**TITLE:**

**Targeted RNA Sequencing Assay to Characterize Gene Expression and Genomic Alterations**

**AUTHORS:**

Dorrelyn P. Martin1, Jharna Miya1, Julie W. Reeser1, Sameek Roychowdhury1,2,3

**AUTHORS: INSTITUTIONS(S)/AFFILIATION(S) FOR EACH AUTHOR:**

Patacsil Martin, Dorrelyn

Department of Internal Medicine, Division of Medical Oncology

Comprehensive Cancer Center

The Ohio State University

Columbus, Ohio

[Dorrelyn.Martin@osumc.edu](mailto:Dorrelyn.martin@osumc.edu)

Miya, Jharna

Department of Internal Medicine, Division of Medical Oncology

Comprehensive Cancer Center

The Ohio State University

Columbus, Ohio

[Jharna.Miya@osumc.edu](mailto:Jharna.Miya@osumc.edu)

Reeser, Julie W.

Department of Internal Medicine, Division of Medical Oncology

Comprehensive Cancer Center

The Ohio State University

Columbus, Ohio

[Julie.Reeser@osumc.edu](mailto:Julie.Reeser@osumc.edu)

Roychowdhury, Sameek

Department of Internal Medicine, Division of Medical Oncology

Comprehensive Cancer Center

The Ohio State University

Columbus, Ohio

[Sameek.Roychowdhury@osumc.edu](mailto:Sameek.Roychowdhury@osumc.edu)

**CORRESPONDING AUTHOR:**

Sameek Roychowdhury, MD, PhD

Assistant Professor

Department of Internal Medicine, Division of Medical Oncology

Department of Pharmacology

Comprehensive Cancer Center

The Ohio State University

460 West 12th Avenue, Room 508

Columbus, Ohio 43210

Phone: (614) 685-5842

Fax: (614) 292-6356

Email: [sameek.roychowdhury@osumc.edu](mailto:sameek.roychowdhury@osumc.edu)

**KEYWORDS:**

RNA sequencing, hybridization, capture, gene fusions, expression

**SHORT ABSTRACT:**

We describe a targeted RNA sequencing-based method that includes preparation of indexed cDNA libraries, hybridization and capture with custom probes and data analysis to interrogate selected transcripts for gene expression, mutations, and gene fusions. Targeted RNAseq permits cost-effective, rapid evaluation of selected transcripts on a desktop sequencer.

**LONG ABSTRACT:**

RNA sequencing (RNASeq) is a versatile method that can be utilized to detect and characterize gene expression, mutations, gene fusions, and noncoding RNAs. Standard RNAseq requires 30-100 million sequencing reads and can include multiple RNA products such as mRNA and noncoding RNAs. We demonstrate how targeted RNAseq (capture) permits a focused study on selected RNA products using a desktop sequencer. RNAseq capture can characterize unannotated, low, or transiently expressed transcripts that may otherwise be missed using traditional RNAseq methods. Here we describe the extraction of RNA from cell lines, ribosomal RNA depletion, cDNA synthesis, preparation of barcoded libraries, hybridization and capture of targeted transcripts and multiplex sequencing on a desktop sequencer. We also outline the computational analysis pipeline, which includes quality control assessment, alignment, fusion detection, gene expression quantification and identification of single nucleotide variants. This assay allows for targeted transcript sequencing to characterize gene expression, gene fusions, and mutations.

**INTRODUCTION:**

Whole transcriptome or RNA sequencing (RNAseq) is an unbiased sequencing method to assess all RNA products. The goal of targeted RNAseq (Capture) is a focused evaluation of selected transcripts with increased sensitivity, dynamic range, reduced cost or scale, and increased throughput compared to standard RNAseq. Similar to standard RNAseq, targeted enrichment approaches can be used to evaluate gene expression, multiple RNA species such as mRNA, microRNA (miRNA), lncRNA[1](#_ENREF_1" \o "Wang, 2009 #1), other noncoding RNAs[2](#_ENREF_2), gene fusions[3](#_ENREF_3), and mutations[4-6](#_ENREF_4).

Capture involves hybridization of complementary oligonucleotides to enrich cDNA libraries for sequencing. The rationale for RNAseq Capture is similar to microarray approaches where complementary oligonucleotides or probes are hybridized to samples and then measured for relative abundance. For microarray technologies, expression is based on relative signal measured for transcripts binding to these probes. Microarrays are thus limited by range, potential background noise from non-specific binding, and cross-hybridization of probes. Furthermore, arrays have limited dynamic range for low and highly expressed transcripts compared to RNAseq[1](#_ENREF_1). Microarrays are widely utilized due to their reduced cost and high throughput capacity compared to RNAseq.

Here, we demonstrate a method for RNAseq Capture that offers a middle ground between RNAseq and microarray approaches for evaluating the transcriptome. RNAseq Capture has intermediate throughput, greater dynamic range and sensitivity, and is scaled for fast turnaround on desktop sequencers. RNAseq Capture also requires reduced computational resources in terms of storage space and data processing.

