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

Roi Avraham[1,6], Nathan Haseley[1], Amy Fan[1], Zohar Bloom-Ackermann[1], Jonathan Livny[1,2] & Deborah T Hung[1,3–5]

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [4]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [5]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts, USA. [6]Present address: Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel. Correspondence should be addressed to D.T.H. (hung@molbio.mgh.harvard.edu).

Published online 21 July 2016; doi:10.1038/nprot.2016.090

The ability to simultaneously characterize the bacterial and host expression programs during infection would facilitate a comprehensive understanding of pathogen–host interactions. Although RNA sequencing (RNA-seq) has greatly advanced our ability to study the transcriptomes of prokaryotes and eukaryotes separately, limitations in existing protocols for the generation and analysis of 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 for analysis of 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 that is suitable for large-scale studies and 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 pneumophilia* and *Neisseria gonorrhea*, spend a substantial 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[1] and activation of an orchestrated defense response by the host to counter the pathogen[2]. Genomic approaches have been used in recent years to uncover substantial molecular details of this rich host–pathogen biology[3,4]. 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[5]. The limitations of conventional protocols for simultaneous analysis of host and pathogen transcripts include (i) the inability to obtain high-quality RNA with efficient lysis of both bacterial and mammalian host cells; (ii) inefficient depletion of both microbial and mammalian rRNA species; (iii) inability to simultaneously process polyadenylated and nonpolyadenylated transcripts from low-yield samples (>10ng of RNA); and (iv) 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[6–8] 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[9].

Development of the protocol

To simultaneously analyze both host and bacterial pathogen transcripts, we extensively modified a conventional RNAtag-seq protocol[10] to overcome its limitations while preserving its advantages such as pooling of samples. RNAtag-seq starts 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 adaptor 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[10].

To adapt RNAtag-seq for simultaneous profiling of host and pathogen transcriptomes, the development of the following capabilities were required (**Fig. 1**): first, 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 lyses bacteria; second, 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 rRNA; and finally, the ability to analyze the sequencing data, including demultiplexing of pooled samples using inline barcodes, 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.

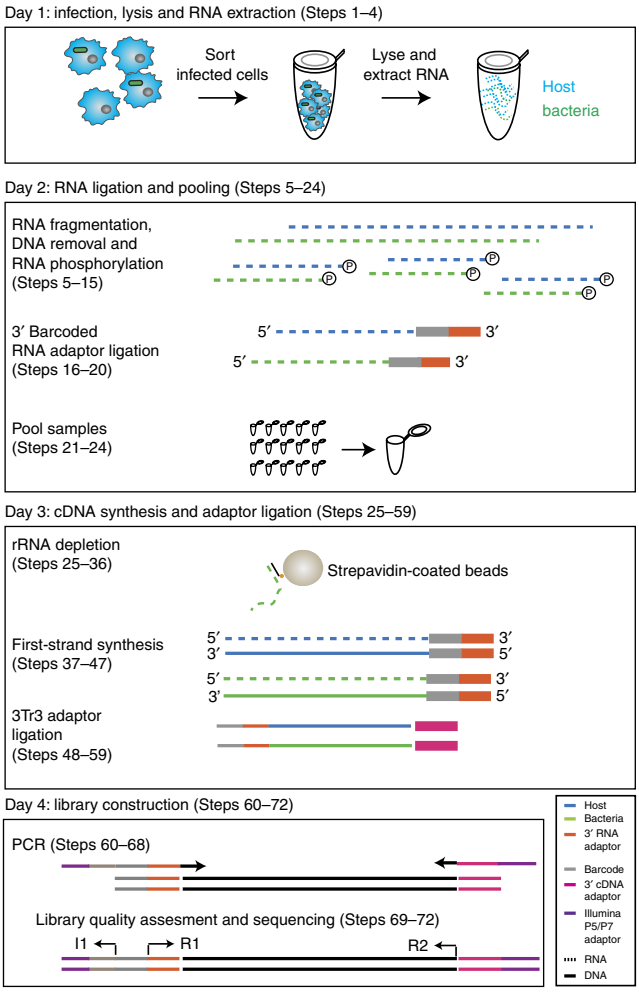

Figure 1 | Overview of the simultaneous host–pathogen RNA-seq analysis protocol. Dotted and solid lines correspond to RNA and DNA, respectively. The library is sequenced with paired-end read (R1 and R2). Inline RNA adaptors are read as part of the sequence read (R1). The P7 Illumina barcode is read from the index read (I1).

Applications, advantages and limitations

The protocol is applicable to most infection models that aim to delineate the interaction between an intracellular bacterial pathogen and a 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 rRNA in order to enrich for 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 regard to sample processing and library preparation per sample. Although the advantages 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 with other currently available methods

RNA-seq analysis of host–pathogen interactions has already been successfully applied for eukaryotic pathogens[11–15]. Transcripts of eukaryotic pathogens, unlike bacteria, are polyadenylated and occur with comparable abundance to that of the hosts. Thus, conventional RNA-seq methods such as Tru-seq are readily applicable to host–fungal pathogen analysis. Other current available methods that allow simultaneous analysis of host and bacterial pathogen by RNA-seq are differential RNA-seq (dRNA-seq)[16,17] and the methods discussed in refs. 6–8. Compared with our protocol, these methods require a high amount of input RNA (e.g., at least 15 µg of total high-quality RNA is required for dRNA-seq[16]), and thus they 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

Box 1 | Labeling of bacteria with the fluorescent dye pHrodo. ● TIMING variable

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

1. Centrifuge bacteria at 13,000*g* for 1 min at RT.
2. Remove the supernatant.

▲ **CRITICAL STEP** Be careful not to disturb the cell pellet.

