# Supporting Appendix 2:
# Biological Description of the Origons
# in the *E. coli* Transcription-Regulatory Network

# Table of Contents

# Origons containing FFLs

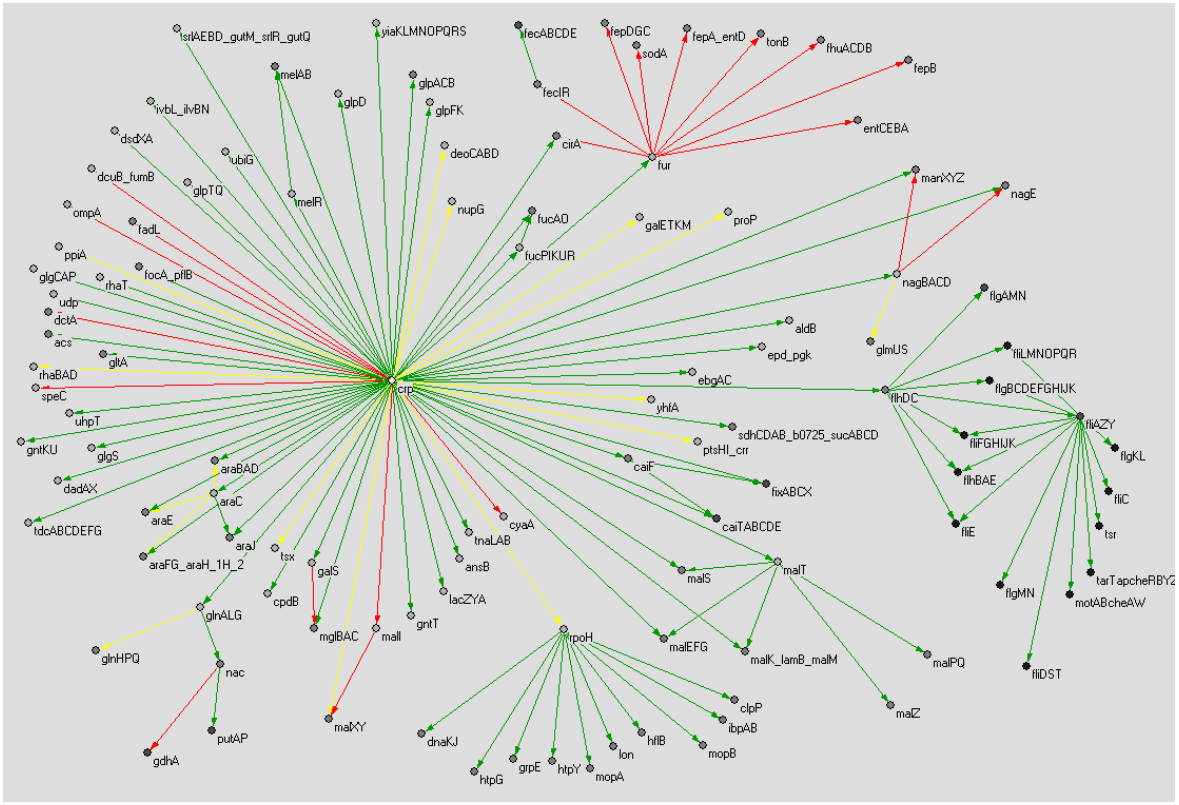


**Fig. 12. The crp origon.**
Activating (green), repressing (red) and dual function (yellow) links are shown

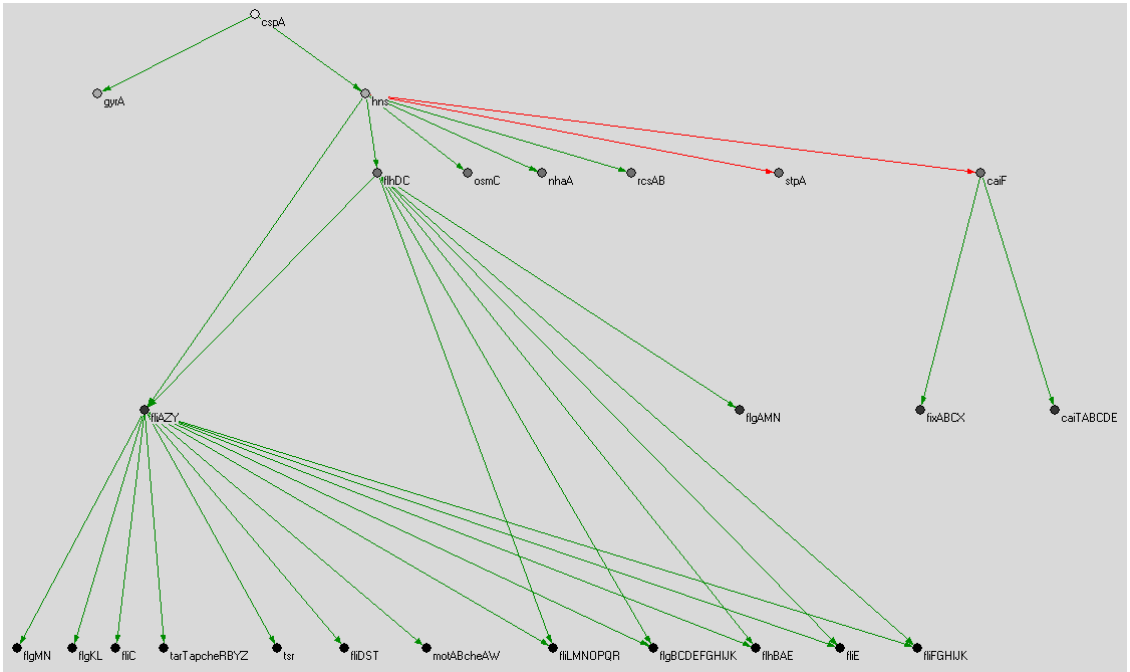


**Fig. 13. The cspA origon**
Activating (green), and repressing (red) links are shown

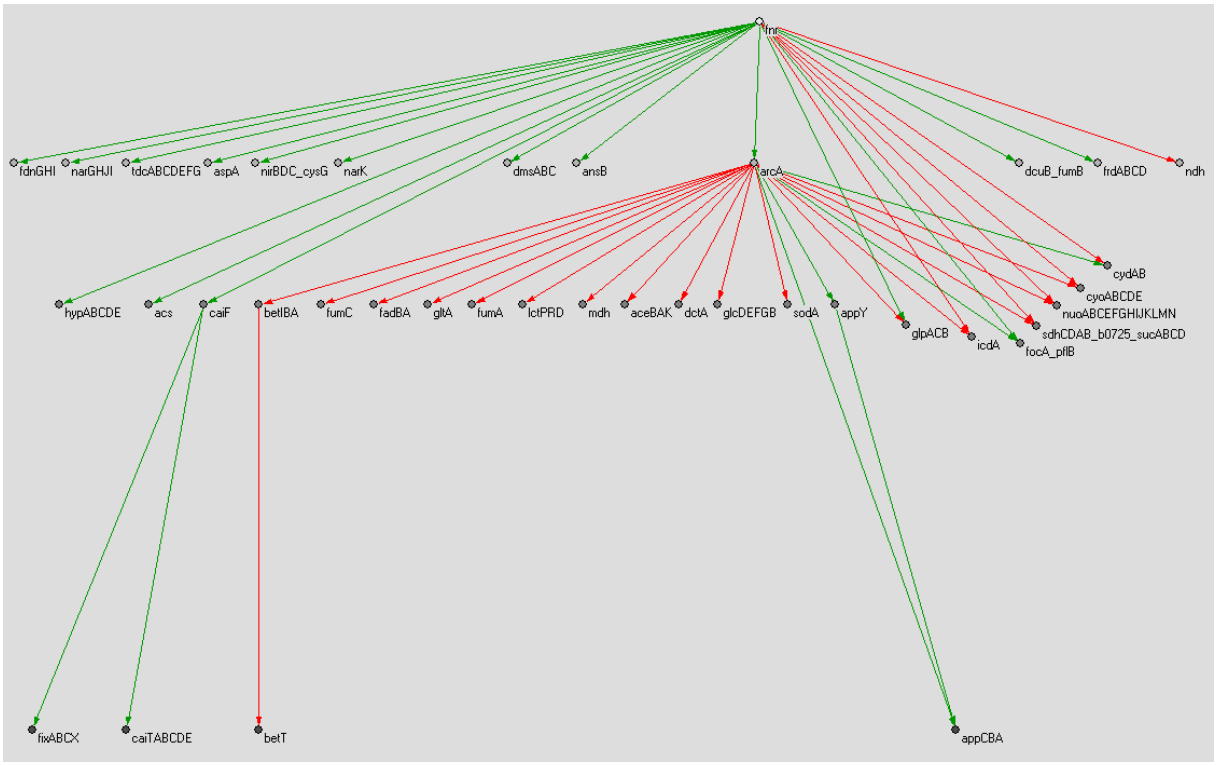


**Fig. 14. The fnr origon.**
Activating (green), and repressing (red) links are shown

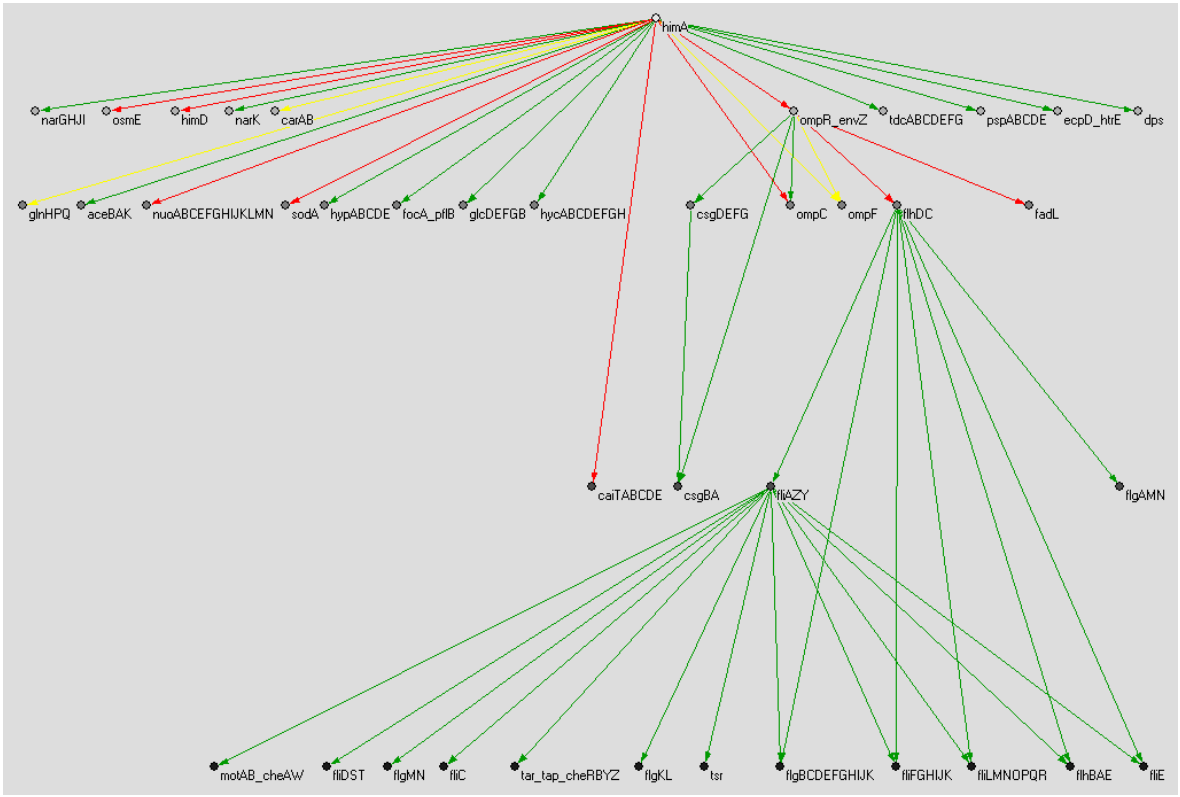


**Fig. 15. The himA origon.**
Activating (green), repressing (red) and dual function (yellow) links are shown

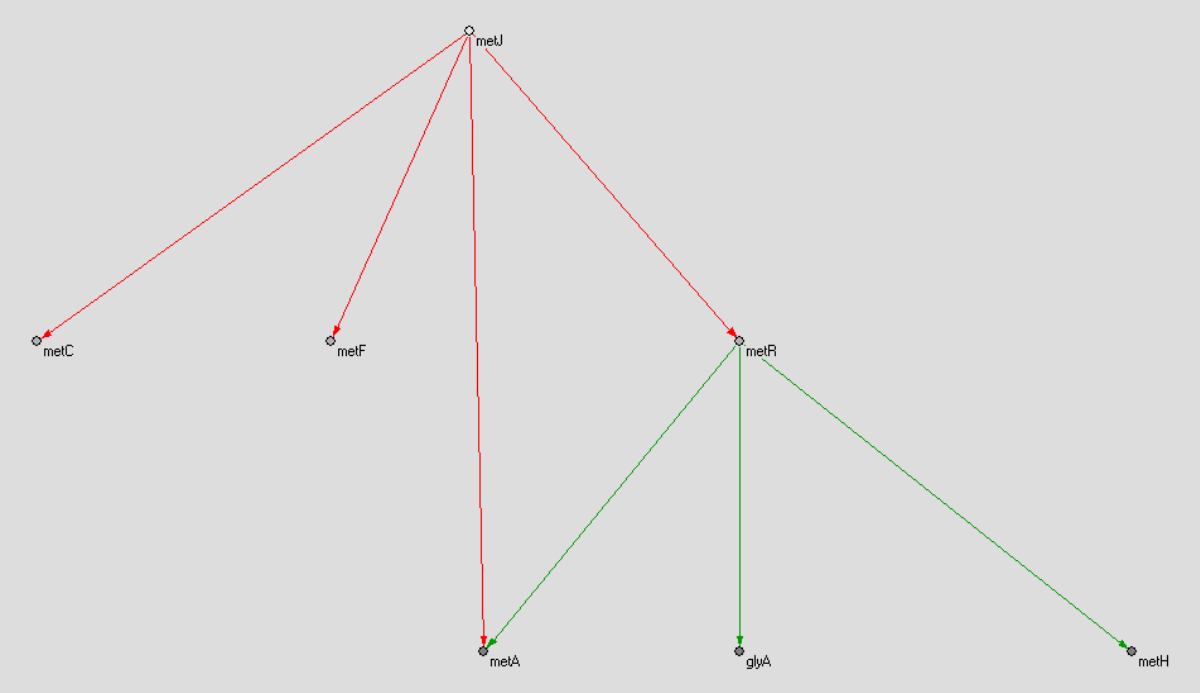


**Fig. 16. The metJ origon.**
Activating (green), an repressing (red) links are shown

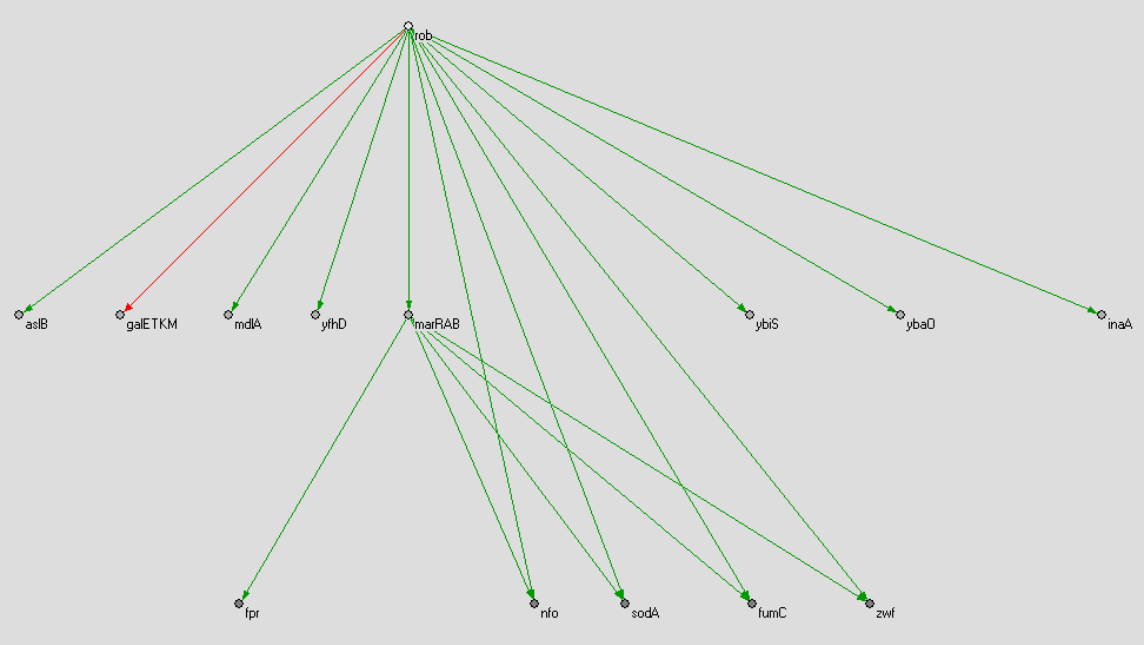


**Fig. 17. The rob origon.**
Activating (green), an repressing (red) links are shown

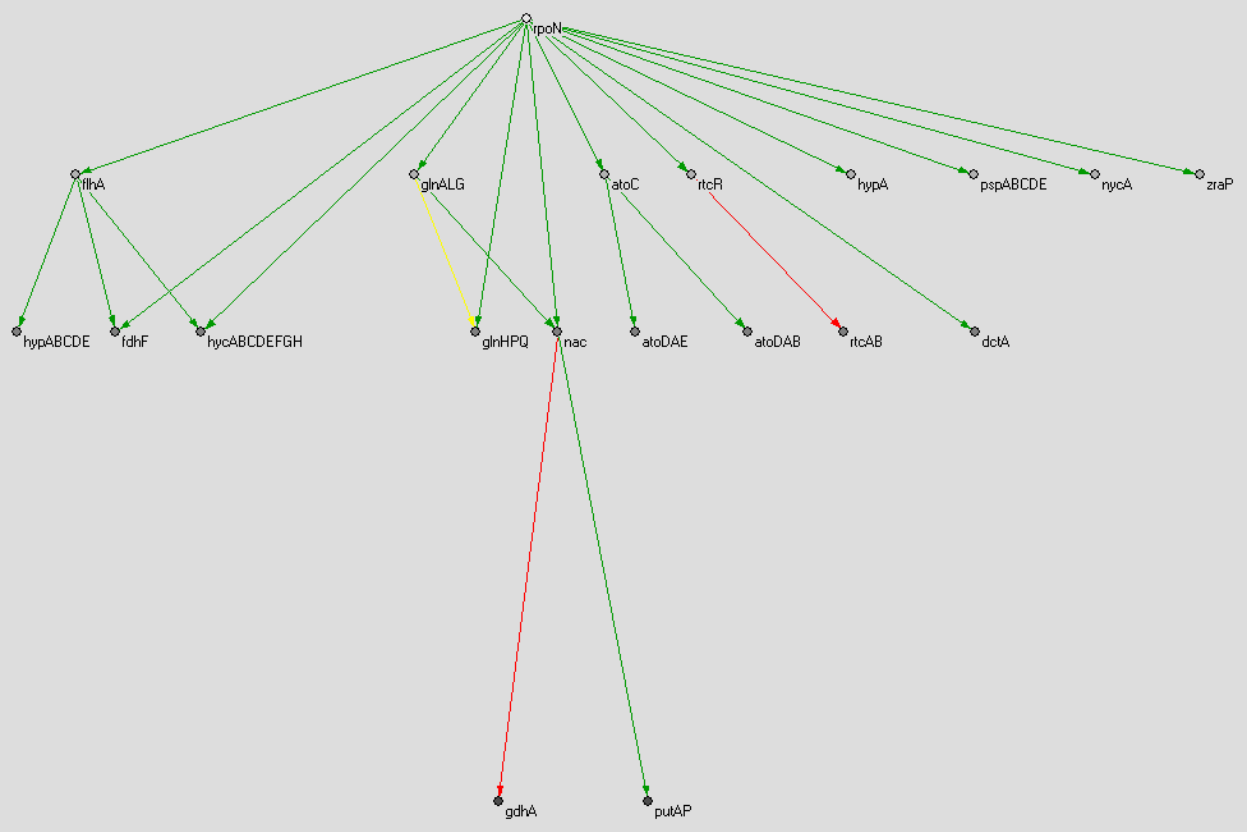


**Fig. 18. The rpoN origon.**
Activating (green), an repressing (red) links are shown

