Published in final edited form as:

Wiley Interdiscip Rev RNA. 2013 November ; 4(6): 617–630. doi:10.1002/wrna.1173.

Translation regulation gets its ‘omics’ moment

Scott Kuersten^{1,#}, Agnes Radek¹, Christine Vogel^{2,*}, and Luiz O. F. Penalva^{3,#,*}

¹Epicentre (An Illumina Company), Madison, WI

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

³Greehey Childrens Cancer Research Institute, University of Texas, San Antonio, TX

Abstract

The fate of cellular RNA is largely determined by complex networks of protein-RNA interactions through ribonucleoprotein (RNPs) complexes. Despite their relatively short half-life, transcripts associate with many different proteins that process, modify, translate, and degrade the RNA. Following biogenesis some mRNPs are immediately directed to translation and produce proteins, but many are diverted and regulated by processes including miRNA-mediated mechanisms, transport and localization as well as turnover. Because of this complex interplay estimates of steady state expression by methods such as RNAseq alone cannot capture critical aspects of cellular fate, environmental response, tumorigenesis or gene expression regulation. More selective and integrative tools are needed to measure protein-RNA complexes and the regulatory processes involved. One focus area are measurements of the transcriptome associated with ribosomes and translation. These so-called polysome or ribosome profiling techniques can evaluate translation efficiency as well as the interplay between translation initiation, elongation and termination - subject areas not well understood at a systems biology level. Ribosome profiling is a highly promising technique which provides mRNA positional information of ribosome occupancy, potentially bridging the gap between gene expression (i.e. RNAseq and microarray analysis) and protein quantification (i.e. mass spectrometry). In combination with methods such as RNA immunoprecipitation, miRNA profiling, or proteomics, we obtain a fresh view of global post-transcriptional and translational gene regulation. In addition, these techniques also provide new insight into new regulatory elements, such as alternative open reading frames, and translation regulation under different conditions.

Keywords

Ribosome footprinting; polysome profile; RNA-seq; next-generation sequencing; RIP-seq

Introduction

The advent of high throughput sequencing has prompted the discovery of a plethora of new RNA types present in whole transcriptome samples, e.g. we find long and short non-coding RNAs, miRNAs, piRNAs, lincRNAs and many others¹. The use of RNAseq technologies has evolved to the stage at which we routinely examine transcriptomes at essentially saturating resolutions, thus providing a detailed understanding of the full complement of transcripts present at a particular time and place. Transcriptomics is often used to derive a proxy of the final protein concentrations in the cell, assuming a correlation between mRNA and protein levels. As proteins represent the major group of catalytically active cellular

[#]Corresponding authors: scott.kuersten@epicentre.com; penalva@uthscsa.edu.

^{*}Co-last authors

molecules, estimates of protein levels are critical to understanding cellular processes. However, due to the extremely variable modes of RNA modification and stability, the correlation between protein and mRNA levels is in fact relatively low². For example, some RNAs are simply not translated or contain secondary structure elements that prevent them from being as efficiently translated into proteins. In a medulloblastoma cell line, only ~30% of the variation in protein expression levels can be explained by mRNA concentrations, leaving ~65% of the variation to noise and regulation of translation and protein degradation². This estimate has essentially been confirmed for other cell lines in independent studies, estimating translation to account for as much as 40% of the variation²⁷. Further, several molecular mechanisms can potentially affect translation and mRNA stability, including translation initiation, elongation, termination and even protein degradation. The human genome encodes ~2,000 transcription factors that help to control gene expression and ~1,000 potential RNA binding proteins that putatively regulate transcript stability, localization and translation³. Indeed, there is a vast network of regulatory mechanisms that influence gene expression from transcription to protein synthesis and whole transcriptome sequencing alone does not provide the resolution needed to estimate the protein potential of a cell. For that reason, recent method development has targeted approaches to measure the position and distribution of translating ribosomes which can help bridge the gap between the synthesis of RNA and protein production.

A typical mRNA undergoing translation is bound by a few or many ribosomes. Indeed, perhaps one of the largest and most complex cellular structures is the ‘polysome’, the full complement of ribosomes associated with given mRNAs. Analysis of polysome versus monosome content (called polysome profiling) is a well-established method to characterize and quantitate the mRNA population associated with the ribosomes as a readout of translation efficiency. The method was developed several years ago⁴ and typically uses microarrays as the detection readout – thus it is very robust and comparatively simple to perform, but does not provide information on *where* in the mRNA the ribosomes bind. A more recent approach called ribosome profiling (or footprinting) involves RNA-seq instead of microarrays and therefore provides exact *positional* information of ribosome binding along the mRNA^{5–7}.

