UC Riverside
UC Riverside Previously Published Works

Title

Analyzing RNA-Seq data from Chlamydia with super broad transcriptomic activation: challenges, solutions, and implications for other systems.

Permalink

Journal

Authors

Wan, Danny
Cheng, Andrew
Wang, Yuxuan
et al.

Publication Date

DOI

Peer reviewed

Analyzing RNA-Seq data from *Chlamydia* with super broad transcriptomic activation: challenges, solutions, and implications for other systems

Danny Wan[1], Andrew Cheng[1], Yuxuan Wang[1], Guangming Zhong[2], Wei Vivian Li[3*] and Huizhou Fan[1*]

Abstract

Background RNA sequencing (RNA-Seq) offers profound insights into the complex transcriptomes of diverse biological systems. However, standard differential expression analysis pipelines based on DESeq2 and edgeR encounter challenges when applied to the immediate early transcriptomes of *Chlamydia* spp., obligate intracellular bacteria. These challenges arise from their reliance on assumptions that do not hold in scenarios characterized by extensive transcriptomic activation and limited repression.

Results Standard analyses using unique chlamydial RNA-Seq reads alone identify nearly 300 upregulated and about 300 downregulated genes, significantly deviating from actual RNA-Seq read trends. By incorporating both chlamydial and host reads or adjusting for total sequencing depth, the revised normalization methods each detected over 700 upregulated genes and 30 or fewer downregulated genes, closely aligned with observed RNA-Seq data. Further validation through qRT-PCR analysis confirmed the effectiveness of these adjusted approaches in capturing the true extent of transcriptomic activation during the immediate early phase of chlamydial infection.

Conclusions This study highlights the limitations of standard RNA-Seq analysis tools in scenarios with extensive transcriptomic activation, such as in *Chlamydia* spp. during early infection. Our revised normalization methods, incorporating host reads or total sequencing depth, provide a more accurate representation of gene expression dynamics. These approaches may inform similar adjustments in other systems with unbalanced gene expression dynamics, enhancing the accuracy of transcriptomic analysis.

*Correspondence:
Wei Vivian Li
weil@ucr.edu
Huizhou Fan
fanhu@rwjms.rutgers.edu
[1]Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
[2]Department of Microbiology and Immunology, University of Texas Health San Antonio, San Antonio, TX 78229, USA
[3]Department of Statistics, University of California Riverside, Riverside, CA 92521, USA

Background

RNA sequencing (RNA-Seq) technology has revolutionized the field of molecular biology, offering unprecedented insights into the complexity of transcriptomes across a wide array of biological systems [1, 2]. This high-throughput technology not only facilitates the quantification of gene expression but also enables the detection of novel transcripts and post-transcriptional modifications. The comprehensive nature of RNA-Seq data has been instrumental in elucidating the regulatory mechanisms underpinning growth, development, and response to environmental changes.

Central to the processing and analysis of RNA-Seq data are bioinformatic pipelines. By analyzing the quality of sequencing data, aligning sequencing reads to reference genomes or transcriptomes, quantifying gene or transcript expression levels, and performing differential expression analysis, these tools enable researchers to decipher complex transcriptomic data and identify gene expression patterns critical for biological functions [2–4].

A critical aspect of RNA-Seq analysis pipelines is data normalization [5, 6]. This process aims to distinguish true biological differences in gene expression from technical artifacts (e.g., variation in sequencing depth). By removing technical variability and adjusting for differences in sequencing depth and RNA composition, normalization allows researchers to directly compare transcript abundance without the confounding effects of sample-to-sample variations in library size.

DESeq2 and edgeR are the two most-widely used bioinformatics tools designed for analyzing differential gene expression from RNA-Seq data [7, 8]. They employ size factor normalization and trimmed mean of M-values normalization, respectively, to adjust for library size differences and RNA composition. Both methods assume that, across different conditions, the differential expression of the majority of genes is either negligible or symmetrically balanced between up- and down-regulation, which is true in most biological systems.

