

SCIENTIFIC REPORTS



OPEN

Temporal dynamics in meta longitudinal RNA-Seq data

Sunghee Oh¹ , Congjun Li², Ransom L. Baldwin², Seongho Song³, Fang Liu^{2,4} & Robert W. Li²

Identification of differentially expressed genes has been a high priority task of downstream analyses to further advances in biomedical research. Investigators have been faced with an array of issues in dealing with more complicated experiments and metadata, including batch effects, normalization, temporal dynamics (temporally differential expression), and isoform diversity (isoform-level quantification and differential splicing events). To date, there are currently no standard approaches to precisely and efficiently analyze these moderate or large-scale experimental designs, especially with combined metadata. In this report, we propose comprehensive analytical pipelines to precisely characterize temporal dynamics in differential expression of genes and other genomic features, i.e., the variability of transcripts, isoforms and exons, by controlling batch effects and other nuisance factors that could have significant confounding effects on the main effects of interest in comparative models and may result in misleading interpretations.

Similar to microarrays, investigators have been increasingly conducting experiments that focus on ontological alterations across a series of time periods^{1–8}. Perhaps even more popular is the use of longitudinally repeated measurements at different time points in relation to some baseline stimuli or perturbation. Prior to the main downstream analyses, a prerequisite step must be the removal of experimental artifacts and unwanted sample-to-sample variation using appropriately proposed methods in pipelines. While this has long been recognized as an important step in the analysis of high-throughput data, it has largely been overlooked in the detection of significantly differential expression^{9–18}.

The purpose of this research is to develop data management procedures for the increasing wealth of data being generated by new approaches and to deepen the characterization of temporal dynamics by including isoform diversity in addition to gene-level analyses. In this study, we describe how to incorporate improved strategies to remove systematic biases and to fully characterize temporal dynamics by accounting for data-driven inherent features. This is based on our large-scale time course longitudinal stimuli-response data at every step, along with a panorama snapshot of the entire workflow (Supplementary Fig. S1).

Results

Exploratory and differential expression analysis in temporal dynamics. Our meta-framed longitudinal data have been sequenced twice in different sequencing dates and temporal expression levels have been measured over the 5 distinct time points, from Day 0 (D0) through Day 14 (D14). And 8 biological replicates at each time point have been utilized in this study. To characterize the complexity of temporal dynamics, our proposed Bayesian dynamic AR method with batch correction and isoform diversity, compared to existing static and other dynamic methods have been employed. A schematic illustration of the entire analytical strategy in the detailed analytical pipelines at each window is depicted in the section of 4. METHODS and in Supplementary Fig. S1. A description of the samples in the experimental design is also presented in the section of 4. METHODS, Fig. 1 and Dataset S1.

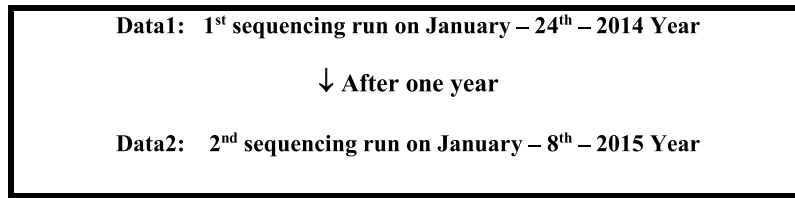
After pooling replicates for a biological individual sample at each time point. Based on the results where the first pipeline was used to analyze our raw data, without regard to the sophisticated diagnosis in the exploratory analyses (See Supplementary Fig. S1), we observed very low sensitivity for detection of temporally differentially expressed genes in Dataset S3-(1), despite the high quality of the sequencing samples in Dataset S1 and S2, and

¹Department of Computer Science and Statistics, Jeju National University,, Jeju City, Jeju Do, S., 690-756, Korea.

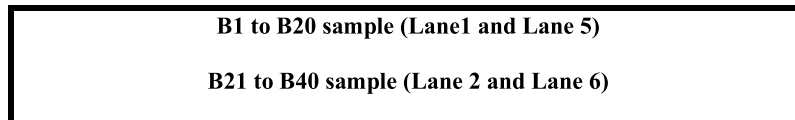
²United States Department of Agriculture, Agriculture Research Service (USDA-ARS), Animal Genomics and Improvement Laboratory, Beltsville, MD, 20705, USA. ³Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, 45221-0025, USA. ⁴College of Food Science and Engineering, Ocean University of China, Qingdao, 266003, China. Correspondence and requests for materials should be addressed to S.O. (email: sshshoh1105@gmail.com) or R.W.L. (email: robert.li@ars.usda.gov)

Experimental design performed in this study

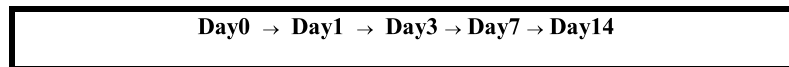
- Different sequencing runs (batch effect on different dates)



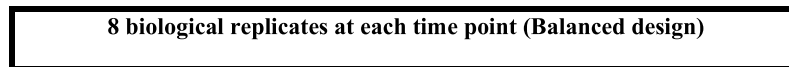
- Different lanes within a sequencing run



- Different time points



- Different biological replicates



- Different technical replicates

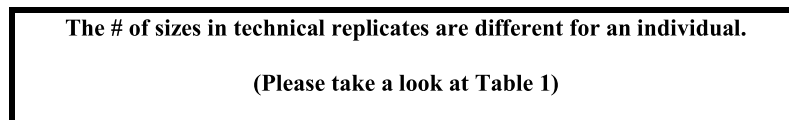


Figure 1. Description of experimental design performed in this meta study.

Fig. 2. These observations can be explained by several possibilities, including the impact of outlier samples, pooling issues without adjustment for batch effects on different sequencing dates or other experimental technical replicates, normalization issues, exploratory analysis-free preprocessing steps, and the selection of inappropriate differential expression methods, resulting in large confounding effects of batch and other experimental nuisance factors in comparative models between baseline time point versus later time points.

Hence, we excluded sample B29, which was highly suspected to be an outlier, as shown in Fig. 2, and we also employed dynamic specific methods for pooled data in analytical pipeline 2. The objective of this pipeline was to evaluate whether there was any significant effect of sophisticated exploratory analysis with sample diagnostic tools and the use of dynamic specific methods, compared to pipeline 1. Based on the normalization strategies in edgeR and DESeq^{13,19–21}, our data still showed the extra variability of samples, indicating that there could be other significant lurking factors in the data that would need further correction of systematic artifacts (Left panel of Fig. 3). In Fig. 2, after the removal of sample B29, we saw a distinguished pattern that was more likely grouped with samples within each lane than by time point, implying a substantial lane effect between samples (B1 to B20 versus B21 to B20) as is shown in the 2D multidimensional scaling plot. Furthermore, we confirmed the unwanted systematic artifacts due to different technical lanes by using a 3D movie of a principal component analysis plot (Supplementary Movie S1).