**PROTOCOL:**

Note: This protocol describes the simultaneous processing and analysis of four samples. This method is compatible with RNA isolated from cells, fresh frozen tissue and formalin-fixed paraffin-embedded tissue (FFPE). This protocol begins with 50-1000ng (250ng recommended) of starting RNA input for each sample.

1. **rRNA Depletion and Fragmentation of RNA Procedure**
   1. **rRNA Depletion**
      1. Remove elute, prime, fragment mix, rRNA removal mix, rRNA binding buffer and resuspension buffer from -20°C and thaw at room temperature. Remove elution buffer, rRNA removal beads and RNA/cDNA specific paramagnetic beads from 4°C and bring to room temperature.
      2. Add 0.25µg of RNA to PCR tube. Dilute the total RNA with nuclease-free ultra-pure water to a total final volume of 10µL. Add 5µL of rRNA binding buffer to each tube then add 5µL of rRNA removal mix and gently pipette up and down to mix. Place tubes in thermal cycler with lid pre-heated to 100°C and program thermal cycler to 68 °C for 5 minutes to denature the RNA.
      3. After RNA denaturation, remove tubes from thermal cycler and incubate at room temperature for 1 minute. Vortex rRNA removal bead tube vigorously to resuspend the beads. Add 35µL of rRNA removal beads to new tubes and transfer the RNA denaturation reaction (20µl) to tubes containing rRNA removal beads.
      4. Adjust pipette to 45µL and pipette quickly up and down 20X to mix. Incubate tubes at room temperature for 1 minute. Then place tubes in magnetic stand at room temperature for 1 minute. Transfer supernatant to newly labeled PCR tubes and place these tubes again in magnetic stand at room temperature for 1 minute. This ensures that no beads are transferred. Transfer supernatant to newly labeled PCR tubes.
      5. Vortex RNA/cDNA specific paramagnetic beads and add 99µL of beads to each tube. Gently pipette entire volume up and down 10X to mix. If starting with degraded-RNA, add 193µL of well-mixed RNA/cDNA specific paramagnetic beads to each tube. Incubate at room temperature for 15 minutes. Then place tubes on the magnetic stand at room temperature for 5 minutes. Remove and discard the supernatants.
      6. Add 200µL of freshly prepared 70% Ethanol (EtOH). Keep tubes on magnetic stand and take care to not disturb the beads. Incubate at room temperature for 30 seconds, then remove and discard supernatant. Allow tubes to stand at room temperature for 5-10 minutes to dry.
      7. Centrifuge thawed room temperature elution buffer at 600 × g for 5 seconds. Add 11µL of elution buffer to each tube and gently pipette up and down 10X to mix. Incubate tubes at room temperature for 2 minutes. Place tubes on magnetic stand at room temperature for 5 minutes. Transfer 8.5µL of the supernatant to newly labeled PCR tubes.
   2. **Fragmentation of rRNA depleted RNA**
      1. Add 8.5µL elute, 8.5µL prime, and 8.5µL fragment mix to each tube. Fently pipette up and down 10X to mix. Place tubes in thermal cycler with following program: pre-heated lid to 100°C, 94°C for 8 minutes, then 4°C hold.

NOTE: This step is designed to generate an average insert size of 155bp. If average fragment size of RNA sample is lower than 200bp, skip this step and proceed to First Strand cDNA Synthesis.

