



# A highly multiplexed and sensitive RNA-seq protocol for simultaneous analysis of host and pathogen transcriptomes

## Citation

Avraham, Roi, Nathan Haseley, Amy Fan, Zohar Bloom-Ackermann, Jonathan Livny, and Deborah T Hung. 2016. "A Highly Multiplexed and Sensitive RNA-Seq Protocol for Simultaneous Analysis of Host and Pathogen Transcriptomes." *Nature Protocols* 11 (8) (July 21): 1477–1491. doi:10.1038/nprot.2016.090.

## Published Version

doi:10.1038/nprot.2016.090

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33766540>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## **A highly multiplexed and sensitive RNA-Seq protocol for simultaneous analysis of host and pathogen transcriptomes.**

Roi Avraham<sup>1,7</sup>, Nathan Haseley<sup>1</sup>, Amy Fan<sup>1</sup>, Zohar Bloom-Ackermann<sup>1</sup>, Jonathan Livny<sup>1,2</sup>, Deborah T Hung<sup>1,3,4,5,6</sup>

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA, 02142.

<sup>2</sup>Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

<sup>3</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

<sup>4</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114, USA

<sup>5</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA

<sup>6</sup>correspondance to: hung@molbio.mgh.harvard.edu

<sup>7</sup>Current address: Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel, 76100.

Keywords: Host-pathogen interactions, RNA-Seq, bioinformatics.

**EDITORIAL SUMMARY:** This protocol enables simultaneous analysis of host and bacterial transcripts by RNA-Seq. Including procedures for efficient host and bacterial cell lysis, barcoding of samples, and analysis of both mammalian and microbial reads from mixed host-pathogen RNA-Seq data.

**TWEET: Simultaneous analysis of host and pathogen transcriptomes by RNA-Seq.**

### **Abstract**

The ability to simultaneously characterize the bacterial and host expression programs during infection would facilitate a comprehensive understanding of pathogen-host interactions. While RNA-Seq has greatly advanced our ability to study the transcriptomes of prokaryote and eukaryotes separately, limitations in existing protocols for generating and analyzing RNA-Seq data have hindered simultaneous profiling of host and bacterial pathogen transcripts from the same sample. Here we provide a detailed protocol for simultaneous analysis of host and bacterial transcripts by RNA-Seq. Importantly, this protocol details the steps required for efficient host and bacteria lysis, barcoding of samples, technical advances in sample preparation for low yield sample inputs and a computational pipeline to analyze both mammalian and microbial reads from mixed host-pathogen RNA-Seq data. Sample preparation takes 3 d from cultured cells to pooled libraries. Data analysis takes an additional day. Compared with previous methods, the protocol detailed here provides a sensitive, facile and generalizable approach, suitable for large-scale studies, which will enable the field to obtain in-depth analysis of host-pathogen interactions in infection models.

## Introduction

Intracellular bacterial pathogens, such as *Mycobacterium tuberculosis*, *Salmonella enterica*, *Legionella pneumophila*, and *Neisseria gonorrhoea*, spend a significant portion of their life-cycle surviving and replicating within host cells. The cellular interaction entails both a complex virulence program executed by the bacterial pathogen during infection<sup>1</sup> and activation of an orchestrated defense response by the host to counter the pathogen<sup>2</sup>. Genomic approaches have been employed in recent years to uncover substantial molecular details of this rich host-pathogen biology<sup>3,4</sup>. However, technical constraints limit these studies to profiling either the host or the pathogen – while a comprehensive understanding of host–pathogen interactions requires simultaneous analysis of the associated gene expression changes in both the pathogen and the host<sup>5</sup>. The limitations of conventional protocols for simultaneous analysis of host and pathogen transcripts include 1) the inability to obtain high quality RNA with efficient lysis of both bacterial and mammalian host cells; 2) inefficient depletion of both microbial and mammalian ribosomal RNA species; 3) inability to simultaneously process polyadenylated and non-polyadenylated transcripts from low yield samples (>10ng of RNA); and 4) a lack of robust computational approaches for analyzing the often small subset of bacterial transcripts in infected cells or tissue. Recently, several approaches were introduced<sup>6-8</sup> to approach simultaneous processing of host and bacterial pathogen transcripts, but they generally do not provide solutions to all of the mentioned limitations. Here, we provide a protocol that overcomes these limitations and allows inference of inter-species gene regulatory networks based on simultaneous gene expression data from samples of host cells infected with pathogenic bacteria<sup>9</sup>.

## Development of the protocol

In order to simultaneously analyze both host and bacterial pathogen transcripts, we extensively modified a conventional RNAtag-seq protocol<sup>10</sup> to overcome its limitations while preserving its advantages such as pooling of samples. RNAtag-seq initiates with immediate barcoding of RNA-purified samples through direct ligation of adaptors to RNA, followed by pooling of the samples. Next, species-specific rRNA removal and reverse transcription enables ligation of a second adapter to the 3' end of the cDNA. The last step of PCR amplification then allows for strand-specific, quantitative sequencing and assembly of full-length transcripts in prokaryotic or eukaryotic species<sup>10</sup>.

In order to adapt RNAtag-seq for simultaneous profiling of host and pathogen transcriptomes, the development of the following capabilities were required (**Figure 1**): firstly, the ability to effectively lyse both host and bacterial pathogen in a manner that allows recovery of RNA with high integrity and efficiency. For this, we have optimized two options: a gentle bead beating protocol that does not decrease the integrity of host RNA, and an efficient enzymatic reaction that also effectively lyse bacteria; Secondly, the ability to obtain host and bacterial transcripts with minimal RNA input. To achieve this, we have optimized the efficiency of the RNA ligation step; Next, the ability to simultaneously deplete both host and bacterial ribosomal RNA; and finally, the ability to analyze the sequencing data including demultiplexing of pooled samples by inline barcode, aligning reads to a composite transcriptome of both host and pathogen, and

using pathway-level enrichment to identify biologically relevant trends in low abundance bacterial data.

### **Applications, advantages and limitations**

The protocol is applicable to most infection models that aim to delineate the interaction between an intracellular bacterial pathogen and mammalian host cell. It overcomes major challenges including the need to lyse both host and bacterial cells in a manner that preserves the integrity of the RNA, to construct RNA-seq libraries with very low bacterial RNA input, to deplete host and bacterial ribosomal RNA in order to enrich for the mRNA signals, and to simultaneously analyze host and bacterial expression through alignment to both genomes. This protocol is also highly cost and time efficient with regards to sample processing and library preparation per sample. While the advantage of this method indeed include the ability to simultaneously capture both host and bacterial transcriptomes, one of its major limitations is that it cannot overcome the problem of the low abundance of bacterial mRNA within total RNA of infected cells except through high sequencing depth and dedicated analysis algorithms (provided below).

### **Comparison to other currently available methods**

RNAseq analysis of host-pathogen interactions has already been successfully applied for eukaryotic pathogens<sup>11-15</sup>. Transcripts of eukaryotic pathogens, unlike bacteria, are polyadenylated, and occur with comparable abundance to the hosts. Thus, conventional RNA-seq methods, such as Tru-seq, are readily applicable to host-fungal pathogen analysis. Other current available method that allows simultaneous analysis of host and bacterial pathogen by RNA-seq is the differential RNA-Seq (dRNA-Seq)<sup>16, 17</sup> as well as<sup>6-8</sup>. Compared to our protocol, these methods require a high amount of input RNA (*e.g.*, at least 15 µg total RNA of high quality is required for dRNA-seq<sup>16</sup>), and thus cannot be used for analyzing samples with limited RNA quantity (*e.g.*, FACS sorted populations or limiting clinical samples). Moreover, our protocol introduces inline RNA barcodes as one of the first stages of the procedure, which makes sample processing more compatible with large-scale studies, and reduces the cost for sample processing significantly. Last, the detailed protocol here provides extensive optimized solutions for efficient lysis of both host and bacteria.

