# The *E. coli* molecular phenotype under different growth conditions

Mehmet U. Caglar, John R. Houser, Walter F. Lenoir, Drew Vander Wood, Dariya K. Sydykova, Bartram L. Smith, Dan Boutz … other authors … Edward M. Marcotte, Jeffrey E. Barrick, Claus O. Wilke

Affiliations

## Abstract

Modern systems biology requires extensive, carefully curated measurement of cellular components in response to different environmental conditions. While high-throughput methods have made transcriptomics and proteomics datasets widely accessible and relatively cheap to generate, systematic measurements of both mRNA and protein abundances under a wide range of different conditions are still relatively rare. Here we present a detailed, genome wide transcriptomics and proteomics dataset of *E. coli* grown under 34 different conditions. We manipulate concentrations of sodium and magnesium in the growth media, and we consider four different carbon sources glucose, gluconate, lactate, and glycerol. Moreover, samples are taken both in exponential and stationary phase, and we include two extensive time-courses, with multiple samples taken between 3h and 2 weeks. We find that exponential-phase samples systematically differ from stationary-phase samples, in particular at the level of mRNA. Regulatory responses to different carbon sources or salt stresses are more moderate, but we find numerous differentially expressed genes for growth on gluconate and under salt and magnesium stress. Our data set provides a rich resource for future computational modeling of *E. coli* gene regulation, transcription, and translation.

## Introduction

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 (Botstein and Risch 2003). Computational models are particularly well suited to studying this problem, as they can synthesize and organize diverse and complex data in a predictive framework, but there is a need for more detailed experimental studies including more samples in this area to understand interactions between different levels of omics (Zhang, Li, and Nie 2010). Much effort is currently being spent on understanding how to best integrate multiple cellular subsystems. For example, there are many proposed approaches to combining gene expression with protein abundances focusing on integrative, whole-cell models (Joyce and Palsson 2006; Zhang, Li, and Nie 2010; Ideker et al. 2001; Vogel and Marcotte 2012).

Given the growing interest in integrative modeling approaches, there is a pressing need for high quality genome-wide data that is comparable across cellular subsystems under many different external conditions. *E. coli* is an ideal organism to study genome-wide, multi-level regulatory effects of external conditions, since it is well adapted to the laboratory environment (Lee 1996) and one of the first organisms studied at the whole-genome level (Blattner et al. 1997). There have been a number of studies of the *E. coli* transcriptome and/or proteome in response to different growth conditions. For example, if cells are grown at high density, expression of most amino-acid biosynthesis genes is down-regulated and expression of chaperones is up-regulated, suggesting that high cell density is stressful for cells (Yoon et al. 2003). Exposure of *E. coli* to reduced temperature leads to changes in gene-expression patterns consistent with reduced metabolism and growth (Gadgil, Kapur, and Hu 2005). At the single-cell level, the *E. coli* transcriptomes and proteomes are noisy, and mRNA and protein copy numbers tend to be uncorrelated (Taniguchi et al. 2010). Under long-term glucose starvation, mRNAs are generally down-regulated while the protein response is more varied (Houser et al. 2015). Copy numbers of proteins involved in energy-intensive processes decline whereas those of proteins involved in nutrient metabolism remain constant, likely to provide the cell with the ability to jump-start metabolism when nutrients become available again. Larger-scale studies that measure mRNA and/or protein abundances under multiple different conditions include (Soufi et al. 2015; Schmidt et al. 2015; Lewis et al. 2010; Lewis et al. 2009).

Here, we provide a systematic analysis of *E. coli* gene expression under a wide variety of different conditions. We measure both mRNA and protein abundances, at exponential and stationary phases, for growth conditions including different carbon sources and different salt stresses. We find that mRNAs and proteins display divergent responses to the different growth conditions. Further, growth phase yields more systematic differences in gene expression than does either carbon source or salt stress, though this effect is more pronounced in mRNAs than in proteins. We expect that our data set will provide a rich resource for future modeling work.