# Transcriptional Response of *E. coli* to Stimuli: Further Details

In the following, we analyze the transcriptional response of *E. coli* to four different types of stimuli, arguing for the functional importance of origons. For each of the perturbations, we plot:

a. the shape of the applied external stimulus as perceived by the cell
b. the double z-scores calculated for all 76 origons
c. the distribution of z-scores within the significantly affected origons, compared to the same number of randomly chose nodes
d. the expression profile of the most affected operons within the most affected origons.

The ideal dataset for our methods would contain a large number of experiments, changing as few factors in the medium as possible. The cell culture should therefore be grown in a chemostat, or biofermentor, maintaining pH, temperature and gas pressures constant (except when the perturbation is imposed through one of these factors).

**Aerobic-Anaerobic Shift Experiment Set.** We start our analysis with the aerobic-anaerobic shift experiment set (1), discussed in the manuscript. Despite not being a dynamic dataset (as the culture was left to reach equilibrium after each perturbation), it is most suitable for our methods, due to the high number of experiments and the simplicity of the perturbation (most likely, only oxygen and nitrogen concentration changed in the medium). The external signal profile is shown on the left in Fig. 19, while the double Z-scores are shown on the right. As these figures indicate, the number of significantly affected origons ($Z_{NS}$>2) is small.

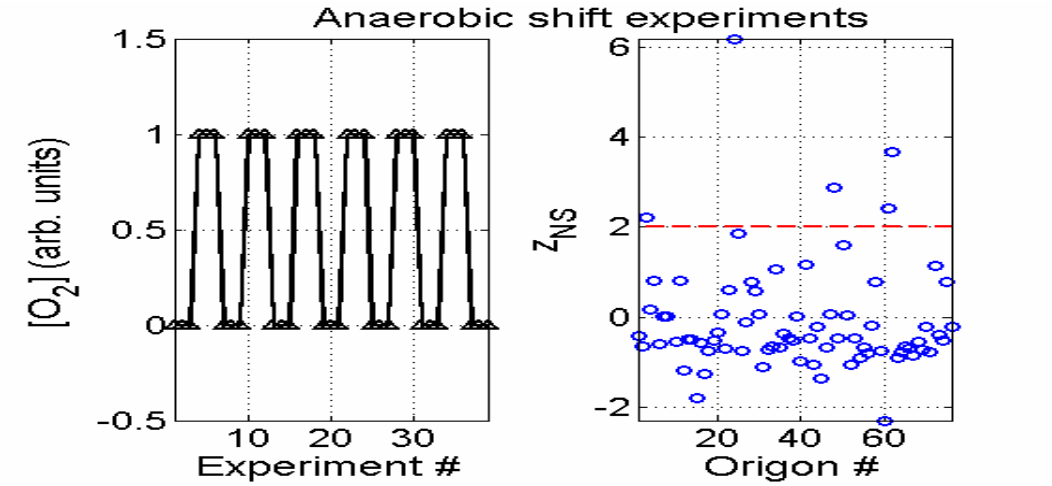


**Fig. 19. Anaerobic-aerobic shift experiments:**
The external signal (left), and the magnitude of the elicited response (right) are shown

The most affected origon for the aerobic-anaerobic shift experiment set was the one rooted at *fnr*. The FNR protein (encoded by *fnr*) is a known sensor of anaerobic conditions, and a large number of transcription units (TUs) within its origon are affected by these perturbations (as indicated on the left in Fig. 20). The four most affected TUs are *sodA*, *cyoABCDE*, *appCBA* and *frdABCD* (shown on the right in Fig. 20). Note that the response of some TUs correlates with the

external signal, while the response of others anti-correlates with it.