Figure 1 provides an overview of these profiling techniques and how they differ from traditional RNAseq methods. The principle method behind polysome and ribosome profiling is to first fractionate cellular extract using a sucrose gradient and centrifugation, and then isolate the intact polysome-containing mRNA to identify bound RNA⁸. The polysome fraction of RNA can be directly compared to the whole transcriptome, i.e. the entirety of mRNA present in the crude cell extract, to determine the proportion and identity of the transcripts involved in active translation. Thus, polysome profiling provides two measurements: the ribosome *occupancy*, i.e. the number of mRNA molecules belonging to a specific gene that are bound by ribosomes; and the ribosome *density*, i.e. the (average) number of ribosomes bound per mRNA (for a given gene). The technique cannot, however, report on the region within the mRNA that is occupied by ribosomes. Both measures, ribosome occupancy and density, have been used as measures of translation efficiency. In literature (and in this review) the term *translation efficiency* is used loosely as an indirect and semi-quantitative measure of the number of proteins produced per mRNA species, which is clearly a function of the number and density of bound ribosomes. However, the term translation efficiency does not allow conclusions about the actual *translation rate*, i.e. the number of proteins produced (per mRNA) per unit time – a measurement that is still very difficult to obtain through experiments.

In the case of ribosome profiling, the polysome-containing lysate is first treated with nuclease to digest any RNA not protected by bound protein. The mRNA sequences involved

in active translation are protected from nuclease digestion due to the highly stable association of ribosomes. They are ~28–35 nt in length (depending on the organism) and are purified and converted into RNAseq libraries for sequence analysis⁹. Ribosome profiling provides information on ribosome occupancy and density similar to polysome profiling, but it differs from polysome analysis in a number of ways. First, ribosome profiling provides essentially single nucleotide resolution to the exact position of a ribosome (the ribosome ‘footprint’) when the lysate was produced. The result is a defined map of ribosome densities *along* transcripts that can be used for identification of translation initiation sites or determination of translational pause sites in transcripts (see below). It has also been used to analyze the impact of codon composition in various regions of the mRNA^{10–12}. Furthermore, since actively translating ribosomes are primarily localized to coding regions of mRNA, ribosome footprints are highly enriched for coding sequences and typically contain very little 5’ or 3’ untranslated region (UTR) (Figure 2). Both profiling methods suffer from a major technical hurdle in that they rely on the use of sucrose gradient ultracentrifugation techniques to purify the large complexes needed for analysis. Sucrose gradient analysis is technically difficult and labor intensive; a primary reason that these approaches have not been readily adopted as a principle tool for analysis. Simplified methods for the isolation of polysomes or nuclease-treated ribosomes would greatly facilitate more wide-spread use of these techniques. Indeed, more recent work has made use of sucrose cushions to pellet nuclease treated ribosomes to isolate bound mRNA fragments⁹ instead of the more complex gradient analysis. An even more simplified technique that involves the use of size exclusion chromatography spin columns is available commercially (e.g. ARTseq™ Ribosome Profiling Kits, Epicentre) and helps to highlight how ribosome profiling methods are quickly evolving and becoming more streamlined and user friendly.

Despite their differences, both polysome analysis and ribosome profiling techniques are extremely useful approaches for analysis of translation efficiency at a global level toward a better understanding of the coding potential of the cell. Studies focusing on the mechanics of translation initiation, elongation and termination as well as translational pausing have applied profiling as a general tool to yield fresh understanding of these processes. In addition, miRNA mediated regulation of translation and the translation activity associated with novel open reading frames (ORFs), in particular upstream of the main ORFs, have also greatly benefited by the use of ribosome footprinting techniques. Below we discuss details of these recent developments.

