PLOS | PATHOGENS

REVIEW

Resolving host–pathogen interactions by dual RNA-seq

Alexander J. Westermann[1☯], Lars Barquist[1☯], Jörg Vogel[1,2*]

1 RNA Biology Group, Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany,
2 Helmholtz Institute for RNA-based Infection Research (HIRI), Würzburg, Germany

☯ These authors contributed equally to this work.
* joerg.vogel@uni-wuerzburg.de

Abstract

The transcriptome is a powerful proxy for the physiological state of a cell, healthy or diseased. As a result, transcriptome analysis has become a key tool in understanding the molecular changes that accompany bacterial infections of eukaryotic cells. Until recently, such transcriptomic studies have been technically limited to analyzing mRNA expression changes in either the bacterial pathogen or the infected eukaryotic host cell. However, the increasing sensitivity of high-throughput RNA sequencing now enables "dual RNA-seq" studies, simultaneously capturing all classes of coding and noncoding transcripts in both the pathogen and the host. In the five years since the concept of dual RNA-seq was introduced, the technique has been applied to a range of infection models. This has not only led to a better understanding of the physiological changes in pathogen and host during the course of an infection but has also revealed hidden molecular phenotypes of virulence-associated small noncoding RNAs that were not visible in standard infection assays. Here, we use the knowledge gained from these recent studies to suggest experimental and computational guidelines for the design of future dual RNA-seq studies. We conclude this review by discussing prospective applications of the technique.

Check for updates

OPEN ACCESS

Citation: Westermann AJ, Barquist L, Vogel J (2017) Resolving host–pathogen interactions by dual RNA-seq. PLoS Pathog 13(2): e1006033. doi:10.1371/journal.ppat.1006033

Editor: James B. Bliska, Stony Brook University, UNITED STATES

Published: February 16, 2017

Copyright: © 2017 Westermann et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: LB was supported by a Research Fellowship from the Alexander von Humboldt Stiftung/Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The application of high-throughput sequencing–based transcriptomic technologies has delivered major advances in our understanding of biological processes in essentially every organism analyzed [1]. The high resolution of RNA-seq down to the single nucleotide level, however, also allows for a parallel analysis of different organisms interacting with each other—for example, during infection processes (Fig 1A). Simultaneous RNA-seq of host–pathogen models was initiated in the fields of viral [2,3], fungal [4], and parasite infection [5–7], in which the transcriptome structure of the pathogen resembles that of its host. In contrast, bacterial transcriptomes differ dramatically from their eukaryotic counterparts in terms of both the quantity and composition of their RNA (summarized in [8]), which necessitated the use of dedicated protocols to capture bacterial or eukaryotic transcriptomes in isolation. Typically, to profile bacterial gene expression during infection, the overwhelming host material was depleted prior to analysis (Fig 1B). Consequently, until recently transcriptome analyses of bacterial infections were

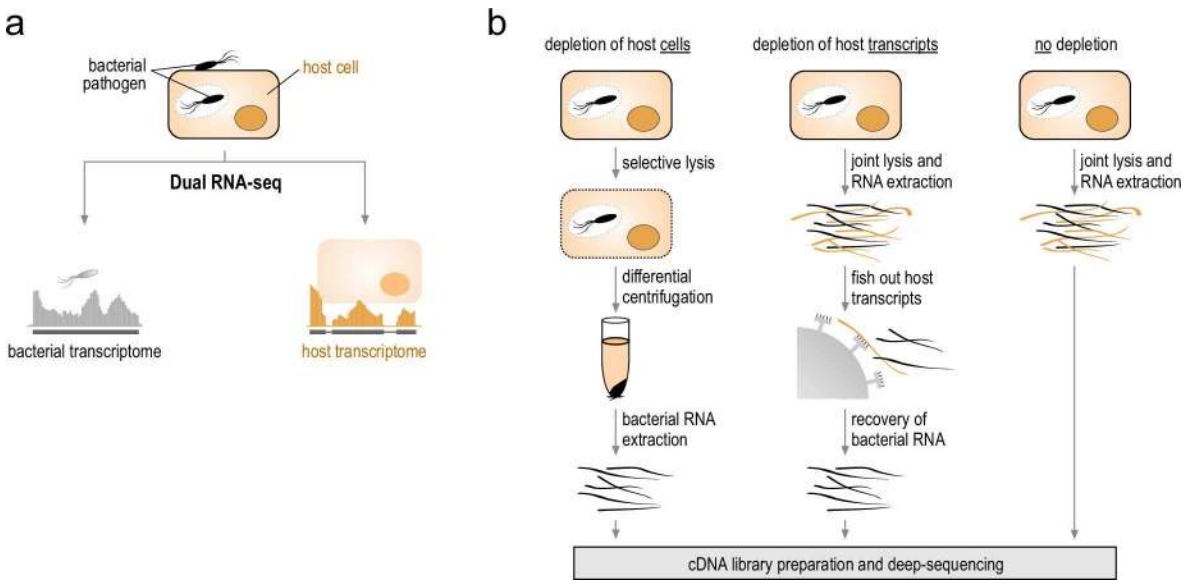

Fig 1. Methods for RNA sequencing of bacterial infections. A. Concept of dual RNA-seq. Total RNA is extracted from infected cells and analyzed by RNA-seq. The mixed sequencing reads are assigned to their originating genomes in silico. **B**. Different approaches to quantify gene expression of bacteria in context with mammalian host cells. Traditionally, host material was depleted prior to analysis, either by detergent-mediated lysis of host cells (left) or by sequence-specific removal of host transcripts (middle). Instead, dual RNA-seq omits host depletion (right) and analyzes pathogen and host gene expression in parallel.

doi:10.1371/journal.ppat.1006033.g001

necessarily one-sided, limiting our ability to understand the interactions between pathogen and host.

Five years ago, we coined the term "dual RNA-seq" to refer to the simultaneous RNA-seq analysis of a bacterial pathogen and its infected host (Fig 1A) and theoretically evaluated its feasibility [8]. The key technical issue we identified was the different nature and content of RNA between bacterial and eukaryotic cells. For example, a typical mammalian cell contains on the order of 20 picograms of RNA, which is roughly two orders of magnitude more than a single bacterial cell [9]. Accounting for the prevalence of rRNA transcripts and variable infection rates, this would leave a minute fraction of informative bacterial transcripts in a mixed RNA pool, compromising accurate quantification. This hurdle has now been overcome in a variety of ways (Table 1): by sequencing cDNA libraries to high depth [10], by partially enriching bacterial transcripts prior to sequencing [10,11], by enriching for invaded host cells by fluorescence-activated cell sorting (FACS) [12,13] or laser capture microdissection [14], by depleting rRNA of the bacterium and host either in series or in parallel [10–13,15,16], and by combinations thereof. As a result, most of the current dual RNA-seq protocols [12–15] can provide informative data with as few as ~25 million reads per sample of mixed pathogen–host RNA, making them practical on current sequencing platforms. Importantly, dual RNA-seq of total mixed RNA following double rRNA depletion (see Fig 2) has now become an affordable, straightforward approach that can be generically applied to any bacterial infection model [13].

This newfound feasibility has led to a variety of emerging applications of dual RNA-seq to bacterial infection models, including the direct correlation of bacterial gene activity with a specific host response and the identification of "molecular phenotypes" of pathogen genes that are invisible in standard virulence assays [13]. Here, we update our earlier theoretical considerations [8] based on the biological insights gained from recent dual RNA-seq studies of diverse

Table 1. Overview of dual RNA-seq and related studies published to date. "Dual SAGE" refers to the simultaneous analysis of host and pathogen by Serial Analysis of Gene Expression (SAGE), and "Multi RNA-seq" refers to a metatranscriptomic analysis of bacterial species constituting the airway microbiota in conjunction with nasal epithelial host cells. "M," million; "TPM," transcripts per million; "RPKMO," reads per kilobase pairs of a gene per million reads aligning to annotated ORFs. Databases containing raw sequencing data: NCBI (National Center for Biotechnology Information), ENA (European Nucleotide Archive), GEO (Gene Expression Omnibus).

