Model-Based Quantification of Induced Gene Expression Time Courses

Sam Hokin & M. Kathryn Barton
shokin@carnegiescience.edu

CARNEGIE SCIENCE | Plant Biology

INTRODUCTION

We model gene expression time courses following induction of a glucocorticoid receptor-bound transcription factor into the nucleus after exposure to dexamethasone in Arabidopsis thaliana.

In contrast to single-time measurements of induced expression, which result in a single fold-change value, typically long after the induction event, our model reproduces the dynamics of TF translocation into the nucleus followed by the transcriptional dynamics of its direct and indirect targets.

The model quantifies an observed time course with two rate coefficients:

1. $\hat{\eta}_p$ = the normalized initial rate of rise (+) or fall (-) of target mRNA levels
2. γ_p = the mRNA loss rate

These two processes compete to determine the final transcript fold change: a highly-induced transcript may saturate at a low fold change if it also decays quickly, while weakly-induced transcripts may rise to a high level if they have very low loss.

In addition, some targets exhibit "turn-off", or auto-regulation. These time courses require a turn-off time, t_{off}, after which the mRNA level decays exponentially at γ_p.

We present analysis of experiments with GR-TFs involved in leaf regulation:

- GR-STM: AT1G62360 *SHOOT MERISTEMLESS/STM*
- GR-AS2: AT1G65620 *ASYMMETRIC LEAVES 2/AS2/LBD6*
- GR-KAN: AT5G16560 *KANADI 1/KAN1*
- GR-TINY: AT5G25810 *TINY*
- GR-REV: AT5G60690 *REVOLUTA/REV*

Our experiments measure expression time courses for up to two hours after DEX exposure with both RNA-seq and microarray assays. Since secondary targets of the induced transcription factor are expected to rise after some delay, during which the primary transcript's protein level rises, we explore the use of our model to distinguish between primary and secondary targets.

OBJECTIVES

1. Model the translocation of the GR-TF into the nucleus after DEX exposure
2. Model the modified transcription of GR-TF primary targets (subscript p)
3. Model the modified transcription of GR-TF secondary targets (subscript s), which are primary targets of GR-TF primary targets
4. Compare the model-fitted rate coefficients with standard fold-change analysis

MODEL AND EXPERIMENTAL METHOD

Nuclear import model

Transport of a glucocorticoid receptor (GR) into the nucleus after DEX application has been measured in animal tissue by several groups.

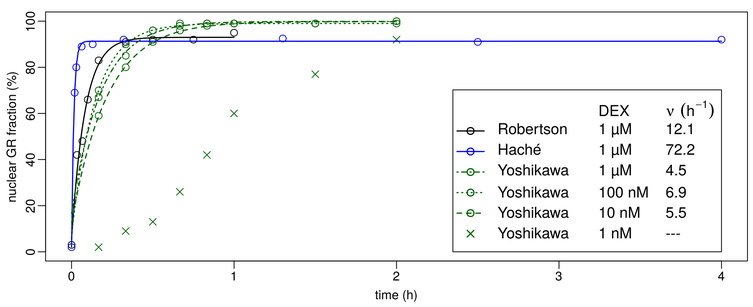

Figure 1: The measured nuclear GR concentration follows a saturated exponential time course, except at very low DEX concentration. Data sources in References.

The measured time courses are well described by a pair of linear ODEs:

$$\dot{\rho}_c = -\nu\rho_c + \gamma_e\rho_n$$
$$\dot{\rho}_n = \nu\rho_c - \gamma_e\rho_n - \gamma_n\rho_n$$

ρ_c = cytoplasmic GR protein concentration
ρ_n = nuclear GR protein concentration
ν = nuclear import rate
γ_e = nuclear export rate
γ_n = rate of loss of GR-TF from other causes

We neglect nuclear export since we measure over a time period $t \ll 1/\gamma_e$. We assume that import of chimeric GR-TF is similar to the published GR import data.

