A streamlined workflow for targeted RNA-seq using the xGen™ RNA Library Prep Kit with xGen Custom Hyb Panels

Abstract

RNA-seq provides important information about gene expression and can be used to better understand the expression signatures of various cancers. However, researchers generally must perform either ribosomal RNA (rRNA) depletion or poly-A selection to sequence messenger RNA (mRNA) deeply enough to study the characteristics of cellular expression profiles. The combined use of the xGen RNA Library Prep Kit with xGen Custom Hyb Panels in a target enrichment workflow allows researchers to use total RNA as input material without depleting rRNA or using poly-A selection. More specifically, the IDT xGen Custom Hyb Panels can be designed to target and enrich genetic regions of interest, thus eliminating the need for rRNA depletion and poly-A selection. The data presented here compares two approaches for designing xGen Custom Hyb Panels for targeted RNA-seq, which both provide a streamlined, accurate, and reproducible approach to RNA-seq using total RNA.

Introduction

RNA-seq is a high-throughput technology to profile transcriptomes, allowing researchers to better understand how genes are expressed under different conditions. RNA-seq uses Next Generation Sequencing (NGS) to generate an immense amount of data that allows for the quantification of transcripts, and the identification of sequence alterations, both of which are critical in understanding transcriptome dynamics (gene expression) during diseased and normal states. Cancer is a well-known example of what can happen to cellular functions when expression changes. These changes can be attributed to various mechanisms such as copy number variants, single nucleotide variants, splicing alterations, and gene fusions. Different cancers have different associated expression signatures, and by researching these differences, it will be possible to identify cancer and therefore potential therapeutic options earlier [1].

One of the difficulties of RNA-seq is the cost associated with sequencing deep enough to get enough coverage for low-expression transcripts and rare structural variants [2]. A common hurdle to generating enough sequences for these transcripts of interest is to avoid sequencing highly expressed transcripts from ribosomal RNA (rRNA) [3]. To do so, rRNA-depletion or poly-A selection is often a necessary step of library preparation before sequencing. However, this step can be cumbersome, inefficient, and can result in the loss of valuable transcripts [3]. An alternative technique that does not require rRNA depletion and can lower cost while maintaining the high-throughput capabilities of RNA-seq is a targeted approach. Targeted RNA-seq allows researchers to focus on specific genetic regions of interest, while not sequencing unnecessary areas of the transcriptome.

IDT has designed a targeted xGen RNA-seq workflow that includes a fast and low-cost prep kit for constructing RNA libraries from the 1st strand cDNA–without needing the 2nd strand cDNA (**Figure 1**). Because of the integration of the xGen Hyb Capture Panels in the workflow, total RNA can be utilized without the need for rRNA depletion or poly-A selection. The prepared fragments are compatible with the xGen Hyb Capture workflow, which targets and enriches the material of interest (**Figure 2**).