3. Resuspend the bacteria in 1 ml of PBS, and repeat steps 1 and 2.
4. Resuspend the bacteria in 100 mM sodium bicarbonate, pH 8.0.
5. Resuspend an aliquot of pHrodo dye in 10 µl of DMSO.

▲ **CRITICAL STEP** Resuspend 1 mg of pHrodo in 1 ml of acetyl nitryl, and make 20 aliquots of 50 µl per tube. Speed-vacuum until they are dry (10 min), in the 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 the bacteria in the dark at RT for 1 h.
8. Resuspend the bacteria from step 7 in 1 ml of Hank's Balanced Salt Solution (HBSS), and repeat steps 1 and 2.
9. Repeat step 8 two more times.

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

10. After the final wash, resuspend the bacteria in 140 µl of HBSS.
11. Measure the OD_{600} of the bacteria after this procedure and before infection. Carry out infections as described in Avraham *et al.*[9].

? TROUBLESHOOTING

sample processing more compatible with large-scale studies and reduces the cost of sample processing. Finally, the detailed protocol described 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*[9]. 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 (Avraham *et al.*[9]). We also recommend optimizing the conditions of the experiment with regard to muliplicity of infection (MOI). In general, we found that when bone marrow macrophages are infected with *Salmonella* at high MOIs (>10:1), host transcriptional responses are dominated by the response to extracellular ligand[9,18]. However, at low MOIs (<5:1), most host cells remain uninfected, and it is important to distinguish between the responses of infected and uninfected cells[19]. Thus, to avoid analyzing 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 of the protocol for processing of sorted cell populations. For example, at low MOIs (<5), 1–5% of mouse bone marrow macrophages are infected with *Salmonella*. We recommend starting with ~10^6 host cells, to achieve a final sorted population of >10,000 infected cells for analysis. We also offer a method for prelabeling bacteria before infection to allow separation of infected and uninfected cell populations (**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-µm nozzle, so as not to apply any mechanical pressure to 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[10] 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 numbers; >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 adaptors, pooled, depleted of host and bacterial rRNA, and converted to Illumina cDNA libraries. Each infected cell sample is given 1 of the 36 inline RNA barcodes provided (**Table 1**), such that the same adaptor will be ligated to both host and bacterial RNAs from that sample. If you are using <36 samples, each barcode within a group of eight consecutive barcodes is divergent enough

TABLE 1 | List of barcoded RNA adaptors

	Barcode sequence (5′–3′)	Adaptor sequence (5′–3′)	Barcode read
1	ACATTATT	AACATTATTAGATCGGAAGAGCGTCGTGTA	AATAATGT
2	ACCCATGT	AACCCATGTAGATCGGAAGAGCGTCGTGTA	ACATGGGT
3	AAGTGTTG	AAAGTGTTGAGATCGGAAGAGCGTCGTGTA	CAACACTT
4	AGAATTAT	AAGAATTATAGATCGGAAGAGCGTCGTGTA	ATAATTCT
5	ATATGGAC	AATATGGACAGATCGGAAGAGCGTCGTGTA	GTCCATAT
6	ATCACTTG	AATCACTTGAGATCGGAAGAGCGTCGTGTA	CAAGTGAT
7	CCAAGTCG	ACCAAGTCGAGATCGGAAGAGCGTCGTGTA	CGACTTGG
8	CAACTCGC	ACAACTCGCAGATCGGAAGAGCGTCGTGTA	GCGAGTTG
9	CCCGTCTT	ACCCGTCTTAGATCGGAAGAGCGTCGTGTA	AAGACGGG
10	CCCTACAG	ACCCTACAGAGATCGGAAGAGCGTCGTGTA	CTGTAGGG
11	CCCTCGGC	ACCCTCGGCAGATCGGAAGAGCGTCGTGTA	GCCGAGGG
12	CCGGTACC	ACCGGTACCAGATCGGAAGAGCGTCGTGTA	GGTACCGG
13	CGGAGGGC	ACGGAGGGCAGATCGGAAGAGCGTCGTGTA	GCCCTCCG
14	CTCGGTAC	ACTCGGTACAGATCGGAAGAGCGTCGTGTA	GTACCGAG
15	CGGCACTT	ACGGCACTTAGATCGGAAGAGCGTCGTGTA	AAGTGCCG
16	CTCTAACT	ACTCTAACTAGATCGGAAGAGCGTCGTGTA	AGTTAGAG
17	CTGGATCG	ACTGGATCGAGATCGGAAGAGCGTCGTGTA	CGATCCAG
18	GCAGCCAC	AGCAGCCACAGATCGGAAGAGCGTCGTGTA	GTGGCTGC
19	GCCTGTAT	AGCCTGTATAGATCGGAAGAGCGTCGTGTA	ATACAGGC
20	GAGATTGT	AGAGATTGTAGATCGGAAGAGCGTCGTGTA	ACAATCTC
21	GAGCCATC	AGAGCCATCAGATCGGAAGAGCGTCGTGTA	GATGGCTC
22	GTAACTGC	AGTAACTGCAGATCGGAAGAGCGTCGTGTA	GCAGTTAC
23	GGCCCAAG	AGGCCCAAGAGATCGGAAGAGCGTCGTGTA	CTTGGGCC
24	GTCTGGCG	AGTCTGGCGAGATCGGAAGAGCGTCGTGTA	CGCCAGAC
25	GGTCCTCT	AGGTCCTCTAGATCGGAAGAGCGTCGTGTA	AGAGGACC
26	GGTCTGGG	AGGTCTGGGAGATCGGAAGAGCGTCGTGTA	CCCAGACC
27	TCATCGTG	ATCATCGTGAGATCGGAAGAGCGTCGTGTA	CACGATGA
28	TACAACAT	ATACAACATAGATCGGAAGAGCGTCGTGTA	ATGTTGTA
29	TCCCGCGG	ATCCCGCGGAGATCGGAAGAGCGTCGTGTA	CCGCGGGA
30	TACAGATG	ATACAGATGAGATCGGAAGAGCGTCGTGTA	CATCTGTA
31	TACCGGCC	ATACCGGCCAGATCGGAAGAGCGTCGTGTA	GGCCGGTA
32	TAGCTACA	ATAGCTACAAGATCGGAAGAGCGTCGTGTA	TGTAGCTA
33	TTACCACG	ATTACCACGAGATCGGAAGAGCGTCGTGTA	CGTGGTAA
34	TGAACCAG	ATGAACCAGAGATCGGAAGAGCGTCGTGTA	CTGGTTCA
35	TGGGAGAC	ATGGGAGACAGATCGGAAGAGCGTCGTGTA	GTCTCCCA
36	TTTCTAAC	ATTTCTAACAGATCGGAAGAGCGTCGTGTA	GTTAGAAA

