# Controlled Measurement of Cellular Components in *E. coli* Reveals Broad Changes in Cellular Composition under Long-Term Starvation

John R. Houser1,2, Craig Barnhart1,2, Daniel R. Boutz1,2, Sean M. Carroll3,@, Aurko Dasgupta1,2 Joshua K. Michener3,\*, Brittany D. Needham1,2, Ophelia Papoulas1,2, Viswanadham Sridhara4, Dariya K. Sydykova1, Christopher J. Marx3,5,6,7, M. Stephen Trent1,2,8, Jeffrey E. Barrick1,2,4,8#, Edward M. Marcotte1,2,4,8#, Claus O. Wilke1,2,4,9#

1Institute for Cellular and Molecular Biology, The University of Texas at Austin  
2Center for Systems and Synthetic Biology, The University of Texas at Austin  
3Department of Organismic and Evolutionary Biology, Harvard University  
4Center for Computational Biology and Bioinformatics, The University of Texas at Austin  
5Faculty of Arts and Sciences Center for Systems Biology, Harvard University

6Department of Biological Sciences, University of Idaho

7Institute for Bioinformatics and Evolutionary Studies, University of Idaho  
8Department of Molecular Biosciences, The University of Texas at Austin  
9Department of Integrative Biology, The University of Texas at Austin

@Current address: KnipBio Inc., Harvard, MA

\*Current address: Department of Biological Engineering, Massachusetts Institute of Technology

## Abstract

How do bacteria regulate their cellular physiology in response to changes in growth conditions? Several prior works have investigated the physiological response of cells to environmental changes, but previous studies on this topic were far from comprehensive. Many studies only measure one type of genome-wide data and, due to technical limitations, often don’t include certain subpopulations of molecules within the types of genome-wide data they measure. Here, we present a detailed and comprehensive characterization of *E. coli* growth and starvation over a time-course lasting two weeks. We measured multiple cellular components, including RNA and proteins, with deep genomic coverage. In addition, we measured lipid modifications and flux through central metabolism, further expanding the completeness of our dataset. We chose glucose-limited minimal medium conditions, such that most cells survive starvation and mutant, dividing populations do not emerge. Thus, our study focuses on the physiological response of *Escherichia. coli* to starvation, not on the adaptation of *E. coli* to utilize alternative nutrients as was the case in previous long-term starvation studies. By comparing and contrasting our transcriptomic and proteomic data, we have found that post-transcriptional regulation, such as protein degradation, can contribute to important differences in protein expression and physiological regulation. We have also found that mRNAs are widely down-regulated in response to glucose starvation, presumably as a strategy for reducing new protein synthesis. Finally, we outline a novel strategy for sorting expression profiles based upon basic behaviors in an unbiased manner that also takes into account experimental error.

# Introduction

Many global changes in cellular physiology occur during the growth of a typical laboratory culture of a microorganism, such as *Escherichia coli,* as it transitions from an exponential rate of cell division to a stationary phase where it eventually ceases dividing as nutrients become exhausted(Neidhardt and Curtiss, 1996). However, how these changes affect specific cellular components and processes is not fully known. Existing surveys, even if conducted at the genome scale, tend to have limited completeness, in at least two ways. First, most studies consist of the piece-wise collection of only one type of genome-scale data. For example, they either measure changes in expression, through RNA or protein levels, or they measure changes in metabolites. Second, technological limitations often prevent the detection of some subset of molecules in a category of interest. For example, small bacterial RNA species with key roles in regulation are lost when using typical purification methods (Stead et al., 2012). Furthermore, DNA microarray-based methods for profiling gene expression can only detect specific RNA sequences depending on the design of their probes, whereas RNA-seq transcriptomic methods theoretically recover all RNA species in a sample (Wang et al., 2009). Similarly, in proteomics, 2-D gel electrophoresis approaches typically detect many fewer proteins than newer mass spectrometry based shotgun methods (Soares et al., 2013; Wiśniewski and Rakus, 2014).

Moreover, while the short-term changes in cellular physiology that occur in a laboratory culture of *E. coli* have been the subject of intensive study, considerably less is known about the changes in cellular composition that occur during the long-term survival of *E. coli* and other non-spore-forming microbes under starvation, despite the likely prevalence of this condition in nature (Morita, 1990). Most studies of this metabolic state have concentrated on using rich media in which growth ceased and there was an ecological catastrophe accompanied by the death of most cells and mutants emerged that continued to divide (Finkel and Kolter, 1999). Thus, these were often studies of adaptation to new nutrient sources rather than purely of cellular physiology in stressed and starving, but genetically wild-type, cells. The specific physiologic changes that occur in *E. coli* under long-term starvation thus are not fully understood.

Here we performed a time course experiment of *E. coli* B REL606 growth and starvation up to two weeks. We used a chemically defined glucose-limited medium in which cells entered a starvation state but did not die. We collected genome-wide RNA and protein levels at multiple time points, as well as lipid modifications and metabolic-flux data, all under identical, controlled experimental conditions. The resultant data set serves as a rich resource for computational models that span and integrate cellular sub-systems and for cataloguing and correlating the responses of specific genes and/or molecules across cellular subsystems during growth and long-term starvation. We found that post-translational regulation can play a role in buffering proteins with consequences to regulation of *E. coli* physiology, and that the mRNA pool was drastically reduced during starvation, possibly to limit new protein synthesis overall. Additionally, we used a novel general approach for unbiased classification of expression time courses that could aid in analyzing future gene expression and proteomic time courses.

# Results

## Controlled measurements of multiple cellular components yield highly reproducible data and unprecedented depth of coverage.

We grew multiple culturesof *E. coli* REL606, from the same stock, under identical growth conditions of long-term glucose starvation, in the same media. The samples were subsequently distributed to different laboratories that measured RNA, protein, lipids, and metabolic flux ratios (Figure 1A). Freezer stocks of the REL606strain where revived for 24 hrs, diluted and preconditioned for another 24 hrs, and diluted again at the start of the experimental time course (Figure 1B). Each biological replicate was performed on separate days. In a pilot experiment a growth curve was taken to determine informative time points for analysis (Figure 1C). Time points spanning 3 hrs to 2 weeks where collected and equitably distributed to measure RNA via RNAseq, proteins via LC/MS, lipids via MALDI-TOF MS and ESI MS, and central metabolic fluxes via 13C labeled glucose and GC-MS (Figure 1D). In our conditions, the optical density at 600nm (OD600) changed little once cells entered stationary phase (Figure 1B). Additionally, cell viability remained constant after the cells entered stationary phase at 24 h for up to a week. From one to two weeks, the number of viable cells per culture count decreased by just 38% (Figure 1B).

We first assessed reproducibility of our data, and found that it was highly reproducible. Biological replicates of protein and RNA measurements correlated highly with each other. We saw Spearman correlations of 0.95 and 0.92 between biological repeats of raw proteomics counts and correlations of 0.93 for raw RNAseq counts between biological replicates (Figure 1-supplemental 1). We next compared how many different RNA and protein species we detected to previous studies, and found that we obtained more coverage compared to other similar studies that measured both protein and RNA (Table 1). Yoon et al. used 2D gels and microarrays to measure 60 significantly changing proteins and 4144 mRNAs in the *E. coli* REL606 strain, the same strain used in this study (Yoon et al., 2012). By comparison, we observed over 2600 proteins with ~1200 that changed significantly through the duration of our time course along with 4116 mRNAs, 89 sRNAs, and 85 tRNAs (at early exponential phase). Even though the total number of proteins Yoon et al. observed at early exponential phase was not reported, it was likely an order of magnitude less than our observations, if it followed the same pattern as the proteins undergoing fold change. Taniguch et al. measured protein and mRNA content of single cells using YFP fusions and FISH, resulting in the measurement of 1,018 proteins and 137 transcripts in an *E. coli* K12 variant (Taniguchi et al., 2010). Lewis et al. also measured ~1,000 proteins and RNA expression of 4428 genes. Although these data sets were published separately, they were performed in the same lab and under similar conditions and thus were also comparable to a degree (Lewis et al., 2010, 2009). Our proteomics measurements were far more complete than other comparison studies, providing more than 1,000 additional protein observations than the previous top comparative study, as many mRNAs as other studies, and additional data on tRNAs and sRNAs.