Transcription model

We model altered transcription of target mRNA due to DEX-induced import of the GR-TF with a set of coupled ODEs:

$$\dot{\rho}_p = \eta_p\rho_n - \gamma_p\rho_p$$
$$\dot{\rho}_s = \eta_s\rho_p - \gamma_s\rho_s$$

ρ_p = expressed mRNA of a primary target of the GR-TF
ρ_s = expressed mRNA of a secondary target of the GR-TF, i.e. target of a primary target
γ_p = loss rate of primary mRNA
γ_s = loss rate of secondary mRNA
η_p = transcription rate between the GR-TF and primary target
η_s = transcription rate between the primary and secondary targets

ρ_p and ρ_s are measured with microarray or RNA-seq assays. γ_p and γ_s represent loss of mRNA due to translation into proteins or other mechanisms. We assume that the concentration of primary target protein is proportional to ρ_p so that ρ_p can be used directly in the equation driving ρ_s (eliminating an extra translation rate parameter).

These equations can be solved analytically, and we assume $\gamma_n = 0$ throughout, and set $\nu = 10\text{h}^{-1}$, based on the animal tissue studies (we need to measure it!).

Primary target transcription time course metrics

The model provides two useful, independent expression metrics:

- $\hat{\eta}_p = \eta_p\rho_n(0)/\rho_p(0)$ characterizes the initial change of mRNA level
- γ_p characterizes the loss of mRNA

In addition, we can calculate an asymptotic ($t \to \infty$) fold change:

$$\text{logFC}_\infty \equiv \lim_{t\to\infty} \log_2 \frac{\rho_p(t)}{\rho_p(0)} = \log_2\left(1 + \frac{\hat{\eta}_p}{\gamma_p}\frac{\rho_c(0)}{\rho_n(0)}\right)$$

The ratio $\hat{\eta}_p/\gamma_p$ therefore determines the asymptotic fold change of a transcript: **mRNA transcription and loss play equal roles in the final transcript level in this model**.

Time-course expression measurements in *Arabidopsis thaliana*

The five GR-TF plant lines plus WT plants (all Col-0) were grown to seedlings, and tissue was exposed to 50 μM of DEX and then flash frozen 0, 0.5, 1 and 2 hours later. Almost all samples had at least three biological replicates per time. RNA was extracted and assayed with the Affymetrix ATH1 microarray and Illumina HiSeq RNA-seq. The microarray data were analyzed using the R packages `affy` and `limma`; the RNA-seq data were analyzed using TopHat2, Cufflinks and Cuffdiff. Sequences were mapped to the TAIR10 reference genome.

RESULTS

Expression time courses exhibit qualitative variation

It matters greatly <u>when</u> you measure differential expression in an induced expression experiment!

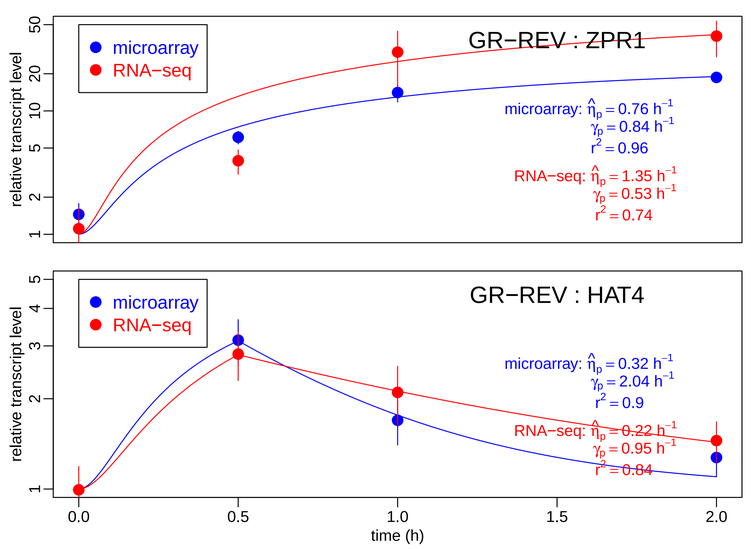

Figure 2: Expression time courses vary greatly, from strong GR-REV monotonic risers like *ZPR1* to HD-ZIPII self-regulators like *HAT4*. Lines represent primary target model fits with transcription turned off at $t = 0.5$ h for *HAT4*.

