Optimizing RNA-seq data quality and costs for FFPE samples with the xGen™ Broad-Range RNA Library Preparation Kit

Summary

This application note illustrates how the xGen Broad-Range RNA Library Preparation Kit enables multiple workflow options to maximize the output of RNA-seq data from FFPE samples, including upstream ribodepletion of total RNA, and downstream enrichment of total RNA libraries by hybridization capture. For the analysis of coding sequence information, a four-fold savings in sequencing cost per sample can be achieved using the hybridization capture workflow. The details and methods presented here allow core facilities, production labs, and research scientists to choose the best workflow for their sequencing goals while reducing costs and time associated with RNA-seq library preparation and sequencing.

- For FFPE samples, DV_{200} is more informative than RIN score
- Ribodepletion preserves coding and non-coding RNA but is not cost-effective for coding sequence analysis of FFPE samples because it can only obtain approximately 20–30% exonic reads
- Hybridization capture provides a more cost-effective workflow for coding sequence analysis, with 6-plex library capture resulting in a two-fold savings for FFPE RNA-to-FASTQ costs
- Comparative analysis of both workflows including RNA recovery, kits and reagents used, preparation time, and costs per sample

Introduction

Most of today's next generation sequencing (NGS) libraries are made from RNA samples for the quantitative evaluation of biological function. Among the most prevalent RNA-seq samples are RNA isolated from human tissues preserved through formalin-fixation and paraffin-embedding (FFPE). Although this practice is effective for long-term sample preservation, it introduces several challenges for NGS such as, extracting high-integrity RNA, achieving sufficient library yield, and generating quality data. Commonly utilized techniques for RNA-seq library preparation from FFPE samples are vulnerable to crosslinked or fragmented RNA substrates, as well as to an increased representation of intronic reads in the final library, an observation that is consistently seen for formalin-fixed samples[1]. In addition, libraries produced from limited material suffer from high duplication rates and adapter dimers, along with lower mapping rates and fewer genes and transcripts detected.

As a solution for FFPE samples, we present the xGen Broad-Range RNA Library Preparation Kit. This stranded RNA-seq kit leverages patented Adaptase™ technology—a method for ligating adapters to single-stranded substrates to directly convert first-strand cDNA into NGS libraries in about 4.5 hours (Figure 1). This approach saves time, guarantees stranded libraries without requiring conventional second-strand cDNA synthesis or degradation, and maximizes yield and complexity while minimizing bias and undesired library byproducts. Moreover, the xGen Broad-Range RNA Library Prep Kit can make quality libraries from as little as 10 ng total FFPE RNA, allowing it to accommodate samples that have a limited starting input.