Our experiments also provided coverage comparable to or better than other experiments that focus on proteomics or RNA measurements alone. Using stable isotope labeling of amino acids (SILAC), Soares et al. observed 2,053 proteins in at least 1 of 2 biological repeats, at a false discovery rate (FDR) of <1% (Soares et al., 2013). We measured 2,658 proteins in at least 1 of 3 biological repeats with around 2,200 protein IDs per sample using the same FDR cutoff. A more recent study, using the filter aided sample preparation (FASP) method, also observed around 2,200 proteins per sample, comparable to our recovery (Wiśniewski and Rakus, 2014). Additionally, using RNAseq, we recovered as many mRNAs as microarray approaches do, with the added benefit of measuring 89 sRNAs and 85 tRNAs at the same time (Raghavan et al., 2011). As a point of reference, previous RNAseq experiments on the *E. coli* K-12 strain found 133 sRNAs and 4161 genes. Thus our recovery of both proteins and RNA represents the state of the art of the field, far outperforming recent comparative studies. As an added benefit of our study, we also simultaneously measured lipid A, phospholipids, and flux ratios in central metabolism, covering a wider range of cellular components than previous comparison studies.

## Measured mRNAs are regulated in a relatively more uniform manner compared to proteins.

We next investigated changes in mRNA and protein abundance over time. Due to translational and post-translational regulation we expected differences in the response of mRNA transcripts and proteins in response to glucose starvation. To visualize these differences we compared and contrasted the general trends in the response of mRNA and proteins by way of *K*-means clustering. To simplify the analysis we focused on only those mRNAs and proteins that were changing significantly (as measured by false discovery rate and fold change cutoff, respectively) throughout the time course, yielding ~1900 significantly changing transcripts/proteins. To perform *K*-means clustering, an arbitrary choice for the number of clusters must be made such that the profiles are well separated into groups with unique and distinct behaviors. (An alternative classification approach that does not depend on such an arbitrary choice will be presented below.) We varied the number of clusters for both mRNA and protein profiles, and we found the best clustering performance to be around 15 clusters for the mRNA profiles and 25 clusters for the protein profiles. Thus, the mRNAs appeared to respond in a more uniform manner than the proteins did. This finding is illustrated by the heatmap of the cluster centroids of mRNA and protein (Figure 2A and B, respectively). A vast majority of the mRNAs were down-regulated while the protein response was much less uniform, likely due to differences in post-translational modification such as degradation. Additionally, the mRNA profiles showed a clear separation between early and late time points with a transition period around 6-8 h. After this transitional period of entry to stationary phase, the transcription profiles remained relatively constant, with only minor changes in expression. At two weeks some of the transcripts began changing again, perhaps signaling a further shift in cell state.

As the cells entered stationary phase, overall demand for new protein synthesis was significantly decreased, demand for certain stress response proteins increased, and resources became limiting. New protein synthesis can be globally limited in at least three ways: by reducing the overall accessible rRNA, the charged tRNAs, or mRNA. To understand how these different RNA pools changed relative to each other, we calculated the relative amount of mRNA, tRNA, sRNA, and rRNA present in both ribosome depleted and non-ribosome depleted samples (Figure 2-supplemental figure 1A and B). In the non-ribosomal depleted case the fraction of rRNA changed very little throughout the course of the experiment while the tRNA fraction increased and the mRNA fraction decreased. In the ribosome depleted samples, the tRNA fraction also increased as the mRNA fraction decreased, confirming that this effect was not due to sensitivity, or sampling bias, issues resulting from rRNA dominating the RNA pool in the non-ribosome depleted sample.

We would like to emphasize that the above clustering of the RNA and protein abundances were independent. Therefore, we could not directly compare individual clusters between Figures 2A and B. The next section addresses the correlation between absolute and relative changes in abundance of individual proteins and their transcripts.

## Differences in post-transcriptional regulation leads to differences in correlation between individual mRNA and protein time courses.

It has been observed that absolute levels of proteins do not correlate strongly, relatively speaking, with their corresponding transcripts. In addition to a weak absolute correlation between mRNA and protein at a given time point we also expected, due to potentially long protein degradation rates in stationary phase, a weak relative correlation within a single time course between a protein and its transcript. To put proteins and RNA within a given sample on comparable scales, we normalized protein counts using the Apex method (Lu et al., 2007) for absolute quantification, and we normalized mRNA counts to the length of each transcript. Both protein and mRNA levels were then averaged across all three biological replicates. Additionally, all proteins and mRNAs were scaled by the average of all proteins and mRNA.

To relate the relative levels of a protein to its transcript we had to account for the underlying dynamics of the time courses. We considered two limiting cases: At one extreme we assumed each protein had a degradation rate slower than the time scale of the experiment. At the other extreme we assumed each protein was degraded on a time scale that was fast compared to the time scale of the experiment. In the first limiting case proteins integrate their transcript levels over time. In the second limiting case (relative) protein levels track with their (relative) transcript level. For proteins that do not integrally or proportionally relate to their transcripts, potential causes can be intermediate protein degradation rates, saturation effects, or some other post-translational regulation or nonlinearity.

The strongest absolute correlation, across the time course, between mRNA and protein occurred at three hours (Figure 2F, Spearman ρ=0.71, *P*=10-224). Absolute correlation between proteins and their corresponding transcripts were relatively strong for time points <=8 hrs, with a correlation coefficient of ~0.71. After 8 hrs, when cells had entered stationary phase, the correlation was much weaker, with correlations around 0.3-0.4 (Figure 2 – Supplemental Figure 2). The correlation at three hours was somewhat higher than is usually observed for correlations between RNA and protein for other measured prokaryotes and eukaryotes, which typically have Spearman correlations around 0.5 between proteins and their transcripts (Abreu et al., 2009; Gygi et al., 1999; Vogel et al., 2010; Vogel and Marcotte, 2012; Washburn et al., 2003).

Plotted in Figure 2C and D are histograms of the Spearman correlation coefficients (ρ) calculated for proteins vs. the integrals of their transcripts (integral regulation) and protein vs. their transcripts (proportional regulation), respectively. Approximately 15% of the proteins correlated highly (ρ>0.70) with the integrals of their transcripts whereas approximately 20% correlated highly with their transcript levels. There was little overlap between the two sets, as can be seen by the strong anti-correlation in the 2D histogram in Figure 2E of protein versus the integral and proportional levels of mRNA. Genes that were proportionally regulated were enriched for, among other things, locomotion and cell division. Genes that were integrally regulated were enriched for glycerol, aditol, and polyol metabolism. For a full list of proteins that were either proportionally or integrally related to their transcripts see Figure 2-supplemental table 1 and 2, respectively.

## RNAs within an operon correlated strongly while proteins within an operon did not necessarily correlate strongly with each other.

Genes within an operon were co-transcribed as a single RNA and thus were likely to be under the same transcriptional control. Differences in translational efficiency between genes, localization, feedback, or other post-translational modifications often lead to differences in protein expression of genes in the same operon (Lim et al., 2011a; Mattheakis and Nomura, 1988; Wek et al., 1987; Yamada and Saier Jr, 1988). It has been suggested that there is a correlation between the translational efficiency of a gene and its genetic distance from the end of the operon (Lim et al., 2011a). We expected to see very high correlation between RNAs within an operon as they were under the same transcriptional control; however, we expected there to be less correlation between proteins within an operon as proteins within an operon were not guaranteed to be under the same translational/post-translational regulation.

As a measure of correlation of gene expression within an operon we took the average of the pairwise Spearman correlation coefficient for all possible pairs of transcripts and proteins within an operon. Approximately eighty percent of transcripts had a mean pairwise correlation coefficient greater than 0.8 within an operon (Figure 3A). On the other hand, less than fourteen percent of proteins had a mean pairwise correlation coefficient greater than 0.8 within an operon (Figure 3B). Genes closer together within an operon were more likely to have correlated protein profiles (see Figure 3C), which we took as evidence that distance between genes was a strong indicator of translational regulation. Also shown are a few examples of highly correlated transcripts and proteins for individual operons (Figure 3D, E, and F, respectively).

## Energy-intensive processes are transcriptionally down-regulated while stress-related proteins are up-regulated in response to starvation.