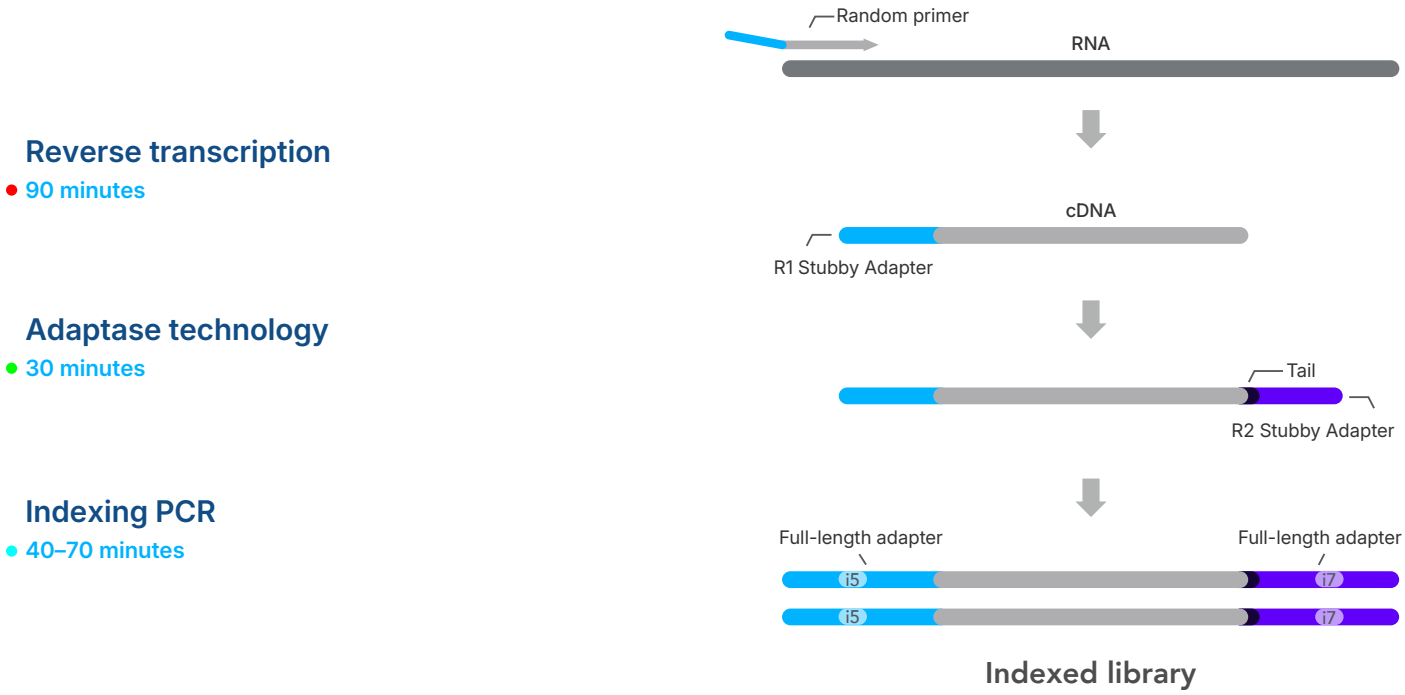

Figure 1. Streamlined xGen RNA Library Prep workflow. RNA-seq libraries are produced following three main steps. First, fragmentation and reverse transcription with a tailed random primer incorporates the i5 adapter. Next, the i7 adapter is added using Adaptase technology. Finally, xGen UDI Primers or xGen Normalase UDI primers are used for indexing.

