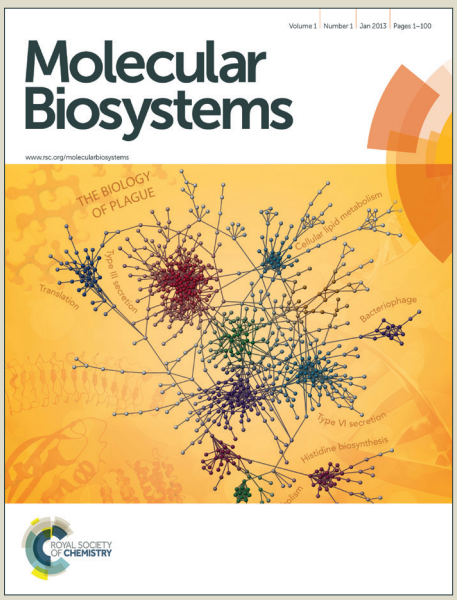
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Next-generation analysis of gene expression regulation – comparing the roles of synthesis and degradation

Joel McManus[1], Zhe Cheng[2], Christine Vogel[2]

1 Carnegie Mellon University, Department of Biological Sciences, Pittsburgh, PA, USA

2 New York University, Department of Biology and Center for Genomics and Systems Biology, New York NY, USA

Corresponding authors: cvogel@nyu.edu, mcmanus@andrew.cmu.edu

Abstract

Technological advances now enable routine measurement of mRNA and protein abundances, and estimates of their rates of synthesis and degradation that inform on their values and the degree of change in response to stimuli. Importantly, more and more data on time-series experiments are emerging, e.g. of cells responding to stress, enabling first insights into a new dimension of gene expression regulation - its dynamics and how it allows for very different response signals across genes. This review discusses recently published methods and datasets, their impact on what we now know about the relationships between concentrations and synthesis rates of mRNAs and proteins in yeast and mammalian cells, their evolution, and new hypotheses on translation regulatory mechanisms generated by approaches that involve ribosome footprinting.

The different stages of protein expression regulation

The seemingly simple task of producing a protein molecule from its gene is in fact highly complex. Protein production is regulated in multiple, diverse ways which all act in a controlled, but stochastic and highly dynamic manner in what we collectively call 'gene expression regulation'. Gene expression regulation involves synthesis of mRNA and protein via transcription and translation, respectively, and degradation of the molecules. Both transcription and translation are coordinated by many participating factors and pathways. Roughly 2,000 of the ~20,000 protein-coding genes in the human genome encode are transcription factors[1]. A similar fraction of the genome appears to regulate the second stage of protein synthesis: the human genome may encode as many as ~1,000 RNA-binding proteins and ~1,000 miRNAs which putatively regulate both RNA degradation and translation[2-4].

Many additional processes add further complexity to gene expression regulation. Alternative pre-mRNA splicing generates an average of four transcript variants per human gene[5-7]. Alternative translation initiation and termination can create additional variants. Once a protein is made, ~200 unique post-translational modifications, including phosphorylation, acetylation, ubiquitination, and SUMOylation, can be attached to target it for degradation, change its localization, interactions, and functions. Consequently, the Uniprot sequence database comprises >68,000 human protein variants, produced from just over 20,000 genes[8].

While sometimes overlooked, the degradation of mRNA and protein molecules is as much regulated as is their synthesis. mRNA turnover regulation is highly complex, occurring through two major pathways. In rapidly growing cells, most mRNA decay is initiated by removal of the m^7G cap found on the 5' end. However, in some cases decay is initiated by removal of the polyA tail – a process called deadenylation [9]. Furthermore, the vast majority of protein degradation in eukaryotic cells is managed by the proteasome which itself consists of a protease core and regulatory caps. Proteasomal degradation is initiated by lysine-48-linked polyubiquitination of the target protein – a process regulated by more than 100 ubiquitinating and deubiquitinating enzymes in yeast, and hundreds in mammalian cells[8, 10, 11]. The targets and condition-specific activities of these enzymes are only known for a small subset.

These regulatory processes are further complicated by feedback mechanisms and coupling between individual processes[12]. For example, mRNA degradation has been reported to be coupled to both transcription [13] and translation [9, 14]. Other work suggests that RNA-binding proteins and miRNAs, two entirely different regulators of RNA translation and degradation, can jointly regulate the same pathway [15]. Therefore, the 'one gene – one protein' hypothesis is far from describing gene

expression regulation in its entirety, ignoring the plethora of different protein products, their interactions, combinatorial regulation, and changes in response to stimuli.

This review first outlines recent methods that enable large-scale measurements of concentrations and rates. We place special emphasis on an approach called ribosome footprinting, which provides estimates of translation efficiency and has received much attention with respect to both the insights it provides and its limitations. We then discuss new insights into the principles and evolution of gene expression regulation from studies using these techniques on yeast and mammalian cells. We finish by describing our view of where the field of systems biology of gene regulation is headed and what questions are likely to be addressed in the near future.

Experimental approaches to characterize gene expression regulation

Excitingly, the last decade has seen enormous technological and methodological advances that enable large-scale measurements of the above-described multiple dimensions of gene expression regulation – both with respect to measurements of concentrations and rates (**Table S1**). While modifications and interactions can also be measured, they are not the focus of this review. For comprehensive reviews of other 'dimensions' of gene expression dynamics, see refs. [16, 17].

Measuring genome-wide mRNA and protein concentrations

Standard methods to estimate mRNA and protein concentrations are high-throughput RNA sequencing and shotgun proteomics, respectively (**Table S1**). In most cases, molecular concentrations of genes are estimated relative to each other. However, both approaches can be used with spike-in reference samples if absolute copy numbers per cell are desired. While RNA-seq is truly genome-wide, shotgun proteomics has yet to cross that threshold. Individual proteomics efforts from laboratories with highest-end instrumentation are now able to identify up to ~12,000 proteins in mammalian cells, e.g. ref. [18], but routine measurements identify fewer proteins. However, compared to just a few years ago, proteomics has advanced enough to allow time-series experiments, measuring the abundance of hundreds to thousands of proteins across multiple time points. As a result, the first integrative studies that combine mRNA and matching protein concentration measurements have recently been published (**Table 1**).

Similarly, new computational tools have emerged that enable statistical significance analysis of these time course experiments to extract regulatory information (for focused reviews, see [19, 20]). For example, a recent study by Jovanovic et al. used a mathematical model involving differential equations to estimate translation and protein degradation rates from time-series, pulsed labeling data [21]. Earlier work in yeast [22] and human [23] applied similar approaches: the change of protein concentration over time is modeled as a linear function of protein synthesis based on the mRNA concentration and a translation rate, and degradation, based on the existing protein concentration and a degradation rate. While a linear model is a very simple approach, it can still reproduce

comparatively complex concentration changes for a large fraction of observed patterns [24]. To quantify the contributions of the different regulatory levels and identify genes and time points at which these significant changes occur, we recently developed a statistical framework called Protein Expression Control Analysis (PECA). PECA transforms time-course mRNA and matching protein expression data into significance measures of regulation at both the RNA or protein level, resolved at a per-time-point basis [25].

Such computational analyses of time-series data are vital to progress to the next stage of gene expression analysis: the dynamics of regulatory systems. They provide highly specific types of information at the level of individual genes, but also, in conjunction with other, orthogonal information, describe emerging properties of the networks that regulate gene expression. For example, time series data can verify the functional interactions between regulators and their targets and is instrumental to identification of causal relationships [20]. Further, when we applied PECA to various yeast time series datasets [25], we detected significantly changing genes at a per-time-point basis and some cases of regulatory 'buffering', i.e. synthesis of mRNA molecules that were counteracted by degradation of proteins (and *vice versa*) – observations that would be hidden in static data.