	One-sided (focus on bacterial gene expression)					Dual RNA-seq								Dual SAGE	"Multi" RNA-seq
	Mandlik et al. [17]	**Lamont et al. [18]**	**Szafranska et al. [19]**	**Srikumar et al. [20]**	**Avican et al. [21]**	**Humphrys et al. [10]**	**Vannucci et al. [14]**	**Mavromatis et al. [11]**	**Rienksma et al. [15]**	**Baddal et al. [16]**	**Avraham et al. [12]**	**Westermann et al. [13]**	**Aprianto et al. [22]**	**Afonso-Grunz et al. [23]**	**Pérez-Losada et al. [24]**
Bacterial species	*Vibrio cholerae*	*Mycobacterium avium subsp. paratuberculosis*	*Staphylococcus aureus*	*Salmonella* Typhimurium	*Yersinia pseudotuberculosis*	*Chlamydia trachomatis* serovar E	*Lawsonia intracellularis*	uropathogenic *Escherichia coli* (UPEC)	*Mycobacterium bovis* Bacillus Calmette–Guérin	nontypeable *Haemophilus influenzae*	*Salmonella* Typhimurium	*Salmonella* Typhimurium	*Streptococcus pneumoniae*	*Salmonella* Typhimurium	airway microbiota
Host model	infant rabbits and mice	bovine epithelial cells (MAC-T); primary bovine monocyte–derived macrophages	mouse model of osteomyelitis	murine monocytic cells (RAW 264.7)	FVB/N mice	human epithelial cells (HEp-2)	primary porcine enterocytes	mouse bone marrow–derived macrophages	human monocytic cells (THP-1)	primary normal human bronchial epithelial cells	mouse bone marrow–derived macrophages	diverse cell culture models (human, murine, porcine)	human lung alveolar epithelial cells (A549)	human epithelial cells (HeLa)	nasal epithelium from human donors
Intracellular/ extracellular	extracellular	intracellular	extracellular	intracellular	extracellular	obligate intracellular	obligate intracellular	intracellular	intracellular	extracellular	intracellular	intracellular	extracellular	intracellular	extracellular
Sample fixation?	RNA*later*	-	murine tibiae incubated in RNA*later*	RNA stabilization solution (0.2% SDS/19% ethanol/1% acidic phenol)	cryosections incubated in RNA*later*	-	treated with RNase inhibitor prior to embedding	-	-	-	-	RNA*later*	saturated ammonium sulfate solution	-	-
Enrichment of invaded cells?	n.a.	n.a.	n.a.	n.a.	n.a.	-	laser capture microdissection	-	-	n.a.	FACS-based (upon lipopolysaccharide [LPS] staining)	FACS-based (green fluorescent protein [GFP]-expressing bacteria)	n.a.	-	-
Lysis technique	tissue homogenized; cells lysed in TRIzol or lysis/ binding buffer (*mir*Vana kit)	TRIzol + zirconium beads	lysostaphin treatment followed by the addition of RLT buffer and mechanic disruption	0.2% SDS/ 19% ethanol/ 1% acidic phenol to lyse host cells; TRIzol-based bacterial lysis	Dispomix Drive; glass beads	freeze-thaw + Lysis Solution (MasterPure RNA Purification kit)	Extraction Buffer (PicoPure kit)	Buffer RLT (RN*easy* kit)	TRIzol + bead beating	TRIzol	freeze-thaw	lysis/binding buffer (*mir*Vana kit)	bead beating and phenol-chloroform	lysis/binding buffer (*mir*Vana kit)	TRIzol
RNA extraction technique	TRIzol or *mir*Vana PARIS	TRIzol	RN*easy*	TRIzol	hot phenol	MasterPure RNA Purification	PicoPure	RN*easy*	TRIzol	TRIzol	RNAClean SPRI beads	*mir*Vana	High Pure RNA Isolation	*mir*Vana	TRIzol
Enrichment of bacterial cells/ transcripts?	MICROB*Enrich*	MICROB*Enrich* + MassageAmpII	anti-*S. aureus* immunoglobulin G (IgG) antibodies coupled to magnetic beads	selective lysis and differential centrifugation	MICROB*Enrich*	with or without polyA-depletion (Poly (A) Purist Mag) to enrich bacterial transcripts; re-combined both RNA samples prior to sequencing	-	MICROB*Enrich*	with or without differential lysis (with guanidine thiocyanate)	-	-	-	-	with or without polyA-enrichment (Dynabeads Oligo dT_{25}); $poly(A)^+$ and $poly(A)^-$ samples analyzed separately	-
rRNA depletion?	MICROB*Express*	-	Terminator Exonuclease	-	MICROB*Express*	RiboZero (gram-negative bacteria; human/mouse/ rat)	-	RiboZero (gram-negative bacteria; human/mouse/ rat)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (epidemiology)	RiboZero (gram-positive bacteria; human/mouse/ rat)	RiboZero (gram-negative bacteria; human/ mouse/rat)	RiboZero
cDNA library preparation	Illumina: strand-specific ds-cDNA; Helicos: ss-cDNA	mRNA Seq library preparation kit (Illumina)	ScriptSeq	Illumina-based protocol	TruSeq	TruSeq	Ovation RNA-Seq System V2	Digital Gene Expression Tag Profiling kit	TruSeq	ScriptSeq	RNAtag protocol (generation of multiple RNA-seq libraries in a single reaction)	Illumina-based protocol	TruSeq	SuperSAGE libraries for $poly(A)^+$ and $poly(A)^-$ fractions	TruSeq
Sequencing platform	Illumina (paired-end), Helicos	GA IIx (paired-end)	HiSeq 2500 (single-end)	HiSeq 2000	HiSeq 2000 (paired-end)	HiSeq 2000 (paired-end)	GA IIx (paired-end)	HiSeq 2000 (paired-end)	HiSeq 1500 (paired-end)	HiSeq 2500 (paired-end)	HiSeq 2500	HiSeq; NextSeq 500 (single-end)	NextSeq 500 (single-end)	HiSeq 2000 (single-end)	HiSeq 2500 (single-end)
Sequencing depth/library	~50 M (Illumina); 1–5 M (Helicos)	20 M for bovine samples; 7.5 M for bacterial samples	53–105 M	~20 M	~20–250 M	~14–353 M	~22 M	~15–30 M	~22–40 M (for infection samples)	~60–180 M	on average 6 M	varies (~25 M for main time-course experiment)	on average 70 M	~1–7.5 M	~40 M

(*Continued*)

Table 1. (*Continued*)

	One-sided (focus on bacterial gene expression)					Dual RNA-seq								Dual SAGE	"Multi" RNA-seq
	Mandlik et al. [17]	**Lamont et al. [18]**	**Szafranska et al. [19]**	**Srikumar et al. [20]**	**Avican et al. [21]**	**Humphrys et al. [10]**	**Vannucci et al. [14]**	**Mavromatis et al. [11]**	**Rienksma et al. [15]**	**Baddal et al. [16]**	**Avraham et al. [12]**	**Westermann et al. [13]**	**Aprianto et al. [22]**	**Afonso-Grunz et al. [23]**	**Pérez-Losada et al. [24]**
Fraction of bacterial reads (of all aligned reads in infection samples)	n.a.	n.a.	0.7%–16.5%	n.a.	0.002%–19%	~0.02% (1 h postinfection); ~30% (24 h postinfection)	~5%	~0.03%–58%	2–4% (nonenriched); 11%–25% (enriched)	~0.2%–1.5%	on average 0.28%	~1%–10%	on average 67%	~0.7% (30 min postinfection); ~2% (24 h postinfection)	~5%
Differential expression analysis tool	DESeq	Cufflinks	edgeR, DESeq, SAMseq	TPM	RPKMO	DESeq	Cuffdiff (Cufflinks)	Cuffdiff (Cufflinks)	edgeR	limma	TPM; DESeq	edgeR	DESeq	TPM	Cufflinks
Data availability	n.a.	PRJNA218473 (NCBI)	PRJEB6003 (ENA)	GSM1462575–1462579, GSM1914919 (GEO)	GSE55292 (GEO)	GSE44253 (GEO)	n.a.	PRJNA256028 (NCBI)	PRJEB6552 (ENA)	GSE63900 (GEO)	GSE65528–31 (GEO)	GSE60144 (GEO)	GSE79595 (GEO)	GSE61730 (GEO)	n.a.