## Results

### Experimental design and data collection

We grew multiple cultures of *E. coli* REL606, from the same stock, under a variety of different growth conditions. We measured RNA abundances under all conditions and matching protein abundances for approximately 2/3 of the conditions (Figure 1). Results from one of these conditions, long-term glucose starvation, have been presented previously (Houser et al. 2015). Conditions not previously described include one additional starvation experiment, using glycerol instead of glucose as carbon source, exponential and stationary phase cultures using either gluconate or lactate as carbon source, and conditions including Mg+2 and Na+1 stress.

Measurements of RNA and protein abundances were carried out as previously described (Houser et al. 2015). All resulting data sets were carefully checked for quality, normalized, and log-transformed. We discarded 6 RNA samples from further analysis due to inconsistencies in biological replicates, so that our final data set consisted of 143 RNA samples and 101 protein samples (Supplementary Table 1).

Our raw RNA-seq data covers 4279 distinct mRNAs and our protein data covers 4201 distinct proteins. All raw data files are available in appropriate repositories (see Methods for details), and final processed data are available as Supplementary Tables 3 and 4.

**Broad trends of gene expression differ between mRNA and proteins**

To identify broad trends of gene expression among the different growth conditions, we performed hierarchical clustering on both mRNA and protein abundances (Figures 2 and 3). For mRNA, we found that differences in gene expression were primarily driven by the growth phase (exponential vs. stationary/late stationary). Nearly all exponential samples clustered together in one group, separate from the vast majority of stationary and late-stationary samples (Figure 2). Mg+2 levels, Na+1 levels, and carbon source had less influence on the clustering results. We also found a similar result for protein abundances (Figure 3). The exponential-phase samples grouped together, separated from stationary and late stationary samples. Similarly, Na+1 levels and carbon sources also seemed to be grouped together upon clustering.

To quantify the clustering patterns of mRNA and protein abundances, we defined a metric that measured how strongly clustered a given variable of the growth environment (growth phase, Mg+2 level, Na+1 level, carbon source) was relative to the random expectation of no clustering. For each variable, we calculated the mean cophenetic distance between all pairs corresponding to the same condition (e.g., for growth phase, all pairs sampled at exponential phase and all pairs sampled at stationary/late stationary phase). The cophenetic distance is defined as the height of the dendrogram produced by the hierarchical clustering from the two selected leafs to the point where the two branches merge. We then converted each mean cophenetic distance into a *z*-score, by resampling mean cophenetic distances from dendograms with reshuffled leaf assignments. A *z*-score below −1.96 indicates that the mRNA or protein abundances are clustered significantly by the corresponding variable.

### We found that mRNA abundances were significantly clustered by growth phase, with a *z*-score of −23.99 (Table 1). Na+1 and Mg+2 levels displayed the next-largest *z*-scores by magnitude, of −1.54 and −1.46, but these were not significantly different from zero. The *z*-score for carbon source was 1.16, which implies that there is no significant clustering by carbon source in the mRNA data. Importantly, when we calculated a *z*-score for batch number, we found that batch effects also significantly influenced mRNA abundances, with *z* = −2.82. Batch number here represents cultures grown at the same time; all bacterial samples with the same batch number were grown in parallel.

### For protein abundances, the variables Na+1 level, growth phase, and carbon source were all significantly clustered, with *z*-scores of −4.78, −4.21, and −3.15, respectively (Table 1). Batch number had a *z*-score of −23.29, which implies that there were strong batch effects present in the protein data.

### In summary, the largest effect in mRNA abundances, growth phase, was similarly present in proteins. However, protein abundances clustered also by Na+1 and carbon source, effects that weren’t present in the mRNA data. Finally, both mRNA and protein data were influenced by batch effects, and the effect was much more pronounced for proteins than for mRNA (Table 1).

### Identification of differentially expressed genes

We next asked under which conditions and to what extent RNA and protein expression were altered. To identify differentially expressed mRNAs and proteins, we used DeSeq2 (Love, Huber, and Anders 2014). Since our data clustered significantly by growth phase, we analyzed RNA and protein expression separately for exponential and stationary phase. For each growth phase, we defined a reference condition of glucose as carbon source, with 5mM Na+1 and 0.8mM Mg+2. We then compared RNA and protein abundances between this reference condition and the alternative conditions (different carbon sources, elevated Na+1, and elevated or reduced Mg+2). Note that a detailed comparison of reference exponential phase vs. reference stationary phase has already been published (Houser et al. 2015).

