Application of Sigmoidal Models to Elucidate the Timing of the RpoS Regulon in an E. Coli Cell Starvation RNA-Seq Time Course

Ethan Ashby[1,4], Annie Cohen[2,4], Lian Morales[3,4], Prof. Jo Hardin[1], Prof. Dan Stoebel[3], Prof. Danae Schulz[3]

Affiliation: [1] Department of Mathematics, Pomona College [2] Department of Mathematics, Scripps College, [3] Department of Biology, Harvey Mudd College [4] NSF Data Science REU at Harvey Mudd College

Abstract

E. coli possesses a general stress response that coordinates physiological responses to a variety of stressful stimuli including cell starvation during exponential growth. A key transcription factor in the general stress response, RpoS, is involved in the transcription of approximately one quarter of *E. coli's* genome. Groups of genes were previously classified by their kinetics with respect to RpoS concentration: genes with expressions that increase linearly with RpoS concentration ('linear' genes), genes that are transcribed *more* than anticipated under the linear hypothesis at low RpoS concentrations ('sensitive' genes), genes that are transcribed *less* than anticipated under the linear hypothesis at low RpoS concentrations ('insensitive' genes). Wong *et al.* proposed that the graded RpoS sensitivity of these genes could function as a mechanism to control the *timing* of genes invovled in *E. coli's* response to stress (Wong *et al* 2017). To address this question, DEGs were determined using a thoughtfully-constructed pipeline, and gene-wise sigmoidal models were fit to significant *E. coli* expression trajectories using ImpulseDE2. The dispersion robustness, impact of number of time points, and outlier robustness of the models' parameters were assessed by simulations. Visualization of the biologically meaningful "onset time" parameter of the sigmoidal model indicated that the *E. coli* genes all begin transcription at roughly the same time. While a significant difference between onset times for sensitive and insensitive genes was observed, the difference in onset times was small and not descriptive of the global trend of transcriptomic response. Implications and future directions are discussed.

Introduction

E. coli possesses a general stress response to a variety of environmental stresses (Battesti *et al* 2011, Hengge 2011) ranging from osmotic shock to nutrient starvation. A key transcription factor coordinating this response is RpoS, which regulates one quarter of the bacteria's genome (preliminary data from Professor Dan Stoebel). Simple interpretations of transcriptional networks often invoke an analogy of an on/off switch, in which the presence of a stimulus turns some genes on and other genes off. However, these simple interpretations don't adequately describe the complex, dyanmic processes underlying many transcriptional responses to stimuli. Currently, there exists a limited understanding regarding the dynamic nature of transcriptional responses, and the well-annotated, heavily-studied genome of E. coli presents an excellent model to study these intricate regulatory circuits.

The RpoS regulon is not a static, 'switch-like' network; rather, the RpoS regulon is a highly complex regulatory circuit influenced by several factors including the duration/degree of stress (Lange and Hengge-Aronis 1994), processes like transcription, translation, and mRNA degradation (Lange and Hengge-Aronis 1994), other proteins (Pratt and Silhavy 1998), competition between transcription factors for RNA polymerase (Farewell *et al.* 1998), and even strain type (Hryckowian *et al.* 2014, Chiang *et al.* 2011). Wong and colleagues previously showed that several genes' expression increases linearly with Rpos concentration: these were dubbed 'linear' genes. However, several genes didn't follow this linear trend. Several genes were transcribed *more* than anticipated under the linear hypothesis at low RpoS concentrations: these were called 'sensitive' genes.

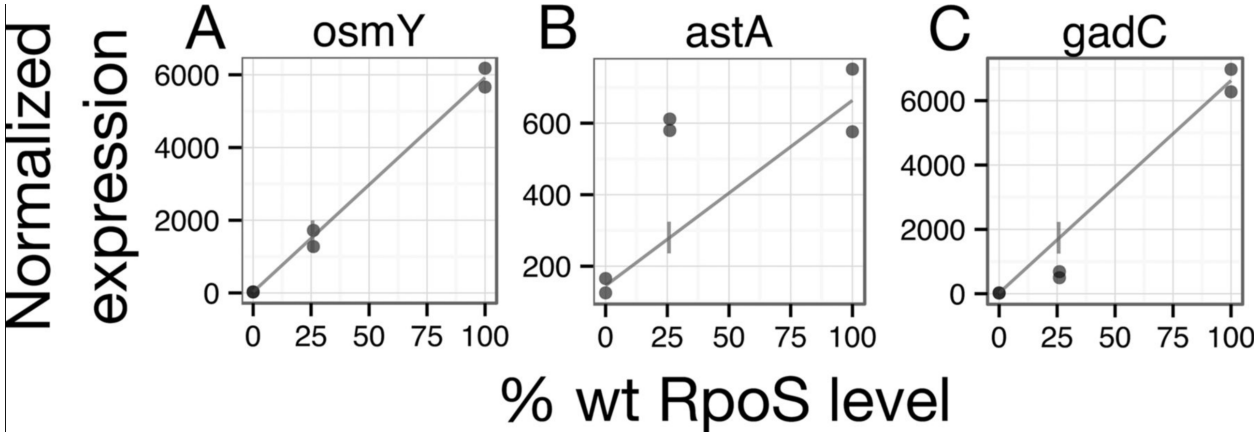


Figure 1: (A) osmY: a linear gene (B) astA: a sensitive gene, (C) gadC: an insenstive gene. Figure obtained from Wong *et al.* 2017

Other genes were transcribed *less* than anticipated under the linear hypothesis at low RpoS concentrations: these were dubbed 'insensitive' genes (Figure 1) (Wong *et al* 2017). Wong and colleagues hypothesized that sensitivity to Rpos could be a mechanism to control the **timing** of genes involved in the general stress response.

We aimed to investigate this hypothesis using an *in silico* tool, ImpulseDE2. ImpulseDE2 is a serial Time Course (TC) Differential Expression tool which fits impulse/sigmoidal models to expression trajectories and compares these dynamics models to constant/reduced models to determine differential expression over time or between time courses. The key advantage of ImpulseDE2 is that its models are parametrized by *biologically meaningful* parameters (ex. onset time), allowing researchers to leverage the parameter values to directly address biological questions (Figure 2).

We applied ImpulseDE2 to an time course (TC) RNA-Seq dataset generated by Professor Dan Stoebel. Two strains of *E. coli* - a WT strain and another strain, designated delta_RpoS, with the *rpos* gene knocked out - were allowed to grow exponentially for 150 minutes, inducing cell starvation and RpoS production (Figure 3). A thoughtful pipeline employing best-performing time course differential expression methods (DESeq2, NEXT maSigPro, and ImpulseDE2) was run on the dataset to identify differentially expressed genes. The onset time (t) parameter fitted by ImpulseDE2 was extracted for monotonically differentially expressed genes, and simulations were conducted to better understand and visualize the effects of dispersion, number of time points, and outliers on all sigmoidal parameters. Then, statistical tests were applied to the timing parameters of sensitive and insensitive genes to test the hypothesis that graded sensitivity to RpoS coordinates the timing of gene expression in response to stress.

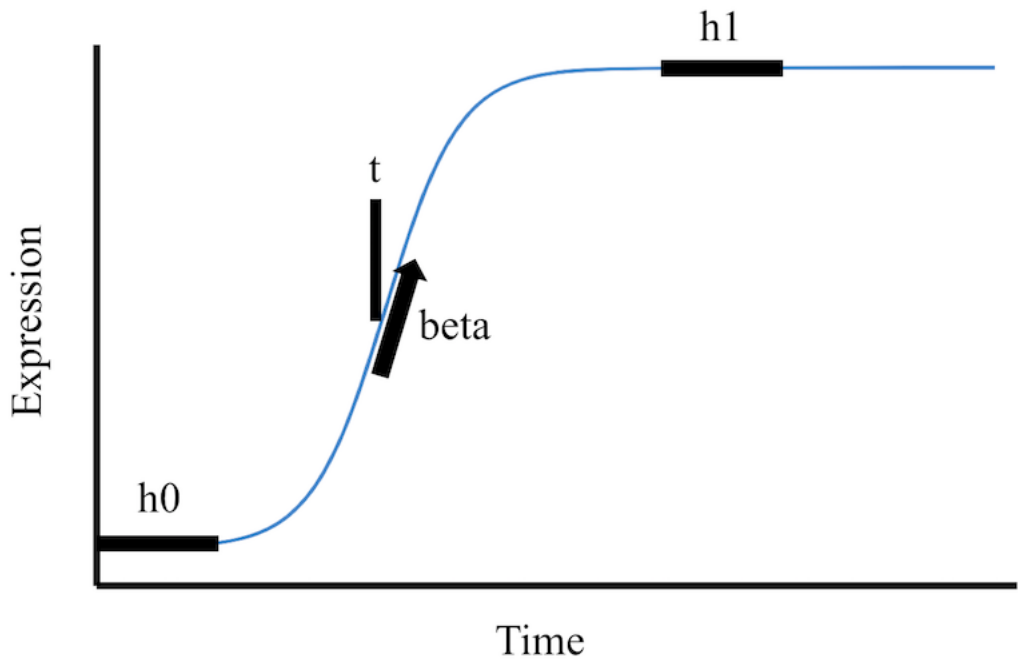
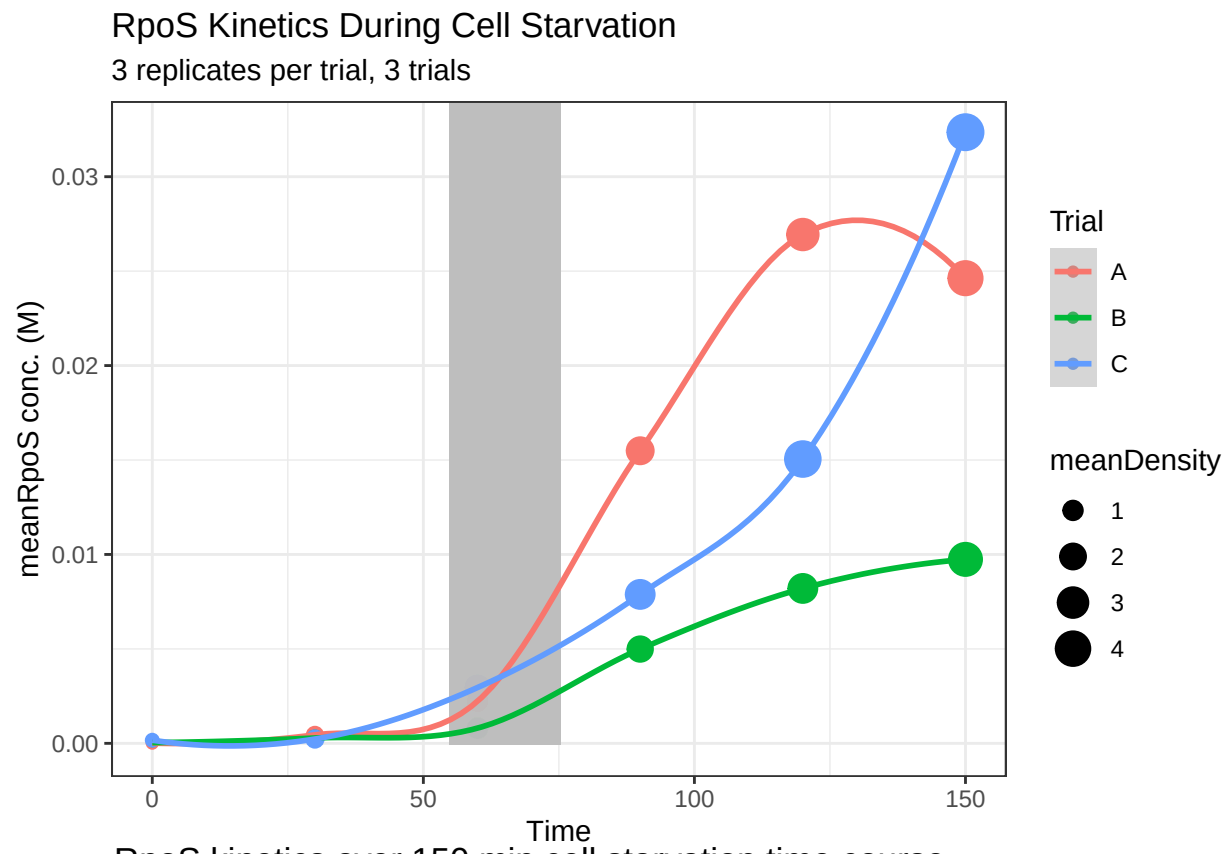


Figure 2: Sigmoid model for expression trajectories, parametrized by initial expression level (h0), onset time (t), onset slope (beta), and peak expression level (h1)



