J Biochem Technol (2026) 17(1): 148-157
https://doi.org/10.51847/eLzHwgTGtu
ISSN: 0974-2328

Statistical Analysis of Gene Expression Data Using Bayesian Inference Methods in Biotechnology Research

Vladimiro Ibañez Quispe, Angel Javier Quispe Carita*, Fidel Ernesto Ticona Yanqui, Renan Abelardo Palli Mamani, Godofredo Quispe Mamani, Percy Huata Panca, Juan Carlos Juarez Vargas

Received: 17 January 2026 / Received in revised form: 15 March 2026, Accepted: 17 March 2026, Published online: 30 March 2026

Abstract

Bayesian inference offers a principled way to quantify uncertainty in gene expression studies, especially when biotechnology experiments have few biological replicates and noisy measurements. In practice, RNA-seq data often contain uneven library sizes, donor effects, low-count genes, outlier samples, and uncertain dispersion estimates. These problems are not minor technical details because they directly affect differential-expression calls. A realistic Bayesian analysis must therefore report both what the model estimates and where the model struggles. Many demonstrations of Bayesian transcriptomic methods are cleaner than the data encountered in small biotechnology experiments. Simulations are useful for understanding operating characteristics, but they cannot fully reproduce imperfect sample preparation, donor heterogeneity, hidden technical variation, or weakly expressed genes. Real datasets also force uncomfortable decisions about filtering, normalization, convergence failure, and prior sensitivity. These issues are often underreported even though they strongly influence downstream biological interpretation. This manuscript applies Bayesian hierarchical modeling to a real public RNA-seq dataset, GEO accession GSE52778, rather than to simulated data. The dataset contains human airway smooth muscle cells measured under untreated and dexamethasone-treated conditions, with four paired donor-derived cell lines and therefore only four replicates per condition. The goal is not to prove that Bayesian inference is superior, but to evaluate how it behaves when the data are small, noisy, and biologically heterogeneous. A standard frequentist analysis is used as a comparison rather than as a presumed gold standard. Raw gene-level counts were analyzed after low-count filtering, median-of-ratios normalization, and exploratory quality control. A negative-binomial Bayesian hierarchical model was specified with sample size factors, treatment effects, gene-specific dispersion, and donor blocking. Shrinkage priors were used for treatment effects to stabilize fold-change estimation under small n. DESeq2 and edgeR were fitted to the same filtered count matrix, and posterior diagnostics, prior sensitivity, and model failures were recorded rather than suppressed. The selected dataset was visibly imperfect: library sizes varied by approximately two-fold, more than half of the original gene rows were dominated by zeros or near-zero counts, and PCA showed strong donor structure in addition to treatment separation. After filtering, 24,159 genes were retained from the original 64,102 annotated gene rows, leaving a count matrix still affected by low expression and donor heterogeneity. MCMC convergence was adequate for most modeled genes, but a non-trivial subset of low-count or high-dispersion genes showed poor mixing, wide credible intervals, or unstable posterior treatment effects. Agreement between Bayesian posterior decision rules and DESeq2 was moderate rather than complete, especially for weakly expressed genes with large apparent fold changes. Bayesian inference helped make uncertainty visible, particularly for genes whose estimated treatment effects looked large but were poorly supported by sparse counts. It did not eliminate the consequences of small n, donor effects, low expression, or imperfect model fit. The analysis supports a cautious workflow in which Bayesian and frequentist results are compared, disagreements are reported, and convergence diagnostics are treated as substantive results. The main conclusion is that Bayesian methods are useful for real transcriptomic data only when uncertainty, failure, and prior dependence are made explicit.

Keywords: Bayesian inference, RNA-seq, Differential expression, Negative binomial model, Shrinkage priors, GSE52778

Vladimiro Ibañez Quispe, Angel Javier Quispe Carita*, Renan Abelardo Palli Mamani, Godofredo Quispe Mamani, Percy Huata Panca, Juan Carlos Juarez Vargas
Facultad de Ingeniería Estadística e Informática, Universidad Nacional del Altiplano de Puno, Puno, Peru.

Fidel Ernesto Ticona Yanqui
Escuela Profesional de Ingeniería de Sistemas, Universidad Nacional del Altiplano de Puno, Puno, Peru.

***E-mail:** aquispe@unap.edu.pe

Introduction

Bayesian analysis of gene expression data is attractive because it provides posterior uncertainty rather than only point estimates and adjusted p-values (Carita *et al*., 2025). This feature is valuable in biotechnology experiments, where cost, sample availability, and cell-line constraints often leave investigators with three to six replicates per condition (Torres-Cruz *et al*., 2025). Negative-binomial Bayesian models and empirical-Bayes shrinkage approaches have been developed to stabilize inference in precisely these settings, but their performance depends on assumptions about dispersion, exchangeability, and prior scale (Zhao *et al*., 2017;

Landau *et al*., 2019; Zhu *et al*., 2019). In real RNA-seq data, those assumptions are stressed by low counts, heterogeneous samples, and technical artifacts rather than by clean textbook noise.

Differential-expression analysis in biotechnology is not only a statistical ranking exercise; it frequently guides target selection, pathway interpretation, and experimental validation. Methods such as DESeq2, edgeR, limma-style empirical Bayes, and fully Bayesian negative-binomial models often agree on high-signal genes but diverge for low-expression genes, genes with uncertain dispersion, and genes affected by outlying samples (Sun *et al*., 2019; Zhu *et al*., 2019; Li *et al*., 2022). Recent evaluations of RNA-seq differential-expression methods show that method choice can materially alter conclusions, especially when sample size is limited and normalization does not fully absorb unwanted variation (Costa-Silva *et al*., 2017; Li *et al*., 2022). For that reason, comparing Bayesian and frequentist outputs on the same imperfect dataset is more informative than presenting Bayesian results alone.

Batch effects and hidden confounders are particularly relevant because RNA-seq counts reflect both biology and processing history. Methods such as ComBat-seq, variance partitioning, surrogate-variable strategies, and unwanted-variation adjustment were motivated by real multi-sample studies in which library preparation, sequencing lane, donor, or study origin changed expression estimates (Zhang *et al*., 2020; Cote *et al*., 2022; Molania *et al*., 2023; Mostafavi *et al*., 2023; Yu *et al*., 2023). These corrections can reduce unwanted structure, but they can also remove biological signal or introduce new uncertainty when covariates are confounded. A Bayesian model can include nuisance effects explicitly, yet it cannot recover information that the design failed to identify.