Measuring rates of synthesis and degradation

Recent technological developments now allow researchers to move beyond descriptions of concentrations to experimental measurements of molecular synthesis and degradation rates (**Table S1**). Classic approaches to estimate rates of mRNA degradation involve shutting off transcription with either drugs (e.g. actinomycin and thiolutin) or temperature sensitive yeast mutants of RNA polymerase II (*rpb1-1*)[9](**Table S1**). After inhibiting synthesis, the decreasing molecule concentrations are fit to a decay function to estimate degradation rates. However, these approaches have two main disadvantages. First, degradation may not follow the assumed decay function. Second, these approaches severely disrupt cellular homeostasis, and therefore provide poor estimates of the actual degradation rates. For example, thiolutin has been shown to inhibit both mRNA synthesis *and* degradation in yeast[26, 27]. Even the well-regarded *rpb1-1* mutation system in yeast appears to decrease transcription only 3-fold at the non-permissive temperature, with rates recovering after about one hour [28]. As a result, there is little concordance among mRNA degradation rates measured in different studies.

To circumvent these problems, methods have been developed to measure degradation rates in the absence of inhibitors. Instead, pulsed labeling is used to mark preexisting and newly synthesized mRNAs or proteins. In the case of RNA, several techniques pulse-label RNA with 4-thio-uridine, a non-disruptive analog that can be biochemically enriched after RNA preparation [29, 30](**Table S1**). A time-resolved comparison of the labeled and unlabeled molecules can then be used to estimate

both transcription and RNA degradation rates. Datasets first emerged for yeast, but are also now available for mammalian cells [23].

Similar logic applies to measuring protein synthesis and degradation. Time-resolved proteomics measurements of differently labeled amino acids, in approaches such as pulsed- and dynamic-SILAC[31, 32], have provided rate estimates for both yeast and mammalian cells **(Table S1)**. One challenge with proteomics-based measurements is insufficient coverage, which is further decreased when several time points are required. Another challenge lies in the rate of label incorporation into the proteins, which can lead to small mass spectrometric peaks below the detection limit. Therefore it is very difficult to measure synthesis rates for proteins that are rarely translated. To circumvent this challenge, novel approaches have been developed which use a methionine-analog or a tagged puromycin translation inhibitor to specifically enrich for newly synthesized proteins (**Table S1**). Such approaches, increase sensitivity and coverage of the translation rate measurements – but have they have their own disadvantages by the methionine analog affecting cellular homeostasis and not easily penetrating thick cell walls such as those of yeast.

Ribosome profiling – generating hypotheses on translation regulation

Often, it is not the actual rate that is most interesting in an experiment, but an estimate of the *efficiency* of the process – it may be more informative to learn how a rate is regulated in the cell rather than to compare two rate values. Recent years have seen much excitement about a new method that combines the resolution and deep coverage of next-generation sequencing with measurements of translation efficiency – and to generate hypotheses on the mechanisms underlying translation regulation. The method is interchangeably called ribosome profiling, ribosome footprinting, or Ribo-seq (**Table S1**) [33].

The basic approach is outlined in **Table S1** and **Figure 2**. The data generated by ribosome profiling can provide both bulk translation efficiency estimates per gene and nuanced mechanistic details of translation with respect to regulatory sequence elements in the mRNA. While standardized computational methods for data analysis are still under development, the relative number of ribosomes translating an open reading frame (ORF) can be estimated by tallying the number of reads that cover each ORF. By comparing these estimates of ribosome load with RNA-abundance estimates, the relative efficiency of translation (RPF/mRNA) can be measured. No other method provides this information simultaneously at system-wide scale.

However, ribosome footprinting also has some disadvantages. First, it is one of the most difficult experimental methods in RNA biology, involving many different steps. Second, unless used in combination with different translation inhibitors and time-resolved measurements[34], it does not provide actual *rate* estimates. Furthermore, ribosome footprinting assumes that the number of

mRNA-bound ribosomes correlates with translation efficiencies. Although this is likely to be a reasonable assumption, a direct comparison of ribosome footprinting and proteomics-based identification of newly synthesized proteins is still needed. Finally, the slow uptake of translation inhibitors (which may be an essential part of the method) can complicate studies in yeast [35].

New insights into gene expression regulation

For many years, comparisons of protein and mRNA concentrations have been limited to steady-state measurements. Under these conditions, the population averages of protein and mRNA concentrations in unperturbed cells do not vary over time, with the molecules produced and degraded simultaneously at equilibrium rates. Excitingly, for both baker's and fission yeast and mammalian cells, several time-series datasets of protein and mRNA expression are now available, complemented by a few measurements of synthesis and degradation rates (**Table 1**). These data allow for early insights into the principles governing gene regulation in dynamic systems, i.e. cells responding to a stimulus. The following sections highlight what we think might be general trends and future perspectives in light of these recent advances.

Relationships between concentrations and rates

The first question that has been asked for many years addresses the correlation between mRNA and protein concentrations within one organism growing at steady state[36]. A perfect correlation between mRNA and protein concentrations would suggest no gene-specific differences in translation or protein degradation – but reality is far from that. By 2009, several estimates for the protein-mRNA correlation under steady state were available[36], but no common trend across bacteria and yeast was observed. For mammalian cells, both a computational study in 2010 and an experimental approach in 2011 concluded that "transcription is only half the story"[37] and protein-level regulation may be as important as that of the RNA-level[23, 38]. Similarly, in synchronized cells at different cell cycle stages, post-transcriptional regulation has been observed for much of the proteome[39]. However, these findings have been disputed, and reanalysis of the 2011 data showed that transcription may indeed do the majority of the regulatory workload, accounting for 56 to 81% of the overall variation in gene expression[40].

Moving from steady-state to dynamic systems, an experimental study in dendritic cells responding to lipopolysaccharide (LPS) treatment estimated that mRNA levels (set by both transcription and degradation) explain 59 to 68% of the variance in protein levels[21] – again placing the main workload in gene expression regulation on transcription. The authors find that RNA-level regulation governs the response of newly synthesized proteins, while protein-level regulation is more important for concentration adjustments of preexisting proteins with basic cellular functions [21]. During the LPS response, mRNA changes drive the overall expression response, and even for down-regulated mRNAs the discrepancy to changes in protein concentrations can be explained by

a combination of slow and delayed protein degradation and translation[21]. The study partially contrasts earlier findings in yeast responding to osmotic stress, where the correlation between protein and mRNA abundances was stronger for up-regulated than for down-regulated genes, suggesting that protein-level regulation is dominant with respect to protein removal[22]. One of the reasons for these drastic reprogramming choices might lie in the fact that these studies subjected cells to very large and rapid perturbations. Further nuances and implications of the relationship between protein and mRNA concentrations in the cell are discussed in an excellent review in ref. [41].

A second, very basic question involves the ranges of concentrations and rates that were found – and again a somewhat surprising picture emerged. For example, although RNA-seq is arguably more sensitive compared to proteomics – collecting millions of reads compared to tens of thousands of spectra – the dynamic range of mRNA concentrations seems consistently smaller than that of protein concentrations [23, 42]. For example, RNA concentrations in mammalian cells vary over four to five orders of magnitude, starting from 0.1 molecule per mammalian cell on average – while protein concentrations have been found to cover up to six or even up to 12 orders of magnitude [18, 42]. For example, ribosomes each contain dozens of ribosomal proteins and exist in copy numbers as high as ten million per mammalian cell [43].