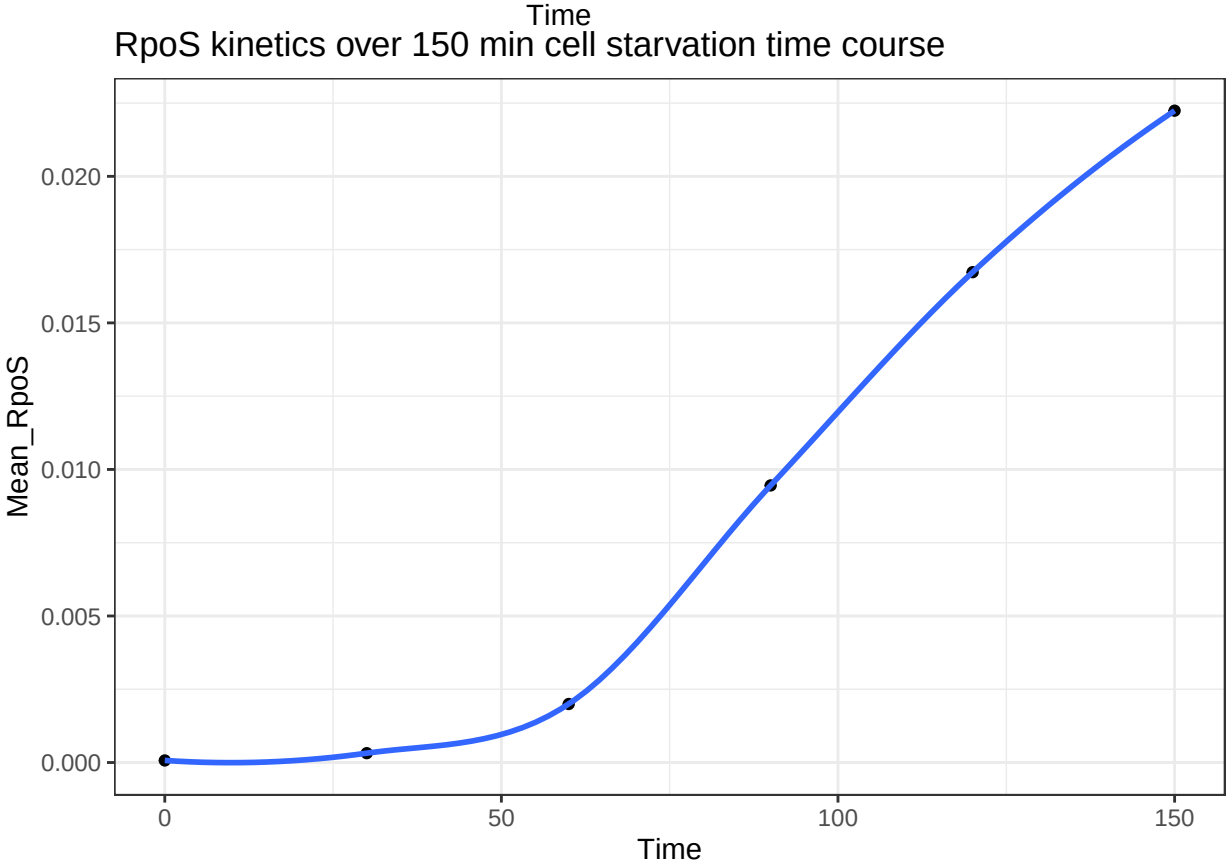


Figure 3: RpoS kinetics over 150 min starvation time course.

Methods

Data Pre-Processing

RNA-Seq data was obtained in the form of read count data. Code developed by Madison Hobbes was used to parse the gene identifiers, extract sequence type, gene names, and other useful information. Since this experiment focuses on the kinetics of gene expression, the read count data was filtered for Coding Sequences (CDS's). Rows with NA counts and 0 counts were filtered out, and duplicate genes were resolved.

Identifying Differentially Expressed Genes

Three *in silico* differential expression tools were employed to identify DEGs between the WT and delta_RpoS time courses. Previous research shows that two-sample comparisons using DESeq2 and serial tools maSigPro and ImpulseDE2 are the best performing methods to identify differentially expressed genes in TC experiments (Spies *et al.* 2019). Morever, combining gene lists generated by multiple tools reduces false positives without compromising sensitivity and is thus recommended for optimal DEG identification in TC experiments (Spies *et al.* 2019).

DESeq2 (Love *et al* 2014) is an algorithm that models the read counts using a negative binomial distribution and uses a relative log expression (RLE) method to normalize the read count data across between-sample effects. DESeq2's main innovation is its Bayesian Shrunken Dispersion and LFC estimates, which improve stability and interpretability of these estimates (Love *et al* 2014). DESeq2 is one of the most widely accepted and empolyed algorithms for differential expression analysis. However, DESeq2 is not built to analyze serial (time course) data, and its categorical treatment of serial data leads to loss in power to identify DEGs. Thus, DESeq2 alone is an insufficient method to identify DEGs in TC experiments.

NEXT maSigPro (Nueda *et al.* 2014) is a differential expression algorithm designed to analyze serial/TC RNA-Seq data. maSigPro models the gene expression value using polynomial regression with a NB Generalized Linear Model. The model's parameters are fit using maximum likelihood, and the log likelihood ratio test is applied to the full model and the null model (a flat/alternative model fit to a control time course). In the second step of maSigPro, the goodness of fit, R^2, is computed for each optimized gene model, allowing filtering based on genes with clear expression trends. maSigPro requires data to be normalized *a priori* (we used the RLE method implemented through DESeq2) and contains a naive dispersion estimate of 10, which according to the authors does not have a substantial impact on downstream DEG identification (Nueda *et al.* 2014). maSigPro is a robust DE tool designed for analysis and interpretation of serial/TC RNA-Seq data.

ImpulseDE2 (Fischer *et al.* 2018) is another differential expression algorithm designed to analyze serial/TC RNA-Seq data. ImpulseDE2 fits impulse/sigmoidal models to the read count data by estimating the parameters using the Broyden–Fletcher– Goldfarb–Shanno algorithm. Gene-wise dispersion estimates are generated using the bayesian shrinkage approach used in DESeq2. Differential expression is assessed using the log likelihood ratio test comparing the likelihood of the alternative and null (flat or control fit) models.

Using a shiny app created by Annie Cohen, we noticed that many of E. Coli's DEGs appeared montonically differentially expessed. To capture this global expression change, we ran the categorical DE tool DESeq2 a two-sample contrast across the delta_Rpos and WT strains at the final time point (150min), when expression differences between the two time courses were putatively maximized. maSigPro was run with default arguments and results were filtered by R^2 of 0.6 to obtain DEGs with relatively clear expression trends. ImpulseDE2 was run with defaults and with the specification to differentiate between transient (impulse) and monotonic (sigmoidal/linear) DEGs.

Genes that were identified as significantly differentially expressed at a Benjamini-Hochberg corrected p value of less than 0.01 by ImpulseDE2 and at least one additional tool were included in our DEG list. 1007 genes were labeled as DE and were subjected to further analysis (Figure 4).

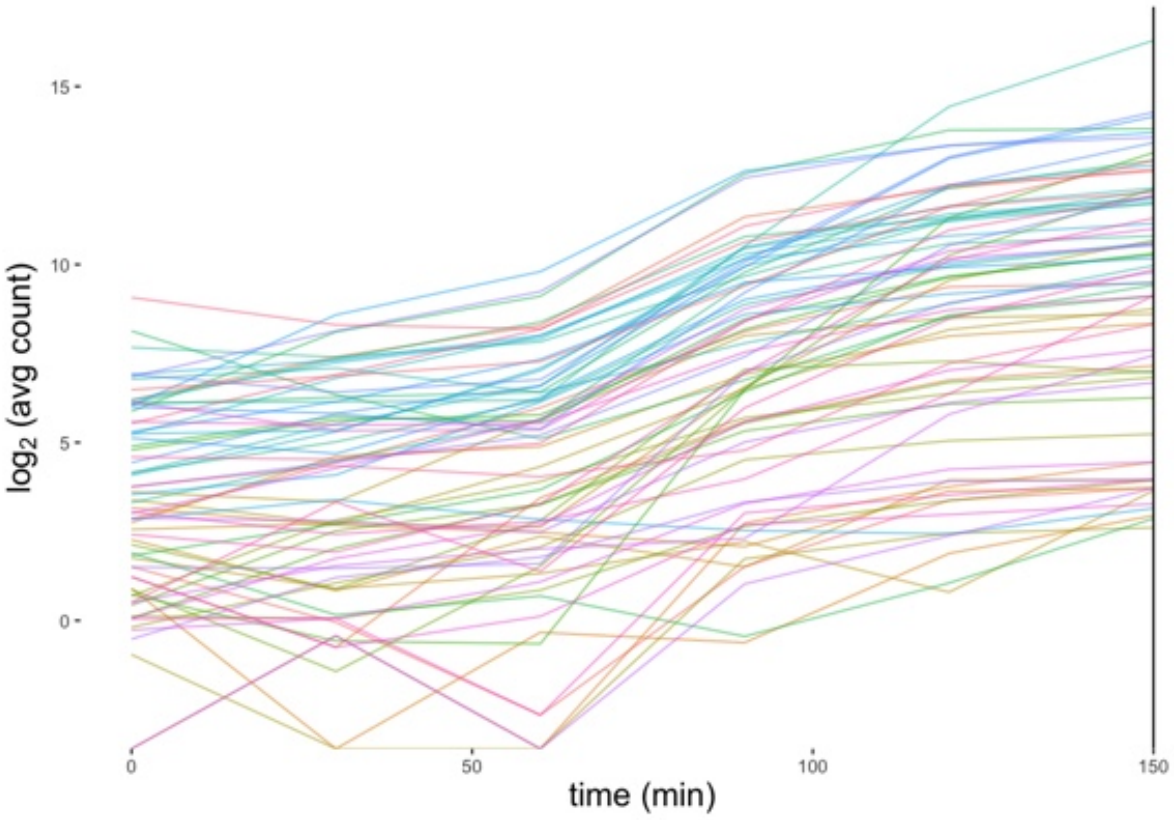
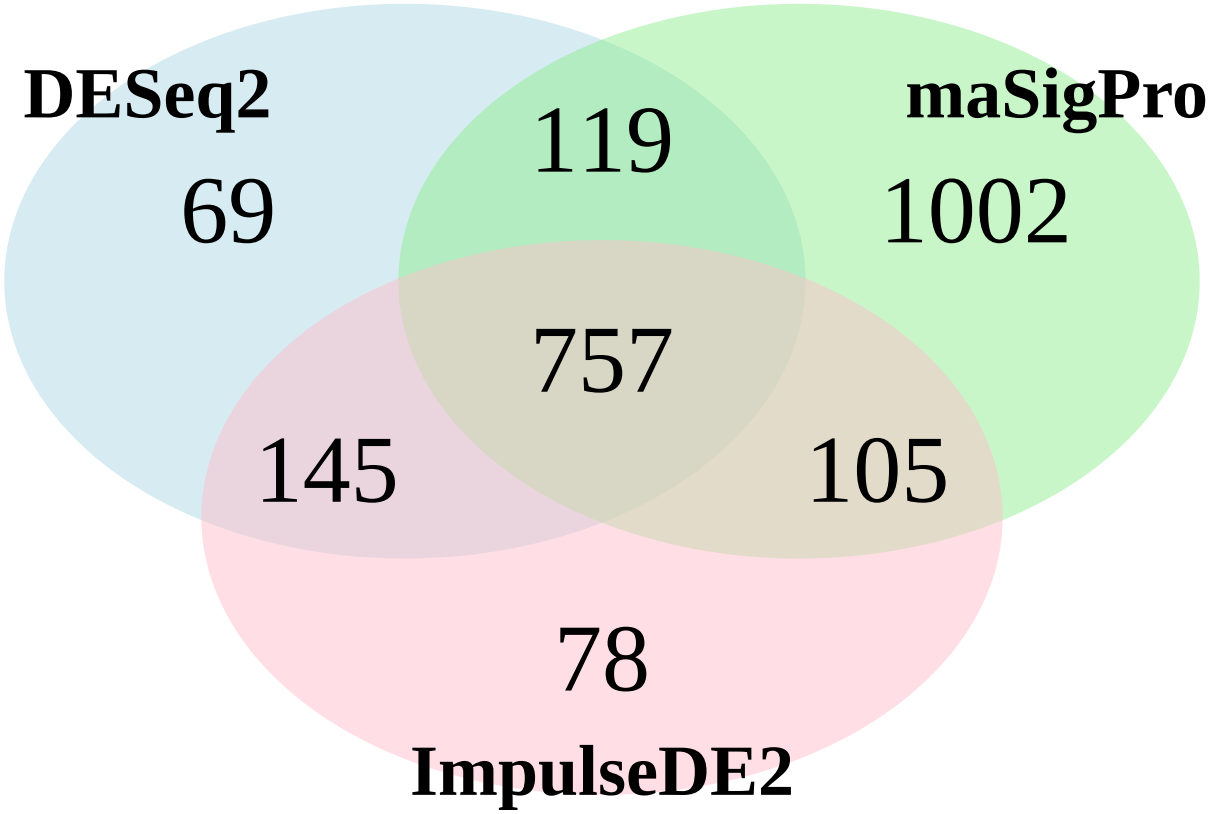


Figure 3: Shiny App image illustrating the monotonic expression profiles for DEGs



Simulations for Assessment of Sigmoidal Model Parameters

Simulations were conducted to understand the effects of dispersion, number of time points, and outliers on the sigmoid model's parameters. The basic simulation structure took in mean read counts estimated from the fit of a sigmoidal model to a real *E. coli* expression trajectory, and generated 100 simulated expression trajectories by adding appropriate NB noise to each profile. The amount of noise added to each profile depended on the DESeq2-estimated dispersion parameter obtained for each gene. Then, sigmoidal models were fit to the simulated trajectories using ImpulseDE2, and the distribution of simulated parameters were compared to the initial parameter values.