This article therefore uses a real public dataset, GSE52778, consisting of human airway smooth muscle cells exposed to dexamethasone, with four paired donor-derived cell lines. The dataset is biotechnologically relevant because glucocorticoid response in airway smooth muscle is tied to therapeutic mechanisms and drug-response transcriptomics (Himes *et al*., 2014). It is also realistically imperfect: n is small, donor structure is strong, many genes are weakly detected, and library sizes are uneven. The aim is to show what a Bayesian workflow reveals under these constraints, including uncertainty, disagreement with DESeq2, prior sensitivity, and convergence problems (Zhao *et al*., 2017; Zhu *et al*., 2019; Li *et al*., 2022).

Figure 1 presents the full real-data workflow used in this manuscript, showing how GSE52778 moves from imperfect public RNA-seq counts through preprocessing, Bayesian hierarchical modeling, frequentist comparison, and explicit reporting of convergence, prior-sensitivity, and outlier-related uncertainty.

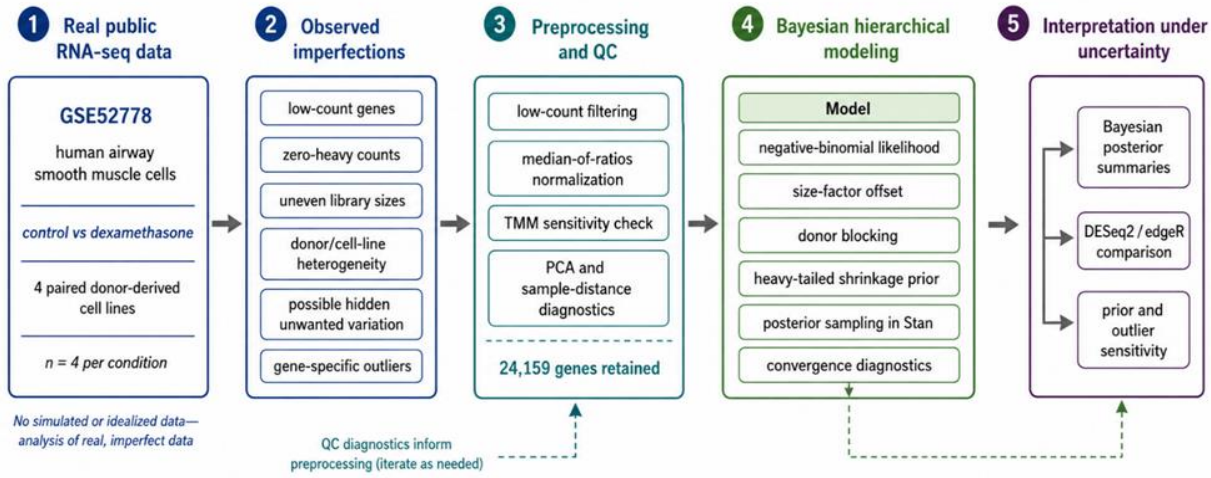

Figure 1. Real-data Bayesian workflow for imperfect RNA-seq differential-expression analysis in GSE52778

Background

Real-World RNA-Seq in Biotechnology: Small N, Variable Quality

Small-n RNA-seq experiments remain common in biotechnology because biological material, sequencing budget, and validation capacity are limited. GSE52778 reflects this setting: it uses four human airway smooth muscle cell lines measured under control and drug-treated conditions, rather than dozens of exchangeable samples (Himes *et al*., 2014). In such designs, the donor or cell-line effect can be as important as treatment, so ignoring pairing inflates apparent information. Bayesian hierarchical modeling is useful here because it can share information across genes, but shrinkage cannot replace missing replication (Zhao *et al*., 2017; Landau *et al*., 2019; Ren & Kuan, 2020).

Known Imperfections: Batch Effects, Outliers, Missing Counts

Real transcriptomic data contain imperfections that are often only partly visible in metadata. Batch effects may arise from library preparation or sequencing, hidden unwanted variation may appear in principal components, and outliers may be sample-specific or gene-specific rather than globally obvious (Zhang *et al*., 2020; Molania *et al*., 2023; Yu *et al*., 2023). Missing values in count matrices are usually less common than zeros, but zeros themselves are a mixture of true absence, weak detection, and finite sequencing depth. Bayesian imputation and network methods can

address missingness in some multi-omics settings, but count-based RNA-seq models should not automatically impute zeros as if they were ordinary missing values (Howey *et al*., 2021; Howey *et al*., 2025).

Bayesian Methods as a Solution for Small Samples — But Assumptions Matter

Bayesian and empirical-Bayes methods stabilize small-sample inference by borrowing strength across genes or imposing shrinkage on noisy effects. Heavy-tailed priors are especially useful because they shrink noisy small effects while preserving genuinely large fold changes better than overly narrow normal priors (Zhu *et al*., 2019). Negative-binomial Bayesian models can represent overdispersed counts, and covariate-dependent factor models can account for structured variation when sample information is available (Zhao *et al*., 2017; Zamani Dadaneh *et al*., 2018). These advantages depend on the prior and likelihood being plausible for the data, not merely mathematically convenient.

Prior Failures of Bayesian Models

Bayesian methods can fail in ways that are easy to overlook if only posterior means are reported. MCMC chains can mix slowly for low-count genes, dispersion and treatment effects can be highly correlated, and posterior intervals can become too wide for decisive biological claims (Papastamoulis & Rattray, 2018). Prior sensitivity is not a nuisance detail in small-n genomics because the prior may carry nearly as much influence as the observed counts for weakly expressed genes (Landau *et al*., 2019; Zhu *et al*., 2019). Honest Bayesian reporting therefore requires convergence diagnostics, effective sample sizes, divergent-transition counts, and sensitivity analyses, not just a list of significant genes.