Similarly, the rates of synthesis and degradation vary enormously between genes, but are in general much larger at the protein level compared to the RNA level (**Figure 1**). For example, yeast transcription rates range between 0.03 and 0.5 mRNA copies per minute and RNA half-lives range from 2 to 60 minutes, but vary under different conditions (**Table S1**)[26, 28, 30]. Transcription rates in mammalian cells range between 0.1 and 100 mRNA copies per cell per hour [23]. In comparison, proteins are much more stable than mRNA, with median half-lives of hours if not days, and mammalian translation rates vary from 0.1 to 10^5 proteins per mRNA per hour across genes (**Table 1**). However, estimates gained from ribosome profiling analyses placed this range much smaller, spanning only two orders of magnitude[33] – a finding that might be due to technical limitations. In sum, proteins, compared to mRNA, have a larger dynamic range in rates of synthesis and degradation, delivering one explanation for the fact that overall protein concentrations are much larger than mRNA concentrations.

Speculations on reasons and consequences of different rates

As a refined view of gene expression regulation slowly emerges (**Figure 1**), we can begin to ask questions as to why rates of mRNA and protein synthesis and degradation may have evolved to their current values, and how, these rates produce very different expression response patterns to internal or environmental challenges on a per-gene basis. This question is especially interesting if one assumes that well-adapted biological systems must be capable of mounting large and rapid responses to the fluctuations found in the wild, while remaining robust to transient perturbations.

The ability to change concentrations rapidly and drastically depends greatly on the absolute concentration: it is much easier to change small concentrations than large ones. As described above, mRNAs are usually expressed at concentrations lower than those of proteins (**Figure 1**), which allows for very fast and large fold-changes while not requiring much absolute synthesis and degradation due to low molecule numbers. In contrast, changing the concentrations of proteins, which are often higher than those of the corresponding mRNAs, by only a small amount requires enormous efforts with respect to translation or degradation and a large energy expense. Therefore unsurprisingly, stimulus-dependent fold-changes observed at the mRNA-level are often much larger than those seen for proteins[21, 22, 44]. This consideration is perhaps one explanation for the small effects on translation often seen by regulators such as miRNAs and RNA-binding proteins[45, 46].

Gene functions may place particular constraints on mRNA and protein synthesis and degradation rates. For example, consider the differing requirements of transcription factors (TFs) and ribosomal proteins (RPs). During vegetative growth, TFs often function with few protein copies while RPs are required at very high copy number. However, the two classes of genes may have similar mRNA concentrations (**Figure 3,** example I). Upon a stimulus, transcription factors are often the first responders to cellular signals, requiring rapid production and fold changes in a switch-like manner. Such a fast response could be enabled by rapid translation or slow protein degradation. Therefore, it is unsurprising that ribosomes and other highly expressed proteins have large translation rates which change drastically during a stimulus to enable changes in absolute protein copy numbers[34, 47]. Similarly, the temporal response patterns might differ based on gene function. As a first line of response to signals, transcription factors often require rapid production and large fold-changes in a switch-like manner, while house-keeping genes such as ribosomes or enzymes may not need such a fast response. The low total concentrations of TFs also enables their rapid degradation, leading to pulse-like responses.

In addition to the desired signal shape, concentration, and response time, different genes may differ in their requirements for accuracy and robustness to variation in gene expression levels – which can be achieved by specific combinations of transcription and translation rates (**Figure 3**, example II). For some genes, the accuracy in producing the correct amount might be crucial for their function, while others function well even if their cell-to-cell concentration varies. Therefore, the concentrations of some proteins may change stochastically over time more than those for others. A computational study in yeast suggested such a scenario: high transcription rates combined with low translation rates lead to less noise in final protein concentrations than the opposite case[48].

Interestingly, recent work suggests that gene expression noise is subject to purifying selection for at least one yeast gene [49]. Thus, "Nothing in Biology Makes Sense Except in the Light of Evolution"

(Theodosius Dobzhansky, 1973) may also apply to the evolution of different rates of synthesis and degradation – and the time is ripe for models and hypotheses that explain these selection processes.

Ribosome profiling glances into mechanisms of translation regulation

Ribosome profiling reports ribosome positions at the level of codons or even nucleotides – providing unique insights into possible mechanisms that affect translation efficiency of genes, for example with respect to codon usage. Although many codons can be used to specify the same amino acid, most organisms show clear usage preferences. Genes with abundant mRNA generally use codons that are decoded by abundant tRNA, suggesting that this codon usage bias contributes to increased translation efficiency of abundant mRNA transcripts. To examine this relationship, many researchers have attempted to extract codon-specific translation elongation rates from ribosome profiling data (**Table 1**), however these studies resulted in markedly varying conclusions. Some suggest that rare codons stall translation[50], but this effect is not seen after correcting for amplification and sequencing biases present in ribosome profiling datasets[51]. However, all of this work was done with datasets from yeast grown in log-phase at steady-state, and, similar to the discussions above, it is possible that the codon usage bias becomes more important during dynamic gene expression regulation, e.g. during meiosis or mating.

Further, mapping the genomic locations of ribosome profiling reads has revealed that ribosomes often bind to unexpected regions in the RNA. The most prominent of these include eukaryotic mRNA transcript leader sequences (TLSs, also known as 5' UTRs). Translation within TLSs often occurs in regions termed upstream ORFs (uORFs). While uORF translation has been known for decades anecdotally[52], ribosome profiling suggested that it may be much more widespread than previously appreciated and that uORFs often make use of start codons differing from the canonical AUG sequence [53]. In fact, roughly ~10,000 uORFs are proposed to function in translation regulation during yeast meiosis[54].

Other ribosome profiling experiments revealed that ribosomes often associate with RNAs thought to be non-coding. For example, Ingolia and colleagues reported that candidate long non-coding RNAs (lncRNAs) from mouse embryonic stem cells were often bound by ribosomes[34]. Other data support translation of short ORFs in non-coding RNAs [55, 56], however further evidence indicates that most ncRNAs do not encode functional proteins[57, 58] – raising the question whether binding by ribosomes indeed leads to translation of the mRNA or not. One possibility is that ribosome association may function as a surveillance mechanism to degrade improperly localized ncRNA via nonsense mediated decay.

Conservation and divergence of concentrations and rates

Above we discussed the impact and consequences of the variation in mRNA and protein concentrations across genes within one organism. We now examine the role of these processes across organisms, during evolution (**Figure 4**). For example, mRNA abundances of orthologous genes vary greatly between species. This variation has been observed in comparisons across kingdoms, including yeast[59], *Drosophila*[60, 61], mice[62], and humans[63]. Despite this widespread variation in mRNA abundance, more recent work suggests that protein abundance is less divergent. In a comparison of *D. melanogaster* and *C. elegans* proteome and transcriptome expression data, Schrimpf and colleagues discovered that the correlation between these species' protein abundance was higher than that seen for their mRNA abundance[64]. Soon after, Laurent et al. expanded this analysis to eight organisms and reported that increased conservation of protein abundance could be found in comparisons across all domains of life, including, *E. coli*, *S. cerevisiae*, and humans[65]. More recently, the same phenomenon was observed in comparisons of lymphoblastoid cell lines derived from humans, chimpanzees, and rhesus macaques[66]: selection pressures to conserve protein concentrations across organisms appear to be higher than those on mRNAs. A study in mouse and human cells showed that strong conservation may also apply to protein degradation rates [67] – although rates appear to vary across subcellular localizations[68].

