# Auto-regulation with no other input

What other questions should be answered to help us further analyze the data?

* Production rate vs. degradation rate. How do these combine?
* ANOVA p-value for within strain
  1. Magnitude (large dynamics)?
  2. Variance (spread of the data points)?
  3. Some combination of the two?
* Fit of the model vs. parameter value stability

## MBP1

This gene appears to only be a repressor

Represses itself, MAL33 and SWI4

### Production rate vs. degradation rate

Production rates all seem to be within the same range 0.08 – 0.13 with one outside of that range at 0.17

### ANOVA Within strain p-value

MPB1 p-value:

* B&H: 0.5240
* Unadjusted: 0.3760

More Analysis

* Magnitude: it seems to be up-regulated
* Variance: data point seem to show quite a bit of variance within the time points
  1. However in analyzing the graphical output, between the time points, there is little variance
  2. Higher data points could have caused the dynamics to increase if the average was used to model the dynamics

### Fit of the model vs. parameter value stability

Wild-type

* Fits the data fairly well; there seems to be increasing gaps between the first and second data points for t30 and t60
* Up-regulation is seen

dCIN5

* The model falls between the first and second data point for t15 and t30 and hits the average for t60
* The data points are widespread and the dynamics are slightly increasing

dGLN3

* The data points for t15 and t60 are wide and vary while t30 is tight and compacted
* The model falls near the average of t15 and comes close to hitting the data for t30
  + T60 is poorly fit and appears that the model should have dipped down for dGLN3

dHMO1

* The data points are widespread with no overlapping
* The model did its best to fit the data and its dynamics are slightly up-regulated

dZAP1

* The data points are closer together for t30, but t15 andt60 still show greater variances
* The model has larger dynamics than the previous runs with a steeper increase with its slope
  + The overall effect of taking the production – degradation has a value of 0.139, higher than all the other overall effects of 0.08 or below

wt + dCIN5

* There is a steady, slight increase for the dynamics
* The data point are spread out and the model took the average of all the data points

wt + dGLN3

* The data points for dGLN3 are more compacted than those of the wt; however, that does little to change the fit of the model for the two strains when visualized together

wt + dHMO1

* There is overlap in the top data points for wt and dHMO1 which could contribute to the average of the data to create the curve of the model

wt + dZAP1

* The dynamics of the curve are again larger than any of the other individual or comparison ones (with excpetion to dZAP1 individual)
* It fits the average of the data points
  + Its overall effect is clser to 0.1 than any other strain/run with a value of 0.0949

wt + dCIN5 +dZAP1

* The average of all data points are fit well with slight dynamics
  + Overlap in the data for t60 as well as compactness of data when all run together

wt + dCIN5 + dZAP1 + dGLN3

* The dGLN3 data should have different dynamics than the rest of the strains
* Model again averages all the data points instead of trying to graph those with different dynamics separately (looking at you dGLN3)

all strain

* Up-regulation seen in the dynamics of this gene

Overall parameter stability & fit of the model:

* Production rates seem to be stable/consistent across the strains
* The threshold values tend to have some stability (negative numbers) except for two strains (dCIN5 & dHMO1)

Inputs: the model follows trends of the data

* It is not affected by any deletion strain majorly
* Could be a connection between ZAP1 🡪 it would be a repressor

## SKN7

### Production rate vs. degradation rate

* Initial increase is related to the production/overall effect on the gene
  + It has to create product before its degradation rate can kick in
* Steeper the initial climb, the larger its overall effect
  + Its degradation rate also takes an effect earlier in the time points, which stabilizes the gene expression

### ANOVA Within strain p-value

SKN7 p-value:

* B& H p-value: 0.0228
* Unadjusted: 0.00455
* dHAP4
  + B&H p-value: 0.0018
  + Unadjusted: 5.07E-5
* Magnitude
  1. Large changes in data points suggest that its significance comes from changes in its dynamics
* Variance
  1. Wide variance, so the significant p-value must be due to the significant change in expression/dynamics of gene expression
* Some combination of the two?