We defined significantly differentially expressed genes as those whose abundance had at least a two-fold change (log2 fold change > 1) between the reference condition and a chosen experimental condition, at an FDR-corrected *P* value < 0.05. We found that the number of significantly differentially expressed mRNAs and proteins varied substantially between exponential and stationary phase and between mRNAs and proteins (Figure 4). In general, there were fewer differentially expressed genes in stationary phase than in exponential phase. Further, protein abundances showed the most differential regulation for high Na+1 and for the carbon source glycerol, whereas mRNA showed the most differential regulation for altered Mg+2 levels and again for the carbon source glycerol (Figure 4).

Next, we asked how much overlap there was among differentially expressed genes between the various growth conditions. To simplify this analysis, we did not distinguish between up- or down-regulated genes, and we combined low and high Mg+2 into one group “Mg stress” and glycerol, lactate, and gluconate into one group “carbon source”. (Note that differentially expressed genes were still identified for individual conditions, as described above, and were combined into “Mg stress” and “carbon source” only for the final comparison.) At the mRNA level, there was little overlap between Na stress and either Mg stress or carbon source, and moderate overlap between Mg stress and carbon source, in particular in exponential phase (Figure 5). By contrast, at the protein level, there was more overlap between Na stress and either Mg stress or carbon source, in particular in exponential phase (Figure 5).

To see the response of *E. coli* to different external conditions we look at the significantly altered biological pathways and molecular activities of gene products. We use Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) for biological pathways and annotations from Gene Ontology (GO) Consortium for molecular functions (Ashburner et al. 2000). To find the significantly altered GO annotations and KEGG pathways we use DAVID (Huang, Sherman, and Lempicki 2008) web tools, and use *P' = 0.05* to filter significantly altered molecular activities and biological pathways. Supplementary tables 6 and 7 show the top 5 significantly altered biological pathways indicated by KEGG and molecular function as indicated by GO annotations under different conditions. We also individually display the changed pathways/annotations and associated, significantly altered mRNAs and proteins under supplementary figures 1–39. Significantly altered KEGG pathways with high Na+1 and high Mg+2 concentrations in exponential phase associated with mRNA concentrations are shown in figures 6 and 7. As one can see, almost all of the most altered pathways are significantly down regulated in both situations

We also found, in contrast to other salt stresses we observed, high Mg concentrations in stationary phase causes up regulation of many metabolism related pathways (Figure 8), which seems like an exception when considering other up and down regulation trends (Figure 9).

## Discussion

We have studied the regulatory response of *E. coli* under a wide variety of different growth conditions. The experimental conditions we have considered include four different carbon sources, variations of Na+1 and Mg+2 stress levels, and growth into deep stationary phase, up to two weeks post inoculation. We have found that gene regulation changes the most with respect to growth phase; in general, the exponential phase under one condition is more similar to the exponential phase under another condition than to the stationary phase under the same condition. Further, we have found that the smallest number of differentially expressed genes for different carbon sources and larger numbers for different salt stresses. Under salt stress, most genes tend to be down-regulated, with the exception of elevated Mg+2 in stationary phase, where we have found over twice as many up-regulated mRNAs as down-regulated mRNAs. Finally, our measurements provide a comprehensive dataset that will be useful for future modeling of the *E. coli* metabolism.

Our data provides a comprehensive picture of *E. coli* in terms of number, range, and depth of different stresses, comparable and complementary to other recently published datasets. For example, (Schmidt et al. 2015) considered 22 unique conditions and measured abundances of >2300 proteins. mRNA abundances were not measured. (Soufi et al. 2015) considered 10 unique conditions and also measured abundances of >2300 proteins. They were interested primarily in up- and down-regulated proteins under different ethanol stresses. However, similar to our study, they found down-regulation of genes associated with arginine biosynthesis and metallic ion binding under stress. (Lewis et al. 2010) considered only 3 different carbon sources but measured mRNA and protein abundances in different strains adapted to these growth conditions. Finally, (Lewis et al. 2009) compiled a database of 213 mRNA expression profiles covering 70 unique conditions, including different carbon sources, terminal electron acceptor, growth phase, and genotype. In comparison, we considered 34 unique conditions, measured 143 mRNA expression profiles and 101 protein expression profiles, and used the same genotype throughout.