The fact that variation in mRNA abundance between species is not necessarily mirrored at the protein level suggests that post-transcriptional processes act to buffer evolutionary changes in expression regulation[69]. Indeed, comparisons of mRNA abundance and turnover rates in yeast revealed that mRNA degradation often offsets evolutionary differences in mRNA levels[70]. Other work has found that translation regulation buffers species differences in mRNA abundance[71, 72]. These studies compared translation efficiency and mRNA abundance in *S. cerevisae* and *S. paradoxus* using ribosome profiling and found that roughly a quarter of the transcriptome exhibited changes in translation efficiency that were biased toward reducing interspecies differences in protein production. While the observed extent of this "translational buffering" varies between studies[73], a more recent comparison of *S. cerevisiae* and *S. uvarum (bayanus)* identified even more translational buffering than earlier studies[74], potentially due to precise control of environmental differences by co-culturing the two species.

The molecular mechanisms underlying post-transcriptional buffering remain unclear. Interestingly, many *trans*-acting factors have been found responsible for buffering via mRNA turnover[70] and translation[72]. In many of these cases, buffering is likely mediated by proteins that function at multiple levels of gene expression. For example, the yeast RNA binding protein Rpb4 appear to affect both transcription and translation, as do the CCR4/Not complex and Not5[ref.75]. In other cases, buffering of species differences in mRNA levels was mediated by *cis*-acting factors. For example, swapping promoter elements between *S. cerevisiae* and *S. paradoxus* was sufficient

to reproduce species differences in mRNA turnover[76]. Interestingly, promoter sequences have also been shown to determine the subcellular localization and translation efficiency of mRNAs induced during glucose starvation in yeast[77]. Regardless of the exact mechanisms responsible, it appears that transcription, mRNA turnover, and translation are intimately coupled in yeast in a manner that generally increases the robustness of gene expression, i.e. the conservation of protein expression levels across organisms.

Outlook

Large-scale methods are now in place to measure both mRNA and protein concentrations and their rates of synthesis and degradation (**Table S1**), and first datasets have become available that describe these aspects of gene expression regulation in dynamic systems (**Table 1**)[16]. After some earlier insights, we may wonder where the field might be headed. One of the next goals should be to obtain more time-series datasets from different organisms and tissue types under a variety of conditions. These data will help evaluate the general trends already observed in yeast and mammalian cells (**Table 1**). For example, studies in yeast have shown that under stress, the cells appear to strictly regulate either synthesis or degradation of a given protein, but not both[24], and we do not know if this finding is conserved across organisms.

Additional datasets would also help inform the discussion on the relative importance of transcription versus translation. As often in biology, the answer may be a diplomatic "it depends" – on the type of stimulus, the amount of protein that is needed, the response time required, the desired signal shape, or the required accuracy in exact copy numbers (**Figure 3**). Different rates can achieve these scenarios. For example, in cells responding to LPS treatment, transcription plays enables rapid synthesis of functionally relevant proteins, while protein degradation acts more slowly and removes pre-existing functions[21]. In comparison, translation regulation plays major roles in both yeast and mammalian cell responses to environmental challenges or the circadian rhythm[44, 78-80]. Evidence for coupling among transcription, translation, and mRNA turnover processes further complicates this picture – and we are only beginning to understand the impact of such coupling[12].

To evaluate how much these first insights apply in general – across conditions or organisms – we need not only more datasets, but also tools to analyze the data efficiently, specifically incorporating the dynamics of gene expression changes. While the quantitative analysis of RNA-sequencing and shotgun proteomics data is comparatively standardized, complex methods, such as ribosome profiling still lag behind with respect to informatics tools. The largest challenges lie in appropriate normalization and internal calibration to obtain, for example, reliable quantification of changes in ribosome binding between different regions of an mRNA or cross-normalization of transcription and translation data.

The future will likely also bring more integrative studies that combine several different techniques to examine multiple aspects of gene expression within one system, e.g. proteomics and transcriptomics measurements combined with translation and degradation assays. Importantly, theses studies need to carefully consider which type of measurement fits the biological question best and the relative sources of bias and error inherent to each method. While providing essential information on the mechanisms of translation regulation, ribosome profiling does not necessarily provide translation rates; a type of information gained from pulsed-labeling proteomics techniques (**Table S1**). Future integrative studies will demand increasingly advanced analysis techniques – an area that will also see much future development. Mathematical techniques, such as higher order singular value decomposition[81-83], may be used to extract patterns that are common and specific to sets of diverse data matrices. We are at the beginning of an exciting era that moves towards a new dimension of gene expression analysis: that of the dynamics of a response, and the intricate interplay between the multiple regulatory processes that control protein production.

Acknowledgements

We thank Hyungwon Choi and Justin Rendleman for valuable feedback on the manuscript. C.V. acknowledges funding by the NSF (EAGER), the NIH (R01 GM113237), NYU Whitehead Foundation, the NYU URCF, and the Zegar Family Foundation Fund for Genomics Research at New York University. We apologize to all authors whose work was not cited here due to length restrictions.

Figures/ Tables

Figure 1. Annotating the Central Dogma of Molecular Biology

An illustrated version of the Central Dogma of Molecular Biology shows that, thanks to the emergence of new technologies, we can now quantify the concentrations and rates that produce and degrade mRNAs and proteins. Estimates are for mammalian cells, taken from different sources (see text and mainly ref. [23]). Numbers are for illustration purposes and represent overall estimates.

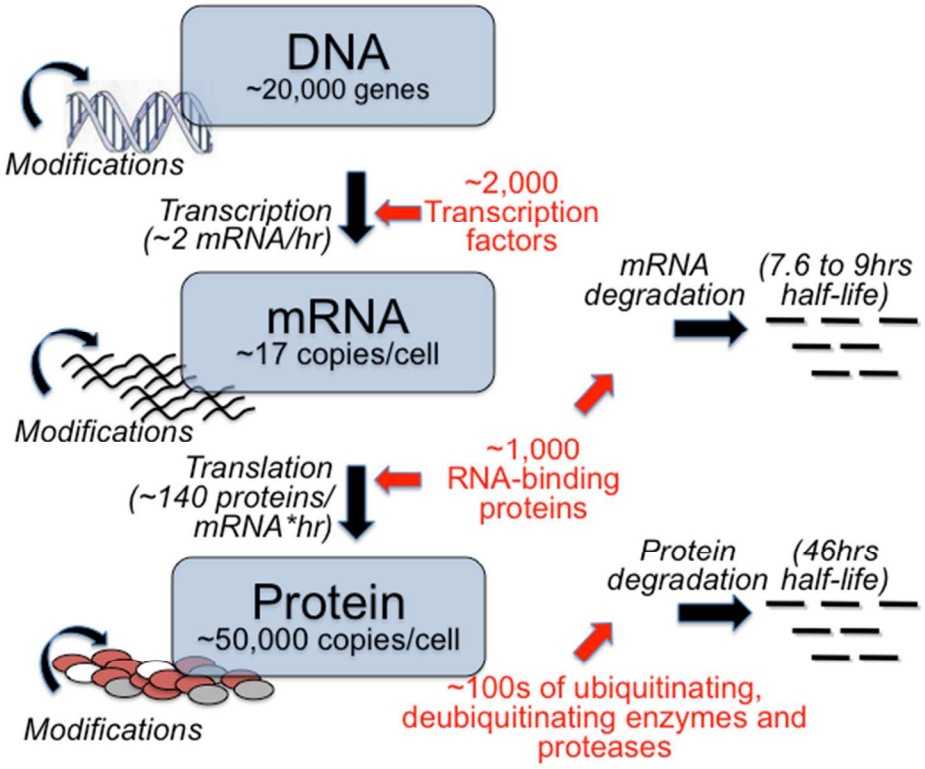


Figure 2. Ribosome profiling

Cell cultures are treated with cycloheximide and lysed in a buffer that maintains ribosome/mRNA associated polysomes. Polysomes are split into two fractions. Nuclease digestion of one fraction removes mRNA fragments not protected by ribosomes. Ribosome protected fragments (RPFs) are then purified and cloned into high-throughput sequencing libraries (left). mRNA is purified from the second fraction, fragmented by base hydrolysis, and cloned into sequencing libraries (right). Libraries are then sequenced to deep coverage.