Typical analysis of RNA expression data often involves performing a hierarchical clustering of profiles followed by a term enrichment of subsets of genes found in the emerging patterns. In this approach the patterning that comes from hierarchical clustering can be arbitrary, depending on the level of the hierarchy one chooses to focus on. Here, instead, we sought to sort the time courses into general behaviors in an unbiased manner. To accomplish this goal we fit each individual mRNA and protein to a piecewise continuous curve (Figure 4- supplemental figure 1A). This curve was defined by 4 free time parameters and 3 free amplitude parameters. To fit the curve we used a population based differential evolution (DE) algorithm. Furthermore, the fitness function used in minimization was scaled to the experimental error (see methods). Thus, our algorithm provided confidence intervals for our fit based upon the variability in biological replicates. To demonstrate the effectiveness of our fitting strategy we randomly selected five mRNA profiles and their respective fits (Figure 4-supplemental figure 1B-E). Green circles show the average of three biological replicates with their standard deviations (green bars) and the blue line and bar show the average and standard deviation of the population of fits, respectively. Both the data and fit were normalized to the average of the time course. We also plotted histograms of the time scale parameters we found by fitting the piecewise continuous curve to our data (Figure 4-supplemental figure 2). The most informative time scales were *t*1, the time to first inflection, and *t*2+*t*3+*t*4, the time it takes for the profile to stop changing. The majority of proteins and their transcripts began changing before the 10 hr mark (or just after the cells enter stationary phase). Once the profiles began to change it took >10 hrs before it stopped changing again. However, the apparent long time scale of proteins and transcripts changing could be due to the low time resolution of our experiment after the cells have entered stationary phase.

As can be seen in Figure 4B-E, for mRNAs we generally had good agreement between the data and model. Thus, the fits gave us reasonable estimates of the distribution of time scales involved in the response. Figure 4F shows the distribution of *t*1, the time to first inflection. Most of the mRNAs responded between 3-8 hrs, with a strong peak at around 6 hrs (when cells begin entry to stationary phase). To better understand the regulation of cellular processes (and mRNAs) in our dataset, we sorted the mRNA profiles into five general categories, defined on the basis of our fitted parameters: up-regulated, down-regulated, temporarily up-regulated, temporarily down-regulated, or unknown. The confidence intervals for our fits allowed sorting individual mRNAs into these five categories with high confidence. The mRNAs in the categories “down-regulated” and “up-regulated” showed significant enrichment for GO terms. The average of the mRNAs in each of these terms is shown in Figure 5A, B. Terms enriched in the set of down-regulated transcripts were involved in translation, carboxylic acid biosynthetic process, and nitrogen compound biosynthetic process. These processes were likely down-regulated for energy conservation purposes in the face of limiting resources. Terms enriched in the set of up-regulated transcripts were involved in the carbohydrate catabolic process.

To characterize the protein response we followed the same general strategy of fitting, classification, and GO enrichment as we did for the RNA profiles. The distribution of the time to first inflection for the proteins was a little broader than the mRNAs. However, the first-inflection times still mostly fell into the range of 3-8 hrs, and very few proteins had not responded by the time the cells entered stationary phase. There were many proteins that were turned on for the duration of the time course, compared to the mRNAs where very few remained on for the entire duration of the experiment. Figure 4C shows the average abundance of the proteins in a given GO term that were enriched in the set of proteins that were being up-regulated. As in the case of down-regulated RNAs these proteins were likely down-regulated to conserve energy, and they included proteins involved in translation and locomotion. Up-regulated proteins were, like the up-regulated transcripts, involved in carbohydrate catabolism but also included terms involved in stress response and metabolism of glycerol. The average protein abundances for GO terms being down-regulated had a much wider distribution of decay times compared to the RNAs being down-regulated, likely due to differing protein degradation rates (and/or thermodynamic stability) (Figure 4D).

As a complementary approach we also averaged all proteins in a given Kegg pathway regardless of their behavior. Many pathways showed little to no differential regulation, on average, in their protein levels. Pathways that changed cohesively are plotted in Figure 4E and F depending on whether they were down- or up-regulated, respectively. As in the previous term-enrichment analysis, we saw motility to be down-regulated, as well as other energy consuming processes involved in metabolism and biosynthesis. Interestingly, biosynthesis of siderophores was up-regulated, likely due to do increased demands for, or reduced supply of, iron.

## Metabolic fluxes ratios show possible changes at 2 weeks

We used flux ratio analysis to measure the relative metabolic fluxes passing through different branches of the central metabolism (Zamboni et al., 2009). To measure flux ratios we used the FiatFlux software that fits a metabolic model to the amino acid labeling pattern (Zamboni et al., 2005). As there was little *ab-initio* protein synthesis after the cells stopped growing (after ~8 hrs), we didn’t include the flux ratios after this point. Our major observation was that there was little change in flux ratios throughout growth, and for most of the experiment this initial labeling remained (Figure 1-supplemental figure 2A-I). Interestingly, we observed changes at two weeks in the flux ratio in P5P from G6P lower branch (Figure 1-supplemental figure 2G). Given that there was no net synthesis of amino acids after growth ceased, we cannot use the steady-state approach to interpret these data. They do suggest, however, that either internal amino acid recycling or some de novo synthesis from cross-feeding occurred after 1 week.

## Lipids are modified in response to starvation for up to two weeks

Using negative-ion MALDI-TOF and ESI mass spectrometry (MS), we analyzed lipid A and phospholipid profiles of cells at each time point, respectively. Beginning before one week, we observed an appearance of an MS peak associated with the acylation of lipid A with a C16 chain (Figure 5A, C). In the phospholipid analysis, a notable increase began around 8 hr in the cyclopropanation of one unsaturated double bond within molecules of the major phospholipids, phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). This change was identified by the gradual relative increase of peaks at ~702.5 m/z and ~733.5 m/z, respectively (representative data for PE is shown in Figure 5D). Both the modifications to lipid A and phospholipids continued to increase up to the 2 week time point. In fact, the 702.5 m/z peak corresponding to cyclopropanation of phospholipid was barely detectable before six hours but became the predominant peak by the end of the time course.

The enzymes relevant to the above lipid A and phospholipid modifications are PagP and cycloproponated fatty acid synthase (CFA), respectively (Bishop et al., 2000; Grogan and Cronan, 1997). PagP is known to be constitutively transcribed at low levels and remain latent in the outer membrane until enzyme activation (Jia et al., 2004). It is also up-regulated by the transcriptional regulator, PhoP, under various stressful conditions encountered by a cell (Needham and Trent, 2013). However, during our time course, transcript levels of PagP and PhoP did not change significantly. Furthermore, neither PagP nor PhoP was observed at the protein level. In the case of PagP, this could be due to the difficulty in detecting outer membrane beta-barrel proteins. With respect to phospholipid modification, CFA synthase protein levels increased between 3-6 hrs before decreasing again. This observation agreed with prior data showing that CFA synthase was important for transition to stationary phase of growth (Grogan and Cronan, 1997). CFA synthase RNA levels increased again around 1 week, which was consistent with the activity observed in phospholipid modification, although it is not known why we did not observe a corresponding increase in protein levels at this point (Figure 5B).

# Discussion

Our results provide a coherent picture of *E. coli* starvation, as summarized in Figure 6. *E. coli* could survive for over a week when starved for glucose, with little change in cell viability and optical density at 600nM (Figure 6A). The fraction of mRNA relative to all RNA was down regulated after cells entered stationary phase (Figure 6B). As cells ceased to divide the demand for new protein synthesis was reduced, and the reduction in the overall pool of mRNA could have helped limit new protein synthesis. In response to starvation, lipid A and phospholipids were modified by PagP and CFA synthetase, respectively (Figure 6C). Modification of lipids continued gradually until eventually the modified lipids dominated the lipid population at two weeks. All genes started to change in expression by 10 hrs, and mRNA expression clustered temporally into two regimes, before and after 10 hrs (cells entered stationary phase at around 8 hrs) with some late changes in expression beginning around 2 weeks (Figure 6D). We found that 20% of observed proteins were regulated in proportion to their transcripts (Figure 6F) allowing for rapid down-regulation of the processes they were involved in. On the other hand, 15% of the observed proteins were integrally related to their transcripts (Figure 6F) and likely served to buffer against environmental changes. In addition to measuring and characterizing RNA and protein changes under starvation, we also proposed a novel fitting strategy that allowed us to classify expression profiles into different categories in an unbiased manner. The enriched terms in the resulting classification were reasonably aligned with what was known about, or at least consistent with, cells coping with starvation (Figure 6E). Importantly, this classification was accomplished in an unbiased manner, without any ad-hoc determination of the number of clusters existing in the data.