Chlamydia is an obligate intracellular bacterium that undergoes a unique developmental cycle alternating between the elementary body (EB) and reticulate body (RB) [9]. EBs, adapted for extracellular survival and host cell infection, exhibit limited metabolic activity due to non-conducive external environments. In contrast, RBs replicate within host cells, eventually producing progeny EBs [10]. This cycle suggests that the primary EB-to-RB differentiation and the secondary RB-to-EB differentiation involve extensive transcriptomic activation and repression, respectively. Such broad activation and repression challenge the assumptions underlying standard normalization techniques used for transcriptomic research.

In our recent study aimed at understanding transcriptomic changes during the first hour of infection [11], DESeq2 analysis of only chlamydial RNA-Seq reads failed to accurately capture transcriptomic changes during the immediate early phase (i.e., the first hour of infection). To address this, we integrated mapped host RNA-Seq reads for normalization. However, the necessity and effectiveness of this strategy were not explicitly discussed in that report.

In this report, we build upon our previous findings by further illustrating the necessity of using mitigation strategies with either DESeq2 or edgeR analysis. By incorporating mapped host RNA-Seq reads or using overall sequencing depth for read count normalization, we address the limitations of these tools when analyzing systems with unbalanced gene expression dynamics. These adjustments provide a more accurate representation of gene expression dynamics, as validated by quantitative reverse transcription PCR (qRT-PCR) analysis. This highlights the importance of these integration strategies in accurately interpreting early-stage transcriptomic dynamics in Chlamydia. Additionally, we propose potential strategies for analyzing systems with similar transcriptomic activation or repression challenges where host RNA-Seq reads are not available, offering broader implications for research on free-living microbes and other isolated biological systems.

Methods

RNA-Seq dataset

The RNA-Seq datasets analyzed in this study were obtained from two independent RNA-Seq studies are available under NCBI Gene Expression Omnibus accession number GSE248988 [11]. EB stocks used for the RNA-Seq studies were prepared from mouse L929 cells [12]. Briefly, chlamydiae were first purified by ultracentrifugation through a 35% MD-76R gradient. EBs were further separated from RBs by ultracentrifugation through a 40%/44%/52% MD-76R gradient [13]. The multiplicity of infection (MOI) was 50 inclusion-forming units per host cell in the first RNA-Seq study and 200 inclusion-forming units per host cell in the second study. The high infectious doses were necessary to generate sufficient transcriptome coverage [11]. To synchronize infection, cells were centrifuged at 900 X *g* at room temperature for 10 min. EBs in the medium and on the cell surface but not committed to entry were removed by three washes with Hank's balanced salt solution containing heparin. Triplicate cultures were harvested at the end of last wash (i.e., 0 hpi). Another set of triplicate samples were harvested following 1 h at 37 ℃. The infectivity of the EB stocks was performed by infecting a separate plate of cells at the MOI of 1 inclusion-forming unit per cell, which resulted in the detection of inclusions in 80% of host cells

the following day. The trend of transcriptomic changes of the 50-MOI and 200-MOI studies were similar [11]. This study was conducted using RNA-Seq reads pooled from the two studies.

DESeq2 and edgeR analyses

The reads for each sample were first analyzed using FastQC (version 0.12.1) [14]. Trimming of adapter sequences and removal of low-quality reads were performed using Trimmomatic (version 0.38) [15]. Short read sequences were aligned to the CtL2 434/Bu genome including the chromosome (GCF_000068585.1_ASM6858v1) and the pL2 plasmid (AM886278) using TopHat (version 2.1.1) [16] and then quantified for gene expression by HTSeq (version 0.13.5) [17] to obtain raw read counts per gene. For the second normalization approach to be introduced below, the read sequences were also aligned to the mouse genome (GCF_000001635.27) using TopHat and qualified using HTSeq.

DESeq2 (version 1.36.0) [7] and edgeR (version 3.38.4) [8] were used to carry out the differential expression analysis and identify differentially expression genes between 0 and 1 hpi based two criteria: adjusted P value (by Benjamini-Hochberg method) $\leq$ 0.05 and fold change $\geq$ 2. When RNA-Seq data normalization was performed using unique chlamydial reads alone, the default analysis pipeline in DESeq2 and edgeR was used. When normalization was performed using unique chlamydial and host reads, we replaced the library size factors in both methods with the recalculated ones using the approach below. First, we estimated the library size as the total number of unique chlamydial and host reads in each sample. Second, we calculated the library size factor of a sample as its library size divided by the geometric mean across all samples. When normalization was performed using sequencing depth, we recalculated the library size factors as the sequencing depth of a sample divided by the geometric mean of sequencing depth across samples.