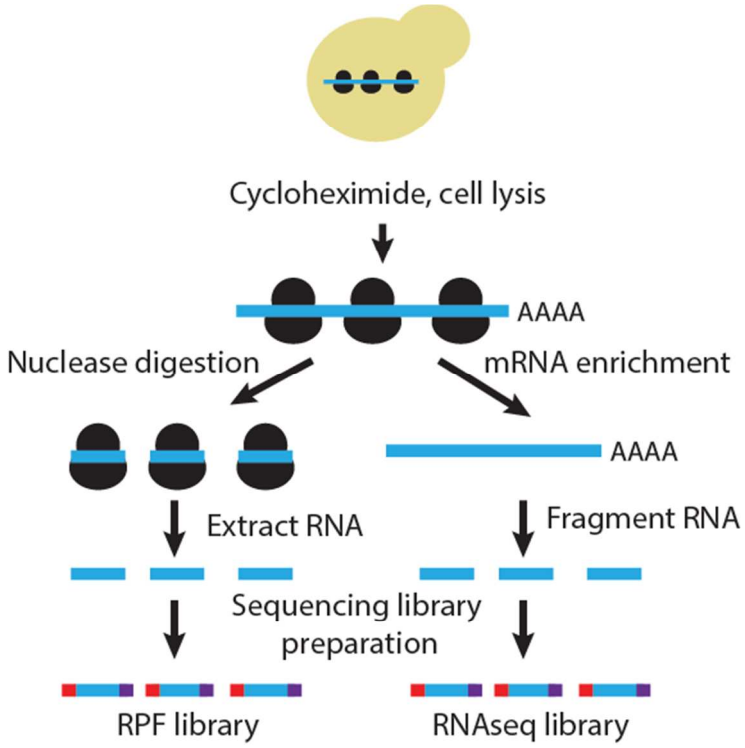


Figure 3. Possible outcomes of different rates of synthesis and degradation

Different rates of synthesis and degradation can result in different concentrations and concentration changes over time. Example I. While having similar RNA concentrations, protein B (e.g. a transcription factor) is much more abundant than protein A (e.g. a ribosomal protein). Upon a stimulus (red arrow), RNA concentrations switch to a new steady-state through increased transcription. For protein B, despite large changes in translation rate, the response time to double the protein concentration (red bracket) is much slower compared to protein A. Example II. At steady-state, transcription for protein B is faster than for protein A which, at similar RNA degradation rates, results in B's RNA being more abundant than that for A. The resulting protein concentrations might be very similar. However, due to comparatively slow translation, B's protein concentration is much noisier over time than that for A.

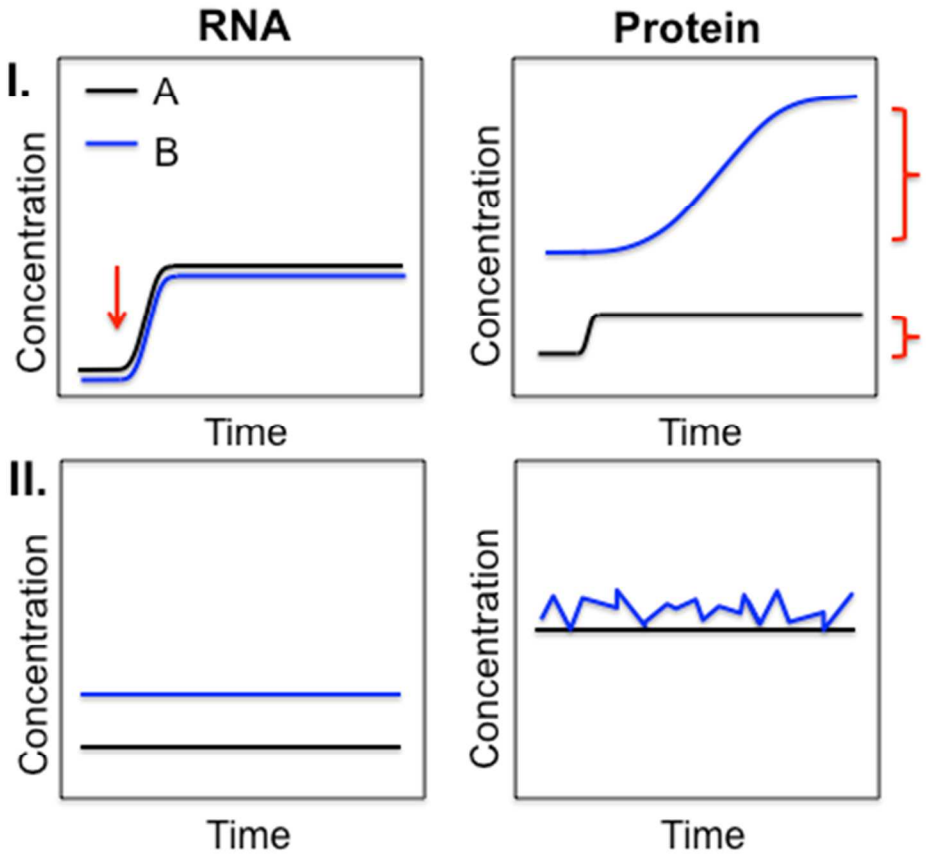


Figure 4. Comparisons of mRNA and protein concentrations among and between species
Shown are simple comparisons of mRNA and protein concentrations, which revealed a surprising observation when compared across organisms (see text, ref. [84]). A. mRNA and protein levels are positively correlated in both *C. elegans* and *H. sapiens*. B. Interspecies comparisons show higher conservation of protein levels than of mRNA. As more datasets become available, we are beginning to understand how mRNA and protein synthesis and degradation rates contribute to the evolution of gene expression.