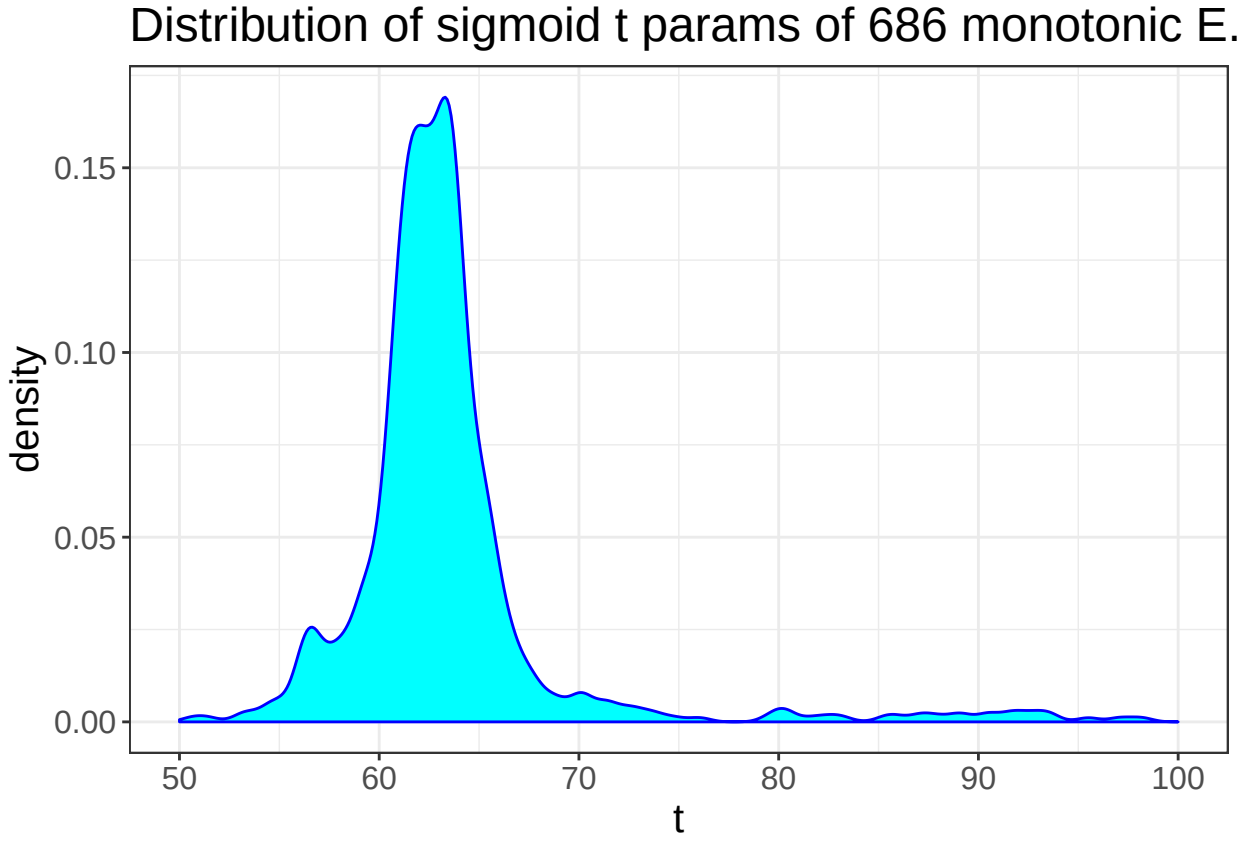
Results

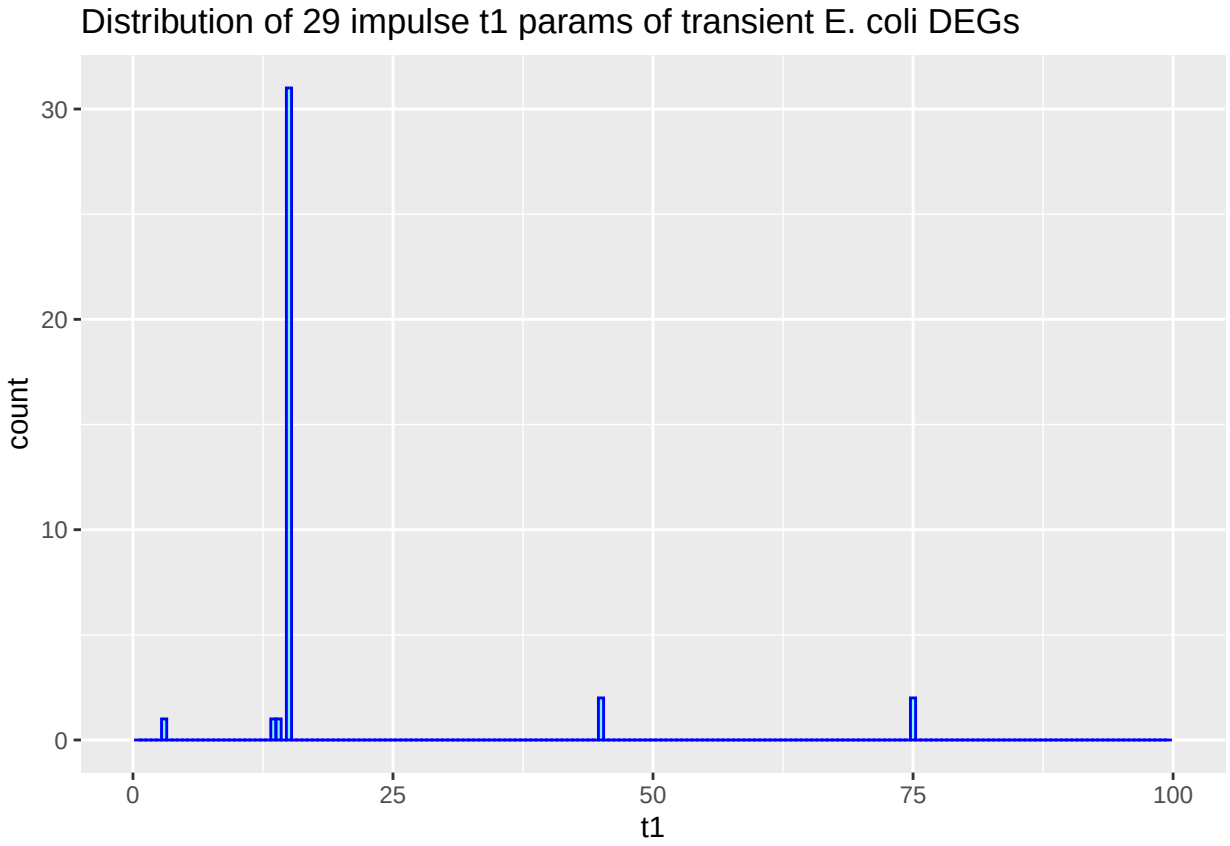
Visualizing onset times for DEGs and Sensitive/Insensitive Genes

For genes identfied as differentially expressed by our pipeline, onset timing parameters were extracted and plotted in a density plot. Differentially expressed genes were divided into three groups according to their trajectories: monotonic, transient, and unknown trajectories. Monotonic DEGs were the most abundant (686), followed by unknown (282), followed by transient (39). Monotonic expression trajectories were described using the *sigmoid model* parametrized by four parameters. Transient trajectories were described using the *impulse model*, an extension of the sigmoid model that describes the on/off dyanmics by taking the scaled product of two sigmoid functions.

The distribution of t params for monotonic DEGs were plotted. Around 92% of the monotonic DEGs have onset times between 55 and 75 minutes, suggesting that the vast majority of monotonic DEGs turn on at around the same time. However, there exists weak bimodality/trimodality in the density curve, perhaps suggestive of waves of *E. coli* genes with coordinated expression times.

Interestingly, the vast majority of onset time params for genes identified as *transiently* differentially expressed fell around the 14 minute mark.





Simulations

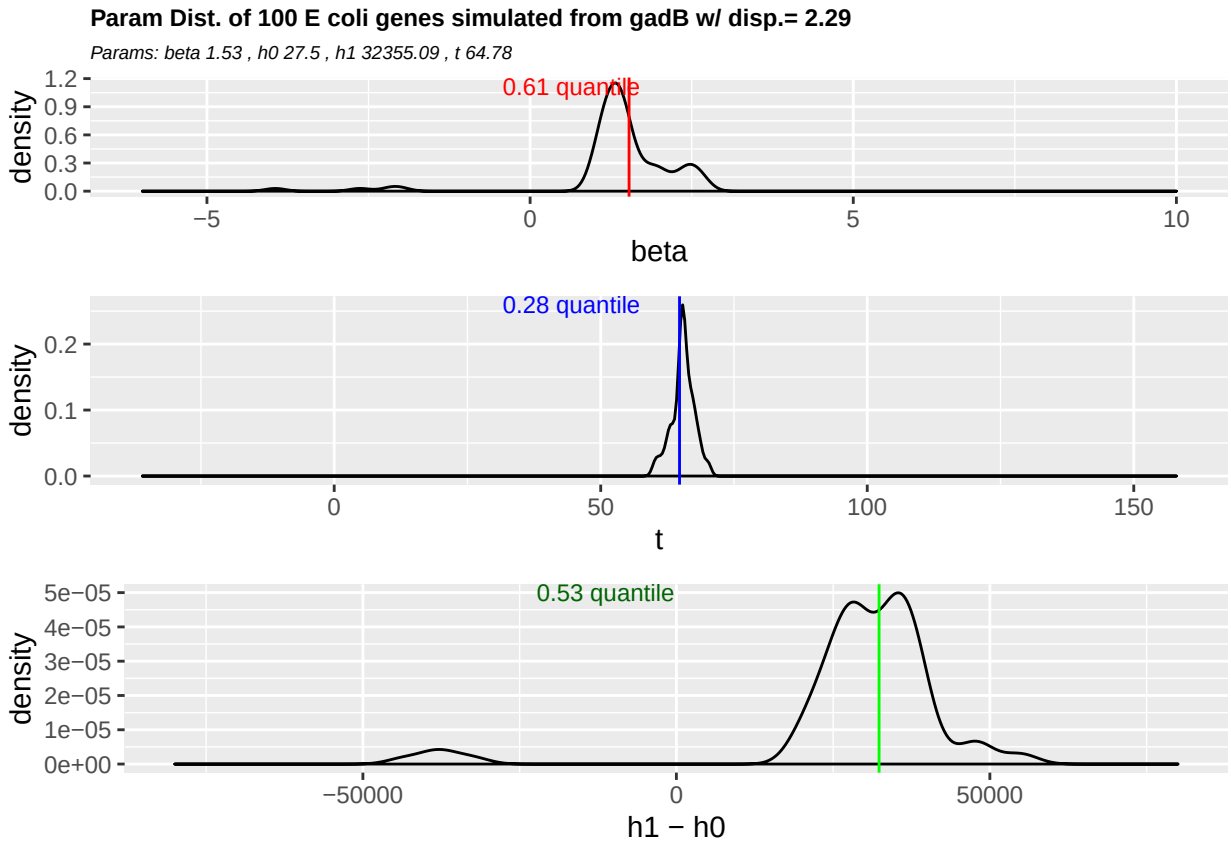
To better understand the effects of dispersion, number of time points, and outliers on the sigmoid model's biologically-meaningful parameters, simulations were conducted from real data from *E. coli* DEGs. In the following plots of simulated sigmoid trajectories, the verical dark-blue lines represent the initial onset time (t) parameter used to generate all the simulated sigmoid functions. The vertical magenta line represents the median simulated onset time parameter.