In this study, our experimental dataset is a meta-framed longitudinal time course with repeated measurements, i.e., a within-subject stimuli-response dataset. In principle, within-subject longitudinal stimuli-response data is applied to a Bayesian dynamic autoregressive model (AR), which more precisely targets the repeatedly measured time course RNA-Seq data, as was proposed in our previous study^{3,5}. Another dynamic-specific method, maSigPro, which is implemented in between-subject factorial time courses and is designed for the comparison of cases (single or multiple cases) versus control group over time, was also conducted for our within-subject time series data for a comparison. The latter method showed less sensitivity than our proposed Bayesian dynamic

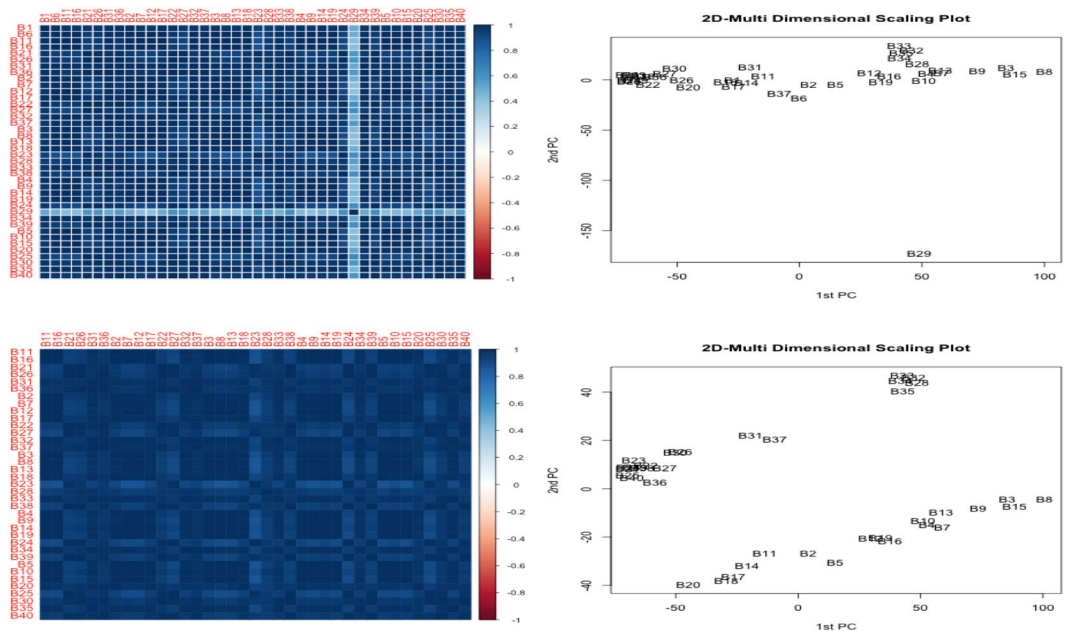


Figure 2. (1) and (2) represents correlation heatmap and multi-dimensional plot with B29 sample after pooling replicates of genes (3) and (4) represents correlation heatmap and multi-dimensional scaling plot without B29 sample after pooling replicates of genes, respectively.

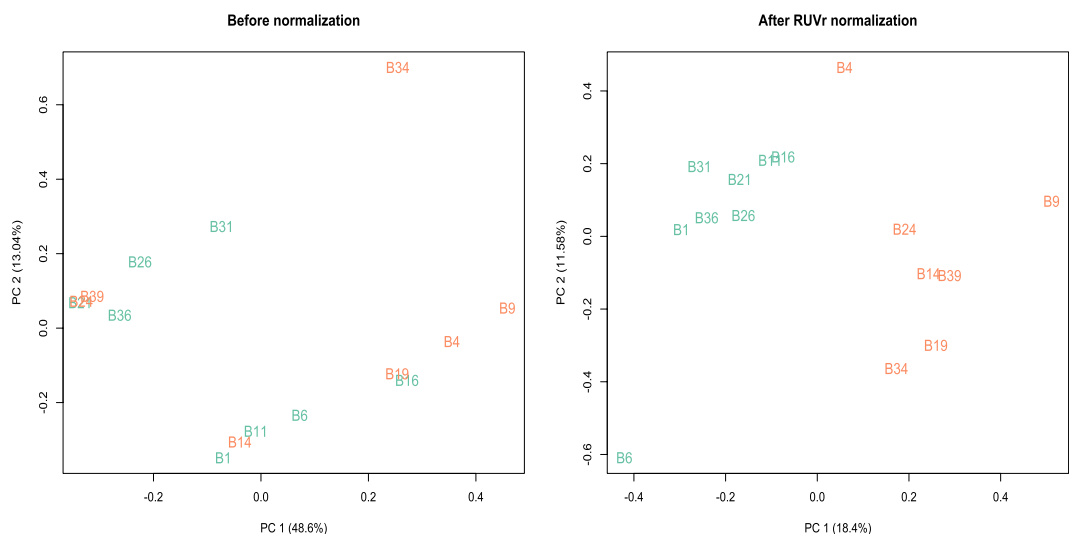


Figure 3. PCA plot after pooling replicates before and after correction of systematic biases in pairwise comparisons of time points (baseline time point D0 versus D7) without sample B29, respectively.

autoregressive model because our experimental setting is designed in a within-subject stimuli-response (data not shown).

Conclusively, when performing differential expression analysis with well-qualified samples after correction for systematic artifacts and making use of the desired dynamic method that was implemented for longitudinal data, we clearly detected more highly significant genes and isoforms that are insignificant in pipeline 1 due to the significant lurking factor (lane effect) in our experiment and confirmed in the multiple results (Dataset S3 and Supplementary Fig. S2). And more specifically, for Bayesian dynamic AR method, the detected genes (before correction of unwanted biases) are all included in those genes (after correction) as the nested sets (Supplementary Fig. S2-(3) to S2-(10)). In addition, the correction of systematic artifacts and filtering out sample B29 revealed temporal patterns in 147 and 251 genes (detected by edgeR and DESeq, respectively) between the stimulated D7 and control D0 groups in Supplementary Figs S2-(1) and S3, indicating a higher sensitivity in differential expression analysis with the improved analytical pipeline.

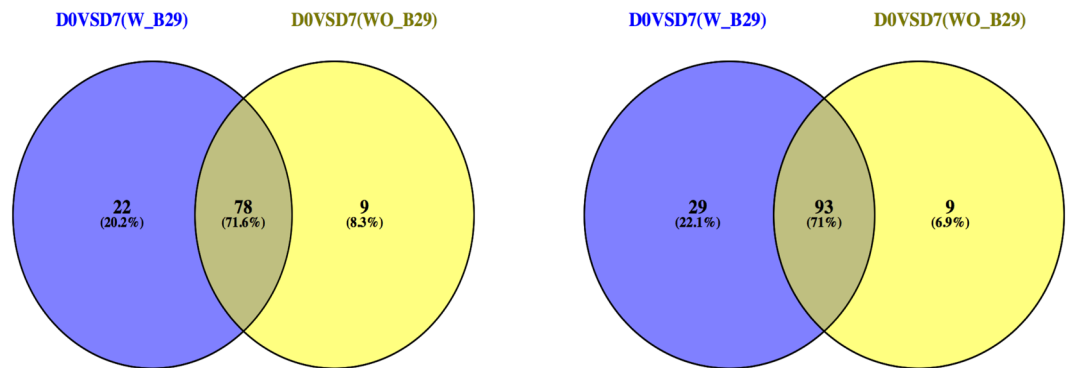


Figure 4. Comparison before correction of unwanted biases with and without B29 when comparing D0 versus D7 at the tail probability of 0.1 after pooling samples for the Bayesian dynamic AR model of genes and transcripts, respectively.