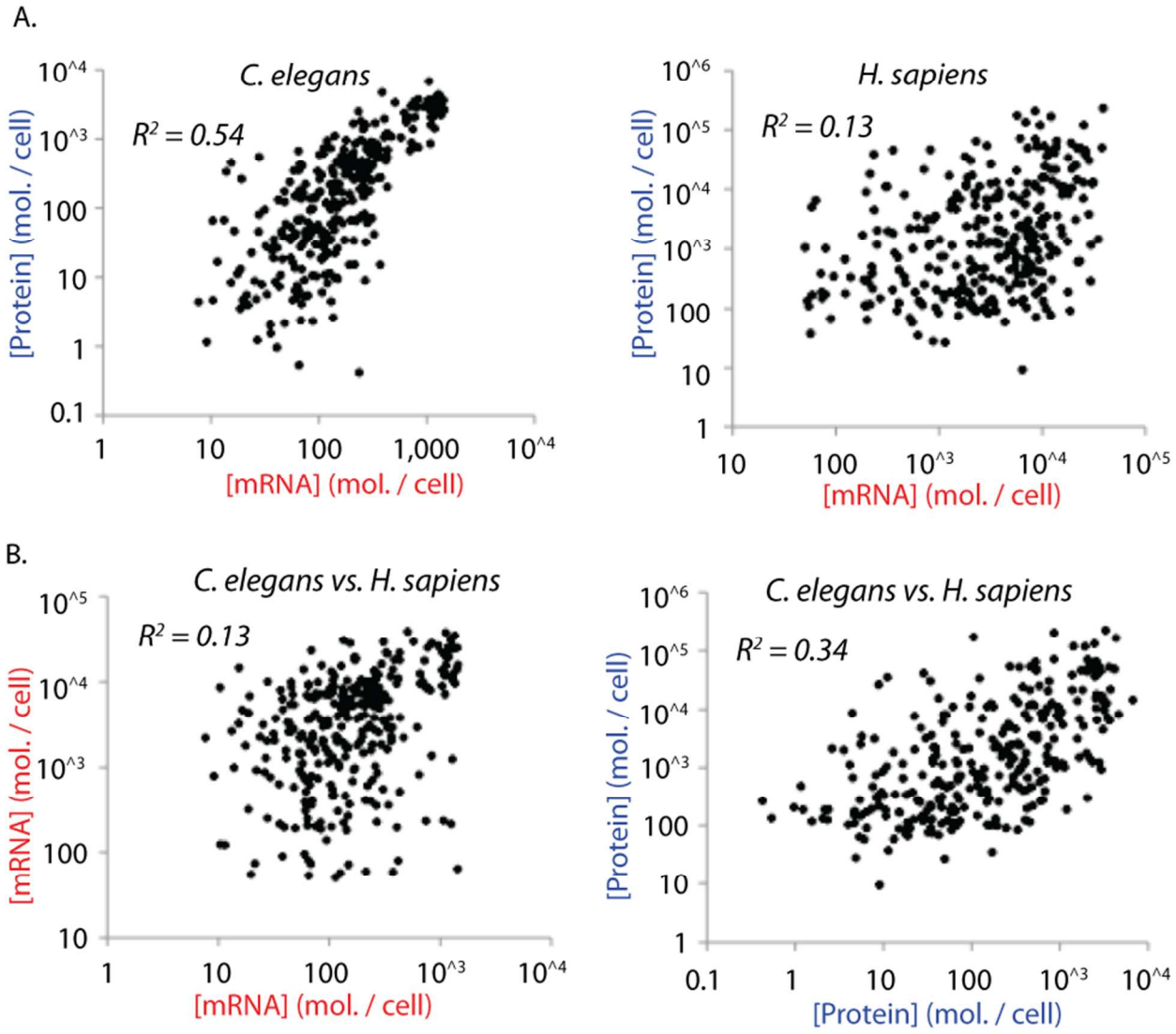


Table 1. Example datasets on dynamic gene expression regulation

Examples of recently published datasets providing first insights into the dynamics of eukaryotic gene expression regulation are shown below. While attempting to cover a range of published datasets, this collection may not be comprehensive.

mRNA and matching protein time-series analysis	Organism	Condition	Number of time points	Approx. #genes	Ref
Fournier, Mol Cell Proteomics 2009	Yeast	Rapamycin treatment	7	6,000	85
Vogel, Mol Cell Proteomics 2011	Yeast	Oxidative stress	8	800	44
Lee, Mol Sys Bio 2011	Yeast	Osmotic stress	6	2,500	22
Lackner, Genome Biology 2012	Yeast (S. pombe)	Oxidative stress	5	2,100	86
Gruen, Cell Reports 2014	Nematode worms	Development	7	3,000	87
Ly, eLIFE 2014	Mammalian cells	Cell cycle	3-6	6,000	88
Eichelbaum, Mol Cell Prot 2014	Mammalian cells	LPS treatment	3-4	4,800	89
Robles, PLoS Gen 2014	Mammalian cells	Circadian rhythm	16	3,000	80
Kristensen, Mol Sys Bio 2013	Mammalian cells	Differentiation	3-5	1,900	90
Jovanovic, Science 2015	Mammalian cells	LPS treatment	6	2,300	21
Transcription rates			**Rate (median or typical range)**		
Pelechano, PLoS One, 2010	Yeast	Normal	2 to 30 mRNAs/hr	4,700	26
Miller, Mol Sys Bio 2011	Yeast	Normal and osmotic stress	1 to 600 mRNAs/cell cycle	5,200	29
Schwanhaeusser, Nature 2011	Mammalian cells	Normal	2 mRNA/hr	5,000	23
RNA degradation			**Half-life (median or typical range)**		
Wang, PNAS, 2002	Yeast	Normal	20 min (3 to 90 min)	4,700	91
Neymotin, RNA, 2004	Yeast	Normal	15 min	5,200	30
Miller, Mol Sys Bio, 2011	Yeast	Normal and osmotic stress	11 min	5,200	29
Munchel, Mol Sys Bio, 2011	Yeast	Normal and stress	20 min	5,200	92
Yang, Genome Res, 2003	Mammalian cells	Normal	2 hrs	1,000s	93
Dolken, RNA, 2008	Mammalian cells	Normal	20 min to 48 hrs	1,000s	94
Friedel, Nucl Acid Res, 2009	Mammalian cells	Normal	4.5 to 5.1 hrs	8,000	95
Schwanhaeusser, Nature 2011	Mammalian cells	Normal	7.6 to 9 hrs	5,000	23
Translation rates			**Rate (median or typical range)**		
Schwanhaeusser, Nature 2011	Mammalian cells	Normal	140 proteins/(mRNA*hr)	5,000	23

Ingolia, Cell 2011	Mammalian cells	Normal	5.6 codons/sec	20,000	34
Protein degradation			**Half-life (median or typical range)**		
Belle, PNAS 2006	Yeast	Normal	4 to 161 min	3,750	96
Christiano, Cell Reports 2014	Yeast	Normal	8.8 to 12.0 hrs	4,000	97
Yen, Science 2008	Mammalian cells	Normal	0.5 to 2 hrs	8,000	98
Doherty, J Proteome Res 2009	Mammalian cells	Normal	6 min to 10s of hrs	600	31
Price, PNAS 2010	Mammalian cells	Normal	72 to 216 hrs	2,500	99
Cambridge, J Prot Res 2011	Mammalian cells	Normal	-	4,100	67
Schwanhaeusser, Nature 2011	Mammalian cells	Normal	46 hrs	5,000	23
Boisvert, Mol Cell Proteomics 2012	Mammalian cells	Normal	20 hrs	8,000	100

References

1 J.M. Vaquerizas, S.K. Kummerfeld, S.A. Teichmann, and N.M. Luscombe, *Nature Publishing Group*, 2009, **10**(4), 252-263.

2 A.G. Baltz, M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler, *Mol Cell*, 2012, **46**(5), 674-690.

3 N.E. Davey, D.T. Humphreys, T. Preiss, and L.M. Steinmetz, *Cell*, 2012.

4 M.R. Friedländer, E. Lizano, A.J.S. Houben, D. Bezdan, M. Báñez-Coronel, G. Kudla, E. Mateu-Huertas, B. Kagerbauer, J. González, K.C. Chen, E.M. Leproust, E. Martí, and X. Estivill, *Genome biology*, 2014, **15**(4), R57.

5 D. Komander, M.J. Clague, and S. Urbe, *Nat Rev Mol Cell Biol*, 2009, **10**(8), 550-563.

6 J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo, and T.J. Hubbard, *Genome Research*, 2012, **22**(9), 1760-1774.

7 N.L. Barbosa-Morais, M. Irimia, Q. Pan, H.Y. Xiong, S. Gueroussov, L.J. Lee, V. Slobodeniuc, C. Kutter, S. Watt, R. Colak, T. Kim, C.M. Misquitta-Ali, M.D. Wilson, P.M. Kim, D.T. Odom, B.J. Frey, and B.J. Blencowe, *Science*, 2012, **338**(6114), 1587-1593.