We found that, as cells entered stationary phase, the total pool of mRNA was depleted compared to all other RNAs and many individual transcripts were down-regulated, possibly as part of a broader strategy to reduce the production of new protein. Reducing overall protein production could also be achieved by limiting the available ribosomes or by limiting the pool of available tRNA. Previous studies have reported a stringent response activated, through ppGpp, and down regulation of new rRNA synthesis , which could help reduce new protein production (Magnusson et al., 2005). In our data, the fraction of relative rRNA changed little while the tRNA fraction increased with time. Thus new protein synthesis may be limited more by the reduced mRNA pool than by reduced translational efficiency.

It has been suggested that the degradation rate of many proteins in *E. coli* is much slower than the division rate of the bacteria (Nath and Koch, 1971, 1970). As a consequence, when cells cease to divide, such as in the case of glucose starvation, not all proteins can respond immediately to possible changes in transcript levels. In effect, some proteins may be buffered to relatively fast changes in nutrient availability. At the same time certain proteins may need to be rapidly regulated to ensure survival upon starvation. We found that a subset of the proteome, ~20% of proteins, fell into the rapidly regulated category and were proportional to their transcripts. Another subset, ~15% of proteins, tended to integrate the response of their transcripts over the time scale of our experiment. For example, several of the flagella proteins were proportional to their transcripts, whereas proteins involved in metabolism and energy production integrated their transcript levels over time. Turning off proteins involved in cell division and the flagellum machinery, both energy intensive processes, needs to happen relatively fast. By contrast, the proteins that were relatively stable were enriched for energy production terms. Thus, these proteins presumably persist so that if nutrients were to become available again the cell can immediately take advantage of them.

We also found that many genes (both among the protein and the RNA profiles) that were significantly up- or down-regulated during starvation were uncharacterized or had no annotation. A subset of these uncharacterized proteins have computationally predicted functions (Hu et al., 2009) that were consistent with our findings for annotated genes. For instance, several uncharacterized proteins that were up-regulated were predicted to be involved in stress response and cell-wall bio-genesis. Other predictions seemed to be inconsistent with our observations for annotated genes. For example, some uncharacterized proteins that were up-regulated were predicted to be involved in translation, even though translation was heavily enriched in down-regulated genes. Lists of proteins and transcripts that were significantly regulated in our time course are provided in the supplemental materials (Figure 5-supplemental tables 1 and 2) and include behavior (e.g. up/down-regulated), curated annotations, and functional predictions (where available).

Even though mRNA abundances within an operon were highly correlated (as expected), in many cases their protein profiles were only weakly correlated. This finding could be due to different translation efficiencies between proteins (Lim et al., 2011) as well as differing degradation rates. In support of the former we saw a tendency for proteins separated by a larger genetic distance to be less correlated than those closer genetically. However, it was likely that different protein degradation rates also played a role in the low correlation between proteins within an operon. Indeed, many proteins that were close genetically showed poor correlation in their profiles (Figure 3C)

In addition to the expected disparities between RNA and protein levels, we also observed surprising changes in enzyme activity that did not correspond to the respective RNAseq and proteomics analysis. For example, we saw striking levels of lipid modification late during the time course. These modifications were easily explained by their association with adaptation to stressful environments such as depleted nutrients and cations as well as increase acid resistance during starvation (Grogan and Cronan, 1997; Needham and Trent, 2013). However, the stark differences in RNA, protein, and activity trends of the enzymes responsible for the lipid modifications, PagP and CFA synthase, highlights the fact that activation does not necessarily follow abundance measurements.

Metabolic fluxes stayed surprisingly constant throughout the experiment, with the exception of the two-week time point, where the flux ratio corresponding to P5P from G6P lower branch declined. This change so late in the experiment was unexpected, since we did not anticipate substantial turnover in cellular composition that late into the starvation state. The observation suggests that either internal amino acid recycling or some de novo amino-acid synthesis from cross-feeding occurs past the 1 week time point.

A goal of systems biology has been to understand how phenotype originates from genotype. The phenotype of a cell is determined by complex regulation of cell signaling, gene regulation, metabolism, and lipid biochemistry. Understanding the connection between phenotype and genotype is crucial to understanding disease and for synthetic engineering of biology. Computational models have been particularly well suited to studying this problem, as they can synthesize and organize diverse and complex data in a predictive framework. Even though computational models of individual component subsystems, such as flux models of metabolism (Duarte et al., 2004; Feist et al., 2007; Feist and Palsson, 2008), have enjoyed a long history of success, they remain limited in their application. Much effort is currently being spent on understanding how to best integrate multiple subsystems. For example, there are many proposed approaches to combining gene expression with metabolic flux networks (Åkesson et al., 2004; Colijn et al., 2009; Collins et al., 2012; Fang et al., 2012; Lee et al., 2012; Mahadevan et al., 2002; O’Brien et al., 2013; Van Berlo et al., 2011; Vogel and Marcotte, 2012) while other studies focused on integrative, whole-cell, models (Carrera et al., 2014; Karr et al., 2012). Given the growing interest in integrative modeling approaches, there is a pressing need for studies that collect high quality genome-wide data across multiple cellular subsystems from the same biological samples. Our data set is a rich resource for comparing and contrasting the response of multiple cellular subsystems. Additionally, in the future we plan to use the techniques developed in this paper to measure the response of *E. coli* to several other environmental conditions, which will allow for more detailed models of regulation.

Despite the completeness and quality of our data set, however, there were a few key limitations concerning our approach. Our analysis via RNAseq and shotgun MS allowed for high confidence when comparing the relative levels of a particular transcript or protein over time. However, due to potential differences in detection efficiency between individual RNAs or peptides, care should be taken when comparing absolute abundances. In our analysis we used the APEX method to account for differences in detection efficiency. We normalized RNA by the length of a transcript as an estimate of RNA detection efficiency, for a particular experiment. This resulted in a correlation coefficient of ~0.71 between proteins and their transcripts, a finding on the high end for such correlation measurements. Thus, even straightforward means of correcting our experimental bias led to reasonable comparisons of levels between individual RNAs or proteins.

The stepwise linear function we used for modeling works for a majority of our expression profiles. However, in some cases it over-fits the data and in other cases the function was unable to capture the underlying behavior. An example of a profile that may be under-constrained is a gene that is up- or down-regulated without further changes to expression. In this case the free time parameters *t*2 through *t*4, along with amplitude parameter *A*2, may be under-constrained. Even in this case, however, the parameters *t*1, *t*2+*t*3+*t*4, *A*1, and *A*3 were still well constrained, providing enough information to reliably sort the profiles based upon behavior. More complicated behavior, such as multiple peaks separated in time, cannot be captured by our function. The presence of these more complicated behaviors was rare enough as to not warrant special consideration.

Our study is the most complete measurement, to our knowledge, of multiple cellular components in a changing environment (of which there are few). This work represents an important step toward understanding how regulation of a cells physiology is coordinated by multiple cellular subsystems.

# Methods

## Cell Growth

*E. coli* B REL606 was inoculated from freezer stock in 50mL DM500 and incubated at 37°C overnight. 500 uL of the overnight culture was diluted in 50mL of DM500 at 37°C and grown for 24 hrs. On the day of the experiment, 500 μL of the 24 hr culture was added to 10 flasks containing 50mL Davis-Mingioli (DM) minimal medium supplemented with limiting glucose at 0.5 g/L (DM500) each, grown at 37C. Each biological replicate was performed on separate days. At each time point 1ml was removed, washed with 0.7% NaCl, spun down, the supernatant was removed, and the remaining cell pellet was flash frozen using liquid nitrogen and stored at –80°C. Samples for each experiment where taken from the same batch of culture except for the samples used for flux analysis which was grown separately in 13C labeled glucose.

To measure colony-forming units (CFU), the OD600 at each time point was taken relative to sterile DM500 glucose, cultures were diluted in sterile saline, and finally plated on DM agar supplemented with 0.2g/L glucose. Colonies were counted after incubation at 37°C for 24 hr. Cultures for measuring CFU were grown separately from the main culture but in identical conditions.