doi:10.1371/journal.ppat.1006033.t001

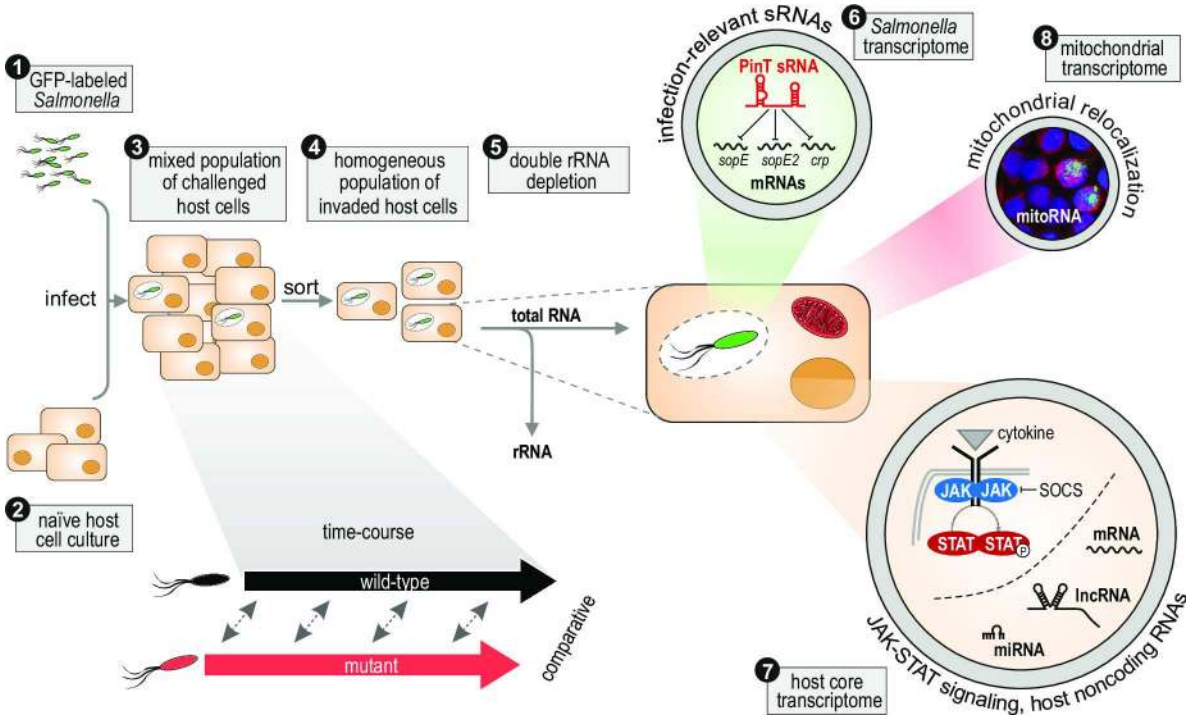

Fig 2. A generic dual RNA-seq workflow analyzing total mixed RNA after double rRNA depletion that discovered the role of PinT small regulatory RNA (sRNA) during *Salmonella* infection of host cells [13]. *Salmonella* having *gfp* stably integrated in their chromosome and expressed from a constitutive promoter were used to infect cultures of HeLa cells. RNA-seq of the bacterial input (1) or mock-infected HeLa cells (2) served as reference controls for *Salmonella* or human gene expression analysis, respectively. Infection was carried out in parallel with wild-type and sRNA mutant *Salmonella* strains, and samples were taken over a time-course of infection. The resulting cell samples constituted a mixed population consisting of both invaded (GFP-positive) and uninfected bystander (GFP-negative) cells (3). To obtain a homogeneous population of invaded cells, the samples were sorted based on the emitted GFP fluorescence (4). Total RNA was extracted from the thus enriched cells, rRNA from both infection partners was depleted (5), and rRNA-free samples were converted into cDNA libraries and sequenced. The resulting sequencing reads were mapped in parallel against the *Salmonella* and human (core and mitochondrial) genome. Differential expression analysis of the time course revealed the strong induction over time of a novel *Salmonella* sRNA, PinT, and comparative analysis unraveled the molecular footprint of this sRNA in the bacterial transcriptome (6). Likewise, comparison of the host transcriptome between wild-type and Δ*pinT* infections revealed PinT-dependent changes in the immune response, including a differential activation of Janus kinase-Signal Transducer and Activator of Transcription (JAK-STAT) signaling as well as changes with respect to the expression of host long noncoding RNAs (lncRNAs) and microRNAs (miRNAs) (7). In addition, the *pinT* status of the infecting bacterium influenced mitochondrial gene expression, and infection with Δ*pinT Salmonella* led to the relocalization of mitochondria in invaded host cells (8).

doi:10.1371/journal.ppat.1006033.g002

bacterial infection models, aiming to provide experimental and biocomputational guidelines for future dual RNA-seq assays.

Emerging Applications of Dual RNA-seq

Most dual RNA-seq analyses so far have been exploratory, characterizing the transcriptional dynamics of a particular infection system. An early dual RNA-seq study of HEp-2 epithelial carcinoma cells infected with the obligate intracellular pathogen *Chlamydia trachomatis* [10] revealed the induction of numerous metabolic mechanisms early after invasion—for example, riboflavin biosynthesis genes (*ribBA*) responding to extracellular reduction of iron (Fig 3). These changes previously escaped detection because of the few individual *Chlamydia* cells in the infected culture. Host transcripts, on the other hand, revealed an active response to invading *Chlamydia* (albeit with a currently unexplained dampening of immune signaling), in

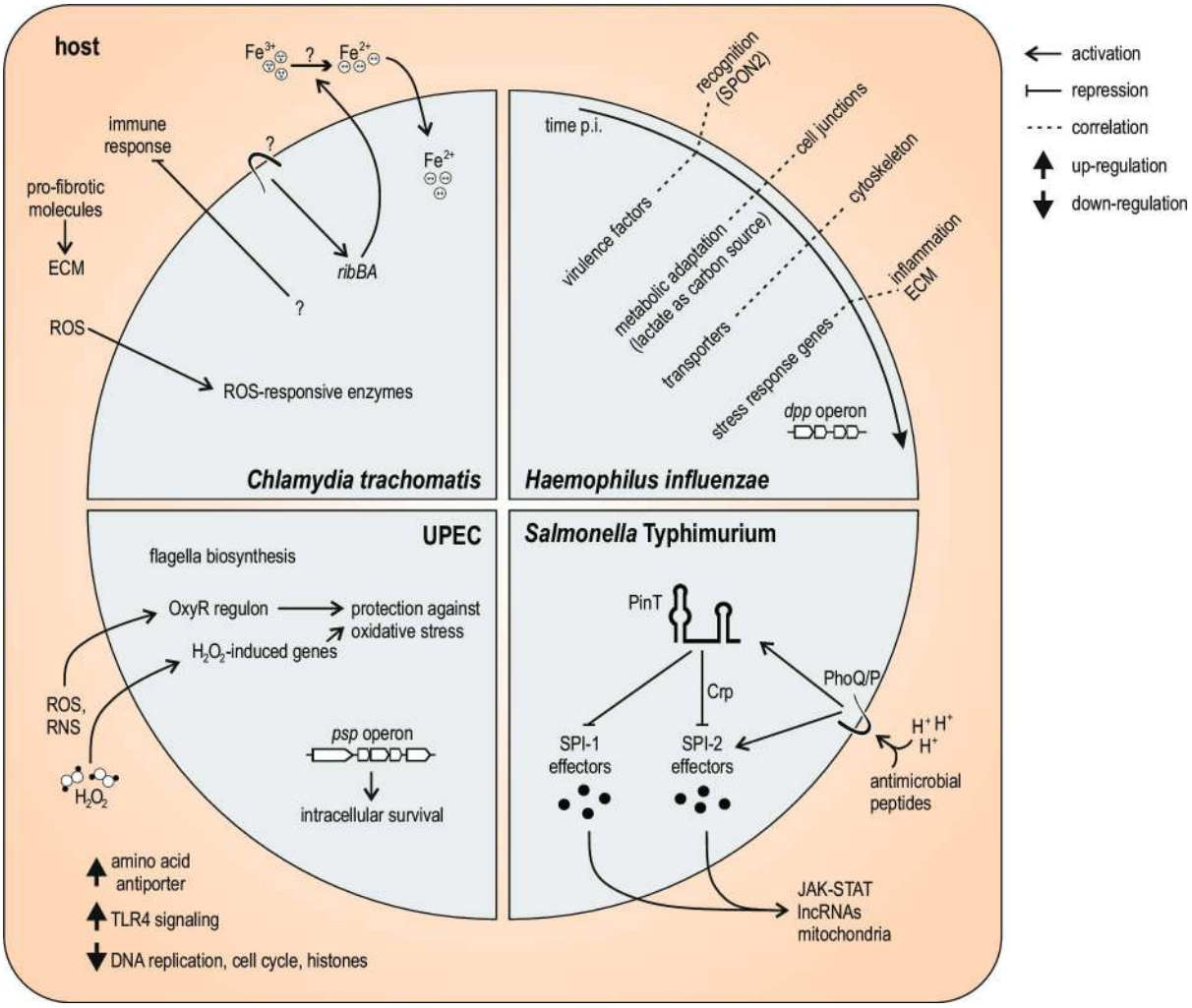

Fig 3. Illustration of biological insights obtained from dual RNA-seq studies in four different bacterial infection models. HEp-2 cells infected with obligate intracellular *Chlamydia trachomatis* [10], primary airway epithelial cells with nontypeable *Haemophilus influenzae* [16], primary murine bone marrow macrophages with uropathogenic *E. coli* (UPEC) [11], and diverse human, mouse, and porcine cell lines with *Salmonella* Typhimurium [13]. See main text for details.

doi:10.1371/journal.ppat.1006033.g003

contrast to an earlier microarray-based report in which only few changes in host transcription were observed during early infection [25].

This exploratory concept has been expanded into higher-resolution time-courses covering longer periods of infection. A study following host and pathogen gene expression over 72 hours in primary airway epithelial cells infected with nontypeable *Haemophilus influenzae* [16] revealed a strong early induction in the host of the extracellular pathogen recognition receptor Spondin 2 (SPON2), which acts as an opsonin that promotes macrophage phagocytosis of bacteria in the extracellular matrix [26]. The bacterial transcriptome reflected defined stress responses such as the induction of the *dipeptide transport system permease protein* (*dpp*) operon, whose gene products contribute to the protection against oxidative stress (Fig 3).

A comparative dual RNA-seq approach was taken to study two isolates of uropathogenic *Escherichia coli* (UPEC) strains—one being replication-competent and the other susceptible to killing by the host—in primary murine bone marrow–derived macrophages [11]. While the

host transcriptome was broadly similar, bacterial gene expression varied markedly between the two isolates. Several genes were induced exclusively in the replicating isolate, suggesting that some of these might encode for essential virulence factors. Indeed, deletion of one of these genes, *phage shock protein A* (*pspA*), led to a survival defect compared to the cognate wild-type strain (Fig 3).