Quantitative reverse transcription real-time PCR (qRT-PCR)

qRT-PCR was performed using QuantStudio 5 real-time PCR System (Thermo Fisher Bioscientific) and Luna Universal one-step qRT-PCR kit (New England BioLabs) as previously described [12]. *t*-tests were conducted using

Microsoft Office Excel to identify differentially expressed genes from the qRT-PCR results. *P* values were adjusted for multiple comparisons by Benjamini-Hochberg procedure to control the false discovery rate [18].

Results

Discrepancy between unnormalized and normalized differential expression analyses of immediate early transcriptomic changes of *C. trachomatis* infection

We analyzed RNA-Seq data from cultures of *C. trachomatis*-infected mouse fibroblast cells at 0 and 1 h post-inoculation (hpi) [11]. The mapped RNA-Seq read data, as summarized in Table 1, reveal a notable 6-fold increase in the percentage of uniquely mapped *C. trachomatis* reads from the total combined reads of chlamydial and mouse origins, rising from 0.9 to 5.4% from 0 to 1 hpi (Table 1). This increase translates to a 6.3-fold rise in *C. trachomatis* transcriptome coverage, from 17.1- to 107.7-fold (Table 1). Among the 980 *C. trachomatis* genes, 792 (80.8%) exhibited a ≥2-fold increase in read counts from 0 to 1 hpi, while only 33 genes (3.4%) showed a ≥2-fold decrease (Table 2). These observations signify a broad activation of the *C. trachomatis* transcriptome during the immediate early developmental phase, despite a constant number of chlamydial cells.

This extensive transcriptomic activation, alongside limited repression, challenges the standard assumptions underlying DESeq2 and edgeR analyses because these tools generally assume that most genes do not exhibit significant differential expression across conditions or that the numbers of upregulated and downregulated genes are balanced [7, 8]. When analyses were confined to *C. trachomatis* reads, both DESeq2 and edgeR failed to accurately capture the transcriptomic dynamics indicated by the read counts. Specifically, DESeq2 identified only 305 genes (31.1%) with ≥2-fold upregulation and 303 genes (30.9%) with downregulation from 0 to 1 hpi (Table 2). Similarly, edgeR recognized 349 upregulated genes (35.6%) and 271 downregulated genes (27.7%) (Table 2). The numbers of upregulated genes are less than half of those suggested by direct read counts, while the reported downregulated genes are nearly 10-fold more (Table 2). Notably, DESeq2 identified 228 genes as downregulated despite their read count showing a ≥2-fold increase, while edgeR reported 111 such genes. This discrepancy

Table 1 Summary of RNA-Seq reads obtained from cultures of *Chlamydia trachomatis* (Ct)-infected mouse cells

	0 hpi				1 hpi			
	Replicate 1	Replicate 2	Replicate 3	Average	Replicate 1	Replicate 2	Replicate 3	Average
Reads uniqely mapped to Ct genes	261,249	388,612	319,552	323,138	1,744,834	1,582,744	2,786,959	2,038,179
Fold Ct transcriptome coverage (depth)	13.8	20.5	16.9	17.1	92.2	83.7	147.3	107.7
Reads uniqely mapped to mouse genes	29,157,294	36,976,293	35,248,060	33,793,882	35,537,308	30,667,452	40,757,064	35,653,941
Uniquely mapped Ct reads/total reads (%)	0.9	1.0	0.9	0.9	4.7	4.9	6.4	5.4

The 945,817 bp *C. trachomatis* transcriptome was used to calculate transcriptome coverage

Table 2 Numbers of differentially expressed genes identified with direct RNA-Seq read counting or bioinformatic programs, DESeq2 or edgeR, using uniquely mapped ct reads, unique mapped ct and host reads, or ct reads with sequencing depth normalization

