

Discovering *Yersinia*-Host Interactions by Tissue Dual RNA-seq

Maria Kusmirek¹⁺, Ann Kathrin Heroven¹⁺, Michael Beckstette¹, Aaron M. Nuss¹, and Petra Dersch^{1*}

¹Department of Molecular Infection Biology,
Helmholtz Centre for Infection Research, Braunschweig, Germany

*Corresponding author:

Petra Dersch
Dept. of Molecular Infection Biology,
Helmholtz Centre for Infection Research
38124 Braunschweig
Germany
Phone: +49-531-6181-5700
FAX: +49-531-6181-5709
e-mail: petra.dersch@helmholtz-hzi.de

+: These authors contributed equally to this work.

Running head: *Yersinia*-Host Interaction Identified by Tissue Dual RNA-seq

Keywords: Tissue Dual RNA-seq, Host-Pathogen interaction, *Yersinia* infection, host-adapted metabolism, non-coding RNAs, Transcriptomics

Abstract

A detailed knowledge about virulence-relevant genes, as well as where and when they are expressed during the course of an infection is required to obtain a comprehensive understanding of the complex host-pathogen interactions. The development of unbiased probe-independent RNA sequencing (RNA-seq) approaches has dramatically changed transcriptomics. It allows simultaneous monitoring of genome-wide, infection-linked transcriptional alterations of the host tissue and colonizing pathogens. Here, we provide a detailed protocol for the preparation and analysis of lymphatic tissue infected with the mainly extracellularly growing pathogen *Yersinia pseudotuberculosis*. This method can be used as a powerful tool for the discovery of *Yersinia*-induced host responses, colonization and persistence strategies of the pathogen, and underlying regulatory processes. Furthermore, we describe computational methods with which we analyzed obtained datasets.

1 Introduction

The interplay between enteric pathogens, such as the enteropathogenic *Yersinia* species *Y. enterocolitica* and *Y. pseudotuberculosis*, and their host is a very complex and highly dynamic process. Both interacting organisms initiate intricate programs to adapt rapidly to the new situation and ensure their survival. For instance, enteric *Yersiniae* express multiple adhesion and invasion factors to colonize the intestinal tract. After the passage of the intestinal epithelial layer and entry into the underlying lymphatic tissues, they initiate a sophisticated host defense system to prevent their elimination by innate immune cells. Vice versa, the host triggers a complex multi-layered defense program. This includes a very strong inflammatory and acute phase response, an activation of the coagulation cascade for fibrin matrices formation around the bacterial colonies and initiation of a T_H1/T_H17 response to eliminate the bacteria [1]. During infection, however, multiple outcomes are possible: host cell death, bacterial persistence without symptoms, or full bacterial clearance [2,3]. This suggests a heterogeneous programming of host-pathogen interactions during the course of the infection. To obtain a comprehensive understanding of the underlying processes, detailed knowledge of the associated gene expression profiles and their infection-linked changes during the distinct stages of the infection of both the host and the pathogen are required. Here, we describe a recently established probe-independent RNA-sequencing approach termed *Tissue Dual RNA-seq* that allows simultaneous profiling of a pathogen and its infected host to tackle this task [1,3].

This very sensitive approach allowed us:

(a) to reveal host immune responses that are triggered during the acute phase leading either to death of mice, or to clearance of the bacteria, or to the initiation of the persistence stage;

(b) to discover virulence-relevant bacterial functions that are crucial for the survival of the pathogen in different host niches or during persistency that were not previously reported in the context of a *Yersinia* infection;

(c) to identify regulatory factors as well as novel regulatory mechanisms and processes that drive different pathogenicity programs (early/late acute or persistent phase).

To this end, *Yersinia*-infected lymphatic tissue of the ileum (Peyer's patches) was lysed, total RNA of the bacterial and host cells was isolated and the ribosomal RNA of this RNA mixture was depleted. Subsequently, remaining RNA pools were converted into cDNA libraries and subjected to Illumina sequencing (**Figure 1**). Sequencing reads were consecutively mapped to the *Yersinia* genome and the virulence plasmid, and the host core genome. Data normalization and analysis of the aligned reads allowed us to obtain highly accurate and sensitive transcriptional profiles of both the host and the pathogen from the data subsets. Moreover, a detailed pathway analysis combined with a differential expression analysis between *in vitro* grown bacterial cultures or uninfected mice has permitted identification of differentially expressed (a) bacterial and host genes during different stages of infection (acute versus chronic) or between wild-type or mutant infected mice (*Yersinia* wildtype versus *cnfY* mutant), and (b) numerous non-coding RNAs of the pathogen [1,3]. The following detailed Tissue Dual RNA-seq protocol was established for the detection of bacterial (*Y. pseudotuberculosis* IP32953) and host (BALB/c mice) RNAs expressed in the Peyer's patches three days post infection (**Figure 1**).

