Sensitive capture of full-length transcript information with targeted RNA-seq

Rachel N. Fish[1], Sally Zhang[1], Yevgeniy Gindin[2], Sarah Taylor[1], and Andrew Farmer[1]

1 Takara Bio USA, Inc., Mountain View, California, USA
2 Gilead Sciences, Inc., Foster City, California, USA

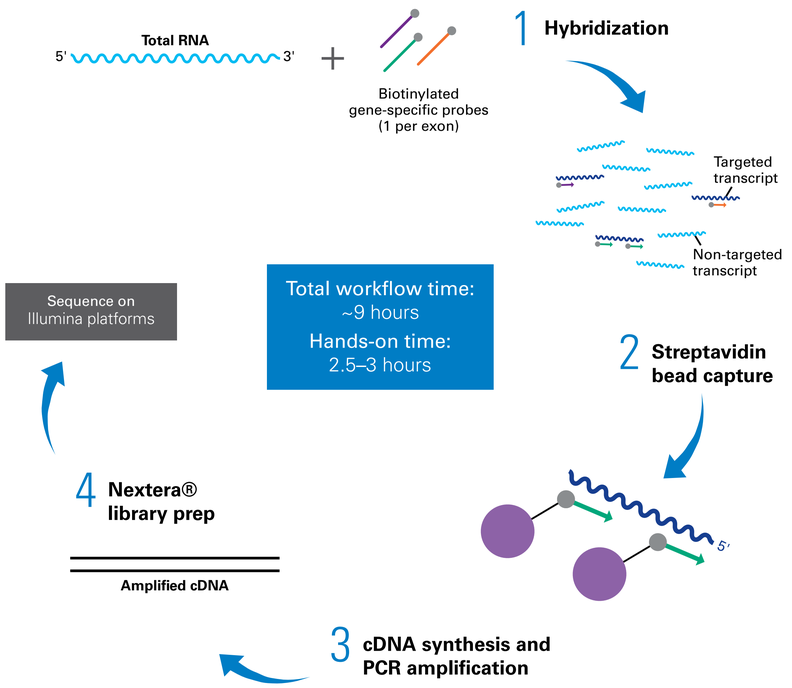
Abstract

While recent advances in technology have greatly improved research in whole transcriptome RNA-seq, several challenges still remain, stemming from the complexity inherent in such large-scale sequencing. The large dynamic range of the transcriptome often means that a few highly abundant transcripts account for the majority of sequencing reads, while less-abundant transcripts (representing a majority of RNAs) account for only a small percentage of sequencing reads. Targeted RNA-seq aims to overcome this problem by improving sequence coverage of transcripts of interest that may be present in low amounts, thus saving costs and simplifying analysis. Sensitive targeted enrichment of RNA enables the capture of information about transcripts that would otherwise be missed or would require a much greater number of sequencing reads to be detected, including chimeric gene fusions, transcript isoforms, and splice variants.

We combined a streamlined method for direct capture of full-length transcripts from total RNA with the cDNA synthesis technology of the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing to develop a target-specific protocol with high sensitivity and low background. The resulting RNA capture method demonstrates consistent enrichment and coverage from 1 ng to 1 µg of a variety of total RNA inputs. By capturing full-length transcripts, the protocol enables the detection of structural variation in expressed RNAs. For example, we identified a gene fusion event present at a frequency lower than 0.5% in a sequencing library of fewer than 2 million reads while targeting only one partner of the gene fusion.

Additionally, we were able to maintain relative expression levels for targeted genes post-enrichment, providing confidence in differential expression analysis of transcripts of interest. Furthermore, the substantial increase in coverage of only a subset of genes of interest lowers the required sequencing depth, thereby reducing analysis time and experimental costs.