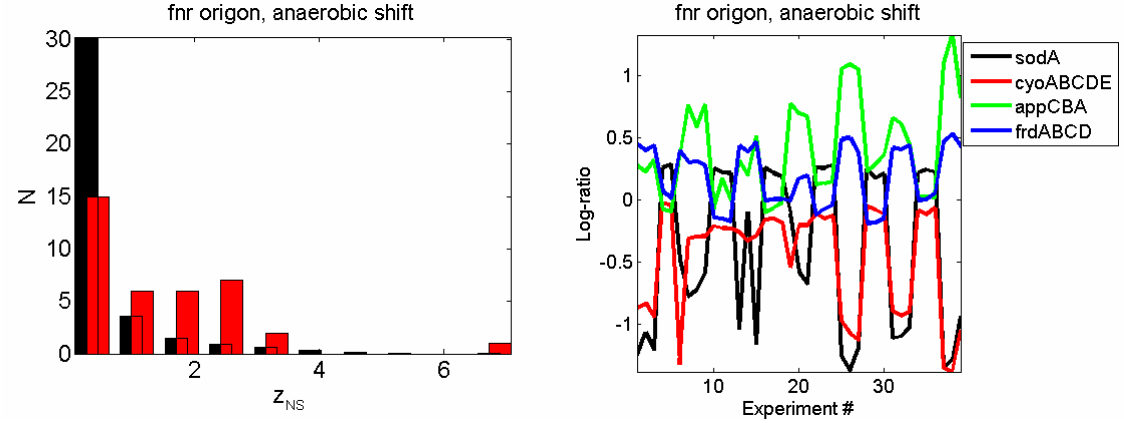


**Fig. 20. Response within the *fnr* origon, $Z_{NS}$ = 5.99.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The second most affected origon for the aerobic-anaerobic shift experiment set was the one rooted at *soxR* (Fig. 21). SoxR is known to be a sensor of oxygen concentration – in fact it uses an iron-sulfur cluster to monitor the oxygen concentration just like FNR.

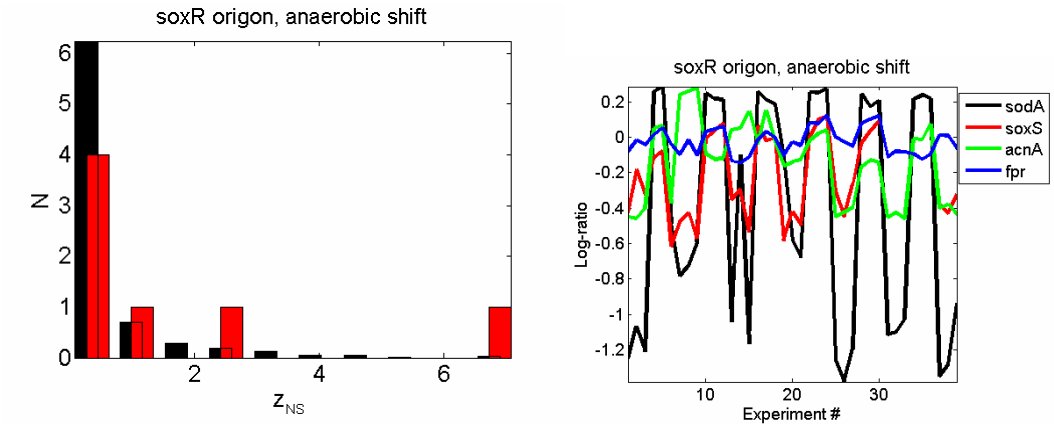


**Fig. 21. Response within the *soxR* origon, $Z_{NS}$ = 3.95.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The root gene of the third most significantly affected origon (Fig. 22), *narL* encodes NarL, a transcriptional regulator of anaerobic respiration and fermentation, which forms a two-component regulatory system with NarX (or NarQ). It activates/represses several operons in response to nitrate/nitrite induction signals, which are transmitted by NarX or NarQ.

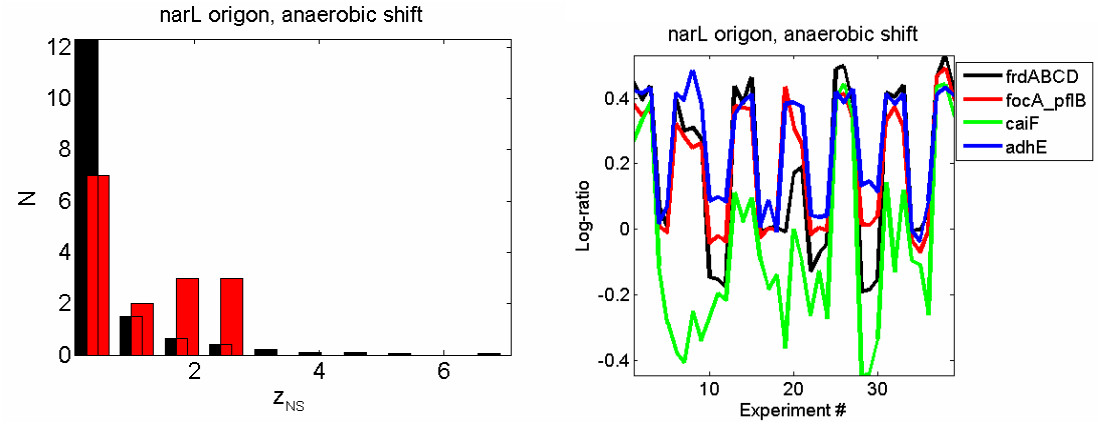


**Fig. 22. Response within the *narL* origon, $Z_{NS}$ = 2.74.**
The distribution of *zNS* within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The root gene of the fourth most affected origon is *rpoN* (Fig. 23), encoding the sigma 54 factor RpoN, which is responsible for gene nitrogen-dependent gene expression. The fact that the *rpoN* origon is affected indicates that aerobic conditions were achieved by replacing oxygen with nitrogen in the medium.

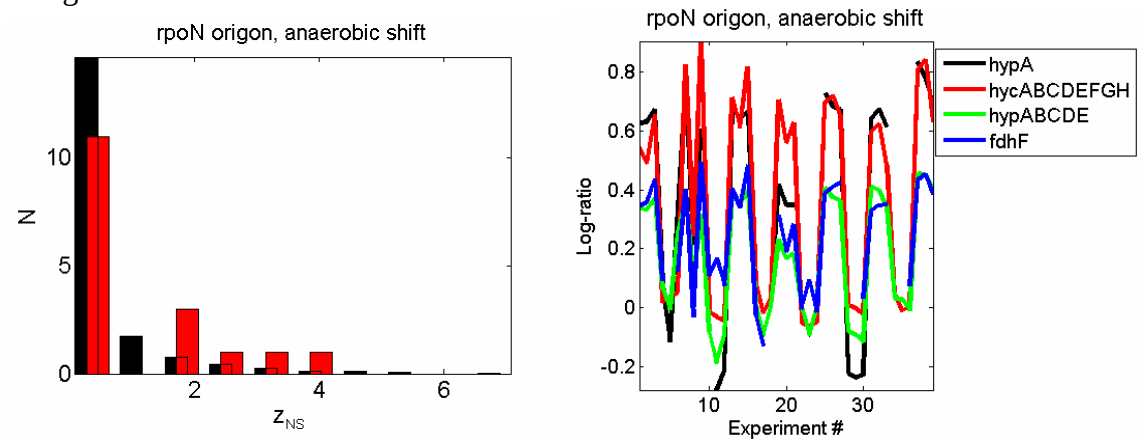


**Fig. 23. Response within the *rpoN* origon, $Z_{NS}$ = 2.47.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

*E. coli* has two types of arginine decarboxylase. At the root of the fifth most affected origon (Fig. 24) is *adiAY*, encoding the components of degradative arginine decarboxylase, which is induced in rich medium, at a low pH in the presence of excess substrate under anaerobic conditions. This origon is very small: it contains only two TUs.

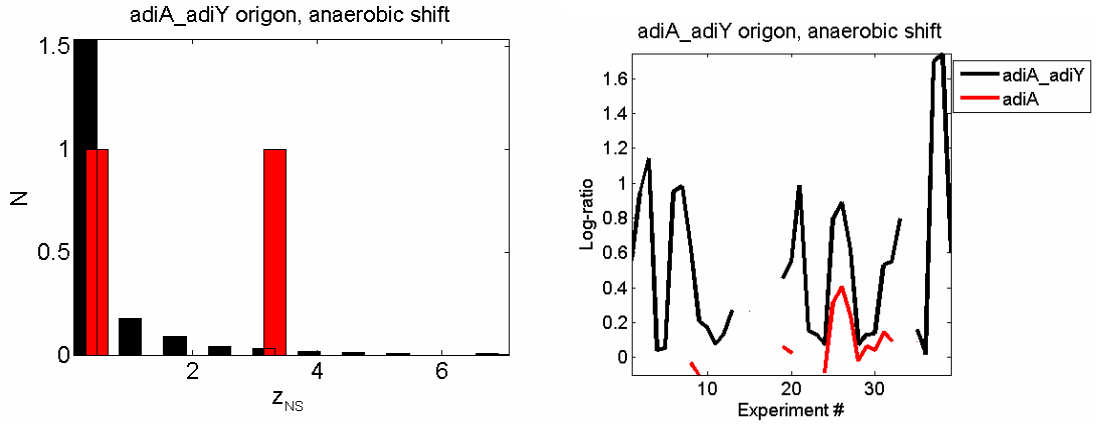


**Fig. 24. Response within the *adiAY* origon, $Z_{NS}$ = 2.11.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The last significantly affected origon originates at *fruR* (Fig. 25), encoding FruR - a dual transcriptional regulator that is implicated in the regulation of a large number of operons that encode enzymes which comprise central pathways of carbon metabolism. The nodes (TUs) *aceBAK* and *icdA* are also part of the *fnr* origon.