### **Experimental design**

**Cell type, infection and sorting.** In general, this protocol should be applicable to any host cell type. The efficiency of lysis by the methods provided here was verified for both gram-negative and gram-positive bacteria. We have previously successfully applied this protocol for mouse bone marrow macrophages infected with *Salmonella*<sup>9</sup>. For these *in vitro* infection experiments, we recommend performing a time course experiment with triplicate samples for each time point to enable statistical significance to be reached (see<sup>9</sup>). We also recommend optimizing the conditions of the experiment with regards to Multiplicity Of Infection (MOI). In general, we found that infection of bone marrow macrophages with *Salmonella* at high MOIs (higher than 10:1), host transcriptional responses are dominated by the response to extracellular ligand<sup>9, 18</sup>. However, at low MOIs (lower than 5:1) most host cells remain uninfected, and it is important to distinguish between the response of infected and uninfected cells<sup>19</sup>. Thus, in order not to

analyze mixed infected and uninfected populations, and to avoid diluting the intracellular bacterial transcripts with host transcripts, we recommend sorting of the uninfected and infected cell populations. For this we have optimized the sensitivity the protocol for processing of sorted cell populations. For example, at low MOIs (lower than 5), 1-5% of mouse bone marrow macrophages are infected with *Salmonella*. We recommend starting with  $\sim 10^6$  host cells, to achieve a final sorted population of  $>10,000$  infected cells for analysis. We also offer a method to pre-label bacteria before infection to allow separation of infected and uninfected cell populations (see **Box 1**). We also recommend the following FACS settings to minimize sorting effects on gene expression of the cells: the input sample is maintained at 4°C while sorting; we use a 130 micron nozzle, so as not to apply any mechanical pressure on the cells; cells are sorted directly into the lysis buffer maintained at 4°C and are snap frozen immediately after sorting is done.

**Cell lysis.** The original RNAtag-seq protocol<sup>10</sup> was optimized for bacterial transcript analysis, and lysis methods involved enzymatic pretreatment and mechanical disruption, both of which can be deleterious to host RNA and inadequate for low yield samples. In this protocol, we have optimized the lysis of both host and bacterial cells either by using gentle bead disruption (for higher cell number;  $>10^6$ ) or by enzymatic treatment (for lower cell numbers;  $<10^5$ ). It is important to follow the lysis instructions detailed here as other bead beating methods or other cell wall lytic enzymes may degrade the host RNA, will not provide efficient bacterial lysis, or are not suitable for downstream RNA processing.

**Pooled cDNA library construction and sequencing.** Using this modified RNAtag-seq protocol, RNA samples are ligated to barcoded adapters, pooled, depleted of host and bacterial rRNA, and converted to Illumina cDNA libraries. Importantly, Illumina indexes can be added to each pool during library construction to enable demultiplexing of samples both by inline barcodes and Illumina indexes in data derived from the same sequencing lane. The relatively low abundance of bacterial transcripts does limit the level of multiplexing that can be used for these samples. In our recent study<sup>9</sup> in which we analyzed sorted infected populations of 10,000 cells, we loaded up to 12 libraries per lane of an Illumina HiSeq 2000 that is capable of  $\sim 300$  million reads. We found this to be sufficient to achieve near saturation of depth: To detect most expressed genes, RNA-seq libraries have to be sequenced to a depth of about  $10^7$  reads for most libraries. Libraries of particularly high interest can be resequenced to obtain additional reads.

**Analysis of mixed host-pathogen RNA-Seq data.** Many of the challenges of analyzing mixed host-pathogen data stem from the fact that bacterial transcripts often constitute a tiny fraction of the total RNA isolated from infected tissues or cells. Indeed, in our study of *Salmonella* infected macrophages<sup>9</sup>, over 99% of the reads were derived from host transcripts. In these cases, even infrequent misalignment of reads derived from highly abundant host RNA to the bacterial genome can greatly skew read counts for bacterial transcripts. Thus, robust approaches to minimize spurious read alignments is essential. Moreover, even with accurate alignment of reads to their cognate reference sequences, the low read counts for bacterial genes makes it difficult to ascertain the statistical significance of their differential abundance among samples.

To address these challenges, all reads derived from our host-pathogen RNA-seq data are aligned to a composite host and pathogen reference sequence database and host and bacterial transcripts are quantified separately to account for differences in pathogen burdens and variations in cell isolation efficiencies. These data are then mined for coordinated changes in the abundances of functionally related and/or transcriptionally co-regulated genes, enabling identification of biologically relevant transcriptional changes even when the differential expression of individual genes does not achieve statistical significance due to low read counts.

**Differential expression analysis of low abundance bacterial RNA-Seq data** It is frequently the case for the systems that we have worked with that the number of reads mapping to the pathogen genome will be exceedingly small relative to the number of host transcripts. Frequently these low read counts will prevent traditional forms of differential expression analysis from being effective as statistical significance can be very hard to achieve. To overcome this problem, we used a method detailed below relying on transcription factor network information. We describe this here to offer readers one possible avenue forward, realizing that there are many other alternative approaches that could be employed such as Motif-based sequence analysis (e.g MEME, <http://meme-suite.org/>), and the prediction of functional regulatory modules<sup>12</sup> if sufficient transcriptional data is available. We leave it to the reader to investigate these approaches.

## MATERIALS

### REAGENTS

- Optional: pHrodo™ Red, succinimidyl ester (Thermo Fischer, Cat#:P36600)
- Fresh mammalian cells (e.g. mouse bone marrow macrophages) infected with intracellular bacteria (e.g. *Salmonella*) (>10<sup>5</sup> cells are preferable to start with).  
**Optional:** Pre-label bacteria with the fluorescent dye pHrodo before infection for sorting of infected vs. uninfected cells (see **Box 1**).
- Optional: ERCC spike-in mix 1 (Ambion, cat.# 4456740)
- RNAGEM tissue; zyGEM, Cat.# RTI0050.
- Turbo DNase; Ambion/Applied Biosystems, Cat.# AM2239
- FastAP Thermosensitive Alkaline Phosphatase; Thermo Scientific, Cat.# EF0651
- RLT buffer; Qiagen, Cat.# 79216 (220mL). RLT is a lysis buffer, based on high concentration of guanidine isothiocyanate. The exact [composition is proprietary of Qiagen](#).
- T4 RNA Ligase 1 (Custom Order from NEB); 30,000U/mL (0.5mL) - 3x concentrated, normally 10,000U/mL, Cat.# M0204L
- T4 PNK (10U/μl) (NEB, Cat.# M0201S)
- ATP (100mM); Roche, Cat.# 11140965001 (aliquot and store at -80°C, always use fresh aliquot)
- DMSO (100%); Sigma, Cat.# D8418-50ML for Molecular Biology grade.

