Purification System and a commercially available column-based kit. Adjacent scrolls were harvested from each block and processed separately by each method. The DNA purified by each method was quantified prior to NGS. The Ionic system produced a higher average yield from each sample to enable library preparation without additional PCR amplification with a resulting lower rate of PCR duplicates during next generation sequencing.

NUCLEIC ACID PURIFICATION
PURE AND SIMPLE™

PURIGEN™ BIOSYSTEMS

Ionic® Purification System

The Ionic Purification System is a benchtop platform that provides automated purification of DNA and RNA from biological samples using Ionic® Fluidic Chips and Purification Kits. Using a cutting-edge technology based on isotachophoresis (**FIGURE 1**), the Ionic system separates nucleic acids freely in solution, without binding to or stripping nucleic acid from physical surfaces. Taking as little as 3 minutes of hands-on time per sample to purify nucleic acid from a range of sample types including cells and FFPE tissue, the Ionic system is a highly efficient platform that produces more nucleic acid of higher purity, consistently and reliably.

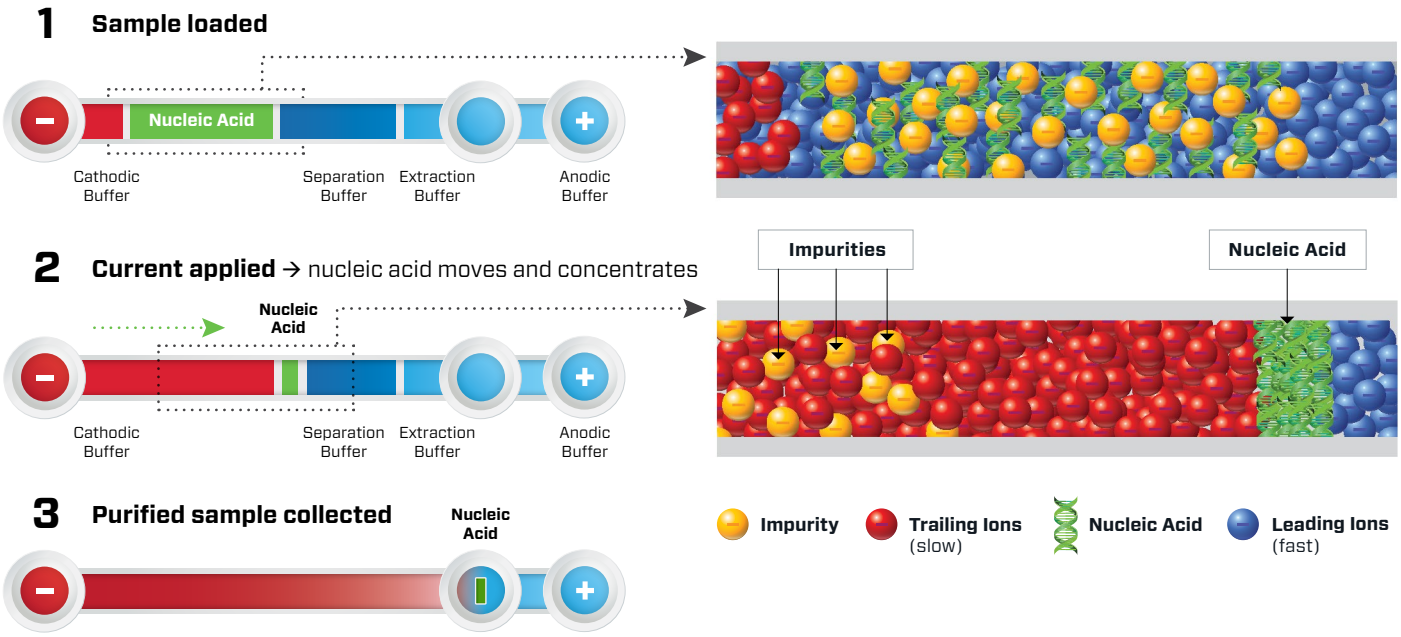

FIGURE 1: The Ionic system uses isotachophoresis to separate and isolate nucleic acid from samples. Isotachophoresis is conceptually represented in the following steps: **1)** Lysate containing nucleic acids is loaded into the microfluidic channel between a leading electrolyte containing ions with a faster electrophoretic mobility than nucleic acids and a trailing electrolyte containing ions with a slightly slower electrophoretic mobility. **2)** Electrical current is applied to the channel, causing nucleic acids to move through the solution toward the positive electrode. As this happens, the leading and trailing ions form a sharp electric field gradient, and nucleic acids focus toward the gradient. This is the ITP zone. Impurities are not collected into the zone because their mobilities are not bracketed by the leading and trailing ions. **3)** The band moves into an extraction well. The field gradient is detected by a sensor, and the system shuts down the electric current. The result is purified nucleic acid in a well accessible by a pipette tip.