2 Materials

2.1 Bacterial Cultures

1. *Yersinia pseudotuberculosis* IP32953 [4]
2. 100 ml Erlenmeyer flasks
3. Shaker for Erlenmeyer flasks and incubator for bacterial plates
4. Luria Bertani Broth (LB): 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 85.6 mM sodium chloride; LB plates: LB medium (see above), 1.8% (w/v) agar

2.2 Mouse Infection Model

1. Six groups of five weight-matched 7-8 weeks old female BALB/c mice
2. Individually ventilated cages on a 12 h light:dark cycle under specific pathogen-free conditions according to the FELASA recommendations
3. Falcon tubes, Eppendorf tubes, pipets
4. Spectrophotometer or cell density meter, and cuvettes
5. Benchtop microcentrifuge
6. Phosphate-buffered saline (PBS)
7. Vortex
8. Gavage needle

2.3 Tissue Preparation

1. Tissue preparation tools (scalpel, scissor, tweezer etc.)
2. Sterile **homogenization** tubes (12 or 15 ml)
3. Liquid nitrogen
4. Pre-cooled Solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7, 0.5% (w/v) N-lauroylsarcosine, 0.1 M 2-mercaptoethanol)
5. Tissue homogenizer (Polytron PT 2100 homogenizer, Kinematica, Switzerland)

6. 2 ml safe-lock tubes
7. 100 µm and 600 µm **acid washed glass beads** (Sigma Aldrich)
8. TissueLyser (Qiagen)
9. 70% ethanol (EtOH)
10. Phosphate-buffered saline (PBS)
11. Benchtop microcentrifuge

2.4 RNA Isolation, Stabilization and rRNA Depletion

1. 1.5 and 2 ml safe-lock tubes
2. Resuspension buffer: 0.01 M sodium acetate (NaOAc) pH 4.5, 0.3 M saccharose
3. Lysis buffer: 0.01 M NaOAc pH 4.5, 2% SDS
4. Water-saturated phenol, pH 4.5-5.0
5. Water bath or heat block
6. Liquid nitrogen
7. Benchtop microcentrifuge (refridgerated)
8. **Vortex**
9. **3 M NaOAc pH 4.5**
10. **100%** pure EtOH
11. **70% EtOH, ice-cold**
12. **Nuclease-free water**
13. **TURBO DNase I (Thermo Fisher)**
14. **Phenol:chloroform:isoamyl alcohol (25:24:1) and chloroform:isoamyl alcohol (24:1)**
15. **Agilent RNA 6000 Nano Kit**
16. **Agilent 2100 Bioanalyzer**

17. Ribo-Zero Gold rRNA Removal Kit Human/Rat/Mouse (Illumina)
18. MICROBExpress Bacterial mRNA Enrichment Kit (Thermo Fisher).
19. TE buffer, pH 8.0
20. ERCC ExFold RNA Spike-In Mix 1 or Mix 2 (Thermo Fisher)

2.5 Library Preparation and Illumina Sequencing

1. Sonicator (Covaris Adaptive Focused Acoustics device, Covaris)
2. RNeasy MinElute Cleanup Kit (Qiagen)
3. Agilent RNA 6000 Nano Kit or Pico Kit, High Sensitivity DNA Kit
4. Agilent 2100 Bioanalyzer,
5. T4 polynucleotide kinase (Thermo Fisher)
6. 3'- and 5'-RNA-adaptor oligonucleotides (for adapter sequences see **Table 1**, [5])
7. T4 RNA ligase (Thermo Fisher)
8. SuperScript III Reverse Transcriptase (Invitrogen)
9. DNA primer complementary to the 3'-adapter sequence to generate first strand cDNA (for primers see **Table 1**, [5])
10. Phusion® high fidelity DNA polymerase (NEB)
11. PCR primers which are identical to corresponding Illumina primers to ensure full compatibility with GenomeAnalyzer flowcells (**Table 1**, [5]).
12. QIAquick Gel Extraction Kit (Qiagen) or BluePippin (Biozym)
13. Capillary electrophoresis (Shimadzu MultiNA microchip electrophoresis system)
14. HiSeq2500 (Illumina)

2.6 Bioinformatics tools

1. Genome Analyzer Pipeline Analysis software 1.8.2 (Illumina)

2. Fastq-mcf and fastq-multxtool of ea-utils [6].
(http://hannonlab.cshl.edu/fastx_toolkit/),
3. *fastx_trimmer* from the FASTX-toolkit version 0.0.13
4. *FastQC* program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)
5. *Bowtie2* (version 2.1.0) [7]
6. *TopHat2* [8]
7. *SAMtools* [9]
8. *htseq-count* program [10]
9. *DESeq2* [11], see **Note 1**
10. KEGG database ([12])
11. R package GOstats [13]
12. Signaling Pathway Impact Analysis (SPIA) algorithm [14]
13. ERCC RNA-Spike-In Control Mixes User Guide:
https://tools.thermofisher.com/content/sfs/manuals/cms_086340.pdf
14. TSSAR [15]
15. WebLogo software [16]
16. MEME software [17]
17. Integrated Genome Browser [18]

3 Methods

3.1 Mouse Infection with *Y. pseudotuberculosis* and non-infected control mice

1. Prior to infection, starve 2 x 3 groups of five weight-matched BALB/c female mice aged between 7-8 weeks for several hours.