- PEG 8000, 50% (v/v) in water; Sigma, Cat.# 83271-100ML-F.
- RNase Inhibitor, Murine (40U/μl); NEB, Cat.# M0314S (3,000 units)
- SUPERase-IN 20U/μl (Ambion; AM2694; 2500U)
- AffinityScript Multiple Temperature cDNA Synthesis Kit, 50 rxn; Agilent, Cat.# 200436 – includes the dNTPs, 10x RT buffer, RNase Block Ribonuclease Inhibitor (40U/μl)
- Affinity Script RT Enzyme; Agilent, Cat.# 930107-51
- 10x AffinityScript RT buffer; Agilent, Cat.# 600100-52
- RNA Clean & Concentrator™-5 columns; Zymo Research, Cat.# R1015.
- RNAClean XP beads; Agencourt/Beckman, Cat.# A63987.
- AMPure XP beads; Agencourt/Beckman, Cat.# A63880
- **CRITICAL:** AMPure XP and RNAClean XP beads are stored at 4 °C, but make sure that beads are added to samples at room temperature (~25 °C). For convenience, make an aliquot to equilibrate it to room temperature in advance.
- Ribo-Zero™ Magnetic Kit (epidemiology); Epicentre/Illumina, Cat.# MRZE706
- 5N NaOH (Sigma, Cat# 656046-6X1L)
- Acetic Acid 20% (v/v) 1 mL aliquot (Sigma, Cat#537020)
- Nuclease free water
- 1x low TE (10mM Tris pH 8.0, 0.1mM EDTA); (Affymetrix/USB cat#75793)
- Flat PCR 12-cap strips, Optically clear; USA Scientific, Cat.# 1400-3120
- TempPlate No-skirt PCR Plates, 96 wells, 10 plates; USA Scientific, Cat.# 1402-9608

**Oligos (IDT):**

- 3Tr3 adapter: 5'-/5Phos/AGA TCG GAA GAG CAC ACG TCT G/3SpC3/-3'
- AR2 primer: 5' TAC ACG ACG CTC TTC CGA T 3'.
- Barcoded adapter: Oligonucleotides for sequencing barcodes are shown in **Table 1**. Barcoded adapters were chosen from a set designed and vetted as part of the development of RNAtag-Seq<sup>10</sup>. These adapters were found to yield similar numbers of reads per sample and produce highly correlated transcriptional profiles when used to barcode replicate RNA samples derived both from bacteria and from mammals
- Illumina Primer P5/P7 (12.5 μM) – see **Table 2**.

**CRITICAL:** Dissolve each adapter to 100 μM in water, and make aliquots of 15μl. These aliquots can be stored at -20 °C for up to 2–3 years.

**REAGENT SETUP**

**1N NaOH**

To prepare 1N NaOH, dilute 10 μl 5N NaOH + 40 μl water = 50 μl 1N NaOH

**CRITICAL:** Freshly prepare on the day of usage.

**0.5M Acetic Acid.**

To prepare 0.5M Acetic Acid, dilute 10 μl 3.5M stock + 60 μl water = 70 μl 0.5M Acetic Acid.

**CRITICAL:** Freshly prepare on the day of usage.

## EQUIPMENT

- DynaMag-96 side magnet (Life Technologies, cat. no. 12331D)
- 2100 Bioanalyzer system (Agilent Technologies)
- Qubit 3.0 fluorometer (Thermofisher scientific, cat.# Q33216)
- Benchtop centrifuge (Eppendorf 5417R)
- Thermocycler (MJ mini, Biorad).
- R software package (<http://www.r-project.org>) with gplots installed: the scripts used in this protocol have been tested with R v3.1.2
- Trimmomatic software tool: <http://www.usadellab.org/cms/?page=trimmomatic>.
- Piccard software tool: <http://broadinstitute.github.io/picard/>.
- BWA software tool: <http://bio-bwa.sourceforge.net/>

## PROCEDURE

### Lysis, RNA extraction and quality control

1. RNA extraction can be performed using Option A for limited number of cells, or Option B for  $>10^5$  cells:

**CRITICAL STEP:** RNA extraction should be carried out under RNase-free conditions.

**A: lysis and RNA extraction from a limited number of cells ( $\sim 10^5$  cells). Timing: 45 min to 1 hour.**

- i. Make extraction mix:

1	Cell pellet	up to 17 $\mu$ l
2	Silver buffer	2 $\mu$ l
3	RNAgem	1 $\mu$ l
	Total	20 $\mu$ l

- ii. Vortex and incubate at 75°C in a thermal cycler for 5 minutes.
- iii. Move immediately to ice.
- iv. Add 2.2  $\mu$ l of 10xTE buffer.
- v. Add 40  $\mu$ l of RNAClean XP beads. Pipette entire volume up/down 10 times, and leave for 15 minutes to bind the RNA.
- vi. Place tubes on magnet for at least 5 minutes (until liquid appears clear), and carefully remove and discard supernatant. Take care not to disturb the beads. Keep tubes on magnet.
- vii. Add 200  $\mu$ l freshly prepared 80% ethanol (vol/vol) without disturbing beads, leave for 30 seconds and remove and discard all of the supernatant from tube, take care not to disturb the beads.
- viii. Repeat previous Step.
- ix. Air dry on magnet for 15 minutes, until bead pellet dries.  
**CRITICAL STEP:** A sign of dry pellets is a loss of glossy surface and visible cracks.
- x. Once the pellet is dry, immediately add 14  $\mu$ l of RNase/DNase free water to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting and leave for 5 min.



- xi. Put the tube back on the magnet for 2–3 min, and then carefully transfer 12  $\mu$ l of the supernatant to a new tube.

**B: Lysis and RNA extraction for  $>10^5$  cells. Timing:** 1 to 2 hours

- i. Resuspend cell pellets in a freshly prepared mix of 250  $\mu$ l PBS and 250  $\mu$ l of buffer RLT.
- ii. Attach the OmniLyse cartridge to the Battery Pack via the blue connectors. Make sure to line up the red arrows.
- iii. Using the provided syringe, fill the chamber with PBS, and expel all the fluid.
- iv. Insert the syringe to the cell lysate and fill the chamber with 0.25 ml of cell lysate.
- v. Turn on the battery and slowly draw the majority of the remaining sample through the cartridge such that the cartridge remains filled with sample. Dispense the sample through the cartridge into the sample tube. Continue withdrawing and infusing the sample for exactly 2 minutes.  
**CRITICAL STEP:** Do not introduce air into the cartridge to avoid foaming of the sample.
- vi. Add equal volume (500  $\mu$ l) of EtOH (100%) to the sample tube.
- vii. Transfer up to 700  $\mu$ l of the sample to a Zymo-spin column, and spin for 30 seconds at  $>15,000 \times g$  at room temperature. Discard the flow-through. If sample is more than 700  $\mu$ l, repeat until the entire sample is loaded onto the column.
- viii. Add 400  $\mu$ l RNA Prep Buffer to the column and centrifuge for 30 seconds at  $>15,000 \times g$  at room temperature. Discard the flow-through.
- ix. Add 700  $\mu$ l RNA Wash Buffer to the column and centrifuge for 30 seconds at  $>15,000 \times g$  at room temperature. Discard the flow-through.
- x. Add 400  $\mu$ l RNA Wash Buffer to the column and centrifuge for 2 minutes at  $>15,000 \times g$  at room temperature to ensure complete removal of the wash buffer. Transfer the column carefully into an RNasefree tube.
- xi. Elute with 25  $\mu$ l nuclease free water.

**Pause point:** The RNA can be frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  for several months.

2. Check RNA quality using the Agilent bioanalyzer and verify the presence of the 18s and 28s rRNA peaks to confirm high integrity of the RNA and efficient lysis (see **Figure 2**; for concentrated samples use RNA Nano; for low concentration samples use RNA Pico kit). Quantify the RNA by qubit RNA HS assay.