### Fit of the model vs. parameter value stability

Wild-type

* The model fits the data points provided from the lab

dCIN5

* The model does not seem to fit the data well
  + Hits below the average for t15 and above for t60
  + It seems to predict higher values for the latter two time points, which could be due to the overall effect of the production – the degradation rate
    - 9.3…?

dGLN3

* The model fits the data points poorly
  + The model sits below t15 and t30 and above for t60
* The data covers a large amount of ground among the three time points
* Seems that deletion of GLN3 has an effect on the dynamics of SKN7 such that there needs to be some type of direct or indirect control from GLN3

dHMO1

* Model fits data well in that it hits the average of the data points for all three times

dZAP1

* Model fits the data well in that again, it hits the average of the data points
  + The points seem to be closer to each other so that the model can better visualize the dynamics

wt + dCIN5

* Fits the average of the data points well

wt + dGLN3

* Average of the data points are used to model the dynamics
* There should be a split between the two where one goes up while the other diverges and moves down

wt + dHMO1

* Again the average of the data points seem to be taken
  + The data follows the same upward trend, so it appears ok

wt + dZAP1

* Data points at t15 are wider, but the model finds their average
* The data points become tighter at the next two time points – 30 and 60 – such that the curves would converge into one

wt + dCIN5 +dZAP1

* Model fits the data points well
  + There is a lot of overall/close points for all three runs such that one line is appropriate

wt + dCIN5 + dZAP1 + dGLN3

* The model should have two different curves, one for the first three strains and a separate one for dGLN3
  + In going through all the individual strains, it appears that the dGLN3 model which had a poor fit was just lied on top of the data points from the first three strains
  + Model needs to be adjusted for dGLN3

all strains

* Fits the model poorly due to interference from dGLN3

Overall parameter stability and fit of model:

* Exceptional fit of the model in all strains; however, the parameter values are everywhere for SKN7
  + Production and threshold values vary widely and do not seem to have an “average value”
* P-value vs. fit of model: wt p-value is significant; the model fits the data a lot better for this individual strain vs. other strains where the model followed its own trajectory

Inputs: follows the average values of the data at the time points

* A connection between GLN3 and SKN7 should exist

## ZAP1

### Production rate vs. degradation rate

Production rates seem to vary widely

* dGLN3 appears to have a lower production rate than dZAP1, the TF that was deleted
* All values lie below 0.1

There is an overall positive effect with two negative effects 🡪 dGLN3 and dZAP1

### ANOVA Within strain p-value

ZAP1 p-value:

* B&H: 0.0086
* Unadjusted: 0.0012

### Fit of the model vs. parameter value stability

Wild-type

* The overall effect was 0.09 which appeared to have a great impact on the overall dynamics of ZAP1
* The data points suggest that there is a great difference in the magnitude for the wt
* The model fits the data well finding the average value for each time point

dCIN5

* The dynamics for dCIN5 are slight with a moderately fitting model for the data point
  + It misses t15 entirely
* The t30 data is spread, but the model achieves in hitting the average point

dGLN3

* The model fits the data adequately, missing t60, but finding the average of t30 (which has some variance w/ the data points)
* It is down regulated instead of up-regulated, seen in wt and dCIN5
  + This could be due to a negative value for its overall effect at -0.004

dHMO1

* The model fits the data pretty well and seems to have small, negligible down-regulation
* The data points are fairly close to each other
  + These small dynamics could be due to a small overall effect of 0.0015

dZAP1

* This gene was deleted
* Its dynamics fit with the data

wt + dCIN5

* The dynamics fit that over the wt/dCIN5 which is more positive than the dynamics for dCIN5
* There should be two lines however because the curves should diverge at t60

wt + dGLN3

* The dynamics are slightly increasing
  + However the trend is that there should be negative regulation for dGLN3/slight down-regulation for the gene while wt is sharply up-regulated