2. In parallel, grow *Y. pseudotuberculosis* strain IP32953 over night (OD₆₀₀ approx. 4.0) in 20 ml LB broth in a 100 ml Erlenmeyer flask at 25°C under aeration (2.24 x g) in a shaker.
3. Transfer a volume of 10 ml of the culture to a 50 ml reaction tube, fill up with sterile PBS to 50 ml and pellet the bacteria (2,773 x g) for 10 min at 25°C.
4. Remove the supernatant completely and wash the pellet again in 50 ml PBS.
5. Resuspend the bacteria in 10 ml PBS, determine the optical density and adjust the cells with PBS to 1x10⁹ CFU/ml (OD₆₀₀ = 1 corresponds to ca. 1x10⁹ IP32953 CFU/ml).
6. Infect three groups of five mice intragastrically using a gavage needle with approximately 2 x 10⁸ CFUs of *Y. pseudotuberculosis* IP32953 (200 µl/mouse), **see Note 2**. For uninfected controls, treat three groups in parallel with PBS only.
7. Plate serial dilutions of the inoculum in triplicate to assess the actual CFU used for the infection.

3.2 Harvest of Infected and Non-Infected Lymphatic Tissue

1. Sacrifice mice three days post infection, dissect, isolate and transfer Peyer's patches of the infected and non-infected groups in precooled 2 ml test tubes and immediately snap freeze the samples in liquid nitrogen. Store at - 80°C or directly proceed with homogenization.

3.3 Tissue (Peyer's Patches) Homogenization

1. Cool homogenization tubes on ice.
2. Transfer the organ/tissue (Peyer's patches) from -80°C freezer on ice.
3. Add 1.5 ml solution D to the organ and transfer to the homogenization tube.

4. Homogenize the tissue on ice at **1,220 x g** for 10 sec. Wash the homogenizer before and in between **tissues** with 4 ml **cold 70% EtOH** and 4 ml **PBS**.
5. Transfer the tissue lysate to a 2 ml reaction tube.
6. In case of foamy samples, centrifuge the tissue lysate at **17,949 x g** at 4°C for 1 min.
7. Aliquot the lysate/supernatant in 2 x 0.75 ml to pre-cooled 2 ml safe-lock tubes containing 200 mg small (100 µm) and 200 mg big (600 µm) acid washed glass beads (i.e. 2 tubes for each sample)
8. Treat the sample with TissueLyser at fixed speed for 1 min (30 Hz), see **Note 3**.
9. Cool on ice for 1 minute
10. Repeat step 8 and 9 two times.

3.4 Total RNA Isolation of Bacteria and Host Cells from Infected Tissue

1. Transfer 0.5 ml of the tissue lysate (without glass beads) to a new 2 ml reaction tube.
2. First, add 50 µl of 3 M sodium acetate pH 4.5 and mix the solution by inversion
3. Add 500 µl water-saturated phenol and mix thoroughly by inversion.
4. Add 100 µl chloroform:isoamyl alcohol (24:1) and **vortex** the samples for 10 sec, cool the samples on ice for 15 min.
5. **Centrifuge the samples** at **17,949 x g** for 20 min at 4°C.
6. Transfer the **top layer** to a new reaction tube and add 300 µl of chloroform:isoamylalcohol (24:1). Vortex the mixture for 30 sec and **centrifuge** the samples for 3 min at **17,949 x g** at 4°C.
7. Repeat **steps 4-5**.
8. Transfer the supernatant to a new 1.5 ml reaction tube and add 0.1 volumes of 3 M sodium acetate pH 4.5 and 2.5 volumes of 100% **ice-cold** EtOH. Invert the

tubes several times and incubate the samples for at least 1 h at -20°C to precipitate the RNA.

9. Pellet the RNA by centrifugation for 30 min at 17,949 x g at 4°C.
10. Discard the supernatants and wash the pellets with 500 µl of 70% ice-cold EtOH.
11. Pellet the samples by centrifugation for 5 min at 17,949 x g at 4°C.
12. Remove the supernatants and air-dry the pellets for 5-20 min at room temperature. Resuspend the pellets in 50 µl nuclease-free water.
13. Assess the final RNA concentration and RNA integrity with the Agilent RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer as described by the manufacturer.
14. Digest the residual DNA using TURBO DNase I (see 3.3.4) and analyze the RNA integrity again on the Agilent 2100 Bioanalyzer (**Figure 2**), **see Note 4**.
15. **rRNA Depletion and Spiking of the RNA of the Tissue Samples.**
 - 15.1 Combine high quality RNA (RIN ≥ 9) from Peyer's Patches of 5 mice to one RNA pool in equal amounts.
 - 15.2 To reduce sequencing depth requirements, deplete mouse rRNA from 3 independent RNA pools using the Ribo-Zero Gold rRNA Removal Kit and purify the RNA by EtOH precipitation according to the instructions of the manufacturer.
 - 15.3 For subsequent fragmentation, **resuspend 800 ng RNA in TE buffer to reach the final volume of 130 µl (pH 8.0)**. Analyze the RNA integrity again on the Agilent 2100 Bioanalyzer (**Figure 2**).
16. Add 1 µl of either 1:10 diluted ERCC ExFold RNA Spike-In Mix 1 or Mix 2 to 1 µg of mouse rRNA depleted RNA.