All RNA adaptors require a 5′ phosphate (5Phos) and a 3′ blocking group (3SpC3; for first ligation (Step 18), red letters indicate barcode sequence).

TABLE 2 | P5/P7 primer sequences (for PCR, Step 60).

	Illumina P7 barcode name	Barcode sequence (5′–3′)	Sequence (5′–3′)
P5	—	—	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
P7	tagged_880	TCGCCAGA	CAAGCAGAAGACGGCATACGAGATTCGCCAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_367	TCGCTATG	CAAGCAGAAGACGGCATACGAGATTCGCTATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_357	GGCTCCTG	CAAGCAGAAGACGGCATACGAGATGGCTCCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_927	ATCCGACA	CAAGCAGAAGACGGCATACGAGATATCCGACAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_289	AACATAAT	CAAGCAGAAGACGGCATACGAGATAACATAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_422	ATGGTAGG	CAAGCAGAAGACGGCATACGAGATATGGTAGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P7	tagged_800	GCTAAGTA	CAAGCAGAAGACGGCATACGAGATGCTAAGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

to allow for correct sample assignment for sequencing (i.e., for eight samples use barcodes 1–8 or 9–16 or 17–24 or 25–32, and for ten samples use barcodes 1–10 or 17–26 and so on). Further, different Illumina P7 indexes (**Table 2**) can be added to each pool of 36 samples during library construction to enable further multiplexing of samples on each sequencing run and demultiplexing of samples both by inline barcodes and Illumina indexes in data derived from the same sequencing lane. Optimization of the PCR in the library construction step (Step 62) is required to lower amplification bias and to improve representation of all library fragments while retaining sufficient material for sequencing. In our experience, an optimal yield is achieved with 12 cycles of amplification for a pool of 16 samples, each sample starting with ~50 ng of total RNA. The number of PCR cycles should be tested for other sample inputs.

The relatively low abundance of bacterial transcripts does limit the level of multiplexing that can be used for these samples. In our recent study[9] 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 ~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 ~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[9], more than 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—are essential. Moreover, even with accurate alignment of reads to their cognate reference sequences, the low read counts for bacterial genes make 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 because of 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 the method detailed below, which relies 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 used, such as motif-based sequence analysis (e.g., MEME, http://meme-suite.org/) and the prediction of functional regulatory modules[12] if sufficient transcriptional data are available. We leave it to the reader to investigate these approaches.

MATERIALS

REAGENTS

- (Optional) pHrodo Red, succinimidyl ester (Thermo Fischer, cat. no. P36600)
- Fresh mammalian cells (e.g., mouse bone marrow macrophages extracted from femurs and tibiae of C57BL/6J mice (Jackson labs, stock no. 000664)) infected with intracellular bacteria (e.g., *Salmonella typhimurium* strain 14028 (ATCC, cat. no. 14028)) (>10^5 cells are preferable to start with; infections and macrophage extractions can be carried out as escribed in Avraham *et al.*[9]). Optionally, prelabel bacteria with the fluorescent dye pHrodo before infection for sorting of infected versus uninfected cells (**Box 1**)
- (Optional) External RNA Controls Consortium (ERCC) spike-in mix 1 (Ambion, cat. no. 4456740)
- Omnilyse lysis kit (Claremontbio, cat. no. 01.340.06)
- RNAGEM tissue (ZyGEM, cat. no. RTI0050)