PURIGEN™ BIOSYSTEMS

Improved Yield

To compare the yield of purified DNA from both the Ionic system and a column-based kit, two adjacent 10 µm thick sections were harvested from eight blocks (3 breast, 3 colon, 2 lung tissue). This provided a total of 3 replicate pairs of FFPE samples of breast and colon tissue and 2 replicate pairs of FFPE samples of lung tissue. High-quality human reference Coriell DNA from the GM24385 cell line was used as a "gold standard" control as input for library preparation. Up to 200 ng of the purified extracts from each method were brought up to a final volume of 55 µL in low TE buffer and sheared to a size range of 150 to 200 base pairs on a Covaris ultrasonicator using a Covaris microTUBE-50. All purified DNA was sheared for samples yielding <200 ng. Sheared DNA yield was quantified using the Qubit 1X dsDNA High Sensitivity Assay to determine precise input amounts for library preparation. For samples from FFPE blocks BQ, BS, CK, CW and CX, the sheared yield from samples purified by the Ionic system was over 300% greater than samples purified by the column-based kit (**TABLE 1**). Libraries were prepared according to the recommendations of the Agilent SureSelect XT HS Target Enrichment System for Illumina Paired-End Multiplexed Sequencing using the Cancer All-in-One Solid Tumor hybridization panel. Following the SureSelect XT HS library preparation recommendations, samples with a sheared yield between 100 ng and 200 ng were amplified with 11 cycles of PCR. Samples with a sheared yield between 50 ng and 100 ng were amplified with 12 cycles of PCR. Samples with a sheared yield between 10 ng and 50 ng were amplified with 14 cycles of PCR. The resulting amplified yields were pooled at equimolar concentrations and purified using Beckman AMPure XP purification reagent.

Block	Sample Type	Method	Purified DNA (ng)	Sheared DNA (ng)	Pre-capture PCR Cycles	Hybrid Capture Input (ng)	Final Concentration (nM)
C1	GM24385	Control	N/A	172.0	11	1000	15.4
C2	GM24385	Control	N/A	172.0	11	1000	11.7
BI	Breast	Ionic system	111.60	53.0	12	1000	13.5
BI	Breast	Column kit	124.65	60.5	12	1000	12.1
BQ	Breast	Ionic system	277.78	187.0	11	1000	21.4
BQ	Breast	Column kit	50.74	41.5	14	1000	13.1
BS	Breast	Ionic system	670.8	145.0	11	1000	10.1
BS	Breast	Column kit	24.596	15.6	14	1000	11.3
CK	Colon	Ionic system	722.4	200.0	11	1000	20.1
CK	Colon	Column kit	70.65	30.7	14	1000	21.1
CW	Colon	Ionic system	199.09	190.0	11	1000	7.29
CW	Colon	Column kit	27.563	24.1	14	1000	9.39
CX	Colon	Ionic system	203.82	170.0	11	803	15.2
CX	Colon	Column kit	32.25	30.4	14	1000	11.9
LW	Lung	Ionic system	219.73	194.0	11	1000	15.7
LW	Lung	Column kit	193.93	139.0	11	1000	13.6
LX	Lung	Ionic system	261.01	200.0	11	1000	15.4
LX	Lung	Column kit	86.86	81.0	12	1000	11.7

TABLE 1: Summary of properties for replicate sections from 8 FFPE tissue blocks purified by the Ionic system or a column-based kit for hybrid-capture based library preparation. Purified and sheared amounts were quantified by Qubit 1X dsDNA High Sensitivity Assay. PCR cycles were performed prior to hybridization capture according to SureSelect XT HS protocol recommendations. High-quality human reference Coriell DNA was used as a "gold standard" control.