3.5 Total RNA Isolation from bacterial cultures *in vitro*

1. Grow *Y. pseudotuberculosis* IP32953 in 20 ml LB broth in a 100 ml Erlenmeyer flask under aeration (2.24 x g) in a shaker in 3 separate cultures to exponential phase (OD₆₀₀ 0.5) or stationary phase (OD₆₀₀ approx. 4.0) at 25°C or 37°C.
2. Pellet the culture by centrifugation (2,773 x g for 3 min at 25°C or 37°C) and use 5x10⁹ bacteria for total RNA preparation per growth condition and sample. Extract total RNA from bacterial pellets by the hot phenol extraction protocol described by Sambrook 2001 [19], see **Note 5**.
3. To remove residual genomic and *Yersinia* virulence plasmid DNA remnants, treat the samples with 0.008 U of TURBO DNase I per 100 ng of RNA for 1 h at 37°C in a total volume of 240 µl.
4. Perform a phenol-chloroform purification and EtOH precipitation with the DNA-free total RNA of the samples.
 - 4.1. Add one volume of phenol:chloroform:isoamyl alcohol (25:24:1) to each sample. Mix the tubes by vortexing for 30 s and centrifuge for 3 min at 17,949 x g at 4°C.
 - 4.2. Extract the RNA containing supernatants twice with 1 volume of chloroform:isoamyl alcohol.
 - 4.3. Transfer the aqueous phases into fresh 1.5 ml reaction tubes, supplemented with 0.1 volume of sodium acetate (pH 4.5) and 2.5 volumes of 100% ice-cold EtOH, and incubate the samples overnight at -20°C to precipitate the RNA.
 - 4.4. Pellet precipitated RNA by centrifugation for 30 min at 17,949 x g at 4°C.
 - 4.5. Remove the supernatants from each sample and wash the pellets with 1 ml of 70% ice-cold EtOH.

- 4.6. Centrifuge the samples for 10 min at 17,949 x g, remove the supernatants and air-dry the pellets at room temperature (RT).
- 4.7. Resuspend the pellets in RNase-free water and resolve the RNA for 15 min at room temperature.
- 4.8. Assess the RNA quality on a Agilent 2100 Bioanalyzer.
5. To deplete ribosomal RNA, treat 8 µg of isolated total RNA with the MICROBExpress Bacterial mRNA Enrichment Kit according to the manufacturer's instructions.
6. For subsequent fragmentation, resuspend 800 ng RNA in TE buffer to reach the final volume of 130 µl (pH 8.0).
7. Add 1 µl of either 1:10 diluted ERCC ExFold RNA Spike-In Mix 1 or Mix 2 to 1 µg of rRNA depleted RNA.

3.6 cDNA Library Preparation

modified from Dötsch *et al.* [5], see **Note 6**.

1. For strand-specific RNA-seq cDNA library preparation and barcode introduction, perform a fragmentation of the rRNA-depleted RNA by sonication to a median size of 200 nt.
 - 1.1. Fragment the RNA (from steps 3.4.11 and 3.5.6) to a median size of 200 nt by sonification with the following settings: processing time - 150 sec; fragment size range – 200 bases; intensity - 5; duty cycle - 10%.
 - 1.2. Precipitate the RNA by ethanol precipitation (see manual Ribo-Zero Gold rRNA Removal Kit) and resuspend the pellet in 16 µl nuclease-free water.
 - 1.3. Use 1 µl for the subsequent analysis of quality and size distribution of the fragmented RNA on a Bioanalyzer Pico Chip.