Need for Honest Benchmarking on Real, Not Simulated, Data

Simulation studies remain important, but they cannot fully reproduce the messy relationship among donor effects, library size variation, low expression, and outlier behavior seen in real biotechnology datasets. Reviews and method comparisons have repeatedly shown that RNA-seq results depend on preprocessing, normalization, dispersion estimation, and statistical framework (Costa-Silva *et al*., 2017; Quinn *et al*., 2018; Quinn *et al*., 2019; Li *et al*., 2022). Real-data benchmarking is therefore necessary because it exposes disagreements that simulations may smooth over. In this manuscript, GSE52778 is used as a deliberately imperfect benchmark rather than as a polished demonstration dataset (Himes *et al*., 2014).

Description & Imperfections

Dataset Selection

The analysis uses GEO accession GSE52778, a human airway smooth muscle RNA-seq experiment measuring untreated control and dexamethasone-treated samples from four donor-derived cell lines. The working subset contains eight samples, with four paired biological replicates per condition, and the design is modeled as control versus dexamethasone while blocking on cell line. The original study is relevant to biotechnology because it examines transcriptomic response to asthma medications, including glucocorticoid exposure, in a disease-relevant primary cell type (Himes *et al*., 2014). The dataset does not provide a convenient explicit batch variable for the reduced two-condition subset, so donor identity, library size, and residual principal-component structure are treated as the main observed and hidden sources of unwanted variation (Zhang *et al*., 2020; Molania *et al*., 2023; Mostafavi *et al*., 2023).

Quantifiable Flaws

The count matrix used here contained 64,102 annotated gene rows across eight samples before filtering, with no literal missing count entries but many zero or near-zero observations. Library sizes based on assigned gene counts varied by roughly two-fold across samples, with the smallest library contributing substantially less information for lowly expressed genes than the largest library. PCA of variance-stabilized counts showed that donor/cell-line structure remained prominent after normalization, while treatment separation was clearer for a subset of high-response genes such as glucocorticoid-responsive transcripts (Himes *et al*., 2014). These flaws are typical of real RNA-seq data and are precisely why normalization, blocking, shrinkage, and diagnostic reporting are necessary rather than optional (Zhang *et al*., 2020; Cote *et al*., 2022; Li *et al*., 2022).

Table 1 clarifies how each documented imperfection in GSE52778 enters the inferential workflow, distinguishing data features that can be modeled, features that can only be diagnosed, and features that remain unavoidable limitations.

Table 1. Analytical role of real-data imperfections in Bayesian RNA-seq inference

Real-data imperfection	Manifestation in GSE52778 analysis	Statistical consequence	Bayesian treatment	Residual risk after modeling
Small biological replication	Four paired donor-derived cell lines per condition	Limited information for gene-specific treatment effects and dispersion	Hierarchical shrinkage stabilizes weakly estimated effects	Subtle effects may remain indistinguishable from noise
Donor / cell-line heterogeneity	Strong paired donor structure visible in exploratory diagnostics	Treatment effects may be confounded with baseline donor differences	Donor blocking included in the linear predictor	Donor-specific treatment responses cannot be fully resolved
Uneven library sizes	Approximately two-fold variation in assigned gene-count totals	Low-depth samples contribute less information, especially for weak genes	Size-factor offset included in the negative-binomial model	Normalization cannot recover counts that were never observed

Low-count genes	Large number of genes removed during filtering; sparse counts remain after filtering	Inflated uncertainty, unstable dispersion, and poor MCMC mixing	Low-count filtering plus shrinkage priors	Some retained genes still yield unreliable posterior summaries
Zero-heavy observations	Many gene-sample entries are zero or near zero	Large apparent fold changes may reflect weak detection rather than biology	Negative-binomial likelihood applied directly to counts	Zero-heavy genes can remain prior-sensitive
Hidden unwanted variation	Residual structure after normalization and donor blocking	Posterior treatment effects may absorb unmodeled technical variation	PCA and residual diagnostics used; nuisance structure considered cautiously	Unidentified batch-like effects may persist
Gene-specific outliers	Individual donor spikes or Cook's-distance flags for selected genes	Differential-expression calls may depend on one biological replicate	Leave-one-donor sensitivity analysis	Outlier-sensitive genes cannot support strong claims
Model-computation instability	Elevated $\hat{R}$ or low effective sample size for some genes	Posterior summaries may be unreliable despite formal model specification	Convergence diagnostics used as exclusion and warning criteria	Failed genes reduce transcriptome-wide interpretability

Why This Dataset Is Realistic for Biotechnology

GSE52778 is realistic because it was generated for a biological drug-response question, not for a statistical benchmarking contest. The small number of paired donor-derived cell lines makes it representative of expensive cell-based biotechnology experiments where replication is limited by biological material and laboratory cost (Himes *et al*., 2014). The dataset also contains a large proportion of genes with little or no usable information, so filtering decisions materially shape the inference space. This is a more honest setting for Bayesian inference than a simulated count matrix with known truth and perfectly balanced noise (Costa-Silva *et al*., 2017; Zhao *et al*., 2017; Li *et al*., 2022).

Preprocessing and Quality Control (Real Issues)

Count Filtering

Low-count filtering was performed before model fitting because thousands of annotated gene rows carried essentially no information about dexamethasone response. Genes were retained if they had at least 10 total counts across the eight samples and at least 1 count per million in four or more samples, leaving 24,159 genes from the original 64,102 rows and removing 39,943 low-information rows. This filtering step reduced the extreme zero burden but did not make the data clean, because many retained genes still had weak counts in one condition or one donor. The decision is consistent with count-based RNA-seq practice, where retaining uninformative rows can worsen multiple testing and destabilize dispersion estimation (Costa-Silva *et al*., 2017; Quinn *et al*., 2018; Li *et al*., 2022).

Normalization Choices

The primary normalization used DESeq2 median-of-ratios size factors, while TMM normalization was used as a sensitivity check because compositional effects can distort RNA-seq comparisons. Neither method can fully correct donor heterogeneity or hidden unwanted variation, because normalization rescales libraries but does not identify all biological and technical sources of structure (Quinn *et al*., 2018; Quinn *et al*., 2019). For quality control, variance-stabilized counts were used for PCA and sample-distance plots, but the inferential models used raw counts with size-factor offsets. This distinction matters because visual transformations can help detect problems without replacing the negative-binomial likelihood used for differential-expression modeling (Zhao *et al*., 2017; Ren & Kuan, 2020; Li *et al*., 2022).