Add blocking oligos

Hybridize targets to capture probes

Incubate with magnetic streptavidin beads

Isolate targets with magnet

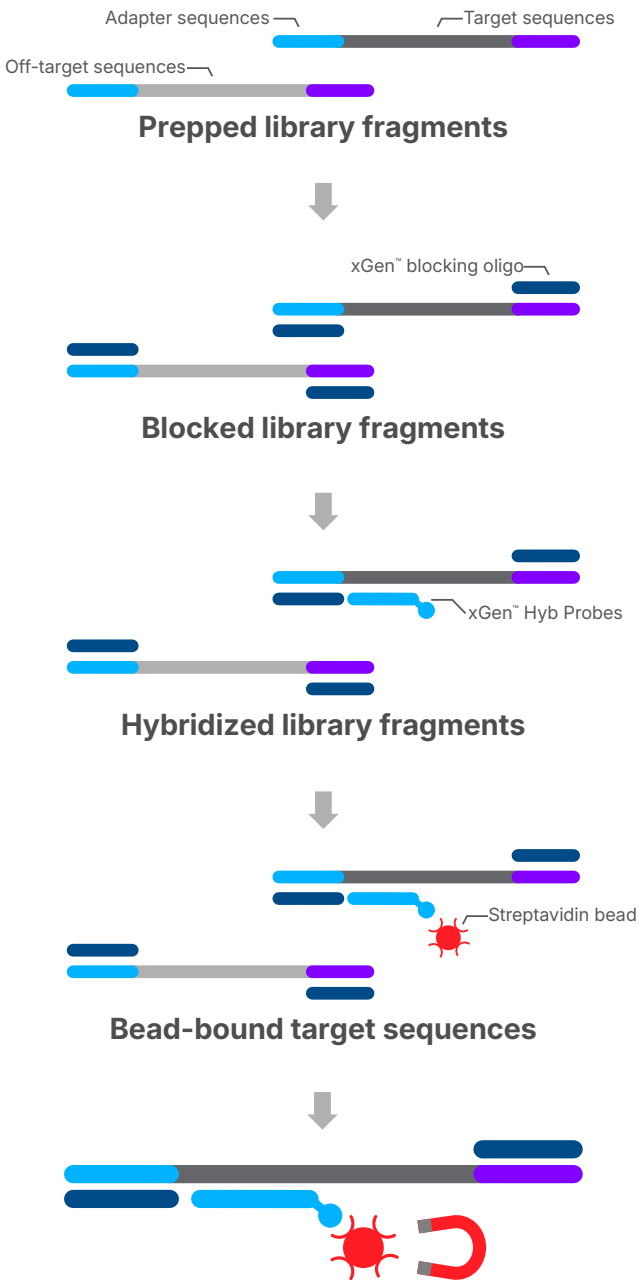

Figure 2. The xGen Hyb Capture workflow. Desired prepared library fragments are separated from off-target fragments using hybridization capture. First, xGen Universal Blockers are mixed with prepared library fragments to prevent adapter-to-adapter hybridization. Blocked library fragments are then annealed to the 5′ biotinylated oligonucleotide probes from an xGen Predesigned Hyb Panel or an xGen Custom Hyb Panel. The probe and fragment duplexes are then separated from the unbound fragments by streptavidin-coated magnetic bead purification. The resulting library is highly enriched for fragments of interest.

The performance of the targeted RNA-seq workflow using the IDT xGen RNA Prep Library Kit and two xGen Custom Hyb Panels was evaluated. Both panels were designed based on the xGen Pan-Cancer Panel which was generated to enrich for significantly mutated genes across multiple tumor types [**4**]. Two different design approaches were evaluated to capture the same target space. One design is RNA specific while the other could be used with RNA or DNA. The data presented here, compares the performance of these two designs, resulting in consistent and retained gene expression data when compared to the transcriptome.

Methods

xGen Custom Hyb Panel design

Two xGen Custom Hyb Panels were designed based on the IDT **xGen Pan-Cancer Panel**, which targets 127 mutated genes implicated across 12 tumor tissue types but customized to be used for RNA-seq applications.

The first panel, the Pan-Cancer Genomic Panel, was designed to target gene coding DNA sequences (CDS) and are extracted using transcript gDNA coordinates. This design, which totaled 8,763 probes, phenocopies the online version of the xGen Pan-Cancer Panel, thus it is compatible for both DNA and RNA applications.

The second, the Pan-Cancer Transcript Panel, was designed to capture transcripts associated with a gene symbol from the University of California Santa Cruz Human Genome Browser database [**5**]. To capture all the desired transcripts with the fewest number of probes, duplicate transcripts (100% sequence identity) were removed. Additionally, an in-house multistrain analysis tool removed functionally redundant probes using sequences with >90% homology. This panel contained a total of 3,952 probes and is only intended to be used with RNA applications.

Targeted RNA-seq library preparation using the xGen RNA Library Prep Kit

To test the xGen Custom Hyb Panels, six RNA libraries were prepared using 100 ng or 500 ng total Universal Human Reference (UHR) RNA (Invitrogen™) using the xGen RNA Library Prep protocol following modifications for xGen Custom Hyb Panel capture (see **xGen RNA Library Prep Appendix B**). Three libraries were multiplexed per hybridization capture following the xGen Hyb Capture Panel protocol using one of the two xGen Custom Hyb Panels.

Sequencing was performed on a NextSeq™ (Illumina) using 2 × 150 paired end (PE) reads, and each library was subsampled to 2 million reads.

Whole-transcriptome RNA-seq library prepration

It is important that target capture does not alter the expression profile when compared to the whole transcriptome. To test this, whole-transcriptome sequencing data was generated using three RNA-seq libraries prepared using 10 ng of poly-A selected UHR (Invitrogen) purified mRNA according to the xGen RNA Library protocol. Poly-A selection was completed using the Poly(A)mRNA Magnetic Isolation Module (New England Biolabs) according to the manufactuer's protocol.

Sequencing was performed on a NextSeq instrument using 2 × 150 PE reads and each library was subsampled to 40 million reads.

Sequence processing and data analyses

Alignment and filtering of sequences was performed using STAR (V2.6.1b) [**6**] with hg38 as the reference genome. General sequencing metrics were generated using the Picard software (v2.18.9) [**7**], and expression data was calculated with featureCounts (v2.0.1) [**8**].

Results

High-quality sequencing from total RNA enriched with xGen Custom Hyb Capture Panels

Target enrichment using either of the xGen Custom Hyb Panel designs resulted in a mapping rate of >78% (**Figure 3A**), a high on-target rate of >98% (**Figure 3B**) and correct strandedness of ≥99.8% (**Figure 3C**). The HS library size is slightly higher for the Pan-Cancer Transcript than for the Pan-Cancer Genomic design. As expected, the higher mass inputs have higher library size than the lower mass inputs within a panel design (**Figure 3D**). The mapping rate achieved from both xGen Custom Hyb Panels are well within the expected range [**3**] and the nearly complete rates of on-target rate and correct strandedness, show that the xGen Custom Hyb Panels provide high-quality sequences at the end of the workflow. Taken together, these results illustrate that the xGen RNA Library Prep used with the xGen Custom Hyb Panel can provide researchers with a time- and cost-saving approach to targeted RNA-seq.