1. **cDNA Synthesis**
   1. **Synthesize First Strand cDNA**
      1. Remove first strand synthesis mix from -20°C and thaw at room temperature.
      2. Pre-program the thermal cycler with the following settings: pre-heat lid option and set to 100°C, then 25°C for 10 minutes, 42°C for 15 minutes, 70°C for 15 minutes and hold at 4°C. Save this program as “Synthesize 1st Strand.”
      3. Centrifuge thawed-first strand synthesis mix tube at 600 × g for 5 seconds. Mix 1µL reverse transcriptase with 9µL of first strand synthesis mix.
      4. Add 8µL of first strand synthesis and reverse transcriptase mix to each tube, gently pipette up and down 6X to mix. Centrifuge tubes for 4 seconds using a fixed speed mini tabletop centrifuge at 6.0 x g to bring liquid to bottom. Place tubes in the thermal cycler and select Synthesize 1st Strand. When the thermal cycler reaches 4°C, remove tubes and proceed immediately to Synthesize Second Strand cDNA.
   2. **Synthesize Second Strand cDNA**
      1. Pre-heat thermal cycler to 16°C with lid pre-heated to 30°C. Thaw second strand master mix and re-suspension buffer on ice. In advance, remove the bottle of paramagnetic beads from 4°C and let stand for at least 30 minutes to bring them to room temperature.
      2. Add 5µL of resuspension buffer to each PCR tube. Centrifuge second strand master mix at 600 × g for 5 seconds and add 20µL to each PCR tube. Place tubes in pre-heated thermal cycler at 16°C for 1 hour. When incubation is completed, remove from thermal cycler and allow tubes to come to room temperature.
      3. Vortex the paramagnetic beads until they are well dispersed. Add 90µL of well-mixed paramagnetic beads to each tube. Gently pipette the entire volume up and down 10X to mix. Incubate tubes at room temperature for 15 minutes. Place tubes on the magnetic stand at room temperature for 5 minutes.
      4. Remove and discard 135µL of supernatant then leave tubes on the magnetic stand to perform EtOH wash. Add 200µL freshly prepared 80% EtOH to each tube without disturbing the beads, then incubate at room temperature for 30 seconds. Remove and discard all of the supernatant from each. Repeat for a total of two 80% EtOH washes.
      5. Let tubes stand at room temperature for 5-10 minutes to dry on the magnetic stand. Centrifuge the thawed room temperature resuspension buffer at 600 × g for 5 seconds. Remove PCR tubes from magnetic stand. Add 17.5µL resuspension buffer to each PCR tube and gently pipette the entire volume up and down 10X to mix thoroughly. Incubate tubes for 2 minutes at room temperature.
      6. Place tubes on the magnetic stand at room temperature for 5 minutes, then transfer 15µL supernatant (double stranded cDNA) to 0.2ml PCR strip tubes. NOTE: This is a safe stopping point as cDNA can be stored at -20°C for up to 7 days.
2. **Library Preparation**
   1. **Adenylate 3’ Ends** 
      1. Remove A-tailing mix from -20°C and thaw at room temperature. Pre-program the thermal cycler with the following settings: choose the pre-heat lid option and set to 100°C, then set at 37°C for 30 minutes, 70°C for 5 minutes and hold at 4°C. Save this program as “ATAIL70.”
      2. Add 2.5µL of resuspension buffer to each tube. Then add 12.5µL of thawed A-tailing mix. Gently pipette the entire volume up and down 10X to mix thoroughly. Place the tubes in the thermal cycler and select ATAIL70. When the thermal cycler temperature is at 4°C, remove the PCR tubes from the thermal cycler and proceed immediately to ligate adapters.
   2. **Ligate Adapters**
      1. Remove appropriate RNA adapter tubes, stop ligation buffer and resuspension buffer from -20°C and thaw at room temperature. Do not remove the ligation mix tube from -20°C until instructed to do so in the protocol. Remove the bottle of paramagnetic beads from 4°C and let stand for at least 30 minutes to bring to room temperature.
      2. Pre-heat the thermal cycler to 30°C and choose the pre-heat lid option and set to 100°C. Centrifuge the thawed RNA adapter tubes at 600 × g for 5 seconds. Immediately before use, remove the ligation mix tube from -20°C storage.
      3. Add 2.5µL of resuspension buffer to each sample tube. Add 2.5µL of ligation mix. Return the ligation mix tube to -20°C storage immediately after use. Add 2.5µL of the thawed RNA adapter index to each sample tube. Gently pipette the entire volume up and down 10X to mix thoroughly.
      4. Centrifuge tubes for 4 seconds using a fixed speed mini tabletop centrifuge at 6.0 x g. Place tubes in the pre-heated thermal cycler. Close the lid and incubate at 30°C for 10 minutes.
      5. Remove tubes from thermal cycler and add 5µL of stop ligation buffer to each tube to inactivate the ligation. Gently pipette the entire volume up and down 10X to mix thoroughly.
      6. Repeat wash as described in 2.2.3-2.2.4, using 42µL of mixed paramagnetic beads, and discarding 79.5µL supernatant. With the tubes on the magnetic stand, let the samples air-dry at room temperature for 5-10 minutes.
      7. Remove the PCR strip tubes from the magnetic stand and add 52.5µL resuspension buffer to each tube. Gently pipette the entire volume up and down 10X to mix thoroughly. Incubate the tubes at room temperature for 2 minutes. Place the tubes on the magnetic stand at room temperature for 5 minutes or until the liquid is clear. Transfer 50µL of supernatant from each tube to new 0.2 ml PCR strip tubes. Take care not to disturb the beads.
      8. Repeat wash as described in 2.2.3-2.2.4, using 50µL of mixed paramagnetic beads, and discarding 95µL supernatant. With the tubes on the magnetic stand, let the samples air-dry at room temperature for 5-10 minutes.
      9. Remove the PCR strip tubes from the magnetic stand and add 22.5µL resuspension buffer to each tube. Gently pipette the entire volume up and down 10X to mix thoroughly.
      10. Incubate the tubes at room temperature for 2 minutes. Place the tubes on the magnetic stand at room temperature for 5 minutes or until the liquid is clear. Transfer 20µL supernatant from each tube to a new 0.2ml PCR strip tube. Take care not to disturb the beads. This is a safe stopping point and cDNA can be stored at -20°C for up to 7 days.
3. **Library Amplification**
   1. **Enrich DNA Fragments**
      1. Remove the PCR master mix, PCR primer cocktail and resuspension buffer from -20°C storage and thaw at room temperature. Remove the bottle of paramagnetic beads from 4°C storage and let stand for at least 30 minutes to bring to room temperature.
      2. Pre-program the thermal cycler with the following settings: choose the pre-heat lid option and set to 100°C, then set initial denaturation at 98°C for 30 seconds, 15 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 30 seconds, one final extension cycle at 72°C for 5 minutes and hold at 4°C. Save this program as “PCR.”
      3. Centrifuge the thawed PCR master mix and PCR primers tubes at 600 × g for 5 seconds. Add 5µL of thawed PCR primers to each sample tube. Add 25µL of thawed PCR master mix to each sample tube. Gently pipette the entire volume up and down 10X to mix thoroughly.
      4. Place the capped PCR strip tubes in the pre-programmed thermal cycler. Close the lid and run PCR program. When PCR is completed, remove tubes from thermal cycler and keep on ice.
      5. Repeat wash as described in 2.2.3-2.2.4, using 50µL of mixed paramagnetic beads, and discarding 95µL supernatant. With the tubes on the magnetic stand, let the samples air-dry at room temperature for 5-10 minutes.
      6. Remove tubes from the magnetic stand and add 32.5µL resuspension buffer to each sample tube. Gently pipette the entire volume up and down 10X to mix thoroughly. Incubate at room temperature for 2 minutes. Place the PCR tubes on the magnetic stand at room temperature for 5 minutes or until the liquid is clear. Then transfer 30µL supernatant to fresh 0.2ml PCR tubes.
      7. Perform quantification of cDNA using a fluorometer[7](#_ENREF_7" \o "Invitrogen, 2010 #23) and determine cDNA quality using a capillary electrophoresis system[8](#_ENREF_8). NOTE: This is a safe stopping point and cDNA can be stored at -20°C for up to 7 days. Note: This library is considered a RNAseq library.