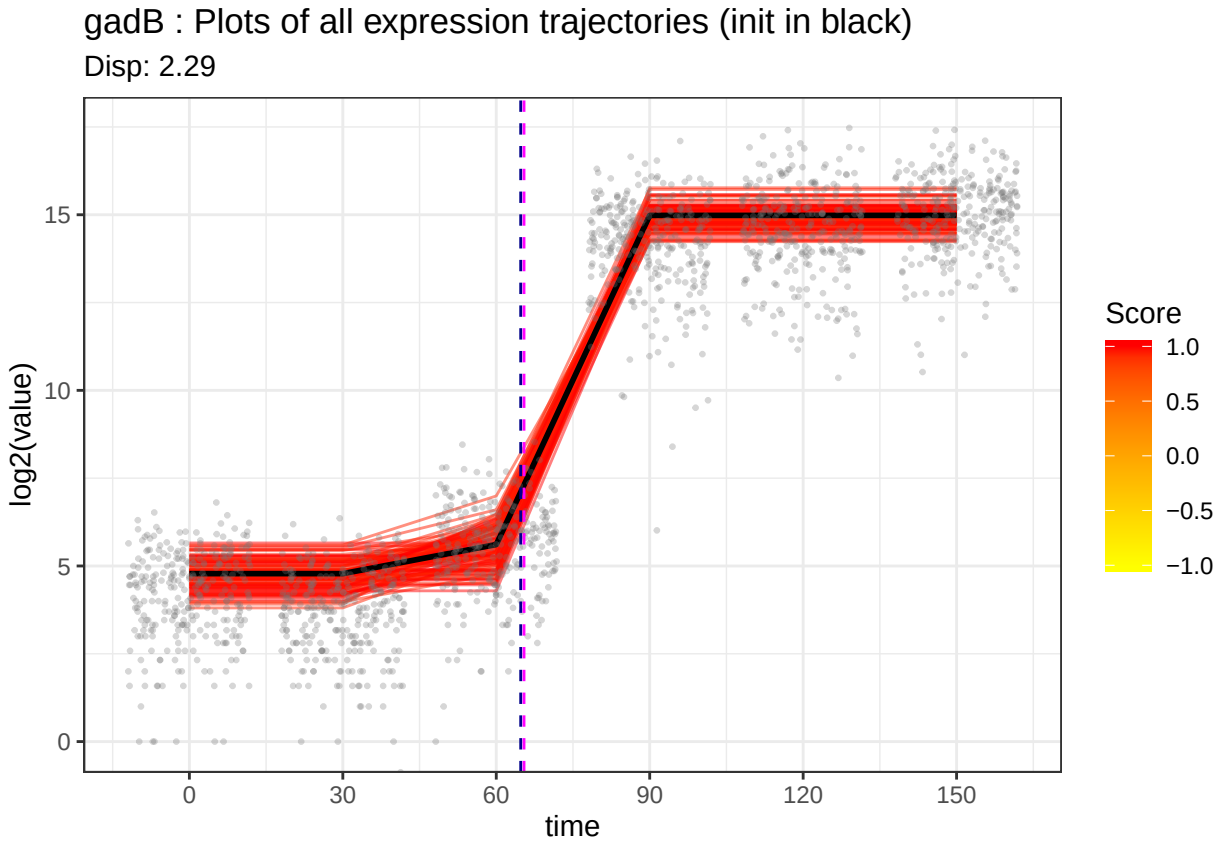
Effect of Dispersion on Sigmoid Parameters

The simulation pipeline was conducted under three conditions: with the *a priori* DESeq2-shrunk dispersion parameter estimates, a 100-fold *increase* in the dispersion parameter estimate (i.e. a 100-fold reduction in the noise caused by dispersion), and 100-fold *decrease* in the dispersion parameter estimate (i.e. a 100-fold increase in the noise caused by dispersion).

With unchanged dispersion, beta, t, h0, and h1 all appeared highly reproducible, as the initial parameter values were relatively central to the distributions of simulated values. All expression trajectories were highly correlated (>0.94) with the initial trajectory, and the simulated t-params showed low variability with an IQR of 2.11 minutes.The difference between the initial t param and median simulated t param was 0.64, suggestive of stability of the t parameter.

```
## Warning: Removed 12 rows containing non-finite values (stat_density).
```

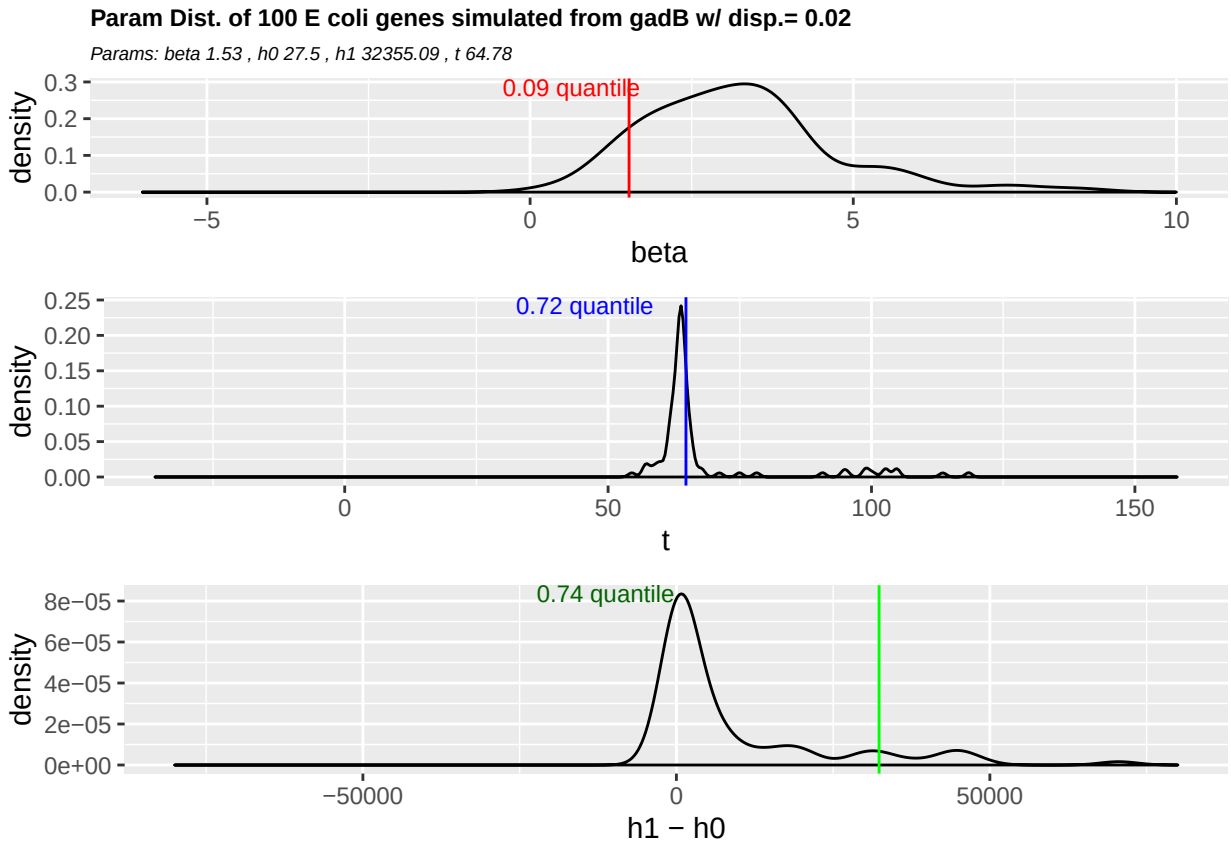


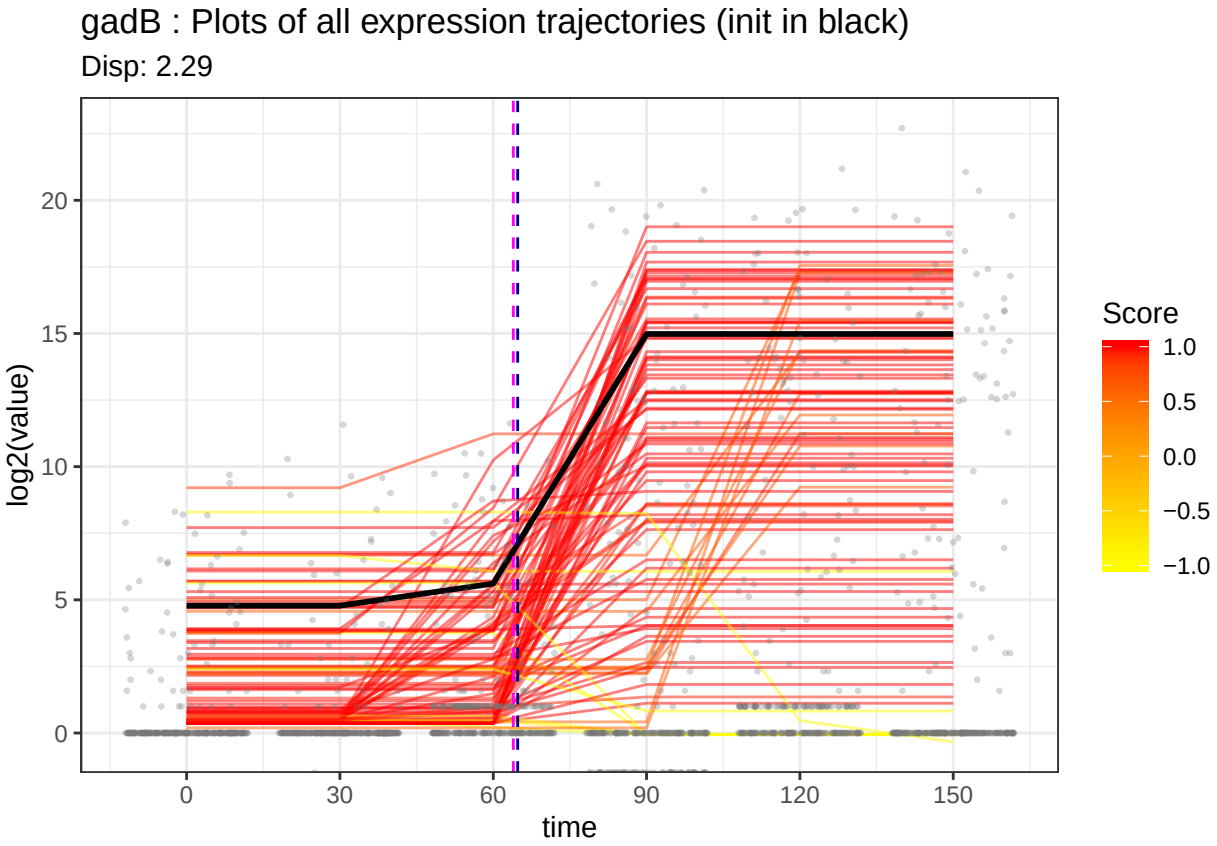


Increasing the dispersion-based noise lead to profound variability in the resulting simulated trajectories. The

slope parameter (beta) was most affected by the additional noise; this distribution of simulated betas was not centered around the initial value (9th quantile) had a substantial right skew suggestive of high variabilities in the simulated slope parameters. Amplitude parameters h0 and h1 were also affected by the dispersion-based noise, as the distribution of their difference (h1-h0) shifted considerably left from the initial value (over 30,000 units).

However, the effect of dispersion-based noise was minimal in the onset time (t) parameter; the distribution of simulated onset times did not profoundly shift or widen. The variability of the simulated t parameter remained quite low: IQR=2.34 (an 10.9% increase from the standard dispersion simulation). The stability of the t parameter was also illustrated by a small difference between initial t and median simulated t: 0.85 (32.8% increase from standard dispersion simulation). Thus, t appears to be the parameter most robust to dispersion-based noise in the sigmoidal model, an encouraging finding for RNA-seq experiments with few replicates hoping to address timing-related questions.