Improved Next-Generation Sequencing Performance

All libraries were sequenced on the Illumina Nextseq system. Sequencing reads were down sampled for breast (15,844,370), colon (14,136,466) and lung (15,844,370) and analyzed through the Agilent SureCall v4.1.2 analysis pipeline using the hg38 human reference. On average, the total number of high-quality reads produced from extracts generated on the Ionic system that mapped to the human reference are 34% higher than extracts from a commercially available column-based kit (**TABLE 2**). This metric represents the number of reads retained after the removal of low-quality reads, reads that align to more than one region of the reference genome and identified PCR duplicate artifacts.

Block	Sample Type	Method	Reads in BAM File Passing Mapping Quality Filters	% Reads in Covered Regions	% Duplicate Reads	Sequenced Bases in Analyzable Target Regions	Avg. Read Depth in Analyzable Target Regions	Median Read Depth in Analyzable Target Regions
C1	GM24385	Control	1.24E+07	77.00%	15.92%	1.08E+09	1,469	1,447
C2	GM24385	Control	1.23E+07	74.78%	17.08%	1.04E+09	1,416	1,416
BI	Breast	Ionic system	6.78E+06	34.44%	49.68%	2.24E+08	306	294
BI	Breast	Column kit	7.12E+06	27.00%	39.98%	1.82E+08	248	219
BQ	Breast	Ionic system	1.08E+07	39.50%	20.10%	4.46E+08	608	612
BQ	Breast	Column kit	6.23E+06	26.18%	48.65%	1.57E+08	213	200
BS	Breast	Ionic system	4.81E+06	20.42%	61.68%	5.30E+08	724	735
BS	Breast	Column kit	4.81E+06	20.42%	61.68%	9.51E+07	129	122
CK	Colon	Ionic system	9.01E+06	46.37%	27.02%	4.23E+08	577	561
CK	Colon	Column kit	7.74E+06	36.59%	37.63%	3.00E+08	409	412
CW	Colon	Ionic system	7.02E+06	44.81%	41.09%	3.05E+08	416	416
CW	Colon	Column kit	5.43E+06	10.62%	42.17%	5.30E+07	72	67
CX	Colon	Ionic system	9.00E+06	28.16%	23.65%	2.45E+08	335	330
CX	Colon	Column kit	5.68E+06	13.76%	44.15%	7.07E+07	96	88
LW	Lung	Ionic system	1.05E+07	61.29%	26.74%	6.86E+08	936	936
LW	Lung	Column kit	9.51E+06	50.96%	30.67%	5.11E+08	697	661
LX	Lung	Ionic system	1.04E+07	56.29%	27.01%	6.20E+08	846	851
LX	Lung	Column kit	8.19E+06	50.14%	40.58%	4.29E+08	586	562

TABLE 2: SureCall summary statistics for libraries prepared from replicate sections of 8 FFPE tissue blocks purified by the Ionic system or a column-based kit.

The correlation between input amount, PCR cycles performed prior to hybridization capture, and the percentage of PCR duplicates generated by each library can be observed in **FIGURE 2**. Libraries from Ionic system extracts generated 42% fewer PCR duplicates than extracts from the column-based kit. This observation was most pronounced in the samples from FFPE blocks containing breast tissue where the libraries Ionic system extracts generated 51% fewer PCR duplicates by comparison.