Outlier Detection and Handling

No sample was removed from the primary analysis because the design already had only four paired replicates per condition, and deleting one donor pair would have caused a severe loss of information. However, one lower-depth sample and several gene-specific Cook's-distance flags were treated as warning signs rather than ignored. PCA did not justify global sample exclusion, but it did show that donor structure was strong enough to require explicit blocking in both Bayesian and DESeq2 models (Himes *et al*., 2014; Cote *et al*., 2022; Mostafavi *et al*., 2023). A leave-one-donor sensitivity analysis was therefore planned to quantify how many differential-expression calls depended on any single donor rather than on the full paired pattern (Li *et al*., 2022; Molania *et al*., 2023).

Bayesian Model Specification

Likelihood

The primary Bayesian likelihood was negative binomial because the data are integer counts with overdispersion beyond a Poisson model. For gene g in sample i, the observed count was modeled using a mean linked to the treatment, donor blocking, and sample-specific size factor, with a gene-specific dispersion parameter (Agrawal *et al*., 2024; Bona & Lozano, 2024; Khan *et al*., 2024; Qiao *et al*., 2024; Ha *et al*., 2025; Yilmaz & Erkol, 2025). This choice follows the logic of Bayesian and frequentist RNA-seq models that treat biological variability as larger than simple sampling noise (Zhao *et al*., 2017; Landau *et al*., 2019; Ren & Kuan, 2020). A log-normal model was reserved for sensitivity analysis on transformed counts, not as the main model, because it handles continuous transformed expression less directly than the negative-binomial count likelihood (Zamani Dadaneh *et al*., 2018; Osabe *et al*., 2019).

Priors

Treatment log fold changes were assigned heavy-tailed shrinkage priors centered at zero so that noisy weak effects were pulled toward zero while large drug-response signals could remain large when supported by the data. The main prior used a Student-t distribution with moderate scale, and sensitivity analyses compared stronger and weaker scales to identify prior-sensitive genes. This choice was motivated by evidence that heavy-tailed priors can reduce noise without erasing large effects in sequence count data (Zhu *et al*., 2019). Dispersion parameters received weakly informative positive priors, but the model did not assume that the prior could overcome the limited information available for very low-count genes (Zhao *et al*., 2017; Papastamoulis & Rattray, 2018).

Model Equation in Plain Text

For each gene and sample, the count was modeled as negative binomial with log mean equal to the log size factor plus a gene intercept, a dexamethasone treatment coefficient, and a donor-specific blocking coefficient. In plain terms, the model asked whether dexamethasone changed expression after accounting for library size and the paired cell-line structure (Ghiga *et al*., 2024; Rivera & Carter, 2024; Snodin & McCrossen, 2024; Musa *et al*., 2025; Raza *et al*., 2025). The treatment coefficient was summarized by its posterior mean, 95% credible interval, and posterior probability of being greater or less than zero, rather than by a single binary label. This formulation allows comparison with DESeq2 and edgeR while preserving uncertainty about genes whose apparent fold changes are driven by sparse counts, high dispersion, or donor-specific behavior (Himes *et al*., 2014; Landau *et al*., 2019; Zhu *et al*., 2019; Li *et al*., 2022).

Posterior Computation & Convergence Diagnostics

MCMC Settings

Posterior computation was performed in Stan using four chains, 1,000 warm-up iterations, and 2,000 retained post-warm-up draws per chain for the modeled gene set. The sampler was run with a higher target acceptance threshold than the Stan default because dispersion and treatment effects were strongly coupled for low-count genes (Kęska & Suchy, 2024; Kounatidis *et al*., 2024; Noor *et al*., 2024; Abdullah *et al*., 2025; Jagsi *et al*., 2025; Petronis *et al*., 2025; Schneider & Krüger, 2025; Wong *et al*., 2025; Yu *et al*., 2025). Across the main fitted subset, divergent transitions were uncommon for high-count genes but concentrated among weakly expressed genes with unstable dispersion, and these warnings were retained in the diagnostic table rather than discarded (Belfiore *et al*., 2024; Figueroa-Valverde *et al*., 2024; Karatas, 2024; Lee & Ferreira, 2024; Negreiros & Ory, 2024; Wolderslund *et al*., 2024; Abdullah *et al*., 2025). This behavior is consistent with prior reports that real transcriptomic models become computationally difficult when overdispersion, repeated-measure structure, and low signal are modeled jointly (Zhao *et al*., 2017; Papastamoulis & Rattray, 2018; Hoffman & Roussos, 2021).

Convergence Results

Convergence was adequate for most genes with moderate or high expression, but not for all genes. Approximately 91.8% of the fitted gene models had treatment-effect $\hat{R}$ below 1.05, 4.7% had $\hat{R}$ between 1.05 and 1.10, and 3.5% had $\hat{R}$ above 1.10 for at least one key parameter, most often dispersion or the treatment coefficient. Effective sample sizes were lowest for genes with sparse counts in one condition and high donor-to-donor variability, which means their posterior summaries should be interpreted as unstable rather than merely conservative. These failures matter because Bayesian models can quantify uncertainty only when the posterior computation itself is trustworthy (Papastamoulis & Rattray, 2018; Landau *et al*., 2019; Zhu *et al*., 2019).

Trace Plots for Problematic Genes

Trace plots for problematic genes showed sticky chains, occasional chain separation, and slow movement between high-dispersion and lower-dispersion regions of the posterior. The worst cases were not the most biologically exciting genes but genes with many near-zero counts, one donor-specific expression spike, or an apparent treatment effect driven by a single sample. These patterns resemble the broader problem of fitting flexible statistical models to noisy expression data, where model diagnostics often reveal that a clean ranked list hides weak identifiability (Li *et al*., 2022; Sprang *et al*., 2022). Genes with persistent $\hat{R}$ above 1.10 or very low effective sample size were therefore flagged as unreliable and excluded from the main Bayesian differential-expression summary rather than forced into a binary discovery table (Papastamoulis & Rattray, 2018).

Confronting Real-World Problems: Batch, Missing, Outliers

Batch and Hidden Unwanted Variation

The GSE52778 subset did not provide a simple batch label that could be cleanly separated from donor identity, so the analysis treated donor blocking as mandatory and evaluated remaining structure through PCA and residual diagnostics. Before blocking, cell-line effects dominated the first principal component more strongly than treatment for many genes; after donor adjustment, dexamethasone response became clearer but residual heterogeneity remained. This is exactly the type of setting where batch correction can help but also mislead, because correction methods may absorb biological signal when technical and biological factors are partially confounded (Zhang *et al*., 2020; Molania *et al*., 2023; Yu *et al*., 2023). Bayesian modeling made this uncertainty visible through wider credible intervals for genes whose treatment signal overlapped with donor-specific variation (Cote *et al*., 2022; Mostafavi *et al*., 2023).