## RNAseq

Total RNA was isolated from cell pellets using the RNAsnap method. After extraction, RNA was ethanol precipitated and resuspended in 100ul ultra pure water. The RNA was then cleaned and DNase treated using Zymo Clean and Concentrator 25 (Zymo Research). RNA was then eluted in 50ul ultra pure water. Total RNA concentration was determined using Qubit 2.0 Fluorometer (Life Technologies). Ribosomal RNA was depleted using Gram-negative bacteria RiboZero rRNA removal kit (Epicentre). After depletion the RNA was ethanol precipitated and resuspended in 20ul ultra-pure water. RNA concentration was determined using Qubit 2.0 Fluorometer (Life Technologies). RNA was then fragmented to approximately 250 bp using NEBNext Magnesium RNA Fragmentation Module (New England Biolabs). After fragmentation the RNA was ethanol precipitated and resuspended in 20ul ultra-pure water. After fragmentation, the RNA was Kinase treated using T4 PNK (New England Biolabs). The RNA was ethanol precipitated and resuspended in 20ul ultra-pure water. Library preparation was done using NEBNext Small RNA Library Prep Set for Illumina, Multiplex Compatible (New England Biolabs). Each sample was multiplexed using NEBNext Multiplex Oligos for Illumina (New England Biolabs). After library preparation samples were ethanol precipitated and then run on a 4% agarose gel. All DNA greater than 100 bp was excised and cleaned using Zymoclean Gel DNA Recovery kit (Zymo Research). Final library concentration was determined using Qubit 2.0 Fluorometer (Life Technologies). Libraries were sequenced using an Illumina HiSeq 2500 at the Genomic Sequencing and Analysis Facility (GSAF) at the University of Texas at Austin.

For RNA-seq analysis, we used our own custom analysis pipeline using the REL606 *Escherichia coli* B strain genome as the reference sequence (Jeong et al., 2009). Prior to mapping, all of the single-end reads have had the adapters trimmed using FLEXBAR 2.31 (Dodt et al., 2012). Mapping was carried out using bowtie2 2.1.0 with the –k 1 option to achieve one unique mapping location per read (Langmead and Salzberg, 2012). Raw number of reads mapping per gene were counted using htseq-count 0.6.0 (Anders et al., 2014). The exact pipeline is available at <https://github.com/clauswilke/Ecoli_RNAseq>.

## Proteomics

E. coli cell pellets were resuspended in 50mM Tris-HCl pH 8.0, 10mM DTT. 2,2,2-trifluoroethanol (Sigma) was added to 50% (v/v) final concentration and samples were incubated at 56°C for 45 min. Following incubation, iodoacetamide was added to a concentration of 25mM and samples were incubated at room temperature in the dark for 30 min. Samples were diluted 10-fold with 2mM CaCl2, 50mM Tris-HCl, pH 8.0. Samples were digested with trypsin (Pierce) at 37°C for 5 hr. Digestion was quenched by adding formic acid to 1% concentration. Tryptic peptides were filtered through Amicon Ultra 30kD spin filtration columns and bound, washed, and eluted from HyperSep C18 SpinTips (Thermo Scientific). Eluted peptides were dried by speed-vac and resuspended in Buffer C (5% acetonitrile, 0.1% formic acid) for analysis by LC-MS/MS.

For LC-MS/MS analysis, peptides were subjected to separation by C18 reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo Scientific). Peptides were loaded onto an Acclaim C18 PepMap RSLC column (Dionex; Thermo Scientific) and eluted using a 5-40% acetonitrile gradient over 250 min at 300nl/min flow rate. Eluted peptides were directly injected into an Orbitrap Elite mass spectrometer (Thermo Scientific) by nano-electrospray and subject to data-dependent tandem mass spectrometry, with full precursor ion scans (MS1) collected at 60,0000 resolution. Monoisotopic precursor selection and charge-state screening were enabled, with ions of charge >+1 selected for collision-induced dissociation (CID). Up to 20 fragmentation scans (MS2) were collected per MS1. Dynamic exclusion was active with 45 s exclusion for ions selected twice within a 30 s window.

Spectra were searched against an *Escherichia coli* strain REL606 protein sequence database and common contaminant proteins (MaxQuant using SEQUEST (Proteome Discoverer 1.4; Thermo Scientific). Fully-tryptic peptides were considered, with up to two missed cleavages. Tolerances of 10 ppm (MS1) and 0.5 Da (MS2), carbamidomethylation of cysteine as static modification, and oxidized methionine as dynamic modification were used. High-confidence peptide-spectral matches (PSMs) were filtered at <1% false discovery rate determined by Percolator (Proteome Discoverer 1.4; Thermo Scientific).

## Flux analysis

Flux ratios were obtained from the samples grown with 13C labeled glucose, using the methods of (Zamboni et al., 2009). Cell pellets were resuspended in 200 mL of 6 N HCl, hydrolyzed at 105 °C overnight, and dried at 95 °C for up to 24 hours. To the hydrolyzed cell material we added 40 mL of dimethylformamide (DMF) and gently mixed until a “light straw” color was obtained. The DMF resuspension was transferred to a GC-MS vial with plastic insert and 40 mL of N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide with 1% tert-butyldimethyl-chlorosilane; vials were capped and baked at 85°C for 2 hours, and samples were analyzed within 2 days of derivitization.

Analysis of derivitized samples was performed on a Shimadzu QP2010 Plus GC-MS (Columbia, MD) with autosampler. The GC-MS protocol included: 1 mL of sample injected with 1:10 split mode at 230°C; an oven gradient of 160°C for 1 min, ramp to 310°C at 20°C/min, and hold at 310°C for 0.5 min; and flow rate was 1 mL/min in helium. A total of five runs were performed for each sample: a blank injection of DMF to waste, a blank injection of DMF to the column, and three technical replicates of each vial.

Flux inference was performed using the fiat-flux software as described (Zamboni et al., 2009, 2005).

## Lipid analysis

Lipid A and phospholipids were isolated from bacterial pellets containing 3-9 x 109 cells. Pellets were resuspended in 5ml 1:2:08 CHCl3: MeOH: H2O for 20 minutes and spun at 10,000 xg for 10 minutes. Pellets containing lipid A were further purified by the Bligh/Dyer method as previously described (Hankins et al., 2013). Phospholipids in the supernatant were further purified by extractions as previously described (Giles et al., 2011). Mass analysis of purified lipid A fractions was performed using a MALDI-TOF/TOF (ABI 4700 Proteomics Analyzer) mass spectrometer in the negative ion linear mode as previously described (Hankins et al., 2013). Phospholipid analysis was performed by liquid chromatography/ESI-mass spectrometry as previously described (Giles et al 2011).

## Expression profile data analysis

We analyzed raw counts from the proteomics and RNAseq experiments as follows. Initially, proteins with low counts (<10) over the entire duration of the time course where filtered out. Each time point was then normalized to the read depth (e.g. the sum of all counts for that particular time point). Only proteins with a fold change of >=1.5 where considered for further analysis. Protein profiles where then normalized to the maximum value for a given protein time course. To estimate the absolute protein abundance we made use of the APEX normalization method (Vogel and Marcotte, 2008).

Raw RNA counts where normalized using the DeSeq method (Anders and Huber, 2010). RNAs with a *p*-value less than 0.05, at some point during the time course, where considered to have changed significantly and kept for further analysis. To compare RNA abundances within a single time point, raw RNA counts where normalized to their read depth (sum of all counts for that sample). Finally, normalized RNA and protein profiles, both relative and absolute, where averaged across all three biological replicates.

Clustering of protein profiles was performed using the python library scipy (Jones et al., 2001). We used the *k*-means clustering algorithm with the number of protein clusters set to 25 and RNA clusters to 15. To compare relative protein profiles with the integral of their relative transcript levels we integrated each of the transcript profiles, from the initial time to each additional time point, using the trapezoidal method implemented by the python library numpy (Walt et al., 2011).

We used a piecewise continuous curve to fit both RNA and protein profiles. This curve was defined by seven free parameters, 4 free time parameters, and 3 free amplitude parameters. To fit the profiles we used a custom implementation of a differential evolution (DE) algorithm (Price et al., 2005). Briefly, the DE algorithm initially generates an ensemble of random parameter guesses within a predefined range; subsequently, vectors of individual parameter sets (sometimes called agents) are mixed together at a predefined crossover rate, only those crossover events that yield a smaller error (defined by a predefined cost function) are kept, and the process was iterated until a convergence criterion was met. In our fits we used an ensemble of 15 agents with a crossover frequency of 0.75 and a mixing strength of 0.6. The crossover frequency determines the probability that an agent will be changed at any given iteration and the mixing strength determines how large a change an agent undergoes if it was chosen to be altered. The crossover frequency and mixing strength where picked based upon an empirical study of the dependence of convergence efficiency on these parameters (Pedersen, Magnus Erik Hvass, 2010) for some standard optimization problems. The cost function is given by



where *di*(*tj*), *σi*(*tj*), and *si*(*tj*) are the average of all experimental repeats of protein (or mRNA) *i* at time *tj*, the standard deviation of the experiments of the protein (or mRNA) *i* at time *tj*, and the average of the ensemble simulations *i* at time *tj*, respectively. Scaling by the standard deviation places a relatively lower weight on data points with relatively larger errors for a given protein or mRNA.