**?Troubleshooting**

3. Calculate up to 100 ng of RNA and place in a tube. Increase the volume to 15  $\mu$ l with Nuclease free water. **Optional:** make a dilution of 1:2,000 of ERCC spike-in mix 1, and add 1  $\mu$ l to RNA with 14  $\mu$ l Nuclease free water.

**CRITICAL STEP:** The ERCC spike-ins are processed along with sample RNA through the RT reaction and subsequent cDNA amplification and sequencing, and can be used for the normalization of the RNA expression between samples<sup>20</sup>.

**CRITICAL STEP:** Each sample should have identical quantity of RNA.

4. Add 1  $\mu$ l of SUPERase-IN (20U/ $\mu$ l) to a Final total volume of 16  $\mu$ l.

**RNA fragmentation, DNA removal and RNA phosphorylation. Timing:** 1 hour and 30 minutes.

5. Set up the following reaction:

1	FastAP buffer (10x)	4 $\mu$ l
2	RNA from Step 4	16 $\mu$ l
	Total	20 $\mu$ l

6. Mix well. Incubate on pre-heated thermal cycler for 3 min at 94°C, and place immediately on ice.

7. Set up the following reaction mix

2	SUPERase-IN (20U/ $\mu$ l)	1 $\mu$ l
3	TURBO DNase (2U/ $\mu$ l)	4 $\mu$ l
4	FastAP (1U/ $\mu$ l)	10 $\mu$ l
5	T4 PNK (10U/ $\mu$ l)	2 $\mu$ l
6	Nuclease free water	3 $\mu$ l
	Total	20 $\mu$ l

**CRITICAL STEP:** For multiple samples, always prepare an excess (~10%) to ensure that all samples have the exact amount of mixture indicated in the protocol.

8. Add 20  $\mu$ l of the mix from Step 7 to the fragmented RNA of Step 6 and mix well.
9. Incubate on preheated thermal cycler for 30 min at 37°C to dephosphorylate the RNA.
10. Add 40  $\mu$ l of nuclease free water to each reaction, and then add 160  $\mu$ l of RNAClean XP beads. Pipette entire volume up/down 10 times, and leave for 15 minutes to bind the RNA.
11. Repeat Steps 1A-vi-ix.
12. Once the pellet is dry, immediately add 11  $\mu$ l of RNase/DNase free water to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting and leave for 5 min.
13. Put the tube back on the magnet for 2–3 min, and then carefully transfer 10  $\mu$ l of the supernatant to a new tube.
14. Take 5  $\mu$ l of each sample and proceed to first ligation (Step 15).
15. Add 1  $\mu$ l SUPERase-IN (20U/ $\mu$ l) to the remaining material and store at -80°C. RNA can be stored under these conditions for up to six months.

CRITICAL STEP: This can be used as back up if it is necessary to repeat process

**First ligation (RNA/DNA) 3' Linker Ligation. Timing:** 2 hours.

**CRITICAL:** Set up RNA/barcoded adapter in single tube or use the TempPlate No-skirt PCR Plates for batched samples (these rigid plates are easy to handle and can mix well by flicking by hand) – with a razor, cut a column (for 8 samples) or row (for 12 samples) and use as strip and cover with the Flat PCR 12-cap strips (these strips fit tightly on these plates and will not leak)

16. Add 1  $\mu$ l of barcoded adapter (100 pmole) to 5  $\mu$ l of dephosphorylated RNA from Step 13.
17. Heat at 70°C for 2 min and place immediately in cold block on ice.
18. Set up Ligation mix below

1	T4 RNA Ligase Buffer	2 $\mu$ l
2	DMSO (100%)	1.8 $\mu$ l
3	ATP (100 mM)	0.2 $\mu$ l
4	PEG 8000 (50%)	8 $\mu$ l
5	Barcoded Adapter 5' Phosphorylated/3' blocked 100 pmoles	1 $\mu$ l
6	RNase inhibitor, Murine (40U/ $\mu$ l)	0.3 $\mu$ l
7	T4 RNA Ligase 1 (30,000U/mL)	1.8 $\mu$ l
	Total	14.1 $\mu$ l

**CRITICAL STEP:** Make up mix at room temp so the reagents don't start precipitating when combined. Cut bottom of tip with sterile blade for easier aspiration of PEG (very viscous).

**CRITICAL STEP:** When setting up mix for multiple reactions include 25% extra to account for pipetting error due to the viscosity of the PEG in the mix

**CRITICAL STEP:** Mix really well by flicking tube since the solution is very viscous, Quick spin mastermix tube.

19. Add 14.1  $\mu$ l of ligation mix to each well containing 6  $\mu$ l denatured RNA + adapter from Step 16.

**CRITICAL STEP:** Mix well MANY times; mix by flicking since the solution is very viscous

20. Incubate at 22°C in a thermal cycler for 1 hour 30 minutes

#### **Pooling of barcoded samples using RNA Clean & Concentrator™-5 columns.**

**Timing:** 45 minutes.

**CRITICAL:** Maximum binding capacity of columns is 5  $\mu$ g; do not exceed when pooling samples

21. Add 60  $\mu$ l of RLT buffer to each sample to inhibit ligase and mix well (80  $\mu$ l total)

22. Add 80  $\mu$ l of binding buffer and 80  $\mu$ l EtOH (100%) to each reaction (binding buffer and Ethanol can be made as a mastermix and added simultaneously if working with multiple samples).

23. Repeat Steps 1Bvii-x

24. **CRITICAL STEP:** When pooling multiple samples >700  $\mu$ l onto Zymo column use a vacuum manifold. Elute 2 times with 16  $\mu$ l nuclease free water for a total volume of 32  $\mu$ l.

**CRITICAL STEP:** 2 elutions help improve recovery/yield of RNA.

**Optional:** Save 2  $\mu$ l for QC- Run on Agilent RNA pico chip to check the fragmentation profile of the pool and quantify with qubit RNA HS assay.

**Pause point:** The pooled RNA can be frozen in liquid nitrogen and stored at -80 °C for several months.

**rRNA depletion using Ribo-Zero™ Magnetic Kit (Epidemiology). Timing:** 1 hour and ten minutes.

**CRITICAL:** Ribo-zero (Epidemiology) rRNA depletion method was chosen based on<sup>21</sup> which evaluated rRNA depletion methods and chose a protocol that eliminates rRNA reads efficiently and robustly, largely irrespective of the quality of the RNA input sample.

25. Prepare magnetic beads: add 225 µl Magnetic Beads to RNase-free tube, and put on magnet for 1 minute.
26. Carefully remove and discard supernatant. Take care not to disturb the beads. Remove tubes from magnet and add 225 µl of RNase-free water.
27. Place beads on magnet for 1 minute and repeat previous Step.
28. Resuspend beads in 65 µl Resuspension Solution. Add 1 µl RiboGuard RNase Inhibitor from the Ribo-Zero kit.
29. Treat sample with rRNA Removal Solution:

1	RNA from Step 23	26 µl
2	rRNA Removal Solution	10 µl
3	Reaction Buffer	4 µl
	Total	40 µl

**CRITICAL STEP:** If using RNA extracted from a limited number of cells (~10<sup>5</sup> cells), use 28 µl of RNA, and 8 µl of rRNA Removal Solution.