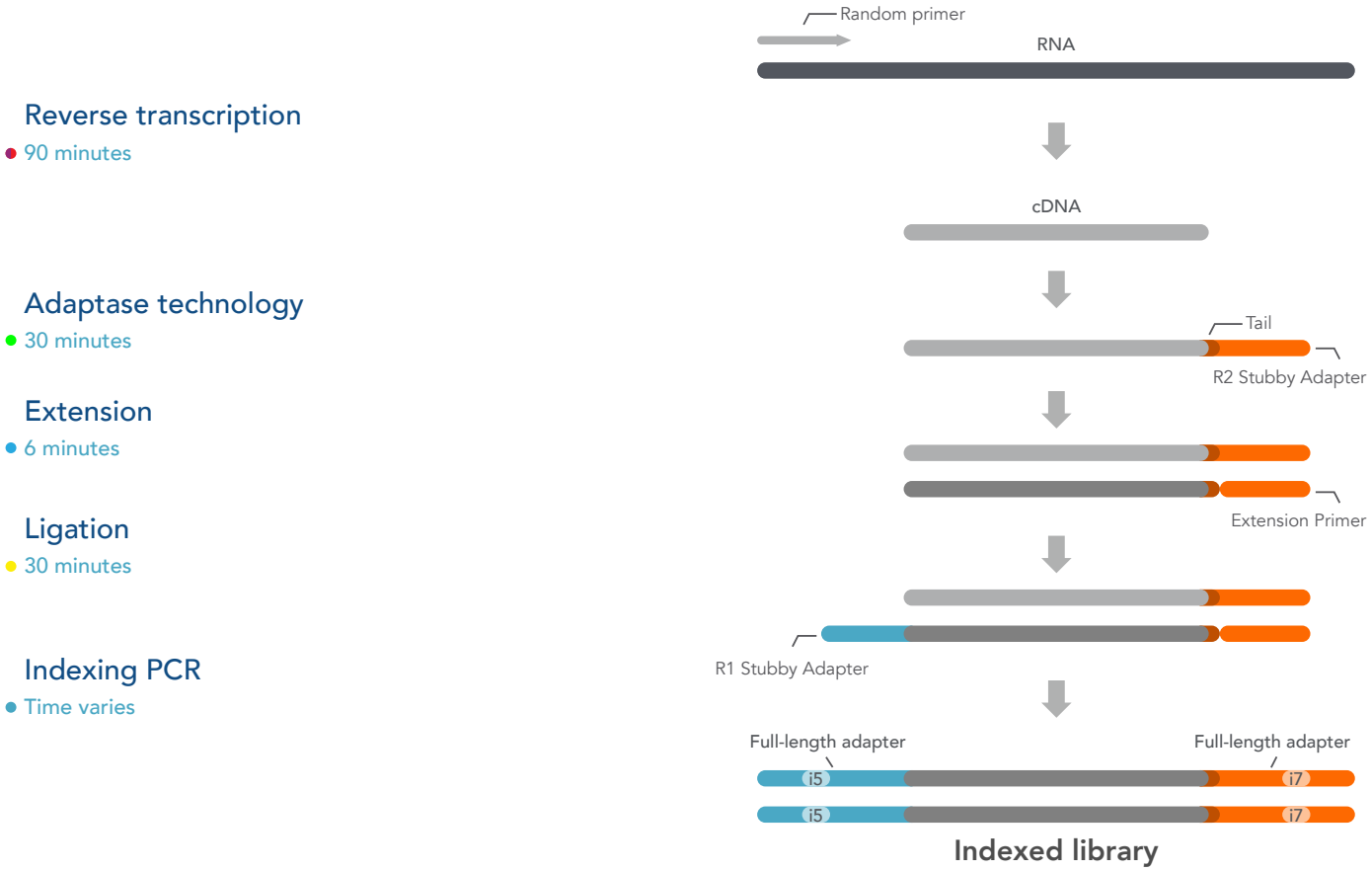

Figure 1. IDT xGen Broad-Range RNA Library Preparation Kit workflow.

Most labs consider the depletion of ribosomal RNA (rRNA), also known as ribodepletion, as the default approach for processing FFPE RNA samples, despite its high costs in reagents and sequencing. Beside the sequencing costs, another drawback is that the poly(A)-selection would only provide 3′ coverage of fragmented transcripts. Alternatively, hybridization capture can instead be used to produce quality data from FFPE RNA-seq libraries. The IDT Broad-Range RNA Library Prep Kit is compatible with both FFPE workflows (Figure 2). The choice between ribodepletion and hybridization capture should be informed by the RNA types and sequencing data of interest. Here, we discuss each workflow and provide representative data from Universal Human Reference (UHR) RNA and multiple breast cancer tumor FFPE samples. We also provide a comparison of sample types, workflow time, and library and sequencing costs to help inform your decision for maximizing recovery and data quality while minimizing costs.

RIBODEPLETION WORKFLOW

Depletion	• Kit: Lexogen RiboCop V1.3* • Purpose: removes rRNA (makes up 80–90% of total RNA), leaving mRNA & non-coding RNA
Library Prep	• Kit: **xGen Broad-Range RNA Library Preparation Kit** • Purpose: converts mRNA & non-coding RNA to NGS libraries
Sequencing	• Kit: Illumina® Sequencing Kits • Purpose: generates sequencing data from ribodepleted Swift RNA libraries

HYB CAPTURE WORKFLOW

Library Prep	• Kit: **xGen Broad-Range RNA Library Preparation Kit** • Purpose: converts total RNA, including rRNA and non-coding RNA, into NGS libraries
Hyb Capture	• Kit: xGen Exome v2 Hyb Panel* • Purpose: captures exonic library molecules and washes away rRNA fragments
Sequencing	• Kit: Illumina® Sequencing Kits • Purpose: generates sequencing data from enriched IDT RNA libraries

Figure 2. RNA-seq library workflows: upstream ribodepletion vs. downstream enrichment.