1 SMARTer targeted RNA-capture workflow



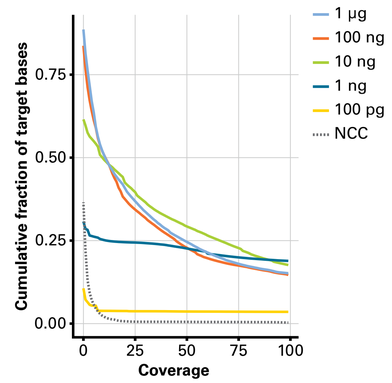
Protocol overview. The SMARTer® Target RNA Capture for Illumina® protocol (through library quantification and quality determination) is completed over two days, with just 2.5–3 hours of hands-on time and without the need for additional rRNA removal methods or kits. Recommended input levels range from 10 ng–1 µg of high-quality total RNA. **Step 1:** Target-specific, 60-mer biotinylated probes are hybridized to transcripts of interest. Probes are designed at a frequency of one per exon, and hybridization takes only 2.5 hours. **Step 2:** The resulting RNA-DNA hybrids are captured with streptavidin-coated Capture Beads, which have been specifically selected for their low adsorption of protein and nucleic acids, and lack of interference with downstream reactions. **Step 3:** SMART-Seq v4 technology is used for highly sensitive first-strand cDNA synthesis, performed while the RNA-DNA complexes are immobilized on the Capture Beads. The reverse transcriptase efficiently displaces the biotinylated probes, making the first-strand cDNA available for PCR amplification in solution. **Step 4:** Following amplification, library quantification, and quality determination of the PCR product, Illumina sequencing adapters are added. The final library is then ready for sequencing on Illumina platforms.

2 Consistent enrichment and coverage across input levels

A

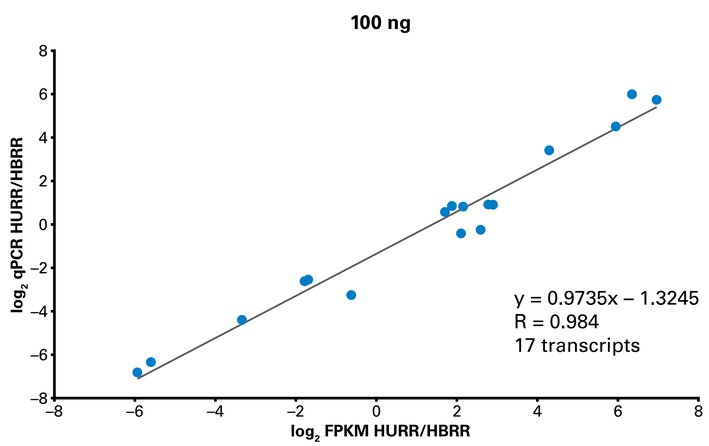
Input	On-target reads	Fold change	Target genes identified
HBRR			
100 pg	17%	47	32
1 ng	31%	91	33
10 ng	28%	89	51
100 ng	22%	67	58
1 µg	24%	78	58
HURR			
100 pg	43%	91	29
1 ng	46%	107	46
10 ng	41%	93	55
100 ng	51%	120	60
500 ng	1%	33	59

B



Capture across input levels for HBRR and HURR total RNA with ~1 M reads per sample. HBRR (Thermo Fisher Scientific, Cat. # AM6050) or HURR (Agilent, Cat. # 740000) total RNA was used as input into the SMARTer Target RNA Capture for Illumina protocol, with amounts ranging from 100 pg to 1 µg, where 10 ng is the recommended lower input. 1 pmol of probes against a 60-gene panel were used (1,200 exons, ~0.46 Mb). Sequencing libraries were run on an Illumina MiSeq® (75 x 75 paired-end reads), and sequencing reads were aligned with STAR to the GRCh38 release 23 with Gencode annotations. The number of read counts and FPKMs were determined for each targeted gene after downsampling to the same number of reads (~1 million). **Panel A.** The percentage of on-target reads was calculated as the total number of sequencing reads against the targeted genes, divided by the number of uniquely mapped reads as calculated by STAR. Fold enrichment was calculated as the total number of sequencing reads against the target obtained after capture, divided by the number of sequencing reads obtained without capture. The 500-ng input for HURR displayed difficulties due to contaminants present in the RNA buffer that inhibited the hybridization reaction. **Panel B.** On-target cumulative coverage plot for a titration of HBRR samples with ~1 M reads per sample. The lines show the cumulative fraction of the target bases along the y-axis with a read coverage of at least x. With lower input, the probability of capturing any specific gene decreases.