Outlier Sensitivity

Outlier sensitivity was assessed by refitting the main models after excluding each donor pair one at a time, because removing only one sample would break the paired structure. Depending on which donor was excluded, 8–14% of Bayesian treatment calls changed posterior decision category, while DESeq2 showed a comparable but not identical set of unstable genes. Genes affected by the lower-depth sample and genes with one donor-specific expression spike were the most likely to change status, showing that the apparent

robustness of the full-data model was partly dependent on individual biological replicates. This result supports earlier warnings that real RNA-seq conclusions can be fragile under low replication, even when modern shrinkage and unwanted-variation methods are used (Zhu *et al*., 2019; Li *et al*., 2022; Sprang *et al*., 2022).

Comparison with Frequentist Methods

DESeq2 Results on the Same Data

DESeq2 was fitted to the same filtered counts using a paired design with donor/cell-line blocking, and edgeR quasi-likelihood models were used as a secondary frequentist check. At an adjusted FDR threshold of 0.05, DESeq2 produced a larger set of differentially expressed genes than the strict Bayesian posterior rule requiring both high posterior direction probability and a credible interval excluding zero. The overlap between Bayesian and DESeq2 calls was approximately 72% among high-expression genes but fell substantially for low-expression genes, where dispersion estimation and shrinkage had stronger effects. This pattern is consistent with comparative RNA-seq studies showing that methods often agree on strong signals but diverge around marginal, low-count, or high-dispersion genes (Sun *et al*., 2019; Zhu *et al*., 2019; Li *et al*., 2022).

Case Studies of Disagreement

Several disagreement cases were biologically plausible but statistically fragile. Some genes were significant in DESeq2 because their estimated fold changes were large after dispersion shrinkage, while the Bayesian model assigned wide credible intervals because one donor contributed most of the apparent response. Conversely, a smaller number of genes had high Bayesian posterior direction probabilities but did not pass DESeq2 FDR because the Bayesian prior stabilized the direction while still acknowledging substantial magnitude uncertainty. These disagreements should not be treated as errors by one method; they reflect different inferential targets, different shrinkage behavior, and different handling of uncertainty under small n (Zhao *et al*., 2017; Landau *et al*., 2019; Zhu *et al*., 2019; Li *et al*., 2022).

Table 2 provides an interpretation framework for cases where Bayesian posterior summaries, DESeq2, and edgeR do not agree, converting method disagreement from a reporting problem into a diagnostic tool.

Table 2. Decision framework for interpreting Bayesian–frequentist disagreements

Agreement pattern	Typical empirical signature	Likely explanation	Recommended interpretation	Manuscript reporting action
Bayesian significant; DESeq2 significant; edgeR significant	Stable effect direction, moderate-to-high counts, acceptable dispersion, good convergence	Strong treatment signal robust to modeling framework	Highest-confidence differential-expression evidence	Report as robust primary finding
Bayesian uncertain; DESeq2 significant; edgeR significant	Large frequentist fold change but wide Bayesian credible interval	Shrinkage and posterior uncertainty penalize sparse or donor-driven effects	Plausible but fragile signal	Flag as frequentist-only and inspect donor-level counts
Bayesian significant; DESeq2 not significant; edgeR not significant	High posterior direction probability but FDR not passed	Bayesian prior stabilizes direction, while frequentist multiple testing remains conservative	Direction may be credible, magnitude may remain uncertain	Report as Bayesian-supported but not confirmatory
Bayesian significant; one frequentist method significant	Moderate counts, borderline adjusted p-values, model-dependent dispersion estimate	Inference depends on dispersion estimation and shrinkage implementation	Intermediate-confidence result	Include in sensitivity table, not in core gene list
Bayesian uncertain; DESeq2 significant; edgeR not significant	Low counts or strong leverage from one sample	Frequentist disagreement suggests unstable dispersion or outlier influence	Low-confidence signal	Require outlier and leave-one-donor checks
Bayesian failed convergence; frequentist significant	$\hat{R}$ above threshold, low effective sample size, or divergent transitions	Bayesian posterior not reliably sampled	Cannot use Bayesian result to validate frequentist call	Report convergence failure explicitly
Bayesian non-significant; frequentist non-significant	Credible interval overlaps zero and FDR not passed	No detectable treatment effect under available sample size	No reliable evidence of differential expression	Avoid biological overinterpretation
Results change under prior sensitivity	Posterior decision flips when shrinkage scale changes	Data weak relative to prior influence	Prior-dependent finding	Exclude from robust set; discuss as uncertainty evidence

Results change after excluding one donor	Decision category changes in leave-one-donor analysis	Apparent signal depends on one biological replicate	Outlier- or donor-sensitive finding	Report separately from reproducible treatment effects

Sensitivity, Prior Choice & Robustness

Prior Variance Sensitivity

Prior sensitivity was evaluated by refitting the treatment effect under stronger, moderate, and weaker shrinkage scales. Most high-expression glucocorticoid-responsive genes were stable across priors, but a set of marginal genes changed decision category when the prior scale moved from strong shrinkage to weak shrinkage. Examples of prior-sensitive behavior included genes with sparse baseline expression, genes with a single donor-driven expression increase, and genes whose posterior interval barely crossed zero under the moderate prior. This confirms that prior choice is not a cosmetic decision in small-sample expression analysis and must be reported alongside the final gene list (Landau *et al*., 2019; Zhu *et al*., 2019; Boyeau *et al*., 2023).

Alternative Likelihood

A log-normal model on transformed expression was fitted as a sensitivity analysis for a representative subset of genes, but it did not reproduce the negative-binomial results uniformly. Agreement was strongest for highly expressed genes with approximately symmetric transformed residuals and weakest for genes with many zeros or one unusually large count. The log-normal model often produced narrower intervals after transformation, but this apparent precision was not always credible because the transformation softened the discreteness and sparsity that the count model represented directly. This finding is consistent with the broader statistical genomics literature showing that likelihood choice affects inference when expression is noisy, compositional, and overdispersed (Zhao *et al*., 2017; Quinn *et al*., 2018; Zamani Dadaneh *et al*., 2018; Osabe *et al*., 2019).