RNA characterization

The quality of input RNA into an RNA-seq workflow can substantially impact the library yield and resulting data quality. Because FFPE RNA is often degraded, assessing the starting quality is important for adjusting the RNA-seq workflow to optimize results. RNA quality and integrity can be evaluated using these two metrics:

- **RIN score** (RNA Integrity Number; ratio of the 28S to 18S rRNA peaks)
- **DV_{200}** (percentage of RNA fragments that are greater than 200 nucleotides)

RIN scores are useful for determining whether a poly(A)-selection module can be used, as they are typically only recommended for samples with RIN ≥ 7. High-quality samples typically have RIN scores ≥7, whereas FFPE samples often have RIN scores <7 (Figure 3). For samples with RIN <7, ribodepletion and hybridization capture are the preferred workflows. For these samples, more information about sample quality can be obtained through the DV_{200} score which can help inform the fragmentation time, SPRI ratio, and PCR cycling. In general, it is not advised to make libraries using samples with a RIN < 2 or a DV_{200} < 30.

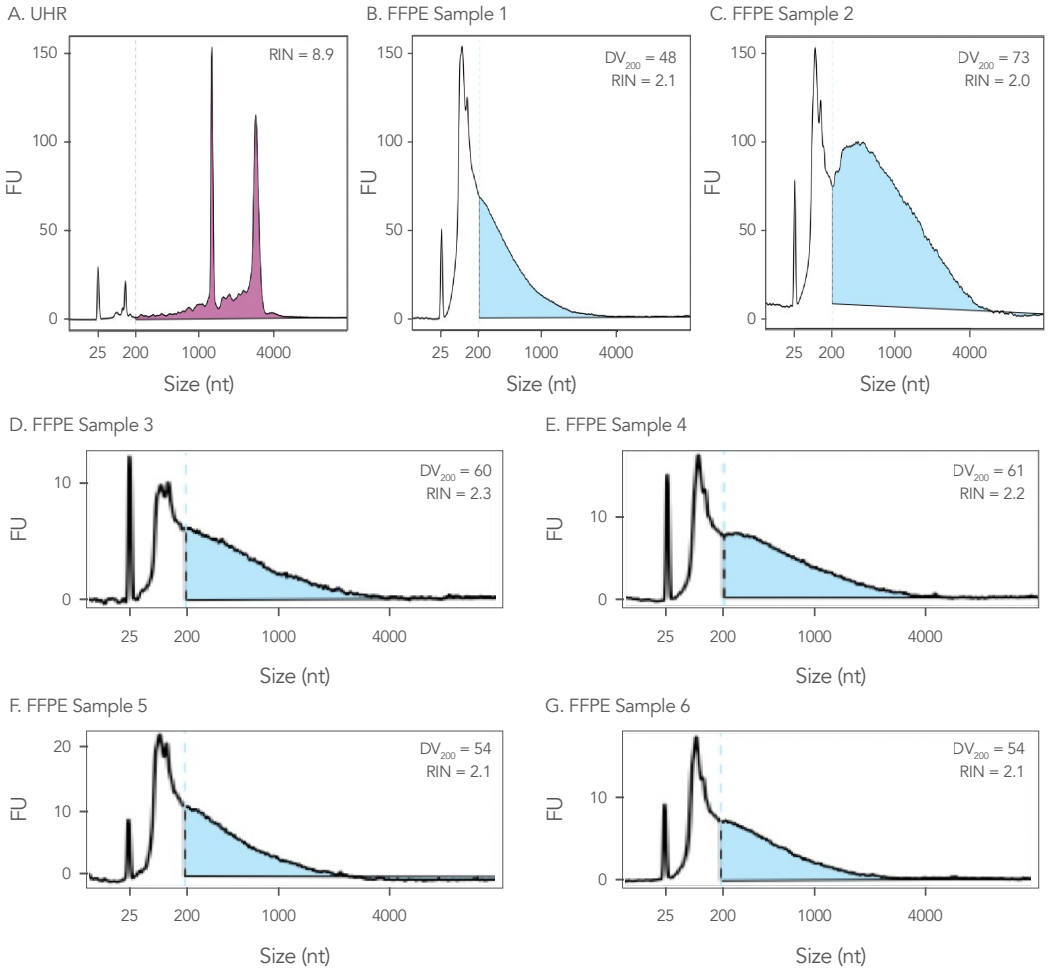

Figure 3. RNA Bioanalyzer traces for UHR and FFPE RNAs.

RNA characterization method

We obtained FFPE curls from breast cancer tumors (Spectrum Health; Grand Rapids, MI) and extracted the RNA using the RNeasy FFPE Kit (Qiagen 73504). Isolated RNA was analyzed using an RNA 6000 Pico Kit (Agilent 5067-1513) on the Agilent Bioanalyzer. Trace analyses show all samples have a RIN score near 2, whereas the DV_{200} scores range from 48 to 73 (Figure 3), providing a more accurate depiction of RNA integrity.