8 T.U. Consortium, *Nucleic acids research*, 2015, **43**(D1), D204-D212.

9 R. Parker, *Genetics*, 2012, **191**(3), 671-702.

10 D. Komander, *Biochem Soc Trans*, 2009, **37**(Pt 5), 937-953.

11 D. Komander, *Sub-cellular biochemistry*, 2010, **54**, 69-87.

12 O. Dahan, H. Gingold, and Y. Pilpel, *Trends Genet*, 2011, **27**(8), 316-322.

13 K.A. Braun and E.T. Young, *Molecular and cellular biology*, 2014.

14 V. Presnyak, N. Alhusaini, Y.-H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K.E. Baker, B.R. Graveley, and J. Coller, *Cell*, 2015, **160**(6), 1111-1124.

15 A. Joshi, Y. Van de Peer, and T. Michoel, *Nucleic Acids Res*, 2011, **39**(21), 9108-9117.

16 M. Larance and A.I. Lamond, *Nat Rev Mol Cell Biol*, 2015, **16**(5), 269-280.

17 C. Vogel and E.M. Marcotte, *Nat Rev Genet*, 2012, **13**(4), 227-232.

18 N. Nagaraj, J.R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Paabo, and M. Mann, *Mol Syst Biol*, 2011, **7**, 548.

19 Z. Bar-Joseph, A. Gitter, and I. Simon, *Nat Rev Genet*, 2012, **13**(8), 552-564.

20 R. Bonneau, *Nat Chem Biol*, 2008, **4**(11), 658-664.

21 M. Jovanovic, M.S. Rooney, P. Mertins, D. Przybylski, N. Chevrier, R. Satija, E.H. Rodriguez, A.P. Fields, S. Schwartz, R. Raychowdhury, M.R. Mumbach, T. Eisenhaure, M. Rabani, D. Gennert, D. Lu, T. Delorey, J.S. Weissman, S.A. Carr, N. Hacohen, and A. Regev, *Science (New York, NY)*, 2015, **347**(6226), 1259038.

22 M.V. Lee, S.E. Topper, S.L. Hubler, J. Hose, C.D. Wenger, J.J. Coon, and A.P. Gasch, *Mol Syst Biol*, 2011, **7**, 514.

23 B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, *Nature*, 2011.

24 K. Tchourine, C.S. Poultney, L. Wang, G.M. Silva, S. Manohar, C.L. Mueller, R. Bonneau, and C. Vogel, *Mol Biosyst*, 2014, **10**(11), 2850-2862.

25 G. Teo, C. Vogel, D. Ghosh, S. Kim, and H. Choi, 2013.

26 V. Pelechano, S. Chávez, and J.E. Pérez-Ortín, *PLoS ONE*, 2010, **5**(11), e15442.

27 V. Pelechano and J.E. Pérez-Ortín, *Yeast*, 2008, **25**(2), 85-92.

28 M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivière, K.C. Maier, M. Seizl, A. Tresch, and P. Cramer, *Genome Research*, 2012, **22**(7), 1350-1359.

29 C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D.E. Martin, A. Tresch, and P. Cramer, *Molecular Systems Biology*, 2011, **7**, 458.

30 B. Neymotin, R. Athanasiadou, and D. Gresham, *RNA (New York, NY)*, 2014, **20**(10), 1645-1652.

31 M.K. Doherty, D.E. Hammond, M.J. Clague, S.J. Gaskell, and R.J. Beynon, *J Proteome Res*, 2009, **8**(1), 104-112.

32 B. Schwanhausser, M. Gossen, G. Dittmar, and M. Selbach, *Proteomics*, 2009, **9**(1), 205-209.

33 N.T. Ingolia, S. Ghaemmaghami, J.R.S. Newman, and J.S. Weissman, *Science (New York, NY)*, 2009, **324**(5924), 218-223.

34 N.T. Ingolia, L.F. Lareau, and J.S. Weissman, *Cell*, 2011, **147**(4), 789-802.

35 M.V. Gerashchenko and V.N. Gladyshev, *Nucleic Acids Res*, 2014, **42**(17), e134.

36 R. de Sousa Abreu, L.O. Penalva, E. Marcotte, and C. Vogel, *Mol Biosyst*, 2009, **5**(12), 1512.

37 J.B. Plotkin, *Mol Syst Biol*, 2010, **6**, 406.

38 C. Vogel, R. de Sousa Abreu, D. Ko, S.-Y. Le, B.A. Shapiro, S.C. Burns, D. Sandhu, D.R. Boutz, E.M. Marcotte, and L.O. Penalva, *Molecular Systems Biology*, 2010, **6**, 1-9.

39 R. Aviner, T. Geiger, and O. Elroy-Stein, *Genes & Development*, 2013, **27**(16), 1834-1844.

40 J.J. Li, P.J. Bickel, and M.D. Biggin, *PeerJ*, 2014, **2**, e270.

41 S.H. Payne, *Trends Biochem Sci*, 2015, **40**(1), 1-3.

42 S. Marguerat, A. Schmidt, S. Codlin, W. Chen, R. Aebersold, and J. Bahler, *Cell*, 2012, **151**(3), 671-683.

43 B.S.f.C. Biology.

44 C. Vogel, G.M. Silva, and E.M. Marcotte, *Mol Cell Proteomics*, 2011.

45 M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, *Nature*, 2008, **455**(7209), 58-63.

46 D. Baek, J. Villen, C. Shin, F.D. Camargo, S.P. Gygi, and D.P. Bartel, *Nature*, 2008, **455**(7209), 64-71.

47 A.C. Hsieh, Y. Liu, M.P. Edlind, N.T. Ingolia, M.R. Janes, A. Sher, E.Y. Shi, C.R. Stumpf, C. Christensen, M.J. Bonham, S. Wang, P. Ren, M. Martin, K. Jessen, M.E. Feldman, J.S. Weissman, K.M. Shokat, C. Rommel, and D. Ruggero, *Nature*, 2012, **485**(7396), 55-61.

48 H.B. Fraser, A.E. Hirsh, G. Giaever, J. Kumm, and M.B. Eisen, *PLoS Biol*, 2004, **2**(6), e137.

49 B.P. Metzger, D.C. Yuan, J.D. Gruber, F. Duveau, and P.J. Wittkopp, *Nature*, 2015, **521**(7552), 344-347.

50 T. Tuller, I. Veksler-Lublinsky, N. Gazit, M. Kupiec, E. Ruppin, and M. Ziv-Ukelson, *Genome biology*, 2011, **12**(11), R110.

51 C.G. Artieri and H.B. Fraser, *Genome Research*, 2014, **24**(12), 2011-2021.

52 A.G. Hinnebusch, *CRC critical reviews in biochemistry*, 1986, **21**(3), 277-317.

53 N.T. Ingolia, S. Ghaemmaghami, J.R. Newman, and J.S. Weissman, *Science*, 2009.

54 G.A. Brar, M. Yassour, N. Friedman, A. Regev, N.T. Ingolia, and J.S. Weissman, *Science (New York, NY)*, 2011, **335**(6068), 552-557.

55 N.T. Ingolia, *Nature Publishing Group*, 2014, **15**(3), 205-213.

56 J.E. Smith, J.R. Alvarez-Dominguez, N. Kline, N.J. Huynh, S. Geisler, W. Hu, J. Coller, and K.E. Baker, *CellReports*, 2014, **7**(6), 1858-1866.

57 M. Guttman, P. Russell, N.T. Ingolia, J.S. Weissman, and E.S. Lander, *Cell*, 2013, **154**(1), 240-251.

58 B. Bánfai, H. Jia, J. Khatun, E. Wood, B. Risk, W.E. Gundling, A. Kundaje, H.P. Gunawardena, Y. Yu, L. Xie, K. Krajewski, B.D. Strahl, X. Chen, P. Bickel, M.C. Giddings, J.B. Brown, and L. Lipovich, *Genome Research*, 2012, **22**(9), 1646-1657.