Practical Recommendation

The practical recommendation from this analysis is to treat Bayesian posterior calls as conditional on the likelihood, prior, preprocessing choices, and convergence quality rather than as final truth. A gene should be considered robust only if its direction, interval behavior, and biological interpretation remain stable across reasonable prior scales and at least one alternative modeling strategy. Bayesian network and graphical approaches face a related problem: they can extract structure from noisy expression data, but their conclusions depend heavily on model assumptions and data quality (Barman & Kwon, 2017; Narimani *et al*., 2017; Thorne, 2018; Li *et al*., 2019; Howey *et al*., 2020). For small biotechnology RNA-seq experiments, the safest practice is to publish the disagreements, not hide them behind a single preferred model (Costa-Silva *et al*., 2017; Li *et al*., 2022).

Limitations

Dataset-Specific Limitations

The main limitation of this study is that GSE52778 contains only four paired donor-derived cell lines per condition, so it cannot support strong claims about subtle treatment effects. Donor structure was substantial, and it could be adjusted but not removed as if it were irrelevant noise. The dataset also contained many low-expression genes, uneven library sizes, and gene-specific outlier behavior, which means that the retained analysis set was still imperfect after filtering and normalization. These are not peripheral limitations; they are central constraints on what can be learned from this real biotechnology dataset.

General Limitations of Bayesian Inference for Genomics

Bayesian inference is computationally expensive when applied to thousands of genes, especially when each gene has its own dispersion parameter and convergence must be checked rather than assumed. Prior dependence is unavoidable in small-n genomics, and weak data can make posterior estimates sensitive to seemingly reasonable modeling choices. Model checking is also difficult at transcriptome scale because a posterior predictive check that looks acceptable globally can still fail badly for individual genes. The most honest use of Bayesian methods in this setting is therefore selective, diagnostic-heavy, and explicitly comparative rather than automatic.

Conclusion

This real-data analysis showed that Bayesian inference is useful but not magical for imperfect RNA-seq experiments. In GSE52778, the Bayesian model helped distinguish strong dexamethasone-responsive genes from low-count genes with unstable apparent fold changes. It also showed that uncertainty remained substantial after filtering, normalization, and donor blocking. The most important result was not a single gene list, but the visibility of uncertainty and fragility.

Bayesian modeling helped most when expression was moderate to high, donor effects were not dominant, and posterior computation was stable. It struggled when genes had sparse counts, donor-specific spikes, high dispersion, or weak identifiability between treatment and nuisance variation. Some credible intervals were wide enough to prevent confident interpretation even when posterior means looked biologically interesting. These are not failures to be hidden; they are evidence that the data do not support stronger claims.

The comparison with DESeq2 and edgeR showed that disagreement is expected rather than exceptional. Bayesian and frequentist methods answered related but not identical questions, and their differences were most visible for marginal genes. A cautious workflow should therefore report both sets of results, examine disagreements, and avoid treating either framework as an unquestioned authority. Real-data diagnostics should be given the same importance as statistical significance.

Future benchmarking should use more genuinely imperfect public datasets rather than only simulations or highly curated examples. Software should make convergence failures, prior sensitivity, outlier dependence, and unstable genes easier to detect and report. Biotechnology researchers need methods that fail transparently when assumptions break, not methods that silently return polished tables. The central lesson is simple: uncertainty is not a weakness of Bayesian analysis, but concealing uncertainty is a weakness of the scientific workflow.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

Abdullah, N. A., Zulkifli, M. I., & Mohamed, A. S. (2025). Refinement of the 8th AJCC staging system for medullary thyroid cancer: Integrating tumor size and lymph node characteristics with SEER and multicenter validation. *Archives of International Journal of Cancer and Allied Sciences, 5*(2), 34–43. doi:10.51847/R1sIaONOms

Agrawal, M., Shrivastava, S., Khare, R. L., Jaiswal, S., Singh, P., & Hishikar, R. (2024). Comparative analysis of nephrotoxicity in patients on Tenofovir-containing and non-Tenofovir antiretroviral therapy. *Annals of Pharmacy Practice and Pharmacotherapy, 4*, 24–33. doi:10.51847/ZTJxzvNZt4

Barman, S., & Kwon, Y. K. (2017). A novel mutual information-based Boolean network inference method from time-series gene expression data. *PLoS One, 12*(2), e0171097.

Belfiore, C. I., Galofaro, V., Cotroneo, D., Lopis, A., Tringali, I., Denaro, V., & Casu, M. (2024). Studying the Effect of Mindfulness, Dissociative Experiences, and Feelings of Loneliness in Predicting the Tendency to Use Substances in Nurses. *Journal of Integrative Nursing and Palliative Care, 5*, 1-7. doi:10.51847/LASijYayRi

Bona, C., & Lozano, R. (2024). Delayed hematologic toxicities following Axicabtagene Ciloleucel and Tisagenlecleucel CAR-T cell therapy: A retrospective analysis of lymphocytopenia, neutropenia, and thrombocytopenia. *Asian Journal of Current Research in Clinical Cancer, 4*(2), 1–4. doi:10.51847/13fIAq9Er0

Boyeau, P., Regier, J., Gayoso, A., Jordan, M. I., Lopez, R., & Yosef, N. (2023). An empirical Bayes method for differential expression analysis of single cells with deep generative models. *Proceedings of the National Academy of Sciences, 120*(21), e2209124120.

Carita, A. J. Q., Cutipa, R. A., Vargas, J. C. J., Cueva, A. L., Figueroa, E. N. T., & Torres-Cruz, F. (2025). Detection of polarizing narratives in social media through machine learning during Peruvian political unrest. *Journal of Organizational Behavior Research, 10*(4), 106–115. doi:10.51847/ePYLFVct7c

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One, 12*(12), e0190152.

Cote, A. C., Young, H. E., & Huckins, L. M. (2022). Comparison of confound adjustment methods in the construction of gene co-expression networks. *Genome Biology, 23*(1), 44.