Ribodepletion workflow

Ribodepletion of FFPE RNA is the recommended workflow when coding and noncoding sequence information is desired. For intact RNA samples, poly(A) selection is considered a simpler and lower-cost approach to enrich the sample for mRNA. However, because FFPE samples are damaged, mRNA transcripts may be broken away from their poly(A) tails, preventing this method from enriching for transcripts in their entirety. Ribodepletion, however, can work effectively on FFPE samples even if they are damaged because it depletes rRNA rather than selects for mRNA, therefore enabling the recovery of fragments without poly(A) tails. Additionally, because ribodepletion leaves behind all non-rRNA molecules, it is the workflow to choose for those interested in intronic regions, such as those from nuclear pre-mRNA and long non-coding RNAs.

Here, we present data from UHR RNA and two FFPE RNA samples with variable DV_{200} scores (Samples 1 and 2, Figure 3) that underwent ribodepletion prior to library prep using the xGen Broad-Range RNA Library Preparation Kit in duplicate. Using 100 ng of total RNA, all samples showed robust library yields with high mapping rates and preservation of transcript expression, as evidenced by a high correlation of the ERCC spike-in2 transcripts (Table 1). Further, FFPE libraries detected an ample and consistent number of genes, despite variability in DV_{200} scores (Table 1). More genes were detected for UHR RNA, likely due to increased RNA integrity, higher sample complexity, and a higher proportion of exonic reads.

Table 1. RNA-seq data for ribodepleted FFPE samples.

Sample	Library yield (nM)	STAR mapping rate (%)	Genes detected	Exonic rate (%)	rRNA rate (%)	ERCC rate (%)
UHR RNA	44.2	86.2	17,132	46.0	3.4	0.94
	33.6	87.5	17,180	46.2	3.2	0.92
FFPE Sample 1	14.0	83.6	15,095	22.0	6.7	0.91
	19.8	85.7	15,112	21.2	6.0	0.92
FFPE Sample 2	24.9	84.2	15,274	24.6	4.8	0.92
	23.9	85.5	15,325	25.7	4.5	0.94

Exonic rates are typically low for ribodepleted RNA (Table 1). Because ribodepletion only removes rRNA, it allows intronic and intergenic RNA to become library molecules, thereby reducing the representation of exonic fragments (40–50% for UHR RNA). The exonic rate for FFPE samples is even lower due to the bias toward intronic reads that comes as a byproduct of formalin-fixation [1], as well as tissue-specific profiles, such as an increased prevalence of intronic reads in breast cancer tumors [3] (Figure 4).

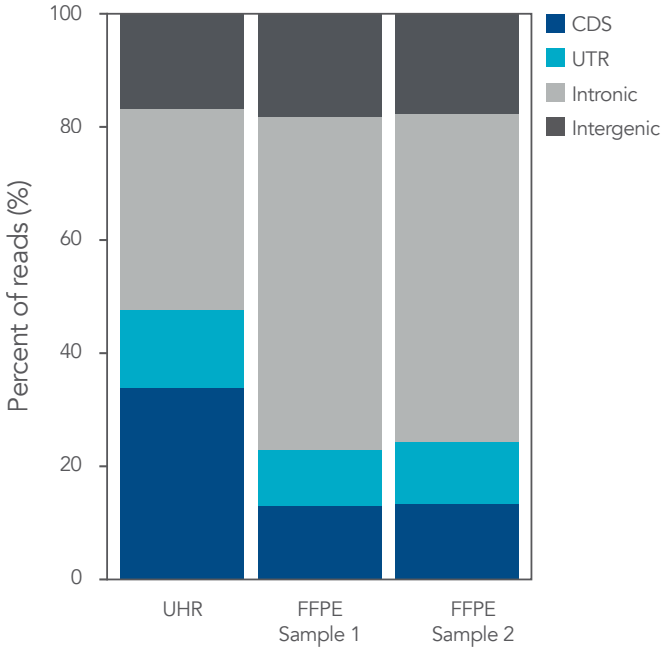

Figure 4. Read distribution for ribodepleted RNA-seq libraries

Ribodepletion workflow method

50–100 ng of UHR and FFPE Samples 1–2 were ribodepleted using Lexogen Ribocop4 V1.3 and eluted in 5 µL dH_2O for input into the xGen Broad Range RNA Library Preparation Kit [5]. RNA libraries were made with the following adjustments: fragmentation at 65°C for 5 min and a 1.2X post-RT SPRI clean-up (see Protocol considerations). All libraries were amplified using 12 PCR cycles. Libraries were sequenced on a MiniSeq 2 x 75 bp and downsampled to 3 million reads before analysis with STAR [6], Picard [7], RNA-SeQC [8], and fgbio [9]. CDS = coding sequence; UTR = untranslated region (5′ and 3′).