Model parameters $\hat{\eta}_p$ and γ_p distinguish expression time course shapes

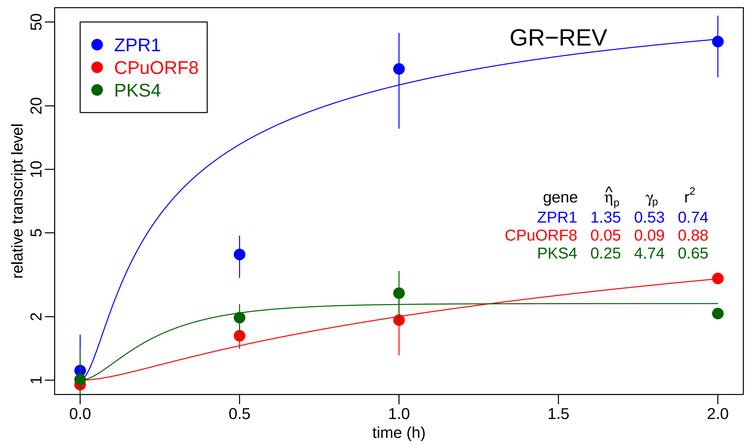

Figure 3: Three distinct time course shapes result from: high $\hat{\eta}_p$ and low γ_p for *ZPR1*, the most strongly regulated GR-REV target; low $\hat{\eta}_p$ and low γ_p for *CPuORF8*, which reaches a significant fold-change due to low loss; and high $\hat{\eta}_p$ and high γ_p for *PKS4*, which saturates quickly due to high loss.

Three examples that exhibit very different transcriptional responses to DEX application:

- *ZPR1* is the most strongly up-regulated target of GR-REV — a strong, steady riser with a strong initial transition (κ is large) leading to a large asymptotic DE level (logFC$_\infty$ is large).
- *CRK30* is the most strongly down-regulated target of GR-KAN in the RNA-seq series — it is predicted to reach a larger DE level further out in time.
- *HAT22* is a strong autoregulating GR-REV target — an early rise with a large κ value is followed by "turn-off" after $t = 0.5$ h; without turn-off, the model predicts it would reach logFC$_\infty = 1.8$.

	logFC			model	
gene (assay)	0.5h	1.0h	2.0h	κ	logFC$_\infty$
ZPR1 (microarray)	2.09	3.29	3.71	122 h^{-2}	4.06
CRK30 (RNA-seq)	-0.72	-1.93	-2.51	-15 h^{-2}	-3.01
HAT22 (microarray)	1.53	×	×	96 h^{-2}	1.77

Table 1: `limma` and Cuffdiff DE levels along with model-derived metrics κ and logFC$_\infty$.

Intermediate DE values, given by logFC, depend on the time of measurement while κ and logFC$_\infty$ characterize the entire time course.

Early-peaking primary transcripts are well fit by mid-course transcription turn-off in the model

When $\gamma_n = 0$, only group L time courses can be fit by the model with constant η_p. Group E and M genes appear to exhibit transcription "turn-off" — after a certain amount of time (or beyond a certain level of expression) transcription of those genes' mRNA "turns off" and the subsequent transcript level drops as $e^{-\gamma_p t}$. We model transcription turn-off by setting $\eta_p = 0$ at $t = 0.5$h for E time courses and $t = 1.0$h for M time courses.

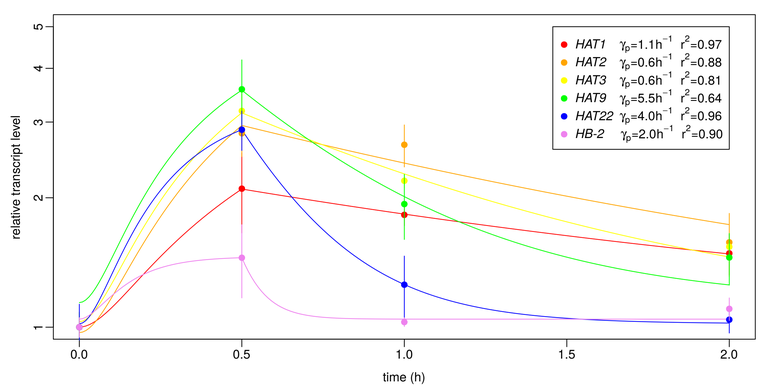

Figure 4: Many early turn-off genes are found in the HD-ZIP family, especially in the HD-ZIPII class, shown here.

The rapidity of turn-off of *HAT9* and *HAT22* may indicate that they play a leading role in the HD-ZIPII autoregulatory network [Ohgishi. et al.].