Defining "Molecular" Phenotypes by Dual RNA-seq

Our recent dual RNA-seq profiling of *Salmonella* Typhimurium infection of human epithelial cells and porcine macrophages [13] combined these above two strategies of exploratory and hypothesis-driven comparative design (Fig 2). This study, for the first time, also analyzed all major coding and noncoding RNA classes of the bacterial pathogen and its host cell. Within the class of *Salmonella* small noncoding RNAs (sRNAs), the previously uncharacterized *Salmonella* PinT sRNA was consistently and highly induced during infection of 14 distinct cell types. Biocomputational clustering of expression kinetics along a high-resolution time-course of infected HeLa cells predicted that PinT is activated by the PhoP/Q two-component system, which regulates intracellular virulence (Fig 3). Subsequently, a comparative dual RNA-seq time-course with a *pinT* deletion mutant unraveled the function of PinT as a posttranscriptional regulator of the expression of important virulence genes of *Salmonella* inside both human and porcine cell lines. The activity of PinT has widespread effects on the host response, with ~10% of all detected human mRNAs as well as various noncoding transcripts being differentially expressed between the two infections.

Importantly, the generic dual RNA-seq protocol used in this study also detects mitochondrial transcripts, which are typically neglected in host RNA profiling. Analysis of this "third transcriptome" showed that mitochondrial transcripts were hyperexpressed in HeLa cells infected with Δ*pinT* compared to wild-type *Salmonella*. This observation guided the discovery of altered subcellular distributions of mitochondria (Fig 2), an sRNA phenotype that would have likely been missed in standard analyses.

Intriguingly, while PinT does not produce a robust "macroscopic" replication phenotype in cell culture, the dual RNA-seq results show that PinT activity times *Salmonella* virulence gene expression shortly after invasion. We refer to this transcriptional signature as a "molecular phenotype" [27], which may represent a new approach to characterizing the role of gene products in infection. Of note, previous transposon mutagenesis studies in large animal models, including pigs, showed that *pinT* disruption is attenuating [28] despite the absence of an obvious phenotype in cell culture, illustrating the relevance of molecular phenotypes to studying disease in the absence of accessible model systems.

On Designing a Dual RNA-seq Experiment

While the technical feasibility of dual RNA-seq has now been firmly established, a near-infinite variety of infection models wait to be explored. The complexity of these systems introduces significant challenges for the analysis of the resulting datasets. Next, we will review challenges in planning and analyzing dual RNA-seq experiments.

(a) Obtaining RNA

Dual RNA-seq requires sufficient starting material for sequencing, particularly for the infecting bacterium. Current protocols are based on at least 10,000 infected cells [12,13,29]. Frequently, only a minor fraction of eukaryotic cells in a sample will be infected, approximately 2%–5% in our study of HeLa cells infected with green fluorescent protein (GFP)-expressing *Salmonella* [13]. Therefore, to enrich for bacterial RNA and to distinguish the host response of

infected from noninfected bystander cells, these populations must be separated before analysis. Of the six current dual RNA-seq studies of intracellular bacteria (Table 1), three enriched invaded cells either by laser capture microdissection [14] or via FACS [12,13] using endogenously expressed fluorescent markers and/or cell wall–binding dyes. To minimize unwanted transcriptomic changes during sample acquisition, cells should be kept at low temperature (e.g., sorted under continuous cooling to 4˚C [12,29]) until they are lysed. However, when many time points or strains are being compared, it may be challenging to sort the cells immediately upon harvest. In such cases, the transcriptomes should be "frozen" by fixation (Box 1). For example, we have optimized fixation conditions for *Salmonella* infections that leave cells physically intact and do not bleach fluorescent signals or interfere with RNA isolation [13], and recently a similar approach has been used for pneumococcal infections [22].

Once the infected cells are collected, they must be lysed to extract RNA. Importantly, many standard commercial lysis buffers are optimized only for particular organisms and may, for example, fail to break the thick envelope of gram-positive pathogens. In a study of human THP-1 cells infected with gram-positive *Mycobacterium bovis* [15], total RNA was obtained after mechanically breaking the cells with beads in a benchtop homogenizer. After lysis, a number of RNA isolation methods have been successfully used for dual RNA-seq (Table 1). Before sequencing, it is advisable to first estimate the relative concentration of bacterial and host RNA in the sample (for instance, by quantitative real-time PCR [qRT-PCR] [13]); this can inform decisions about required read depth or whether changes need to be made to the infection protocol to increase bacterial counts, such as increasing the multiplicity of infection.

Box 1. RNA Preservation

To minimize unwanted transcriptomic changes during sample processing, the RNA content of infected cells may be stabilized. Two preservation strategies exist: Alcohol- or ammonium sulfate–based preservatives inactivate RNases and RNA polymerases by denaturing cellular proteins through the removal of water. In contrast, formaldehyde-containing fixatives induce intra- and intermolecular cross-links between amino groups and thereby block de novo transcription or RNA decay. In the context of dual RNA-seq, besides leaving cells physically intact to enable cell sorting, transcriptome stabilization must avoid quenching fluorescent signals (as is typically the case for phenol- or alcohol-containing reagents) or interfering with high-quality RNA isolation (which is problematic with cross-linked samples). In our recent study of *Salmonella*-infected HeLa cells, we evaluated eight commonly used transcriptome stabilization techniques (see supplementary material of [13]). For this model system, the ammonium sulfate–based RNA*later* reagent (Qiagen), previously been used to fix infected tissue samples [17,19,21] (Tab. 1) or prokaryotic cells alone [30], performed best. However, this is unlikely to represent a generic protocol. For example, we have seen ex vivo that primary cells, which are more fragile than immortalized cell lines, tend to lyse in RNA*later*. Therefore, transcriptome stabilization should be optimized empirically for any infection model. Promising recent studies have demonstrated the compatibility of paraformaldehyde-based fixation with cell sorting and, importantly, high-quality RNA isolation [31,32]. Detailed discussions of transcriptome fixation are available in the literature [33–35].

This naturally raises the question: how many cDNA reads are enough? Differential expression power analyses universally favor biological replication over sequencing depth, particularly once a minimum depth threshold has been attained. For eukaryotes, increasing sequencing depth appears to have diminishing returns after around 10–20 million nonribosomal RNA reads [36,37]—though accurate quantification of low-abundance transcripts may require >80 million reads [38]—while for bacteria this threshold seems to be 3–5 million nonribosomal reads [39]. With current technology, this number of bacterial reads may necessitate specific enrichment, particularly at early time points before intracellular bacteria have undergone replication. However, analysis of subsampled RNA-seq data from a *Vibrio cholera* infection in a juvenal rabbit model [17] showed that differential expression of major virulence and colonization factors could already be detected with as few as 40,000–60,000 nonribosomal RNA reads, in agreement with results for *Salmonella* at early time points of infection [13]. Thus, while low read depth is not ideal, low-coverage data still have value, particularly in the case of poorly characterized pathogens for which basic virulence mechanisms are largely unknown. Clearly, more subtle effects, such as adaptation of bacterial metabolism to the intracellular environment, demand greater sequencing depth.

(b) Mapping and Normalization

The broad strokes of dual RNA-seq analysis differ little from conventional RNA-seq [40,41]: sequencing reads must be cleaned, mapped, and normalized; differentially expressed transcripts must be identified; and then further functional analyses must be performed to aid in interpretation of the data (Fig 4). However, the complexity of dual RNA-seq designs introduces additional complications at each step as well as entirely new analytical problems.

Much complexity derives from working simultaneously with multiple genome sequences. Although this can be done easily by including all replicons of both organisms as references during mapping, it is important to determine the selectivity of read mapping to both genomes, as cross-mapping reads will affect transcript quantification. In practice, with standard Illumina read lengths of 75–150 bases, we observe negligible cross-mapping in the case of *Salmonella* and mammalian hosts [13], with most of this originating from rRNA and tRNA loci. However, since genome composition varies tremendously across the bacterial phylogeny, potential cross-mapping should be a routine quality control step. The READemption RNA-seq analysis pipeline [42], which relies on the segemehl read mapper [43], contains alignment subcommands implementing such cross-mapping analysis. In principle, any read mapper capable of spliced alignment [44] can be used for read alignment, though some studies have chosen to use separate spliced and nonspliced aligners for mapping to the eukaryotic and bacterial genomes, respectively.

Once mapping has been completed, normalization and quantification are required. Within-sample normalization methods, such as transcripts per million (TPM) [45], often suffice to obtain a qualitative overview of transcriptional dynamics [12,20,46] but should be interpreted with caution since methods are currently lacking for incorporating replicate measurements in these analyses. Most analyses of interest require robust comparisons between samples. The most commonly used (and best performing [47]) RNA-seq normalization techniques address this problem by attempting to scale read counts by a factor derived from a set of putatively invariant genes identified through excluding genes with extreme differences in expression [48], the use of robust median statistics [49], or comparisons of quantiles [50]. These normalization methods make the common assumption that a core set of genes are not differentially expressed and may fail when this assumption is violated. Scenarios violating this assumption have been observed in eukaryotes [51] and can similarly

