A hybridization capture-based approach for gene fusion discovery using the xGen™ Exome Hybridization Panel

Katelyn Larkin*, Jessica Sheu, Kevin Lai, Ashley Dvorak

Integrated DNA Technologies

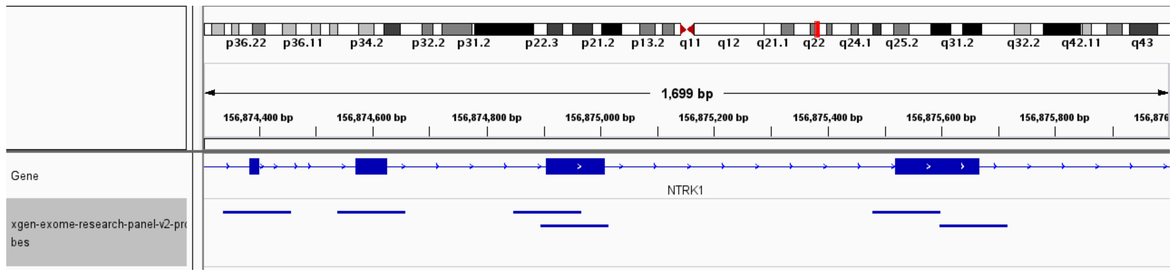
* Corresponding author: klarkin@idtdna.com

Introduction

RNA-seq allows researchers to quantify gene expression and identify sequence alterations such as gene fusions. One of the challenges with whole transcriptome sequencing (WTS) is the sequencing cost associated with the high coverage needs of low-expressed transcripts or rare structural variations. By using a targeted sequencing approach, sequencing reads can be focused on specified regions of interest in the transcriptome, allowing for deeper sequencing of these areas while omitting undesired regions that often result in a disproportionate number of sequencing reads such as ribosomal RNA (rRNA). Gene fusions are important biomarkers for researchers as they are known to be cancer drivers. The capability for deeper sequencing with lower cost enables researchers to use targeted RNA-seq as a tool for the identification of rare fusion events and discovery of gene fusion transcripts that may otherwise be too costly and difficult to identify. Here, we show that the xGen Exome Hyb Panel v2 can be used in a targeted RNA-seq workflow for the discovery of gene fusions (in FFPE reference material), despite the capture panel design assuming no prior knowledge of breakpoints, while still allowing for quantitative assessment of gene expression from degraded FFPE samples.

Methods

Total RNA was extracted from Seraseq® FFPE Tumor Fusion RNA v4 Reference Material (RM) using the Qiagen RNeasy® FFPE RNA Extraction Kit. A RIN of 1.5 and DV_{200} of 75 were determined using Agilent TapeStation 2100. A subset of samples were rRNA-depleted using the NEBNext® rRNA Depletion Kit v2. Library preparation was performed on the rRNA-depleted RNA (10 ng input) in triplicate and total RNA (50 ng input) using the IDT xGen Broad-Range RNA Library Prep Kit. Hybridization capture was performed on total RNA libraries individually (n=3) or as a single 4-plex with the xGen Exome Hyb Panel v2 for 16 hrs, following the supplier's protocol. Transcriptome and enriched libraries were sequenced on an Illumina® NextSeq™ using 2 x 150 paired-end reads and subsampled to 40 million reads/sample. All libraries were analyzed using Picard, Kallisto, STARFusion, and CTAT-splicing. Transcriptome analysis used Genecodev37 as the target space. The exome panel covers 19,433 genes and libraries were analyzed against its associated target BED file.



The xGen Exome Hyb Panel v2 consists of 415,115 probes spanning 34 Mb target space of the human genome, only targeting the coding sequences (CDS) in the RefSeq 109 database. Above, an Integrative Genomics Viewer (v2.8.9) shows an example how the predesigned capture panel covers regions of NTRK1 which are common fusions breakpoints. This design strategy allows for the panel to be used with DNA, RNA, and for gene fusion discovery.