Interestingly, when compared to naïve static methods, the advantageous nature of our proposed Bayesian dynamic AR method has the capability to detect significantly differentially expressed genes before correction. This suggests that our Bayesian dynamic AR model explicitly captures the variability of replicates within a group and the extra variability due to the experimental systematic artifact of lane effect on the given data. In other words, as shown in Dataset S3-(2), S4-(2) and Fig. 4, most of the genes detected by static methods after correction are overlapped with the list of our dynamic method before correction.

Moreover, another advantageous property of our Bayesian dynamic AR model compared to other employed static methods in this study is that the majority of differentially expressed genes and isoforms are more often detected in highly expressed patterns than in zero and low-expressed noise sets. Generally, the Fisher exact pairwise static test, which is a robust simple pairwise static method, also tends to identify more likely highly expressed genes, as is confirmed in previous studies^{22,23}. Therefore, we confirmed in this study that the temporal dynamic genes significantly identified by the Bayesian dynamic AR model have mostly been included in the list from the static Fisher exact test, which was based on before-correction data (Dataset S3-(1) and Dataset S4).

Before pooling replicates for an individual biological sample at each time point. At the beginning of this study, we expected to find a significant batch effect between different sequencing dates (See the labels of samples in the batch column in Dataset S1). However, the variability between the different lanes was much greater than that of the batch effects of two different sequencing dates. For example, the correlation values among samples within each lane at the baseline time point, Day0, represent the overall very high reproducibility of greater than ~0.9, the values of the reproducibility of samples B1_L1_1 versus B21_L2_1, representing two different lanes at Day0, dramatically drops to less than ~0.5 (See Dataset S2 and Supplementary Fig. S4). This confirms the extra variability of samples due to the systematic artifact of lane effect between samples B1 to B20 versus B21 to B40. Compared to this explicit extra variability of samples in correlation metrics on unpooled data based on pipeline 3 (Supplementary Fig. S4), the lane effect was implicitly a lurking pattern in the pooled data (Fig. 2 and Dataset S2-(2)). That is, see the correlation heatmaps (after pooling replicates) in Fig. 2 and corresponding heatmaps (before pooling) in Supplementary Fig. S4 simultaneously.

In summary, our integrated meta longitudinal dataset shows that technical variability in library prep exerts a much greater effect compared to batch effects between different sequencing runs with a time-lag. This suggests that such unwanted nuisance factors should be removed prior to subsequent major analyses to reduce significant confounding effects on the main effects of interest and misleading results of statistical testing of the model^{14-18,24}. In the analyses of pipeline 3, we first examined the effect of the correction of systematic artifacts using the static methods, edgeR and DESeq. These methods showed less sensitivity for truly differential expression before the correction compared to certain sets of differential expression identified after correction. In particular, edgeR tends to more often identify falsely differentially expressed genes that have zero expression for most samples than does DESeq when analyzing data that contains unadjusted biases and unpooled replicates, as in Dataset S3-(3) and Supplementary Fig. S5. Regarding the impact of the outlier sample B29, overall the D0 versus D7 comparison with sample B29 identifies nearly three times more differentially expressed genes and isoforms than without this sample, indicating the higher false calls in differential expression analysis (Dataset S3-(3) and S3-(4)).

Importantly, for unpooled data, a biological sample is split into subdivided technical replicates with varied sizes between two consecutive time points (baseline versus stimulated group) (See the detailed different number of technical replicates in unpooled data from Dataset S1). Hence, it is not meaningful to systematically apply our proposed Bayesian dynamic AR model in longitudinal settings on unpooled data. However, pipeline 3 confirmed that the more significant effect of systematic unwanted biases due to lane effect than pooled data has been observed in Fig. 2 versus Supplementary Fig. S4. Additionally, there are currently no computational methods that incorporate those effects in lower-level preprocessing procedures prior to downstream analyses. Taken together, we simply pool technical replicates of multiple *.sam (or *.bam) files to increase the power of mapped read counts for an individual biological sample after correction of systematic artifacts for our proposed Bayesian dynamic AR method.

Gene Symbol	Fold change		Adjusted P value	
	RNA-Seq (AR model)	qRT PCR	RNA-Seq (AR model)	qRT PCR
BSP30C	1.05	0.92	0.0000	0.1655
FOS	2.79	2.99	0.0060	0.0011
S100A12	0.91	1.17	0.0000	0.0473
TGM1	1.20	1.54	0.0000	0.0000

Figure 5. qRT PCR validation for selected gold standard gene list that has been detected by Bayesian dynamic AR method, but not by other static methods.

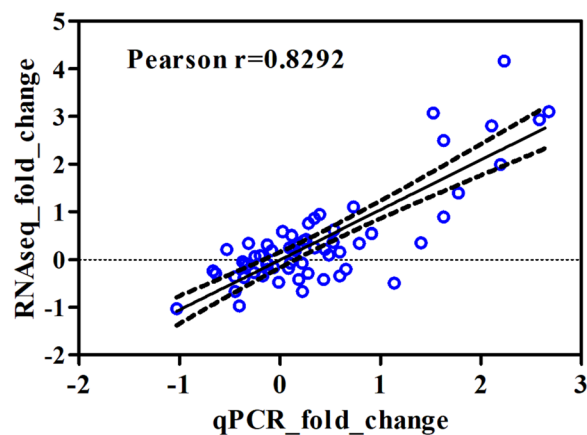


Figure 6. Pearson correlation value between the log₂-transformed fold changes of RNA-Seq data and qRT-PCR validation. Each dot (in blue) represents a data point for the selected genes and each corresponding sample (4 genes and 8 replicates in control (D0) and later time point, D1 and D14, respectively). And the dashed lines are 95% confidence interval.

Quantitative Real-time Reverse Transcription (RT) PCR. Our proposed Bayesian dynamic AR specific differential expression that has not been detected by other static and dynamic methods has been validated with qRT-PCR validation procedure whether those differential expression patterns are truly temporally differentially expressed genes in terms of biological perspectives or not. Two selected genes for each comparison (D0 versus D1 and D14, respectively) that have been uniquely identified using the AR model were validated independently by qRT-PCR (Fig. 5 and Dataset S7). Our selected genes confirmed the highly significant correlation between AR specific differential expression and validation results, suggesting that the higher sensitivity when using AR method in differential expression analysis and after correction of unwanted systematic artifacts that could be significant confounding effects in differential testing. For example, comparing to time zero (D0), a 14-day post-infusion (D14) induced a significant change in the mRNA expression of FOS ($P < 0.01$), in a good agreement with the AR-based RNaseq results. There existed a strong positive correlation between log₂ transformed fold changes calculated from the normalized hit counts (RNaseq) and those derived from the gene copy numbers detected by qRT-PCR (Correlation coefficient $R = 0.8292$, $P = 4.71 \times 10^{-17}$) as shown in Fig. 6. Moreover, we also further validated the gene list that has been identified by our proposed dynamic AR method and after correction of unwanted systematic biases with our previous study using simple pairwise static method²⁵ (Dataset S7). The highlighted genes represent the overlap between current and previous study.