Similar to our prior study (Houser et al. 2015), we observed clear differences in the differential expression of mRNAs and proteins. In particular, (Houser et al. 2015) reported that mRNAs are widely down-regulated in stationary phase whereas only select proteins are down-regulated. Consistent with that observation, we found here that mRNAs were significantly and strongly clustered by growth phase (*z* = −23.99) whereas proteins were much less so (*z* = −4.21). By contrast, at the protein level we saw significant clustering by Na+1 levels (*z* = −4.78) and carbon source (*z* = −3.25), which we didn’t see at the mRNA level. Overall these observations are consistent with the conclusion from (Houser et al. 2015) that mRNA synthesis is widely shut-off in stationary phase, likely to preserve energy, and that proteins are selectively retained to guarantee cell survival and ability to jump-start metabolic activity once nutrients become available again.

*Combine with previous paragraph:* In 2015 (Houser et al. 2015) study founds an up regulation of stress related proteins and down regulation of energy intensive processes, with our extended data set we also found down regulation of metabolism related processes almost in all stress conditions and in all growth phases. Some similar down-regulated pathways under stresses other than glucose starvation are translation, flagellar assembly, nucleotide related pathways such as RNA degredation, aminoacyl-tRNA biosynthesis. Similar stress related down regulation of pathways associated with energy intensive processes were also reported in several previous studies. (Yoon et al. 2003; Gadgil, Kapur, and Hu 2005)

For all growth conditions we observed patterns of altered gene expression consistent with the known function of the differentially expressed genes. For example, under Na+1 stress we saw up-regulation of tktB and talA, as previously reported (Weber, Kögl, and Jung 2006). We also saw many changes in ABC transporter expression under both Na+1 stress and Mg+2 stress. In particular, low Mg+2 concentrations caused a decrease in the fec operon involved in iron transport. A possible explanation is that magnesium transporters are permeable to ferrous iron at low magnesium concentrations (Hantke 1997), and a down-regulation of iron transporters may be needed to compensate. The fact that many ABC transporters were affected by salt concentrations may be indicative of other transporters that are leaky to off-target ions.

Providing glycerol instead of glucose as the sole carbon source increased expression of the glp operon, which is involved in glycerol uptake (Weissenborn, Wittekindt, and Larson 1992). Gluconate as a carbon source increased expression of the gtn and idn operons, both involved in gluconate metabolism (Fujita et al. 1986; Bausch et al. 1998). Finally, for both altered salt concentrations and altered carbon sources we saw differential expression in a number of genes involved in amino-acid and nucleotide biosynthesis, as well as genes encoding ribosomal proteins. These regulatory changes may reflect the overall metabolic activity of *E. coli*, and how the bacteria generally respond to more or less challenging conditions, rather than any specific induction caused by experimental conditions.

It is well known from large-scale, high-throughput gene-expression studies that measurements can be confounded by batch effects (refs) A third problem is the batch effects; the experiments were done in a large time intervals and in different laboratories if there are some affects that are related with these uncontrolled variables it is hard to fix them since batches also correlated with experiments. The batch effect for mRNA and proteins are both significant with z scores of −2.82 and −23.39 respectively. There are plenty of methods pre or post computational discussed associated with reducing the batch effects. One of the recent studies discuss multiple methods related with removing batch effects (Kim, Zorraquino, and Tagkopoulos 2015), where they generate an iterative machine learning algorithm to cope with batch effects.

In summary, our study provides a large and comprehensive data set investigating the gene-regulatory response of *E. coli* under different growth conditions, both at the mRNA and the protein level. We have found systematic differences in gene-expression response between exponential and stationary phase, and between mRNAs and proteins. We believe that our data set provides a rich resource for future modeling of the *E. coli* metabolism.