30. Incubate in a preheated thermocycler at 68°C for 10 min then 5 min at RT
31. Add RNA mixture to Magnetic Beads from Step 27, mix well by pipetting.
32. Incubate 5 min at RT, vortex for 10 seconds.
33. Incubate 5 min in a preheated thermocycler at 50°C.
34. Transfer samples to a magnet, leave for 5 minutes and transfer supernatant (rRNA-depleted sample) to a new RNase-free tube (~90 µl). Take care not to disturb the beads.
35. Add 160 µl of RNAClean XP beads. Pipette entire volume up/down 10 times, and leave for 15 minutes to bind the RNA.
36. Repeat Steps 1Avi-xi

**First Strand cDNA synthesis. Timing:** 1 hour and five minutes.

37. Add 2 µl of AR2 primer to 12 µl of rRNA depleted RNA from Step 34 and mix well.
38. Incubate in a preheated thermocycler at 70°C for 2 min and immediately place on cold block on ice.
39. Make RT mix below:

1	AffinityScript RT Buffer	2 µl
2	DTT (0.1M)	2 µl
3	25mM dNTP Mix (25mM each)	0.8 µl
4	RNAse inhibitor, murine (40U/µl)	0.4 µl
5	AffinityScript RT Enzyme	0.8 µl

	Total	6 $\mu$ l
--	-------	-----------

40. Add 6  $\mu$ l of the above mix to each rRNA depleted RNA sample and mix well. Spin for 5 sec.
41. Place in a preheated thermocycler and incubate at 55°C for 55 minutes.
42. Add 2  $\mu$ l of 1N NaOH to each reaction, and incubate in a preheated thermocycler at 70°C for 12 minutes.
43. Neutralize with 4  $\mu$ l of 0.5M Acetic Acid and mix well.
44. Add 14  $\mu$ l of sterile water, and transfer to new tubes  
**CRITICAL STEP:** Do this step quickly as NaOH will start degrading tubes.
45. Add 80  $\mu$ l of AMPure XP beads. Pipette entire volume up/down 10 times, and leave for 15 minutes to bind the cDNA.
46. Repeat Steps 1Avi-ix.
47. Once the pellet is dry, immediately add 5  $\mu$ l of RNase/DNase free water to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting.

**CRITICAL STEP:** Do not transfer to new tubes and keep the samples with beads.

**Second ligation. Timing:** fifty-five minutes hands-on and overnight incubation.

48. Add 2  $\mu$ l of 3Tr3 adapter to cDNA with beads from Step 45.
49. Place in thermocycler preheated to 75°C for 3 min; remove and place immediately on cold block on ice
50. Make 2nd ligation reaction mix:

1	T4 Ligase Buffer	2 $\mu$ l
2	DMSO (100%)	0.8 $\mu$ l
3	ATP (100 mM)	0.2 $\mu$ l
4	PEG 8000 (50%)	8.5 $\mu$ l
5	T4 RNA Ligase 1 (30,000U/mL)	1.5 $\mu$ l
	Total	20 $\mu$ l

51. Swirl the cDNA with the adapter and beads from Step 47 with pipet tip and dispense 13  $\mu$ l of the ligation mix into the tube. Mix well by capping tubes and flick several times; solution is viscous. Quick spin.
52. Incubate overnight (~16 hours) in a thermocycler at 22°C.  
**CRITICAL STEP:** This second ligation is not as efficient as the first ligation (~50% after ~2hrs, more with overnight ligation).

53. Add 20  $\mu$ l RNase-free water, and then add 80  $\mu$ l of AMPure XP beads and mix up/down 15 times.
54. Leave for 20 minutes to bind the cDNA.
55. Repeat Steps 1Avi-ix, using 70% ethanol and air drying for 10 mins.
56. Once the pellet is dry, immediately add 27  $\mu$ l of RNase/DNase free water to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting.

57. Put the tube back on the magnet for 2–3 min, and then carefully transfer 25  $\mu\text{l}$  of the supernatant to a new tube.
58. Add 37.5  $\mu\text{l}$  of AMPure XP beads and Mix well 15 times.
59. Repeat Steps 54-57.

**PCR amplification, sequencing, and demultiplexing of Illumina libraries. Timing: 4 hours and 30 minutes.**

60. Make a PCR mix:

1	Rnase-free water	15.4 $\mu\text{l}$
2	10x AccuPrime PCR buffer 1	2.5 $\mu\text{l}$
3	Primer P5	1 $\mu\text{l}$
4	Primer P7	1 $\mu\text{l}$
5	AccuPrime HiFi Taq (5U/ $\mu\text{L}$ )	0.1 $\mu\text{l}$
	Total	20 $\mu\text{l}$

61. Mix 5  $\mu\text{l}$  of the cDNA from step 59 and add 20  $\mu\text{l}$  of the above mastermix and mix well.

**CRITICAL STEP:** Include a negative control water sample for each primer set.

**CRITICAL STEP:** divide the reaction from Step 61 into 4 separate tubes of 6  $\mu\text{l}$  and run them in separate wells.

62. Place in a thermal cycler and use the following cycling conditions.

Number of cycles	Temperature	Time
1	98°C	3 min.
12	98°C	30sec
	55°C	30sec
	65°C	60sec
1	65°C	10min
1	4°C	hold

Critical: Number of cycles and amount of cDNA indicated is optimized for a pool of 16 reactions, each starting with ~50ng total RNA-adjust the number of cycles or amount of input cDNA accordingly.

63. Pool the 4 wells together and increase reaction volume to 50  $\mu\text{l}$  with sterile water.
64. Add 75  $\mu\text{l}$  of AMPure XP beads and Mix well 15 times.
65. Repeat Steps 54-55.
66. Once the pellet is dry, immediately add 27  $\mu\text{l}$  1x low TE (10 mM Tris, 0.1M EDTA) to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting.
67. Put the tube back on the magnet for 2–3 min, and then carefully transfer 25  $\mu\text{l}$  of the supernatant to a new tube.
68. **CRITICAL STEP:** For the final PCR or following pooling of multiple library pools with unique P7 adapter barcodes, a second clean up is necessary to eliminate contaminants (adapter dimers, PCR primers  $\leq 160$  bp). To do this, add 17.5  $\mu\text{l}$  of AMPure XP beads and Mix well 15 times, then repeat Steps 54-55 followed by Steps 66-67. Assess the quality and size distribution of the library with an Agilent

2100 Bioanalyzer system (see **Figure 3** for the correct distribution of library size).

### ?Troubleshooting

69. If needed, go back to the cDNA sample from Step 59 and repeat Steps 60-68 and change the cycle numbers.
70. If the size distribution of the library is acceptable for sequencing, quantify the library using KAPA Illumina library quantification kit (which comes with all necessary reagents) and real-time PCR system.
71. Sequence the library, 50-bp paired-end reads, according to the manufacturer's recommendations. This will generate two .fastq files: one file contains the read 1 sequences and the second file contains the read 2 sequences.
72. Demultiplex samples based on inline barcodes using fastx\_clipper ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) or similar tools.