- Turbo DNase (Ambion/Applied Biosystems, cat. no. AM2239)
- FastAP thermosensitive alkaline phosphatase (Thermo Scientific, cat. no. EF0651)
- RLT buffer, 220 ml (Qiagen, cat. no. 79216) RLT is a lysis buffer, based on high concentration of guanidine isothiocycanate. The exact composition is proprietary information
- T4 RNA ligase 1 (custom order from NEB), 30,000 U/ml (0.5 ml)—3× concentrated, normally 10,000 U/ml (cat. no. M0204L)
- T4 polynucleotide kinase (PNK), 10 U/μl (NEB, cat. no. M0201S)
- ATP 100 mM (Roche, cat. no. 11140965001)
- Dimethyl sulfoxide (DMSO), 100% (vol/vol) (Sigma-Aldrich, cat. no. D8418-50ML for molecular biology grade
- PEG 8,000, 50% (vol/vol) in water (Sigma-Aldrich, cat. no. 83271-100ML-F)
- Ethanol (EtOH; Sigma-Aldrich, cat. no. 459844)
- RNase inhibitor, murine, 40 U/μl (NEB, cat. no. M0314S (3,000 units))
- SUPERase-IN, 20 U/μl (Ambion, cat. no. AM2694, 2,500U)
- AffinityScript Multiple Temperature cDNA Synthesis Kit, 50 reactions (Agilent, cat. no. 200436)—includes the dNTPs, 10× RT buffer, RNase Block Ribonuclease Inhibitor (40 U/μl)
- Affinity script RT enzyme (Agilent, cat. no. 930107-51)
- 10× AffinityScript RT buffer (Agilent, cat. no. 600100-52)
- RNA Clean & Concentrator-5 columns (Zymo Research, cat. no. R1015)
- RNAClean XP beads (Agencourt/Beckman, cat. no. A63987)
- AMPure XP beads (Agencourt/Beckman, cat. no. A63880)
  ▲ **CRITICAL** AMPure XP and RNAClean XP beads can be stored at 4 °C for up to 18 months, but make sure that the beads are warmed to room temperature (RT; ~25 °C) before being added to samples. For convenience, make an aliquot and equilibrate it to RT in advance.
- Ribo-Zero Magnetic Kit (epidemiology) (Epicentre/Illumina, cat. no. MRZE706)
- 5 N NaOH (Sigma-Aldrich, cat. no. 656046-6X1L)
- 20% (vol/vol) Acetic acid, 1-ml aliquot (Sigma-Aldrich, cat. no. 537020)
- Nuclease-free water
- 1× Low TE (10 mM Tris, pH 8.0; 0.1 mM EDTA) (Affymetrix/USB, cat. no. 75793)
- Flat PCR 12-cap strips, optically clear (USA Scientific, cat. no. 1400-3120)
- TempPlate no-skirt PCR plates, 96 wells, ten plates (USA Scientific, cat. no. 1402-9608)
- RNA pico kit (Agilent, cat. no. 5067-1513)
- RNA nano kit (Agilent, cat. no. 5067-1511)
- KAPA Illumina library quantification kit (Kapabiosystems, cat. no. 07960140001)

Oligos (IDT):

- 3Tr3 adaptor: 5′-/5Phos/AGA TCG GAA GAG CAC ACG TCT G/3SpC3/-3′ (the 3Tr3 adaptor requires a 5′ phosphate (5Phos) and a 3′ blocking group (3SpC3))
- AR2 primer: 5′-TACACGACGCTCTTCCGAT-3′
- Barcoded RNA adaptors: oligonucleotides for sequencing barcodes are shown in **Table 1**. These barcoded adaptors were chosen from a set designed and vetted as part of the development of RNAtag-seq[10]. These adaptors were found to yield similar numbers of reads per sample and to produce highly correlated transcriptional profiles when used to barcode replicate RNA samples derived from both bacteria and mammals. The RNA adaptors require a 5′ phosphate (5Phos) and a 3′ blocking group (3SpC3)
- Illumina P5/P7 primers (12.5 μM)—see **Table 2**. ▲ **CRITICAL** Dilute each adaptor or primer to 100 μM in water and make 15-μl aliquots. These aliquots can be stored at −20 °C for up to 2–3 years.

EQUIPMENT

- DynaMag-96 side magnet (Life Technologies, cat. no. 12331D)
- 2100 Bioanalyzer system (Agilent Technologies)
- Qubit 3.0 fluorometer (Thermo Fisher Scientific, cat. no. Q33216)
- Benchtop centrifuge (Eppendorf 5417R)
- Thermocycler (MJ Mini, Biorad)
- Vacuum manifold (Promega Vac-Man)
- Real-time PCR system (Roche LightCycler 480)
- 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/
- fastx_clipper: http://hannonlab.cshl.edu/fastx_toolkit/

REAGENT SETUP

1 N NaOH To prepare 1 N NaOH, dilute 10 μl of 5 N NaOH with 40 μl of water to obtain 50 μl of 1 N NaOH. ▲ **CRITICAL** Freshly prepare the solution on the day of use.

ATP Prepare 1.5-μl aliquots of ATP and store them at −80 °C; always use a fresh aliquot.

0.5 M Acetic acid To prepare 0.5 M acetic acid, dilute 10 μl of 3.5 M stock with 60 μl of water to obtain 70 μl of 0.5 M acetic acid. ▲ **CRITICAL** Freshly prepare the solution on the day of use.

PROCEDURE

Lysis and RNA extraction

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 (~10^5 cells) ● **TIMING 45 min to 1 h**

(i) Make an extraction mix from 2 μl of silver buffer (from the RNAGEM kit), 1 μl of RNAGEM and a cell pellet of up to 17 μl, giving a total extraction mix volume of 20 μl.

(ii) Vortex and incubate the mixture at 75 °C in a thermal cycler for 5 min.

(iii) Move the mixture immediately to ice.

(iv) Add 2.2 μl of 10× TE buffer.

(v) Add 40 μl of RNAClean XP beads. Pipette the entire volume up/down ten times and leave it for 15 min to bind the RNA.

(vi) Place the tubes on a magnet for at least 5 min (until liquid appears clear), and carefully remove and discard the supernatant. Take care not to disturb the beads. Keep the tubes on the magnet.

(vii) Add 200 μl of freshly prepared 80% (vol/vol) ethanol without disturbing the beads; leave them for 30 s and remove and discard all of the supernatant from the tube; take care not to disturb the beads.

(viii) Repeat the previous step.

(ix) Air-dry the beads on a magnet for 15 min, until the bead pellet dries.
▲ **CRITICAL STEP** Signs of dry pellets are a loss of glossy surface and visible cracks.

(x) Once the pellet is dry, immediately add 14 μl of RNase/DNase-free water to the pellet, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting and leave the mixture for 5 min.

(xi) Put the tube back on the magnet for 2–3 min and then carefully transfer 12 μl of the supernatant to a new tube.