		Unique Ct reads only	Normalization with unique Ct and host reads	Normalization with sequencing depth
RNA-Seq read counting	Upregulated	792	NA	NA
	Downregulated	33	NA	NA
DESeq2	Upregulated	305	717	718
	Downregulated	303	26	25
edgeR	Upregulated	349	727	746
	Downregulated	271	30	26

Genes with expression changes of ≥ 2-fold expression change from 0 to 1 hpi ($P < 0.05$) are defined as DEGs

highlights the limitations of using DESeq2 or edgeR to analyze *C. trachomatis* RNA-Seq reads in isolation for accurately reporting transcriptomic changes from 0 to 1 hpi.

Improved transcriptomic predictions following inclusion of host RNA-Seq reads

Given that host mouse reads constitute 95 to 99% of all uniquely mapped reads and considering that extensive changes in the host transcriptome from 0 to 1 h post-inoculation (hpi) were not anticipated [19, 20], we hypothesized that incorporating host read counts into the normalization step of DESeq2 and edgeR analyses would yield a more accurate report of chlamydial transcriptomic changes during this period. This hypothesis was based on the premise that, despite the transcriptomic activation occurring within chlamydial cells, incorporating the vast majority of host reads enables a more precise estimation of the true library size.

By including mapped host RNA-Seq read counts in the analysis (see Methods for details), the number of genes identified as upregulated by DESeq2 rose from 305 to 717, and the count of downregulated genes sharply declined from 303 to 26. EdgeR analysis showed a similar trend, with upregulated genes increasing from 349 to 727, and downregulated genes decreasing from 271 to 30 (Table 2). These significant revisions in gene regulation patterns align closely with the read count trends observed, affirming the broad activation of the chlamydial transcriptome during the immediate early phase of infection.

### Improved transcriptomic predictions through sequencing depth normalization

As an alternative to including aligned host reads in the DESeq2 and edgeR analyses, we utilized sequencing depth (total read counts) to calculate the library size factor and perform normalization (Methods). This method does not specifically rely on uniquely mapped mouse reads and offers a computationally more efficient approach to account for the sample's overall complexity.

The DESeq2 analysis with sequencing depth normalization resulted in the identification of one additional upregulated gene and one fewer downregulated gene compared to the analysis using all uniquely mapped counts (Table 2). The edgeR analysis with sequencing depth normalization also demonstrated a high consistency with the previous analysis, revealing 19 additional upregulated genes and four fewer downregulated genes. Results of expression analysis for all 980 *C. trachomatis* genes at 0 and 1 hpi determined by different strategies are presented in Fig. 1, which suggests that both strategies, using uniquely mapped *C. trachomatis* and host reads or the overall sequencing depth for read count normalization, are effective in accurately identifying differentially regulated genes.

Validation of broad transcriptomic activation by qRT-PCR analysis

To validate the RNA-Seq findings derived from the two normalization approaches, by uniquely mapped reads or the sequencing depth, we conducted qRT-PCR analysis on a subset of *C. trachomatis* genes. Specifically, we analyzed 15 genes out of approximately 350 that appeared unchanged when analyzed using only chlamydial reads but were identified as upregulated through the two proposed normalization approaches. Additionally, we tested 5 genes that were initially categorized as downregulated in the analysis using only chlamydial reads but were recognized as upregulated upon the inclusion of mapped host reads or overall sequencing depth.