## Materials and Methods

### Cell Growth, RNA seq, and proteomics

Growth and harvesting of *E. coli* B REL606 cell pellets for the multiomic analysis was as previously described (Houser et al. 2015) with the following additional details. For tests of different carbon sources, the Davis Minimal (DM) medium used was supplemented with 0.5 g/L of the specified compound (glycerol, lactate, or gluconate) instead of glucose. Mg2+ concentrations were varied by changing the amount of MgSO4 added to DM media from the concentration of 0.83 mM that is normally present. For tests of different Na+ concentrations, NaCl was added to achieve the final concentration. The base recipe for DM already contains ~5 mM Na+ due to the inclusion of sodium citrate, so 95 mM NaCl was added for the 100 mM Na+ condition, for example. Exponential phase samples were taken during growth when the OD600 reached 20-60% of the maximum achieved after saturating growth. Stationary phase samples were collect 20-24 hours after the corresponding exponential sample. The exact sampling times for each condition are provided in Supplementary Table S1.

After sample collection, RNA seq and mass-spec proteomics were performed exactly as described in (Houser et al. 2015).

### Data Preparation

Our raw input data consisted of RNA and protein counts. Protein counts can be fractional, because some peptide spectra cannot be uniquely mapped to a single protein. We rounded all protein counts to the nearest neighbor for subsequent analysis. We set the counts of all unobserved proteins to zero. Subsequently, all RNA and protein counts were analyzed in the same manner.

We next performed quality control, by checking replicates of the same condition for consistency. For all pairs of replicate samples, we made histograms of the log-differences of RNA or protein counts. If the two samples differ only by experimental noise, then the resulting histogram should have a mode at 0 and be approximately bell-shaped. If a sample consistently shows deviations from this expectation when compared to other samples, then there are likely systematic problems with this sample, and we excluded all such samples. In total, we excluded XXX samples. (Is Fig. 1 before or after samples have been excluded?)

After quality control, we normalized read counts using size-factors calculated via DeSeq (reference). Because we had many RNAs and proteins with counts of zero at some condition, we added pseudo-counts of +1 to all counts before calculating size factors. However, we then used those size factors to normalize raw counts (i.e., without pseudo-counts).

### Clustering

We clustered normalized counts based on their Eucledian distance, using the complete linkage method implemented in the hclust function in R. This method defines the cluster distance between two clusters as the maximum distance between their individual components (Madhulatha 2012). At every stage of the clustering process, the two closest clusters are merged into the next bigger cluster. The final outcome of this process is a dendogram that measures the closeness of different samples to each other.

To assess whether the clustering process significantly grouped similar samples together, we employed a reshuffling test. For any category that we tested for significant clustering (e.g., carbon source, Na stress, or batch number), we calculated the mean cophenetic distance in the clustering dendogram between all pairs belonging to the same level of the categorical variable tested (e.g., same carbon source). We then repeatedly reshuffled the labeling within each category and recalculated the mean cophenetic distance each time. Finally, we calculated *z* scores of the original cophenetic distance relative to the distribution of reshuffled values.

### Quantification of cell response

The determination of differentially expressed mRNA and proteins contains a process that can be determined in multiple steps.

Firstly we filter out the samples that biological replicates are not poses similar distributions with in each other and sum all the technical replicates to generate one sample, which is the suggested way to handle technical replicates in DeSeq2 [ref]. The next step is to filter out the unnecessary conditions from the data file. For example if we are trying to figure out the affects of high Mg+2 on *E. coli* in exponential phase we first pick the samples that only have glucose as carbon source, Na levels are equal to 0.8mM (base value), and pick the base and high Mg samples in exponential phase. We do not apply any filtering on specific mRNAs or proteins based on raw counts.

Then we apply DeSeq2 to calculate the size factors for samples in order to get rid of the affects of search depth inequalities. While calculating size factors we add "+1" to all samples to get non-zero results for the mRNAs and proteins that have at least one zero read.