(B) Lysis and RNA extraction for >10^5 cells ● **TIMING 1–2 h**

(i) Resuspend cell pellets in a freshly prepared mix of 250 μl of PBS and 250 μl of buffer RLT.

(ii) Attach the OmniLyse cartridge to the battery pack via the blue connectors. Make sure to line up the red arrows.

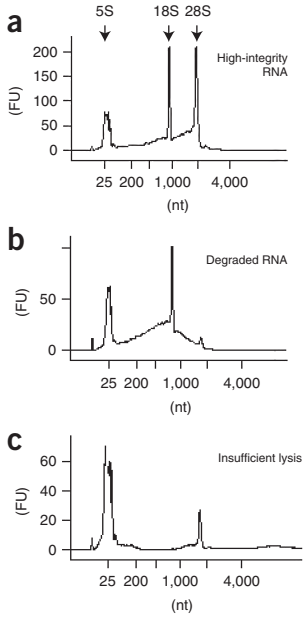

Figure 2 | Electropherogram of total RNA extracted from *Salmonella*-infected macrophages. (**a**) Total RNA with high integrity. For host and bacteria samples, only host RNA is visible because of the low-abundance bacterial RNA. RNA integrity is evident by the 18S and 28S eukaryotic rRNA. (Note that integrity of the RNA is calculated as the ratio of the host 18S and 28S fragment areas; the recommended integrity is >9). (**b**) Total RNA with evident degradation products, as shown by the lack of a 28S peak and little product below the 18S peak. (**c**) Insufficient lysis is indicated by accumulation of small RNA around the 5S rRNA peak. (Note that peak at 15 is a size marker.)

(iii) Using the provided syringe, fill the chamber with PBS, and expel all of the fluid.

(iv) Insert the syringe into 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 min.

▲ **CRITICAL STEP** Do not introduce air into the cartridge to avoid foaming of the sample.

(vi) Add an equal volume (500 µl) of EtOH (100%) to the sample tube.

(viii) Transfer up to 700 µl of the sample to a Zymo-Spin column, and centrifuge it for 30 s at >15,000*g* at RT. Discard the flow-through. If the sample is >700 µl, repeat until the entire sample is loaded onto the column.

(viii) Add 400 µl of RNA prep buffer to the column and centrifuge the mixture for 30 s at >15,000*g* at RT. Discard the flow-through.

(ix) Add 700 µl of RNA wash buffer to the column and centrifuge the mixture for 30 s at >15,000*g* at RT. Discard the flow-through.

(x) Add 400 µl of RNA wash buffer to the column, and centrifuge the mixture for 2 min at >15,000*g* at RT to ensure complete removal of the wash buffer. Transfer the column carefully into an RNase-free tube.

(xi) Elute with 25 µl of nuclease-free water.

■ **PAUSE POINT** The RNA can be frozen in liquid nitrogen and stored at –80 °C for several months.

RNA quality control ● TIMING 1 h

2| Check the 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 **Fig. 2**; for concentrated samples use RNA Nano; for low concentration samples use the RNA Pico kit). Quantify the RNA by qubit RNA HS assay.

? TROUBLESHOOTING

3| Calculate up to 100 ng of RNA and place it in a tube. Increase the volume to 15 µl with nuclease-free water. (Optional) Make a dilution of 1:2,000 of ERCC spike-in mix 1, add 1 µl to the 100 ng of RNA and increase the volume to 15 µl with 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 they can be used for the normalization of the RNA expression between samples[20].

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

4| Add 1 µl of SUPERase-IN (20 U/µl) to a final total volume of 16 µl.

RNA fragmentation, DNA removal and RNA phosphorylation ● TIMING 1 h and 30 min

5| Set up the following reaction:

Component	Volume (µl)
FastAP buffer (10×)	4
RNA from Step 4	16
Total	20

6| Mix the contents well. Incubate the mixture on a preheated thermal cycler for 3 min at 94 °C, and place it immediately on ice.

7| Set up the following reaction mix:

Component	Volume (µl)	Final concentration (U/µl)
SUPERase-IN (20 U/µl)	1	1
TURBO DNase (2 U/µl)	4	0.4
FastAP (1 U/µl)	10	0.5
T4 PNK (10 U/µl)	2	1
Nuclease-free water	3	
Total	20	

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

8| Add 20 µl of the mix from Step 7 to the fragmented RNA of Step 6 and mix well.

9| Incubate the mixture on a preheated thermal cycler for 30 min at 37 °C to dephosphorylate the RNA.

10| Add 40 µl of nuclease-free water to each reaction and then add 160 µl of RNAClean XP beads. Pipette the entire volume up/down ten times and leave it for 15 min to bind the RNA.

11| Repeat Step 1A(vi–ix).

12| Once the pellet is dry, immediately add 11 µl of RNase/DNase-free water to it, remove the tubes from the magnet, resuspend the pellet well by gentle pipetting and leave the solution for 5 min.

13| Put the tube back on the magnet for 2–3 min, and then carefully transfer 10 µl of the supernatant to a new tube.

14| Take 5 µl of each sample and proceed to first ligation (Step 16).

15| Add 1 µl of SUPERase-IN (20 U/µl) to the remaining material and store it at –80 °C.
■ **PAUSE POINT** RNA can be stored under these conditions for up to 6 months.
▲ **CRITICAL STEP** This can be used if it is necessary to repeat the process.

First ligation (RNA 3′-adaptor ligation) ● **TIMING 2 h**

16| Add 1 µl of barcoded adaptor (100 pmole, **Table 1**) to 5 µl of dephosphorylated RNA from Step 14.
▲ **CRITICAL STEP** Set up RNA/barcoded adaptor ligations in single tubes or use the TempPlate no-skirt PCR plates for batched samples (these rigid plates are easy to handle and the samples can be mixed well by flicking by hand). With a razor, cut a column (for eight 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).