Investigating the mechanics of general translation

Ribosome footprints produce a distinct profile of read densities along the entire length of transcripts. Since footprint densities are theoretically a direct reflection of the number and residence time ribosomes are associated with a particular sequence, the actual patterns within ORF sequences are quite informative. For example, both yeast and mammalian mRNAs have higher ribosome densities near the start and stop codons than in the middle of the sequence, suggesting longer residence times of ribosomes in these locations^{6,7,35}. Using this assumption on ribosome residence time, the Weissman lab characterized translational pause sites in *E. coli*¹³. Indeed, the researchers identified spikes in ribosome densities at unexpected locations within open reading frames (ORFs). These putative ribosome pause sites displayed sequence signatures with similarity to canonical Shine-Dalgarno (SD) translation initiation sequences. Indeed, ~70% of pause sites across the entire genome were associated with SD features, suggesting that these sites are likely a dominant force in controlling translation rates in *E. coli*¹³. (See also ref. 10–12).

Further, ribosome profiling has been used in eukaryote systems to directly estimate translation *rates*, employing the use of inhibitors known to affect translation and time course

analyses. Using a mouse embryonic stem cell line, Ingolia and Weissman used both translation initiation and elongation inhibitors in a set of 'pulse-chase' style experiments to help calculate genome-wide translation rates⁷. Pretreatment of cells with elongation inhibitors such as cycloheximide or emetine prevent ribosome translocation and thereby 'freeze' ribosomes in the process of translation elongation. Ribosomes subsequently 'stall' at initiation sites, unable to proceed further along the sequence. Interestingly, emetine treatment revealed a slightly larger average sizes of ribosome protected mRNA fragments than generated by cycloheximide, suggesting the two drugs stabilize slightly different forms of elongating ribosomes.

In comparison, drugs such as harringtonine block translation initiation, but not elongation. To estimate translation elongation rates, the authors pulsed the cells with harringtonine for various times to prevent new initiation, then chased with cycloheximide or emetine to freeze the ribosome in place. The samples were then footprinted, sequenced and the relative distributions of ribosome densities along the length of transcripts were compared for the different time points and converted into an estimation of elongation rate. The main surprise was simply that the rate was calculated to be quite similar for all detected transcripts at ~5.6 amino acids per second, independent of transcript class, length or even codon usage⁷. Due to these consistent translation rates, the authors suggest that accuracy of protein level estimations using ribosome density measurements should be a reliable and valid tool.

Aside from the use of translation inhibitors for investigating general translation mechanisms, the significance of engineered drugs that target the translation machinery for therapeutic purposes cannot be overemphasized. For example, the mTOR (mammalian Target of Rapamycin) kinase is a well-known regulator of translation and is involved in a number of important cellular processes such as cell motility, cell growth and proliferation, translation and transcription¹⁴. It is implicated in aging, Alzheimer's disease progression, and misregulated in certain forms of cancer; significantly in prostate cancer¹⁵. Better understanding of mTOR's role in translation control and the development of selective tools to modulate its function *in vivo* is an important therapeutic goal. Recently, the Rommel and Ruggero labs made use of ribosome profiling for this exact purpose¹⁶. Using a prostate cancer cell line, PC3, where mTOR is oncogenically hyperactivated, the authors investigated the effects of mTOR inhibitors on downstream translation targets implicated in cancer progression. By ribosome profiling the authors were able to first establish a set of novel metastasis genes and a set of four pro-invasion genes (YB1, MTA1, vimentin and CD44) all mediated by oncogenic mTOR signaling. Building upon an existing mTOR ATP-site inhibitor, PP242, the authors also characterized a new version, called INK128, and investigated their activity using a ribosome profiling strategy. Strikingly, both ATP-site inhibitors strongly and selectively inhibited the translation of the four pro-invasion genes whereas rapamycin had no effect on expression of these genes¹⁶. These results not only highlight the selective nature of mTOR signaling, but also demonstrate the utility of ribosome profiling as a sensitive tool in this process.

Proteomics as a tool to study translation efficiency

Quantitative shotgun proteomics in combination with RNAseq have been used as alternative methods to assess the differential contributions of transcription and translation to gene expression regulation¹⁷. One of most widely used proteomics approaches relies on mass spectrometry which, with modern high-resolution instruments, can identify several thousands of proteins in a single sample¹⁸. To quantify protein abundance, methods involving both isotopic labeling or peptide modification, as well as label-free methods are now well-established^{19–23}. Label-free methods and methods such as AQUA provide estimates of absolute protein concentrations which, when compared to mRNA

concentrations measured by RNAseq, have been used as a proxy for translation efficiency: the higher the protein-per-mRNA ratio for a given gene compared to other genes in the dataset, the higher is (presumably) the mRNA's translation efficiency^{24,25}.