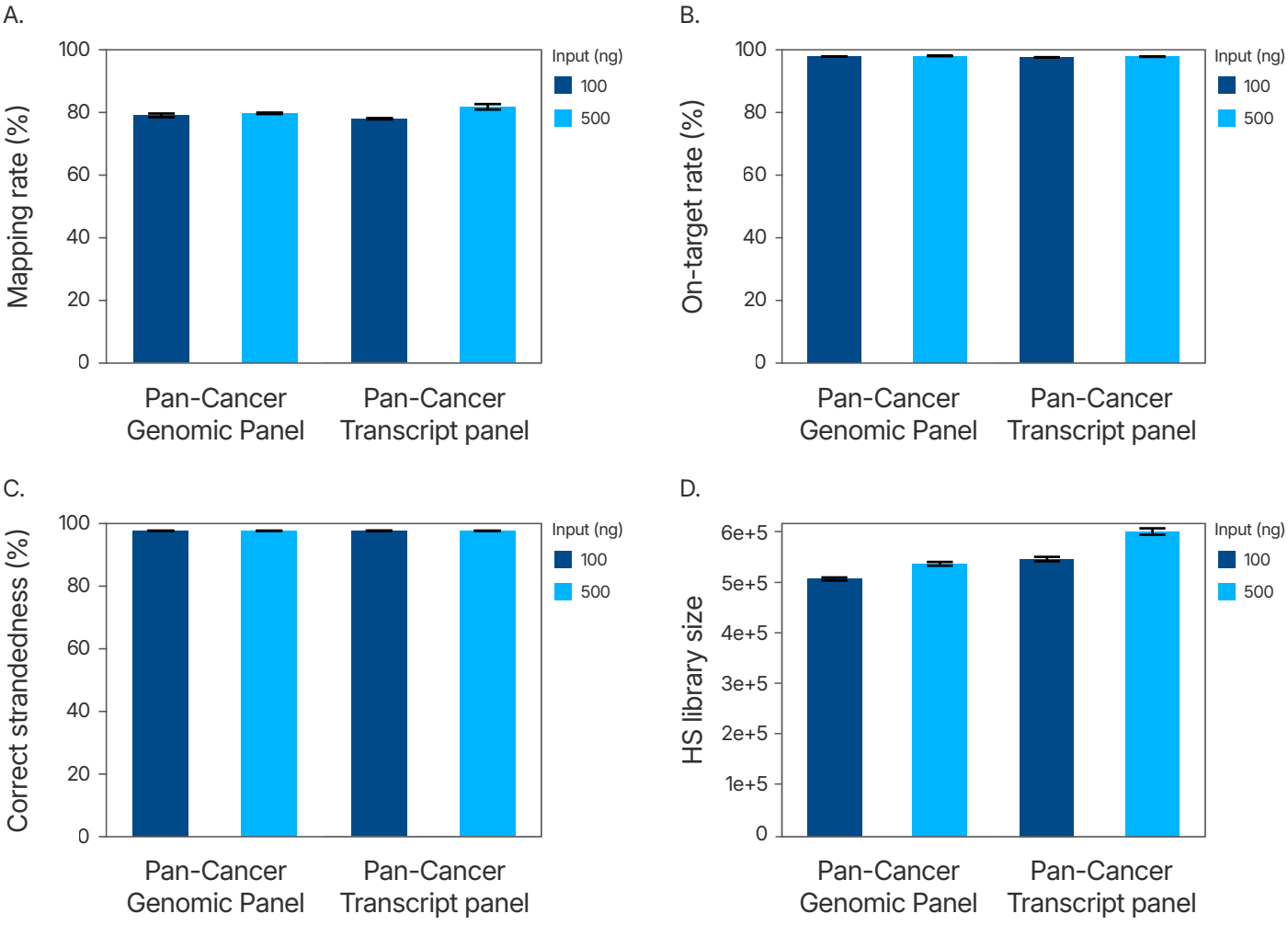

Figure 3. xGen Custom Hyb Panels provide high quality sequencing results according to NGS performance metrics. xGen RNA libraries were generated with 100 ng or 500 ng of total UHR RNA followed by hybridization capture using one of two xGen Custom Hyb Capture Panels (Pan-Cancer Genomic Panel and Pan-Cancer Transcript Panel), and then subsampled to 2 million reads per sample. Sequences were processed and aligned using STAR [**6**] and the hg38 reference genome. This workflow resulted in (**A**) a mapping rate of >78%, (**B**) on-target rate of >98%, and (**C**) and correct strandedness of >99.8%. (**D**) Sequencing shows the number of unique molecules scales as expected based on the RNA input amount into library prep (Picard). Values within the bar chart represent the mean of three libraries multiplexed in one hybridization capture.