wt + dHMO1

* The model fits the average of the data points for all the genes
  + There should have been divergence between the two strains because dHMO1 saw down-regulation

wt + dZAP1

* The model fits the wt data while the flat line occurs for the dZAP1 strain

wt + dCIN5 +dZAP1

* The average was taken for wt and dCIN5 to model the up-regulation for those strains while dZAP1 saw no change

wt + dCIN5 + dZAP1 + dGLN3

* The slight increase for all non-deleted ZAP1 runs were averages of all the data points
  + There should have been some down-regulation for dGLN3
* The fit for this graph is poor

all strain

* The model for the deletion strain was correct
* The other strain dynamics were poorly fit

Overall parameter stability & fit of the model:

* The parameter values: production rates seemed to consistent while the threshold values varied widely
* The overall effects seem to be consistent
  + However when more data is added, the average is taken from their individual values. This results in small dynamics

Inputs: does not fit the data well for many of the graphs

* Seems like all the deleted genes activate ZAP on different levels
* GLN3 would be the strongest activator » HMO1 » CIN5
  + Appears that it might be one of the most connected genes in the network

# General Conclusions

The p-value does not seem to predict the goodness of the fit for the model

* For SKN7 and ZAP1, the model takes the average of the data points and ignores what would be varying trends for the different strains
* The other two genes that did not have significance in their p-values had graphical outputs with decent fits of the model

The production rate and threshold values varied with each other.

* In comparing the behavior of these estimated parameter values, the production rates seemed more stable/consistent in that they fell between a range with 0.05 - .5 variance depending on the estimation
* The b values varied more and had more sign changes for the genes
  + Literally have no idea what could have caused these differences

The production and degradation rates have a relationship with the dynamics seen from the log fold changes/gene expression

* Relative to each individual genes expression, the overall effect for the genes seems to be the rate of change/slope of the slope for the genes
  + The overall effect seems to have its visible effects on the first portion of the graphical output
  + The larger the effect, the steeper the initial dynamic seen for the gene
  + The smaller the effect, the steadier and more gradual the dynamics change for the specific gene
    - It seems to stretch out the curve vs. shrink the curve (larger values)
* Positive numbers have up-regulation while negative values show down-regulation
  + dHMO1 for ZAP1 saw down regulation. Why?
* The closer the number is to 0, the more likely its dynamics are to be slight/small
  + For instance, 0.0015 seen for ZAP1 appeared to have negligible dynamics/dynamics close to 0
  + Smaller numbers closer to 0 may have down regulation. This could be due to an exponential of a negative number, such that the smaller the value, the more likely its dynamics are negative
    - 2^-2 = ¼ so something of that effect

# Auto-regulation with 1-2 additional inputs

## GLN3

### ANOVA Within strain p-value

GLN3 p-value:

* B&H: 0.4125
* Unadjusted: 0.2638

More Analysis

* Magnitude: does not seem to have large dynamics; relatively constant
* Variance: little significance in looking at the differences among the data points

### Production rate vs. degradation rate

Production rate seems to be around 0.35 for every run of MATLAB

* Exceptions: dGLN3, wt + dCIN5 (0.429) and all strain (0.514)
* Visualized Effect
  + Initially thought that 0.13 and above for overall effect was related to the initial dip, but dCIN5 has 0.16 and an initial increase for GLN3
  + There must be another regulator or signal that affects the production/regulation of GLN3 that is not shown in the network
* The overall effects seem to be around the same value as well (0.13)