17| Heat at 70 °C for 2 min and place immediately in a cold block on ice.

18| Set up the ligation mix as follows at RT so that the reagents do not start precipitating when combined. Mix the reagents really well by flicking the tube, as the solution is very viscous; next, briefly centrifuge the tube at RT at 8,000*g*.

Component	Volume (µl)	Final concentration
T4 RNA ligase buffer	2	
DMSO (100%)	1.8	12.7% (vol/vol)
ATP (100 mM)	0.2	1.41 mM
PEG 8000 (50%)	8	28.1% (vol/vol)
RNase inhibitor, murine (40 U/µl)	0.3	0.85 U/µl
T4 RNA ligase 1 (30,000 U/ml)	1.8	3,830 U/µl
Total	14.1	

▲ **CRITICAL STEP** Cut the bottom of the tip with a sterile blade for easier aspiration of PEG (very viscous).
▲ **CRITICAL STEP** When you are setting up the mix for multiple reactions, include 25% extra to account for pipetting error due to the viscosity of the PEG in the mix.

19| Add 14.1 µl of ligation mix to each well containing 6 µl of denatured RNA + adaptor from Step 16.
▲ **CRITICAL STEP** Mix the well contents many times; mix the contents by flicking, as the solution is very viscous.

20| Incubate the mixture at 22 °C in a thermal cycler for 1 h and 30 min.

Pooling of barcoded samples using RNA Clean & Concentrator-5 columns ● TIMING 45 min
▲ **CRITICAL** Maximum binding capacity of the columns is 5 µg; do not exceed when pooling samples.
21| Add 60 µl of RLT buffer to each sample to inhibit ligase, and mix the solution well (80 µl total).

22| Add 80 µl of binding buffer and 80 µl of EtOH (100%) to each reaction (binding buffer and ethanol can be made as a master mix and added simultaneously if you are working with multiple samples).

23| Repeat Step 1B(vii–x).
▲ **CRITICAL STEP** When pooling multiple samples >700 µl onto a Zymo column, use a vacuum manifold.

24| Elute the mixture two times with 16 µl of nuclease-free water for a total volume of 32 µl. Optionally, save 2 µl for quality control: run the mixture on the Agilent RNA pico chip to check the fragmentation profile of the pool, and quantify it with a qubit RNA HS assay.
▲ **CRITICAL STEP** Two elutions help improve recovery/yield of RNA.
■ **PAUSE POINT** The pooled RNA can be frozen in liquid nitrogen and stored at –80 °C for several months.

rRNA depletion using a Ribo-Zero Magnetic Kit (epidemiology) ● TIMING 1 h and 10 min
▲ **CRITICAL** The Ribo-Zero (epidemiology) rRNA depletion method was chosen based on Giannoukos *et al.*[21], 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| Preparation of magnetic beads. Add 225 µl of magnetic beads to an RNase-free tube and put it on a magnet for 1 min.

26| Carefully remove and discard the supernatant. Take care not to disturb the beads. Remove the tubes from the magnet and add 225 µl of RNase-free water.

27| Place the beads on the magnet for 1 min and repeat the previous step.

28| Resuspend the beads in 65 µl of Resuspension Solution from the Ribo-Zero kit. Add 1 µl of RiboGuard RNase Inhibitor from the Ribo-Zero kit.

29| Treat the sample with rRNA Removal Solution from the Ribo-Zero kit:

Component	Volume (µl)
RNA from Step 24	26
rRNA removal solution	10
Reaction buffer	4
Total	40

▲ **CRITICAL STEP** If you are using RNA extracted from a limited number of cells (~10^5 cells), use 28 µl of RNA and 8 µl of rRNA Removal Solution.

30| Incubate the mixture in a preheated thermocycler at 68 °C for 10 min and then for 5 min at RT.

31| Add the RNA mixture to magnetic beads from Step 28 and mix well by pipetting.

32| Incubate the mixture for 5 min at RT and vortex it for 10 s.

33| Incubate the mixture for 5 min in a preheated thermocycler at 50 °C.

34| Transfer the samples to a magnet, leave for them 5 min and transfer the 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 the entire volume up/down ten times, and leave it for 15 min to bind the RNA.

36| Repeat Step 1A(vi–xi).

First-strand cDNA synthesis ● TIMING 1 h and 5 min

37| Add 2 µl of AR2 primer to 12 µl of rRNA-depleted RNA from Step 36 and mix well.

38| Incubate the mixture in a preheated thermocycler at 70 °C for 2 min and immediately place it on a cold block on ice.

39| Make the reverse-transcription mix as follows:

Component	Volume (µl)	Final concentration
AffinityScript RT buffer	2	
DTT (0.1 M)	2	33 mM
25 mM dNTP mix (25 mM each)	0.8	3.33 mM
RNAse inhibitor, murine (40 U/µl)	0.4	2.67 U/µl
AffinityScript RT enzyme	0.8	
Total	6	

40| Add 6 µl of the above mix to each rRNA-depleted RNA sample and mix well. Centrifuge the mixture at RT for 5 s at >8,000*g*.

41| Place the mixture in a preheated thermocycler and incubate it at 55 °C for 55 min.

42| Add 2 µl of 1 N NaOH to each reaction and incubate in a preheated thermocycler at 70 °C for 12 min.

43| Neutralize the reaction with 4 µl of 0.5 M acetic acid and mix well.

44| Add 14 µl of sterile water, and transfer the mixture to new tubes.
▲ **CRITICAL STEP** Do this step quickly, as NaOH will start degrading the tubes.

45| Add 80 µl of AMPure XP beads. Pipette the entire volume up/down ten times, and leave it for 15 min to bind the cDNA.