The qRT-PCR analysis revealed that 14 of the 15 genes tested exhibited more than a 2-fold upregulation from 0 to 1 hpi (with statistical significance), while the other gene displayed nearly a 1.6-fold increase (Fig. 2; Supplementary Table 1). Notably, none of the genes showed downregulation. These findings confirm the effectiveness of including mapped host reads and applying library size normalization in accurately identifying transcriptomic changes during the immediate early phase of *C. trachomatis* infection.
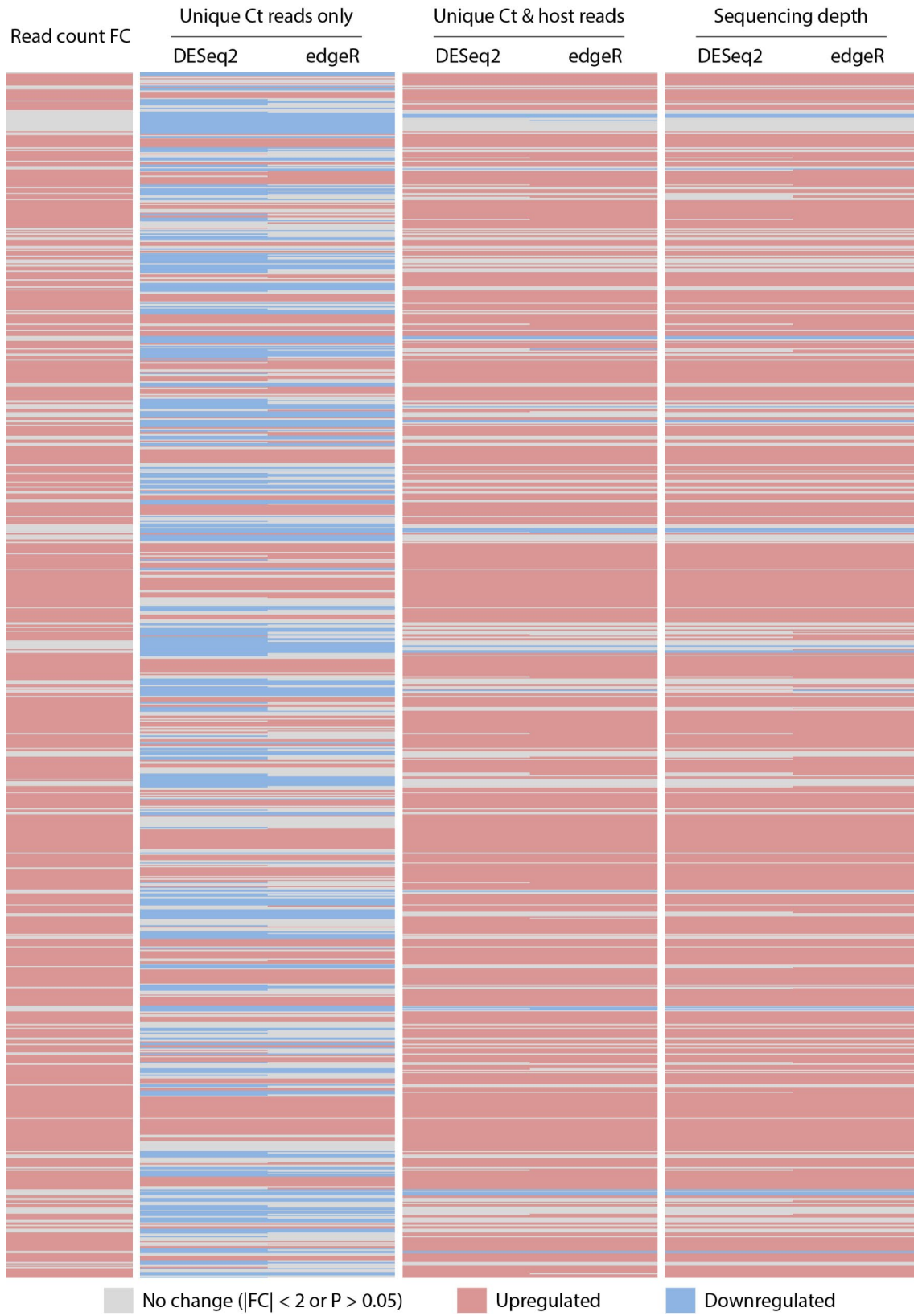
Fig. 1 Expression changes of all 980 *C. trachomatis* genes during the first hour of infection determined by different methods. Increases and decreases in the first column were defined based on fold change of read counts. Abbreviation: FC, fold change