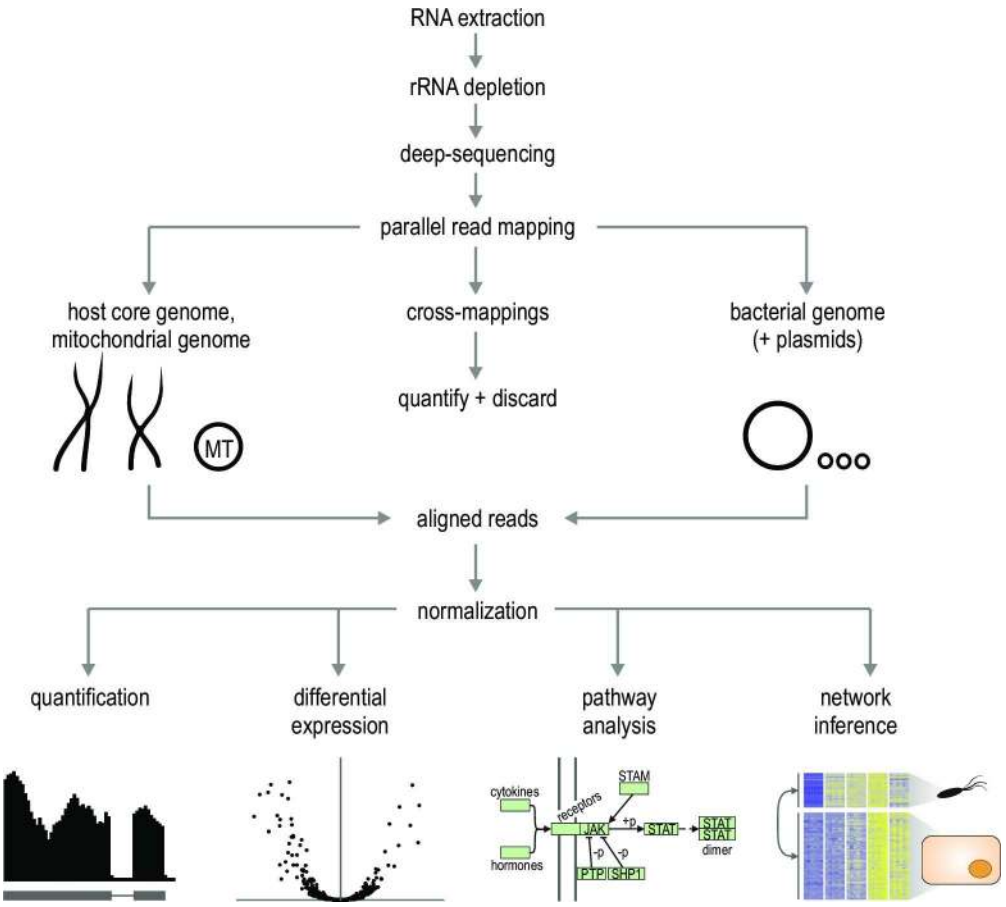

Fig 4. Bioinformatic analysis pipeline for dual RNA-seq datasets. Quality-filtered RNA-seq reads are aligned in parallel against the respective host and pathogen replicons. Reads mapping equally well to both reference organisms ("cross-mappings") are quantified and discarded from downstream analyses. Reads unequivocally mapped to either the bacterial or host reference are used for quantification and functional analyses. Dual RNA-seq enables a wide range of downstream analyses, discussed in detail in the text. "MT," mitochondrial genome.

doi:10.1371/journal.ppat.1006033.g004

be expected to occur in bacteria after major regulatory transitions, such as that from exponential growth to stationary phase. This may be particularly important in certain infections in which dormant or persister cells develop [52,53]. The use of RNA spikes calibrated to cell counts may enable a robust estimation of differences in expression in such cases [51,54–56]. However, the use of spike-ins presents its own problems: a large multicenter study [57] using External RNA Controls Consortium (ERCC) spike-in controls [58] found that biases introduced in library preparation made absolute transcript quantification unreliable, even when identical protocols and platforms are used. The factors driving these biases are unclear, though they appear to be both sequence- and protocol-dependent [57] and thus may be challenging to correct. This also suggests spike-ins should be added as early as possible in sample processing so that any biases from steps such as ribosomal depletion can be captured. These difficulties notwithstanding, ratios between spike-ins in libraries prepared within the same batch are highly reproducible [57,59], indicating that spike-ins should be sufficient for calibrating most differential expression analyses. New spike-in sets have recently been developed that can be used to assess various aspects of

RNA sample processing and analysis [60], such as transcript assembly and isoform quantification, which may be informative in advanced analyses. Alternatively, since dual RNA-seq provides access to two transcriptomes within each pool, if only the host or the bacterium is affected by a global shift in gene expression, a scale factor could be determined for the organism which meets the assumption of the scaling normalization and applied to the other, adjusting for relative population size.

Scaling normalization techniques address differences in sequencing depth between RNA-seq experiments. However, there are many other factors besides read depth which can introduce unwanted variation in high-throughput experiments and lead to reduced power in downstream analyses, commonly referred to as "batch effects" (Box 2). Within the context of dual RNA-seq experiments, myriad opportunities for the introduction of such effects exist: heterogeneity in cell populations and infection, differences in media and reagent batches, variation in laboratory and incubator temperature and oxygen, inaccuracy in cell sorting, and differences in transcriptome stabilization, RNA extraction, library preparation, and sequencing. The prevalence of such effects in high-throughput data has been well documented [61], with lessons to be learned from other fields studying subtle effects in complex model systems, such as stem cell biology and neuroscience [62,63]. We observed similar effects in our study of PinT when comparing wild-type and mutant time-courses [13] and were able to correct for these using recently developed techniques (Box 2).

(c) Differential Expression Analysis

Differential expression analysis forms the backbone of most RNA-seq analyses, most frequently done in the *R* statistical programming language with packages available through the Bioconductor framework [65]. Popular analysis packages include edgeR [66], DESeq2 [67], and limma/voom [68]. These three packages perform well, with slightly different characteristics in benchmarks [69–72]: DESeq generally appears to be more conservative and edgeR more liberal in its *p*-value calculations. While these tools work with predefined annotations and ignore differential isoform usage, RNA-seq also raises the possibility of directly defining boundaries of eukaryotic transcripts, which are typically subject to regulated alternative splicing [73]. A range of algorithms

Box 2. How to Deal with Batch Effects

Traditionally, batch effects were accounted for by incorporating date as a nuisance factor in differential expression analysis [64]. While this may work for simple experiments, in complex experiments (such as dual RNA-seq), samples are likely exposed to many treatments that may vary slightly in their effect, and these will not necessarily be constant even within a single "batch." To solve this problem, recent methods such as RUV-seq and SVA-seq have been developed that perform factor analyses, similar to principal component analysis (PCA), to identify nuisance factors uncorrelated with the experimental factors of interest [54,55]. Nuisance factors can then either be "cleaned" from the read counts directly for the purposes of clustering or other qualitative analyses or incorporated directly as covariates in differential expression analyses. Two excellent case studies provide detailed guidelines for applying methods for evaluating the presence of such confounding batch effects [62,63].

offer isoform discovery, quantification, and differential analysis [74], though generally dedicated pipelines such as the Tuxedo suite [75] have been standard. The recently developed Ballgown utility allows for the easy importation of transcript assemblies and quantifications into *R* [76] and therefore integration of these methods into the Bioconductor RNA-seq analysis ecosystem.

Most of the published dual RNA-seq experiments have involved a time-course and analyzed differential expression by pairwise comparisons; however, this effectively ignores the temporal relationship between samples. Changes in transcript expression can be assumed to be smooth for most genes over time, and this assumption can be used to increase the power of analyses: in effect, contiguous samples act as partial replicates for one another, allowing for more accurate estimation of expression variance. While this does not remove the need for replication, it does raise the possibility of more informative designs than simple replication. For instance, rather than repeatedly sampling the same time points, replicate experiments could be staggered in time so as to provide higher temporal resolution while also demonstrating reproducibility. While not frequently used in the literature, such analyses are possible in analysis packages supporting generalized linear models, such as edgeR and limma/voom, by performing differential expression analysis along fitted curves (see the developmental time-course analysis in *Drosophila* embryos [68]). We hope additional dedicated approaches to time-course analysis will be forthcoming.

(d) Aiding Interpretation: Functional Analyses

The outcome of differential expression analysis is a long list of genes for both bacteria and host, which must be interpreted in terms of gene function to produce testable hypotheses. Several databases provide suits for this purpose, though none provide complete information for either eukaryotic cells or bacteria. Popular databases include the Gene Ontology (GO) [77] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [78] databases, which provide general resources for gene functions and interactions in diverse organisms. More specialized knowledge bases also exist—for example, BioCyc [79], which attempts to reconstruct metabolic networks primarily from genomic information. The innate immunity resource InnateDB is of particular interest for the host response part of dual RNA-seq data [80]; it incorporates interaction data from a variety of sources—complemented with manually curated human, murine, and bovine innate immunity pathways and interactions—and provides a number of tools for analyzing and visualizing functional assays in the context of these data. Furthermore, molecular signatures may be reconstructed from relevant high-throughput experiments as collected by MSigDB [81] from eukaryotic microarray and RNA-seq data. The increasing availability of RNA-seq data for bacteria exposed to simple stress conditions raises the possibility that similar signatures could be constructed for them: for instance, by mining resources like the "*Salmonella* Gene Expression Compendium," which collects expression data for 22 infection-relevant conditions [46].

Dual RNA-seq crucially depends on proper statistical analysis in order to determine gene sets significantly differentially expressed during infection. Originally developed for microarray experiments, many of these techniques remain poorly tested on RNA-seq datasets. Technical issues, such as biases towards detecting differential expression in longer transcripts in sequencing data as compared to array data [82], have made it unclear how applicable these approaches are to RNA-seq. The first benchmarks of gene set enrichment methods on RNA-seq data have recently been published [83] and can provide preliminary guidance.