### **Creating a composite host-bacterial reference sequence database. Timing: 1-2 hours.**

73. Obtain the most up-to-date version of your host genome (.fasta or .fna file) with transcript annotations in .gtf format for the RSEM the aligner we use or .gff for other aligners. For most organisms these can be obtained at NCBI (<ftp://ftp.ncbi.nlm.nih.gov/>) or the UCSC genome browser (<https://genome.ucsc.edu/>).
74. Create an *in silico* transcriptome from the host genome and annotations using the RSEM rsem-prepare-reference command (<http://deweylab.github.io/RSEM/>) including the optional 3' polyadenylation of transcript sequences. **CRITICAL STEP:** Note that reads can be aligned to the whole host genome in lieu of the annotated transcriptome. This will decrease alignment accuracy but enable identification of previously unknown splice variants.
75. We have found that for even well annotated organisms such as mouse, the annotations for ribosomal RNAs are incomplete. To ensure the host reference transcriptome is comprehensive for these highly abundant transcripts, go to "Tools"-> "Table browser" at the UCSC genome browser. Select your organism of interest and then set group to all tables. Select the table "rnsk", then filter for repclass = rRNA. These annotations can be exported as a gtf and a set of non-polyadenylated transcripts can be generated using rsem-prepare-reference as described in Step 73 without polyadenylation.
76. Obtain transcript annotations for the pathogen. If the pathogen is bacterial, it is sufficient to simply use the bacterial genome. If the pathogen is eukaryotic or employs alternative splicing one would need to construct the transcriptome using the genome and gene annotations as described in Step 75.
77. Combine the host transcriptome and pathogen genome/transcriptome into a single file for alignment. In the UNIX/Linux environment this can be done by running: `cat path_ref_file host_ref_file > combined_ref_file`

### **Quality Control, Read alignment and transcript quantification. Timing: 4-5 hours.**

78. For each demultiplexed fastq file from Step 72, use FastQC to produce a basic quality control report. Pay particular attention to poor quality reads, and adaptor sequence contamination. Over 90% of reads should generally be high quality with

relatively consistent average per base quality scores (a score of  $\geq 30$ ) along the length of your read with some possible decrease in quality toward the 3' end of long reads. If this is not the case you may need to troubleshoot your library construction or sequencing protocol.

79. Trim low quality sequences and remove adaptors using programs such as trimmomatic.
80. Run the rsem-prepare-reference command on your composite transcriptome produced in Step 77.
81. For each .fastq file, run the rsem-calculate-expression command using your composite reference database to perform host alignment and transcript quantification. If your reads are paired-end, the first and second strand reads are located in separate .fastq files and this command should be run once for each read/mate pair of files.

**CRITICAL STEP:** Most alignment algorithms developed to date have been optimized for analysis of either eukaryotic or prokaryotic transcripts. Thus while all reads should be aligned to the same composite sequence database described in Step 77, analysis of host and bacterial transcripts should be conducted separately using different tools for alignment and transcript quantification (i.e. see Step 84 for host transcripts and 84-88 for bacterial transcripts).

82. Use the picard tools CollectAlignmentSummaryMetrics command (and optionally the CollectRNASeqMetrics command) or similar tools to check the quality of your alignments. If a significant number of reads ( $< 70\%$ ) are not aligned to any sequence in the composite reference, we recommend extracting a subset of these reads and using BLAST to determine if they correspond to abundant host RNAs not included in the reference database. These sequences can then be added to the composite transcriptome and Steps 80-81 can be repeated.
83. Use the rsem-generate-data-matrix function to extract raw counts for each transcript.
84. Extract read counts for all host transcripts using your programming language or text editor of choice. These can be used for differential expression analysis or other analyses of choice. **CRITICAL STEP:** Note that for analyses other than differential expression analysis, one will typically convert sequence counts to normalized estimates of expression such as TPM (Transcripts per million) or RPKM (reads per kilobase per million) both of which normalize for different transcript lengths and sequencing depths. In general TPM (which normalizes first to transcript length and then normalizes to sequencing depth using length normalized counts) is to be preferred, as it tends to yield more stable comparisons between samples. Regardless of the normalization method chosen, normalization should be based only on reads aligning to the host.
85. Use BWA to align bacterial reads<sup>22</sup>. First, run the bwa index command on the composite genome made in Step 77. **CRITICAL STEP:** BWA can be exchanged in this step for other aligners such as Bowtie<sup>23</sup> that effectively contend with the high frequency of overlapping and paralogous genes in bacterial genomes.
86. For each .fastq file, run the bwa aln command using the composite reference database to generate one .sai file for each .fastq file.



87. For paired end reads, run the `bwa sampe` command using the fastq files and the sai files generated above to generate alignment files for each pair of reads. For single end reads use `bwa samse`.
88. Use a standard method such as FeatureCounts<sup>24</sup> to produce a table of counts per bacterial gene for each sample that can be used for downstream differential expression analysis. If you convert raw counts to normalized counts such as RPKM, this conversion should be done using only reads that align to bacterial ORFs.
89. In many cases, the number of bacterial RNAs will be low and thus it will be important to assess the complexity of the bacterial portion of the RNA-Seq data. Specifically, use the picard tools EstimateLibraryComplexity to determine the proportion of identical reads among reads aligning to the bacterial genome and/or determine the percentage of bacterial genes to which multiple reads have aligned as described in<sup>25</sup>. A high proportion of duplicate bacterial reads and a low percentage of genes detected suggest low library complexity that could arise from a low proportion of bacterial RNA within total RNA isolated from infected cells or tissue and/or inefficient depletion of bacterial rRNA. This low complexity of the bacterial portion of the cDNA library can be confirmed using saturation curves to show that reduction of read numbers does not proportionally reduce the number of genes detected. These plots are generated by downsampling bacterial reads using the picard tools DownsampleSam command and plotting the number of genes detected above a set threshold of reads per gene as the data is downsampled. In cases of low library complexity, additional sequencing may not increase detection of genes and efforts to increase the amount of bacterial RNA, and/or improve rRNA depletion should be taken to improve bacterial library complexity. If these analyses indicate good library complexity with a low percentage of bacterial genes detected, additional sequencing depth may provide improved detection of genes across the dynamic range of abundances.

**Differential expression analysis of low abundance bacterial RNA-Seq data. Timing:** 1-6 hours depending on the time needed to find/construct transcription factor networks for your organism of interest.

90. Identify groups of genes whose expression is regulated by the same transcription factor. For *E. coli* and other well-studied model organism these can be mined from publicly available databases such as reglon DB [<http://www.ccg.unam.mx/en/projects/collado/regulondb>], TRACTOR\_DB (<http://www.tractor.lncc.br/>), CollecTF (<http://collectf.umbc.edu>), and RegPrecise (<http://regprecise.lbl.gov/RegPrecise/>). For strains closely related to strains included in these databases, co-regulated sets of genes can be determined based on homology to the experimentally defined or predicted groups of co-regulated genes reported in these databases. Moreover, numerous transcription factor regulons in diverse bacteria have been experimentally defined or computationally predicted and published. Finally, experimental or computational approaches such as RNA-Seq, CHIP-Seq, or prediction of transcription factor binding sites using published consensus sequences can be used to define *de novo* sets of co-regulated genes in a strain of interest.

91. In addition to defining sets of genes based on their co-regulation, genes can be grouped based on functional annotations such as Gene Ontology (GO)<sup>26</sup> or Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>27</sup>. For some strains GO annotations are available at <http://geneontology.org/page/download-annotations>. For those strains not included in this database or for, GO and KEGG annotations can be derived *de novo* using Blast2GO (<https://www.blast2go.com/>) and BlastKOALA (<http://www.kegg.jp/blastkoala/>).
92. Use DESeq2<sup>28</sup> or other statistical approaches such as edgeR or limma-voom<sup>29</sup> to obtain moderated fold-change estimates between your conditions of interest. If using DESeq2, we recommend including the beta prior distribution when running the R command DESeq. Use these fold change estimates to create a ranked list file for all the genes.
93. Use the ranked list file along with gene set groupings generated in steps 126 and 127 as input in GSEA<sup>30</sup> (<http://software.broadinstitute.org/gsea/index.jsp>) to identify statistically significant, concordant differences in gene expression among the samples.

