



Nov 02, 2021

# 3' mRNA-sequencing

Bryan Yoo<sup>1</sup>, Jessica Griffiths<sup>1</sup>, Sarkis Mazmanian<sup>1</sup>

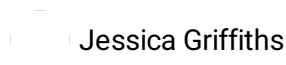
<sup>1</sup>California Institute of Technology

1



[dx.doi.org/10.17504/protocols.io.bzqbp5sn](https://dx.doi.org/10.17504/protocols.io.bzqbp5sn)

Mazmanian Lab



Jessica Griffiths

Protocol for RNAseq used in Yoo et al 2021

DOI

[dx.doi.org/10.17504/protocols.io.bzqbp5sn](https://dx.doi.org/10.17504/protocols.io.bzqbp5sn)

Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian 2021. 3' mRNA-sequencing.

**protocols.io**

<https://dx.doi.org/10.17504/protocols.io.bzqbp5sn>



Center for Environmental and Microbial Interactions

Grant ID: n/a

Emerald Foundation

Grant ID: n/a

Heritage Medical Research Institute

Grant ID: n/a

Department of Defense

Grant ID: PD160030

National Institutes of Health

Grant ID: GM007616 and DK078938

Aligning Science Across Parkinson's

Grant ID: ASAP-000375

protocol ,

Nov 02, 2021

Nov 02, 2021

### Tissue Collection and RNA extraction

- 1 Mice were cervically dislocated and the GI tract was removed.
- 2 1cm of tissue above and below the cecum were dissected and cleaned to represent tissue from the distal SI and proximal colon, respectively.
- 3 Tissue was homogenized in TRIzol (ThermoFisher Scientific, Waltham, MA-Cat. No. 15596018) solution using bead-based homogenizing methods, and total RNA was extracted using chloroform per manufacturer's instructions.

### Library Preparation

- 4 The cDNA libraries were prepared using the QuantSeq 3'mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, Greenland, NH) supplemented with UMI (unique molecular index) as per the manufacturer's instructions. Briefly, total RNA was reverse transcribed using oligo (dT) primers.
- 5 The second cDNA strand was synthesized by random priming, in a manner that DNA polymerase is efficiently stopped when reaching the next hybridized random primer, so only the fragment closed to the 3' end gets captured for later indexed adapter ligation and PCR amplification.
- 6 UMIs were incorporated to the first 6 bases of each read, followed by 4 bases of spacer sequences. UMIs are used to eliminate possible PCR duplicates in sequencing datasets and therefore facilitate unbiased gene expression profiling. The basic principle behind the UMI deduplication step is to collapse reads with identical mapping coordinates and UMI sequences. This step helps increase the accuracy of sequencing read counts for downstream analysis of gene expression levels.
- 7 The processed libraries were assessed for its size distribution and concentration using Bioanalyzer High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA-Cat. No. 5067-4626 and -4627).
- 8 Pooled libraries were diluted to 2 nM in EB buffer (Qiagen, Hilden, Germany, Cat. No. 19086) and then denatured using the Illumina protocol. The libraries were pooled and diluted to 2 nM using 10 mM Tris-HCl, pH 8.5 and then denatured using the Illumina protocol.

### Sequencing

- 9 The denatured libraries were diluted to 10 pM by pre-chilled hybridization buffer and loaded onto an Illumina MiSeq v3 flow cell for 150 cycles using a single-read recipe according to the manufacturer's instructions. Single-end 75bp reads (max 4.5M reads) were obtained. De-multiplexed sequencing reads were generated using Illumina BaseSpace.

#### Analysis

- 10 UMI specific workflows that were developed and distributed by Lexogen were used to extract reads that are free from PCR artifacts (i.e., deduplication).
- 11 First, the umi2index tool was used to add the 6 nucleotide UMI sequence to the identifier of each read and trims the UMI from the start of each read. This generates a new FASTQ file, which is then processed through trimming and alignment.
- 12 Second, after the quality and polyA trimming by BBDuk (Website-<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbdduk-guide/>) and alignment by HISAT2 (version 2.1.0) ([Kim et al., 2015](#)), the mapped reads are collapsed according to the UMI sequence of each read. Reads are collapsed if they have the same mapping coordinates (CIGAR string) and identical UMI sequences. Collapsing reads in this manner removes PCR duplicates.
- 13 Read counts were calculated using HTSeq ([Anders et al., 2015](#)) by supplementing Ensembl gene annotation (GRCm38.78).
- 14 Raw read counts were run through ShinySeq to obtain differentially expressed genes and downstream gene ontology analyses ([Sundararajan et al., 2019](#)).