A hybridization capture-based approach for RNA sequencing and gene fusion discovery using the xGen™ Exome Hyb Panel v2

Abstract

One of the challenges with whole transcriptome sequencing (WTS) is the high cost associated with the coverage needed to study low-expressed transcripts or rare structural variants like gene fusions. A targeted sequencing approach focuses on specified regions of interest within the transcriptome. This approach provides deeper sequencing of targeted regions while omitting the undesired regions that often result in a disproportionate number of sequencing reads not relevant to the research study. The xGen™ Broad-Range Library Prep Kit paired with the xGen Exome Hyb Panel v2 is cost-effective and reliable for performing targeted sequencing, which yields a higher percentage of coding bases and a lower percentage of intronic bases than whole transcriptome data—leading to higher expression profiling efficiency. The exome-enriched RNA libraries confirmed gene fusions with higher read-depth than transcriptome libraries when sampled to the same number of reads.

Introduction

Gene fusions are important biomarkers in oncology research [1]. Using targeted RNA-seq is a cost-effective tool for deeper sequencing of targeted regions that can identify rare fusion events and discover gene fusion transcripts. The xGen Broad-Range Library Prep Kit supports a wide range of inputs, including low quality RNA from Formalin Fixed Parafilm Embedded (FFPE) samples. The xGen Exome Hyb Panel v2 can be used in a targeted RNA-seq workflow for the discovery of gene fusions in FFPE, despite the exome capture panel design assuming no prior knowledge of breakpoints, while still allowing for quantitative assessment of gene expression from degraded FFPE samples.

Methods

As shown in the workflow in Figure 1, total RNA was extracted from Seraseq® FFPE Tumor Fusion RNA v4 Reference Material (RM) using the RNeasy® FFPE RNA extraction kit (QIAGEN). A RIN of 1.5 and DV_{200} of 75 were determined using the TapeStation™ 2100 (Agilent). A subset of samples were rRNA-depleted using the NEBNext® rRNA Depletion kit v2. xGen Broad-Range RNA libraries were prepared using 10 ng of rRNA-depleted RNA (*n* = 3) or 50 ng of total RNA (*n* = 7). Total RNA libraries underwent hybridization capture using the xGen Exome Hyb Panel v2 individually (*n* = 3) or as a single 4-plex (*n* = 1) for 16 h following the vendor's protocol. Transcriptome and exome-enriched libraries were sequenced on a NextSeq® (IIIumina) using 2 x 150 paired end (PE) reads and subsampled to 40 million reads/sample. All libraries were analyzed using Picard, Kallisto, STARFusion, and CTAT-splicing. Transcriptome analysis was performed using Genecodev37 as the target space. Exome-enriched libraries were analyzed using its' associated target BED file (hg38).

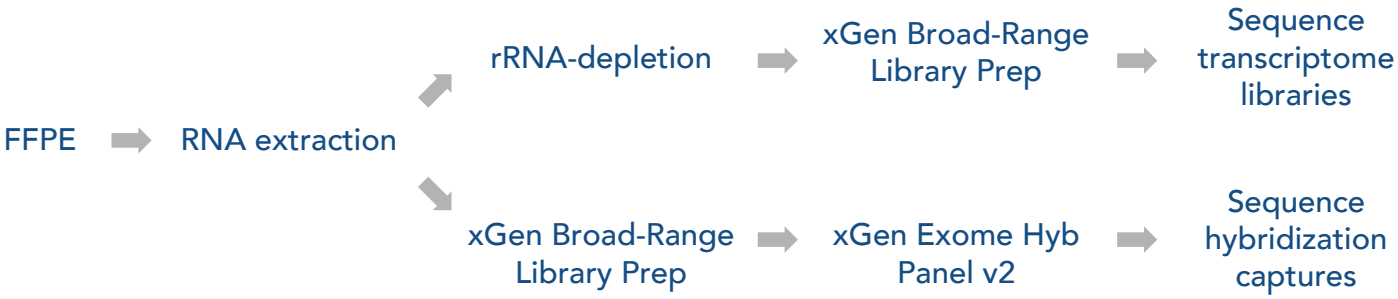

Figure 1. Schematic of methods workflow. Extracted RNA from FFPE samples followed one of two pathways to create either rRNA-depleted transcriptome libraries or hybridization capture libraries using total RNA.