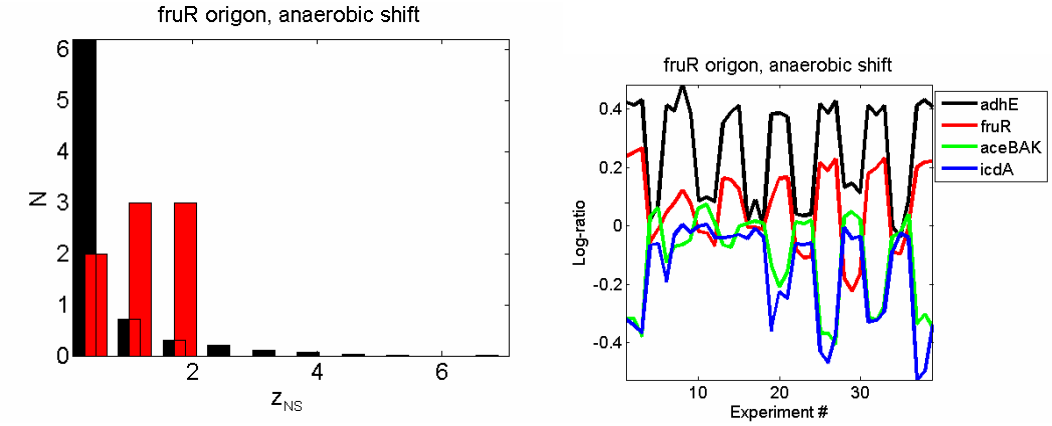


**Fig. 25. Response within the *fruR* origon, $Z_{NS}$ = 2.10.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

**Hydrogen Peroxide Treatment Experiment Set.** Next, we analyzed the transcriptional response of *E. coli* to hydrogen peroxide ($H_2O_2$) treatment (2). The samples were taken from *E. coli* batch cultures in a 2 liter biofermenter. (Batch cultivation is not ideal for the purpose of our analysis, because the medium is changing, the cells shift their nutrition, and the culture ends up in stationary phase.) After taking a sample from a culture in logarithmic growth (at 100 min), $H_2O_2$ was introduced at 140 min, and gene expression was interrogated at 145, 150, 160, 170, 180, 200, 255, 330, 345, and 390 minutes. Unfortunately, $H_2O_2$ concentration (the external signal) was not monitored during the experiment. Therefore, we approximate the signal by a step function from 0 to 1 at t=140 min (see Fig. 26, *Left*). Since the medium is changing and the culture is approaching stationary phase towards the end of the timecourse, we considered only the time points up to 255 min.

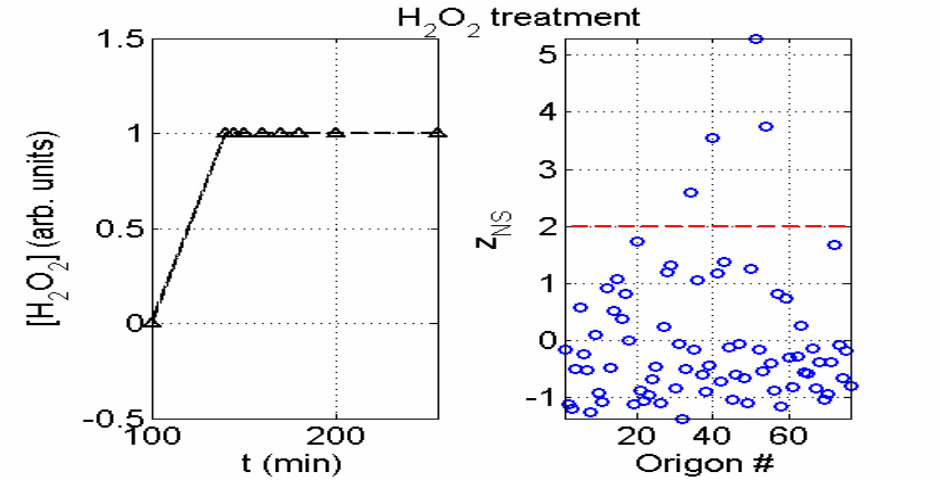


**Fig. 26. $H_2O_2$ treatment:**
The external signal (left), and the magnitude of the elicited response (right) are shown.

The most significantly affected origon following $H_2O_2$ treatment is rooted at *oxyR* (Fig. 27). OxyR participates in controlling several genes involved in the response to oxidative stress and the production of surface proteins that control the colony morphology and auto-aggregation ability. It regulates intracellular $H_2O_2$ levels by activating genes such *dps*, *fur*, and *katG*. The peaked expression profile of the most affected operons within the *oxyR* origon (Fig. 27, *Right*) suggest either that the concentration of $H_2O_2$ might have been decreasing in the medium, or the cells have adapted to the new conditions through changes in mRNA and protein expression.

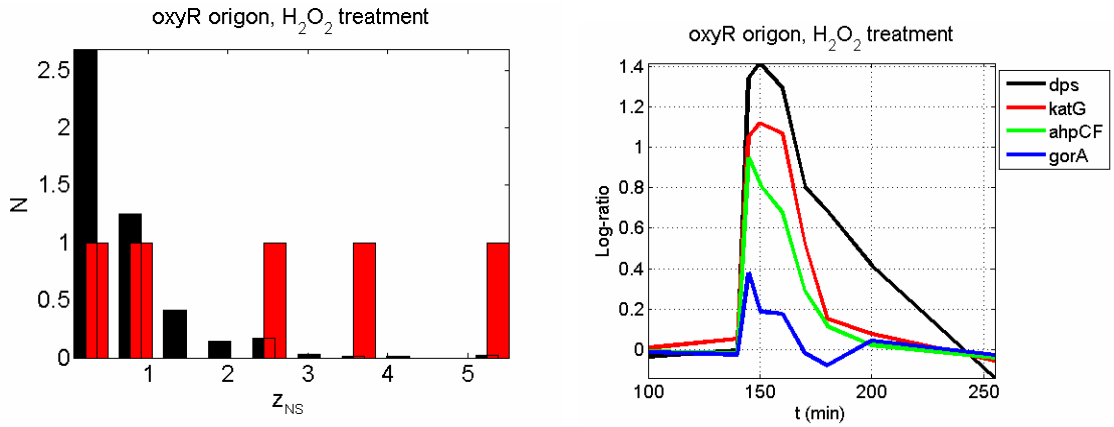


**Fig. 27. Response within the *oxyR* origon, $Z_{NS}$ = 5.29.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The second most affected origon is rooted at *purR* (Fig. 28), encoding PurR, the controller of several genes involved in de novo purine nucleotide biosynthesis. PurR necessarily binds hypoxanthine and guanine to become active. As shown in Fig. 28, *Right*, purine synthesis is shut down following the stress, which correlates with the fact that the cells stop growing (and synthesizing DNA).

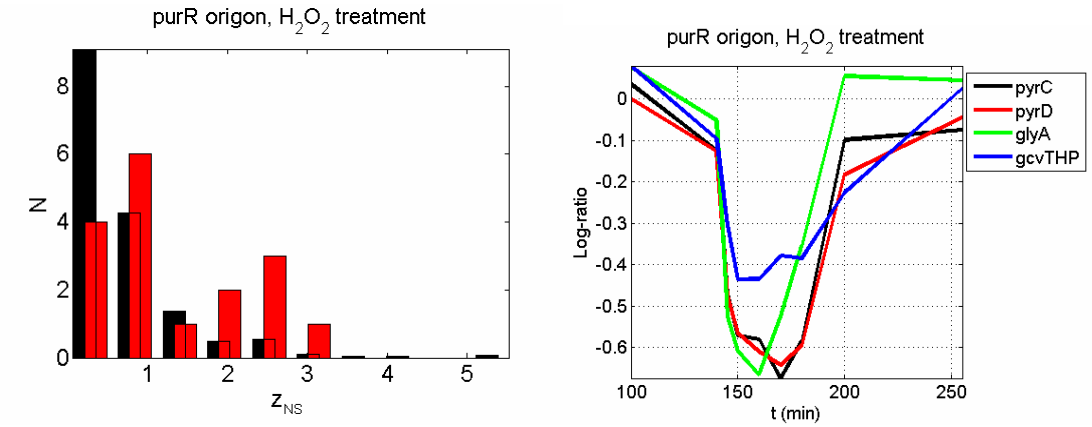


**Fig. 28. Response within the *purR* origon, $Z_{NS}$ = 3.76.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

The third most affected origon to respond is rooted at *lexA_dinF* (Fig. 29). LexA participates in controlling several genes involved in the SOS response, usually induced following DNA damage. The fact that this origon is affected indicates that oxydative stress results in some DNA damage.