### Timing

Step 1-4, RNA extraction option A: 45 min to 1 hour

Step 1-4, RNA extraction Option B: 1 to 2 hours (depending on number of samples).

Steps 5-15, RNA fragmentation, DNA removal and RNA phosphorylation: 1 hour and 30 minutes.

Steps 16-20, first ligation (RNA/DNA): 2 hours.

Steps 21-24, pooling of barcoded samples: 45 minutes.

Steps 25-36, rRNA depletion: 1 hour and ten minutes.

Steps 37-47, first Strand cDNA: 1 hour and five minutes.

Steps 48-59, second ligation: fifty-five minutes hands-on and overnight incubation.

Steps 60-72, PCR Enrichment of Illumina libraries: 4 hours and 30 minutes.

Steps 73-77, creating a composite host-bacterial reference sequence database: 1-2 hours.

Steps 78-89, quality Control, Read alignment and transcript quantification: 4-5 hours.

Steps 90-93, differential expression analysis of low abundance bacterial RNA-Seq data: 1-6 hours depending on the time needed to find/construct transcription factor networks for your organism of interest.

Timing of computational analysis will be highly dependent on sequencing depth and available compute time

### Box 1: Labeling of bacteria with the fluorescent dye pHrodo.

Grow an overnight culture of bacteria (depending on the experimental design, bacteria can then be used immediately or re-cultured to the desired growth stage).

1. Spin down bacteria at 13,000 g for 1 minute.
2. Remove the supernatant.

**CRITICAL STEP:** be careful not to disturb the cell pellet.

3. Resuspend the bacteria in 1 ml PBS and repeat Steps 1-2.
4. Resuspend the bacteria in 100mM sodium bicarbonate pH 8.0.
5. Resuspend an aliquot of pHrodo dye in 10  $\mu$ l of DMSO.

**CRITICAL STEP:** Resuspend 1 mg of pHrodo in 1ml acetyl nitryl, make 20 aliquots of 50 µl per tube. Speed-vac till dry (10 min) in dark. Aliquots can be then stored at -20°C for months.

6. Add 140 µl of bacteria from Step 4 to the pHrodo dye from Step 5 and mix well by pipetting 10 times.
7. Incubate in the dark at room temperature for 1 hour.
8. Resuspend the bacteria from Step 6 in 1ml HBSS and repeat Steps 1-2.
9. Repeat Step 8 two more times.

**CRITICAL STEP:** Supernatant will appear red in the first few washes. Make sure supernatant appears clear after these three washes. If not, repeat washes until clear.

10. Measure the OD600 of the bacteria after this procedure before infection. Carry out infections as described in<sup>9</sup>.

### ?Troubleshooting

### Anticipated results

**Step 23:** Typical results of the bioanalyzer RNA pico assay are presented in Figure 2. Whereas in A intact total RNA is evident by the two rRNA peaks, in B and C, one of the peaks is missing due to either degradation of RNA or insufficient lysis, respectively.

**Step 125:** After size selection, the pooled library is composed of fragments of a range of fragment sizes (usually 200–600 bp), after the last size-selection step. The 100-bp band that appears in Figure 3B is composed of primer dimers. Molarity of peaks does not notably change after the size-selection process, as evident after size selection in Figure 3A.

### Legends to figures:

**Figure 1: Overview of the simultaneous host-pathogen RNA-seq analysis protocol.**

**Figure 2: Electropherogram of total RNA extracted from *Salmonella* infected macrophages.** (A) Total RNA with high integrity. For host and bacteria samples, only host RNA is visible due to the low abundance bacterial RNA. RNA integrity is evident by the 18S and 28S eukaryotic rRNA (Please note – integrity of the RNA is calculated as the ratio of the host 18s and 28s fragment areas, recommended integrity is above 9). (B) Total RNA with evident degradation products, as shown for the lack of the 28s peak and small product below the 18s peak. (C) Insufficient lysis is evident by accumulation of small RNA around the 5s rRNA peak. (Please note that peak at 15 is a size marker).

**Figure 3: Electropherogram of the resulting RNA-seq libraries.** Electropherogram of a library either after (A) or before (B) size selection and purification, which is suitable for sequencing. Please note that peaks at 35 and 10,380 bp are size markers and are not included in the sequencing sample.

### Tables

**Table 1:** List of barcoded RNA adapters. (for 1<sup>st</sup> ligation step 16, red indicates barcode sequence)

	Barcode sequence (5'-3')	Oligo	Barcode read
1	ACATTATT	A <b>ACATTATT</b> AGATCGGAAGAGCGTCGTGTA	AATAATGT
2	ACCCATGT	A <b>ACCCATGT</b> AGATCGGAAGAGCGTCGTGTA	ACATGGGT
3	AAGTGTTG	A <b>AAGTGTTG</b> AGATCGGAAGAGCGTCGTGTA	CAACACTT
4	AGAATTAT	A <b>AGAATTAT</b> AGATCGGAAGAGCGTCGTGTA	ATAATTCT
5	ATATGGAC	A <b>ATATGGAC</b> AGATCGGAAGAGCGTCGTGTA	GTCCATAT
6	ATCACTTG	A <b>ATCACTTG</b> AGATCGGAAGAGCGTCGTGTA	CAAGTGAT
7	CCAAGTCG	A <b>CCAAGTCG</b> AGATCGGAAGAGCGTCGTGTA	CGACTTGG
8	CAACTCGC	A <b>CAACTCGC</b> AGATCGGAAGAGCGTCGTGTA	GCGAGTTG
9	CCCGTCTT	A <b>CCCGTCTT</b> AGATCGGAAGAGCGTCGTGTA	AAGACGGG
10	CCCTACAG	A <b>CCCTACAG</b> AGATCGGAAGAGCGTCGTGTA	CTGTAGGG
11	CCCTCGGC	A <b>CCCTCGGC</b> AGATCGGAAGAGCGTCGTGTA	GCCGAGGG
12	CCGGTACC	A <b>CCGGTACC</b> AGATCGGAAGAGCGTCGTGTA	GGTACCGG
13	CGGAGGGC	A <b>CGGAGGGC</b> AGATCGGAAGAGCGTCGTGTA	GCCCTCCG
14	CTCGGTAC	A <b>CTCGGTAC</b> AGATCGGAAGAGCGTCGTGTA	GTACCGAG
15	CGGCACTT	A <b>CGGCACTT</b> AGATCGGAAGAGCGTCGTGTA	AAGTGCCG
16	CTCTAACT	A <b>CTCTAACT</b> AGATCGGAAGAGCGTCGTGTA	AGTTAGAG
17	CTGGATCG	A <b>CTGGATCG</b> AGATCGGAAGAGCGTCGTGTA	CGATCCAG
18	GCAGCCAC	A <b>GCAGCCAC</b> AGATCGGAAGAGCGTCGTGTA	GTGGCTGC
19	GCCTGTAT	A <b>GCCTGTAT</b> AGATCGGAAGAGCGTCGTGTA	ATACAGGC
20	GAGATTGT	A <b>GAGATTGT</b> AGATCGGAAGAGCGTCGTGTA	ACAATCTC
21	GAGCCATC	A <b>GAGCCATC</b> AGATCGGAAGAGCGTCGTGTA	GATGGCTC
22	GTAAGTGC	A <b>GTAAGTGC</b> AGATCGGAAGAGCGTCGTGTA	GCAGTTAC
23	GGCCCAAG	A <b>GGCCCAAG</b> AGATCGGAAGAGCGTCGTGTA	CTTGGGCC
24	GTCTGGCG	A <b>GTCTGGCG</b> AGATCGGAAGAGCGTCGTGTA	CGCCAGAC
25	GGTCCTCT	A <b>GGTCCTCT</b> AGATCGGAAGAGCGTCGTGTA	AGAGGACC
26	GGTCTGGG	A <b>GGTCTGGG</b> AGATCGGAAGAGCGTCGTGTA	CCCAGACC
27	TCATCGTG	A <b>TCATCGTG</b> AGATCGGAAGAGCGTCGTGTA	CACGATGA
28	TACAACAT	A <b>TACAACAT</b> AGATCGGAAGAGCGTCGTGTA	ATGTTGTA
29	TCCCGCGG	A <b>TCCCGCGG</b> AGATCGGAAGAGCGTCGTGTA	CCGCGGGA
30	TACAGATG	A <b>TACAGATG</b> AGATCGGAAGAGCGTCGTGTA	CATCTGTA
31	TACCGGCC	A <b>TACCGGCC</b> AGATCGGAAGAGCGTCGTGTA	GGCCGGTA
32	TAGCTACA	A <b>TAGCTACA</b> AGATCGGAAGAGCGTCGTGTA	TGTAGCTA
33	TTACCACG	A <b>TTACCACG</b> AGATCGGAAGAGCGTCGTGTA	CGTGGTAA
34	TGAACCAG	A <b>TGAACCAG</b> AGATCGGAAGAGCGTCGTGTA	CTGGTTCA
35	TGGGAGAC	A <b>TGGGAGAC</b> AGATCGGAAGAGCGTCGTGTA	GTCTCCCA
36	TTTCTAAC	A <b>TTTCTAAC</b> AGATCGGAAGAGCGTCGTGTA	GTTAGAAA