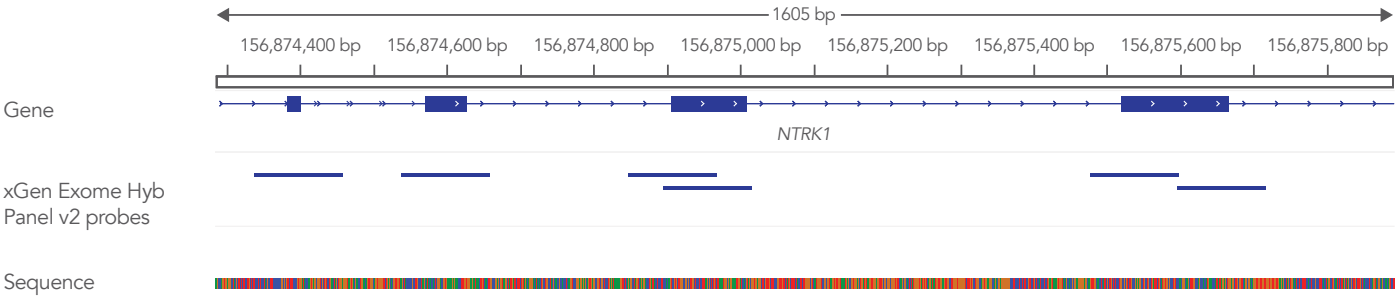

Figure 2. xGen Exome Hyb Panel V2 in Integrative Genomics Viewer (IGV) (v2.8.9). Image shows an example how the pre-designed capture panel probes cover regions of *NTRK1* which are common fusions breakpoints.

Results

The xGen Exome Hyb Panel v2 was designed to capture protein-coding regions of the human genome and targets 19,433 genes using 415,115 probes spanning 34 Mb target space of the human genome targeting the coding sequences (CDS) in the RefSeq 109 database. This design strategy allows for the panel to be used with DNA, RNA, and for gene fusion discovery (an example shown in Figure 2 for the *NTRK* locus). Both transcriptome libraries and the single-plex and 4-plex RNA captured mRNA libraries show high mapping rates and comparable duplication rates (Figure 3a). High intronic rates in FFPE samples are a well-characterized phenomenon attributed to the fixation process [2]. The captured mRNA libraries show higher exonic sequence rates when compared to the whole transcriptome libraries, which show higher intronic rates (Figure 3b). By using an exome panel to enrich total RNA libraries, the number of intronic base reads decreased from ~55% to ~4%, and the number coding bases sequenced increased from ~18% to ~87% irrespective of whether a single-plex or 4-plex hybridization was performed (Figure 3b). Additionally, rRNA bases are <1% for all captured mRNA libraries and transcriptome libraries (Figure 3b), indicating that rRNA depletion is not necessary prior to hybridization capture when enriching with the xGen Exome Hyb Panel v2.