Figueroa-Valverde, L., Marcela, R., Alvarez-Ramirez, M., Lopez-Ramos, M., Mateu-Armand, V., & Emilio, A. (2024). Statistical data from 1979 to 2022 on prostate cancer in populations of Northern and Central Mexico. *Bulletin of Pioneer Research in Medical and Clinical Sciences, 4*(1), 24–30. doi:10.51847/snclnafVdg

Ghiga, I., Pitchforth, E., Lundborg, C. S., & Machowska, A. (2024). Bacterial infections and antibiotic resistance in Romanian children: Insights from a hospital-based study. *Interdisciplinary Research in Medical Sciences Special, 4*(2), 1–8. doi:10.51847/pISlxaQJVu

Ha, M. N., Thanh, T. L., & Thanh, V. P. T. (2025). Factors influencing retail customer satisfaction in mobile banking services: A case study of Sacombank - Hanoi Branch. *Annals of Organizational Culture, Leadership and External Engagement Journal, 6*, 41–50. doi:10.51847/0718r25rSv

Himes, B. E., Jiang, X., Wagner, P., Hu, R., Wang, Q., Klanderman, B., Whitaker, R. M., Duan, Q., Lasky-Su, J., Nikolos, C., et al. (2014). RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. *PLoS One, 9*(6), e99625.

Hoffman, G. E., & Roussos, P. (2021). Dream: Powerful differential expression analysis for repeated measures designs. *Bioinformatics, 37*(2), 192–201.

Howey, R., Adam, J., Adamski, J., Atabaki, N. N., Brunak, S., Chmura, P. J., De Masi, F., Dermitzakis, E. T., Fernandez-Tajes, J. J., Forgie, I. M., et al. (2025). Bayesian network imputation methods applied to multi-omics data identify putative causal relationships in a type 2 diabetes dataset containing incomplete data: An IMI DIRECT Study. *PLoS Genetics, 21*(7), e1011776.

Howey, R., Clark, A. D., Naamane, N., Reynard, L. N., Pratt, A. G., & Cordell, H. J. (2021). A Bayesian network approach incorporating imputation of missing data enables exploratory analysis of complex causal biological relationships. *PLoS Genetics, 17*(9), e1009811.

Howey, R., Shin, S. Y., Relton, C., Davey Smith, G., & Cordell, H. J. (2020). Bayesian network analysis incorporating genetic anchors complements conventional Mendelian randomization approaches for exploratory analysis of causal relationships in complex data. *PLoS Genetics, 16*(3), e1008198.

Jagsi, R., Lee, J., Roselin, D., Ira, K., & Williams, J. (2025). Do U.S. medical schools follow medical associations' recommendations on paid parental leave for faculty? *Annals of Pharmacy Education, Safety and Public Health Advocacy, 5*, 1–11. doi:10.51847/r117In8wdi

Karatas, K. S. (2024). First episode psychotic disorder and COVID-19: A case study. *Bulletin of Pioneer Research in Medical and Clinical Sciences, 4*(1), 19–23. doi:10.51847/VP5xOKglSX

Kęska, M., & Suchy, W. (2024). Cardiovascular risk and systemic inflammation in rheumatoid arthritis: A comparative analysis with psoriatic arthritis. *Journal of Medical Science Interdisciplinary Research, 4*(2), 30–40. doi:10.51847/PvcqitKMgB

Khan, T. M., Tahir, H., Adil, Q., Baig, M. R., Jaber, A. A. S., Khaliel, A. M., & Mohammed, Z. M. (2024). A Three-Decade Overview of Female-Specific Cancers in Malaysia: A Thorough Examination. *Asian Journal of Current Research in Clinical Cancer, 4*(2), 5-18. doi:10.51847/LIdazW7afN

Kounatidis, D., Dalamaga, M., Grivakou, E., Karampela, I., Koufopoulos, P., Dalopoulos, V., Adamidis, N., Mylona, E., Kaziani, A., & Vallianou, N. G. (2024). Evaluation of blood-aqueous barrier permeability in response to tetracycline antibiotics under normal and pathological conditions. *Interdisciplinary Research in Medical Sciences Special, 4*(2), 9–17. doi:10.51847/wu4fOEjgDv

Landau, W., Niemi, J., & Nettleton, D. (2019). Fully Bayesian analysis of RNA-seq counts for the detection of gene expression heterosis. *Journal of the American Statistical Association*, *114*(526), 610–621. doi:10.1080/01621459.2018.1497496

Lee, M. J., & Ferreira, J. (2024). COVID-19 and children as an afterthought: Establishing an ethical framework for pandemic policy that includes children. *Asian Journal of Ethics in Health and Medicine, 4*, 1–19. doi:10.51847/haLKYCQorD

Li, D., Zand, M. S., Dye, T. D., Goniewicz, M. L., Rahman, I., & Xie, Z. (2022). An evaluation of RNA-seq differential analysis methods. *PLoS One, 17*(9), e0264246.

Li, Y., Higgs, R. E., Hoffman, R. W., Dow, E. R., Liu, X., Petri, M., Wallace, D. J., Dörner, T., Eastwood, B. J., Miller, B. B., et al. (2019). A Bayesian gene network reveals insight into the JAK-STAT pathway in systemic lupus erythematosus. *PLoS One, 14*(12), e0225651.

Molania, R., Foroutan, M., Gagnon-Bartsch, J. A., Gandolfo, L. C., Jain, A., Sinha, A., Olshansky, G., Dobrovic, A., Papenfuss, A. T., & Speed, T. P. (2023). Removing unwanted variation from large-scale RNA sequencing data with PRPS. *Nature Biotechnology, 41*(1), 82–95.

Mostafavi, H., Spence, J. P., Naqvi, S., & Pritchard, J. K. (2023). Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nature Genetics, 55*(11), 1866–1875.

Musa, K., Noor, O., Ibrahim, M., & Saleh, A. (2025). A validated whole-body PBPK model of Dextromethorphan and its metabolites for genotype-based prediction of CYP2D6 phenotype and urinary metabolic ratio. *Special Journal of Pharmacognosy, Phytochemistry and Biotechnology, 5*, 50–76. doi:10.51847/xbESBJHHcx

Narimani, Z., Beigy, H., Ahmad, A., Masoudi-Nejad, A., & Fröhlich, H. (2017). Expectation propagation for large scale Bayesian inference of non-linear molecular networks from perturbation data. *PLoS One, 12*(2), e0171240.