Subsequent steps lead to a Capture library.

1. **Hybridization, Capture and Sequencing** 
   1. **Multiplexed Hybridization** 
      1. Remove custom hydrated probes, Cot-1 DNA, universal blocking oligos, and adapter specific blocking oligos from -20°C and thaw on ice.
      2. In a low bind 1.5 mL tube, combine 500ng DNA (125ng per sample when multiplexing 4 samples), 5µL Cot-1 DNA (1µg/µL), 1ul universal blocking oligos, 0.5µL p7 (6 nucleotide) adapter specific blocking oligos (this amount may need to be adjusted depending on multiplex conditions) and 0.5µL p7 (8 nucleotide) adapter specific blocking oligos (this amount may need to be adjusted depending on multiplex conditions).
      3. Place the sample tube in a vacuum concentrator with cap open facing the opposite direction of rotation. Dry contents at 45°C for 20 minutes or until complete evaporation of the liquid.

* + 1. Resuspend dried content with 8.5µL 2X hybridization buffer, 3.4µL hybridization component A and 1.1µL nuclease free water. Allow 10 minutes for resuspension and vortex every 2.5 minutes. Transfer resuspended-material to a 0.2 ml PCR tube and incubate at 95°C for 10 minutes on a thermal cycler.
    2. Remove hybridization sample tube from thermal cycler and add 2µL of custom probes resuspended at a concentration of 1.5pmol/µL. Alternatively, add 4µL of custom probes resuspended at a concentration of 0.75pmol/µL. Incubate hybridization reaction overnight (16-24 hours) at 65°C. NOTE: Probes purchased from various Vendors may be used and manufacturer’s instructions should be followed. The duration of the hybridization step may also vary.
  1. **Bead Preparation and Capture**
     1. Remove bottle of streptavidin-coupled paramagnetic beads from 4°C and equilibrate at room temperature for 30 minutes. Dilute 10X Wash Buffers (I, II, III, and Stringent) and 2.5X Bead Wash Buffer to create 1X working solutions.
     2. Aliquot 140µL of 1X wash buffer I into to a fresh 1.5ml tube. Heat entire amount of 1X stringent buffer and aliquot of 1X wash buffer I at 65°C in a heat block for at least 2 hours.
     3. Aliquot 100µL of streptavidin-coupled paramagnetic beads per capture into a 1.5ml tube. Place on magnet and discard supernatant. Add 200µL bead wash buffer per 100µL beads and vortex for 10 seconds. Place on magnet for 2-5 minutes or until supernatant is clear. Once supernatant is clear, discard it and repeat once more for a total of two washes.
     4. After removal of bead wash buffer, add equal volume bead wash buffer as initial starting volume (*i.e.* 100µL for one capture). Resuspend and transfer to a 0.2ml PCR tube. Place tube in magnetic rack for 2-5 minutes or until supernatant is clear. Discard supernatant.
     5. With both the hybridization sample and beads in the thermal cycler at 65°C, transfer the hybridization mix to the bead tube and pipette up and down 10X to mix. Incubate at 65°C for 45 minutes, vortex and spin down sample for 4 seconds using a fixed speed (6.0 x g) table-top mini centrifuge.
  2. **Bead Wash**
     1. Remove capture tube from thermal cycler and add 100µL pre-heated 1X Wash Buffer I to the tube and vortex for 10 seconds to mix. Transfer the mixture to a fresh low bind 1.5mL tube. Place the tube in a magnetic separation rack and allow 2-5 minutes for separation or until supernatant is clear. Discard supernatant.
     2. Add 200µL preheated 1X stringent wash buffer and pipette up and down 10 times to mix. Incubate at 65°C for 5 min. Place the tube in the magnetic separation rack and allow 2-3 minutes for separation or until supernatant is clear. Discard supernatant. Repeat stringent wash once more for a total of two washes.
     3. Add 200µL room temperature 1X wash buffer I and vortex for 2 minutes to mix. Place the tube in the magnetic separation rack and allowing 2-5 minutes for separation or until supernatant is clear. Discard supernatant.
     4. Add 200µL room temperature 1X wash buffer II and vortex for 1 minute to mix. Place the tube in the magnetic separation rack and allowing 2-5 minutes for separation or until supernatant is clear. Discard supernatant.