Dual RNA-seq can also be directly used to infer links between genes through so-called network inference (NI) approaches which are popular in reconstructing global regulatory networks from large collections of expression data in diverse conditions [84,85]. NI methods

frequently use measures of coexpression, such as correlation or mutual information, to predict interactions between, for instance, transcriptional regulators and their regulons. With dual RNA-seq time-course gene expression data, a time lag can be introduced in coexpression calculations, allowing for the prediction of potentially causal interactions. This approach was pioneered in studies of the cyanobacterium *Synechocystis* and its response to varying light intensities to identify putative directional interactions [86]. We applied such a correlational analysis to a dual RNA-seq time-course of *Salmonella* infection of HeLa cells, linking virulence gene expression in *Salmonella* to the induction of host immune signaling through the Janus kinase-Signal Transducer and Activator of Transcription (JAK-STAT) pathway [13]. More complex NI models utilizing ordinary differential equations (ODEs) with dual RNA-seq data have successfully predicted host–pathogen interactions between the fungal pathogen *Candida albicans* and murine host cells [4,87]. While ODEs are preferable in that they can explicitly model the dynamics of changes in gene expression and incorporate prior information in a principled fashion, they are also computationally demanding, limiting their use to modeling small subnetworks of genes [88].

Future Directions

While the potential of dual RNA-seq in cell culture–based infection models has clearly not yet been exhausted, the next steps in the development of this technique are already on the horizon. For example, "Multi RNA-seq" was applied to characterize the human airway epithelium in conjunction with the commensal bacteria populating it [24]; similarly, a recent study profiling *Yersinia pseudotuberculosis* gene expression in the mouse cecum was able to discriminate between various intestinal bacterial species [21] (Table 1). In the future, such approaches could address frequently occurring coinfections of human hosts with bacterial and viral pathogens, including those of *Streptococcus* spp. and influenza virus [89] or *Chlamydia* spp. and human herpes virus [90,91]. As coinfections are a major risk factor for human health [92], such "Triple RNA-seq" experiments would be of direct medical relevance. Likewise, robotic systems have enabled previously prohibitively laborious applications, such as comprehensive chemical–genetic screens [93] and mapping of large transposon mutant libraries [94]. In combination with ongoing improvements in cDNA library preparation and sequencing technologies, these could provide a foundation for high-throughput dual RNA-seq designs. For instance, with efficient multiplexing techniques [29], systematic virulence screens could be imagined that compare expression changes between infections with defined deletion strains of, say, every gene identified as a hit in transposon mutagenesis screens and the isogenic wild-type strain. Combining such ultra-high-throughput approaches will be a powerful strategy to define the molecular phenotypes of hundreds of pathogen genes in parallel, providing a rich basis for dissecting host–microbe interactions [27].

Two more intermediate and exciting possibilities are the expansion to infected tissue (and eventually animal models), and the development of single-cell dual RNA-seq. With respect to the former, several studies suggest widespread differences in bacterial behavior during the infection of two-dimensional monocultures compared to that of three-dimensional tissue [95,96] and whole animal models [97]. Adapting dual RNA-seq to these more realistic models will require numerous innovations. Simple homogenization of the tissue, as has been done for in vivo bacterial RNA-seq studies [19,21] (Table 1), may provide a first step along this path. While this review was in production, two studies were published that report the successful application of dual RNA-seq to in vivo models of infection with extracellular pathogens. Host and pathogen gene expression was analysed in a murine model of acute pneumonia caused by *Pseudomonas aeruginosa* [98] and in a murine gastroenteritis model with *Yersinia pseudotuberculosis* [99]. In both cases, infected tissues (lungs or Peyer's patches, respectively) were isolated and homogenized prior to total RNA extraction, rRNA depletion, and sequencing. However,

since these complex samples do not consist of a single cell type, dissociation of tissues into single-cell suspensions and the enrichment of defined cell types of interest would provide a more complete picture of this complex environment. As dissociation and antibody staining are time consuming, the transcriptomes of host and pathogen must be stabilized immediately after harvest. Ongoing progress in sample preservation provides a foundation on which to build (Box 1) if these treatments can be made compatible with, say, enzymatic treatment to disrupt cell junctions. Additionally, current in vitro dual RNA-seq studies have been performed with 10,000–50,000 sorted cells. Cell numbers will likely be limiting in tissue and animal models, requiring technical advances in cDNA library preparation. Such advances may come in the development of dual RNA-seq protocols for single cells.

Single-cell dual RNA-seq promises to be a game changer in the study of those many bacterial pathogens that are known to form specific, phenotypically distinct subpopulations during infection [100], often associated with distinct disease outcomes [101]. Eukaryotic single-cell RNA-seq studies have already shown that individual immune cells stimulated with the same concentration of the bacterial cell wall component lipopolysaccharide (LPS) mount disparate responses to the challenge [102,103]. In addition, single-cell RNA-seq has revealed heterogeneity in the host mRNA response as the result of pathogen variability [12]. However, current protocols are unable to sample the bacterial transcriptome, as they generally rely on poly(A)-dependent priming of reverse transcription [104,105]. The literature suggests several solutions to poly(A) dependency, such as direct adapter ligation, which unfortunately currently requires approximately10,000 cells [12,13,29]. Priming with random hexamers or—to selectively deplete rRNAs—"not-so-random" primers [106] may provide a more efficient solution. Finally, a thermostable group II intron reverse transcriptase (TGIRT) has recently been described as a highly sensitive (down to 1 ng input RNA), poly(A)-independent enzyme with template-switching activity that can be used to add sequencing adapters, directly avoiding inefficient ligation steps [107,108]. Dedicated bacterial single-cell RNA-seq protocols have also recently been described [109,110] that rely on rolling circle amplification and have been demonstrated to generate large amounts of double-stranded cDNA product from small amounts of input template. Since reverse transcription in these protocols is mediated by random primers, it might be adopted for single-cell dual RNA-seq, though the efficiency of this remains to be tested.

In summary, dual RNA-seq is an emerging technique to profile gene expression changes that accompany infection of mammalian cells by bacterial pathogens. Unlike traditional approaches, dual RNA-seq has proven capable of capturing host and pathogen transcriptomes simultaneously, providing direct insight into host–pathogen interplay. However, dual RNA-seq is still in its infancy, and future efforts—with respect to both experimental aspects and bioinformatics—will be required to exploit its full potential.

Acknowledgments

The authors thank Dr. Sandy Ramona Pernitzsch for help with the figures and Dr. Antoine-Emmanuel Saliba for critical comments on the manuscript.

References

1. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17: 333–351. doi: 10.1038/nrg.2016.49 PMID: 27184599
2. Woodhouse SD, Narayan R, Latham S, Lee S, Antrobus R, et al. (2010) Transcriptome sequencing, microarray, and proteomic analyses reveal cellular and metabolic impact of hepatitis C virus infection in vitro. Hepatology 52: 443–453. doi: 10.1002/hep.23733 PMID: 20683944
3. Strong MJ, Xu G, Coco J, Baribault C, Vinay DS, et al. (2013) Differences in gastric carcinoma microenvironment stratify according to EBV infection intensity: implications for possible immune adjuvant therapy. PLoS Pathog 9: e1003341. doi: 10.1371/journal.ppat.1003341 PMID: 23671415