High percentage of mRNA captured with minimal rRNA signal

Both the Pan-Cancer Genomic Panel and the Pan-Cancer Transcript Panel, resulted in a limited number of rRNA sequences in the libraries (**Figure 4**). This is an important metric, as this workflow does not use a rRNA depletion or a poly-A selection step. Instead, the xGen Custom Hyb Panel targets and enriches for the regions of interest. The small number of sequences attributed to the rRNA highlights the effectiveness of the xGen Custom Hyb Panels. Especially when considering that rRNA can account for up to 90% of RNA in a sample [**3**]. Regardless of whether the xGen Custom Hyb Panel was designed using the Pan-Cancer Genomic design strategy or Pan-Cancer Transcript strategy, most sequences could be assigned as mRNA (>97%) (**Figure 4**). In comparison, whole transcriptome libraries resulted in ~10% fewer sequences from mRNA and more intergenic and intronic bases (**Figure 4**).

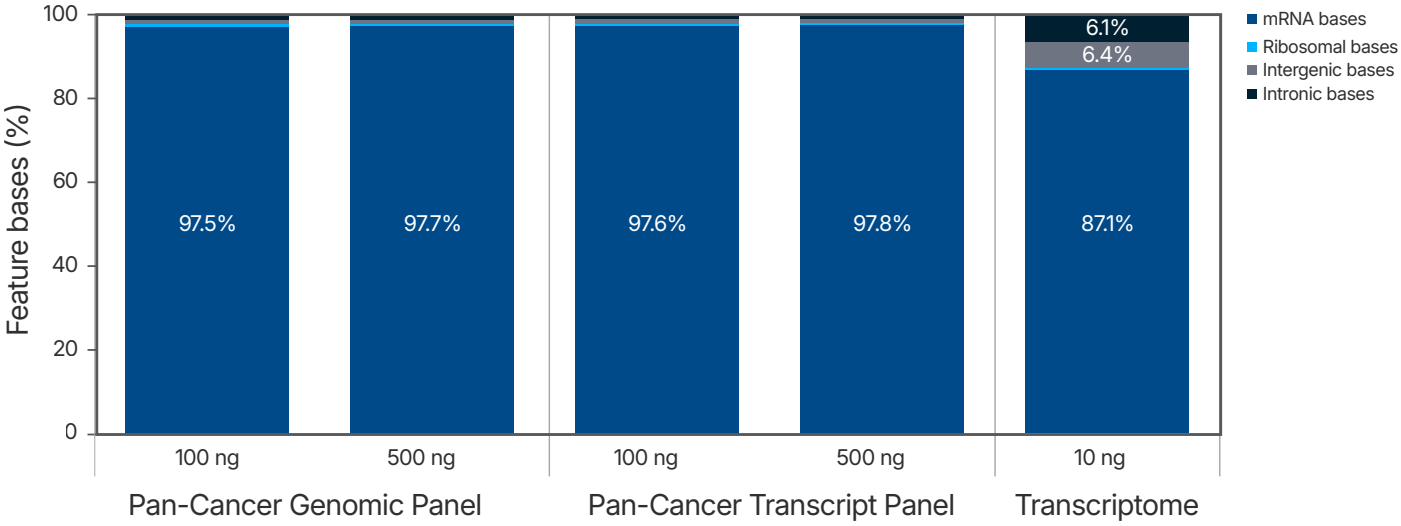

Figure 4. xGen Custom Hyb Panels effectively enrich for mRNA resulting in minimal rRNA transcripts. xGen RNA libraries were generated with 100 ng or 500 ng of total UHR RNA followed by hybridization capture using one of two xGen Custom Hyb Capture Panels (Pan-Cancer Genomic Panel and Pan-Cancer Transcript Panel). Whole-transcriptome libraries were generated with 10 ng of poly-A selected UHR mRNA. All libraries were sequenced on a NextSeq™ platform. The whole transcriptome libraries were subsampled to 40 million reads per sample and the xGen Custom Hyb Panel libraries were subsampled to 2 million reads per sample. Values within the bar chart represent the mean of three libraries multiplexed in one hybridization capture. Sequences were processed and aligned using STAR [**6**] and the hg38 reference genome.