Differential exon and isoform diversity in temporal dynamics. To identify various temporal patterns at the isoform level, we applied the most popular and robust methods, Cuffdiff and DEXSeq^{3,26–43}. Datasets S3-(1), S3-(2), S6-(1) and S6-(2) represent the analysis of differential expression between a reference time point versus later time points for pooled and unpooled data at the transcript isoform level to investigate the impact of correction (before- and after-correction), respectively.

Analogous to gene level analyses before the correction of the systematic artifact of lane effect, for both exon and isoform level analysis, relatively few significant sets were identified by the existing static Cuffdiff and DEXSeq methods, implicating suspected confounding effects (Dataset S3-(1) and Supplementary Movie S2). Therefore, we also performed differential expression at the exon and transcript levels by the static edgeR and DESeq methods and our dynamic AR model, which enables incorporation of other nuisance factors. Thus, after correction, we detected more significant sets of exons and transcripts (Datasets S3-(1), S3-(2), S4-(1) and S4-(3)). The analyses

of transcripts and individual exons for each gene demonstrated that, as expected in the complex architecture of gene-to-multiple isoforms, significant genes do not necessarily guarantee significance for all isoforms of those genes, and vice versa. Additionally, the magnitude of the expression level of a single gene is not similar for all of the corresponding multiple isoforms, implicating the variety of the expression spectrum for each isoform, with exon- and isoform-specific alterations. Thus, the complete features of the temporal dynamics of transcriptomes in gene regulation should be explored within gene-isoform families, because a parent (i.e., gene) is linked with multiple children (i.e., isoforms) to generate distinct functionalities and protein structures.

Gene set enrichment analysis and network modules. To explore the biological processes affected and functional mechanisms involved, we further conducted gene set enrichment analysis and network analysis on the temporal dynamic gene lists obtained from differential expression analyses^{7,44–50}. We carried out these functional analyses based on the data after correction of unwanted biases in both static and dynamic methods by pipeline 2, using pooled data (Dataset S5 and Supplementary Figs S6 and S7). Based on the dynamic data, the extracellular exosome and matrix, chromatin remodeling, DNA-dependent transcription, initiation, hemopoiesis, and RNA binding were significant for D0 versus D1. For the differentially expressed genes in D0 versus D14, the DNA-dependent transcription and initiation, detection of stimulus, nervous system development, and immune response terms were highly significant. Homopoiesis, chromatin remodeling, and Parkinson's disease were consistently detected for D0 versus D1 through D14.

Discussion

We have provided generalized analytical pipelines for meta-framed longitudinal RNA-Seq data in the above sections. It is of primary importance to correct for unwanted systematic artifacts due to the various types of experimental lurking factors, such as lane effects and other batch effects, in combined samples to increase the power of detection. Additionally, it is well known that aberrant patterns of splicing events are highly associated with tissue-, external condition-, and developmental stage-specific alterations^{3,26,28–30,33–35,37–40,42,43,51–56}. Thus, state-of-the-art pipelines facilitate a more complete landscape to better characterize the complexity of temporal dynamics by removing unwanted systematic biases, such as experimental artifacts, by incorporating multiple factors contained in experiments and by defining isoform diversity in our proposed Bayesian dynamic AR model for temporal dynamics.

Despite this importance to correct systematic artifacts in combined meta-framed data prior to major downstream analyses, such as differential expression analysis, in general, each method has unique input and output data, distinct normalization methods and rationale for differential expression analysis. It is timely that a universal analytical pipeline outlining a concrete plan for analytical steps should be proposed, as increasingly complicated datasets and combined metadata have become available in the postgenomic era. Indeed, this pattern will continue to grow with the continued rapid development of new technologies, and as an ever-greater amount of data is being generated, which is leading to a higher-order of data entropy.

Thus, it is certain that the issues have received little attention in more complicated experiments such as longitudinal within-subject and factorial between-subject stimuli-response data^{57–59}. Hence, our deeper level analyses of temporal dynamics of meta-framed RNA-Seq data with longitudinally measured stimuli-response measurements will accelerate the way in which to more precisely identify time-specific and condition-specific multiple external factors of stimuli, as well as tissue- or disease-specific differentially expressed genes and other genomic features. Our methods also provide new insights into analytical pipelines by defining biomarker candidates that are highly relevant to aberrant patterns of functionalities, network modules, and pathways in global temporal dynamics in a variety of clinical and experimental settings, including integrated meta time course data and disease spectrum progressive models.

There still many issues to be further addressed in a variety of more complex experimental designs, such as temporal RNA-Seq data, and especially in combined time course RNA-Seq datasets, in terms of the development of computational methods in the field of community. Examples include significant artifacts such as batch issues^{14–18}, sophisticated normalization^{12,60,61}, and several different types of experimental designs for time points^{2,5–8,14,16,57–59,62} (e.g., short, moderate, or long time series, or those containing multiple external factors at each time point, or not). Additional issues include whether the data is a single or multiseried factorial time course dataset, a circadian rhythmic periodical dataset, a single cell cycle dataset, or a meta-framed time course. Important is the establishment of a gold-standard benchmarking list in databases⁶³.

As an extension of this study, we are currently developing Bayesian dynamic models to define temporal dynamics for between-subject stimuli-response in factorized RNA-Seq time course experiments to better take into account systematic biases, multiple experimental factors, and isoform diversity in a model of differential expression in R and OpenBUGS (Winbugs) (implementation is underway). Another promising study in temporal dynamics would be to characterize a progressive disease model, such as pediatric cancer progression, by targeting initiation, progression time points and period, perturbation of cancer progression, and estimation and prediction of unobserved time points^{64–75}.

More specifically, complete analytical pipelines for noncoding RNAs (ncRNAs) are also needed, as aberrant patterns in (very) long and noncoding RNAs have been explored as important biomarkers for the classification of subtypes of cancers and other diseases, as primary factors in oncogenesis, and for therapeutic effects^{69,71,72,76}.