In contrast to such *absolute* concentration measurements, proteomics methods that involve isotopic labeling usually measure protein expression between an experiment and a control sample, thus reporting the *change* in concentrations. These *changes* in protein concentrations can then be compared to *changes* in RNA concentrations to evaluate how translation and transcription respond to the treatment in the experiment *versus* the control. In the first approach, absolute protein and mRNA concentrations are compared within one sample, while in the second approach, changes in concentrations between two samples are compared. Notably, one group developed a modified method involving a pulse of labeled amino acids (pulsed-SILAC)²⁶. These experiments involve time-course measurements that monitor the changes in ratios of existing versus newly produced proteins²⁷, but can directly give estimates of translation rates. However, the approach has not yet been used widely for translation studies.

Proteomics, in comparison to ribosome footprinting, can be a relatively fast and straightforward method, but has some drawbacks for studies of translation efficiency. First, proteomics only incompletely measures the proteome. Despite recent advances in proteome coverage, proteomics experiments describing >4,000 proteins require significant labor and time¹⁸. Second, to assess translation efficiency, protein concentrations have to be compared to RNA concentrations, requiring additional experiments and involving larger measurement errors. Third, measured protein concentrations reflect the combined result of both translation and protein degradation, and degradation itself is also heavily regulated in the cell. While inhibition of protein degradation to purely analyze translation is a theoretical option, cells usually substantially suffer from such treatment. Proteomics thus provides an indirect measure of translation efficiency which relies on several assumptions (e.g. negligible contribution of protein degradation) that may or may not hold true.

Nevertheless, comparisons of protein and mRNA concentrations have provided useful insights into the roles of translation regulation in both prokaryotic and eukaryotic cells^{24,25}. In general, protein and mRNA concentrations correlate significantly^{24,25}, indicating that a substantial amount of variation in protein concentrations is purely determined by transcription and the resulting mRNA concentrations. Translation is often concordant with transcription: proteins whose transcripts are highly abundant are also more translated than other transcripts. However, even after accounting for the effects of transcription (and transcript degradation), substantial variation in protein concentration remains. In mammalian systems, translation (and also protein degradation) are thought to contribute as much as ~40% to the variation in protein concentrations, while transcription (and transcript degradation) contribute a similar amount^{2,27}.

A comprehensive experiment would include simultaneous transcriptomics, proteomics, and ribosome/polysome profiling in *one* system, but these types of studies are still extremely rare. When publishing the ribosome profiling method in baker's yeast, Ingolia et al. compared their findings to protein concentration measurements available in databases and found that mRNAs bound by ribosomes correlate better in their abundance with proteins than total (including unbound) mRNAs⁵. One possible interpretation of this finding is that different types of mRNA exist: those that are functional and actively translated, and those that are non-functional. Another study directly compared transcript concentrations, translation efficiency, and protein concentrations in fission yeast growing under normal and stress conditions²⁸. The authors found that transcript concentrations and translation efficiency correlate well except for 200 proteins across the two conditions. In contrast,

protein concentrations largely disagree with measures of translation efficiency or transcript levels²⁸. This finding suggests extensive regulation at the level of protein degradation – in addition to translation regulation. Similar suggestions have been made for cells growing under perturbed conditions (non-steady state). For both fission and baker's yeast, up-regulated transcripts show strong agreement between mRNA and protein concentrations, while this is not the case for down-regulated genes^{28,29}. To down-regulate protein concentrations, the cell involves an intricate protein degradation machinery.

miRNA regulation

Since some of the earliest experiments on miRNA-mediated regulation of gene expression in *C. elegans* it was discovered that these tiny 21–23 nucleotide RNAs biochemically co-fractionate with very large polysomal mRNA complexes in sucrose gradient ultracentrifugation^{30,31}. This observation suggested that miRNAs are associated with actively translated mRNAs and are therefore likely to act post-transcriptionally to regulate protein levels of their targets. Indeed, this property has been exploited as a means to isolate miRNAs bound to their target mRNAs. Cloonan et al. noticed that purification of high molecular weight RNA from polysome fractions resulted in co-purification of miRNAs³². Performing RNAseq and quantitative analysis on both the miRNA and mRNA components of these samples suggested an enrichment of target mRNA-miRNA pairs indicating the preservation of a physical interaction during the extraction from cells. Using this characteristic, the authors engineered biotinylated synthetic miRNAs or isomir versions and transfected them into cells, then performed pull-down experiments from polysomal fractions to isolate and identify the enriched target mRNAs. These experiments demonstrated that both miRNAs and their isomirs function synergistically to regulate protein levels. Also, it further confirmed the association of these tiny effectors of gene regulation with very large polysome containing structures in living cells.