     5. Add 200µL room temperature 1X wash buffer III and vortex for 30 seconds to mix. Place the tube in the magnetic separation rack allowing 2-5 minutes for separation or until supernatant is clear. Discard supernatant.
     6. Remove the tube from the magnetic separation rack and add 20µL nuclease-free water to resuspend the beads. Mix thoroughly by pipetting up and down 10 times.
  3. **post Capture PCR Amplification**
     1. Remove paramagnetic beads from 4°C and equilibrate at room temperature for 30 minutes.
     2. Remove the hot start PCR Ready Mix (2X) and PCR Primer Mix from -20°C and thaw at room temperature then place on ice. Prepare library amplification Master Mix by combining 27.5µL of 2X hot start PCR Ready Mix with 2.75µL of PCR Primer 1 and 2.75µL of PCR primer 2 (these volumes are for 1 hybridization library plus 10% excess).
     3. Setup the reaction in PCR tube by adding 20µL of beads plus captured DNA with 30µL of library amplification master mix for a total volume of 50µL. Cap tube properly and vortex to mix. Centrifuge tubes for 4 seconds using a fixed speed mini table-top centrifuge at 6.0 x g.
        1. Set up the following PCR program: choose the pre-heat lid option and set to 100°C, then set initial denaturation at 98°C for 45 seconds, 10-12 cycles of denaturation at 98°C for 15 seconds, annealing at 65°C for 30 seconds and extension at 72°C for 60 seconds, one final extension cycle at 72°C for 60 seconds and hold at 4°C.
     4. Remove sample from thermocycler and add 75µL paramagnetic beads. Mix well and incubate at room temperature for 15 minutes.
     5. Place tubes on magnet at room temperature for 2-3 minutes and then remove supernatant. Wash beads on magnet by adding 200µL 80% Ethanol, incubating for 30 seconds then removing supernatant. Repeat for a total of two 80% washes.
     6. Incubate at room temperature for 5-10 minutes to allow beads to dry. Do not over dry to cracking. Resuspend beads in 22µL of Tris-EDTA pH 8.0 (1X TE Solution) and allow 3 minutes for elution. Place sample on magnet for 3-5 minutes then transfer 20ul of eluted product to a fresh low-bind 1.5 mL tube, ensuring no beads are carried over.
     7. Perform quantification of captured cDNA using fluorometer[7](#_ENREF_7" \o "Invitrogen, 2010 #23) and determine captured cDNA quality using a capillary electrophoresis system [8](#_ENREF_8).
  4. **Desktop Sequencer Loading Procedure** [**9**](#_ENREF_9)
     1. Dilute captured cDNA library to a final concentration of 4nM using 10mM Tris-Cl pH 8.5 with 0.1% Tween 20. Thaw 10N NaOH and Hybridization buffer on ice. Approximately 30 minutes before use, thaw desktop sequencer v2 reagent kit box 1 in room temperature water. Do not fill above MAX FILL LINE [10](#_ENREF_10). Please note this is a sequencing platform specific procedure and may vary per manufacturer’s instructions.
     2. Prepare 1mL of 0.2N NaOH by combining 20µL 10N NaOH with 980µL nuclease free water in a microcentrifuge tube (always prepare fresh). Dilute the PhiX library (library control) to 4nM by combining 2ul of 10nM library control with 3µL 10mM Tris-Cl pH 8.5 with 0.1% Tween 20.
     3. Denature final library and library control by combining 5µL of 4nM library with 5µL of 0.2N NaOH and vortex briefly to mix. Centrifuge tubes for 4 seconds using a fixed speed mini table-top centrifuge at 6.0 x g. Incubate at room temperature for 5 minutes to denature libraries.
     4. Add 990µL of pre-chilled hybridization buffer to the tubes containing 10µL of denatured libraries. This results in a 20pM library. Mark the denatured 20pM library with the date and can be stored for up to 3 weeks at -20°C.
     5. Mix 375µL of 20pM library control with 225µL of pre-chilled hybridization buffer and diluted library control to result in a 12.5pM library control. Invert several times to mix the solution.
     6. Combine 594µL of denatured final library with 6µL of 12.5pM denatured library control and vortex to mix. Set the combined sample library and library control aside on ice until samples are ready to load into the desktop sequencer reagent cartridge.