Some of the profiles may be slightly over fit by our curve (e.g. profiles that are up-regulated or down-regulated once during the time course without further modulation of expression). Thus care needs to be exercised in the interpretation of some of the parameters. However, we found *t*1 to reliably represent the time to first inflection, the sum of *t*2, *t*3,and*t*4 was a decent proxy to how long it took an RNA/protein to reach a steady state after entering stationary phase, and we could reliably sort the behavior into four categories based upon the amplitude parameters. The four categories we used where that of up-regulated, down-regulated, temporarily up-regulated or temporarily down-regulated. Genes that were up (or down) regulated where those genes that increased (or decreased) at some point during the time course and did not decrease (or increase) at some later time. Genes that were temporarily up (or down) regulated where those genes that increased (or decreased) at some point during the time course but decreased (or increased) at some later time. The sorting into categories was aided by our estimate of the distribution of parameters that allow for a good fit within the population of fits. A fit was considered good if it was on average (across the time course) one standard deviation, or less, away from the experimental average.

We used the DAVID database (david.abcc.ncifcrf.gov) to perform Gene Ontology term enrichment on each subset of sorted genes: up-regulated, down-regulated, temporarily up-regulated or temporarily down-regulated. Specifically we made use of DAVID's API, instead of the web interface, to generate the GO-enrichment through a python script. GO terms where clustered based upon genes in a given term to reduce redundancy in the returned results.

As a complementary approach we also enriched, for Kegg pathway terms, the entire set of significantly changing proteins (without presorting) using, again, the DAVID database API. The proteins within each returned Kegg pathway where then averaged to see if there was any consistent response across the entire pathway. Those Kegg terms that gave in-cohesive responses across proteins in that pathway returned a relatively flat average and where filtered out.

All of the scripts used to perform the above analysis can be downloaded at <https://github.com/marcottelab/AG3C>.

# References

Abreu, R. de S., Penalva, L.O., Marcotte, E.M., Vogel, C., 2009. Global signatures of protein and mRNA expression levels. Mol. Biosyst. 5, 1512–1526. doi:10.1039/B908315D

Åkesson, M., Förster, J., Nielsen, J., 2004. Integration of gene expression data into genome-scale metabolic models. Metab. Eng. 6, 285–293. doi:10.1016/j.ymben.2003.12.002

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106. doi:10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P.T., Huber, W., 2014. HTSeq - A Python framework to work with high-throughput sequencing data (No. 002824).

Bishop, R.E., Gibbons, H.S., Guina, T., Trent, M.S., Miller, S.I., Raetz, C.R.H., 2000. Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. EMBO J. 19, 5071–5080. doi:10.1093/emboj/19.19.5071

Carrera, J., Estrela, R., Luo, J., Rai, N., Tsoukalas, A., Tagkopoulos, I., 2014. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of Escherichia coli. Mol. Syst. Biol. 10, n/a–n/a. doi:10.15252/msb.20145108

Colijn, C., Brandes, A., Zucker, J., Lun, D.S., Weiner, B., Farhat, M.R., Cheng, T.-Y., Moody, D.B., Murray, M., Galagan, J.E., 2009. Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production. PLoS Comput Biol 5, e1000489. doi:10.1371/journal.pcbi.1000489

Collins, S.B., Reznik, E., Segrè, D., 2012. Temporal Expression-based Analysis of Metabolism. PLoS Comput Biol 8, e1002781. doi:10.1371/journal.pcbi.1002781

Dodt, M., Roehr, J.T., Ahmed, R., Dieterich, C., 2012. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology 1, 895–905. doi:10.3390/biology1030895

Duarte, N.C., Herrgård, M.J., Palsson, B.Ø., 2004. Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. Genome Res. 14, 1298–1309. doi:10.1101/gr.2250904

Fang, X., Wallqvist, A., Reifman, J., 2012. Modeling Phenotypic Metabolic Adaptations of Mycobacterium tuberculosis H37Rv under Hypoxia. PLoS Comput Biol 8, e1002688. doi:10.1371/journal.pcbi.1002688

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, n/a–n/a. doi:10.1038/msb4100155

Feist, A.M., Palsson, B.Ø., 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat. Biotechnol. 26, 659–667. doi:10.1038/nbt1401

Finkel, S.E., Kolter, R., 1999. Evolution of microbial diversity during prolonged starvation. Proc. Natl. Acad. Sci. 96, 4023–4027. doi:10.1073/pnas.96.7.4023

Giles, D.K., Hankins, J.V., Guan, Z., Trent, M.S., 2011. Remodelling of the Vibrio cholerae membrane by incorporation of exogenous fatty acids from host and aquatic environments. Mol. Microbiol. 79, 716–728. doi:10.1111/j.1365-2958.2010.07476.x

Grogan, D.W., Cronan, J.E., 1997. Cyclopropane ring formation in membrane lipids of bacteria. Microbiol. Mol. Biol. Rev. 61, 429–441.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R., 1999. Correlation between Protein and mRNA Abundance in Yeast. Mol. Cell. Biol. 19, 1720–1730.

Hankins, J.V., Madsen, J.A., Needham, B.D., Brodbelt, J.S., Trent, M.S., 2013. The Outer Membrane of Gram-Negative Bacteria: Lipid A Isolation and Characterization, in: Delcour, A.H. (Ed.), Bacterial Cell Surfaces, Methods in Molecular Biology. Humana Press, pp. 239–258.

Hu, P., Janga, S.C., Babu, M., Díaz-Mejía, J.J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulos, C., Nazarians-Armavil, A., Nasseri, N.K., Musso, G., Ali, M., Nazemof, N., Eroukova, V., Golshani, A., Paccanaro, A., Greenblatt, J.F., Moreno-Hagelsieb, G., Emili, A., 2009. Global Functional Atlas of Escherichia coli Encompassing Previously Uncharacterized Proteins. PLoS Biol 7, e1000096. doi:10.1371/journal.pbio.1000096

Jeong, H., Barbe, V., Lee, C.H., Vallenet, D., Yu, D.S., Choi, S.-H., Couloux, A., Lee, S.-W., Yoon, S.H., Cattolico, L., Hur, C.-G., Park, H.-S., Ségurens, B., Kim, S.C., Oh, T.K., Lenski, R.E., Studier, F.W., Daegelen, P., Kim, J.F., 2009. Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3). J. Mol. Biol. 394, 644–652. doi:10.1016/j.jmb.2009.09.052

Jia, W., Zoeiby, A.E., Petruzziello, T.N., Jayabalasingham, B., Seyedirashti, S., Bishop, R.E., 2004. Lipid Trafficking Controls Endotoxin Acylation in Outer Membranes of Escherichia coli. J. Biol. Chem. 279, 44966–44975. doi:10.1074/jbc.M404963200

Jones, E., Oliphant, T., Peterson, P., 2001. SciPy: Open source scientific tools for Python [WWW Document]. http://www.scipy.org/. URL http://www.scipy.org/Citing\_SciPy (accessed 9.9.14).