46| Repeat Step 1A(vi–ix).

47| Once the pellet is dry, immediately add 5 µl of RNase/DNase-free water to the pellet, remove the tubes from the magnet and resuspend the pellet well by gentle pipetting.
▲ **CRITICAL STEP** Do not transfer the mixture to new tubes; keep the samples with beads.

Second ligation (3′-3Tr3 adaptor ligation) ● TIMING 17 h

48| Add 2 µl of 3Tr3 adaptor to cDNA with beads from Step 47.

49| Place the mixture in a thermocycler preheated to 75 °C for 3 min; remove and place it immediately on a cold block on ice.

50| Make the second ligation reaction mix as follows:

Component	Volume (µl)	Final concentration
T4 Ligase buffer	2	
DMSO (100%)	0.8	6% (vol/vol)
ATP (100 mM)	0.2	1.5 mM
PEG 8000 (50%)	8.5	32.5% (vol/vol)
T4 RNA ligase 1 (30,000 U/ml)	1.5	3460 U/ml
Total	13	

51| Swirl the cDNA with the adaptor and beads from Step 49 with a pipette tip and dispense 13 µl of the ligation mix from Step 50 into the tube. Mix well by capping the tubes and flicking them several times; if the solution is viscous, then briefly centrifuge it at RT and >8,000*g*.

52| Incubate the mixture overnight (~16 h) in a thermocycler at 22 °C.
▲ **CRITICAL STEP** This second ligation is not as efficient as the first ligation (~50% after ~2 h, more with overnight ligation).

53| Add 20 µl of RNase-free water and then add 80 µl of AMPure XP beads and mix up and down 15 times.

54| Leave the mixture for 20 min to bind the cDNA.

55| Repeat Step 1A(vi–ix), using 70% (vol/vol) ethanol and air-drying for 10 min.

56| Once the pellet is dry, immediately add 27 µl of RNase/DNase-free water to the pellet, remove the tubes from the magnet and resuspend the pellet well by gentle pipetting.

57| Put the tube back on the magnet for 2–3 min and then carefully transfer 25 µl of the supernatant to a new tube.

58| Add 37.5 µl of AMPure XP beads and mix the solution well 15 times.

59| Repeat Steps 54–57.

PCR amplification, sequencing and demultiplexing of Illumina libraries ● TIMING 4 h and 30 min
60| Make a PCR mix as follows:

Component	Volume (µl)	Final concentration
RNase-free water	15.4	
10× AccuPrime PCR buffer 1	2.5	1.25×
Primer P5 (**Table 2**)	1	5 µM
Primer P7 (**Table 2**)	1	5 µM
AccuPrime HiFi Taq (5 U/µl)	0.1	0.025 U/µl
Total	20	

61| Take 5 µl of the cDNA from Step 59 and add 20 µl of the above master mix and mix well. Divide the mixture into four separate tubes of 6 µl each and run them in separate wells.
▲ **CRITICAL STEP** Include a negative control water sample for each primer set.

62| Place the tubes in a thermal cycler and use the following cycling conditions:

Number of cycles	Temperature (°C)	Time
1	98	3 min
12	98	30 s
	55	30 s
	65	60 s
1	65	10 min
1	4	Hold

63| Pool the four wells together and increase the reaction volume to 50 µl with sterile water.

64| Add 75 µl of AMPure XP beads and mix well 15 times.

65| Repeat Steps 54 and 55.

66| Once the pellet is dry, immediately add 27 µl of 1× low TE (10 mM Tris, 0.1M EDTA) to the pellet, remove the tubes from the magnet and resuspend the pellet well by gentle pipetting.

67| Put the tube back on the magnet for 2–3 min and then carefully transfer 25 µl of the supernatant to a new tube.

68| Add 17.5 µl of AMPure XP beads and mix well 15 times; next, repeat Steps 54 and 55, followed by Steps 66 and 67.

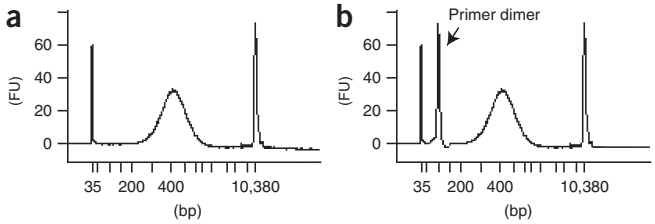

Figure 3 | Electropherogram of the resulting RNA-seq libraries. (**a**,**b**) Electropherogram of a library either after (**a**) or before (**b**) AMPure XP bead cleanup and purification, which is suitable for sequencing. Please note that peaks at 35 and 10,380 bp are size markers and are not included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

69| Assess the quality and size distribution of the library with an Agilent 2100 Bioanalyzer system (see **Fig. 3** for the correct distribution of library size).
? TROUBLESHOOTING

70| If the size distribution of the library is acceptable for sequencing, quantify the library using the KAPA Illumina library quantification kit (which comes with all necessary reagents) and a real-time PCR system. If not, go back to the cDNA sample from Step 59 and repeat Steps 60–69, changing the PCR cycle numbers, before proceeding to sequence the library.

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 other contains the read 2 sequences.

72| Demultiplex the samples based on inline barcodes using fastx_clipper or similar tools.

Creating a composite host–bacterial reference sequence database ● TIMING 1–2 h
73| Obtain the most up-to-date version of your host genome (.fasta or .fna file) with transcript annotations in .gtf format for RSEM, the aligner we use, or .gff for other aligners. For most organisms, these can be obtained at NCBI (http://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 will enable the identification of previously unknown splice variants.