After finishing the DeSeq2 procedure we have log 2 fold change and adjusted p values for individual mRNAs and proteins. Picking up the samples based on *P'<0.05* and *log2 fold change >2* we end up with lists of mRNAs and proteins which responds to external change significantly. We use these lists and send them to the database for annotation, visualization and integrated discovery tool (DAVID) (Huang, Sherman, and Lempicki 2008) and look for results related with KEGG pathways (Kanehisa and Goto 2000) and molecular function related GO annotations (Ashburner et al. 2000).

## References

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. “HTSeq--a Python Framework to Work with High-Throughput Sequencing Data.” *Bioinformatics (Oxford, England)* 31 (2): 166–69. doi:10.1093/bioinformatics/btu638.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, et al. 2000. “Gene Ontology: Tool for the Unification of Biology.” *Nature Genetics* 25 (1): 25–29. doi:10.1038/75556.

Bausch, C., N. Peekhaus, C. Utz, T. Blais, E. Murray, T. Lowary, and T. Conway. 1998. “Sequence Analysis of the GntII (subsidiary) System for Gluconate Metabolism Reveals a Novel Pathway for L-Idonic Acid Catabolism in Escherichia Coli.” *Journal of Bacteriology* 180 (14): 3704–10.

Blattner, Frederick R., Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, et al. 1997. “The Complete Genome Sequence of Escherichia Coli K-12.” *Science* 277 (5331): 1453–62. doi:10.1126/science.277.5331.1453.

Botstein, David, and Neil Risch. 2003. “Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease.” *Nature Genetics* 33 (March): 228–37. doi:10.1038/ng1090.

Burge, Sarah W., Jennifer Daub, Ruth Eberhardt, John Tate, Lars Barquist, Eric P. Nawrocki, Sean R. Eddy, Paul P. Gardner, and Alex Bateman. 2013. “Rfam 11.0: 10 Years of RNA Families.” *Nucleic Acids Research* 41 (Database issue): D226–32. doi:10.1093/nar/gks1005.

Dodt, Matthias, Johannes T. Roehr, Rina Ahmed, and Christoph Dieterich. 2012. “FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms.” *Biology* 1 (3): 895–905. doi:10.3390/biology1030895.

Fujita, Y., T. Fujita, Y. Miwa, J. Nihashi, and Y. Aratani. 1986. “Organization and Transcription of the Gluconate Operon, Gnt, of Bacillus Subtilis.” *Journal of Biological Chemistry* 261 (29): 13744–53.

Gadgil, Mugdha, Vivek Kapur, and Wei-Shou Hu. 2005. “Transcriptional Response of Escherichia Coli to Temperature Shift.” *Biotechnology Progress* 21 (3): 689–99. doi:10.1021/bp049630l.

Hantke, K. 1997. “Ferrous Iron Uptake by a Magnesium Transport System Is Toxic for Escherichia Coli and Salmonella Typhimurium.” *Journal of Bacteriology* 179 (19): 6201–4.

Houser, John R., Craig Barnhart, Daniel R. Boutz, Sean M. Carroll, Aurko Dasgupta, Joshua K. Michener, Brittany D. Needham, et al. 2015. “Controlled Measurement and Comparative Analysis of Cellular Components in E. Coli Reveals Broad Regulatory Changes in Response to Glucose Starvation.” *PLoS Computational Biology* 11 (8): e1004400. doi:10.1371/journal.pcbi.1004400.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2008. “Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources.” *Nature Protocols* 4 (1): 44–57. doi:10.1038/nprot.2008.211.

Ideker, Trey, Vesteinn Thorsson, Jeffrey A. Ranish, Rowan Christmas, Jeremy Buhler, Jimmy K. Eng, Roger Bumgarner, David R. Goodlett, Ruedi Aebersold, and Leroy Hood. 2001. “Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network.” *Science* 292 (5518): 929–34. doi:10.1126/science.292.5518.929.