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival Jr., B., Assad-Garcia, N., Glass, J.I., Covert, M.W., 2012. A Whole-Cell Computational Model Predicts Phenotype from Genotype. Cell 150, 389–401. doi:10.1016/j.cell.2012.05.044

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi:10.1038/nmeth.1923

Lee, D., Smallbone, K., Dunn, W.B., Murabito, E., Winder, C.L., Kell, D.B., Mendes, P., Swainston, N., 2012. Improving metabolic flux predictions using absolute gene expression data. BMC Syst. Biol. 6, 73. doi:10.1186/1752-0509-6-73

Lewis, N.E., Cho, B.-K., Knight, E.M., Palsson, B.O., 2009. Gene expression profiling and the use of genome-scale in silico models of Escherichia coli for analysis: providing context for content. J. Bacteriol. 191, 3437–3444. doi:10.1128/JB.00034-09

Lewis, N.E., Hixson, K.K., Conrad, T.M., Lerman, J.A., Charusanti, P., Polpitiya, A.D., Adkins, J.N., Schramm, G., Purvine, S.O., Lopez-Ferrer, D., Weitz, K.K., Eils, R., König, R., Smith, R.D., Palsson, B.Ø., 2010. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6, n/a–n/a. doi:10.1038/msb.2010.47

Lim, H.N., Lee, Y., Hussein, R., 2011a. Fundamental relationship between operon organization and gene expression. Proc. Natl. Acad. Sci. 108, 10626–10631. doi:10.1073/pnas.1105692108

Lim, H.N., Lee, Y., Hussein, R., 2011b. Fundamental relationship between operon organization and gene expression. Proc. Natl. Acad. Sci. 108, 10626–10631. doi:10.1073/pnas.1105692108

Lu, P., Vogel, C., Wang, R., Yao, X., Marcotte, E.M., 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat. Biotechnol. 25, 117–124. doi:10.1038/nbt1270

Magnusson, L.U., Farewell, A., Nyström, T., 2005. ppGpp: a global regulator in Escherichia coli. Trends Microbiol. 13, 236–242. doi:10.1016/j.tim.2005.03.008

Mahadevan, R., Edwards, J.S., Doyle III, F.J., 2002. Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. Biophys. J. 83, 1331–1340. doi:10.1016/S0006-3495(02)73903-9

Mattheakis, L.C., Nomura, M., 1988. Feedback regulation of the spc operon in Escherichia coli: translational coupling and mRNA processing. J. Bacteriol. 170, 4484–4492.

Morita, R.Y., 1990. The starvation-survival state of microorganisms in nature and its relationship to the bioavailable energy. Experientia 46, 813–817. doi:10.1007/BF01935530

Nath, K., Koch, A.L., 1970. Protein degradation in Escherichia coli. I. Measurement of rapidly and slowly decaying components. J. Biol. Chem. 245, 2889–2900.

Nath, K., Koch, A.L., 1971. Protein degradation in Escherichia coli. II. Strain differences in the degradation of protein and nucleic acid resulting from starvation. J. Biol. Chem. 246, 6956–6967.

Needham, B.D., Trent, M.S., 2013. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. Nat. Rev. Microbiol. 11, 467–481. doi:10.1038/nrmicro3047

Neidhardt, F.C., Curtiss, R., 1996. Escherichia Coli and Salmonella: Cellular and Molecular Biology. ASM Press.

O’Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, n/a–n/a. doi:10.1038/msb.2013.52

Pedersen, Magnus Erik Hvass, n.d. Good parameters for differential evolution.

Price, K., Storn, R.M., Lampinen, J.A., 2005. Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Raghavan, R., Groisman, E.A., Ochman, H., 2011. Genome-wide detection of novel regulatory RNAs in E. coli. Genome Res. 21, 1487–1497. doi:10.1101/gr.119370.110

Soares, N.C., Spät, P., Krug, K., Macek, B., 2013. Global Dynamics of the Escherichia coli Proteome and Phosphoproteome During Growth in Minimal Medium. J. Proteome Res. 12, 2611–2621. doi:10.1021/pr3011843

Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., Kushner, S.R., 2012. RNAsnapTM: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. Nucleic Acids Res. 40, e156. doi:10.1093/nar/gks680

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Science 329, 533–538. doi:10.1126/science.1188308

Van Berlo, R.J.P., de Ridder, D., Daran, J.-M., Daran-Lapujade, P.A., Teusink, B., Reinders, M.J.T., 2011. Predicting Metabolic Fluxes Using Gene Expression Differences As Constraints. IEEE/ACM Trans. Comput. Biol. Bioinform. 8, 206–216. doi:10.1109/TCBB.2009.55

Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.-Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., Penalva, L.O., 2010. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. 6, n/a–n/a. doi:10.1038/msb.2010.59

Vogel, C., Marcotte, E.M., 2008. Calculating absolute and relative protein abundance from mass spectrometry-based protein expression data. Nat. Protoc. 3, 1444–1451. doi:10.1038/nprot.2008.132

Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227–232. doi:10.1038/nrg3185

Walt, S. van der, Colbert, S.C., Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Comput. Sci. Eng. 13, 22–30. doi:10.1109/MCSE.2011.37

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. doi:10.1038/nrg2484

Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzeler, E., Yates, J.R., 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 100, 3107–3112. doi:10.1073/pnas.0634629100

Wek, R.C., Sameshima, J.H., Hatfield, G.W., 1987. Rho-dependent transcriptional polarity in the ilvGMEDA operon of wild-type Escherichia coli K12. J. Biol. Chem. 262, 15256–15261.

Wiśniewski, J.R., Rakus, D., n.d. Quantitative analysis of the Escherichia coli proteome. Data Brief. doi:10.1016/j.dib.2014.08.004

Yamada, M., Saier Jr, M.H., 1988. Positive and negative regulators for glucitol (gut) operon expression in Escherichia coli. J. Mol. Biol. 203, 569–583. doi:10.1016/0022-2836(88)90193-3

Yoon, S.H., Han, M.-J., Jeong, H., Lee, C.H., Xia, X.-X., Lee, D.-H., Shim, J.H., Lee, S.Y., Oh, T.K., Kim, J.F., 2012. Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. Genome Biol. 13, R37. doi:10.1186/gb-2012-13-5-r37

Zamboni, N., Fendt, S.-M., Rühl, M., Sauer, U., 2009. (13)C-based metabolic flux analysis. Nat. Protoc. 4, 878–892. doi:10.1038/nprot.2009.58

Zamboni, N., Fischer, E., Sauer, U., 2005. FiatFlux--a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics 6, 209. doi:10.1186/1471-2105-6-209

## Table 1- Comparison of data set completeness

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Publication** | **# Proteins** | **Cutoff** | **Technique** | **# RNAs** | **Cutoff** | **Technique** |
| This study | 2648 (at 3 hrs) | FDR < 1% | Shotgun MS | 4116 genes, 89 ncRNA, 85 tRNA (at 3 hrs) | Alignment quality>10 | RNA-seq |
| Yoon 2012 | 60 | fold change > 2 | 2-D gel | 4144 gene probes (REL606 strain) | NA | Microarray |
| Taniguchi 2012 | 1018 | NA | YFP fusion | 137 | 95% confidence | FISH |
| Lewis 2010 | ~1,000 | P-value  < 0.05 | Shotgun MS | - |  |  |
| Lewis 2009 |  |  |  | 4428 gene probes | NA | Microarray |
| Soares 2013 | 2118 | FDR < 1% | SILAC | - |  |  |
| Raghavan 2011 |  |  |  | Genes 4161,133 sRNA | %region mapped  >= 50% | RNA-seq |

# Figure Captions

**Figure 1.** **Overview of experimental design.**

Measurements of RNA, protein, lipids, and metabolic flux were taken under uniform growth and environmental conditions. (A) Long-term glucose starvation experiment. The *E. coli* B REL606 strain was taken from freezer stock and revived in 10ml DM500 media with minimal glucose for 24 hrs (day -2). The next day this culture was diluted 1:10 in 500mL DM500 and grown for 24 hrs (day -1). On the day of the experiment (day 0) the previously conditioned culture was diluted again into several individual cultures. (B) The OD @ 600 (blue curve) was measured to assess growth and optimal collection of time points. Nine time points where selected for this experiment spanning 3 hrs – 2 wks. A cell viability assay was performed at each time point to determine the number of colony forming units (CFU, purple curve). (C) For each sample 1ml was removed from the culture for each experiment to be done, spun down, flash frozen, and distributed to the experimental labs in our consortium to measure RNA via RNA seq, protein via LC/MS, lipids via Maldi-TOF MS and ESI MS, and metabolic flux via GC-MS. Metabolic flux samples where grown separately under identical conditions excepting the labeled U-13C glucose. Raw RNA and protein counts, calculated flux ratios, raw phospholipid MS peaks, and LipidA peaks for all time points are available in Figure 1-supplemental file 1.