Hybridization capture workflow

Hybridization capture of FFPE RNA-seq libraries enables the capture and sequencing of specific regions of the transcriptome. For example, whole exome panels capture only exonic fragments, maximizing the representation of coding regions in the sequencing data. Alternately, projects with narrower focus can use research panels designed to target cancer or inherited disease genes (xGen Inherited Diseases Hybridization Panel), among others. Although hybridization capture is a longer workflow compared to ribodepletion, it can dramatically reduce overall costs by capturing and sequencing only regions of interest (see Workflow tradeoffs).

Here, we present data from UHR and FFPE RNA (Samples 1 and 2, Figure 3) processed using the xGen Broad-Range RNA Library Prep Kit with hybridization capture adjustments (see Protocol considerations). Libraries were enriched for human coding sequences using the xGen Hybridization and Wash Kit and the xGen Exome Hybridization Panel [10].

This exome panel contains over 400,000 probes and spans a 39 Mb target region comprising 19,396 genes. Importantly, using probes designed specifically to the exome more than triples the exonic rate compared to ribodepleted samples (>80% compared to ~25%, respectively; see Tables 1–2 and Figure 5). Additionally, hybridization capture provides consistent data metrics for both intact (UHR) and damaged (FFPE) samples with a similar number of genes being detected (Table 2). Note that Table 2 data is at a sequencing depth of 3 million reads per sample; deeper sequencing captures all 19,396 genes targeted by the xGen Exome Hybridization Panel.

Table 2. RNA-seq data for hybridization capture with 100 ng input.

Sample	Library yield (ng/µL)	STAR mapping rate (%)	Genes detected	Exonic rate (%)	rRNA rate (%)	Selected basis (%)
UHR RNA	82	89.6	17,758	91.2	1.3	96.7
FFPE Sample 1	93	88.8	17,009	89.2	0.9	96.5
FFPE Sample 2	73	87.9	17,024	88.1	1.0	95.8

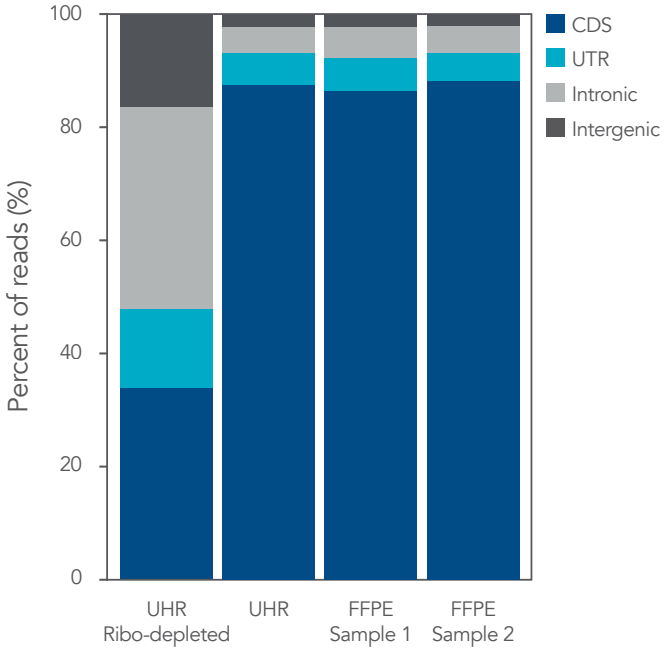

Figure 5. Read distribution for hybridization capture of RNA-seq libraries.

Robust data across FFPE samples

The variability in sample quality, particularly for FFPE samples, often acts as a hurdle to producing consistent, quality libraries. To evaluate a wider range of samples, we performed hybridization capture on RNA libraries from four additional breast cancer tumor FFPE samples with DV_{200} scores ranging from 54 to 62 (Figure 3). Because FFPE samples are often limited in yield following RNA extraction, we also tested the ability of the workflow to support an input of 10 ng FFPE RNA.