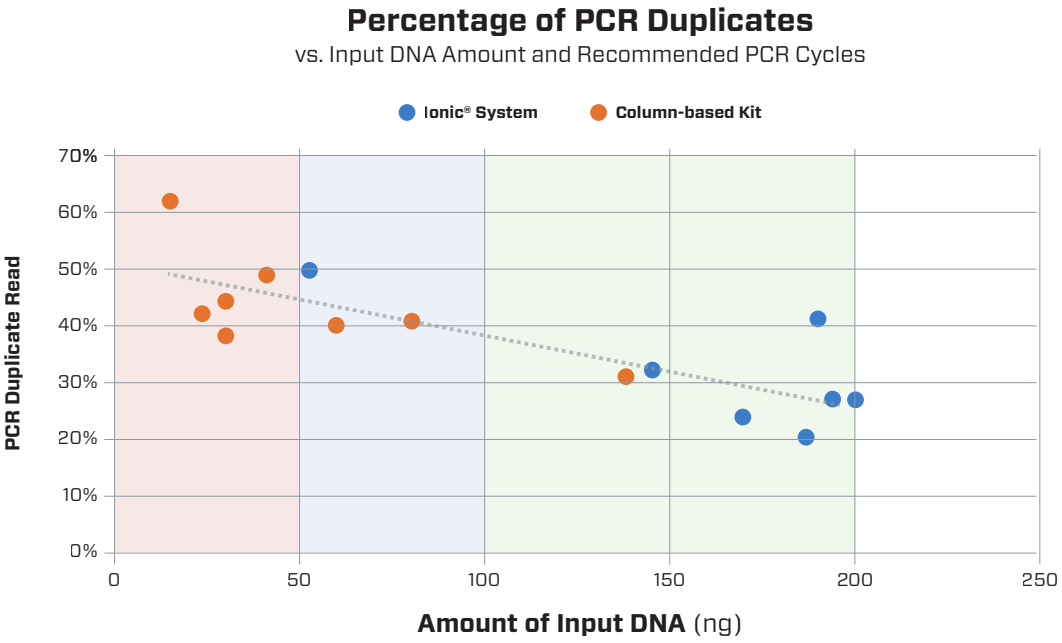

FIGURE 2: Comparison of the percentage of PCR duplicates reported for sequenced libraries from the 16 samples described in Table 1 to the amount of input DNA used for each library and the recommend PCR cycles for each amount. Libraries from extracts purified using the Ionic system are represented with blue dots. Libraries from extracts purified using a column-based kit are represented with orange dots. Shading indicates the recommended number of PCR cycles with red, blue and green representing 14, 12, and 11 respectively.

A common metric used to assess the success of hybridization capture sequencing is the total number of on-target reads. This metric is an indicator of the overall depth of coverage for target regions. A higher value enables a higher resolution for the identification of potential mutations. The ratios of high-quality aligned reads and on-target reads from libraries prepared using DNA extracted and purified by both methods indicates a clear advantage in both metrics using the Ionic system.

A higher number of reads aligned to the human reference was reported for libraries from Ionic system extracts from 7 out of the 8 FFPE blocks sampled, while all libraries generated from Ionic extracts yielded more on-target reads. The libraries generated from samples purified by the Ionic system produced more mapped reads and nearly a 100% average increase in the number of on-target reads, with multiple libraries yielding more than five times that of the column-based kit (**TABLE 2, FIGURE 3**). This increase in on-target coverage can enable detection of SNPs and other features at a much lower allele frequency while reducing costs by increasing the number of samples pooled per sequencing run.

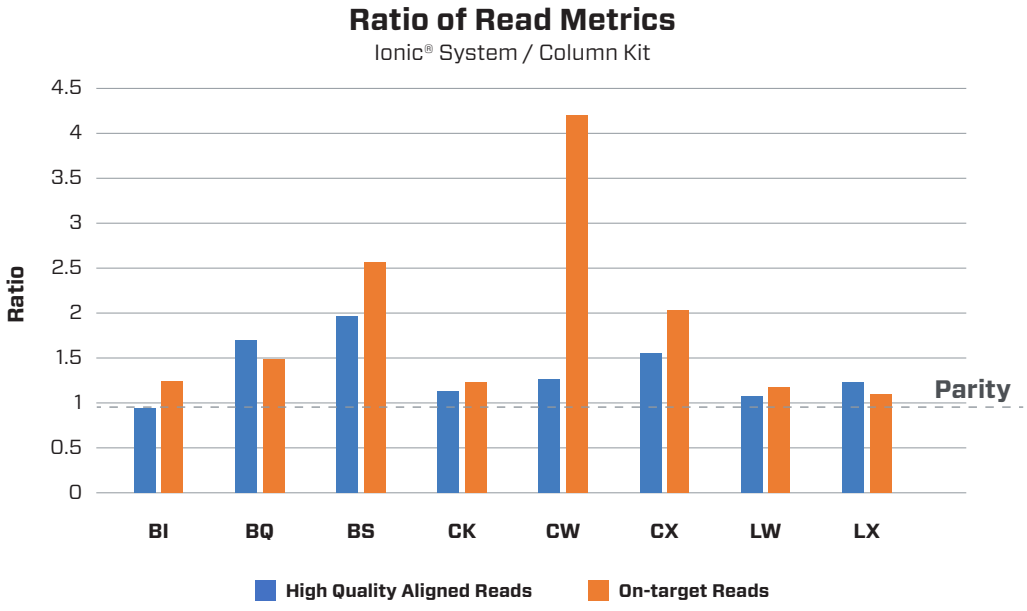

FIGURE 3: Comparison of Aligned Reads and On-Target Reads. The ratio of high-quality reads aligning to the human reference for libraries prepared from Ionic system extracts and libraries prepared from column-based kit extracts are shown with blue bars. The ratio of on-target reads for libraries prepared from Ionic system extracts and libraries prepared from column-based kit extracts are shown with orange bars. A ratio greater than 1 indicates a higher amount coming from the libraries prepared from Ionic system extracts.