Methods

Autoregressive model (AR) incorporated with systematic artifacts and isoform levels. To precisely characterize temporal dynamics in within-subject stimuli-response longitudinal RNA-seq data in the format of a single series time course experiment, we employed an autoregressive model (AR) that had been initially proposed to account for the count property with Poisson gamma (negative binomial distribution) and time

dependency in our previous study⁷⁷. To identify differentially expressed genes between the baseline time point (D0) with ruminal infusion of butyrate versus later time points in the comparative tests, that is, D0 vs D1 (to D14), we computed the tail probability of $(\varphi_i|y_i)$ of gene i , such that $p(\varphi_i > 0|y_i)$ or $p(\varphi_i < 0|y_i)$ for $i = 1, \dots, n$ using Monte Carlo Markov Chain (MCMC) with 10,000 simulations and 8,000 burn-ins, where φ_i denotes the autocorrelation in the Poisson gamma AR model, and y_i represents the expression levels for i -th gene (see the further detailed explanation of our proposed model)⁷⁷. The cutoff value of 0.1 to tail probabilities in this model is selected properly and corresponds to a 0.1 false discovery rate (FDR) for other methods. It indicates the significance of differential expression for each gene-by-gene testing. This method is able to be straightforwardly extended to detect temporally differential expression for the quantification level of isoforms and other genomic features. Strikingly, our proposed dynamic model enables the inclusion of the factor of systematic biases from lane effects, which was estimated from preprocessing procedures in the differential expression analysis. Furthermore, this dynamic model has the capability to infer multiple factors in various experimental and clinical settings simultaneously, such as other additional nuisance factors resulting in unwanted systematic biases in a given dataset.

Comparison with other methods. Prior to down-stream analyses, the following diagnostic analyses on all samples for both the before and after pooling data were performed in the R package with the latest versions: gplots, RUVSeq⁷⁸, multivariate analyses, and Venny online tool. For comparison to our proposed temporal dynamics Bayesian AR method, differential expression analyses of genes, including pairwise comparisons, a generalized linear model incorporating multiple factors, and dynamic methods were carried out using Cuffdiff, the Tuxedo Suite, edgeR, DESeq, voom in limma, and maSigPro, in turn^{1,19,77,79–85}. For the quantification of isoforms (and exons) and the identification of differential expressions of splicing, Cuffdiff and DEXSeq were employed^{8,21,85,86}. For functional analyses, we employed NetworkAnalyst and David^{87,88}.

Experimental design and preprocessing analysis. *Experimental design.* Four Holstein cows were ruminally cannulated, as previously described^{31,89–94}. Cows in mid-lactation were fed ad libitum a total mixed ration consisting of 50% corn silage and 50% concentrate on a dry matter basis. The cows were moved to a tie stall barn for adaptation and acclimation 7 d prior to the experiment. A ruminal infusion of butyrate was initiated immediately following the 0 h sampling (baseline control) and thereafter continued for 168 h at a rate of 5.0 L/d of a 2.5 M solution (representing >10% of the daily anticipated metabolizable energy intake to support lactation) in a buffered saliva solution (pH 7.0; 3.8% KHCO₃, 7.3% NaHCO₃) as a continuous infusion. After 168 h of infusion, the infusion was stopped and the cows were maintained on the basal lactation ration for an additional 168 h for sampling. Rumen epithelial samples were serially collected via biopsy through the rumen fistulae at 0, 24, 72, and 168 h of infusion, and at 24 and 168 h post infusion (post 24 h and post 168 h, respectively). The ruminal pH was monitored using a standard pH meter and recorded at each sampling. Rumen epithelial samples were snap frozen in liquid nitrogen and stored at -80°C until RNA extraction (See Dataset S1). All animal care and handling was conducted according to the guidelines approved by the USDA Beltsville Area Institutional Animal Care Committee.

RNA extraction and sequencing using RNA-Seq. Total RNA was extracted from 24 rumen epithelial samples using Trizol (Invitrogen, Carlsbad, CA, USA), followed by DNase digestion and Qiagen RNeasy column purification (Qiagen, Valencia, CA, USA). The RNA integrity was verified using an Agilent Bioanalyzer 2100 (Agilent, Palo Alto, CA, USA). High-quality RNA (RNA integrity number (RIN) > 8.0) was processed using an Illumina TruSeq RNA sample prep kit following the manufacturer's instructions (Illumina, San Diego, CA, USA). After quality control procedures, individual RNA-seq libraries were pooled based on their respective sample-specific 6-bp adaptors and sequenced at 50 bp/sequence read using an Illumina HiSeq. 2000 sequencer, as previously described²⁵.

Preprocessing and exploratory analysis. All raw sequence fastq files were initially preprocessed against the reference genome, Bos_taurus.UMD3.1.80.gtf, downloaded from tools^{95,96}. In the sample raw data, biological replicates were collected from 8 different cell lines and technical replicates were run in different sequencing dates (Oct-24-2014 and Jan-08-2015) and in distinct lanes. The tuxedo method, bowtie, TopHat, and Cufflinks tools for trimming fastqc were performed using the latest version of tools. For the quantification of expression levels, there were two different types of the pipeline that were compared in this study: after and before pooling of the replicates. Hereafter, we refer to these as after and before pooling data. In the pipelines, the mapped expression FPKM levels quantified by Cufflinks were further utilized for the purpose of detecting temporally differentially expressed genes and isoforms. Prior to the identification of temporally expressed genes and isoforms, we employed exploratory analysis for all individual samples to verify sample reproducibility, variability, and unwanted systematic biases by making use of the diagnostic tools¹².

Quantitative Real-time Reverse Transcription (RT) PCR. Purified total RNA samples were converted to cDNA using an iScript Advanced cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA). *Quantitative Real-time Reverse Transcription (RT) PCR* analysis was performed using an Absolute Quantitation (standard curve) method. Briefly, the reaction was carried out in a SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) using 200 nM of each amplification primer and 100 ng (the input total RNA equivalents) of the first-strand cDNA in a 25 μl reaction volume as previously described^{97,98}. Real-time amplification was conducted on a CFX96 Real-Time PCR Detection System (Bio-Rad) with the following profile: 95 $^\circ\text{C}$ for 120 s, 40 cycles of 95 $^\circ\text{C}$ for 30 s, 60 $^\circ\text{C}$ for 30 s and 72 $^\circ\text{C}$ for 30 s followed by a melting curve analysis for each primer pair. Standards with known quantities (copy numbers) for a single mRNA sequence (gene of interest) were prepared from PCR products purified using Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, Indiana, USA). The expression levels were determined from a

standard curve of known target cDNA copy numbers (1.0×10^1 to 1.0×10^5 molecules per reaction) and analyzed simultaneously with unknown experimental samples on the same plate. The primers used in the study were listed in Supplementary Table S6.

Development of the analytical pipeline. Pipeline 1: Initially, we pooled all of the technical replicates from different lanes and sequencing dates in the preprocessing procedures, resulting in eight biological samples at each time point, identical to the previous study⁹⁹. For the detection of temporally differentially expressed genes and isoforms, we intuitively performed typical pairwise comparison methods, as our main hypothetical testing of interest is to identify any significant changes between treatment groups (later time points, D3 through D14) versus the baseline control group (Day 0). The comparative methods were carried out by Cufflinks and Cuffdiff, edgeR, and DESeq pairwise comparison. We utilized FPKM (fragments per kilobase per million mapped reads) values as the input data of expression values for the direct comparison of methods. When pooling replicates, we clearly observed that the quality of sample B29 was not good due to the sample and library prep. Further examination of this outlier sample explored how discrepant results impact downstream analyses, such as differential expression analysis, when this sample was either excluded or included. This step in our study ensured that poor quality samples were carefully dealt with in the preprocessing procedure as a prerequisite step since these samples can significantly affect the following results of downstream analyses.