The exact mechanism by which miRNAs regulate protein levels has been hotly debated from the time of their initial discovery. Great effort has been made to identify the detailed mechanisms of miRNA action on their targets and their overall impact on gene regulation. Three main concepts have emerged as to their mechanism of action: target mRNA degradation, direct inhibition of translation, and nascent peptide turnover^{33,34}. Since miRNAs and their targets are found associated with polysomal complexes, the method of ribosome profiling is uniquely suited to detect and quantify target mRNA at a distinct point in the regulatory process and potentially provide better understanding of how miRNAs function in gene regulation. It is perhaps not surprising that both polysomal analysis and ribosome profiling have been employed by several groups in an attempt to clarify the mechanism of miRNA action.

Target mRNA degradation

Small interfering RNA (siRNA) mediated gene silencing, and miRNA-mediated gene regulation involve similar RNP complexes to perform their functions, for example, the RNA-Induced Silencing Complex, or RISC, is involved in both processes³³. So it is perhaps logical to assume they both may function in a similar manner. However, RISC complexes involved in miRNA-mediated regulation are typically found with a version of the Argonaute proteins that does not contain the catalytic 'slicer' activity³⁴, suggesting RNA degradation may not be the major means by which miRNAs down-regulate protein levels. Regardless, target mRNA levels are often reduced in the presence of their miRNA effectors, implying that at least some fraction of RNA turnover contributes to protein expression regulation. To help address this issue, the Bartel lab used ribosome profiling as a means to analyze the genome-wide effects of miRNA targets in mammalian cells³⁵. Experiments were designed in which cells were transfected with different synthetic miRNA mimics to monitor and

compare their known and predicted mRNA targets. In parallel, endogenously expressed miRNA activity was specifically suppressed. Both mRNA levels and the corresponding ribosome density profiles were compared using RNAseq approaches. Protein abundances estimated from SILAC experiments were also used, allowing the authors to correlate mRNA levels, ribosome densities and protein levels. The main conclusions from this extensive and detailed work is that miRNAs function primarily through destabilization of their mRNA targets (Figure 3). Reduction in mRNA levels closely matched reduction in the relative amounts of ribosome protected mRNA fragments without a change in footprint patterns along the ORF length. A change in ribosome footprint patterns would have indicated that a specific stage during translation was targeted during miRNA regulation. Thus ribosome profiling allowed the authors a more predictive measurement of miRNA-mediated regulation. However, due to the nature of these experiments it was impossible to look closely at the actual timing of various events leading to target mRNA destabilization.

Translational repression, mRNA deadenylation and turnover

To examine specific stages of miRNA-induced regulation in detail, the Giraldez lab exploited a unique scenario in zebrafish embryogenesis in which at a specific time during development the miRNA pool that exists in the embryo is dominated by one particular miRNA, miR-430, which primarily suppresses maternal mRNAs. By comparing wild-type and dicer mutant embryos at embryonic stages before and after expression of miR-430 they were able to provide a detailed view of specific steps of the miRNA regulatory process³⁶. Using a combination of mRNAseq, ribosome profiling and monitoring target mRNA tail length the authors were able to conclude that the initial stages of miRNA regulation employed translational repression, followed by a shortening of the polyA tail length and eventually resulting in the turnover of miR-430 targets (Figure 3). This model provides a mechanistic explanation for the seemingly contradictory observations that miRNAs function either by suppressing translation initiation or by triggering mRNA turnover in other experimental systems. In zebrafish development, both activities are evident and it is more a matter of the chronological order of the events that is important. Ribosome profiling in particular allowed the authors to experimentally separate the various processing events in both time and space, highlighting the strength of this technique in the context of miRNA regulation. In support of their model, work in *Drosophila* S2 cells monitoring both endogenous and synthetic miRNA targets arrives at similar conclusions; translational repression is initially detected, followed by deadenylation and then decay of the mRNA targets of the regulatory miRNAs³⁷.