59 R.B. Brem, G. Yvert, R. Clinton, and L. Kruglyak, *Science (New York, NY)*, 2002, **296**(5568), 752-755.

60 J.M. Ranz, C.I. Castillo-Davis, C.D. Meiklejohn, and D.L. Hartl, *Science*, 2003, **300**(5626), 1742-1745.

61 S.A. Rifkin, J. Kim, and K.P. White, *Nature Genetics*, 2003, **33**(2), 138-144.

62 E.E. Schadt, S.A. Monks, T.A. Drake, A.J. Lusis, N. Che, V. Colinayo, T.G. Ruff, S.B. Milligan, J.R. Lamb, G. Cavet, P.S. Linsley, M. Mao, R.B. Stoughton, and S.H. Friend, *Nature*, 2003, **422**(6929), 297-302.

63 J.K. Pickrell, J.C. Marioni, A.A. Pai, J.F. Degner, B.E. Engelhardt, E. Nkadori, J.-B. Veyrieras, M. Stephens, Y. Gilad, and J.K. Pritchard, *Nature*, 2010, **464**(7289), 768-772.

64 S.P. Schrimpf, M. Weiss, L. Reiter, C.H. Ahrens, M. Jovanovic, J. Malmström, E. Brunner, S. Mohanty, M.J. Lercher, P.E. Hunziker, R. Aebersold, C. Von Mering, and M.O. Hengartner, *PLoS biology*, 2009, **7**(3), e48.

65 J.M. Laurent, C. Vogel, T. Kwon, S.A. Craig, D.R. Boutz, H.K. Huse, K. Nozue, H. Walia, M. Whiteley, P.C. Ronald, and E.M. Marcotte, *PROTEOMICS*, 2010, **10**(23), 4209-4212.

66 Z. Khan, M.J. Ford, D.A. Cusanovich, A. Mitrano, J.K. Pritchard, and Y. Gilad, *Science*, 2013.

67 S.B. Cambridge, F. Gnad, C. Nguyen, J.L. Bermejo, M. Kruger, and M. Mann, *J Proteome Res*, 2011, **10**(12), 5275-5284.

68 M. Larance, Y. Ahmad, K.J. Kirkwood, T. Ly, and A.I. Lamond, *Mol Cell Proteomics*, 2013, **12**(3), 638-650.

69 C. Vogel, *Science*, 2013, **342**(6162), 1052-1053.

70 M. Dori-Bachash, E. Shema, and I. Tirosh, *PLoS biology*, 2011, **9**(7), e1001106.

71 C.G. Artieri and H.B. Fraser, *Genome Research*, 2013.

72 J. McManus, G.E. May, P. Spealman, and A. Shteyman, *Genome Research*, 2013.

73 F.W. Albert, D. Muzzey, J.S. Weissman, and L. Kruglyak, *PLoS Genet*, 2014, **10**(10), e1004692.

74 Z. Wang, X. Sun, Y. Zhao, X. Guo, H. Jiang, H. Li, and Z. Gu, *Genome Biol Evol*, 2015, **7**(4), 1155-1167.

75 Z. Villanyi, V. Ribaud, S. Kassem, O.O. Panasenko, Z. Pahi, I. Gupta, L. Steinmetz, I. Boros, and M.A. Collart, *PLoS genetics*, 2014, **10**(10), e1004569.

76 M. Dori-Bachash, O. Shalem, Y.S. Manor, Y. Pilpel, and I. Tirosh, *Genome biology*, 2012, **13**(12), R114.

77 B.M. Zid and E.K. O'Shea, *Nature*, 2015, **514**(7520), 117-121.

78 I. Ventoso, A. Kochetov, D. Montaner, J. Dopazo, and J. Santoyo, *PLoS One*, 2012, **7**(5), e35915.

79 M. Stoeckius, D. Grun, M. Kirchner, S. Ayoub, F. Torti, F. Piano, M. Herzog, M. Selbach, and N. Rajewsky, *EMBO J*, 2014, **33**(16), 1751-1766.

80 M.S. Robles, J. Cox, and M. Mann, *PLoS Genet*, 2014, **10**(1), e1004047.

81 O. Alter, P.O. Brown, and D. Botstein, *Proc Natl Acad Sci U S A*, 2003, **100**(6), 3351-3356.

82 L. Omberg, G.H. Golub, and O. Alter, *Proc Natl Acad Sci U S A*, 2007, **104**(47), 18371-18376.

83 S.P. Ponnapalli, M.A. Saunders, C.F. Van Loan, and O. Alter, *PLoS ONE*, 2012, **6**(12), e28072.

84 J. Laurent, C. Vogel, T. Kwon, S. Craig, D.R. Boutz, H. Huse, K. Nozue, H. Walia, M. Whiteley, P. Ronald, and E.M. Marcotte, *Proteomics*, 2010, **10**(23), 4209-4212.

85 M.L. Fournier, A. Paulson, N. Pavelka, A.L. Mosley, K. Gaudenz, W.D. Bradford, E. Glynn, H. Li, M.E. Sardiu, B. Fleharty, C. Seidel, L. Florens, and M.P. Washburn, *Mol Cell Proteomics*, 2009, **9**(2), 271-284.

86 D.H. Lackner, M.W. Schmidt, S. Wu, D.A. Wolf, and J. Bahler, *Genome Biol*, 2012, **13**(4), R25.

87 D. Grun, M. Kirchner, N. Thierfelder, M. Stoeckius, M. Selbach, and N. Rajewsky, *Cell Rep*, 2014, **6**(3), 565-577.

88 T. Ly, Y. Ahmad, A. Shlien, D. Soroka, A. Mills, M.J. Emanuele, M.R. Stratton, and A.I. Lamond, *eLife*, 2014, **3**, e01630.

89 K. Eichelbaum and J. Krijgsveld, *Mol Cell Proteomics*, 2014, **13**(3), 792-810.

90 A.R. Kristensen, J. Gsponer, and L.J. Foster, *Mol Syst Biol*, 2013, **9**, 689.

91 Y. Wang, C.L. Liu, J.D. Storey, R.J. Tibshirani, D. Herschlag, and P.O. Brown, *Proc Natl Acad Sci U S A*, 2002, **99**(9), 5860-5865.

92 S.E. Munchel, R.K. Shultzaberger, N. Takizawa, and K. Weis, *Mol Biol Cell*, 2011, **22**(15), 2787-2795.

93 E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J.E. Darnell, Jr., *Genome Res*, 2003, **13**(8), 1863-1872.

94 L. Dolken, Z. Ruzsics, B. Radle, C.C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal, and U.H. Koszinowski, *RNA*, 2008, **14**(9), 1959-1972.

95 C.C. Friedel, L. Dolken, Z. Ruzsics, U.H. Koszinowski, and R. Zimmer, *Nucleic Acids Res*, 2009, **37**(17), e115.

96 A. Belle, A. Tanay, L. Bitincka, R. Shamir, and E.K. O'Shea, *Proc Natl Acad Sci U S A*, 2006, **103**(35), 13004-13009.

97 R. Christiano, N. Nagaraj, F. Frohlich, and T.C. Walther, *Cell Rep*, 2014, **9**(5), 1959-1965.

98 H.C. Yen, Q. Xu, D.M. Chou, Z. Zhao, and S.J. Elledge, *Science*, 2008, **322**(5903), 918-923.

99 J.C. Price, S. Guan, A. Burlingame, S.B. Prusiner, and S. Ghaemmaghami, *Proc Natl Acad Sci U S A*, 2010, **107**(32), 14508-14513.

100 F.M. Boisvert, Y. Ahmad, M. Gierlinski, F. Charriere, D. Lamont, M. Scott, G. Barton, and A.I. Lamond, *Mol Cell Proteomics*, 2012, **11**(3), M111 011429.