Pipeline 2: As described in pipeline 1, our experimental design attempted to characterize the stimulated alterations across different time points from D0 through D14, and each individual sample was longitudinally observed as the repeated measurement during the given time period. Since the comparison in our study was focused on the D0 baseline time point versus each of the later time points, as the format of simple pairwise comparisons (i.e., before and after stimulus), other than a full time course series, simple pairwise static methods under the independent assumption of samples before and after stimulus might also work to some degree. However, it was evident that the expression levels were highly correlated between two neighboring time points, as the samples were longitudinally measured after an external stimulus of ruminal infusion. Additionally, our analysis also shows time-dependent expressed patterns between consecutive time points, as shown in Supplementary Fig. S2-(2), such that previously differentially expressed genes tended to also be significantly identified at the current time point. Hence, in order to better characterize data-driven features, for this pooled dataset we also performed dynamic-specific methods (as well as static pairwise comparisons) that are more precisely implemented for our within-subject stimulus-response longitudinal data, containing time-dependent structures¹⁻⁵. In addition to sample B29, we also observed extra variability among samples due to the distinct lanes, indicating a significant lane effect in the experiment. Prior to the main downstream analyses, we adjusted the systematic artifacts that could affect the main biological factors of interest (i.e., the time factor), as shown in the study that explored unwanted biases in static data¹², and we further incorporated the effect in our proposed Bayesian dynamic AR method. In other words, to examine how systematic biases can be confounding in the detection of differentially expressed genes and isoforms in this pipeline, we compared the results of differential expression with and without correction of the extra variability of samples before executing the main differential expression analysis. Furthermore, we also performed gene ontology, gene set enrichment analysis, and network module analysis based on the temporal dynamics of gene and isoforms as putative biomarkers.

Pipeline 3: This pipeline aims at exploring whether there are unwanted systematic biases that significantly affect the main factors of biological interest in unpooled samples. Unlike the previous pipelines that directly analyze the pooled data, prior to pooling all of the replicates we carefully examined all samples with diagnostic tools in an exploratory analysis, as described in the statistical analyses section¹² (See also Supplementary Fig. S1 and Dataset S1). In this pipeline, we explore how the nonnegligible effect of outlier samples and extra sample to sample variation due to lane effects behaves in data that contains unpooled replicates. If any unwanted bias is significantly present in the exploration of samples, this effect should be corrected prior to pooling the samples. We then perform typical pairwise static methods by incorporating those effects into the models of differential expression. The purpose of this is to test whether the effect of correction on systematic biases in the preprocessing steps affects the detection outcomes of temporally differential expression of unpooled data. We examined the results that were obtained both from before and after the correction of systematic biases and how (in)consistently they present in differential expression analysis.

References

1. Conesa, A., Nueda, M. J., Ferrer, A. & Talon, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102, <https://doi.org/10.1093/bioinformatics/btl056> (2006).
2. Nueda, M. J., Tarazona, S. & Conesa, A. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* **30**, 2598–2602, <https://doi.org/10.1093/bioinformatics/btu333> (2014).
3. Oh, S. & Song, S. Differential gene expression (DEX) and alternative splicing events (ASE) for temporal dynamic processes using HMMs and hierarchical bayesian modeling approaches. *Methods Mol. Biol.* **1552**, 165–176, https://doi.org/10.1007/978-1-4939-6753-7_12 (2017).
4. Oh, S., Song, S., Dasgupta, N. & Grabowski, G. The analytical landscape of static and temporal dynamics in transcriptome data. *Front. Genet.* **5**, 35, <https://doi.org/10.3389/fgene.2014.00035> (2014).
5. Oh, S., Song, S., Grabowski, G., Zhao, H. & Noonan, J. P. Time series expression analyses using RNA-seq: a statistical approach. *Biomed. Res. Int.* **2013**, 203681, <https://doi.org/10.1155/2013/203681> (2013).
6. Heinonen, M. *et al.* Detecting time periods of differential gene expression using Gaussian processes: an application to endothelial cells exposed to radiotherapy dose fraction. *Bioinformatics* **31**, 728–735, <https://doi.org/10.1093/bioinformatics/btu699> (2015).
7. Sanavia, T., Finotello, F. & Di Camillo, B. FunPat: function-based pattern analysis on RNA-seq time series data. *BMC Genomics* **16**, S2, <https://doi.org/10.1186/1471-2164-16-S6-S2> (2015).
8. Sun, X. *et al.* Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model. *BMC Bioinformatics* **17**, 324, <https://doi.org/10.1186/s12859-016-1180-9> (2016).

9. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet* **11**, 733–739, <https://doi.org/10.1038/nrg2825> (2010).
10. Liu, R. *et al.* Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res.* **43**, e97, <https://doi.org/10.1093/nar/gkv412> (2015).
11. Loven, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482, <https://doi.org/10.1016/j.cell.2012.10.012> (2012).
12. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902, <https://doi.org/10.1038/nbt.2931> (2014).
13. Law, C. W., Alhamdoosh, M., Su, S., Smyth, G. K. & Ritchie, M. E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* **5**, 1408, <https://doi.org/10.12688/f1000research.9005.2> (2016).
14. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427, <https://doi.org/10.1038/nbt.4091> (2018).
15. Nygaard, V., Rodland, E. A. & Hovig, E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* **17**, 29–39, <https://doi.org/10.1093/biostatistics/kxv027> (2016).
16. Shaham, U. *et al.* Removal of batch effects using distribution-matching residual networks. *Bioinformatics* **33**, 2539–2546, <https://doi.org/10.1093/bioinformatics/btx196> (2017).
17. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127, <https://doi.org/10.1093/biostatistics/kxj037> (2007).
18. Leek, J. T. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* **42**, e161, <https://doi.org/10.1093/nar/gku864> (2014).
19. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
20. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106, <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
21. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, <https://doi.org/10.1093/bioinformatics/btu638> (2015).
22. Ayoub, A. E. *et al.* Transcriptional programs in transient embryonic zones of the cerebral cortex defined by high-resolution mRNA sequencing. *Proc. Natl. Acad. Sci. USA* **108**, 14950–14955, <https://doi.org/10.1073/pnas.1112213108> (2011).
23. Cotney, J. *et al.* Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* **22**, 1069–1080, <https://doi.org/10.1101/gr.129817.111> (2012).
24. Towfic, F., Kusko, R. & Zeskind, B. *et al.* Letter to the editor response: Nygaard. *Biostatistics* **18**, 197–199, <https://doi.org/10.1093/biostatistics/kxw031> (2017).
25. Baldwin, R. L. T. *et al.* Quantification of Transcriptome Responses of the Rumen Epithelium to Butyrate Infusion using RNA-seq Technology. *Gene Regul Syst Bio* **6**, 67–80, <https://doi.org/10.4137/GRSB.S9687> (2012).
26. Alamancos, G. P., Agirre, E. & Eyraes, E. Methods to study splicing from high-throughput RNA sequencing data. *Methods Mol. Biol.* **1126**, 357–397, https://doi.org/10.1007/978-1-62703-980-2_26 (2014).
27. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017, <https://doi.org/10.1101/gr.133744.111> (2012).
28. Angelini, C., De Canditiis, D. & De Feis, I. Computational approaches for isoform detection and estimation: good and bad news. *BMC Bioinformatics* **15**, 135, <https://doi.org/10.1186/1471-2105-15-135> (2014).
29. Aschoff, M. *et al.* SplicingCompass: differential splicing detection using RNA-seq data. *Bioinformatics* **29**, 1141–1148, <https://doi.org/10.1093/bioinformatics/btt101> (2013).
30. Beretta, S., Bonizzoni, P., Vedova, G. D., Pirola, Y. & Rizzi, R. Modeling alternative splicing variants from RNA-Seq data with isoform graphs. *J Comput Biol* **21**, 16–40, <https://doi.org/10.1089/cmb.2013.0112> (2014).
31. Bianchi, V. *et al.* DBATE: database of alternative transcripts expression. *Database (Oxford)* **2013**, bat050, <https://doi.org/10.1093/database/bat050> (2013).
32. Griffith, M. *et al.* Alternative expression analysis by RNA sequencing. *Nat. Methods* **7**, 843–847, <https://doi.org/10.1038/nmeth.1503> (2010).
33. Hu, Y. *et al.* DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res.* **41**, e39, <https://doi.org/10.1093/nar/gks1026> (2013).
34. Jiang, H. & Wong, W. H. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics* **25**, 1026–1032, <https://doi.org/10.1093/bioinformatics/btp113> (2009).
35. LeGault, L. H. & Dewey, C. N. Inference of alternative splicing from RNA-Seq data with probabilistic splice graphs. *Bioinformatics* **29**, 2300–2310, <https://doi.org/10.1093/bioinformatics/btt396> (2013).
36. Leng, N. *et al.* EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043, <https://doi.org/10.1093/bioinformatics/btt087> (2013).
37. Liu, X., Shi, X., Chen, C. & Zhang, L. Improving RNA-Seq expression estimation by modeling isoform- and exon-specific read sequencing rate. *BMC Bioinformatics* **16**, 332, <https://doi.org/10.1186/s12859-015-0750-6> (2015).
38. Nariai, N., Hirose, O., Kojima, K. & Nagasaki, M. TIGAR: transcript isoform abundance estimation method with gapped alignment of RNA-Seq data by variational Bayesian inference. *Bioinformatics* **29**, 2292–2299, <https://doi.org/10.1093/bioinformatics/btt381> (2013).
39. Pick, M., Flores-Flores, C. & Soreq, H. From brain to blood: alternative splicing evidence for the cholinergic basis of Mammalian stress responses. *Ann. N. Y. Acad. Sci.* **1018**, 85–98, <https://doi.org/10.1196/annals.1296.010> (2004).
40. Richard, H. *et al.* Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res.* **38**, e112, <https://doi.org/10.1093/nar/gkq041> (2010).
41. Wang, W., Qin, Z., Feng, Z., Wang, X. & Zhang, X. Identifying differentially spliced genes from two groups of RNA-seq samples. *Gene* **518**, 164–170, <https://doi.org/10.1016/j.gene.2012.11.045> (2013).
42. Wu, Z., Wang, X. & Zhang, X. Using non-uniform read distribution models to improve isoform expression inference in RNA-Seq. *Bioinformatics* **27**, 502–508, <https://doi.org/10.1093/bioinformatics/btq696> (2011).
43. Yi, L., Pimentel, H., Bray, N. L. & Pachter, L. Gene-level differential analysis at transcript-level resolution. *Genome Biol.* **19**, 53, <https://doi.org/10.1186/s13059-018-1419-z> (2018).
44. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**(Suppl 1), S7, <https://doi.org/10.1186/1471-2105-7-S1-S7> (2006).
45. Parsa, A. T. A newly identified transcriptional network for mesenchymal transformation of brain tumors: potential targets for therapeutic intervention. *World Neurosurg.* **73**, 424, <https://doi.org/10.1016/j.wneu.2010.05.022> (2010).
46. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550, <https://doi.org/10.1073/pnas.0506580102> (2005).
47. Vinh, N. X., Chetty, M., Coppel, R. & Wangikar, P. P. GlobalMIT: learning globally optimal dynamic bayesian network with the mutual information test criterion. *Bioinformatics* **27**, 2765–2766, <https://doi.org/10.1093/bioinformatics/btr457> (2011).
48. Zoppoli, P., Morganello, S. & Ceccarelli, M. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. *BMC Bioinformatics* **11**, 154, <https://doi.org/10.1186/1471-2105-11-154> (2010).