Quality sequencing metrics for degraded FFPE samples

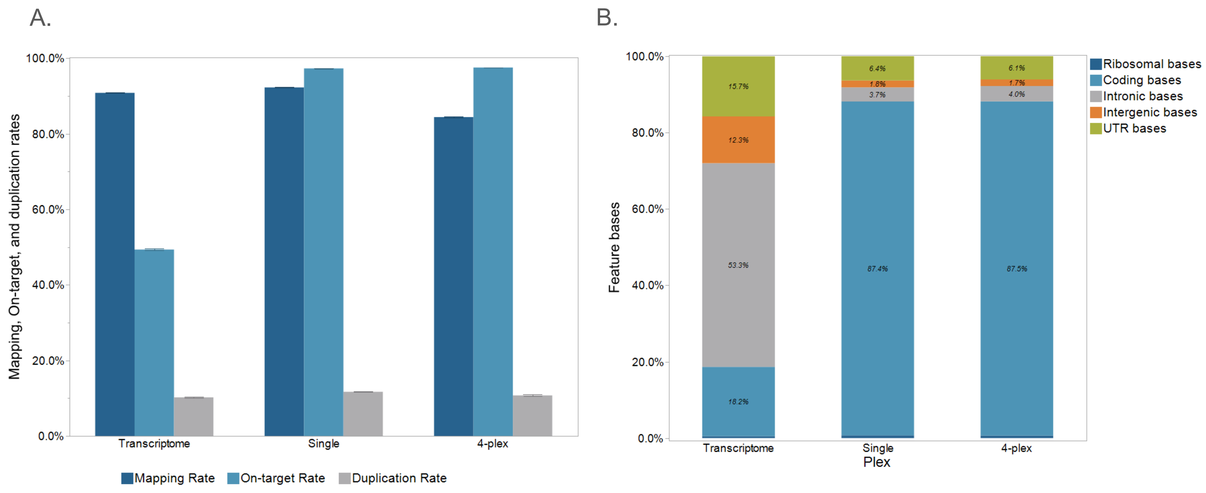


Figure 1. Target enrichment metrics and RNA types. (**A**) Data from both transcriptome, single, and 4-plex exome-enriched libraries show high mapping rates and comparable duplication rates. Exome captures have higher on-target rate when compared to the transcriptome. (**B**) Values represent means of the replicates. By using the xGen Exome Hyb Cap Panel to enrich total RNA libraries, the number of coding bases sequenced increased from ~18% to ~87%, while the number of intronic-base reads decreased from ~55% to ~4% regardless of plexity. rRNA bases are present <1% for all exome-enriched and transcriptome libraries.

Expression correlations between exome captures and transcriptome

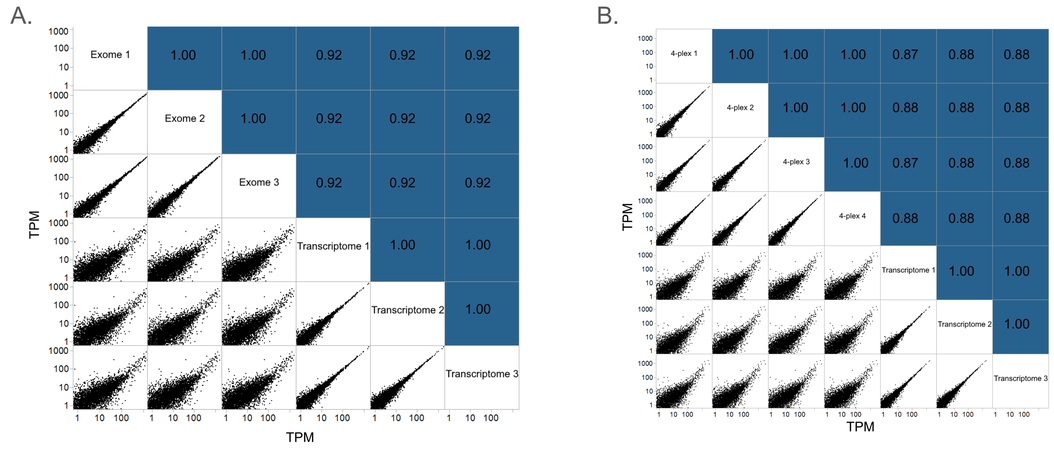


Figure 2. Normalized expression correlations using transcripts per million (TPM). (**A**) WTS vs. single-plex. Replicates show a Pearson's correlation of 1.00 and a strong correlation (r = 0.92) between transcriptome and single-plex xGen Exome Hyb Panel expression data. (**B**) WTS vs. 4-plex capture. xGen Exome Hyb Panel used in 4-plex shows a correlation between replicates of 1.00 and a correlation to the transcriptome of ≥ 0.87.

Fusion discovery using xGen Exome Hybridization panel

Figure 3. SeraSeq FFPE Tumor Fusion RM contains 18 ddPCR-confirmed fusions. (**A**) Junction and spanning read counts for 18 fusions. A junction read is a single read that overlaps the fusion break-point, a spanning read is a paired-end read which maps to both sides of the fusion break-point. A fusion is counted as being identified if it is present in ≥3 total reads (red line). Each bar represents a single library. For all fusions, the hyb capture data (both single-plex and multiplexed libraries) shows higher numeric read values when compared to WTS data. (**B**) Proportion of libraries identifying fusions. All 18 fusions were identified in each of three single-plex and all 4-plex exome-enriched libraries. All fusions were not found in all transcriptome libraries with 3 fusions being missed altogether.

Conclusions

The xGen Broad-Range RNA Library Prep Kit was used to create libraries from a degraded FFPE sample and captured individually or in a 4-plex with the predesigned xGen Exome Hyb Panel v2 and resulted in:

- High mapping, on-target percentages, and low duplication rates as demonstrated by the efficient library prep and hybridization capture workflows.
- Increased number of on-target sequencing reads due to decreased amount of intergenic, intronic, and UTR bases compared to WTS, translating to sequencing cost savings by eliminating wasted reads.
- Removal of >99% of rRNA bases, similar to the rRNA-depleted libraries.
- Retained RNA expression information when compared to WTS libraries.
- Identification of more RNA gene fusions reads than WTS when using the same read sub-sample level.
- Identification of **all** expected gene fusions from FFPE Tumor Fusion RM, despite the panel being designed with no consideration for gene fusion break-points.
- Capability of the predesigned xGen Exome Hyb Panel v2 for gene fusion discovery without the need for any spike-in panel or additional content specific to gene fusions.