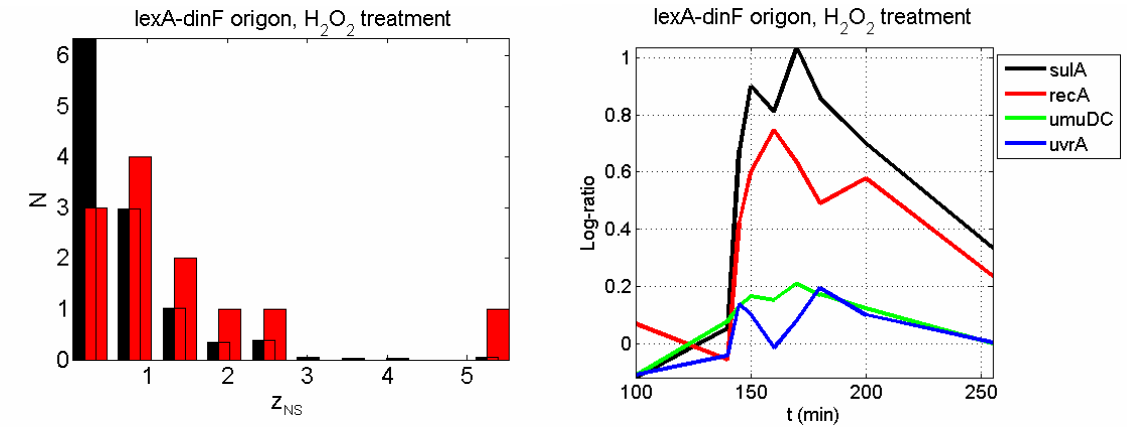


**Fig. 29. Response within the lexA origon, $Z_{NS}$ = 3.54.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

Finally, the fourth most affected origon is rooted at *himA* (Fig. 30), encoding a subunit of IHF (Integration Host Factor). IHF introduces sharp bends in the DNA and is required for site-specific recombination, DNA replication, plus transcriptional and translational control. By affecting DNA replication, IHF and its origon are probably involved in slowing growth down, and repairing damage (by recombination).

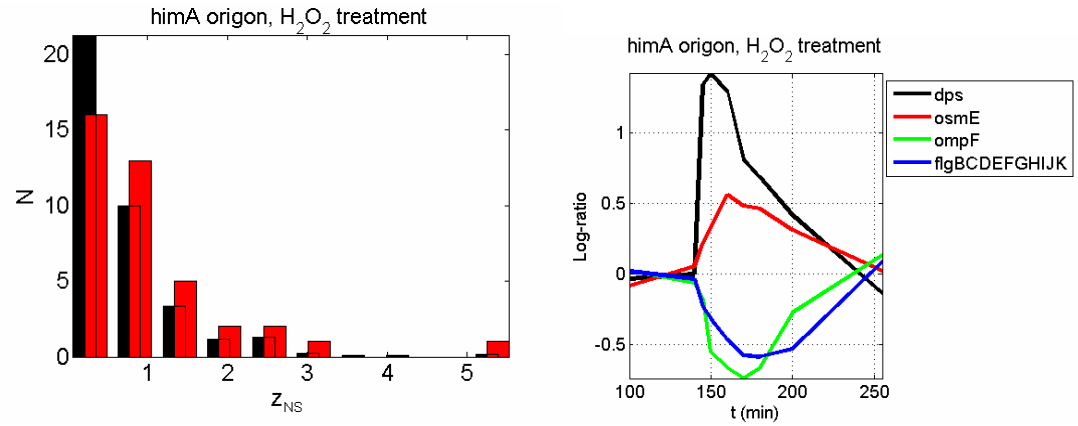


**Fig. 30. Response within the *himA* origon, $Z_{NS}$ = 2.59.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the four most affectetd TUs are shown.

**Diauxic Shift (Glucose to Lactose) Experiment Set.** We also analyzed the transcriptional response of *E. coli* to diauxic shift (2). Cells were grown as in the hydrogen peroxide treatment experiment, in minimal medium initially containing 0.05% glucose and 0.15% lactose. The concentration of lactose was not monitored in the environment, but the concentration of glucose is known (see Fig. 31; Tyrrell Conway, personal communication).

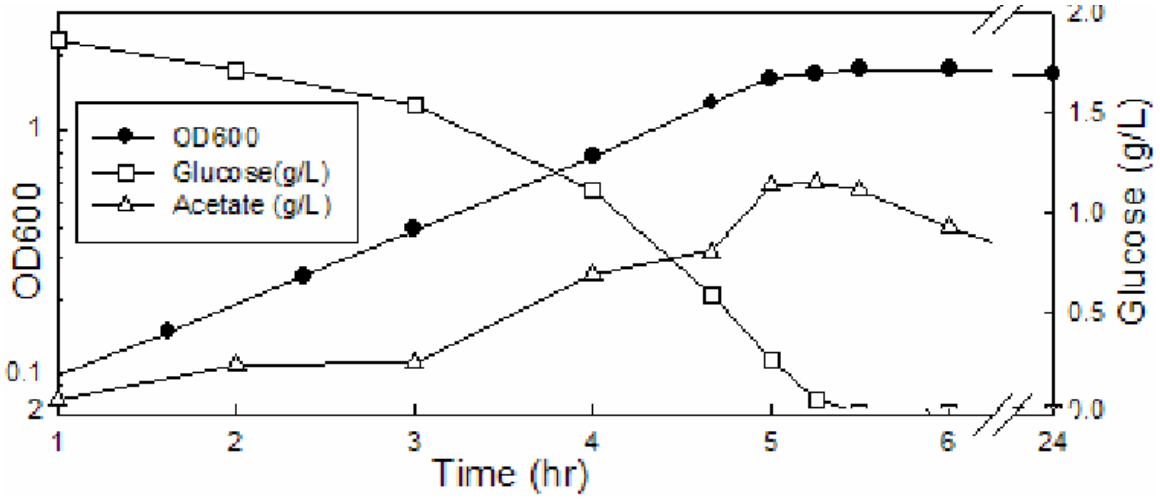


**Fig. 31. Cell density, and the concentrations of glucose and acetate during the diauxic shift experiment. Reproduced with permission from Advances in Microbial Physiology (in press).**

Despite knowing the concentration of glucose, we assumed the external signal to have the shape of a step function (see Fig. 32, *Left*). This assumption is valid if glucose sensing within the cell has a step-like characteristic. We considered the glucose 'as perceived by the cell' drop to zero at the beginning of the diauxic lag. Samples were collected similarly to the $H_2O_2$ experiment, with a sample taken from the logarithmic culture at 90 minutes, and sampling at time points 135, 165, 180, 185, 190, 195, 205, 215, 225, 235, 273, 345, 360, 375, 390, and 420 minutes after onset of the lag. Just as before, due to the culture approaching stationary phase we considered time points up to 273 minutes.

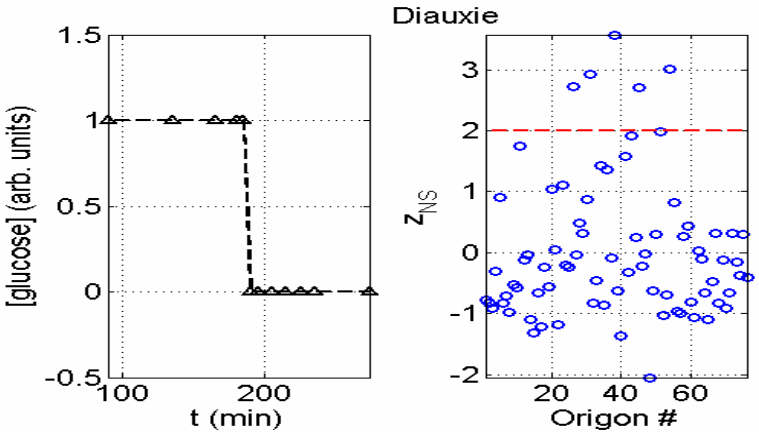


**Fig. 32. Diauxic shift.**
The external signal (left), and the magnitude of the elicited response (right) are shown.

The most significantly affected origon was the one rooted at *lacI* (Fig. 33). LacI participates in controlling several genes involved in lactose catabolism, and therefore it is expected to be significantly affected by a diauxic shift from glucose to lactose.