Unique Ct reads only		Unique Ct & host reads		Sequencing depth		qRT-PCR
DESeq2	edgeR	DESeq2	edgeR	DESeq2	edgeR	
1.56	1.90	22.21	20.31	23.88	23.38	8.44
1.47	1.78	21.53	19.48	23.17	22.45	4.18
1.34	1.63	19.12	17.24	20.60	19.90	27.02
1.25	1.51	18.41	16.14	19.89	18.84	7.01
1.17	1.43	18.00	16.54	19.28	18.95	10.18
1.01	1.22	15.42	13.91	16.58	16.04	4.08
-1.08	1.12	13.33	12.01	14.36	13.88	6.95
-1.13	1.08	12.47	11.35	13.42	13.10	11.33
-1.14	1.07	12.41	11.18	13.38	12.97	11.80
-1.27	-1.03	11.06	9.97	11.91	11.51	13.57
-1.39	-1.15	9.13	8.15	9.80	9.31	4.43
-1.49	-1.21	9.40	8.58	10.11	9.88	4.45
-1.58	-1.31	8.83	8.18	9.44	9.31	11.78
-1.70	-1.42	7.92	6.80	8.59	7.99	6.10
-1.95	-1.59	7.16	6.56	7.69	7.54	14.40
-2.27	-1.86	5.96	5.57	6.39	6.37	9.04
-2.32	-1.90	6.23	5.57	6.73	6.48	33.44
-3.13	-2.58	4.62	4.20	4.97	4.84	15.03
-4.19	-3.45	3.58	3.19	3.86	3.71	2.52
-4.72	-3.91	3.18	2.86	3.43	3.31	1.60

No change (|FC| < 2 or P > 0.05)   Upregulated   Downregulated

Fig. 2 Comparison of differential expression analysis based on RNA-Seq and qRT-PCR data. Displayed values are the fold expression changes calculated by DESeq2 or edgeR based on three normalization methods, using unique Ct reads only, unique Ct and host reads, or the overall sequencing depth, and the fold expression changes calculated based on qRT-PCR experiments. Positive values indicate increased expression and negative values indicate decreased expression from 0 to 1 hpi

Discussion

This study underscores the critical role of context in transcriptomic analyses, particularly for obligate intracellular bacteria like *Chlamydia*. Our findings demonstrate that standard RNA-Seq normalization methods, such as those employed by DESeq2 and edgeR by default, which rely on assumptions of minimal or balanced differential expression across conditions, fall short in accurately capturing the dynamics of transcriptomic changes during the immediate early phase of *C. trachomatis* infection. The significant discrepancy between RNA-Seq counts and differential expression analyses highlights the need for methodological adjustments to account for the extensive transcriptomic activation and limited repression observed in *C. trachomatis*.

The modified normalization strategies based on both chlamydial and host reads or the overall sequencing depth emerged as effective strategies to overcome these analytical challenges. By integrating host reads or

adjusting for total sequencing depth, we were able to align the differential expression analysis more closely with the actual RNA-Seq read trends, validating the extensive activation of the chlamydial transcriptome.

These normalization strategies are particularly suitable for analyzing immediate early and early transcriptomes because, during these phases, host RNA dominates and the number of chlamydiae remain unchanged. Additionally, these strategies can also be applied to analyzing mid-cycle transcriptomes perturbed by transcriptional regulators or other factors, provided host RNA dominate and the chlamydial genome copy number is not significantly affected within the time frame of analysis.

To determine if an alternative normalization strategy that we have utilized for this study is needed for mid-cycle transcriptomic studies, it is essential to first assess whether the transcriptomic changes are limited (based on the numbers of host and chlamydial reads) and balanced (based on the number of up- and down-regulated genes

estimated with individual chlamydial gene reads). If the genome copy number is affected, further studies should include multiple groups with narrow time or treatment dosage increments to ensure the genome copy number remains relatively stable between adjacent groups. This approach would allow for the accurate detection of transcriptomic trends through serial analysis. Additionally, spiking-in synthetic transcripts into RNA-samples prior to RNA-Seq library construction [21] could serve as an alternative strategy.

The secondary RB-to-EB differentiation involves extensive transcriptomic repression and limited activation. Because this differentiation is asynchronous and the amount of host RNA reads no longer dominates, developing alternative strategies will be necessary to analyze late-cycle transcriptomic data effectively.