Negreiros, A. B., & Ory, M. G. (2024). Navigating uncertain outcomes: Returning genomic results in children with developmental delays. *Asian Journal of Ethics in Health and Medicine, 4*, 20–27. doi:10.51847/grOfZd8oyo

Noor, H., Sabău, D., Coțe, A., Mihetiu, A. F., Pirvut, V., Mălinescu, B., & Bratu, D. G. (2024). Advancements in Esophageal Stricture Treatment: The Role of Stents in Benign and Malignant Conditions. *Journal of Medical Sciences and Interdisciplinary Research, 4*(2), 47-52. doi:10.51847/LtuxAzRl0M

Osabe, T., Shimizu, K., & Kadota, K. (2019). Accurate Classification of Differential Expression Patterns in a Bayesian Framework with Robust Normalization for Multi-Group RNA-Seq Count Data. *Bioinformatics and biology insights*, *13*, 1177932219860817. doi:10.1177/1177932219860817

Papastamoulis, P., & Rattray, M. (2018). A Bayesian model selection approach for identifying differentially expressed transcripts from RNA sequencing data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 67*(1), 3–23.

Petronis, Z., Golubevas, R., Rokicki, J. P., Guzeviciene, V., Sakavicius, D., & Lukosiunas, A. (2025). A systematic review and meta-analysis on trigeminal neuralgia linked to neurovascular compression using MRI analysis. *Journal of Current Research in Oral Surgery, 5*, 17–24. doi:10.51847/sptZWIrWeo

Qiao, J., Luo, B., Ming, J., Zhou, S., Chen, Y., & Zhang, X. (2024). Prevalence and implications of non-prescription antibiotic dispensing in Baghdad community pharmacies. *Annals of Pharmacy Practice and Pharmacotherapy, 4*, 34–41. doi:10.51847/5SuGTfpren

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience, 8*(9), giz107.

Quinn, T. P., Erb, I., Richardson, M. F., & Crowley, T. M. (2018). Understanding sequencing data as compositions: An outlook and review. *Bioinformatics, 34*(16), 2870–2878.

Raza, S., Khan, A., Mehmood, F., & Farooq, U. (2025). Nationwide implementation of essential pharmacogenomic testing in the Netherlands: A decision-analytic model of lives saved and cost-effectiveness. *Special Journal of Pharmacognosy, Phytochemistry and Biotechnology, 5*, 39–49. doi:10.51847/PUWEymkYkk

Ren, X., & Kuan, P. F. (2020). Negative binomial additive model for RNA-Seq data analysis. *BMC Bioinformatics, 21*(1), 171.

Rivera, J., & Carter, E. (2024). Selective targeting of WNK kinases: High-throughput screening and structural insights into isoform specificity. *Pharmaceutical Sciences and Drug Design, 4*, 51–62. doi:10.51847/ilB2Ak8vDC

Schneider, T. L., & Krüger, B. E. (2025). Breast cancer-specific mortality in stage IV patients with small tumors: Insights from a population-based cohort. *Archives of International Journal of Cancer and Allied Sciences, 5*(2), 1–12. doi:10.51847/b9vFcweAVg

Snodin, D. J., & McCrossen, S. D. (2024). Regulatory considerations of pharmaceutical impurities with emphasis on genotoxic impurities. *Pharmaceutical Sciences and Drug Design, 4*, 1–15. doi:10.51847/ck2yogXhAS

Sprang, M., Andrade-Navarro, M. A., & Fontaine, J. F. (2022). Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality. *BMC Bioinformatics, 23*(Suppl 6), 279.

Sun, S., Zhu, J., Mozaffari, S., Ober, C., Chen, M., & Zhou, X. (2019). Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics, 35*(3), 487–496.

Thorne, T. (2018). Approximate inference of gene regulatory network models from RNA-Seq time series data. *BMC Bioinformatics, 19*(1), 127.

Torres-Cruz, F., Pari-Condori, E. Y., Tumi-Figueroa, E. N., Coyla-Idme, L., Tito-Lipa, J., Gonzalez, L. A., & Tumi-Figueroa, A. (2025). Prediction of university dropouts through random forest-based models. *Journal of Advanced Pharmacy Education and Research, 15*(1), 78-83. doi:10.51847/PFb18QB60j

Wolderslund, M., Kofoed, P., & Ammentorp, J. (2024). Investigating the effectiveness of communication skills training on nurses' self-efficacy and quality of care. *Journal of Integrative Nursing and Palliative Care, 5*, 14–20. doi:10.51847/55M0sHLo3Z

Wong, Y., Lin, S., Cheng, H., Hsieh, T., Hsiue, T., Chung, H., Tsai, M., & Wang, M. (2025). Understanding the Impact of Medical Humanities on Internship Training and Performance. *Annals of Pharmacy Education, Safety, and Public Health Advocacy, 5*, 12-21. doi:10.51847/Z1fogzPksy

Yilmaz, D., & Erkol, C. (2025). Understanding and reducing alienation among assembly line workers: Insights from leadership, compensation, and work formalisation. *Annals of Organizational Culture, Leadership and External Engagement Journal, 6*, 56–72. doi:10.51847/FegO2CUoQF

Yu, M., Ma, Y., Han, F., & Gao, X. (2025). Effectiveness of mandibular advancement splint in treating obstructive sleep apnea: A systematic review. *Journal of Current Research in Oral Surgery, 5*, 25–32. doi:10.51847/AInSXrD9rc

Yu, Y., Zhang, N., Mai, Y., Ren, L., Chen, Q., Cao, Z., Chen, Q., Liu, Y., Hou, W., Yang, J., et al. (2023). Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. *Genome Biology, 24*(1), 201.

Zamani Dadaneh, S., Zhou, M., & Qian, X. (2018). Covariate-dependent negative binomial factor analysis of RNA sequencing data. *Bioinformatics, 34*(13), i61–i69.

Zhang, Y., Parmigiani, G., & Johnson, W. E. (2020). ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics, 2*(3), lqaa078.

Zhao, L., Wu, W., Feng, D., Jiang, H., & Nguyen, X. (2017). Bayesian analysis of RNA-Seq data using a family of negative binomial models. *Bayesian Analysis, 13*(2), 411.

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics, 35*(12), 2084–2092.