1. **Data Analysis** 
   1. **Sequence Quality Assessment**
      1. Calculate the quality of the raw sequence data (fastq files) using sequence quality assessment[11](#_ENREF_11). NOTE: This step helps to assess the data before it is further subjected to downstream analysis. The software runs with in-built parameters and produces a set of metrics for each fastq file.
   2. **Alignment**
      1. Align sequence reads (fastq files) to the reference human genome hg19 and transcriptome using Tophat2 [12](#_ENREF_12) (version 2.0.10) while providing known transcripts as a GTF file. The output is in the form of a binary alignment format called the BAM file.
      2. Perform post processing steps including sorting and indexing using Samtools[13](#_ENREF_13" \o "Li, 2009 #1850) (version 0.1.19) on the BAM file. Perform duplicate marking, reordering SAM, insert size calculation and adding or replacing read groups using Picard tools[14](#_ENREF_14) (version 1.84).
   3. **RNAseq Quality Assessment**
      1. Compute a series of quality control metrics for RNAseq data using RNAseq quality assessment. The input to this software is a BAM file [15](#_ENREF_15) from the Tophat2 alignment. The output is a HTML file that lists total read count, duplicates, mapped read percentage and rRNA percentage etc, among others.
   4. **Variant Calling**

Use STAR (version 2.4.0)[16](#_ENREF_16) for alignment and then call single nucleotide variants using GATK's (Version 3.3-0) HaplotypeCaller[17](#_ENREF_17).Follow GATK BAM post-processing steps and filtering criteria to flag and remove false positives from the output.

* 1. **Gene expression**
     1. Calculate gene expression using Cufflinks software (version 2.1.1) from Tuxedo suite[18](#_ENREF_18). NOTE: The input is a BAM file from Tophat2 alignment tool. The output is produced at the isoform, gene and transcript level, where expression is calculated as FPKM (Fragments per Kilobase per Million Mapped Reads).
  2. **Fusion calling**
     1. Call fusions from each sample using ChimeraScan[19](#_ENREF_19" \o "Iyer, 2011 #9) (version  0.4.5), Tophat Fusion[20](#_ENREF_20) custom and TRUP[21](#_ENREF_21). Annotate the fusions for domains using Oncofuse[22](#_ENREF_22" \o "Shugay, 2013 #12) (version 1.0.9b2).

**REPRESENTATIVE RESULTS:**

A schematic highlighting key steps in RNAseq Capture is shown in Figure 1. Four cancer cell lines with known mutations were used to demonstrate the effectiveness of the RNAseq Capture technique (K562 with *ABL1* fusion, LC2 with *RET* fusion, EOL1 with *PDGFRalpha* fusion and RT-4 with *FGFR3* fusion). The four samples were pooled together and sequenced with 2x100bp reads on a desktop sequencer, which generates FASTQ files. FASTQ files were run through an RNAseq analysis pipeline, which includes five main components: 1) quality control assessment, 2) alignment to human transcriptome, 3) gene expression quantification, 4) fusion calling, and 5) variant calling. The alignment file (BAM) is used to call single nucleotide variants and calculate gene expression. Fusions are called using fusion callers, such as TopHat Fusion (performing their own alignment) and the output is annotated using fusion detection software.

Comparison of gene expression from RNAseq and capture demonstrates enrichment of targeted transcripts by 10 to 1,000-fold using the capture method (Fig. 2A). Additionally, Fig. 2B shows an increase in the percent of reads mapping to the targeted transcript regions using capture compared to RNAseq. Assessment of quality control measures is represented in Fig. 3. Capture and RNAseq perform equally in terms of alignment to the transcriptome (3A, 94% vs. 93%) and mean insert size (3B, 174bp vs. 162bp). Using the capture method, a higher percentage of exonic regions are sequenced (3C, 77% vs. 60%), and conversely a lower percentage of intronic regions are sequenced (3D, 4% vs. 20%). Total read counts per sample are depicted in 3E, and as expected, RNAseq generated over 50-fold more reads than capture. Finally, the percentage of rRNA sequences present in each sample was lower using the capture method when compared to RNAseq (3F, 4% vs. 15%).

Fusion detection output shown in Table 1 is generated with normalized fusion supporting reads. Capture RNAseq was successful in detecting fusions for all four cell lines. Comparison of single nucleotide variants called in overlapping regions of capture and RNAseq is displayed in Figure 4. This demonstrates a high concordance of variants between Capture and RNAseq within the target region.

**FIGURES and TABLES LEGENDS:**

**Figure 1.** **Schematic of RNAseq Capture steps.** In this experimental demonstration, RNA is first depleted of ribosomal RNA, followed by chemical fragmentation and synthesis of complementary DNA (cDNA) using reverse transcriptase. Next, the cDNA is polyadenylated and ligated on both ends to platform-specific adaptors to generate a library. Only cDNA libraries with proper adaptors are then amplified by PCR. Libraries are then hybridized to custom oligonucleotide probes and captured using magnetic beads. This small amount of captured library must be amplified a second time to have enough for next generation sequencing. Multiple libraries can then be sequenced in parallel. Sequencing data is analyzed for RNA events of interest such as gene fusions, expression or mutations.

**Figure 2.** **Comparison of targeted genes in Capture versus RNAseq.**

**A,** Gene expression comparison between Capture and RNAseq in four cancer cell lines

K562, LC2, EOL1 and RT-4 measured by reads per kilobase per million mapped reads (FPKM)(Log scale). Targeted genes of interest are enriched (blue) compared to non-targeted genes (grey). **B,** Percentage of reads mapping to targeted region is increased in Capture versus RNA seq libraries in four cancer cell lines.