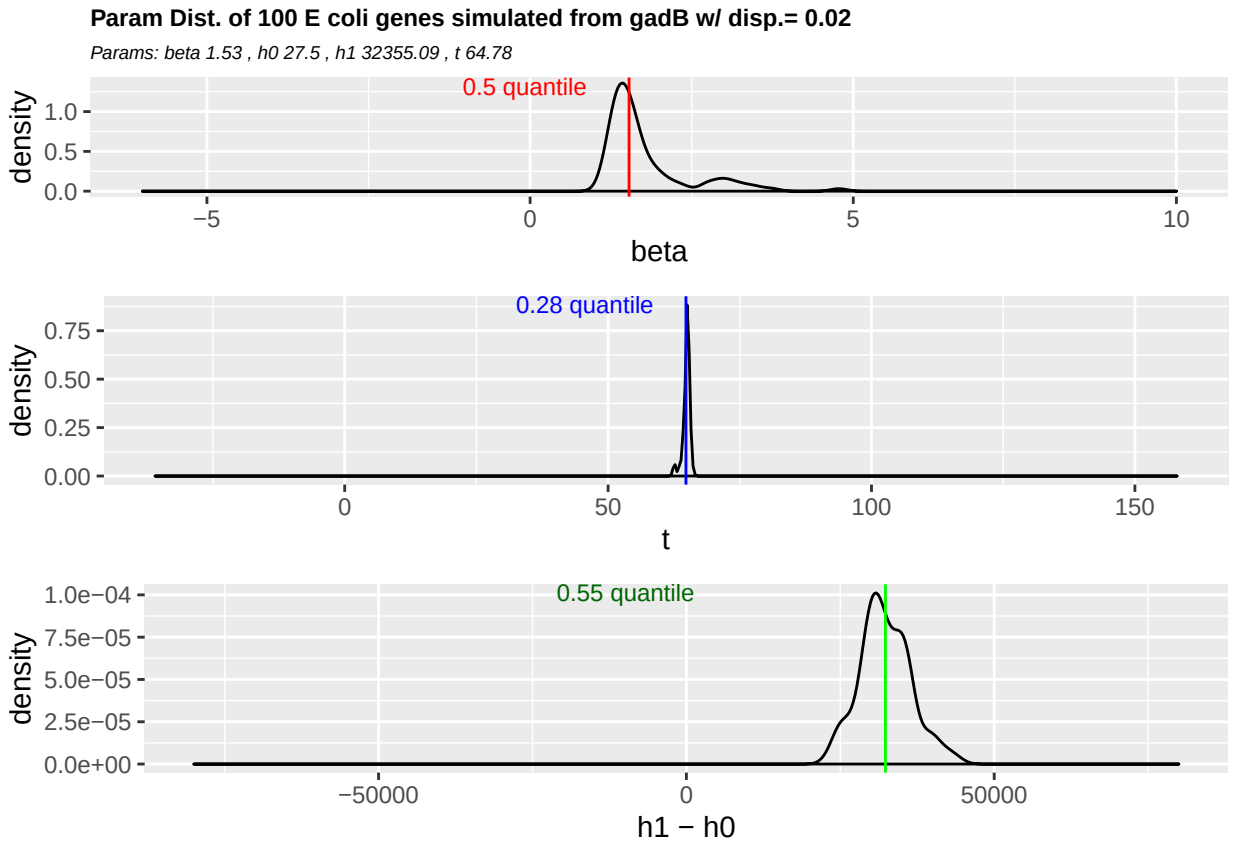


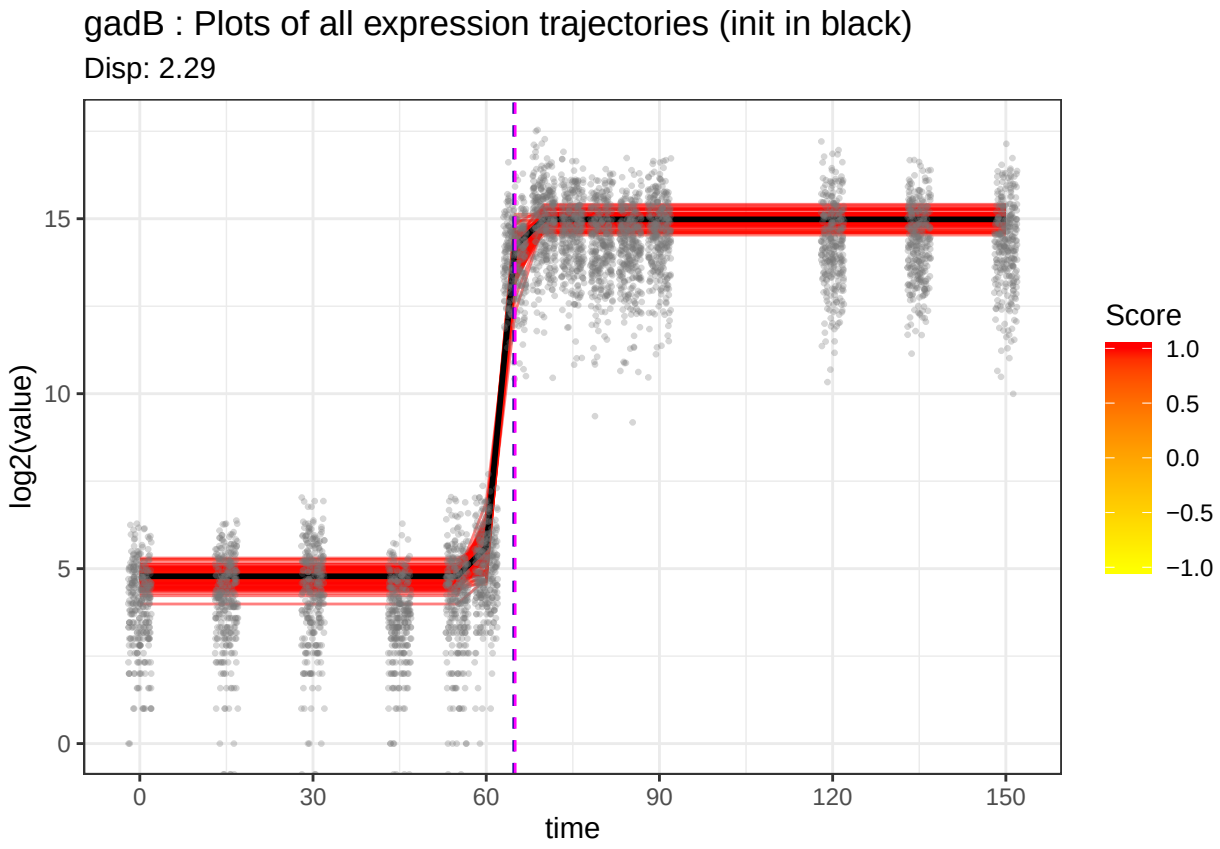


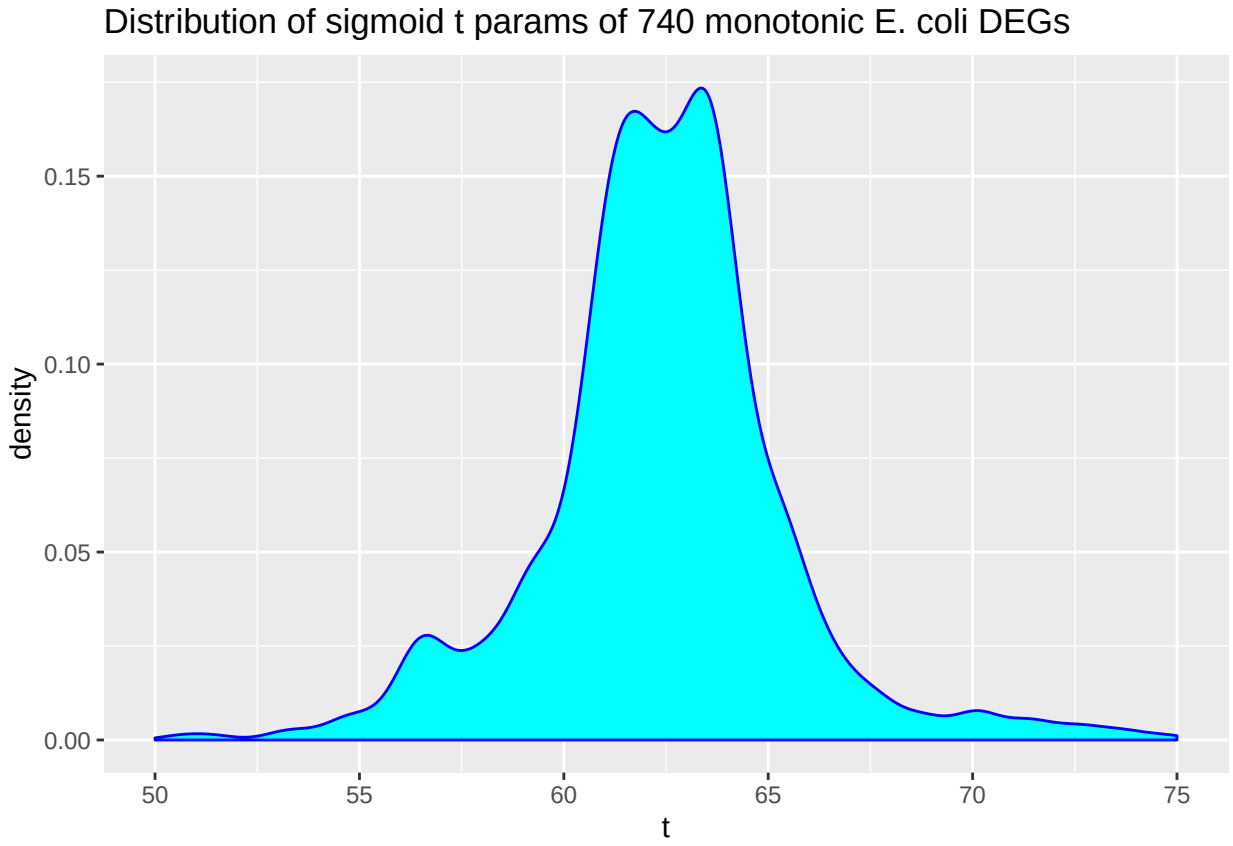
Effect of more TPs on Sigmoid Parameters

For the following simulation, gadB's sigmoid model was extracted and mean read counts were calculated at time points (TPs) 0, 15, 30, 45, 55, 60, 65, 70, 75, 80, 85, 90, 120, 135, and 150 minutes using sigmoidal imputation to simulate data collection at more time points. Simulated data at these time points were generated and sigmoid models were fit. As anticipated, the increase in number of time points afforded better resolution in the onset time (t) and slope (beta) parameters: the width of these distributions of simulated values shrunk and were both centered around the initial values.

To illustrate, the addition of more TPs resulted in a difference between initial onset time and median simulated onset time of 0.22 mins (65.6% reduction compared to the standard simulation). The t-param IQR was also greatly reduced compared to the standard simulation (71.1% reduction). The addition of more time points also eliminated outlier simulated t parameters, as the range of simulated t params was around 4 minutes. This simulation affirms the importance of increasing number of time points in obtaining better timing resolution of gene expression trajectories. According to the distribution of monotonic onset time parameters below, around 92% of the DEGs have onset times between 55 and 75 minutes. Thus, concentrating RNA extractions at TPs between 55 and 75 minutes could yield more resolution in identifying timing-related events in *E. coli* cell starvation.