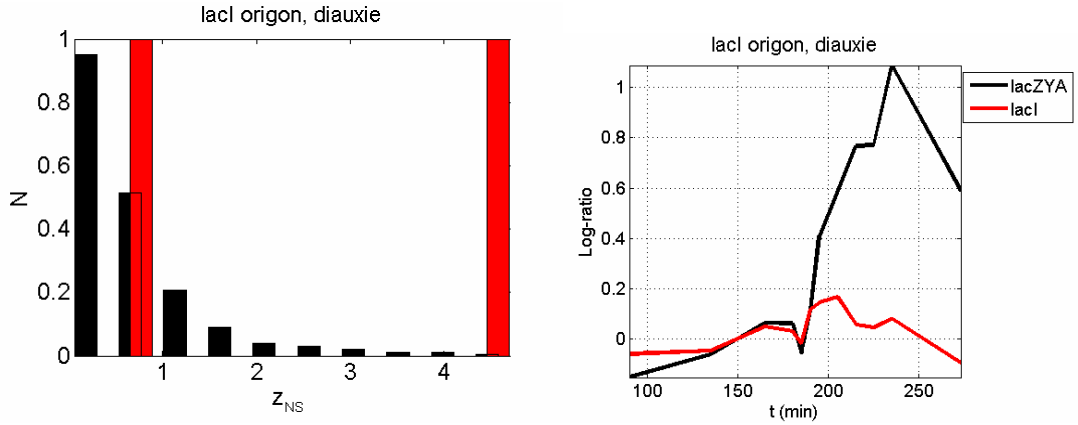


**Fig. 33. Response within the *lacI* origon, $Z_{NS}$ = 3.57.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

Just like in the $H_2O_2$ treatment experiment, the second most affected origon is rooted at *purR* (Fig. 34). Due to the downregulation of purine synthesis and the resulting slower growth, *purR* seems to be involved in stress conditions when growth rate is reduced.

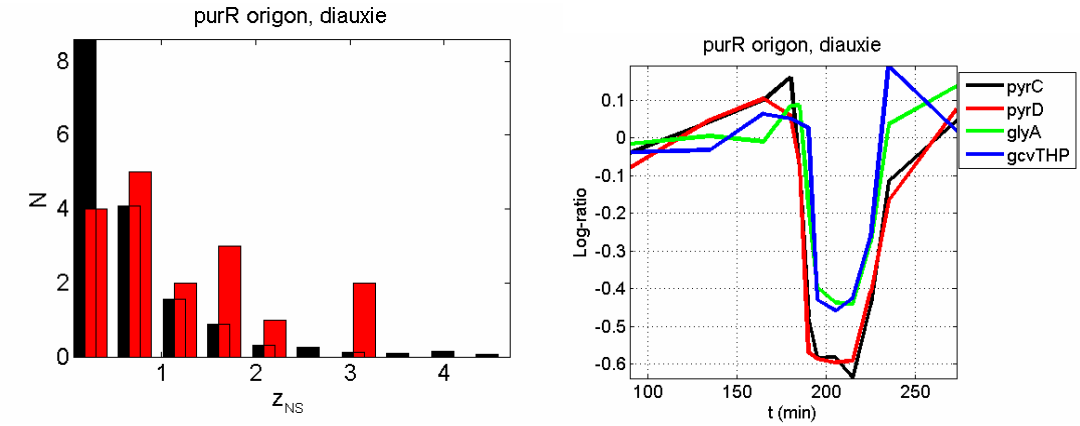


**Fig. 34. Response within the *purR* origon, $Z_{NS}$ = 3.01.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

The third most affected origon is *glpR* (Fig. 35). GlpR is responsible for glycerol uptake and metabolism in various conditions. Most genes in this origon are co-regulated by CRP and GlpR.

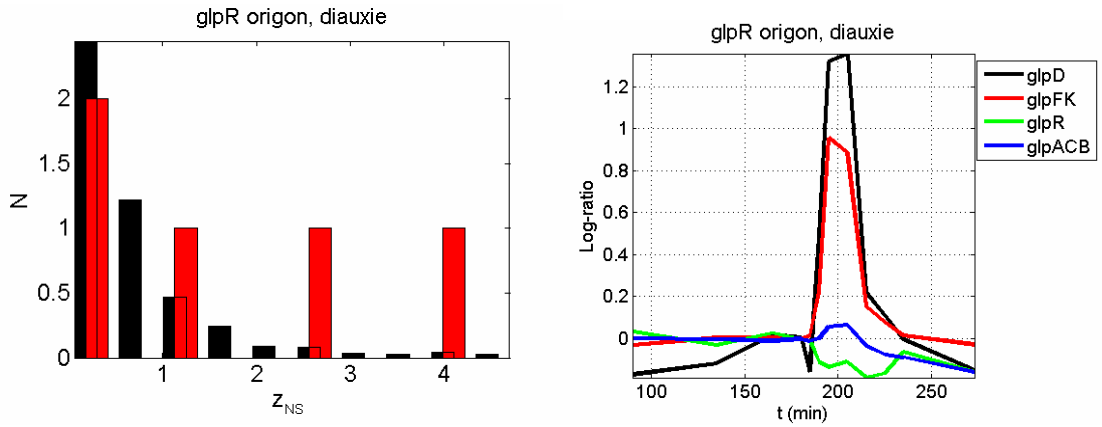


**Fig. 35. Response within the *glpR* origon, $Z_{NS}$ = 2.93.**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

The fourth most affected origon is galR (Fig. 36). GalR is responsible for galactose uptake and metabolism. Most genes in this origon are co-regulated by CRP and GalR.

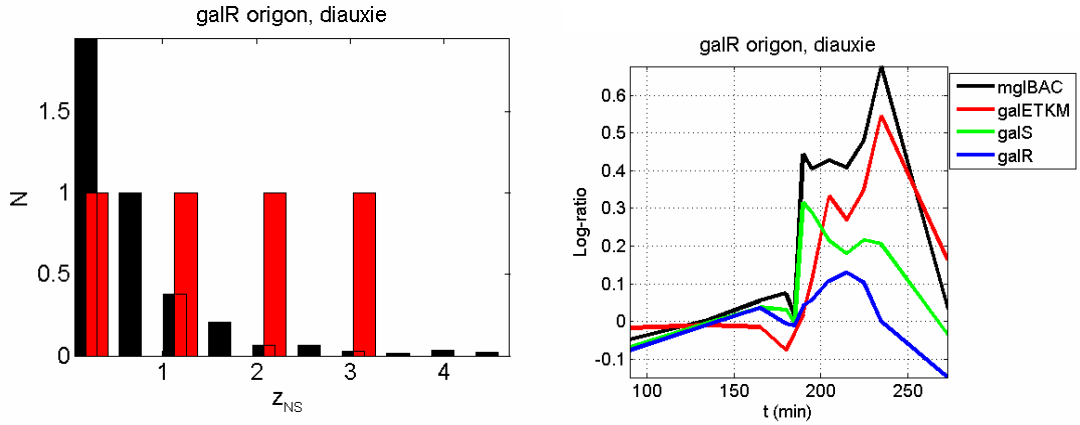


**Fig. 36. Response within the *galR* origon, $Z_{NS}$ = 2.72**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

The fifth most affected origon is *mlc* (Fig. 37). Mlc participates in controlling several genes involved in glucose uptake or glycolisis. CRP co-regulates most genes in this origon as well.