Conventional DE and model fit confidence measures are correlated

For microarray/`limma` and RNA-seq/Cuffdiff DE measurements, p-values and FDR-adjusted q-values are the standard measures of confidence. One typically assumes that $q < 0.05$ represents a 95% probability that the measured DE is real and not a chance occurrence.

In our model fits, the experimental data points, one per sample (3 reps $\times$ 4 times = 12 total), are individually fit, without weighting, by minimizing the variance between data and model values using the core R nonlinear minimization routine `nlm`. We then compute the standard r^2 coefficient of determination to quantify the quality of the fit.

- DE measurements are taken at a single time relative to the $t = 0$ base: three independent logFC, p and q values result for $t = 0.5$, 1 and 2 h, using six data per measurement, reusing the $t = 0$ base data each time.
- The model fits one solution to all twelve data, and is therefore subject to variation of all, not half, of the data
- The model counts the $t = 0$ data equally with the other data.

The model fit r^2 value, therefore, provides an independent measure of DE significance which spans the entire time course in a statistically balanced fashion.

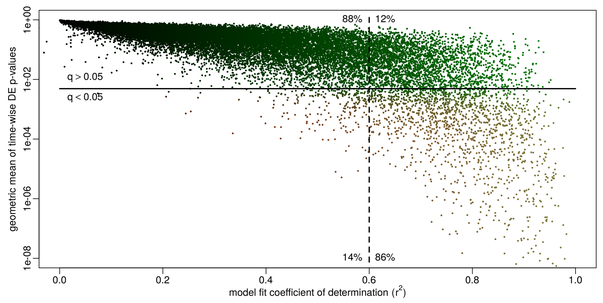

Figure 5: Model r^2 values correlate with the geometric mean of the three p-values from `limma` DE analysis of the 22,585 GR-REV microarray transcripts which result in a model fit of any quality. The horizontal line shows the geometric mean of p-values for which $q = 0.05$, representing the boundary of significant `limma`-determined DE at the $q < 0.05$ level.

The microarray `limma` p and model-fit r^2 values correlate, and we have chosen $r^2 > 0.6$ as the significance level for the model fits, covering 86% of the $q < 0.05$ transcripts. A majority of significant DE time courses meet this model fit level in both microarray and RNA-seq series, boosting confidence in the validity of the model.

Secondary targets are suggested by model fits and transcriptional response across four GR-TF lines

GR-REV:*AHP4* rises very slowly, requiring $\gamma_p < 0$ in the model, which is unphysical. We model the time course of *AHP4* as a *secondary* target with $\gamma_s > 0$, and search for candidate primary targets in the four GR-TF lines which regulate it. If *AHP4* is strictly a secondary target in the GR-TF lines in which it exhibits differential expression, then the corresponding primary target should also exhibit DE in those and only those lines. *AHP4* is an up-regulated group L gene in GR-KAN, GR-REV and GR-STM, further suggesting that it is a secondary, or farther downstream target.

From the RNA-seq series: 44 other genes show all up- or all down-DE in GR-KAN, GR-REV and GR-STM and none in GR-AS2. Of these, model fits using *AHP4* as the secondary target lead to:

- 10 genes as primary targets of GR-REV with $r^2 > 0.6$ and $\gamma_p, \gamma_s > 0$
- 9 genes as primary targets of GR-STM with $r^2 > 0.6$ and $\gamma_p, \gamma_s > 0$
- 5 genes common to the above GR-REV and GR-STM selections

The RNA-seq GR-KAN:*AHP4* data variance is too high to produce good model fits.

From the microarray series, model fits to the five genes above as primaries which target *AHP4* result in three candidates with $r^2 > 0.6$ and $\gamma_p, \gamma_s > 0$: *CRK22*, *JKD* and *GT2L* (At5g28300).