2. To phosphorylate the 5'-ends for adapter ligation and to remove the 3'-phosphates, treat the remaining fragmented RNA with 1 μ l (10 U) T4 polynucleotide kinase for 25 min at 37°C. Perform phenol:chloroform purification and EtOH precipitation as described in step 3.5.4. Resuspend the pellet in 16 μ l nuclease-free water.
3. For the 3' adapter ligation add 5 x concentration of 3' adapter (see Table 1) from a 100 pmol/ μ l adaptor stock solution to RNA (e.g. 10 pmol RNA + 50 pmol adapter) and 1 μ l of T4 RNA ligase.
4. Purify the RNA using the Qiagen RNeasy MinElute Cleanup Kit. Elute the RNA with 16 μ l nuclease-free water.
5. For the 5' adapter ligation add 5 x concentration of the 5' adapters (see Table 1) from a 100 pmol/ μ l adaptor stock solution to RNA (e.g. 10 pmol RNA + 50 pmol adapter) and 1 μ l of T4 RNA ligase.
6. Purify the RNA using the Qiagen RNeasy MinElute Cleanup Kit. Elute the RNA with 16 μ l nuclease-free water.
7. For first strand cDNA synthesis add 1 μ l RT primer (100 μ M) (Table 1) to 4 μ l RNA and perform PCR in thermocycler with the following program: 10 min 70°C, followed by 10 min 25°C. Place samples immediately on ice.
8. Prepare a mastermix for all samples: 2 μ l 5 x first strand buffer; 1 μ l 100 mM DTT; 1 μ l 10 mM dNTP (each); 1 μ l SuperScript III (per sample amounts).
9. Add the 5 μ l of the mix to the 5 μ l RNA/RT primer mix. Perform cDNA synthesis at 50°C for 120 min.
10. To inactivate the reverse transcriptase, incubate the mix for 15 min at 70°C.
11. To amplify the cDNA libraries perform a PCR reaction using Phusion polymerase: 10 μ l of cDNA, 1 x Phusion polymerase HF buffer, 200 μ M dNTPs, 0.5 μ M primer A and primer B and 3% DMSO in a total volume of 50 μ l. For

thermocycling use the following program: 1. 98°C 5 min; 2. 98°C 1 min; 3. 62°C 30 sec; 72°C 30 sec; repeat 2. to 4.: 15 cycles; 72°C 5 min final elongation.

12. Load PCR products on a 2% thin agarose gel and cut out fragments (rather smear) between 150 to 500 bp.
13. Purify DNA using the QIAquick Gel Extraction Kit protocol or by automated DNA Size Selection with the BluePippin. Elute DNA with 30 µl nuclease-free water.
14. Use 1 µl of the samples for a subsequent analysis of quality and size distribution of the PCR products on the Bioanalyzer (High Sensitivity DNA Kit).

3.7 Illumina Sequencing

1. For Illumina HiSeq sequencing, combine all cDNA libraries in approximately equimolar amounts.

1.1 Apply a differential clean-up with Agencourt AMPure kit (Beckman Coulter Genomics) to fractionize cDNA pools in the range of 150-600 bp.

1.2 Analyze cDNA pools by capillary electrophoresis (Shimadzu MultiNA microchip electrophoresis system), see **Note 7**.

2. The Illumina cluster station is used for cluster generation. Single-end sequencing on the HiSeq2500 was performed with 2 nM library denatured with 0,1 N NaOH and diluted with 5 x SSC, 0,05% Tween-20 or HT-1 hybridisation buffer of NextSeq 500 Kit (Illumina) to a final concentration of 12 pM. Cluster generation on HiSeqSR Flow Cell v3 was carried out at cBot using TruSeq SR Cluster Kit v3 - HS to create single molecule DNA templates followed by bridge amplification. Sequencing run was performed at HiSeq 2500 using TruSeq SBS Kit v3 (50 cycle) to run 51 cycles and 7 cycles for the single-indexed read.

Bar-coded 5'-adapters enable the pooling of multiple samples on one lane of the Illumina flow cell. Libraries are sequenced with 36-100 cycles in single-end

mode, which resulted in 36-100 nt long reads including 6 nt of barcode and 30-94 nt of the sequenced 5'-ends of fragments (typically to > 1 million reads/library). Process the fluorescent images to sequences and transform the data to FastQ format using the Genome Analyzer Pipeline Analysis software 1.8.2 (Illumina).

3.8 Read Mapping and Initial Library Quality Assessment

1. Assess all FASTQ files from the sequenced libraries for sufficient read quality and potential contamination using the *FastQC* program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).
2. Remove identified adapter contaminations and remaining artificial sequences (barcodes) from the Illumina reads in FASTQ format using program *fastx_trimmer* from the FASTX-toolkit version 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). At the 3' end, trim reads if the per base Phred score falls short of 20. Discard trimmed reads with a remaining length < 20 nts.

3.9 Bioinformatic Analyses and Statistics

1. Map reads of the libraries to the *Y. pseudotuberculosis* IP32953 genome (NC_006155) and the pYV plasmid (NC_006153.2) using *Bowtie2* (version 2.1.0) [7] in end-to-end alignment mode with default parameterization. Employ *SAMtools* [9] to filter bam files for uniquely mapped reads (both strands). Reads are called uniquely mapped reads with a unique genomic location if and only if they could not be aligned to another location with a higher or same mapping quality. The uniquely mapped reads are classified as reads from *Yersinia*, whereas remaining non-mapping reads are categorized as mouse reads.