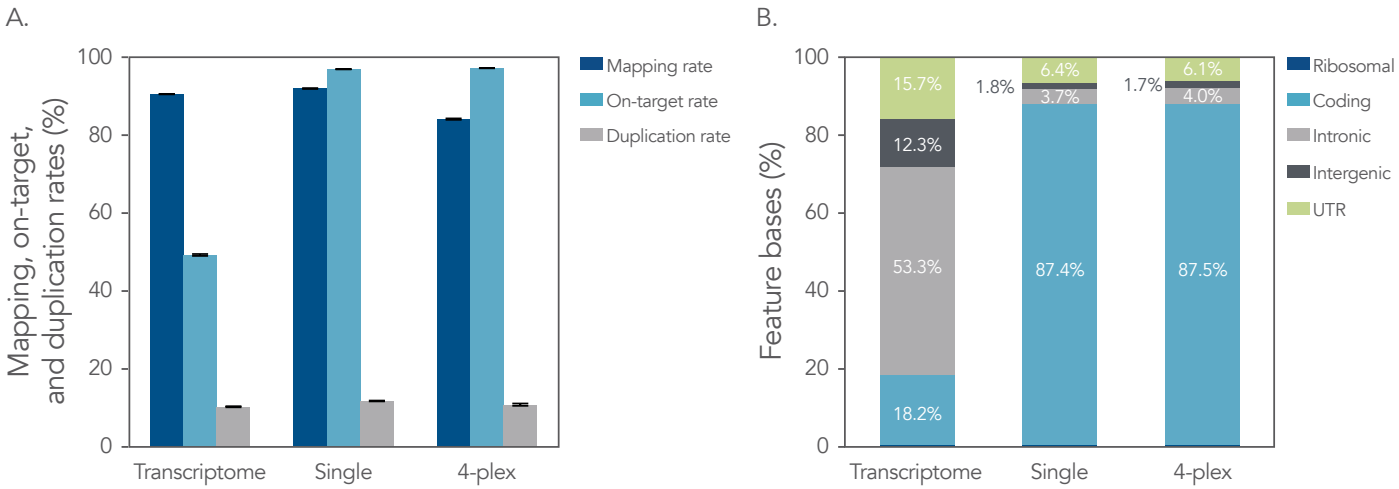

Figure 3. Target enrichment metrics. Extracted FFPE RNA was used to prepare xGen Broad-Range RNA libraries from 10 ng of rRNA-depleted RNA (*n* = 3) or 50 ng of total RNA (*n* = 7). Hybridization capture using the xGen Exome Hyb Panel v2 was performed on total RNA libraries as single-plex (*n* = 3) or in a 4-plex hybridization (*n* = 1). All libraries were sequenced on a NextSeq® (Illumina) 300 cycle kit and subsampled to 40 million reads per sample. (**A**) The data shows comparable mapping rates and duplication rates, and higher exonic reads and lower intronic reads for hybridization captures samples. (**B**) The feature bases chart shows that hybridization capture results in a higher percent of coding bases sequenced, and fewer intronic bases than transcriptome libraries (values within bars represent means of the replicates).

Differential expression analysis allows for the discovery of which genes are expressed at different levels between samples and experimental conditions and is important in understanding biological functions, such as cell differentiation and disease states. It is important that hybridization captured RNA-seq libraries show similar expression levels as transcriptome libraries, as this indicates that expression information is retained. Normalized expression correlations using Transcripts per Million (TPM) were compared between captured mRNA libraries and whole transcriptome libraries. Transcriptome libraries compared to individual captured mRNA libraries show a strong expression correlation (r = 0.92) and the replicates show a Pearson's correlation of 1.00 (Figure 4a). Captured mRNA libraries from the 4-plex hybridization compared to the transcriptome libraries show a slight decrease in expression correlations (r ≥ 0.87) but maintain a Pearson's correlation of 1.00 between replicates (Figure 4b). Depending on the sample type and other conditions, the appropriate plexity level for RNA-seq may differ.