**Figure 2. *K*-means clustering of RNA and protein profiles in long-term glucose starvation reveal trends in transcriptional and post-transcriptional regulation**

RNA and Protein profiles, normalized to each molecule’s maximum value, where clustered by geometric distance using *K*-means clustering with 15 and 25 clusters, respectively. The cluster centroids where then plotted as heat maps with darker blue representing higher RNA or protein levels. (A) RNA levels are largely shut off upon glucose starvation with some RNA being temporarily up-regulated during the transition between exponential and stationary growth. (B) Protein levels showed a much wider range of behaviors with some being up- or down-regulated for the duration of the experiment as well as for a short period of time during the transition from exponential to stationary growth. (C) Histogram of the correlation coefficients between individual protein levels and the time integral of RNA. This was the limit of slow protein degradation where protein levels are proportional to the cumulative sum of their respective transcripts. (D) Histogram of the correlation coefficient between relative protein levels and their corresponding (relative) transcripts. This was a measure of how proportional proteins are to their respective RNAs. (E) 2-D histogram of the correlation coefficient of protein vs. RNA (on the x-axis) and the protein vs. the time integral of RNA (y-axis). Darker colors indicate more genes in that given bin. There is a strong anti-correlation between the two measures of dynamic correlation indicating that these two quantities are largely mutually exclusive. (F) The correlation between all RNA and protein levels for a single time point was strongest at 3 hrs (Spearman correlation coefficient ~0.71). Normalized RNA and protein levels, both relative and absolute, used to generate the above figure are provided in Figure 2-supplemenal file 1.

**Figure 3. RNAs within an operon correlate strongly whereas proteins, generally, do not.**

(A) and (B) histograms of the median pairwise correlation coefficient between all possible pairs of RNA and Protein profiles, respectively, within an operon. Proteins that have a smaller inter-gene distance are more likely to have correlated profiles (C). 2D Histogram of the pairwise correlation between proteins in the same operon, y-axis, and the inter-gene distance between the protein coding regions, x-axis. Darker colors represent higher correlation. (D) Example of RNAs in the same operon that are highly correlated. (E, F) Examples of proteins in the same operon that are highly correlated.

**Figure 4. Flagella genes and other energy intensive processes are down-regulated while stress-response genes are up-regulated.**

Fitting the mRNA and protein profiles allowed us to estimate the underlying dynamics and differential regulation of each gene, sorting them into high confidence categories describing their behavior. Genes were put into categories based upon whether they where up-regulated, down-regulated, temporarily up-regulated, or temporarily down-regulated. The mRNA or proteins in each category were then tested for enrichment of GO terms. (A, B) The average of the mRNAs in a given enriched term for mRNAs that are down- and up-regulated, respectively. (C, D) The average of the proteins in a given enriched term for proteins that are down- and up-regulated, respectively. There was no significant enrichment for either proteins or mRNAs that where temporarily up- or down-regulated. As a complementary approach we took the average of all proteins in a given Kegg pathway annotation. (E, F) The average of proteins in the Kegg pathway that are changing significantly. All the other terms showed no significant change.

**Figure 5. Lipid A and phospholipids are modified starting at 8 hrs and dominate the lipid pool by two weeks.**

Lipid A and phospholipids were extracted from all samples for analysis by negative ion MALDI-TOF and ESI-MS mass spectrometry, respectively, and the 6 hour and 2 week representative samples are shown in this figure. (A) Activation of the acyl-transferase PagP adds a C16 chain to lipid A on the 2-position primary acyl chain, resulting in a m/z of ~2035. (B) Modification of phospholipids by cyclopropanation of one unsaturated double bond is catalyzed by CFA synthase. Transcripts of CFA synthase increase at late times (green) consistent with modification of PE where as the level of CFA protein stays relatively flat at late time points. (C) As represented here by the 6 hr sample, lipid A from all samples collected between 3 and 48 hrs contained one major peak at ~1797 m/z corresponding to wild type, hexa-acylated lipid A. As illustrated on the right by the 2 week sample, the 1 and 2 week (168 and 336 hr) time points show the addition of the C16 chain to lipid A. (D) Phosphatidylethanolamine (PE) is shown here as a representation of similar results obtained in phosphatidylglycerol (PG). The phospholipid profiles of the samples remained relatively consistent with wild type *E. coli* phospholipid profiles until hour 8, when a gradual increase in a peak ~702.5 m/z began. This peak became the predominant species by 2 weeks. Its mass corresponds to the cyclopropanation of one unsaturated double bond within a PE molecule containing acyl chains totaling 33 carbons distributed between the two acyl chains.

**Figure 6. Summary of key results.**

(A) Cells enter stationary phase around  6-8h after which the cell viability (and OD) remain relatively constant up until 2 weeks where there was a decrease of %38 in viability. (B) The relative fraction of rRNA (compared to all RNA) stays fairly constant through the entire time course as does the fraction of tRNAs (in non-rRNA depleted samples). On the other hand, relative levels of mRNA decrease upon entry into stationary phase perhaps as a strategy for reducing overall protein synthesis. As a reference we also showed RNA fractions for rRNA samples that demonstrate reduction of mRNA was not due to low relative counts of mRNA compared to rRNA. (C) Phospholipids and LipidA are modified in a manner consistent with stress response. Modifications begin early in stationary phase and slowly increase during the time course until, at 2 weeks, the modified lipids dominate the lipid population. (D) Transcriptional changes (measured by mRNAs) separate into at least two temporal regions before and after entry to stationary phase with a possible third region corresponding to late transcriptional regulation at 2 weeks. All changes in regulation have begun by 10 h where stress response genes are up-regulated and energy intensive processes are down-regulated.  (E) Approximately 20% of the measured proteins are proportional to their transcript levels over time while 15% of the proteins integrate their transcript’s response over the entire duration of the experiment. This observation highlights that differences in post-transcriptional regulation, such as protein degradation, cause differences in regulation between mRNAs and their expressed proteins.

**Figure 1 - supplemental figure 1.** **RNAseq and MS experiments are highly reproducible.**

Scatter plot between biological replicates 1 and 2 (A, C) and 1 and 3 (B, D) along with their associated Spearman correlation coefficients. P-value for all correlations are <10-100.

**Figure 1 –supplemental figure 2. Flux ratio profiles in long-term glucose growth.**

Flux ratios where computed using fiat flux software from GC-MS derived 13C constraints. Only one independent flux ratio showed a significant change over the course of the experiment. Flux ratios for (A) SER from GLY, (B) OYR from MAL upper branch, (C) PEP through TK upper branch, (D) PEP through PPP upper branch, (E) PEP from OAA, (F) OAA from PEP, (G) P5P from G6P lower branch, (H) E4P through TK, and (I) GLY through serine. P5P from G6P lower branch has a large correlation (~0.85) with the corresponding ratio of proteins that catalyze those reactions while the corresponding mRNA ratio shows a much lower correlation (~0.5) with the flux ratios.

**Figure 2- supplemental figure 1. The mRNA fraction, compared to all other RNA, is strongly down-regulated in response to starvation.**

For each time point the fraction of total RNA that was either non-coding (purple), tRNA (orange), rRNA (green), or mRNA (red). (A) RNA fractions for non-ribosomal RNA depleted samples. (B) RNA fractions for ribosomal depleted fractions. Each bar represents an individual biological repeat.

**Figure 2 - supplemental figure 2. Correlations between the absolute abundance between protein and mRNA is strongest at 3 hrs.**

Both the RNA and protein levels are scaled to their respective averages across all RNAs or proteins for each time point and then log transformed. All P values are <10-43.

**Figure 4 – supplemental figure 1. Fitting of piecewise continuous curve is effective when sorting response curves**

We grouped RNA and protein time courses based on general qualitative behaviors. In response to glucose starvation RNA can be shut off, turned on, pulsed up, and pulsed down. (A) To sort the profiles, a piecewise continuous curve was fit to the data. The parameter *t*0 represents the time at which we start to collect data at 3 hrs into growth. The curve was fit using a differential evolution fitting algorithm that was gradient free and population based, allowing for a range of possible parameter sets that can explain our data given the experimental error. (B-E) Four random examples of measured RNA time courses averaged across 3 biological replicates (green circles) with their standard deviations (green bars) along with the corresponding fits (blue). The blue bars represent the standard deviation of the range of fits that agree with our data. Both experimental time courses and fits are normalized by the average of the time course. (F) Most of the RNA’s begin to change between 6-8 hrs, when the cells begin to enter stationary phase. This is demonstrated by the histogram of *t*1, the time to the first inflection point.

**Figure 4 - supplemental figure 2. Distributions of time scales found by fitting the piecewise continuous curve, described in the main text, to the mRNA and protein profiles.**

(A-C) mRNA distributions of *t*1, time to first inflection (A), the time between the first inflection and the time it takes to stop to changing (B), and the total time it takes for a given profile to stop changing(C). (D-F) Protein distributions of *t*1, time to first inflection (D), the time between the first inflection and the time it takes to stop to changing (E), and the total time it takes for a given profile to stop changing (F).