### Fit of the model vs. parameter value stability

Wild-type

* Model fits the data fairly well with tight/identical values at t60
* Weird initial dip in dynamics before t15
  + In looking at the optimized/visualized network of wt only, see that GLN3 represses itself in addition to having activation from another TF (MAL33)

dCIN5

* Fits the data well w/ tight data points at t30
* It represents the trend of the data & has an initial increase rather than decrease, which was seen in wt only

dGLN3

* No dynamics because it is the deleted gene

dHMO1

* Again, an initial dip shows repression potentially
  + In the visualized network, a grey activation arrow is seen for its self-regulation
* Little change in the dynamics 🡪 no significance in the p-value

dZAP1

* Fits the data well with tight data point (little variance)
* Shows initial up-regulation before slow, minimal degradation
  + Could be due to it up-regulating itself

wt + dCIN5

* Able to see two different dynamics because of a divergence of the model to fit the individual runs to the correct data points
  + dCIN5 appears to have slightly more dynamics than wt, whose curve seems to be approaching at steady state faster than the dCIN5 strain
* do not see the initial dip of wt when it was alone

wt + dGLN3

* Able to see the odd initial decrease when wt is plotted against dGLN3
* The model fits the data fairly well for wt

wt + dHMO1

* The model fits the data points well because there seems to be a lot of overlap, especially at t60
* For both strains, an initial decrease was originally seen when plotted individually
  + Same decrease is seen here when graphed together

wt + dZAP1

* Model fits average of the data
  + Wt’s initial decrease not seen when graphed with dZAP1
* Overlapping of data points at t60

wt + dCIN5 +dZAP1

* Able to see the diverging pattern by t60; however, the initial dip seen in wt is not seen in this figure

wt + dCIN5 + dZAP1 + dGLN3

* See the varying graphs, now there appears to be three
  + dGLN3 straight lines shows that there are dynamics for this particular gene 🡪 up-regulation
* The model fits the data points for the non-zero values/when GLN3 is present in the cell

all strain

* Contains two/three different models for the data
  + Know that code/MATLAB can model things correctly because it shows two different lines when comparing all the strains to the deletion of GLN3
* All the points lie closely together, especially when looking at dZAP1 and dCIN5 runs

Overall parameter stability & fit of the model:

* Production rates seem to be stable at/around 0.35, but the threshold values are what varies widely
* The model seems to fit the data well due to little variance and little change within time points, which suggests that the dynamics are minimal

## AFT2

In most runs, it activates itself while receiving repression cues from SKN7

### Production rate vs. degradation rate

The production rates range from 0.04 to 0.5 while its degradation rate is generally much smaller than the production rate

* The overall effect should see some steep initial climbs with anything above 3 and a more stretched out graph for those between 1 & 3

### ANOVA Within strain p-value

The wt p-value:

* B&H: 0.716
* Unadjusted: 0.602

More Analysis:

* It appears that the data points were not significant because there was wide variance between the data points
* Dynamics must be what gives this gene its significance in the network…?
  + Dynamics: little dynamics for wt and dCIN5 (+) while there is greater positive change for dGLN3 and dZAP1

### Fit of the model vs. parameter value stability

Individual Strains

* It appears that wt has the best fit as well as the best collection for the time points
* There is too much variance with dHMO1 to get a good model fit while the other outputs are standard/decent fits to the data points

Comparison run with wt + dStrain

* Dynamics are slightly positive for all runs
* Initial increase for dCIN5, dGLN3, and dZAP1
* No dynamics for dHMO1 + wt run together

wt + dCIN5 +dZAP1

* Depicts slight up-regulation with the one line 🡪 average data points and slopes of curves to get the singular line

wt + dCIN5 + dZAP1 + dGLN3

* Again like above, however, the initial increase is greater for this run

all strain

* Fits data for other strains better than dHMO1

Overall parameter stability & fit of the model:

* Parameter stability is pretty stable in that it consists of values less than 0 and within a range of 0.01 – 0.1; however, there are a few positive values (dCIN5 and wt + dCIN5)
* The production rates again fall within a specific range, but vary out to the 0.001 decimal place

Inputs: seems to be missing connection with dHMO1

* HMO1 could be an activator because in all other runs, AFT2 is being up-regulated compared to the steady 0 in dHMO1

## CIN5

It appears to strongly activate itself in all runs, except obviously dCIN5 (negligible activation probably from noise), but in dHMO1, it appears to repress itself…?