RNA libraries were obtained by hybridization capture with the xGen Exome Hybridization Panel resulting in sequencing data with high yields, mapping rates, and exonic rates across a range of FFPE samples (Table 3). Further, libraries also produced high-quality data despite the lower input of 10 ng.

Multiplex up to six libraries

RNA libraries can be combined before hybridization capture to maximize the value of each reaction. For DNA libraries, samples can often be multiplexed with up to 16 samples. However, because transcripts are expressed at variable levels, RNA libraries run the risk of depleting probes for highly expressed genes, thereby limiting the number of samples that can be multiplexed per capture.

Here, we show FPKM (**F**ragments **P**er **K**ilobase of transcript per **M**illion mapped reads) plots for UHR and FFPE RNA libraries that have been multiplexed with 2, 4, or 6 samples. Linear regression of the same library with different multiplexing shows very high correlation (Figure 6). Multiplexing can reduce the overall cost of each sample (see Workflow tradeoffs).

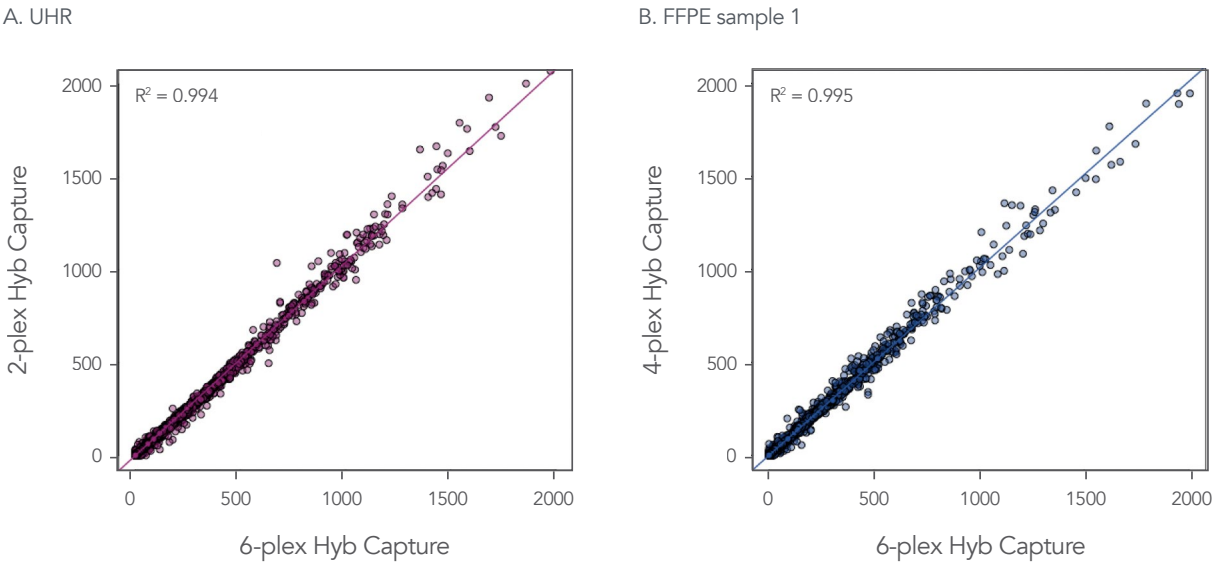

Figure 6. FPKM value correlation of multiplexed hybridization capture libraries.

Hybridization capture methods

Either 100 ng (Table 2) or 10 ng (Table 3) of the respective FFPE Samples (DV_{200} scores from 48–73, Figure 3) were used as input into the xGen Broad-Range RNA Library Preparation Kit [5]. Libraries were made with fragmentation at 94°C for 15 min (UHR), or 2 min (FFPE). All SPRI clean-ups in the protocol were adjusted to 1.8X. 100 ng libraries used 11 pre-hyb PCR cycles and 10 ng libraries used 15 pre-hyb PCR cycles before input into the xGen Exome Hyb Panel [10]. Pooled samples used either 9 (2-plex, Figure 6), 8 (4-plex, Tables 2–3), or 7 (6-plex, Figure 6) post-hyb PCR cycles. Libraries were sequenced on a MiniSeq 2 x 75 bp and downsampled to 3 million reads before analysis with STAR [6], Picard [7], and RNA-SeQC [8]. CDS = coding sequence; UTR = untranslated region (5′ and 3′).

Table 3. RNA-seq data for hybridization capture with 10 ng input.