**Figure 3.** **Sequencing metrics of Capture versus RNAseq in four representative cancer cell lines.**

**A**, Percentage of reads mapping to the transcriptome, **B**, Mean insert size of libraries. **C**, Percentage of reads in exonic regions. **D**, Percentage of reads in intronic regions. **E**, Total sequencing reads. **F**, Percentage of reads mapping to ribosomal RNA.

**Table 1.** **Fusion Detection for Capture versus RNAseq of K562, LC2, EOL1 and RT-4.**

This Table displays four cancer cell lines and three different fusion detection algorithms, TopHat2, Chimerascan, and TRUP utilized in this demonstration. This Table demonstrates the ability to detect fusions with Capture using less than 10 million total reads compared to greater than 60 million reads utilized for RNAseq. Fusion supporting reads were calculated by dividing fusion supporting reads by kinase reads, multiplied by one million.

**Figure 4.** **SNV calling for Capture versus RNAseq.**

These Venn diagrams show the number of Single Nucleotide Variants (SNVs) that were detected by Capture and RNAseq for each of four cell lines (K562, LC2, EOL1 and RT-4). This illustrates high concordance of SNVs between Capture and RNAseq within targeted-region: K562 (81.3%), LC2 (78.3%), EOL1 (89.5%) and RT-4 (73.9%).

**DISCUSSION:**

RNAseq Capture is an intermediate strategy between RNAseq and microarray approaches for evaluating a selected part of the transcriptome. The advantages of Capture include reduced cost, rapid turnaround time on a desktop sequencer, high throughput, and detection of genomic alterations. The method can be adapted to characterize non-coding RNAs[23](#_ENREF_23), detect single nucleotide variants[4-6](#_ENREF_4), examine RNA splicing, and to identify gene fusions or structural rearrangements[24](#_ENREF_24). Further, this approach can be applied to clinical or processed samples that have undergone fixation with formalin and embedded in paraffin blocks[24](#_ENREF_24),[25](#_ENREF_25).

There are several significant benefits of RNAseq capture as compared to microarray, real-time quantitative PCR, Sanger sequencing and DNA sequencing. Microarray is limited by high background due to cross-hybridization and non-specific binding of probes. Quantification of genes with low expression is restricted due to background noise, while highly expressed gene measurements are affected by signal saturation[1](#_ENREF_1). Compared to RNAseq capture, real-time PCR proves difficult to reproduce. Additionally, RNAseq allows for detection of novel transcripts, requires less starting input material and can detect alternative splicing[26](#_ENREF_26). In contrast to Sanger sequencing, RNAseq allows for higher throughput and analysis of low expressed miRNA. Sanger sequencing has proved to be a valuable tool for verification of fusions with known exon-exon junctions and somatic DNA mutations, however identification of novel fusions is hindered by requirements of a priori candidate breakpoint. DNA sequencing is not cost efficient, requires larger storage space for data, and is incapable of detection of post-transcriptional modifications.

There are several critical steps involved in RNAseq Capture. First, to improve yield of library products from the RNA/cDNA specific paramagnetic beads and paramagnetic beads during washes, be cautious not to over dry the beads, which will lead to loss of yield. Also, do not under-dry the beads, ensure all ethanol is removed from the sample tubes, as ethanol can reduce cDNA yield. Second, the hybridization of cDNA libraries with complementary probes is dependent on consistent temperature, we recommend warming Wash Buffer I and Stringent Buffer to 65 **°**C for at least two hours in advance. Further, after hybridization it is essential to maintain 65 **°**C during the binding and wash steps.The probes used here were designed for the exons of genes of interest for drug development including kinases, genes involved in common rearrangements such as transcription factors, and house keeping genes. Moreover, gene content is customizable and capture panel sizes can vary. Further, as new information on genomic regions arises, additional probes can be designed and added to the existing capture panel.

Evaluation of alignment metrics, specifically on-target rate, provides information on how well the targeted region was enriched. A low on-target rate may be due to a failed hybridization and capture, whereby the desired target region was not captured and enriched. In this case, a re-hybridization and capture of the library set must be performed. A low on-target rate may also be due to failure to deplete rRNA, which can be confirmed by calculating the percentage of rRNA in the samples. High rRNA percentage within the sample will require re-preparation of the sample beginning with rRNA depletion. Additionally, if library concentration falls below the requirements for hybridization and capture, it would be advisable to optimize the amount of starting input for the sample type and quality (range: 50-1000ng).

While there are several advantages for targeted RNAseq applications, there are also limitations to consider. Samples with poor RNA quality based on RIN or degree of fragmentation may not yield quality libraries for sequencing. Several groups have demonstrated success with formalin fixed paraffin-embedded samples, however there are samples that will not pass for sequencing[24](#_ENREF_24),[25](#_ENREF_25),[27](#_ENREF_27). Further, since RNAseq Capture focuses on known transcripts, it loses the benefits of unbiased RNAseq for novel or unannotated transcripts. In addition, for SNP detection, RNAseq methods can only detect mutations in expressed transcripts.