Jeong, Haeyoung, Valérie Barbe, Choong Hoon Lee, David Vallenet, Dong Su Yu, Sang-Haeng Choi, Arnaud Couloux, et al. 2009. “Genome Sequences of Escherichia Coli B Strains REL606 and BL21(DE3).” *Journal of Molecular Biology* 394 (4): 644–52. doi:10.1016/j.jmb.2009.09.052.

Joyce, Andrew R., and Bernhard Ø Palsson. 2006. “The Model Organism as a System: Integrating ‘Omics’ Data Sets.” *Nature Reviews Molecular Cell Biology* 7 (3): 198–210. doi:10.1038/nrm1857.

Kanehisa, Minoru, and Susumu Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1): 27–30.

Kim, Minseung, Violeta Zorraquino, and Ilias Tagkopoulos. 2015. “Microbial Forensics: Predicting Phenotypic Characteristics and Environmental Conditions from Large-Scale Gene Expression Profiles.” *PLOS Comput Biol* 11 (3): e1004127. doi:10.1371/journal.pcbi.1004127.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59. doi:10.1038/nmeth.1923.

Lee, Sang Yup. 1996. “High Cell-Density Culture of Escherichia Coli.” *Trends in Biotechnology* 14 (3): 98–105. doi:10.1016/0167-7799(96)80930-9.

Lewis, Nathan E., Byung-Kwan Cho, Eric M. Knight, and Bernhard O. Palsson. 2009. “Gene Expression Profiling and the Use of Genome-Scale In Silico Models of Escherichia Coli for Analysis: Providing Context for Content.” *Journal of Bacteriology* 191 (11): 3437–44. doi:10.1128/JB.00034-09.

Lewis, Nathan E, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, et al. 2010. “Omic Data from Evolved E. Coli Are Consistent with Computed Optimal Growth from Genome-Scale Models.” *Molecular Systems Biology* 6 (July): 390. doi:10.1038/msb.2010.47.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. doi:10.1186/s13059-014-0550-8.

Madhulatha, T. Soni. 2012. “An Overview on Clustering Methods.” *arXiv:1205.1117 [cs]*, May. http://arxiv.org/abs/1205.1117.

Schmidt, Alexander, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. 2015. “The Quantitative and Condition-Dependent Escherichia Coli Proteome.” *Nature Biotechnology* advance online publication (December). doi:10.1038/nbt.3418.

Soufi, Boumediene, Karsten Krug, Andreas Harst, and Boris Macek. 2015. “Characterization of the E. Coli Proteome and Its Modifications during Growth and Ethanol Stress.” *Frontiers in Microbiology* 6 (February). doi:10.3389/fmicb.2015.00103.

Taniguchi, Yuichi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie. 2010. “Quantifying E. Coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells.” *Science* 329 (5991): 533–38. doi:10.1126/science.1188308.

Vogel, Christine, and Edward M. Marcotte. 2012. “Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses.” *Nature Reviews Genetics* 13 (4): 227–32. doi:10.1038/nrg3185.

Weber, Arnim, Stephanie A. Kögl, and Kirsten Jung. 2006. “Time-Dependent Proteome Alterations under Osmotic Stress during Aerobic and Anaerobic Growth in Escherichia Coli.” *Journal of Bacteriology* 188 (20): 7165–75. doi:10.1128/JB.00508-06.

Weissenborn, D. L., N. Wittekindt, and T. J. Larson. 1992. “Structure and Regulation of the glpFK Operon Encoding Glycerol Diffusion Facilitator and Glycerol Kinase of Escherichia Coli K-12.” *Journal of Biological Chemistry* 267 (9): 6122–31.

Yoon, Sung Ho, Mee-Jung Han, Sang Yup Lee, Ki Jun Jeong, and Jong-Shin Yoo. 2003. “Combined Transcriptome and Proteome Analysis of Escherichia Coli during High Cell Density Culture.” *Biotechnology and Bioengineering* 81 (7): 753–67. doi:10.1002/bit.10626.

Zhang, W., F. Li, and L. Nie. 2010. “Integrating Multiple ‘Omics’ Analysis for Microbial Biology: Application and Methodologies.” *Microbiology* 156 (2): 287–301. doi:10.1099/mic.0.034793-0.