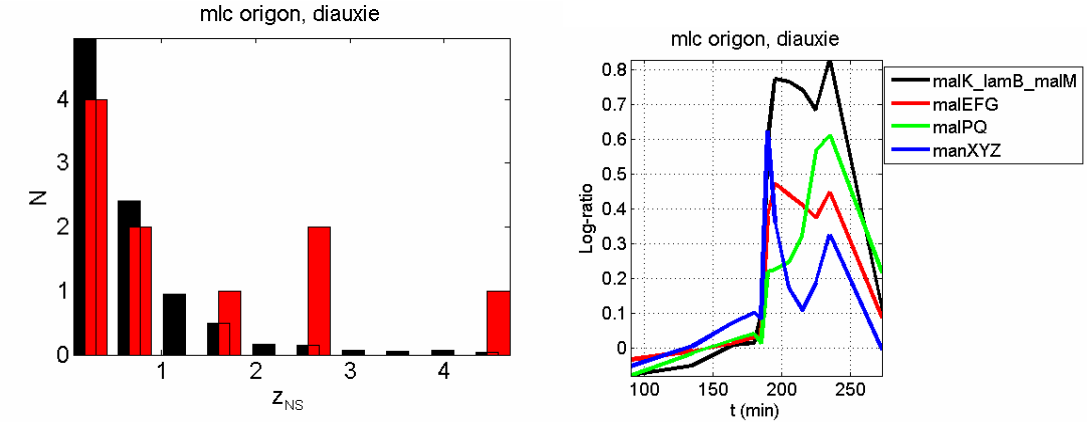


**Fig. 37. Response within the *mlc* origon, $Z_{NS}$ = 2.70**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

Finally, we need to mention that the origon rooted at the encoder of the glucose sensor CRP also ranks high among the affected origons (#7 in the list). The fact that it did not rank among the top five is probably due to its large size, many of its nodes being co-regulated by other TFs. Also, besides *lacI*, other sugar-related origons are affected, indicating that as soon as glucose concentration becomes limited, the cell starts searching for alternative carbon sources.

**UV Exposure Experiment Set.** Finally, we analyzed the transcriptional response of *E. coli* to UV exposure (3). Cells were grown in Davis medium containing 0.4% glucose. A total of 70 ml of culture was placed into a petri dish and irradiated for 60 seconds by a 15-W germicidal lamp. Samples were taken 5, 10, 20, 40, and 60 minutes following UV exposure. In this case, we defined the signal as DNA damage (the UV exposure lasted only for 60 seconds, for which the signal would have had the shape of a spike). Indeed, the cells detect UV by the resulting DNA damage, which in contrast to the actual UV signal is long-lasting signal (as shown in Fig. 38).

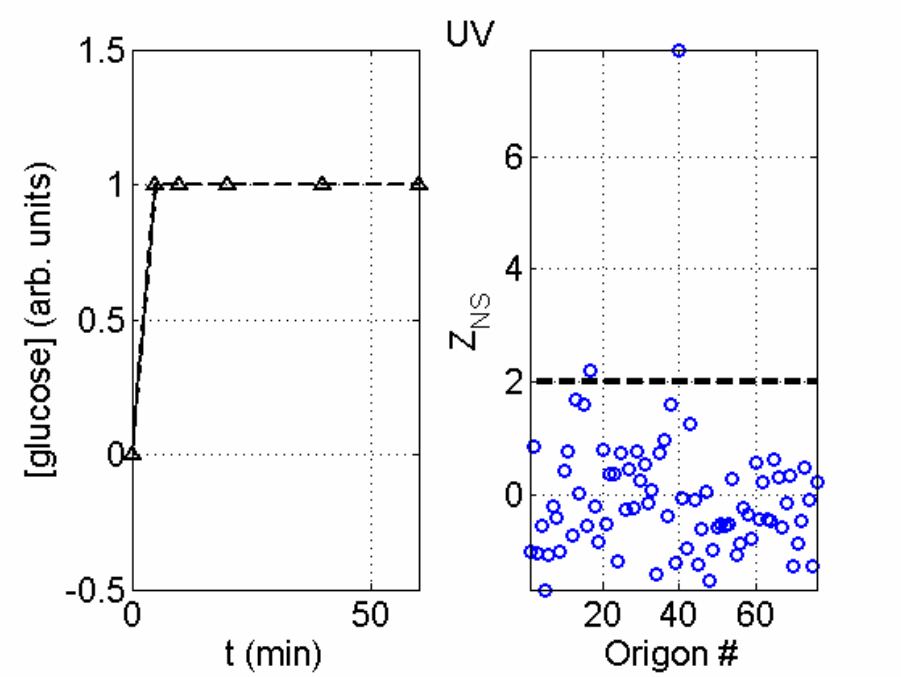


**Fig. 38. UV exposure.**
The detected signal (left), and the magnitude of the elicited response (right) are shown.

Following UV treatment, the *lexA_dinF* origon (Fig. 39) was affected very significantly compared to any other origon. LexA together with RecA is the primary sensor of DNA damage, controlling several genes involved in the SOS response and DNA repair. Therefore, it is expected that this origon should respond strongly to UV exposure.

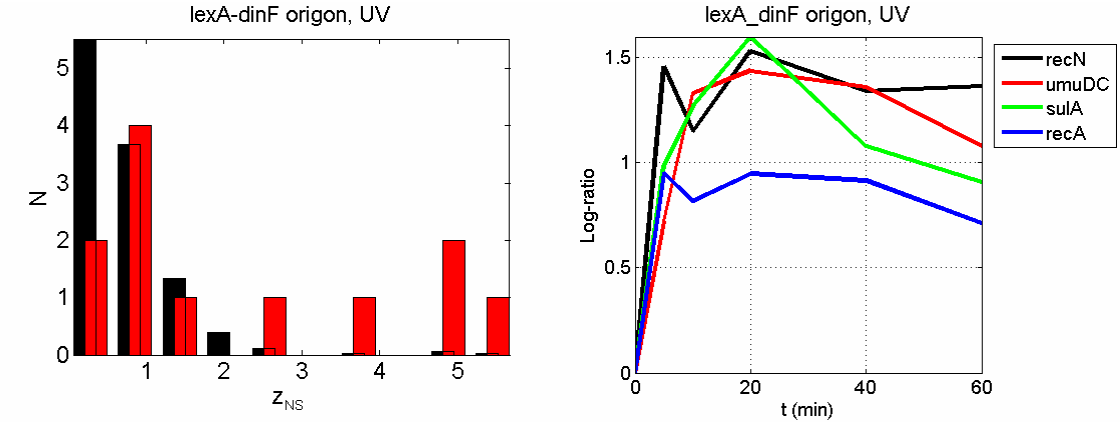


**Fig. 39. Response within the *lexA_dinF* origon, $Z_{NS}$ = 7.45**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

The *dnaA* origon (Fig. 40) was also affected following UV exposure. DnaA is responsible for replication initiation. The perturbation  replication is likely to be stopped following DNA damage causing alterations in DnaA activity.

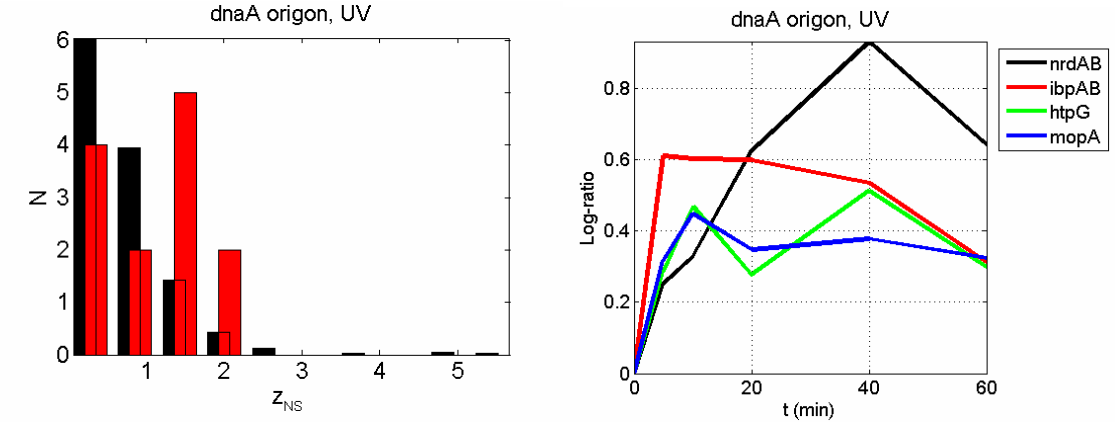


**Fig. 40. Response within the *dnaA* origon, $Z_{NS}$ = 2.20**
The distribution of $z_{NS}$ within the origon versus randomly chosen TUs (left), and the response of the most affectetd TUs are shown.

## References


1. Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J. & Palsson, B. O. (2004) *Nature* **429,** 92-96.
2. Chang, D. E., Smalley, D. J. & Conway, T. (2002) *Mol. Microbiol*. **45,** 289-306.
3. Courcelle, J., Khodursky, A., Peter, B., Brown, P. O. & Hanawalt, P. C. (2001) *Genetics* **158,** 41-64.