Param Dist. of 100 E coli genes simulated from gadB w/ disp.= 0.02
Params: beta 1.53 , h0 27.5 , h1 32355.09 , t 64.78
0.5 quantile
density
beta
0.28 quantile
t
0.55 quantile
h1 − h0

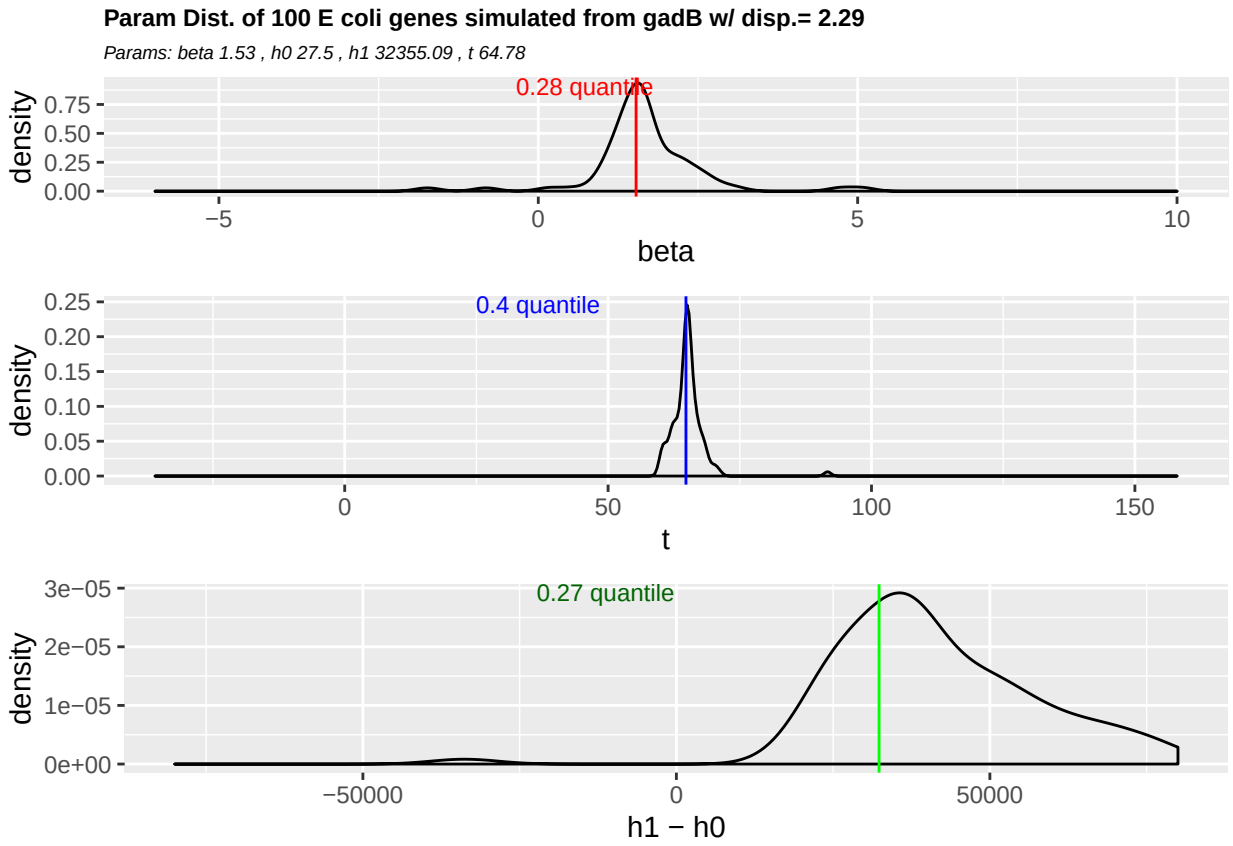

gadB : Plots of all expression trajectories (init in black)
Disp: 2.29
log2(value)
time
Score

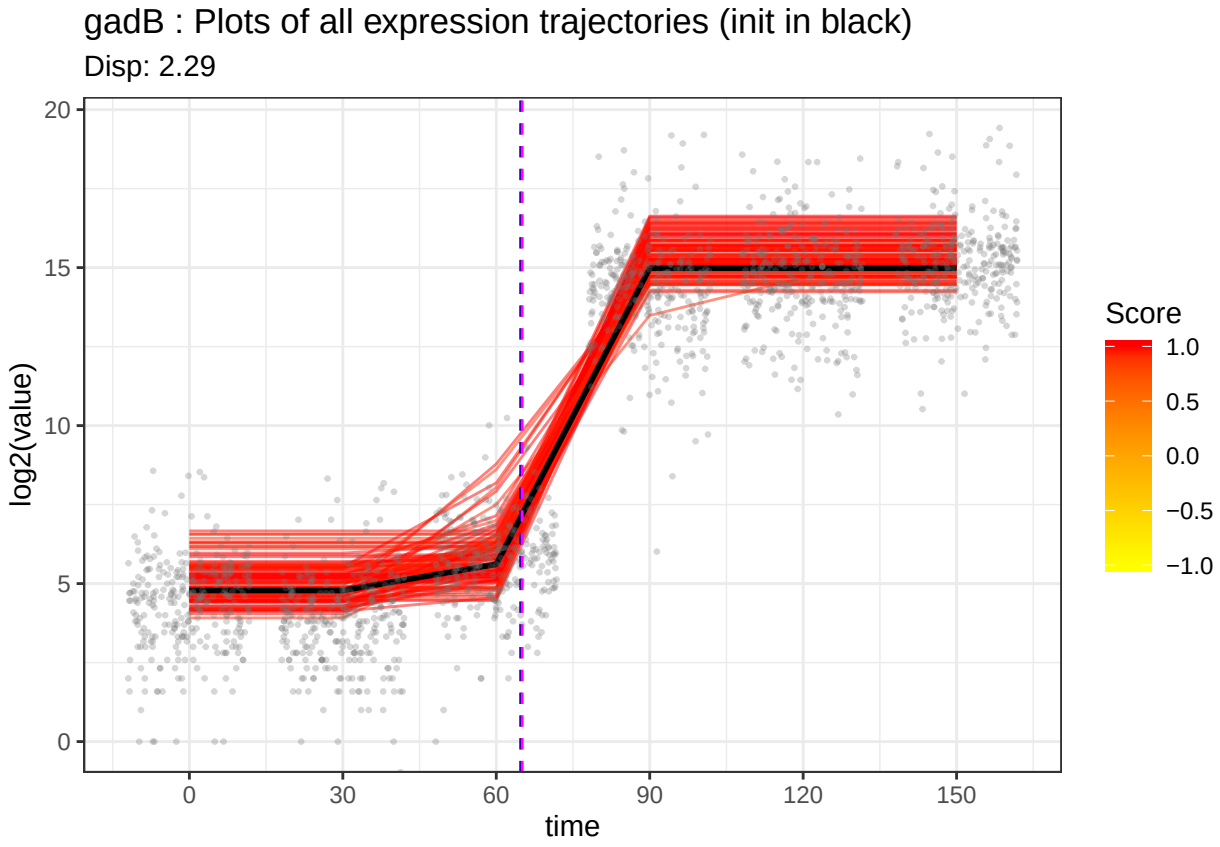


Effect of outliers on Sigmoid Parameters

To ascertain the effect of ouliers on the sigmoid parameters, 90 simulated gene trajectories were divided into 6 groups based on the TP at which the outlier would be generated (ex. Group 1 would have an outlier at the 0 min TP, Group 2 woud have an outlier at the 30 min TP, etc). For the 15 genes in each group, the normalized read count for one replicate corresponding to the group's time point was multipled by a factor of 10. Sigmoid models were fit and variability was assessed.

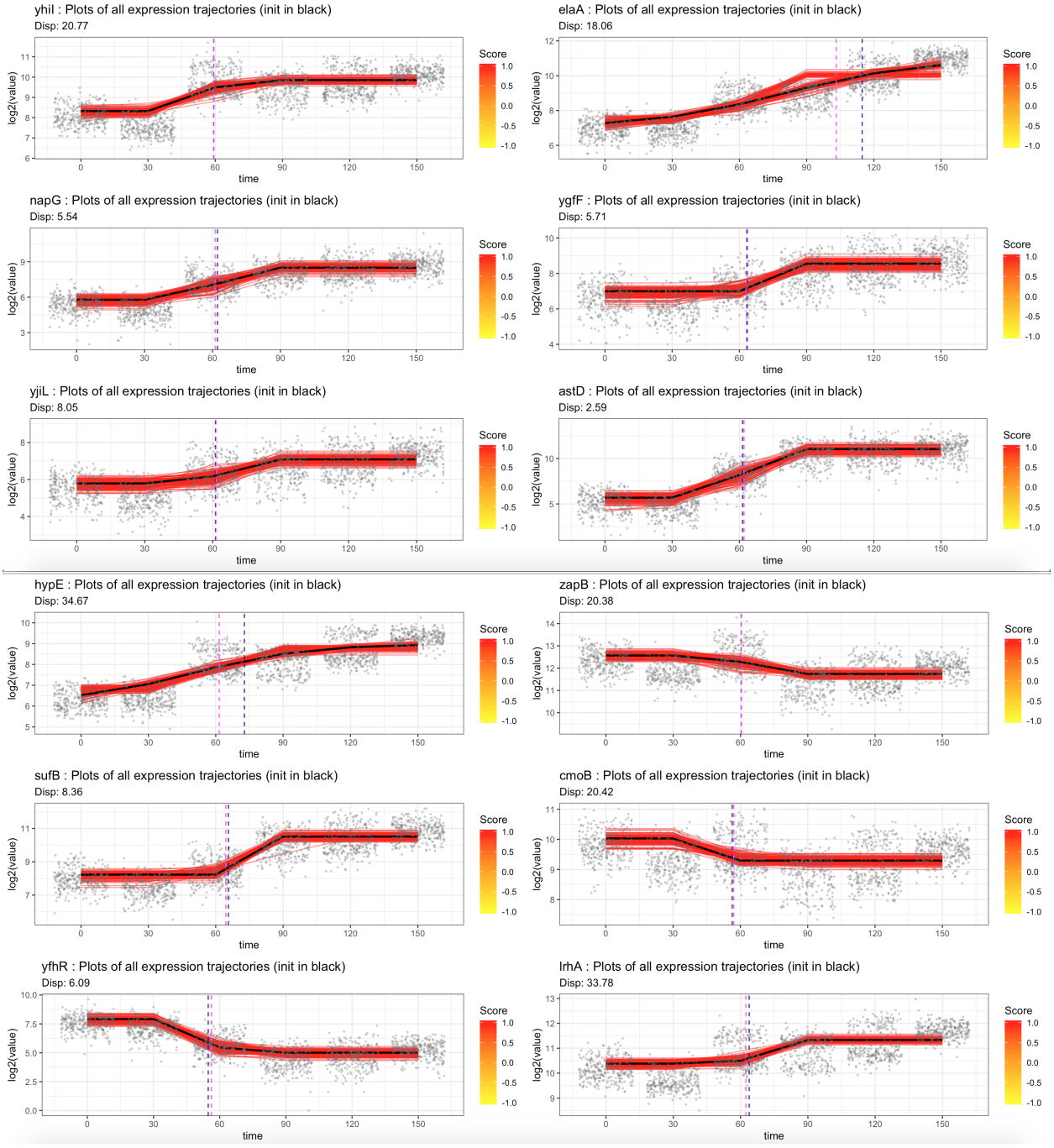
The IQR for the onset-time parameter in the simulation with outliers was only marginally inflated compared to the standard simulation: IQR with outliers was 2.32 (10.0% increase compared to the standard simulation). And rather unexpectedly, the difference between initial t and median simulated t params actually decreased with the addition of outliers: 0.29 (54.7% decrease from standard simulation). Thus, sigmoid onset time params appear quite robust to outliers, another boon for biologists aiming to apply sigmoids to modeling noisy gene expression data.

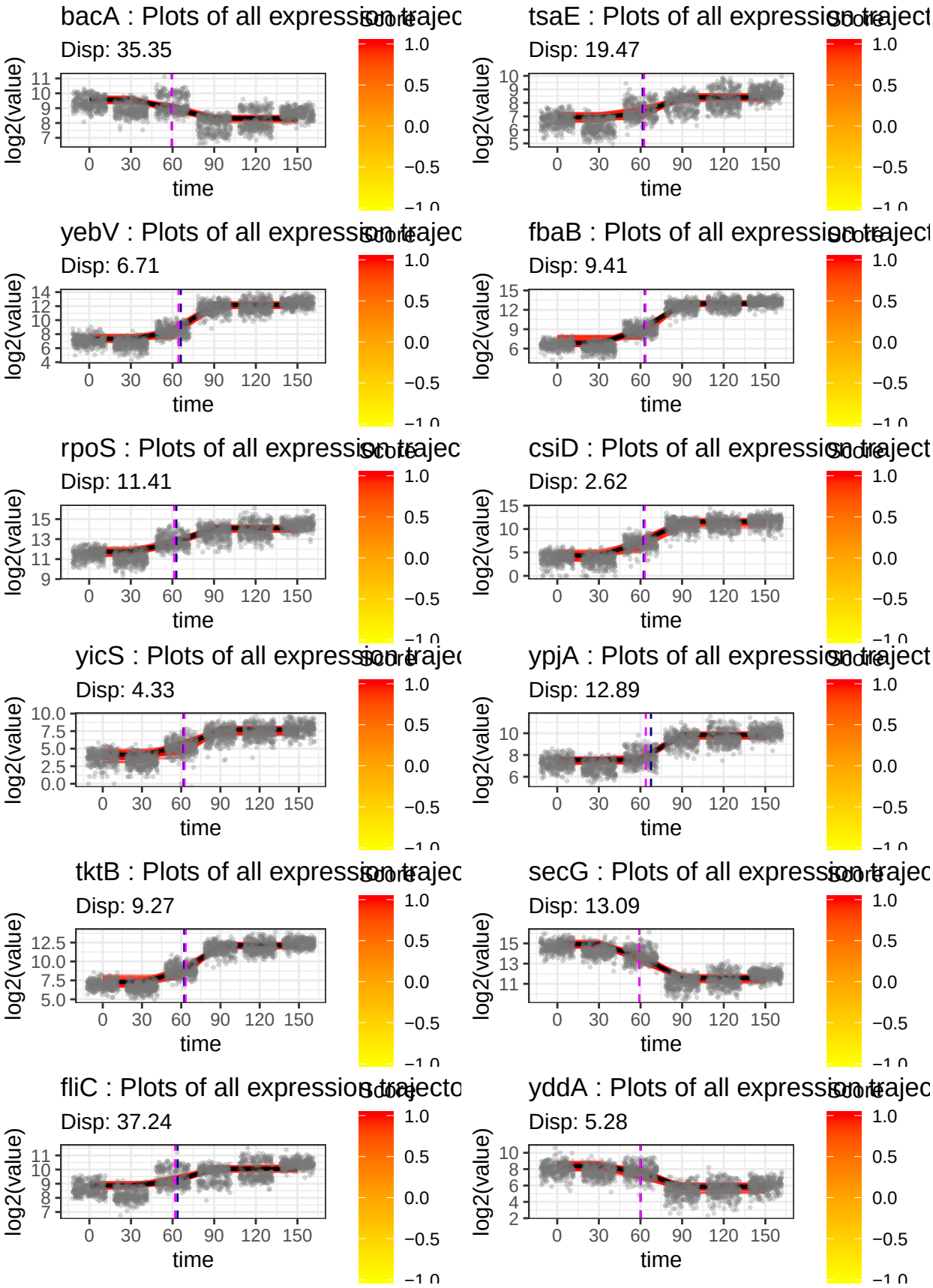

Param Dist. of 100 E coli genes simulated from gadB w/ disp.= 2.29
Params: beta 1.53 , h0 27.5 , h1 32355.09 , t 64.78
0.28 quantile
density
beta
0.4 quantile
t
0.27 quantile
h1 − h0