To compare relative coverage of the targets, target hits were compared between the experimental FFPE samples and a control sample consisting of high-quality DNA (Coriell). To make this comparison, first the total number of reads per library was normalized. Following this, the ratio of reads in the experimental condition to reads in the control was calculated for each target. The $\log_{10}$ of these values were calculated to produce the final result. For this metric, values greater than zero indicate higher coverage in the experimental condition, and values less than zero indicate higher coverage in the control condition. We hypothesize that closely matching the control DNA indicates high sample quality.

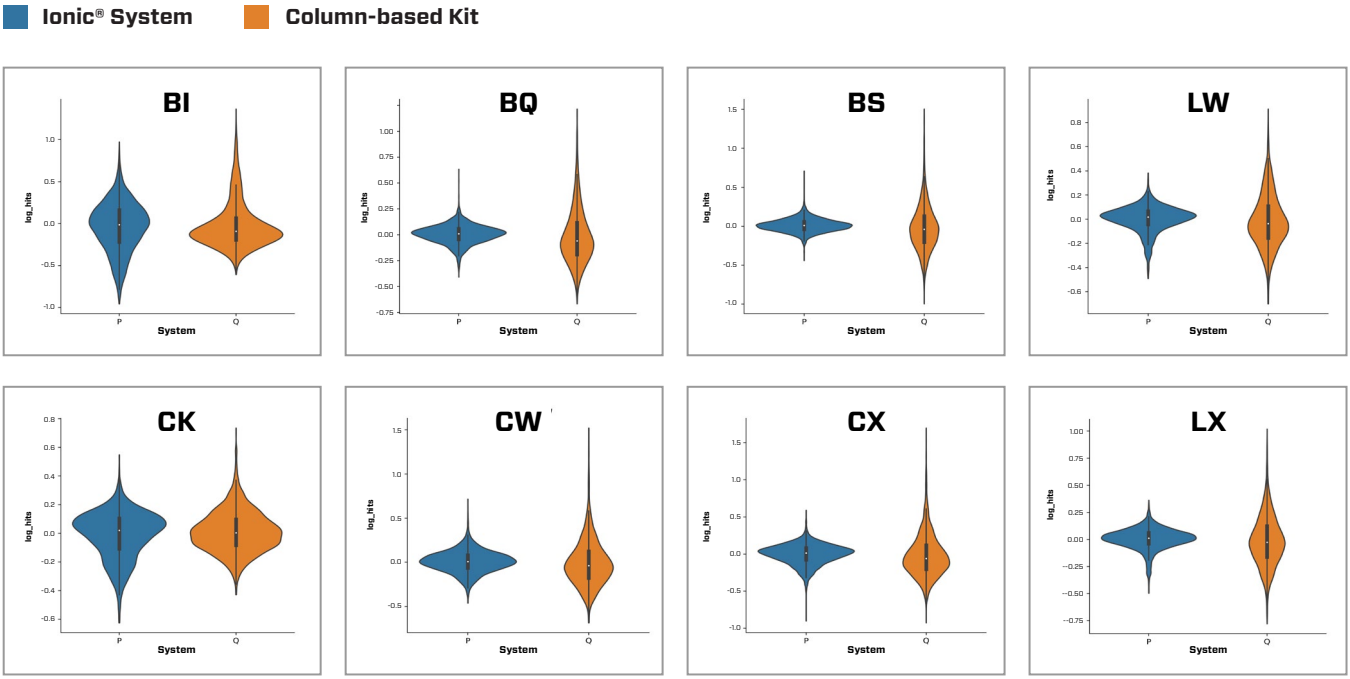

FIGURE 4: Comparison of On-Target Read Distribution Uniformity. The total number of reads per library was normalized for this comparison. The Log_{10} value for the ratio of reads in the experimental condition to a high-quality DNA control (Coriell) is shown for FFPE block processed. The log transformed ratio of sample to control for every target indicates an overall tighter on-target read distribution for libraries generated using the Ionic system extracts in comparison to those generated from the column-based kit extracts, indicating a distribution more similar to that of the control.