Enrichment with xGen Custom Hyb Panel mirrors expression profile of whole transcriptome sequences

It is important that enriching regions of interest for RNA-seq does not result in an altered expression profile. To evaluate this, the gene expression profiles from the libraries enriched with the xGen Custom Hyb Panels were compared to the whole transcriptome libraries. Both panels display high gene expression correlation with the whole transcriptome library (Pearson correlation, r ≥ 0.98) (**Figure 5**). Thus, the use of the xGen Custom Hyb Panels does not change the expression profile from what is expected.

The Pan-Cancer Genomic Panel design shows a slightly higher correlation to whole transcriptome gene expression information than the Pan-Cancer Transcript Panel design. Additionally, when concerning the scales of the graphs, the enrichment with the xGen Custom Hyb Panel results in dramatically higher transcripts per million (TPM) values with 20X fewer reads when compared to the transcriptome libraries highlighting the cost-saving benefit of the xGen Custom Hyb Panels by focusing the sequencing reads on areas of interest (**Figure 5**).

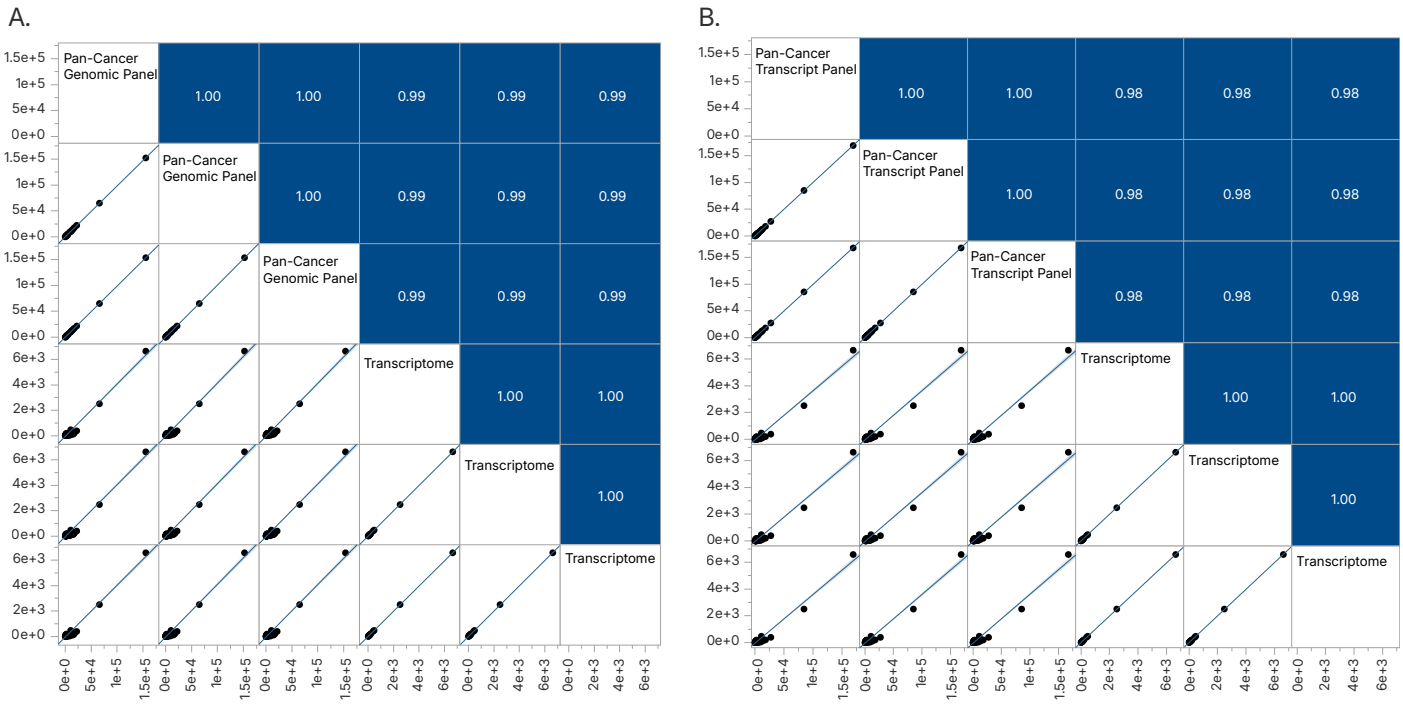

Figure 5. xGen Custom Hyb Panel libraries result in an expression profile that is closely related to that of the whole-transcriptome libraries. xGen RNA libraries were generated with 100 ng or 500 ng of total UHR RNA followed by hybridization capture using one of two xGen Custom Hyb Capture Panels (Pan-Cancer Genomic Panel and Pan-Cancer Transcript Panel). Whole-transcriptome libraries were generated with 10 ng of poly-A selected UHR mRNA. All libraries were sequenced on a NextSeq™ platform. The whole transcriptome libraries were subsampled to 40 million reads per sample and the xGen Custom Hyb Panel libraries were subsampled to 2 million reads per sample. Sequences were processed and aligned using STAR [**6**] and the hg38 reference genome. Expression profiles were calculated using featureCounts [**8**]. Pearson's correlation from pairwise expression level correlation chart normalized to transcripts per million (TPM). (**A**) UHR mRNA uncaptured vs. UHR total RNA captured using the Pan-Cancer Genomic Panel design (N= 3 per library type, r > 0.99). (**B**) UHR mRNA uncaptured vs. UHR total RNA Pan-Cancer Transcript Panel design (n = 6, r > 0.98).

High reproducibility between technical replicates

Finally, the technical replicates between panels were well correlated with both panel design strategies (Pearson correlation, r = 1.00) (**Figure 6**). These results, clearly display the flexibility and reproducibility of the xGen RNA Library Prep Kit and xGen Custom Hyb Panel targeted sequencing workflow.