Libraries with reads classified from mouse are aligned to the mouse genome (assembly: GRCm38/mm10) using the splice junction mapper *TopHat2* [8] with library type *fr-secondstrand*, see **Note 8**.

2. Potentially existing cross-mapping reads (very rare: < 0.001%, reads mapping to mouse and *Yersinia* for 50 bp single-end reads) identified by aligning the classified *Yersinia* reads back to the mouse genome using *Tophat2* [8] are removed from the libraries and excluded from downstream analyses, see **Note 9**.

3.10 Feature normalization, quantification and detection of differential gene expression

1. Quantify reads that align to annotated genes with program *htseq-count* [10]. For gene expression quantification in mouse use up-to-date gene annotations from the Ensembl database. For *Y. pseudotuberculosis* IP32953 use the custom annotation given in Dataset S6 in Nuss *et al.* [1]. This annotation enhances the NCBI annotation for NC_006155 and NC_006153.2 with 197 non-coding RNAs.
2. Use calibrated ERCC RNA spike-in-controls to estimate expression differences, assess the platform dynamic range and the accuracy of fold-change response [20]. See **Note 10**. Add spike-in control sequences to mouse reference genome/annotation prior to read alignment. Read counts for spike-in controls are determined along with normal gene counts with program *htseq-count*. For diagnostic plots of the dose- and fold-change-response follow the description of the manufacturer (ERCC RNA-Spike-In Control Mixes User Guide: https://tools.thermofisher.com/content/sfs/manuals/cms_086340.pdf).
3. Use Circos plot (**Figure 2B**) to visualize RPKM (Reads Per Kilobase transcript length per Million mapped reads) of normalized expression values of *in vitro* and

in vivo RNA-seq data for the IP32953 genome (NC_006155.1) and pYV virulence plasmid (NC_006153.2). Normalize data according to the number of uniquely mapped reads and illustrate read coverage of the RNA-seq analysis of selected loci with the Integrated Genome Browser [18] (**Figure 2C**).

4. Use with *htseq-count* determined gene read counts as input for *DESeq2* [11] for pairwise detection and quantification of differential gene expression, **see Note 1**. Moreover, compute RPKM (reads per kilobase max. transcript length per million mapped reads) values for each library from the raw gene counts, and filter the list of *DESeq2* generated differentially expressed genes (DEGs). Mouse and *Yersinia* genes are classified as being differentially expressed if and only if $|\log_2FC| \geq 2$ and (multiple testing corrected) $p\text{-value} \leq 0.05$ hold.

3.11 Overrepresentation and pathway analyses

1. Further annotate lists of *Yersinia* DEGs with pathway information from the KEGG database [12], and visualize by heat maps (e.g. differentially regulated antisense RNAs, **Figure 3**) or pie charts (differentially regulated metabolic functions, **Figure 4**).
2. Assess the list of DEGs resulting from comparison of infected mice and uninfected mice with functions from the R package GOstats [13] to obtain the association of Gene Ontology (GO) terms and KEGG metabolic pathways to genes. Use a p-value cut-off of 0.001 for the applied conditional hypergeometric test for overrepresentation of GO terms in the three ontologies: molecular function, biological process and cellular component and the annotated KEGG pathways.

3. Use GO annotations from the Bioconductor *Mus musculus* annotation package, and retrieve KEGG pathway annotations directly from KEGG using KEGG's REST API.
4. Apply the Signaling Pathway Impact Analysis algorithm [14] as implemented in the SPIA Bioconductor package with default parametrization to the list of mouse DEGs to identify host signaling processes influence by the *Yersinia* infection.

4 Notes

1. Alternatively, edgeR can be used [21]. Both tools are available as Bioconductor packages [22]. When using *DESeq2*, use a beta prior and disable the Cook distance cut off filtering.
2. Prior to the experiment we tested different infection conditions of the *Yersinia pseudotuberculosis* strain (e.g. varying infection doses, days of infection, preparation and testing of the cfu in the Peyer's patches or other tissues and organs). For the tissue-dual RNA-seq experiments we used the infection condition in which we observed the highest number of bacteria in the Peyer's patches. We further found that genome full-coverage is obtained when $\geq 10^6$ cfu/g tissue are detectable. However, continuous development of new deep sequencing technologies and apparatus (e.g. NOVAseq, Illumina) will certainly improve sequencing depth and might allow us to measure the abundance of rare transcripts across kingdoms in the near future.
3. We used the TissueLyser (Qiagen), but other mechanical lysis methods (e.g. using bead beating) or certain other lysis buffer systems (*mirVanaKit*, Thermo Fisher) could also be used for tissue lysis, but they need to be optimized for the individual *Yersinia* strain.
4. Proceed with the following steps only if the RNA quality is high (RIN ≥ 9).