Sample	Library yield (ng/µL)	STAR mapping rate (%)	Genes detected	Exonic rate (%)	rRNA rate (%)	Selected basis (%)
FFPE Sample 3	23.0	89.4	16,089	87.4	1.5	95.1
FFPE Sample 4	54.0	88.0	16,065	88.7	1.3	95.6
FFPE Sample 5	33.4	87.5	15,653	87.6	1.3	94.6
FFPE Sample 6	56.0	87.6	16,246	88.4	1.2	95.4

Workflow tradeoffs

Two workflows were presented in this application note for the purpose of processing FFPE samples:

1. **Ribodepletion.** Involves the removal of rRNA prior to library prep.
2. **Hybridization capture.** Involves enriching exonic library molecules (e.g., using a whole exome panel) from total RNA libraries.

Table 4 summarizes these two workflows, including the types of RNA recovered, reagents required, workflow details, and costs.

Table 4. Workflow and cost comparison for ribodepletion vs. hybridization capture.

Workflow	Ribodepletion-based	Hyb-capture based
RNA Types Recovered	Coding and non-coding RNA	Coding RNA
Kits Used	1. Lexogen RiboCop rRNA Depletion Kit for Human/ Mouse/Rat (HMR) V2 H/M/R (144.96) 2. xGen Broad-Range RNA Library Preparation Kit (10009865/10009813) 3. xGen UDI Primers Plate 1, 8nt (10005922/X9096-PLATE)	1. xGen Broad-Range RNA Library Preparation Kit (10009865/10009813) 2. xGen UDI Primers Plate 1, 8nt (10005922/X9096-PLATE) xGen Exome Hyb Panel v2 (10005151/10005152/10005153) 3. xGen Hybridization and Wash Kit (1080577/1080584) 4. xGen Universal Blockers (TS 1075474/1075475/1075476; 10bp TS 1081100/1081101/1081102; NXT 1079584/1079585/1079586) 5. xGen Library Amplification Primer Mix (1077675/1077676) 6. KAPA 2X HiFi HotStart ReadyMix (KK2602)
Number of Steps	20	28
Library Prep Time (RNA-to-Library)	6 hours	12 hours
Library Price Per Sample (RNA-to-Library)	$61	$104 (6-plex libraries per capture)
Sequencing Price Per Sample Novaseq SP 100 cycle (Library-to-Fastq, 3 0M exonic reads)	$472	$121
Total Price Per Sample (FFPE RNA-to-Data, $100/hr FTE)	$566	$284

*Estimates are based on US 2020 List prices and assume 24 samples (4 captures) are being prepared simultaneously.

Protocol considerations

FFPE Samples are often degraded and require workflow-specific adjustments to the protocol that will optimize the yield and quality of the final libraries. Follow these guidelines and see Table 5 for more information.

Table 5. xGen Broad-Range RNA Library Preparation Kit adjustments for ribodepletion or hybridization capture.

Step	Ribodepletion	Hybridization
Fragmentation Time	10 min at 94°C (Intact RNA) 5 min at <65°C (FFPE)	15 min at 94°C (Intact RNA) 2 min at 94°C (FFPE)
SPRI Clean-ups	1.2X Post-RT All others the same	1.8X for ALL
PCR Cycles	Follow Total RNA recommendation	Add 4 cycles to mRNA recommendation

Fragmentation time

RNA is fragmented following a chemical- and heat-based process. The amount of fragmentation can be adjusted by increasing or decreasing the fragmentation time and/or temperature. Because FFPE RNA is often degraded, it is already fragmented to a certain extent. The DV_{200} score can help determine how fragmented the RNA is. For samples with low DV_{200} scores, reduce the temperature and time when following the ribodepletion workflow. When following the hybridization capture workflow, reduce only the time (see Table 5). These recommendations can be further adjusted to optimize the fragmentation based on specific samples or insert size requirements.

SPRI cleanup ratios

SPRI cleanups act as a size selection mechanism as well as a buffer exchange. A more relaxed SPRI (i.e., a higher ratio of beads to sample volume) will result in carry over of smaller fragments while a more stringent SPRI (i.e., a lower ratio of beads to sample volume) will select for larger fragment sizes. Because FFPE RNA is often degraded, relaxing the SPRI ratio after reverse transcription (ribodepletion workflow) or throughout the protocol (hybridization capture workflow) preserves smaller fragments that can then be converted into library molecules (see Table 5).