3 Enriched targets maintain relative expression levels



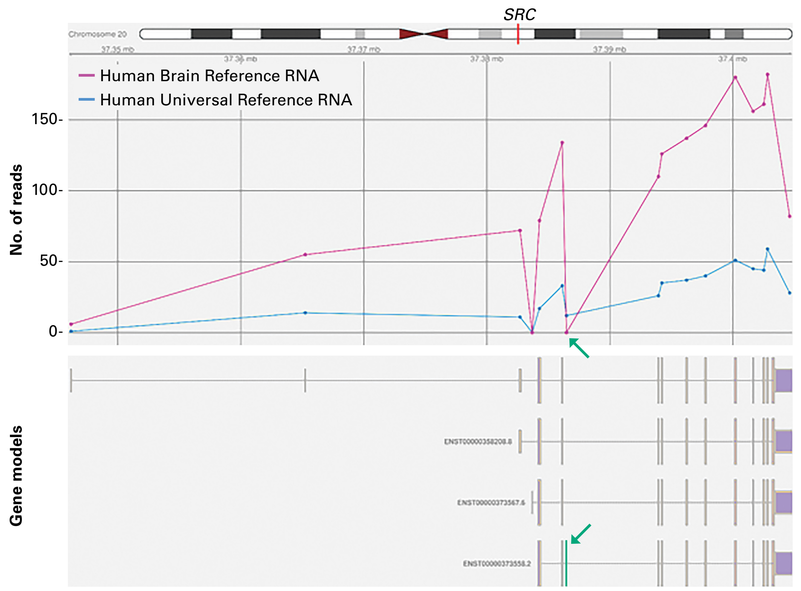
MAQC analysis of targeted capture data. The Microarray Quality Control (MAQC) studies used TaqMan qPCR data to evaluate differential expression of ~1,000 genes between HURR and HBRR on microarray platforms. To evaluate this kit, the NGS expression data (expressed as FPKM) from the experiment described in Figure 2 was compared against the historical TaqMan data. 17 of the 60 genes in the panel also appear in the MAQC data. The y-axis shows the $\log_2$ of the ratio of the TaqMan qPCR Ct values, and the x-axis shows the $\log_2$ of the ratio of FPKM values obtained with SMARTer Target RNA Capture for Illumina. Similar relationships for differential expression ratios were seen with 10 ng or higher (up to 1 µg) of input RNA.

4 Identification of rare gene fusions at low sequencing depth

	10% K562 (2.3 M reads)	1% K562 (2.4 M reads)	0.1% K562 (2.4 M reads)	No-Capture K562 (3.3 M reads)
Fold enrichment of ABL1	197	156	150	1
Reads supporting BCR-ABL1 fusion	628	48	0	14

Identification of a gene fusion event present at a frequency <0.5% while targeting only one partner of the fusion. K562 is a human chronic myelogenous leukemia cell line and harbors a BCR-ABL1 fusion in about 30% of ABL1 transcripts (data not shown). A sequencing library was made from Human Brain Total RNA (Cat. # 636530) spiked with K562 RNA at 10%, 1%, and 0.1%, for 10 ng of combined RNA (1 ng, 100 pg, and 10 pg of K562 RNA, respectively). 1 pmol of probes against only ABL1 were used. Sequencing libraries were run on an Illumina MiSeq (75 x 75 paired-end reads), and sequencing reads were aligned to the human reference with the TopHat aligner and then downsampled to the same number of reads (~2 M). The number of reads to the fusion was determined with a custom script that directly queried reads for evidence of a translocation.

5 Identification of alternative splicing events at low sequencing depth



Full-length cDNA synthesis enables identification of transcript isoforms. HBRR and HURR total RNAs were used as inputs into the SMARTer Target RNA Capture for Illumina, as described in Figure 2. Gene models showing transcript isoforms of the *SRC* gene are shown in the bottom half of the image. The top half shows the number of reads for each exon plotted against exons shown in the gene models. The pink line shows reads for the HBRR sample. Note that *SRC* seems to be more highly expressed in this RNA sample. The blue line shows reads for HURR. For most exons, HBRR and HURR show similar profiles, with more reads from HBRR at each exon. However, at the indicated exon (green arrow), HURR shows reads while HBRR does not, indicating different splice variants present in each sample.

Conclusions

- High sensitivity enables the detection of rare structural events (e.g., chimeric gene fusions, transcript isoforms, and splice variants) at low sequencing depths
- Relative expression levels are maintained after target enrichment
- Performance is consistent across input levels and different RNA types
- Targeted RNA-seq improves sequence coverage of transcripts of interest at lower sequencing depths, thus reducing analysis time and experimental costs

Takara Bio USA, Inc. • A Takara Bio Company
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

Scan to download your copy of this poster and view additional posters, or visit http://info.clontech.com/Targeted-RNA-Seq-2016-Poster-Signup.html

800.662.2566
Visit us at www.clontech.com/NGS

Clontech TaKaRa cellartis

9.16 US (633831)