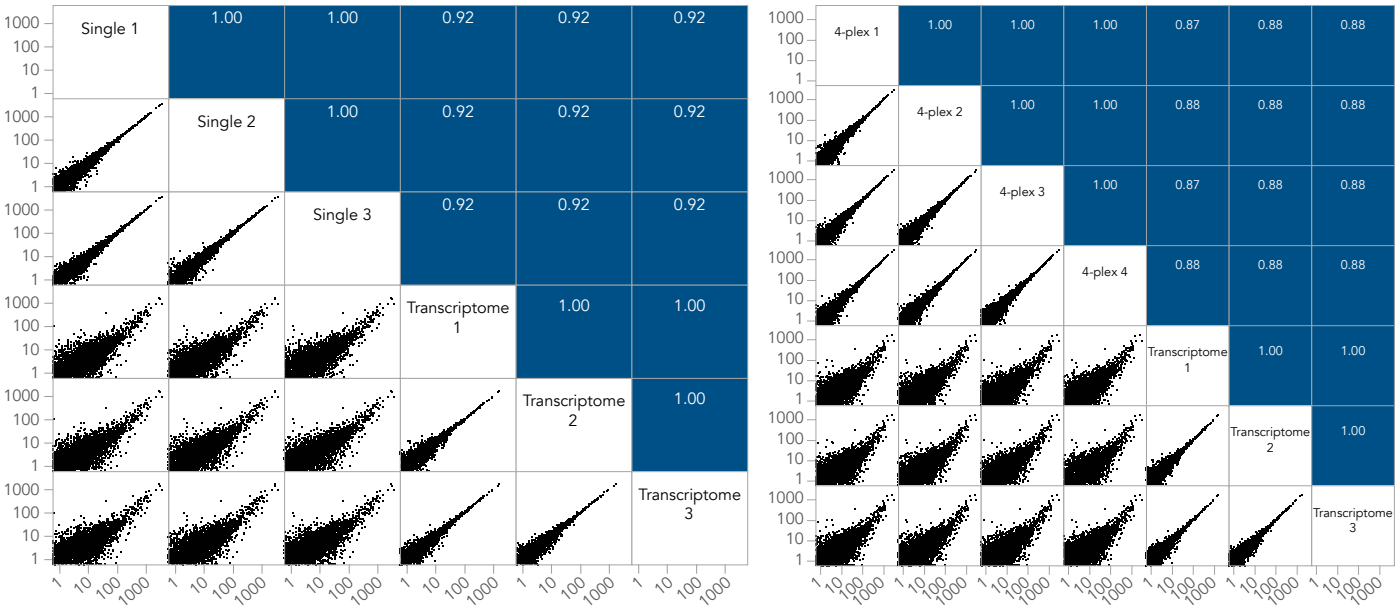

Figure 4. Normalized expression correlations using TPM. FFPE RNA was used to prepare xGen Broad-Range RNA libraries from 10 ng of rRNA-depleted RNA (*n* = 3) or 50 ng of total RNA (*n* = 7). Hybridization capture using the xGen Exome Hyb Panel v2 was performed on total RNA libraries as single-plex (*n* = 3) or in a 4-plex hybridization (*n* = 1). Libraries were sequenced on a NextSeq® (Illumina) 300 cycle kit and subsampled to 40 million reads per sample. (**A**) Data comparing transcriptome and single-plex captured mRNA libraries show is strongly correlated across replicates (r =1.00) and an expression correlation of ≥ 0.92 and (**B**) transcriptome and 4-plex captured mRNA libraries show high levels of consistency across replicates (r =1.00) and an expression correlation of ≥ 0.87.

SeraSeq FFPE Tumor Fusion Reference Material contains 18 ddPCR- confirmed gene fusions which was used to determine if the xGen Exome Hyb Panel v2 could identify fusions without fusion-specific probe designs and without any prior knowledge of fusion breakpoints in the design. A junction read is defined as a single read that overlaps the fusion breakpoint, and a spanning read is a paired end read which maps to both sides of the fusion breakpoint, therefore mapping to both fusion gene partners. A fusion is counted as being identified if it is present in ≥3 total reads. For all fusions, the captured mRNA libraries (both single-plex and multiplexed libraries) showed higher numeric read values when compared to the transcriptome data when subsampled to the same number of reads (Figure 5a). All 18 fusions were identified in each of three single-plex and all 4-plex captured mRNA libraries, while most but not all fusions were found in the transcriptome libraries, where 3 fusions were missed altogether (Figure 5b). This indicates that the xGen Exome Hyb Panel v2 can capture gene fusion events and because of the targeted approach, the hybridization captured libraries can identify more fusions with higher read-depth than transcriptome sequencing. This results in cost-efficiency due to the need for fewer sequencing reads and leaves the potential for high-throughput processing of samples.

Figure 5. Identification of fusion events. FFPE RNA was used to prepare xGen Broad-Range RNA libraries from 10 ng of rRNA-depleted RNA (n = 3) or 50 ng of total RNA (n = 7). Hybridization capture using the xGen Exome Hyb Panel v2 was performed on total RNA libraries as single-plex (n = 3) or in a 4-plex hybridization (n = 1). All libraries were sequenced on a NextSeq® (Illumina) 300 cycle kit and subsampled to 40 million reads per sample. Each bar represents a single library, and a fusion was counted as being identified if it is present in ≥3 total reads (red line). (**A**) For all 18 fusions, the hybridization captured libraries had more junction and spanning reads then the transcriptome sample. (**B**) When looking at the proportion of libraries that identified the expected fusion, each of the libraries in the single-plex and 4-plex correctly identified all 18 fusions, while the transcriptome libraries did not identify all fusions in all libraries.

Conclusions

xGen Broad-Range RNA libraries prepared from a degraded FFPE sample were captured as a single-plex or in a 4-plex hybridization capture with the pre-designed xGen Exome Hyb Panel v2. An efficient library prep and hybridization capture workflow demonstrates high mapping, on target percentages, and low duplication rates. Captured mRNA libraries result in an increased number of exonic sequencing reads due to the decreased amount of intergenic and intronic reads when compared to the transcriptome libraries. This translates to sequencing cost efficiency due to eliminating unusable reads. All libraries had a removal of >99% of rRNA bases, indicating that ribodepletion is unnecessary for hybridization captured libraries. The comparison of TPM of transcriptome and hybridization captured libraries indicate that RNA expression information is retained and reproducible. Furthermore, the xGen Exome Hyb Panel v2 can identify RNA fusions and more gene fusions reads than transcriptome libraries when using the same number of reads. The captured mRNA libraries identified all expected gene fusions despite the panel being designed with no consideration for gene fusion identification, indicating the ability to use the pre-designed xGen Exome Hyb Panel v2 for gene fusion identification without the need for a spike-in panel or additional content specific to gene fusion identification.

References

1. Mitelman F, Johansson B, Mertens F. **The impact of translocations and gene fusions on cancer causation.** Nat Rev Cancer. 2007;7(4):233-245. doi:10.1038/nrc2091

2. Wehmas LC, Hester SD, Wood CE. **Direct formalin fixation induces widespread transcriptomic effects in archival tissue samples.** Sci Rep. 2020;10(1):14497. Published 2020 Sep 2. doi:10.1038/s41598-020-71521-w

A hybridization capture-based approach for RNA sequencing and gene fusion discovery using the xGen™ Exome Hyb Panel v2

Technical support: applicationsupport@idtdna.com