PURIGEN™ BIOSYSTEMS

To explore the differences in on-target read distribution further, a comparison of the GC content ratio compared to the control for samples from FFPE block BS is shown in **FIGURE 5**. Libraries prepared from samples purified by the Ionic system exhibit a distribution more closely resembling the control sample whereas the distribution from samples purified by column-based kits were more diffuse. For samples purified by column-based kits, a majority of targets are either underrepresented or overrepresented relative to the Coriell control which indicates a strong bias for targets with high GC content and against targets with low GC content.

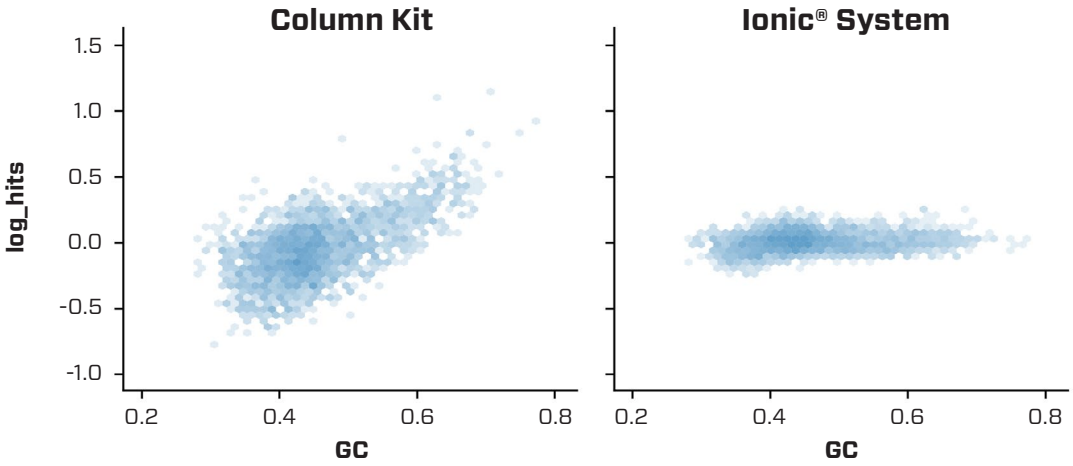

FIGURE 5: Comparison On-Target Read Distribution Uniformity and the GC-Content of each Target for representative block, BS. The distribution of reads for each target on the panel was compared to the high quality Coriell control library and plotted as the log transformed difference as a function of GC content. Darker color indicates a larger number of targets inside that bin. Libraries generated from the Ionic system extracts showed a distribution of reads more consistent with the control library (horizontal clustering centered on zero) when compared to libraries generated from column-based kit extracts.

SUMMARY

Use of excessive PCR cycles to artificially generate enough input DNA from low yielding samples during hybrid capture based NGS sequencing applications is known to increase duplicate reads and downstream analytical overhead, thereby increasing the cost of each read. This can best be addressed by increasing the yield of purified DNA to reduce the number of PCR cycles required to generate sufficient mass during library preparation. The comparison of DNA extracted and purified from a limited input of FFPE material demonstrates that the higher yield achieved by the Ionic system can help to reduce the number of PCR cycles required to enrich the sample prior to hybridization capture for NGS. This reduces the number of PCR duplicates in the resulting sequence data. Furthermore, samples purified using the Ionic system appear to be less biased toward GC-content and result in a more uniform sequence coverage of diverse targets.

NUCLEIC ACID PURIFICATION
PURE AND SIMPLE™

PURIGEN BIOSYSTEMS, INC.
5700 Stoneridge Drive, Suite 100
Pleasanton, CA 94588
TEL: +1 877 PURIGEN (787-4436)

SALES
EMAIL: sales@purigenbio.com

SUPPORT
EMAIL: support@purigenbio.com
TEL: +1 877 PURIGEN (787-4436)

FOR RESEARCH USE ONLY. Not for use in diagnostic procedures.

PURIGEN™ BIOSYSTEMS