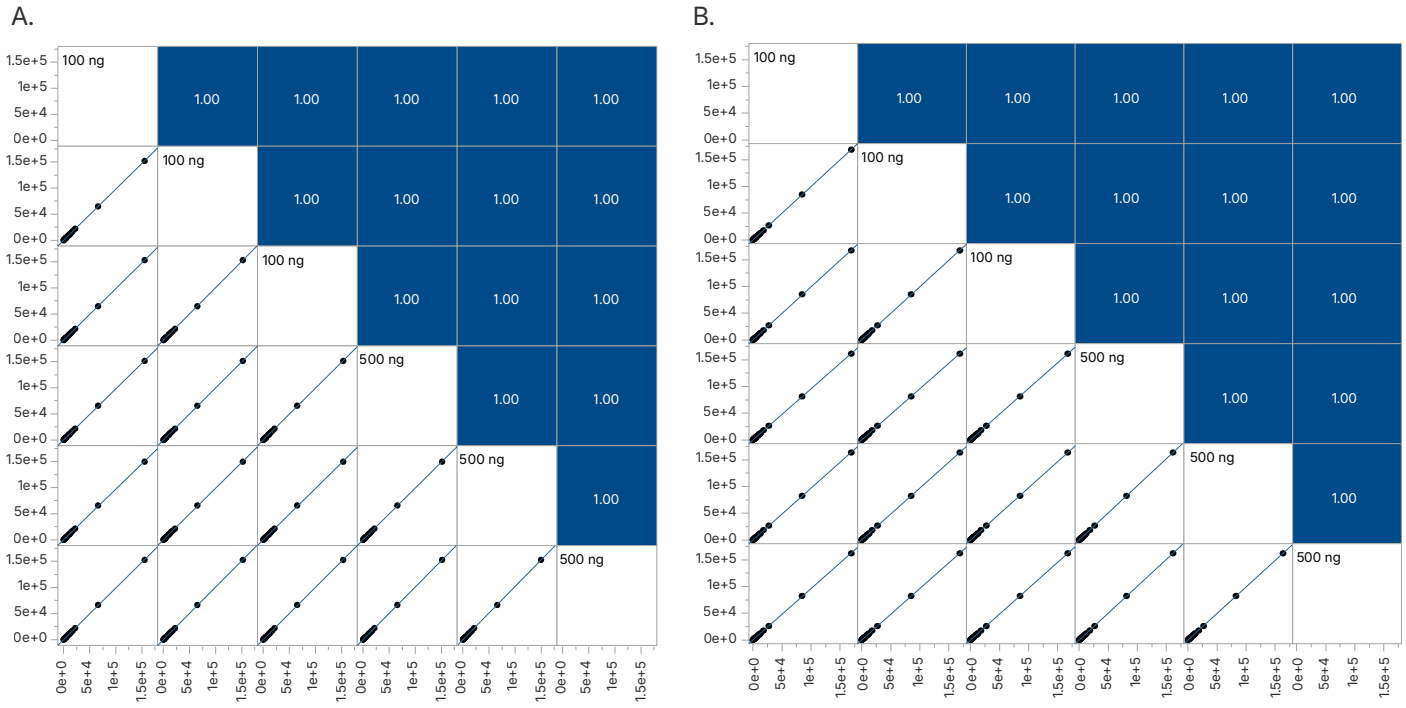

Figure 6. xGen Custom Hyb Panels resulted in minimal variation between technical replicates. xGen RNA libraries were generated with 100 ng or 500 ng of total UHR RNA followed by hybridization capture using one of two xGen Custom Hyb Capture Panels (Pan-Cancer Genomic Panel and Pan-Cancer Transcript Panel). All libraries were sequenced on a NextSeq platform and were subsampled to 2 million reads per sample. Sequences were processed and aligned using STAR [6] and the hg38 reference genome. Expression profiles were calculated using featureCounts [8]. Pearson's correlation from pairwise expression level correlation chart normalized to transcripts per million (TPM). (**A**) The comparison of the Pan-Cancer Genomic Panel between replicates (n = 6, r = 1.00). (**B**) The comparison between the Pan-Cancer Transcript Panel between replicates (n = 6, r = 1.00).

Conclusions

The xGen RNA Library Prep Kit followed by target enrichment with xGen Custom Hyb Panels generates high-quality reproducible libraries that maintain gene expression data while using a streamlined library preparation process and target enrichment with no necessary upstream rRNA depletion or poly-A selection. The low number of rRNA reads, accurate mapping, high on-target and correct strandedness rates indicate these libraries can be used in targeted RNA-seq approaches. More specifically the use of the xGen Custom Hyb Panels to target genes of interest removes intronic, intergenic, and rRNA bases. This saves time and sequencing costs by eliminating the need for rRNA depletion or poly-A selection during library preparation.

Finally, the reproducibility between panels shows that uniquely designed xGen Custom Hyb Panels provide high-quality, dependable results. The Pan-Cancer Genomic Panel which was designed to using gDNA coordinates can be used for RNA-seq applications as well as with DNA, while the Pan-Cancer Transcript design is specific to RNA-seq. Ultimately, the experimental goal will determine which panel design would be best for the application as both designs generate high-quality, consistent data.