49. Meyer, P. E., Lafitte, F. & Bontempi, G. minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* **9**, 461, <https://doi.org/10.1186/1471-2105-9-461> (2008).
50. Castro, M. A., Wang, X., Fletcher, M. N., Meyer, K. B. & Markowetz, F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. *Genome Biol.* **13**, R29, <https://doi.org/10.1186/gb-2012-13-4-r29> (2012).
51. Eyras, E., Caccamo, M., Curwen, V. & Clamp, M. ESTGenes: alternative splicing from ESTs in Ensembl. *Genome Res.* **14**, 976–987, <https://doi.org/10.1101/gr.1862204> (2004).
52. Graveley, B. R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
53. Shen, S. *et al.* MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res.* **40**, e61, <https://doi.org/10.1093/nar/gkr1291> (2012).
54. Sultan, M. *et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960, <https://doi.org/10.1126/science.1160342> (2008).
55. Zhao, K., Lu, Z. X., Park, J. W., Zhou, Q. & Xing, Y. GLiMMPs: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biol.* **14**, R74, <https://doi.org/10.1186/gb-2013-14-7-r74> (2013).
56. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476, <https://doi.org/10.1038/nature07509> (2008).
57. Spies, D., Renz, P. F., Beyer, T. A. & Ciaudo, C. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Brief Bioinform.* <https://doi.org/10.1093/bib/bbx115> (2017).
58. Michna, A. *et al.* Natural Cubic Spline Regression Modeling Followed by Dynamic Network Reconstruction for the Identification of Radiation-Sensitivity Gene Association Networks from Time-Course Transcriptome Data. *PLoS One* **11**, e0160791, <https://doi.org/10.1371/journal.pone.0160791> (2016).
59. Zinkgraf, M., Gerttula, S., Zhao, S., Filkov, V. & Groover, A. Transcriptional and temporal response of Populus stems to gravistimulation. *J. Integr. Plant Biol.* **60**, 578–590, <https://doi.org/10.1111/jipb.12645> (2018).
60. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47, <https://doi.org/10.1093/nar/gkv007> (2015).
61. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **12**, R22, <https://doi.org/10.1186/gb-2011-12-3-r22> (2011).
62. Ren, Y., Hong, C. I., Lim, S. & Song, S. Finding clocks in genes: a Bayesian approach to estimate periodicity. *Biomed. Res. Int.* **2016**, 3017475, <https://doi.org/10.1155/2016/3017475> (2016).
63. Sonesson, C. & Robinson, M. D. iCOBRA: open, reproducible, standardized and live method benchmarking. *Nat. Methods* **13**, 283, <https://doi.org/10.1038/nmeth.3805> (2016).
64. Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer* **15**, 730–745, <https://doi.org/10.1038/nrc4029> (2015).
65. Brait, M. & Sidransky, D. Cancer epigenetics: above and beyond. *Toxicol. Mech. Methods* **21**, 275–288, <https://doi.org/10.3109/15376516.2011.562671> (2011).
66. Chin, L., Hahn, W. C., Getz, G. & Meyerson, M. Making sense of cancer genomic data. *Genes Dev.* **25**, 534–555, <https://doi.org/10.1101/gad.2017311> (2011).
67. Fletcher, M. N. *et al.* Master regulators of FGFR2 signalling and breast cancer risk. *Nat. Commun.* **4**, 2464, <https://doi.org/10.1038/ncomms3464> (2013).
68. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158, <https://doi.org/10.1038/nature05610> (2007).
69. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* **9**, 703–719, <https://doi.org/10.4161/rna.20481> (2012).
70. International Cancer Genome Consortium *et al.* International network of cancer genome projects. *Nature* **464**, 993–998, <https://doi.org/10.1038/nature08987> (2010).
71. Jorge, N. A., Ferreira, C. G. & Passetti, F. Bioinformatics of cancer ncRNA in high throughput sequencing: present state and challenges. *Front. Genet.* **3**, 287, <https://doi.org/10.3389/fgene.2012.00287> (2012).
72. Mitra, S. A., Mitra, A. P. & Triche, T. J. A central role for long non-coding RNA in cancer. *Front. Genet.* **3**, 17, <https://doi.org/10.3389/fgene.2012.00017> (2012).
73. Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196, <https://doi.org/10.1038/nature08658> (2010).
74. Pritchard, J. K. & Cox, N. J. The allelic architecture of human disease genes: common disease-common variant...or not? *Hum. Mol. Genet.* **11**, 2417–2423 (2002).
75. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
76. Isakov, O. *et al.* Novel insight into the non-coding repertoire through deep sequencing analysis. *Nucleic Acids Res.* **40**, e86, <https://doi.org/10.1093/nar/gks228> (2012).
77. Oh, S., Song, S., Grabowski, G., Zhao, H. & Noonan, J. P. Time series expression analyses using RNA-seq: a statistical approach. *Biomed Res Int* **2013**, 203681, <https://doi.org/10.1155/2013/203681> (2013).
78. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**, 896–902, <https://doi.org/10.1038/nbt.2931> (2014).
79. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, <https://doi.org/10.1186/gb-2010-11-10-r106> (2010).
80. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, <https://doi.org/10.1093/nar/gkv007> (2015).
81. Sonesson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14**, 91, <https://doi.org/10.1186/1471-2105-14-91> (2013).
82. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res* **43**, e140, <https://doi.org/10.1093/nar/gkv711> (2015).
83. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* **12**, R22, <https://doi.org/10.1186/gb-2011-12-3-r22> (2011).
84. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46–53, <https://doi.org/10.1038/nbt.2450> (2013).
85. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
86. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008–2017, <https://doi.org/10.1101/gr.133744.111> (2012).
87. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
88. Xia, J., Gill, E. E. & Hancock, R. E. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* **10**, 823–844, <https://doi.org/10.1038/nprot.2015.052> (2015).

89. Li, R. W., Wu, S., Baldwin, R. L. T., Li, W. & Li, C. Perturbation dynamics of the rumen microbiota in response to exogenous butyrate. *PLoS One* **7**, e29392, <https://doi.org/10.1371/journal.pone.0029392> (2012).
90. Die, J. V. *et al.* Selection of internal reference genes for normalization of reverse transcription quantitative polymerase chain reaction (RT-qPCR) analysis in the rumen epithelium. *PLoS One* **12**, e0172674, <https://doi.org/10.1371/journal.pone.0172674> (2017).
91. Li, C. J. & Li, R. W. Butyrate induced cell cycle arrest in bovine cells through targeting gene expression relevant to DNA replication apparatus. *Gene Regul Syst Bio* **2**, 113–123 (2008).
92. Li, C. J., Li, R. W., Wang, Y. H. & Elsasser, T. H. Pathway analysis identifies perturbation of genetic networks induced by butyrate in a bovine kidney epithelial cell line. *Funct Integr Genomics* **7**, 193–205, <https://doi.org/10.1007/s10142-006-0043-2> (2007).
93. Li, R. W. & Li, C. Butyrate induces profound changes in gene expression related to multiple signal pathways in bovine kidney epithelial cells. *BMC Genomics* **7**, 234, <https://doi.org/10.1186/1471-2164-7-234> (2006).
94. Li, R. W. *et al.* Mucin biosynthesis in the bovine goblet cell induced by *Cooperia oncophora* infection. *Vet Parasitol* **165**, 281–289, <https://doi.org/10.1016/j.vetpar.2009.07.008> (2009).
95. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53, <https://doi.org/10.1038/nbt.2450> (2013).
96. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578, <https://doi.org/10.1038/nprot.2012.016> (2012).
97. Guo, Z. *et al.* Possible mechanisms of host resistance to *Haemonchus contortus* infection in sheep breeds native to the Canary Islands. *Sci Rep* **6**, 26200, <https://doi.org/10.1038/srep26200> (2016).
98. Li, R. W. & Gasbarre, L. C. A temporal shift in regulatory networks and pathways in the bovine small intestine during *Cooperia oncophora* infection. *Int J Parasitol* **39**, 813–824, <https://doi.org/10.1016/j.ijpara.2008.11.007> (2009).
99. Baldwin, R. L. *et al.* Quantification of transcriptome responses of the rumen epithelium to butyrate infusion using RNA-seq technology. *Gene Regul. Syst. Bio.* **6**, 67–80, <https://doi.org/10.4137/GRSB.S9687> (2012).

Acknowledgements

This study was partially supported by an internal grant from Jeju National University for principal investigators to SO and was partially supported by the U. S. Department of Agriculture (USDA) National Institute of Food and Agriculture under award no. 2013-67015-22957 to WL and RWL. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Names or commercial products used in this publication are solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. We are also very thankful to American Journal Experts for the valuable comments and editing services in the revision of our manuscript.

Author Contributions

S.O. and R.W.L. conceived this study, R.L.B.V. and R.W.L. designed the experiment. C.L., R.L.B.V. and R.W.L. performed the experiment. F.L. performed the qRT-PCR validation. S.O., S.S., F.L. and R.W.L. analyzed the data. S.O., F.L. and R.L.B.V. wrote the manuscript, and S.O., C.L., R.L.B.V., S.S. and R.W.L. revised the manuscript. All authors reviewed and approved the final version of this manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37397-7>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019