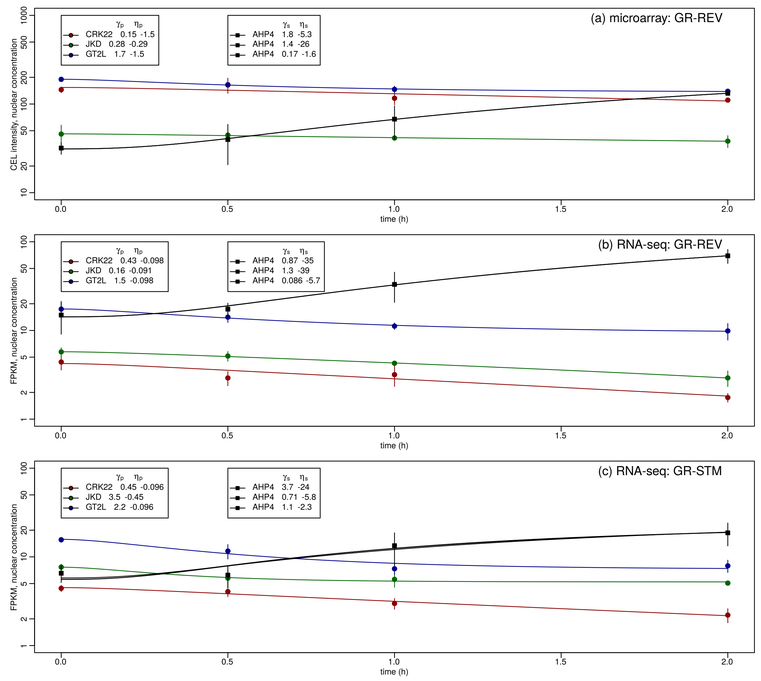

Figure 6: Three genes are well simulated as down-regulated targets of *KAN*, *REV* and *STM*, which in turn down-regulate *AHP4*.

We regard these three genes as candidate primary targets of *KAN*, *REV* and *STM*, which, in turn, target *AHP4*. (One of these three, *GT2L*, is a member of the Trihelix TF family, but *AHP4* has not been reported as a target.) Also note that the three candidate primary targets are *down*-regulated in GR-KAN, GR-REV and GR-STM lines. This suggests that they *repress AHP4*, which then overexpresses its mRNA as their expression is reduced by the induced TF.

CONCLUSIONS

1. DEX-induced GR-TF transcription time courses can be modeled with a set of linear ODEs with parameters (γ_p,η_p) describing the primary target response and (γ_s,η_s) describing the secondary target response.
2. Model fits use data in a balanced fashion, rather than the overemphasis of $t = 0$ data in time-wise DE measurements.
3. Model-determined metrics κ and logFC$_\infty$ characterize the primary target response in a time-independent manner, as opposed to misleading single-time DE measurements.
4. The model simulates transcription turn-off by setting η_p or η_s to 0 at an appropriate time, matching autoregulated genes in the HD-ZIPII family.
5. The model simulates primary and secondary transcriptional responses, supporting target discovery and inference of gene regulatory networks.

Forthcoming Research

1. Improve model to make robust fits for “black box” analysis, perhaps in a Web application, with data and fit diagnostics
2. Measure tighter time course of selected GR-TFs and targets using qPCR
3. Measure GR-TF nuclear import and mRNA accumulation using flourescent probes
4. Measure mRNA induction in real time with CRISPRi activation?

References

R. J. Haché, R. Tse, T. Reich, J. G. Savory, and Y. A. Lefebvre. Nucleocytoplasmic trafficking of steroid-free glucocorticoid receptor. *J. Biol. Chem.*, 274(3):1432–1439, Jan 1999.

M. Ohgishi, A. Oka, G. Morelli, I. Ruberti, and T. Aoyama. Negative autoregulation of the Arabidopsis homeobox gene ATHB-2. *Plant J.*, 25(4):389–398, Feb 2001.

B. J. Reinhart, T. Liu, N. R. Newell, E. Magnani, T. Huang, R. Kerstetter, S. Michaels, and M. K. Barton. Establishing a framework for the Ad/abaxial regulatory network of Arabidopsis: ascertaining targets of class III homeodomain leucine zipper and KANADI regulation. *Plant Cell*, 25(9):3228–3249, Sep 2013.

S. Robertson, J. P. Hapgood, and A. Louw. Glucocorticoid receptor concentration and the ability to dimerize influence nuclear translocation and distribution. *Steroids*, 78(2):182–194, Feb 2013.

N. Yoshikawa, Y. Makino, K. Okamoto, C. Morimoto, I. Makino, and H. Tanaka. Distinct interaction of cortivazol with the ligand binding domain confers glucocorticoid receptor specificity: cortivazol is a specific ligand for the glucocorticoid receptor. *J. Biol. Chem.*, 277(7):5529–5540, Feb 2002.

Acknowledgements

This research was funded by National Science Foundation grant #0929413.