### Production rate vs. degradation rate

The production rate seems to be consistently around 0.15, ranging from 0.10 to 0.17

* There is one production rate outlier at 0.225
* The threshold values vary greatly, jumping from positive to negative randomly
* The overall effect however remains positive even with the negative threshold values. The effect for the various runs seem consistent around a value of 2.5 with a few above and below by an integer of 1

### ANOVA Within strain p-value

CIN5 wt p-value

* B&H: 0.0642
* Unadjusted: 0.0191

More Analysis

* There appears to be variance within the time points, but in a way that there is significance when the criteria for the p-value isn’t as strict
  + Within the time points, the data points tend to be grouped together with points near, for example, 0.5 and others near 1.2
  + This could also be cause for some odd dynamics such that when taking the average of these small and then large values, the model could remain constant
* Dynamics: some, little, great (+/-)

### Fit of the model vs. parameter value stability

Individual Runs

* The dynamics of the curve are appropriate with the data points
* The wt is correct, suggesting that the p-value correlates with better fits of the model to the data points
* The graphical output is correct from the data seen in the other deletion strains; I will say that the initial curve seems to be steeper for the other deletion strains than the wt

Comparison run with wt + dStrain

* Appropriate graph seen for wt + dCIN5
* From what I learned today 🡪 the fact that it graphs what is connected and not what the data points are showing, the other models/curves fit the data
  + CIN5 is not associated with any of the other deletion strains/there is not connection; therefore, the same graph as wt should appear

wt + dCIN5 +dZAP1

* See what would be expected
* The curve for wt/dZAP1 are appropriate for the noted data points and fit them well

wt + dCIN5 + dZAP1 + dGLN3

* See above
* The dynamics do what was seen for the wt and the other individual strains that were not affected by the deletion of dCIN5

all strain

* See what would be expected, but the data points for dHMO1 are not well fit by the curve…

Overall parameter stability & fit of the model:

Inputs:

## FKH2

Appears to repress itself while receiving activation cues from FHL1

### Production rate vs. degradation rate

### ANOVA Within strain p-value

FKH2 p-value for wt:

* B&H: 0.1427
* Unadjusted: 0.0485

More Analysis

* Variance: could be due to clusters of data near each other in wt which gives this gene its significance
* Magnitude: it appears that its dynamics will shift a little, suggesting that significance comes from data points
* Dynamics: some, little, great (+/-)

### Fit of the model vs. parameter value stability

Individual Runs

* The model fits the data well for the wt data 🡪 the unadjusted p-value was significant; however, its B&H was not
* Its own production rate must be less than its degradation rate to achieve the negative slope/dynamics
* The model fits the other deletion strains, having the same curve as wt; however, dHMO1 and dZAP1 have different appearances
  + Its suggests that dHMO1 is a repressor and that there should be some type of connection between the two genes
  + Same with dZAP1 🡪 its deletion appears to have some type of suppressing effect on FKH2

Comparisons of wt +dStrain

* The dynamics match what would be expected from the connections in the GRN
* The models fit the data for each run well

wt + dCIN5 +dZAP1

* The dynamics match that of the data points seen
* However, when compared to dZAP1 individually, one would have expected to see a non-dynamic/unchanging line

wt + dCIN5 + dZAP1 + dGLN3

* See above description

all strain

* Again, you see what would be expected from the constructed GRN and the connections/inputs for FKH2 🡪 only other regulator is FHL1 which doesn’t have an input
  + Look at Grace’s analysis to see which dStrain affected its output to see if it is the same as the graphical outputs for this gene

Overall parameter stability & fit of the model:

Inputs:

# Next Steps

No Input Gene

In Degree:

Out Degree:

Self-Regulating with no other inputs

In:

Out:

Self-Regulating with 1-2 inputs

In:

Out:

Max Input

Max Output

Min Input

Min Output