gadB : Plots of all expression trajectories (init in black)
Disp: 2.29
log2(value)
time
Score

The Sigmoid Model is Descriptive of Many Monotonic DEGs

All previous simulations were done with gadB as the reference gene. As shown below, the sigmoid model is descriptive of many monotonic DEGs. A random sample of 12 monotonic DEGs were selected and their expression profiles, along with their 100 simulated expression profiles were plotted. Onset time parameters between simulated and real trajectories were nearly identical across all simulations.



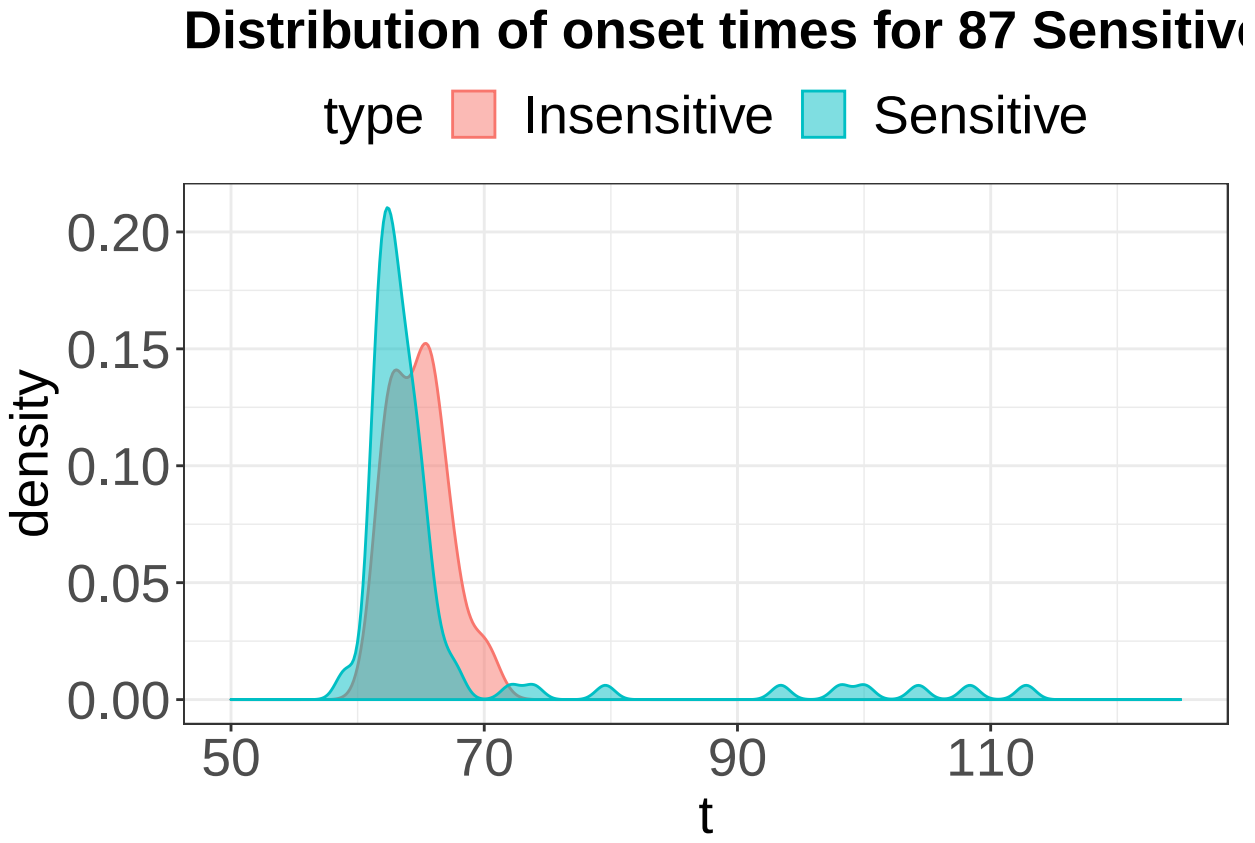

bacA : Plots of all expression trajec
Disp: 35.35
tsaE : Plots of all expression traject
Disp: 19.47
yebV : Plots of all expression trajec
Disp: 6.71
fbaB : Plots of all expression traject
Disp: 9.41
rpoS : Plots of all expression trajec
Disp: 11.41
csiD : Plots of all expression traject
Disp: 2.62
yicS : Plots of all expression trajec
Disp: 4.33
ypjA : Plots of all expression traject
Disp: 12.89
tktB : Plots of all expression trajec
Disp: 9.27
secG : Plots of all expression trajec
Disp: 13.09
fliC : Plots of all expression trajecto
Disp: 37.24
yddA : Plots of all expression trajec
Disp: 5.28
log2(value)
time
Score

Sigmoid parameters of Sensitive and Insensitive Genes

After identifying the limitations and utilities of the sigmoid model, we applied the sigmoid model and its onset time parameter to address whether sensitive genes turn on significantly earlier in the cell starvation time course than insensitive genes.

As shown below, the onset times for 87 sensitive and 15 insensitive monotonic DEGs overlap somewhat, however, the density plot of onset times for the insensitive genes appears bimodal, with one peak overlapping with the sensitive peak and the other peak occuring later than the sensitive peak. Unexpectedly, a number of sensitive genes have onset times later than the large, single peak: 8 sensitive genes turned on later than 75 minutes (prpE, treA, ydcV, ydgD, astB, yeaG, crr, and cysQ).

Wilcox-rank sum tests with unequal variance and confidence levels of 0.95 were conducted on the onset times for sensitive and insensitive genes to ascertain whether there was a signficant difference between the onset times of these two groups. Considering all onset times for the 102 total genes, the true location was nearly significantly different from 0 (p=0.05275). When the 8 potentially misclassified sensitive genes noted above were removed from the analysis, the true location shift was highly significantly different from 0 (p=0.008586), suggesting that sensitive genes may turn on significantly earlier than insensitive genes.