5. It is extremely important to test different experimental set-ups for RNA extraction. We note that the hot phenol extraction protocol worked best for us without addition of RNA stabilizer. In particular when tissue samples are immediately processed, addition of RNA stabilizer seemed expendable. Nonetheless, different RNA stabilization agents were tested, e.g. RNA*later* and RNA*protect* (Qiagen). *Yersinia pseudotuberculosis* growth was immediately stopped in RNA*later*, but the bacteria survived the procedure and started to regrow when the sample was diluted or the buffer was exchanged. Moreover, the overall yield of the RNA was reduced and it is also important to know that it quenched fluorescent signals of Gfp-expressing bacteria in case a sorting process is planned to isolate bacteria from the tissue. Compared to RNA*later*, RNA*protect* was less efficient in *Yersinia* RNA stabilization.
6. We like to note that our protocol for cDNA library was established for sequencing on the HiSeq2500 (Illumina), but currently new technologies, i.e. NOVAseq are used and protocols need to be optimized accordingly.
7. Proceed with Illumina HiSeq sequencing only if the fragments are distributed in the range of 150-600 bps.
8. Alternatively other splice aware alignment programs like e.g. STAR [23] can be used here.
9. Three different categories of mouse and *Yersinia* read sets are classified based on their origin: the infected (IP) or uninfected (PBS) mouse read set, and the *Yersinia* read set from the infected mouse samples. Moreover, four *Yersinia* read sets are generated from the *in vitro* grown pools as a reference (25°C and 37°C exponential phase, 25°C and 37°C stationary phase). The resulting bam files of these read sets constitute the basis for quantification, comparison and visualization.

10. Spike-ins should be added as early as possible to avoid biases in down-stream processing steps.
11. We successively aligned obtained reads to the *Yersinia* and then to the mouse genome. However, also parallel alignment to respective reference genomes is possible, and other pipelines can be used [24]. Moreover, reads cross-mapping to both, the pathogen and the host genome were identified and discarded although the overall proportion was very small to optimize differential expression analysis.

Acknowledgements

We are grateful to M. Fenner for discussions and Robert Geffers and Michael Jarek from the Department of Genome Analytics for Illumina sequencing. This work was supported from grants of the German Research Foundation (DE616/4, DE616/6, SPP1617-young investigator start up funding for A.M. Nuss), and a stipend of the Helmholtz Center for Infection Research Graduate School for M. Kusmierek. P. Dersch is supported by the German Center for Infection Research.

References

1. Nuss AM, Beckstette M, Pimenova M, Schmühl C, Opitz W, Pisano F, Heroven A, Dersch P (2017) Tissue dual RNA-seq: a fast discovery path for infection-specific functions and riboregulators shaping host-pathogen transcriptomes. Proc Natl Acad Sci U S A 114(5) E791-E800
2. Avican K, Fahlgren A, Huss M, Heroven AK, Beckstette M, Dersch P, Fallman M (2015) Reprogramming of *Yersinia* from virulent to persistent mode revealed by

- complex *in vivo* RNA-seq analysis. PLoS Pathog 11 (1):e1004600.
doi:10.1371/journal.ppat.1004600
3. Heine W, Beckstette M, Heroven AK, Thiemann S, Heise U, Nuss AM, Pisano F, Strowig T, Dersch P (2018) Loss of CNFY toxin-induced inflammation drives *Yersinia pseudotuberculosis* into persistency. PLoS Pathog 14 (2):e1006858.
doi:10.1371/journal.ppat.1006858
 4. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M, Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise A, Hauser LJ, Garcia E (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. Proc Natl Acad Sci U S A 101 (38):13826-13831
 5. Dotsch A, Eckweiler D, Schniederjans M, Zimmermann A, Jensen V, Scharfe M, Geffers R, Haussler S (2012) The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. PLoS One 7 (2):e31092. doi:10.1371/journal.pone.0031092
 6. Aronesty E (2011) ea-utils: Command-line tools for processing biological data.
 7. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9 (4):357-359. doi:10.1038/nmeth.1923
 8. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14 (4):R36. doi:10.1186/gb-2013-14-4-r36
 9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25 (16):2078-2079.
doi:10.1093/bioinformatics/btp352

10. Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31 (2):166-169. doi:10.1093/bioinformatics/btu638
11. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15 (12):550. doi:10.1186/s13059-014-0550-8
12. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42 (Database issue):D199-205. doi:10.1093/nar/gkt1076
13. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23 (2):257-258. doi:10.1093/bioinformatics/btl567
14. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, Kim CJ, Kusanovic JP, Romero R (2009) A novel signaling pathway impact analysis. *Bioinformatics* 25 (1):75-82. doi:10.1093/bioinformatics/btn577
15. Amman F, Wolfinger MT, Lorenz R, Hofacker IL, Stadler PF, Findeiss S (2014) TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics* 15:89. doi:10.1186/1471-2105-15-89
16. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14 (6):1188-1190. doi:10.1101/gr.849004
17. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28-36
18. Nicol JW, Helt GA, Blanchard SG, Jr., Raja A, Loraine AE (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25 (20):2730-2731. doi:10.1093/bioinformatics/btp472
19. Sambrook J (2001) *Molecular Cloning: A Laboratory Manual*, . Cold Spring Harbor Laboratories, Cold Spring Harbor, NY,

20. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 21 (9):1543-1551. doi:10.1101/gr.121095.111
21. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1):139-140. doi:10.1093/bioinformatics/btp616
22. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles AK, Pages H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12 (2):115-121. doi:10.1038/nmeth.3252
23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29 (1):15-21. doi:10.1093/bioinformatics/bts635
24. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Muller L, Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529 (7587):496-501. doi:10.1038/nature16547

Table 1: Oligonucleotides used for RNA sequencing**RNA adapters**

Name	sequence (5' > 3')
Adapter-A-BC6-1	UACACGACGCUCUUCCGAUCUACGAGA
Adapter-A-BC6-2	UACACGACGCUCUUCCGAUCUACUUGC
Adapter-A-BC6-3	UACACGACGCUCUUCCGAUCUAUCUGG
Adapter-A-BC6-4	UACACGACGCUCUUCCGAUCUCAUACG
Adapter-A-BC6-5	UACACGACGCUCUUCCGAUCUCGUCAU
Adapter-A-BC6-6	UACACGACGCUCUUCCGAUCUGGUUAC
Adapter-A-BC6-7	UACACGACGCUCUUCCGAUCUGUACCU
Adapter-A-BC6-8	UACACGACGCUCUUCCGAUCUGUCAAG
Adapter-A-BC6-9	UACACGACGCUCUUCCGAUCUUUCUCUG
Adapter-A-BC6-10	UACACGACGCUCUUCCGAUCUUUGGUUC
Adapter-A-BC6-11	UACACGACGCUCUUCCGAUCUUUCCGA
Adapter-A-BC6-12	UACACGACGCUCUUCCGAUCUUUGCAG
Adapter-B ^a	P-AGAUCGGAAGAGCGGUUCAGC-ddC

DNA primer

Name	sequence (5' > 3')
RT-primer	GCTGAACCGCTCTTCCGATCT
PCR-primer-A ^b	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCT
PCR-primer-B ^b	CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTC CTGCTGAACCGCTCTTCCGATCT

Figure legends

Fig. 1. Tissue dual RNA-seq workflow. Five female BALB/c mice are orally infected with 2×10^8 CFUs of *Y. pseudotuberculosis* IP32953 (infected) or 1 x PBS (uninfected) for *Tissue Dual RNA-seq*. For *in vitro* RNA-seq IP32953 is grown in LB to exponential or stationary phase at 25°C or 37°C. Total RNA is isolated from approximately 5×10^9 cfu of *in vitro* grown *Y. pseudotuberculosis* IP32953. In parallel, total RNA is isolated from all isolated Peyer's patches of the infected and uninfected mice. The RNA is further processed for preparation of strand-specific barcoded cDNA libraries and sequenced. cDNA reads were separated *in silico* by mapping to the mouse genome mm10 and the *Y. pseudotuberculosis* IP32953 genome and the *Yersinia* virulence plasmid.

Fig. 2. Global reports of the *Tissue Dual RNA-seq* analysis.

(A) Representative Bioanalyzer profile of total RNA isolated from infected murine Peyer's patches before (left panel) and after (right panel) rRNA depletion. (B) Circos plot visualizing RPKM (Reads Per Kilobase transcript length per Million mapped reads) normalized expression values of *in vitro* and *in vivo* RNA-seq data for the IP32953 genome (NC_006155.1) and pYV virulence plasmid (NC_006153.2). (C) Read coverage of the RNA-seq analysis of the *cscR-icdA* locus is illustrated in the Integrated Genome Browser [18]. The data are normalized according to the number of uniquely mapped reads. E25: exponential phase 25°C; E37: exponential phase 37°C; S25: stationary phase 25°C; S37: stationary phase 37°C; *in vivo*: infected Peyer's patches.

Fig. 3. Comparison of regulatory/antisense RNAs of *Y. pseudotuberculosis* at different environmental conditions and during infection.

Heatmap of the top enriched and depleted regulatory/antisense RNAs of the *in vivo* RNA-seq data from three independent RNA pools of the *Y. pseudotuberculosis* IP32953 core genome and the pYV virulence plasmid at different growth conditions. Color-coding is based on rlog-transformed read count values; enriched (red) or depleted (blue) during infection. E25: exponential phase 25°C; E37: exponential phase 37°C; S25: stationary phase 25°C; S37: stationary phase 37°C; *in vivo*: infected Peyer's patches.

Fig. 4. Bacterial global gene expression analysis uncovered infection-specific metabolic changes.

Pie charts of metabolic functions, which are at least fourfold induced or repressed (p-value 0.05) obtained by comparative RNA-seq using DESeq2 from triplicate experiments (*in vivo* versus all four analyzed *in vitro* conditions).