To find out more about how the xGen RNA Library Prep Kit or xGen Custom Hyb Panels, visit the IDT webpage: **idtdna.com**.

References

1. Group PTC, Calabrese C, Davidson NR, *et al*. **Genomic basis for RNA alterations in cancer**. *Nature*. Feb 2020;578(7793):129-136. doi:10.1038/s41586-020-1970-0
2. Wang Z, Gerstein M, Snyder M. **RNA-seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet*. Jan 2009;10(1):57-63. doi:10.1038/nrg2484
3. Conesa A, Madrigal P, Tarazona S, *et al*. **A survey of best practices for RNA-seq data analysis**. *Genome Biol*. Jan 26 2016;17:13. doi:10.1186/s13059-016-0881-8
4. Bolger AM, Lohse M, Usadel B. **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics*. Aug 1 2014;30(15):2114-20. doi:10.1093/bioinformatics/btu170
5. Kent WJ SC, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. **The human genome browser at UCSC**. *Genome Res*. 2002;12(6):996-1006.
6. Dobin A, Davis CA, Schlesinger F, *et al*. **STAR: ultrafast universal RNA-seq aligner**. *Bioinformatics*. Jan 1 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
7. Institute B (2019) Picard Toolkit. Broad Institute, GitHub repository **http://broadinstitute.github.io/picard/**.
8. Liao Y, Smyth GK, Shi W. featureCounts: **an efficient general purpose program for assigning sequence reads to genomic features**. *Bioinformatics*. Apr 1 2014;30(7):923-30. doi:10.1093/bioinformatics/btt656

A streamlined workflow for targeted RNA-seq using the xGen RNA Library Prep Kit with xGen Custom Hyb Panels

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

Proud to be part of Danaher

Danaher's science and technology leadership puts IDT's solutions at the forefront of the industry, so they can reach more people. Being a part of Danaher means we can offer unparalleled breadth and depth of expertise and solutions to our customers. Together with Danaher's other businesses in Biotechnology, Diagnostics and Life Sciences, we unlock the transformative potential of cutting-edge science and technology to improve billions of lives every day.

Doc ID: RUO22-0857_001 05/25

IDT™ INTEGRATED DNA TECHNOLOGIES