4. Tierney L, Linde J, Muller S, Brunke S, Molina JC, et al. (2012) An Interspecies Regulatory Network Inferred from Simultaneous RNA-seq of Candida albicans Invading Innate Immune Cells. Front Microbiol 3: 85. doi: 10.3389/fmicb.2012.00085 PMID: 22416242
5. Choi YJ, Aliota MT, Mayhew GF, Erickson SM, Christensen BM (2014) Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions. PLoS Negl Trop Dis 8: e2905. doi: 10.1371/journal.pntd.0002905 PMID: 24853112
6. Pittman KJ, Aliota MT, Knoll LJ (2014) Dual transcriptional profiling of mice and Toxoplasma gondii during acute and chronic infection. BMC Genomics 15: 806. doi: 10.1186/1471-2164-15-806 PMID: 25240600
7. Dillon LA, Suresh R, Okrah K, Corrada Bravo H, Mosser DM, et al. (2015) Simultaneous transcriptional profiling of Leishmania major and its murine macrophage host cell reveals insights into host-pathogen interactions. BMC Genomics 16: 1108. doi: 10.1186/s12864-015-2237-2 PMID: 26715493
8. Westermann AJ, Gorski SA, Vogel J (2012) Dual RNA-seq of pathogen and host. Nat Rev Microbiol 10: 618–630. doi: 10.1038/nrmicro2852 PMID: 22890146
9. Alberts B, Bray D, Lewis J, Raff M, Roberts K, et al. (1994) Molecular Biology of the Cell, 3rd edition. Garland Publishing, New York.
10. Humphrys MS, Creasy T, Sun YZ, Shetty AC, Chibucos MC, et al. (2013) Simultaneous Transcriptional Profiling of Bacteria and Their Host Cells. Plos One 8.
11. Mavromatis CH, Bokil NJ, Totsika M, Kakkanat A, Schaale K, et al. (2015) The co-transcriptome of uropathogenic Escherichia coli-infected mouse macrophages reveals new insights into host-pathogen interactions. Cell Microbiol 17: 730–746. doi: 10.1111/cmi.12397 PMID: 25410299
12. Avraham R, Haseley N, Brown D, Penaranda C, Jijon HB, et al. (2015) Pathogen Cell-to-Cell Variability Drives Heterogeneity in Host Immune Responses. Cell 162: 1309–1321. doi: 10.1016/j.cell.2015.08.027 PMID: 26343579
13. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, et al. (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature 529: 496–501. doi: 10.1038/nature16547 PMID: 26789254
14. Vannucci FA, Foster DN, Gebhart CJ (2013) Laser microdissection coupled with RNA-seq analysis of porcine enterocytes infected with an obligate intracellular pathogen (Lawsonia intracellularis). BMC Genomics 14: 421. doi: 10.1186/1471-2164-14-421 PMID: 23800029
15. Rienksma RA, Suarez-Diez M, Mollenkopf HJ, Dolganov GM, Dorhoi A, et al. (2015) Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. BMC Genomics 16: 34. doi: 10.1186/s12864-014-1197-2 PMID: 25649146
16. Baddal B, Muzzi A, Censini S, Calogero RA, Torricelli G, et al. (2015) Dual RNA-seq of Nontypeable Haemophilus influenzae and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. MBio 6: e01765–01715. doi: 10.1128/mBio.01765-15 PMID: 26578681
17. Mandlik A, Livny J, Robins WP, Ritchie JM, Mekalanos JJ, et al. (2011) RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. Cell Host Microbe 10: 165–174. doi: 10.1016/j.chom.2011.07.007 PMID: 21843873
18. Lamont EA, Xu WW, Sreevatsan S (2013) Host-Mycobacterium avium subsp. paratuberculosis interactome reveals a novel iron assimilation mechanism linked to nitric oxide stress during early infection. BMC Genomics 14: 694. doi: 10.1186/1471-2164-14-694 PMID: 24112552
19. Szafranska AK, Oxley AP, Chaves-Moreno D, Horst SA, Rosslenbroich S, et al. (2014) High-resolution transcriptomic analysis of the adaptive response of Staphylococcus aureus during acute and chronic phases of osteomyelitis. MBio 5.
20. Srikumar S, Kroger C, Hebrard M, Colgan A, Owen SV, et al. (2015) RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of Salmonella Typhimurium. PLoS Pathog 11: e1005262. doi: 10.1371/journal.ppat.1005262 PMID: 26561851
21. Avican K, Fahlgren A, Huss M, Heroven AK, Beckstette M, et al. (2015) Reprogramming of Yersinia from virulent to persistent mode revealed by complex in vivo RNA-seq analysis. PLoS Pathog 11: e1004600. doi: 10.1371/journal.ppat.1004600 PMID: 25590628
22. Aprianto R, Slager J, Holsappel S, Veening J-W (2016) Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. Genome Biology 17: 1–16.
23. Afonso-Grunz F, Hoffmeier K, Muller S, Westermann AJ, Rotter B, et al. (2015) Dual 3'Seq using deepSuperSAGE uncovers transcriptomes of interacting Salmonella enterica Typhimurium and human host cells. BMC Genomics 16: 323. doi: 10.1186/s12864-015-1489-1 PMID: 25927313

24. Perez-Losada M, Castro-Nallar E, Bendall ML, Freishtat RJ, Crandall KA (2015) Dual Transcriptomic Profiling of Host and Microbiota during Health and Disease in Pediatric Asthma. PLoS One 10: e0131819. doi: 10.1371/journal.pone.0131819 PMID: 26125632

25. Xia M, Bumgarner RE, Lampe MF, Stamm WE (2003) Chlamydia trachomatis infection alters host cell transcription in diverse cellular pathways. J Infect Dis 187: 424–434. doi: 10.1086/367962 PMID: 12552426

26. He YW, Li H, Zhang J, Hsu CL, Lin E, et al. (2004) The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. Nat Immunol 5: 88–97. doi: 10.1038/ni1021 PMID: 14691481

27. Barquist L, Westermann AJ, Vogel J (2016) Molecular phenotyping of infection-associated small non-coding RNAs. Philosophical Transactions B of the Royal Society.

28. Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, et al. (2013) Comprehensive assignment of roles for Salmonella typhimurium genes in intestinal colonization of food-producing animals. PLoS Genet 9: e1003456. doi: 10.1371/journal.pgen.1003456 PMID: 23637626

29. Avraham R, Haseley N, Fan A, Bloom-Ackermann Z, Livny J, et al. (2016) A highly multiplexed and sensitive RNA-seq protocol for simultaneous analysis of host and pathogen transcriptomes. Nat Protoc 11: 1477–1491. doi: 10.1038/nprot.2016.090 PMID: 27442864

30. Bachoon DS, Chen F, Hodson RE (2001) RNA recovery and detection of mRNA by RT-PCR from preserved prokaryotic samples. FEMS Microbiol Lett 201: 127–132. PMID: 11470350

31. Hrvatin S, Deng F, O'Donnell CW, Gifford DK, Melton DA (2014) MARIS: method for analyzing RNA following intracellular sorting. PLoS One 9: e89459. doi: 10.1371/journal.pone.0089459 PMID: 24594682

32. Evers DL, Fowler CB, Cunningham BR, Mason JT, O'Leary TJ (2011) The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal. J Mol Diagn 13: 282–288. doi: 10.1016/j.jmoldx.2011.01.010 PMID: 21497290

33. Howat WJ, Wilson BA (2014) Tissue fixation and the effect of molecular fixatives on downstream staining procedures. Methods 70: 12–19. doi: 10.1016/j.ymeth.2014.01.022 PMID: 24561827

34. Cox ML, Schray CL, Luster CN, Stewart ZS, Korytko PJ, et al. (2006) Assessment of fixatives, fixation, and tissue processing on morphology and RNA integrity. Exp Mol Pathol 80: 183–191. doi: 10.1016/j.yexmp.2005.10.002 PMID: 16332367

35. Cox ML, Eddy SM, Stewart ZS, Kennel MR, Man MZ, et al. (2008) Investigating fixative-induced changes in RNA quality and utility by microarray analysis. Exp Mol Pathol 84: 156–172. doi: 10.1016/j.yexmp.2007.11.002 PMID: 18291364

36. Ching T, Huang S, Garmire LX (2014) Power analysis and sample size estimation for RNA-Seq differential expression. RNA 20: 1684–1696. doi: 10.1261/rna.046011.114 PMID: 25246651

37. Liu Y, Zhou J, White KP (2014) RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30: 301–304. doi: 10.1093/bioinformatics/btt688 PMID: 24319002

38. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15: 121–132. doi: 10.1038/nrg3642 PMID: 24434847

39. Haas BJ, Chin M, Nusbaum C, Birren BW, Livny J (2013) How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics 13: 734.

40. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. (2016) A survey of best practices for RNA-seq data analysis. Genome Biol 17: 13. doi: 10.1186/s13059-016-0881-8 PMID: 26813401

41. Ravasi T, Mavromatis CH, Bokil NJ, Schembri MA, Sweet MJ (2016) Co-transcriptomic Analysis by RNA Sequencing to Simultaneously Measure Regulated Gene Expression in Host and Bacterial Pathogen. Methods Mol Biol 1390: 145–158. doi: 10.1007/978-1-4939-3335-8_10 PMID: 26803628

42. Forstner KU, Vogel J, Sharma CM (2014) READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. Bioinformatics 30: 3421–3423. doi: 10.1093/bioinformatics/btu533 PMID: 25123900

43. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, et al. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol 5: e1000502. doi: 10.1371/journal.pcbi.1000502 PMID: 19750212

44. Engstrom PG, Steijger T, Sipos B, Grant GR, Kahles A, et al. (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods 10: 1185–1191. doi: 10.1038/nmeth.2722 PMID: 24185836

45. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN (2010) RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26: 493–500. doi: 10.1093/bioinformatics/btp692 PMID: 20022975

46. Kroger C, Colgan A, Srikumar S, Handler K, Sivasankaran SK, et al. (2013) An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. Cell Host Microbe 14: 683–695. doi: 10.1016/j.chom.2013.11.010 PMID: 24331466

47. Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, et al. (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14: 671–683. doi: 10.1093/bib/bbs046 PMID: 22988256

48. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25. doi: 10.1186/gb-2010-11-3-r25 PMID: 20196867

49. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106. doi: 10.1186/gb-2010-11-10-r106 PMID: 20979621

50. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94. doi: 10.1186/1471-2105-11-94 PMID: 20167110

51. Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, et al. (2012) Revisiting global gene expression analysis. Cell 151: 476–482. doi: 10.1016/j.cell.2012.10.012 PMID: 23101621

52. Maisonneuve E, Gerdes K (2014) Molecular mechanisms underlying bacterial persisters. Cell 157: 539–548. doi: 10.1016/j.cell.2014.02.050 PMID: 24766804

53. Helaine S, Holden DW (2013) Heterogeneity of intracellular replication of bacterial pathogens. Curr Opin Microbiol 16: 184–191. doi: 10.1016/j.mib.2012.12.004 PMID: 23485258

54. Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32: 896–902. doi: 10.1038/nbt.2931 PMID: 25150836

55. Leek JT (2014) svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res 42.

56. Chen K, Hu Z, Xia Z, Zhao D, Li W, et al. (2016) The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. Mol Cell Biol 36: 662–667.