A more relaxed SPRI ratio, however, can also enable the carry-over of adapter dimers, library artifacts that arise when two adapter molecules become ligated together. Adapter-dimers can cluster during sequencing to take up valuable sequencing reads and are often exacerbated at lower inputs. For the ribodepletion workflow, if adapter dimers are present in the final libraries (as evidenced by a peak near ~150 bp on an electrophoretic instrument), a second post-PCR SPRI can be performed. For the hybridization capture workflow, adapter dimers do not present an issue as they will not be captured during the hybridization steps. Libraries that are meant for hybridization capture can utilize relaxed 1.8X SPRI clean-ups without adapter dimer carry-over concerns.

PCR cycles

The number of PCR cycles can be optimized to fit your yield requirements. In general, the cycle recommendations for total RNA can be followed when using the ribodepletion workflow. The hybridization capture workflow typically requires 200 ng of each library; thus, 4 PCR cycles should be added to the mRNA recommendations, even though total RNA is used as the input (see Table 5).

Note: Excess PCR cycles, or overamplification of the libraries, can result in the formation of a heteroduplex. This can be evidenced as a second peak >1000 bp. Heteroduplex formation can result in inaccurate quantification of library molecules and negatively impact final data quality. If a heteroduplex is observed, reduce the number of PCR cycles.

Conclusion

The xGen Broad-Range RNA Library Preparation Kit is a fast, consistent, flexible solution for converting FFPE RNA into high-quality NGS libraries. We presented workflow options and summarized the resulting tradeoffs to data and costs, as well as highlighted protocol modifications required to implement each workflow. A compelling price per sample savings advantage is presented through the hybridization capture workflow relative to ribodepletion when exonic transcript information is the desired data output. Due to the high variability in quality and quantity of RNA derived from FFPE, this application note presents data to support the robustness of the xGen Broad-Range RNA Library Preparation Kit with challenging and low input RNA samples.

References

1. Marczyk M, Fu C, Lau R, *et al*. The impact of RNA extraction method on accurate RNA sequencing from formalin-fixed paraffin-embedded tissues. *BMC Cancer*. 2019;19(1):1189.
2. Jiang L, Schlesinger F, Davis CA, *et al*. Synthetic spike-in standards for RNA-seq experiments. *Genome Res*. 2011;21(9):1543-1551.
3. Sinicropi D, Qu K, Collin F, *et al*. Whole transcriptome RNA-seq analysis of breast cancer recurrence risk using formalin-fixed paraffin-embedded tumor tissue. *PLoS One*. 2012;7(7):e40092.
4. RiboCop rRNA Depletion Kit V1.3. Lexogen GmbH. https://www.lexogen.com/store/ribocop-rrna-depletion-kit/
5. xGen™ Broad-Range RNA Library Preparation Kit. Integrated DNA Technologies. https://www.idtdna.com/pages/products/next-generation-sequencing/workflow/xgen-ngs-library-preparation/rna-library-preparation/broad-range-rna-library-prep-kit
6. Dobin A, Davis CA, Schlesinger F, *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
7. "Picard Toolkit.". *Broad Institute*, *GitHub repository*: Broad Institute; 2019.
8. DeLuca DS, Levin JZ, Sivachenko A, *et al*. RNA-seqC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*. 2012;28(11):1530-1532.
9. fgbio tools. *Github repository*.
10. xGen™ Exome Hybridization Panel. Integrated DNA Technologies. https://www.idtdna.com/pages/products/next-generation-sequencing/workflow/xgen-ngs-hybridization-capture/pre-designed-hyb-cap-panels/exome-hyb-panel-v2

Revision history

Version	Release date	Description of changes
2	June 2024	Minor updates on page 9
1	June 2022	Initital release

Optimizing RNA-seq data quality and costs for FFPE samples with the xGen™ Broad-Range RNA Library Preparation Kit

For more information, go to: **www.idtdna.com/ContactUs**

> For more than 30 years, IDT's innovative tools and solutions for genomics applications have been driving advances that inspire scientists to dream big and achieve their next breakthroughs. IDT develops, manufactures, and markets nucleic acid products that support the life sciences industry in the areas of academic and commercial research, agriculture, medical diagnostics, and pharmaceutical development. We have a global reach with personalized customer service.

> SEE WHAT MORE WE CAN DO FOR **YOU** AT **WWW.IDTDNA.COM**.

Doc ID: RUO22-0944_001.1 06/24

INTEGRATED DNA TECHNOLOGIES