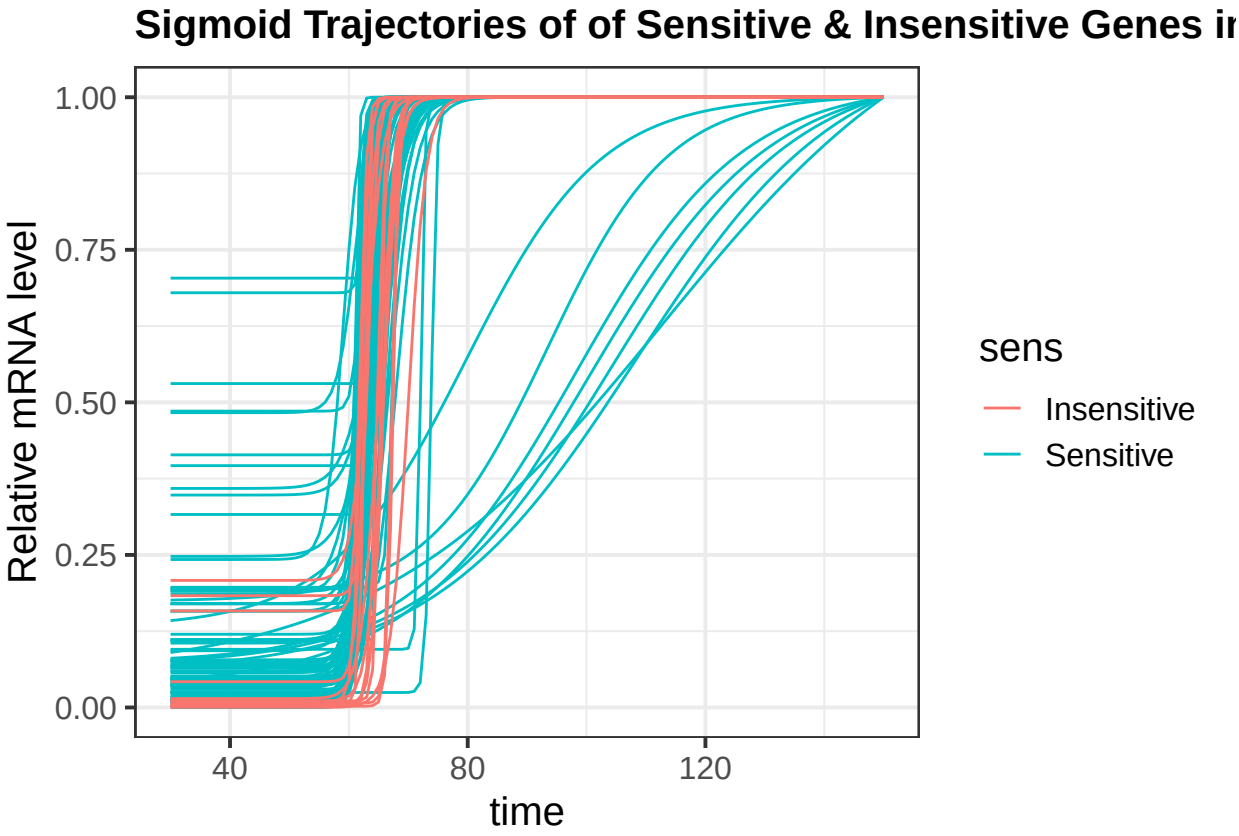

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  sensitivedf$t and insensitivedf$t
## W = 447, p-value = 0.05275
## alternative hypothesis: true location shift is not equal to 0
```

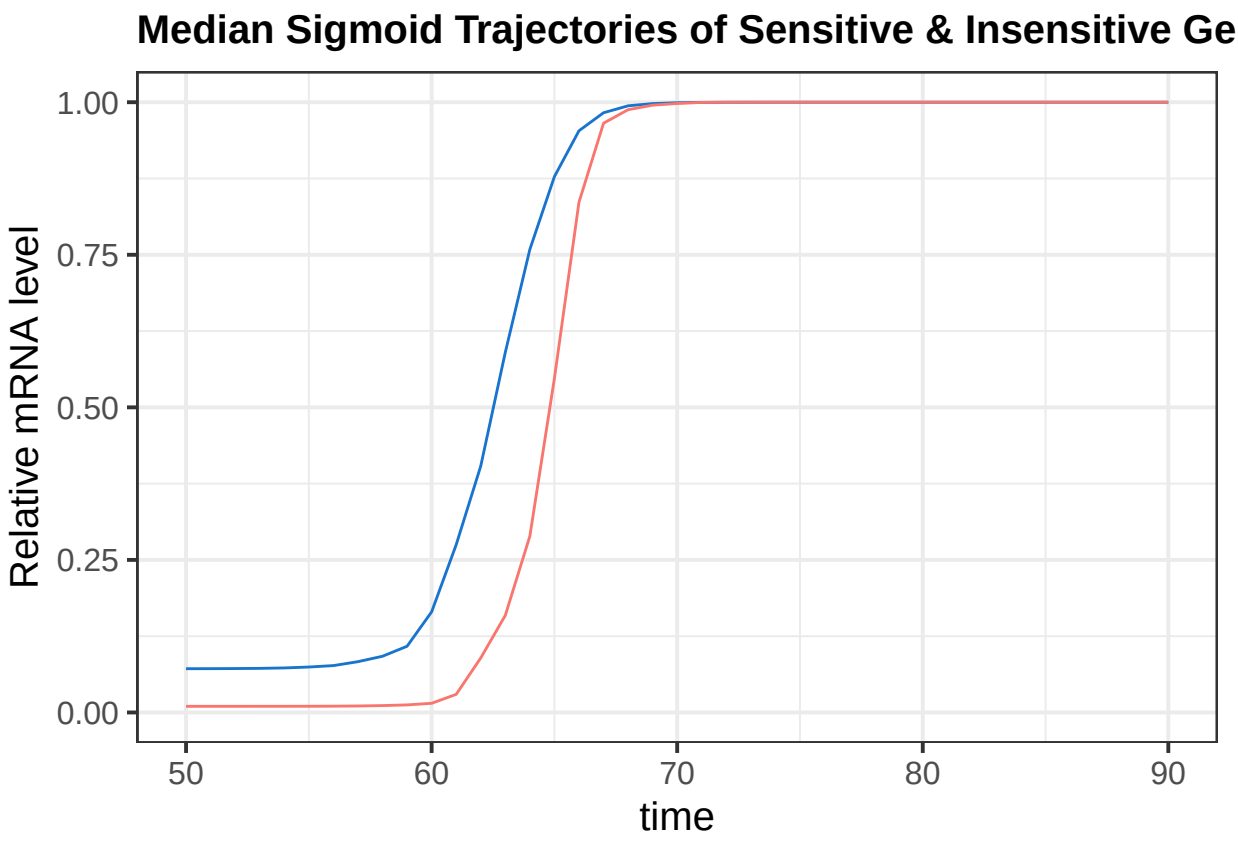
```
##
##  Wilcoxon rank sum test with continuity correction
##
```

```
## data:  sensitivedf$t[sensitivedf$t < 75] and insensitivedf$t
## W = 342, p-value = 0.008586
## alternative hypothesis: true location shift is not equal to 0
```

```
## [1] "prpE" "treA" "ydcV" "ydgD" "astB" "yeaG" "crr"  "cysQ"
```

Using the R package TopGO, sensitive and insensitive genes were enriched for Gene Ontology (GO) terms (at Kolmorgorov-Smirnoff p value<0.05). These GO terms were added to plots of sensitive and insensitive sigmoidal/impulse trajectories faceted by RpoS sensitivity. Sensitive genes were enriched in Membrane, Response to Abiotic Stimuli, and Transporter Terms while insensitive Genes were enriched in Cytoplasmic and Metabolic Processes. The vast majority of genes in both groups appeared monotonically upregulated, however, a number of insensitive genes appeared to achieve higher peak levels of expression than sensitive genes.





Adapting the analysis performed by Wong *et al.* 2017 on data from Conway *et. al* 2014, we generated plots of relative expression levels for median sigmoidal trajectories for sensitive and insensitive genes. Shown above are the sigmoid-determined expression levels scaled by expression at the 90 minute TP. The plot illustrates that the relative transcription of sensitive genes begins earlier than in insensitive genes during cell starvation.

Discussion

In summary, E. coli posesses general stress response to a variety of environmental stresses (Battesti *et al* 2011, Hengge 2011). A key transcription factor coordinating this response is RpoS, which regulates one quarter of the bacteria's genome (preliminary data from Professor Dan Stoebel). Simple interpretations of transcriptional networks as on/off switches don't adequately describe the dynamism of transcriptional responses to environmental change. Professor Stoebel has shown that groups of genes display different levels of sensitivity to RpoS, and it was hypothesized that sensitivity may be mechanism to control expression timing in response to stress *in vivo* (Wong *et al.* 2017). Thus, study of the timing of the RpoS regulon could lead to a better understanding of the complex regulatory circuits at play in the general stress response.

The transcriptomic remodeling of WT and RpoS-knockout bacterial strains were measured during a 150 minute cell starvation time course RNA-Seq experiment. Differentially expressed genes were identified using a thoughtfully-constructed pipeline of TC differential expression tools (Spies *et al.* 2019). Sigmoidal models were fit to DEGs using ImpulseDE2. The robustness to dispersion-based noise, outliers, and number of TPs were asssessed using simulations, and the onset time (t) parameter was found to be stable to high dispersions and outliers. What's more, increasing the number of time points (particularly around the spike in transcription witnessed between 55 and 75 minutes) increased the confidence in the t parameter, which will inform the design of future timing-oriented RNA-seq studies aiming to study *E. coli* responses to cell starvation.

Density plots of onset times for insensitive and sensitive DEGs revealed that while sensitive genes peak onset time occurs earlier than the insensitive group, several sensitive genes turn on much later. A Wilcoxon Rank

Sum test identified a significant difference between the onset times of these two groups (p=0.05). While there was no biological justification for excluding outliers, when the late-to-turn-on sensitive genes were excluded, a highly signficantly different shift between the distributions of sensitive and insensitive genes was discovered using a Wilcox Rank Sum test (p=0.008). Visualization of median sigmoidal curves for each group also suggested that sensitive genes turn on earlier than insensitive genes during cell starvation, analysis consistent with work done by Wong *et al.* on another dataset (Wong *et. al* 2017). What's more, sensitive genes were found to be enriched for membrane and membrane transport GO terms, while insensitive genes were enriched in cytosolic, response to abiotic stimulus, and metabolic terms.

However, RpoS sensitivity does not describe the global trend of transcriptomic response to cell starvation in *E. coli*. Density plots of the onset times for monotonic DEGs indicated that the vast majority of differentially expressed genes turned on at around the same time, with very weak multimodality. The most sensitive genes turn on at around 62 minutes, while most insensitive genes turn on at around 65 minutes. Thus, sensitive and insensitive genes did not turn on abnormally early nor abnormally late respectively compared to all the other monotonic DEGs, and the short time separating their onsets suggests suggests that RpoS sensitivity may not describe the global transcriptional response to cell starvation.

In the future, we hope to conduct an analysis like those who first developed the impulse model for gene expression data (Chechik and Koller, 2009). Chechik and Koller compared the distribution of onset times for genes grouped within GO categories against the distribution of onset times for a baseline category (genes with similar but not identical function). Separate baseline for each category were defined using only genes from sibling categories in the GO hierarchy (other children of its parent category). For each GO category and each condition, they calculated a Wilcoxon score to quantify how significantly a GO category's gene onsets appear earlier or later than the baseline onsets. We could apply this analysis to identify ontologies that are turned on earlier or later than anticipated, helping identify other time-relevant relationships within this dataset.

Based on the results of our simulation studies, we suggest future RNA-Seq experiments interested in the timing of the RpoS regulon in response to cell starvation concentrate their sampling between 55 and 75 minutes. This will improve temporal resolution and strengthen confidence in analysis of t parameters. We also suggest the exploration and timing comparison of other stress types in *E. coli*.

In addition, future studies of the sigmoidal and impulse models are warranted. We would like to assess the stability of the onset time (and other parameters) across many different dispersions, sample sizes, number of time points, outliers, and different genes. We'd also like to explore methods of clustering based on these models, as this could mitigate the problem of clustering to noise (a potential danger when clustering RNA-seq data). We're also interested in possible sophistications we can make to these models: one possibility is including another beta parameter to account for differences in rates of mRNA synthesis and degradation.

Acknowledgements

I would like to sincerely thank Professors Johanna Hardin, Daniel Stoebel, and Danae Schulz for their thoughtful mentorship and insights throughout the research process.

I would also like to thank everyone involved in the Data Science REU at Harvey Mudd College for fostering a supportive and engaging work environment.

Lastly, I would like to deeply thank the Pomona College Department of Mathematics and the Kenneth Cooke Fellowship for affording me this research opportunity and funding my summer experience.

Literature References:

Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in Escherichia coli. Annu Rev Microbiol 65:189–213. doi:10.1146/annurev-micro-090110-102946

Chiang SM, Dong T, Edge TA, Schellhorn HE. 2011. Phenotypic diversity caused by differential RpoS activity among environmental Escherichia coli isolates. Appl Environ Microbiol 77:7915–7923. doi: 10.1128/AEM.05274-11.

Chechik, G., & Koller, D. (2009). Timing of Gene Expression Responses to Environmental Changes. Journal of Computational Biology, 16(2), 279–290. https://doi.org/10.1089/cmb.2008.13tt

Conway T, Creecy JP, Maddox SM, Grissom JE, Conkle TL, Shadid TM, Teramoto J, San Miguel P, Shimada T, Ishihama A, Mori H, Wanner BL. 2014. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. mBio 5:e01442-14. doi:10.1128/mBio.01442-14.

Farewell A, Kvint K, Nyström T. 1998. Negative regulation by RpoS: a case of sigma factor competition. Mol Microbiol 29:1039–1051. doi:10.1046/j.1365-2958.1998.00990.x.

Fischer, D. S., Theis, F. J., & Yosef, N. (2018). Impulse model-based differential expression analysis of time course sequencing data. Nucleic Acids Research, 46(20), 1–10. https://doi.org/10.1093/nar/gky675

Fong, A. J. L., Shull, L. M., Batachari, L. E., Dillon, M., Evans, C., Becker, C. J., ... Daniel, S. (2017). Genome-Wide Transcriptional Response to Varying RpoS Levels in Escherichia coli K-12. Journal of Bacteriology, 199(7), 1–17.

Hengge R. 16 December 2011. Stationary-phase gene regulation in Escherichia coli. EcoSal Plus doi: 10.1128/ecosalplus.5.6.3.

Hryckowian AJ, Battesti A, Lemke JJ, Meyer ZC, Welch RA. 2014. IraL is an RssB anti-adaptor that stabilizes RpoS during logarithmic phase growth in Escherichia coli and Shigella. mBio 5:e01043-14. doi:10.1128/mBio.01043-14.

Lange R, Hengge-Aronis R. 1994. The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. Genes Dev 8:1600–1612. doi:10.1101/gad.8.13.1600.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Nueda, M. J., Tarazona, S., & Conesa, A. (2014). Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics, 30(18), 2598–2602. https://doi.org/10.1093/bioinformatics/btu333

Pratt LA, Silhavy TJ. 1998. Crl stimulates RpoS activity during stationary phase. Mol Microbiol 29:1225–1236. doi:10.1046/j.1365-2958.1998.01007.x.

Spies, D., Renz, P. F., Beyer, T. A., & Ciaudo, C. (2019). Comparative analysis of differential gene expression tools for RNA sequencing time course data. Briefings in Bioinformatics, 20(1), 1–11. https://doi.org/10.1093/bib/bbx115

R Package Citations

Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 Genome Biology 15(12):550 (2014)

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://CRAN.R-project.org/package=tidyverse

David S Fischer (2019). ImpulseDE2: Differential expression analysis of longitudinal count data sets. R package version 1.8.0.

Ana Conesa and Maria Jose Nueda (2019). maSigPro: Significant Gene Expression Profile Differences in Time Course Gene Expression Data. R package version 1.56.0. http://bioinfo.cipf.es/

Futschik M, Carlisle B (2005). "Noise robust clustering of gene expression time-course data." *Journal of Bioinformatics and Computational Biology*, 965-988. <URL:http://mfuzz.sysbiolab.eu>.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2019). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.

Alboukadel Kassambara and Fabian Mundt (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. https://CRAN.R-project.org/package=factoextra

Hanbo Chen (2018). VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. https://CRAN.R-project.org/package=VennDiagram

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Paul Murrell and Zhijian Wen (2019). gridGraphics: Redraw Base Graphics Using 'grid' Graphics. R package version 0.4-1. https://CRAN.R-project.org/package=gridGraphics

Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL http://www.jstatsoft.org/v21/i12/.

Gregory R. Warnes, Ben Bolker, Gregor Gorjanc, Gabor Grothendieck, Ales Korosec, Thomas Lumley, Don MacQueen, Arni Magnusson, Jim Rogers and others (2017). gdata: Various R Programming Tools for Data Manipulation. R package version 2.18.0. https://CRAN.R-project.org/package=gdata

Adrian Alexa and Jorg Rahnenfuhrer (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.36.0.

Marc Carlson (2019). org.EcK12.eg.db: Genome wide annotation for E coli strain K12. R package version 3.8.2.

Matt Dowle and Arun Srinivasan (2019). data.table: Extension of `data.frame`. R package version 1.12.2. https://CRAN.R-project.org/package=data.table