57. Consortium SM-I (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol 32: 903–914. doi: 10.1038/nbt.2957 PMID: 25150838

58. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, et al. (2011) Synthetic spike-in standards for RNA-seq experiments. Genome Res 21: 1543–1551. doi: 10.1101/gr.121095.111 PMID: 21816910

59. Munro SA, Lund SP, Pine PS, Binder H, Clevert DA, et al. (2014) Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. Nat Commun 5: 5125. doi: 10.1038/ncomms6125 PMID: 25254650

60. Hardwick SA, Chen WY, Wong T, Deveson IW, Blackburn J, et al. (2016) Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13: 792–798. doi: 10.1038/nmeth.3958 PMID: 27502218

61. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11: 733–739. doi: 10.1038/nrg2825 PMID: 20838408

62. Peixoto L, Risso D, Poplawski SG, Wimmer ME, Speed TP, et al. (2015) How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. Nucleic Acids Res 43: 7664–7674. doi: 10.1093/nar/gkv736 PMID: 26202970

63. Jaffe AE, Hyde T, Kleinman J, Weinbergern DR, Chenoweth JG, et al. (2015) Practical impacts of genomic data "cleaning" on biological discovery using surrogate variable analysis. BMC Bioinformatics 16: 372. doi: 10.1186/s12859-015-0808-5 PMID: 26545828

64. Krzywinski M, Altman N (2014) Points of significance: Analysis of variance and blocking. Nat Methods 11: 699–700. PMID: 25110779

65. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, et al. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 12: 115–121. doi: 10.1038/nmeth.3252 PMID: 25633503

66. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140. doi: 10.1093/bioinformatics/btp616 PMID: 19910308

67. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550. doi: 10.1186/s13059-014-0550-8 PMID: 25516281

68. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15: R29. doi: 10.1186/gb-2014-15-2-r29 PMID: 24485249

69. Rapaport F, Khanin R, Liang Y, Pirun M, Krek A, et al. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol 14: R95. doi: 10.1186/gb-2013-14-9-r95 PMID: 24020486

70. Soneson C, Delorenzi M (2013) A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics 14: 91. doi: 10.1186/1471-2105-14-91 PMID: 23497356

71. Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform 16: 59–70. doi: 10.1093/bib/bbt086 PMID: 24300110

72. Schurch NJ, Schofield P, Gierlinski M, Cole C, Sherstnev A, et al. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? RNA 22: 839–851. doi: 10.1261/rna.053959.115 PMID: 27022035

73. Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, et al. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol 14: 153–165. doi: 10.1038/nrm3525 PMID: 23385723

74. Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8: 469–477. doi: 10.1038/nmeth.1613 PMID: 21623353

75. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562–578. doi: 10.1038/nprot.2012.016 PMID: 22383036

76. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, et al. (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol 33: 243–246. doi: 10.1038/nbt.3172 PMID: 25748911

77. Gene Ontology C (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res 43: D1049–1056. doi: 10.1093/nar/gku1179 PMID: 25428369

78. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44: D457–462. doi: 10.1093/nar/gkv1070 PMID: 26476454

79. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, et al. (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 44: D471–480. doi: 10.1093/nar/gkv1164 PMID: 26527732

80. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, et al. (2013) InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res 41: D1228–1233. doi: 10.1093/nar/gks1147 PMID: 23180781

81. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, et al. (2011) Molecular signatures database (MSigDB) 3.0. Bioinformatics 27: 1739–1740. doi: 10.1093/bioinformatics/btr260 PMID: 21546393

82. Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol 11: R14. doi: 10.1186/gb-2010-11-2-r14 PMID: 20132535

83. Rahmatallah Y, Emmert-Streib F, Glazko G (2016) Gene set analysis approaches for RNA-seq data: performance evaluation and application guideline. Brief Bioinform 17: 393–407. doi: 10.1093/bib/bbv069 PMID: 26342128

84. De Smet R, Marchal K (2010) Advantages and limitations of current network inference methods. Nat Rev Microbiol 8: 717–729. doi: 10.1038/nrmicro2419 PMID: 20805835

85. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, et al. (2012) Wisdom of crowds for robust gene network inference. Nat Methods 9: 796–804. doi: 10.1038/nmeth.2016 PMID: 22796662

86. Schmitt WA Jr., Raab RM, Stephanopoulos G (2004) Elucidation of gene interaction networks through time-lagged correlation analysis of transcriptional data. Genome Res 14: 1654–1663. doi: 10.1101/gr.2439804 PMID: 15289483

87. Schulze S, Henkel SG, Driesch D, Guthke R, Linde J (2015) Computational prediction of molecular pathogen-host interactions based on dual transcriptome data. Front Microbiol 6: 65. doi: 10.3389/fmicb.2015.00065 PMID: 25705211

88. Schulze S, Schleicher J, Guthke R, Linde J (2016) How to Predict Molecular Interactions between Species? Front Microbiol 7: 442. doi: 10.3389/fmicb.2016.00442 PMID: 27065992

89. Chertow DS, Memoli MJ (2013) Bacterial coinfection in influenza: a grand rounds review. JAMA 309: 275–282. doi: 10.1001/jama.2012.194139 PMID: 23321766

90. Prusty BK, Bohme L, Bergmann B, Siegl C, Krause E, et al. (2012) Imbalanced oxidative stress causes chlamydial persistence during non-productive human herpes virus co-infection. PLoS One 7: e47427. doi: 10.1371/journal.pone.0047427 PMID: 23077614

91. Prusty BK, Siegl C, Hauck P, Hain J, Korhonen SJ, et al. (2013) Chlamydia trachomatis infection induces replication of latent HHV-6. PLoS One 8: e61400. doi: 10.1371/journal.pone.0061400 PMID: 23620749

92. Deng JC (2013) Viral-bacterial interactions-therapeutic implications. Influenza Other Respir Viruses 7 Suppl 3: 24–35.

93. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, et al. (2011) Phenotypic landscape of a bacterial cell. Cell 144: 143–156. doi: 10.1016/j.cell.2010.11.052 PMID: 21185072

94. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6: 279–289. doi: 10.1016/j.chom.2009.08.003 PMID: 19748469

95. Kim SH, Chi M, Yi B, Kim SH, Oh S, et al. (2014) Three-dimensional intestinal villi epithelium enhances protection of human intestinal cells from bacterial infection by inducing mucin expression. Integr Biol (Camb) 6: 1122–1131.

96. Chen Y, Lin Y, Davis KM, Wang Q, Rnjak-Kovacina J, et al. (2015) Robust bioengineered 3D functional human intestinal epithelium. Sci Rep 5: 13708. doi: 10.1038/srep13708 PMID: 26374193

97. Sheppard M, Webb C, Heath F, Mallows V, Emilianus R, et al. (2003) Dynamics of bacterial growth and distribution within the liver during Salmonella infection. Cell Microbiol 5: 593–600. PMID: 12925129

98. Damron FH, Oglesby-Sherrouse AG, Wilks A, Barbier M (2016) Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia. Sci Rep 6: 39172. doi: 10.1038/srep39172 PMID: 27982111

99. Nuss AM, Beckstette M, Pimenova M, Schmuhl C, Opitz W, et al. (2017) Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. Proc Natl Acad Sci U S A. E-pub ahead of print.

100. Ackermann M (2015) A functional perspective on phenotypic heterogeneity in microorganisms. Nat Rev Microbiol 13: 497–508. doi: 10.1038/nrmicro3491 PMID: 26145732

101. Stewart MK, Cookson BT (2012) Non-genetic diversity shapes infectious capacity and host resistance. Trends Microbiol 20: 461–466. doi: 10.1016/j.tim.2012.07.003 PMID: 22889945

102. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498: 236–240. doi: 10.1038/nature12172 PMID: 23685454

103. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 510: 363–369. doi: 10.1038/nature13437 PMID: 24919153

104. Ramskold D, Luo S, Wang YC, Li R, Deng Q, et al. (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 30: 777–782. doi: 10.1038/nbt.2282 PMID: 22820318

105. Trombetta JJ, Gennert D, Lu D, Satija R, Shalek AK, et al. (2014) Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Curr Protoc Mol Biol 107: 4 22 21–17.

106. Armour CD, Castle JC, Chen R, Babak T, Loerch P, et al. (2009) Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nat Methods 6: 647–649. doi: 10.1038/nmeth.1360 PMID: 19668204

107. Qin Y, Yao J, Wu DC, Nottingham RM, Mohr S, et al. (2016) High-throughput sequencing of human plasma RNA by using thermostable group II intron reverse transcriptases. RNA 22: 111–128. doi: 10.1261/rna.054809.115 PMID: 26554030

108. Nottingham RM, Wu DC, Qin Y, Yao J, Hunicke-Smith S, et al. (2016) RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. RNA 22: 597–613. doi: 10.1261/rna.055558.115 PMID: 26826130

109. Kang Y, Norris MH, Zarzycki-Siek J, Nierman WC, Donachie SP, et al. (2011) Transcript amplification from single bacterium for transcriptome analysis. Genome Res 21: 925–935. doi: 10.1101/gr.116103.110 PMID: 21536723

110. Kang Y, McMillan I, Norris MH, Hoang TT (2015) Single prokaryotic cell isolation and total transcript amplification protocol for transcriptomic analysis. Nat Protoc 10: 974–984. doi: 10.1038/nprot.2015.058 PMID: 26042386