**Table 2:** P5/P7 adapter sequences (for PCR step 60).

	<b>illumina P7 barcode name</b>	<b>end</b>	<b>sequence</b>
P5	-	5'	AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCT
P7	tagged_880	5'	CAAGCAGAAGACGGCATAACGAGATTCGCCAGAG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_367	5'	CAAGCAGAAGACGGCATAACGAGATTCGCTATGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_357	5'	CAAGCAGAAGACGGCATAACGAGATGGCTCCTGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_927	5'	CAAGCAGAAGACGGCATAACGAGATATCCGACAG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_289	5'	CAAGCAGAAGACGGCATAACGAGATAACATAATG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_422	5'	CAAGCAGAAGACGGCATAACGAGATATGGTAGGG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_800	5'	CAAGCAGAAGACGGCATAACGAGATGCTAAGTAG TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

**Table 3: Troubleshooting guide.**

<b>Step</b>	<b>Problem</b>	<b>Possible reason</b>	<b>Solution</b>
Box 1	Inefficient pHrodo labeling	1. Suboptimal conditions for labelling.	Make sure to handle bacteria gently during this step. Optimization of pHrodo concentrations may be needed.
		2. pHrodo has degraded	Make sure to aliquot pHrodo, and keep at -20 for several months
2	RNA integrity is low	1. RNA is degraded	Make sure to work in an RNase free environment: clean all surfaces and pipettes with RNase-ZAP, wear clean gloves at all times. Avoid multiple freeze-thaw cycles of the RNA.
		2. Inefficient lysis	2. Make sure to follow steps of lysis carefully. If choosing option A (chemical lysis), option B (mechanical lysis) provides more efficient lysis.
68	High levels of adapter dimers in the pooled libraries	Too low input RNA	The lower the sample concentration, the more adaptors remain after cleanup. Perform an extra SPRI purification cycle after the adaptor ligation cleanup by adding SPRI beads at an 0.7× ratio
	Low	Library	Ensure that fresh dNTP is used in the buffers

	library concentration	enzymatic steps are not optimal	
		Product is lost during SPRI cleanup steps	Perform the SPRI cleanup at the exact indicated SPRI buffer ratio. Be sure to remove all ethanol traces after washing

### Competing financial interests

The authors declare no competing financial interests.

### Contributions

R.A. designed the experiments. R.A., N.H., A.F. and Z.B.A. conducted the experimental work. N.H. and J.L. performed the computational analysis. R.A., J.L. and D.T.H. wrote the manuscript.

### Acknowledgments

This work was supported by NIH grants (HG002295 to N.H.)

### References

1. Galan, J.E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**, 567-573 (2006).
2. Medzhitov, R. TLR-mediated innate immune recognition. *Seminars in immunology* **19**, 1-2 (2007).
3. Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J.C. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. *Mol Microbiol* **47**, 103-118 (2003).
4. Berry, M.P. et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* **466**, 973-977 (2010).
5. Westermann, A.J., Gorski, S.A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* **10**, 618-630 (2012).
6. Baddal, B. et al. Dual RNA-seq of Nontypeable *Haemophilus influenzae* and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. *mBio* **6**, e01765-01715 (2015).
7. Humphrys, M.S. et al. Simultaneous transcriptional profiling of bacteria and their host cells. *PloS one* **8**, e80597 (2013).
8. Mavromatis, C.H. et al. The co-transcriptome of uropathogenic *Escherichia coli*-infected mouse macrophages reveals new insights into host-pathogen interactions. *Cell Microbiol* **17**, 730-746 (2015).

9. Avraham, R. et al. Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. *Cell* **162**, 1309-1321 (2015).
10. Shishkin, A.A. et al. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat Methods* **12**, 323-325 (2015).
11. Dillon, L.A. et al. Simultaneous transcriptional profiling of *Leishmania major* and its murine macrophage host cell reveals insights into host-pathogen interactions. *BMC Genomics* **16**, 1108 (2015).
12. Schulze, S., Henkel, S.G., Driesch, D., Guthke, R. & Linde, J. Computational prediction of molecular pathogen-host interactions based on dual transcriptome data. *Frontiers in microbiology* **6**, 65 (2015).
13. Tierney, L. et al. An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of *Candida albicans* Invading Innate Immune Cells. *Frontiers in microbiology* **3**, 85 (2012).
14. Kawahara, Y. et al. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PloS one* **7**, e49423 (2012).
15. Yazawa, T., Kawahigashi, H., Matsumoto, T. & Mizuno, H. Simultaneous transcriptome analysis of *Sorghum* and *Bipolaris sorghicola* by using RNA-seq in combination with de novo transcriptome assembly. *PloS one* **8**, e62460 (2013).
16. Bischler, T., Tan, H.S., Nieselt, K. & Sharma, C.M. Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods* **86**, 89-101 (2015).
17. Westermann, A.J. et al. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* **529**, 496-501 (2016).
18. Losick, V.P. & Isberg, R.R. NF-kappaB translocation prevents host cell death after low-dose challenge by *Legionella pneumophila*. *J Exp Med* **203**, 2177-2189 (2006).
19. Beattie, L. et al. A transcriptomic network identified in uninfected macrophages responding to inflammation controls intracellular pathogen survival. *Cell host & microbe* **14**, 357-368 (2013).
20. Risso, D., Ngai, J., Speed, T.P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature biotechnology* **32**, 896-902 (2014).
21. Giannoukos, G. et al. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome biology* **13**, R23 (2012).
22. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
23. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25 (2009).
24. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

25. Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W. & Livny, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**, 734 (2012).
26. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
27. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).
28. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550 (2014).
29. Sonesson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics* **14**, 91 (2013).
30. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).