Future opportunities of RNAseq Capture include research and clinical applications. Recent discovery of thousands of long non-coding RNA and their role in biology will require focused characterization. In the clinic, RNAseq Capture may extend beyond research testing and translate into clinical assays to characterize human disease such as cancer, infectious disease, and non-invasive testing. In conjunction with genomic sequencing approaches, RNAseq Capture can be integrated to study and characterize the expressed genome.

**ACKNOWLEDGMENTS:**

We give special thanks to Ezra Lyon, Eliot Zhu, Michele Wing, Esko Kautto and Eric Samorodnitsky for technical support. We would also like to thank Jenny Badillo for her administrative support for our team. We acknowledge the Ohio Supercomputer Center (OSC) for providing disk space, processing capacity, and support to run our analyses. We thank the Comprehensive Cancer Center (CCC) at The Ohio State University Wexner Medical Center for their administrative support of this work. S.R. and Team are supported by the American Cancer Society (MRSG-12-194-01-TBG), a Prostate Cancer Foundation Young Investigator Award, NHGRI (UM1HG006508-01A1), Fore Cancer Research Foundation, American Lung Association, and Pelotonia.

**DISCLOSURES:**

S.R. receives funding from Novartis and Ariad Pharmaceuticals for conduct of clinical trials. S.R. immediate family members own stock in Johnson and Johnson.

**REFERENCES:**

1 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63, doi:10.1038/nrg2484 (2009).

2 Mercer, T. R. *et al.* Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. *Nat Protoc* **9**, 989-1009, doi:10.1038/nprot.2014.058 (2014).

3 Maher, C. A. *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A* **106**, 12353-12358, doi:10.1073/pnas.0904720106 (2009).

4 Piskol, R., Ramaswami, G. & Li, J. B. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet* **93**, 641-651, doi:10.1016/j.ajhg.2013.08.008 (2013).

5 Quinn, E. M. *et al.* Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. *PLoS One* **8**, e58815, doi:10.1371/journal.pone.0058815 (2013).

6 Tang, X. *et al.* The eSNV-detect: a computational system to identify expressed single nucleotide variants from transcriptome sequencing data. *Nucleic Acids Res* **42**, e172, doi:10.1093/nar/gku1005 (2014).

7 Invitrogen. *Qubit dsDNA HS Assay Kit Manual*, <<http://www.science.smith.edu/cmbs/documents/QubitdsDNAHSAssay.pdf>> (2010).

8 Agilent. *Agilent D1000 ScreenTape System Quick Guide*, <<http://www.agilent.com/cs/library/usermanuals/Public/G2964-90032_ScreenTape_D1000_QG.pdf>> (2013).

9 Illumina. *Preparing Libraries for Sequencing on the MiSeq®*, <<https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/preparing-libraries-for-sequencing-on-miseq-15039740-d.pdf>> (2013).

10 Illumina. *MiSeq® Reagent Kit v2 Reagen Preparation Guide*, <<https://support.illumina.com/downloads/miseq_reagent_kit_reagent_preparation_guide.html>> (2012).

11 Andrews, S. *FastQC*, <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>> (2015).

12 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

13 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

14 Institute, B. *Picard*, <<http://picard.sourceforge.net/>> (2014).

15 DeLuca, D. S. *et al.* RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* **28**, 1530-1532, doi:10.1093/bioinformatics/bts196 (2012).

16 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

17 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **11**, 11 10 11-11 10 33, doi:10.1002/0471250953.bi1110s43 (2013).

18 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

19 Iyer, M. K., Chinnaiyan, A. M. & Maher, C. A. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* **27**, 2903-2904, doi:10.1093/bioinformatics/btr467 (2011).

20 Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* **12**, R72, doi:10.1186/gb-2011-12-8-r72 (2011).

21 Fernandez-Cuesta, L. *et al.* Identification of novel fusion genes in lung cancer using breakpoint assembly of transcriptome sequencing data. *Genome Biol* **16**, 7, doi:10.1186/s13059-014-0558-0 (2015).

22 Shugay, M., Ortiz de Mendibil, I., Vizmanos, J. L. & Novo, F. J. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. *Bioinformatics* **29**, 2539-2546, doi:10.1093/bioinformatics/btt445 (2013).

23 Clark, M. B. *et al.* Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat Methods* **12**, 339-342, doi:10.1038/nmeth.3321 (2015).

24 Cieslik, M. *et al.* The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res*, doi:10.1101/gr.189621.115 (2015).

25 Cabanski, C. R. *et al.* cDNA hybrid capture improves transcriptome analysis on low-input and archived samples. *J Mol Diagn* **16**, 440-451, doi:10.1016/j.jmoldx.2014.03.004 (2014).

26 Costa, C., Gimenez-Capitan, A., Karachaliou, N. & Rosell, R. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl Lung Cancer Res* **2**, 87-91, doi:10.3978/j.issn.2218-6751.2013.02.05 (2013).

27 Zhao, W. *et al.* Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419, doi:10.1186/1471-2164-15-419 (2014).