Although our normalization strategies have been developed in the context of *C. trachomatis* infection, they offer a blueprint for analyzing other biological systems with similar challenges. One such analogous system is the germination of bacterial spores, a process characterized by dramatic transcriptomic changes as dormant spores become metabolically active vegetative cells. Like the immediate early transcriptomic activation in *C. trachomatis*, spore germination involves a rapid and broad activation of gene expression. In transcriptomic analysis of bacterial spore germination in environmental samples, we suggest that researchers perform total sequencing depth normalization to accurately identify differentially regulated genes. Additionally, if accompanying organisms in the environment undergo significant changes, increasing the number of experimental groups and spiking-in synthetic transcripts into RNA samples prior to RNA-Seq library construction [21] would be helpful.

Abbreviations
EB	Elementary body
RB	Reticulate body
RNA-Seq	RNA sequencing
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
MOI	Multiplicity of infection
DESeq2	Differential expression analysis for sequence count data 2
edgeR	Empirical analysis of digital gene expression data in R
FastQC	Fast quality control
HTSeq	High-throughput sequence analysis

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10714-3.

Supplementary Material 1

Author contributions
H.F., W.V.L., and G.Z. conceived the study. H.F. and W.V.L wrote the manuscript text. DW, AC, WVL, and H.F. prepared Table 1 and 2, Figs. 1 and 2, and Table S1. 1. Y.W. contributed the qRT-PCR data in Fig. 2 and Table S1.

Data availability
The RNA-Seq datasets analyzed in this study are available under NCBI Gene Expression Omnibus accession number GSE248988. The code for reproducing the presented bioinformatic analysis is available at https://zenodo.org/records/11201379.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
2. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621–8.
3. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. Biol Direct. 2009;4(1):14.
4. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17(1):13.
5. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010;11(1):94.
6. Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2012;14(6):671–83.
7. Love M, Anders S, Huber W. Differential analysis of count data the DESeq2 package. Genome Biol. 2014;15(550):1–41.
8. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–40.
9. Abdelrahman YM, Belland RJ. The chlamydial developmental cycle. FEMS Microbiol Rev. 2005;29(5):949–59.
10. Rockey DD, Wang X, Debrine A, Grieshaber N, Grieshaber SS. Metabolic dormancy in Chlamydia trachomatis treated with different antibiotics. Infect Immun. 2024;92(2):e0033923.
11. Wurihan W, Wang Y, Yeung S, Zou Y, Lai Z, Fondell JD, Li WV, Zhong G, Fan H. Expression activation of over 70% of *Chlamydia trachomatis* genes during the first hour of infection. Infect Immun. 2024;92(3):e0053923.
12. Wurihan W, Zou Y, Weber AM, Weldon K, Huang Y, Bao X, Zhu C, Wu X, Wang Y, Lai Z et al. Identification of a GrgA-Euo-HrcA transcriptional regulatory network in *Chlamydia. mSystems* 2021, 6(4):e0073821.
13. Caldwell HD, Kromhout J, Schachter J. Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. Infect Immun. 1981;31(3):1161–76.
14. Andrews SB. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. *Available online at*: http://www.bioinformaticsbabrahamacuk/projects/fastqc/ 2010.

15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
16. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7(3):562–78.
17. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinf (Oxford England). 2015;31(2):166–9.
18. Benjamini Y, Hochberg Y. Controlling the false Discovery rate: a practical and powerful Approach to multiple testing. J Roy Stat Soc: Ser B (Methodol). 1995;57(1):289–300.
19. Humphrys MS, Creasy T, Sun Y, Shetty AC, Chibucos MC, Drabek EF, Fraser CM, Farooq U, Sengamalay N, Ott S, et al. Simultaneous transcriptional profiling of bacteria and their host cells. PLoS ONE. 2013;8(12):e80597.
20. Hayward RJ, Humphrys MS, Huston WM, Myers GSA. Dual RNA-seq analysis of in vitro infection multiplicity and RNA depletion methods in Chlamydia-infected epithelial cells. Sci Rep. 2021;11(1):10399–10399.
21. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 2011;21(9):1543–51.

Publisher's note