75| We have found that for even well-annotated organisms such as mouse, the annotations for rRNAs are incomplete. To ensure that 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 'rmsk', then filter for repclass = rRNA. These annotations can be exported as a gtf file, and a set of nonpolyadenylated transcripts can be generated using rsem-prepare-reference, as described in Step 74 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 uses alternative splicing, one would need to construct the transcriptome using the genome and gene annotations, as described in Step 74.

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 h
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 to adaptor sequence contamination. Over 90% of reads should generally be high quality, with relatively consistent average per-base quality scores (a score of ≥30) along the length of your read and 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, although 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 Steps 81–84 for host transcripts and Steps 85–89 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 substantial 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 whether 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 and 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[22]. 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[23] 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[24] 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 open reading frames (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 to determine the percentage of bacterial genes to which multiple reads have aligned, as described in Haas *et al.*[25]. A high proportion of duplicate bacterial reads and a low percentage of genes detected suggest low library complexity, which 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. Generate such plots 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 are downsampled.
▲ **CRITICAL STEP** 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 to 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 h

▲ **CRITICAL** These steps take 1–6 h, depending on the time needed to construct transcription factor networks for your organism of interest.

90| Identify groups of genes whose expression is regulated by the same transcription factor. For *Escherichia coli* and other well-studied model organisms, 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 and 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 those from the Gene Ontology (GO) Consortium[26] or the Kyoto Encyclopedia of Genes and Genomes (KEGG)[27]. 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 (ref. 28) or other statistical approaches such as edgeR or limma-voom[29] to obtain moderated fold-change estimates between your conditions of interest. If you are using DESeq2, we recommend including the beta before 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 the gene set groupings generated in Steps 126 and 127 as input in gene set enrichment analysis (GSEA)[30] (http://software.broadinstitute.org/gsea/index.jsp) to identify statistically significant, concordant differences in gene expression among the samples.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 3**.

TABLE 3 | Troubleshooting table.

Step	Problem	Possible reason	Solution
Box 1	Inefficient pHrodo labeling	Suboptimal conditions for labeling	Make sure to handle bacteria gently during this step. Optimization of pHrodo concentrations may be needed
		pHrodo has degraded	Make sure to aliquot pHrodo and keep it at –20 °C for several months
2	RNA integrity is low	RNA is degraded	Make sure to work in an RNAse-free environment: clean all surfaces and pipettes with RNAse-ZAP, and wear clean gloves at all times. Avoid multiple freeze–thaw cycles of the RNA
		Inefficient lysis	Make sure to follow the steps of lysis carefully. If you have chosen option A (chemical lysis), consider option B (mechanical lysis), which provides more efficient lysis
69	High levels of adaptor dimers in the pooled libraries	Concentration of input RNA too low	The lower the sample concentration, the more adaptors remain after cleanup. Perform an extra solid-phase reversible immobilization (SPRI) purification cycle after the adaptor ligation cleanup by adding AMPure XP beads at an 0.7× ratio
	Low library concentration	Library enzymatic steps are not optimal	Ensure that fresh dNTP is used in the buffers
		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

● TIMING

Step 1, Lysis and RNA extraction (option A): 45 min to 1 h
Step 1, Lysis and RNA extraction (option B): 1–2 h (depending on the number of samples)
Steps 2–4, RNA quality control: 1 h
Steps 5–15, RNA fragmentation, DNA removal and RNA phosphorylation: 1 h and 30 min
Steps 16–20, first ligation (RNA 3′-adaptor ligation): 2 h
Steps 21–24, pooling of barcoded samples using RNA Clean & Concentrator-5 columns: 45 min
Steps 25–36, rRNA depletion using a Ribo-Zero Magnetic Kit (epidemiology): 1 h and 10 min
Steps 37–47, first-strand cDNA synthesis: 1 h and 5 min
Steps 48–59, second ligation (3′-3Tr3 adaptor ligation): 17 h (55 min hands-on and overnight incubation)
Steps 60–72, PCR amplification, sequencing and demultiplexing of Illumina libraries: 4 h and 30 min
Steps 73–77, creating a composite host–bacterial reference sequence database: 1–2 h
Steps 78–89, quality control, read alignment and transcript quantification: 4–5 h
Steps 90–93, differential expression analysis of low-abundance bacterial RNA-seq data: 1–6 h, depending on the time needed to find/construct transcription factor networks for your organism of interest
Box 1, Labeling of bacteria with the fluorescent dye pHrodo: variable (timing of computational analysis will be highly dependent on sequencing depth and available computer time)

ANTICIPATED RESULTS

Step 2

Typical results of the bioanalyzer RNA pico assay are presented in **Figure 2**. Whereas in **Figure 2a** intact total RNA is evident from the two rRNA peaks, in **Figure 2b**,**c**, one of the peaks is missing because of either degradation of RNA or insufficient lysis, respectively.

Step 69

After AMPure XP bead cleanup, the pooled library is composed of fragments of a range of fragment sizes (usually 200–600 bp). The 100-bp band that appears in **Figure 3b** is composed of primer dimers. The molarity of the peaks does not notably change after the cleanup process, as is evident in **Figure 3a**.

ACKNOWLEDGMENTS This work was supported by an NIH grant (HG002295 to N.H.)

AUTHOR 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.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

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. *Semin. Immunol.* **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–e01715 (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. *Front. Microbiol.* **6**, 65 (2015).
13. Tierney, L. *et al.* An interspecies regulatory network inferred from simultaneous RNA-seq of *Candida albicans* invading innate immune cells. *Front. Microbiol.* **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. *Nat. Biotechnol.* **32**, 896–902 (2014).

21. Giannoukos, G. *et al.* Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* **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 Biol.* **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 Res.* **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 Biol.* **15**, 550 (2014).
29. Soneson, 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. USA* **102**, 15545–15550 (2005).