Co-translational degradation of nascent peptides

While the above models for miRNA regulation can help explain many observations into their mechanistic function, seemingly contradictory issues remain. One such issue concerns the observation that the relatively low levels of RNA turnover typically seen with miRNA-mediated translation repression are often insufficient to explain comparatively dramatic reductions in protein levels, suggesting that translational repression and mRNA turnover may not be the only mode by which miRNAs control their targets. From some of the early work on miRNA regulatory mechanisms in *C. elegans*^{30,31} it was proposed that *lin-4* miRNA controls protein levels at a step beyond translation initiation, perhaps by induction of an activity that causes turnover of the nascent peptide during protein synthesis. This hypothesis is based on the observation that levels of target mRNA, polyA tail length and the proportion of mRNAs associated with polyribosomes are unchanged before and after the developmental stage of highest *lin-4* activity.

The model involving nascent peptide turnover also garners support from some elegant biochemical experiments where reporter mRNAs were expressed in HeLa cells and were

subjected to control by *let-7* miRNA³⁸. These experiments demonstrated that both control and miRNA-regulated reporter mRNAs co-sediment with polyribosomes and have properties indicative of active translation despite the clear presence of *let-7* miRNA. Furthermore, immunoprecipitation of nascent peptides and their associated transcripts from polysome fractions demonstrated successful co-IP was possible in the absence, but not presence of *let-7* regulation³⁸. This suggested that despite the evidence for active translation, any nascent peptide generated under *let-7* control was either hidden or perhaps quickly degraded during synthesis.

More recent work using ribosome profiling has lent additional support to the nascent peptide turnover model. Making use of a set of *C. elegans* heterochronic mutants in *lin-4* and *let-7* miRNAs and their regulated targets, the Fire lab showed that both target mRNA levels and ribosome footprint densities were reduced upon miRNA regulation, however, these effects were seen to be relatively small compared to the overall protein reduction associated with these transcripts. In addition, little to no reduction in the lengths of target mRNA's polyA tails was observed. These observations suggest that at least in *C. elegans* the mode of miRNA-mediated suppression is not exclusively acting at the level of translation or RNA turnover and leaves open the option of peptide turnover³⁹.

How does one reconcile these seemingly contradictory mechanistic possibilities to explain how miRNAs function in gene regulation? One option is to simply embrace these different models collectively as all being 'correct' in some fashion, but not mutually exclusive. For example, depending on the nature of the mRNA, the organism, cell type, or developmental timing, any or all of these mechanisms may contribute to reducing the target proteins, but experimentally only the most dominant effect is observed due to sensitivity issues of the methods. Another possibility is that despite our best efforts we simply do not have a complete picture (yet) of miRNA-mediated gene regulation. For example, it has been proposed that miRNAs may induce the formation of 'pseudo-polysomes' in a *Drosophila in vitro* translation system designed to mimic miRNA-mediated repression⁴⁰. These structures sediment in sucrose gradient fractionation experiments like normal polysomes even when 60S subunit association was blocked, suggesting these heavy structures were unlike traditional actively translating polysomes⁴⁰. This observation may provide some explanation for the numerous examples discussed here where miRNAs and their targets are found in polysomal fractions. The finding that miRNAs may have various means to regulate protein production is perhaps not so difficult to reconcile. Biological networks often involve overlapping mechanisms to achieve a desired result and may be a key way to adapt flexible modes of gene regulation.

RNA binding proteins and translational control

The human genome contains ~1,000 RNA binding proteins (RBPs) which are implicated in a variety of cellular functions from RNA processing to translation³. Most of these RBPs belong to families containing well-established RNA binding motifs (RRM, KS, dsRBD, etc)⁴¹. However, a new approach developed by the Hentze and Landthaler labs to isolate proteins directly associated with mRNAs revealed a surprising number of novel RNA binding proteins without any of the characteristic binding motifs^{42,43}. These findings suggest that the total number of RBPs may be much larger than assumed so far. Since RBPs are still a poorly studied group of proteins, it is hard to estimate how many RBPs from the current list are truly involved in translation regulation or in processes that indirectly affect translation. The RBPs that have been characterized so far could be placed into three main categories: RBPs that form part of the basic translation machinery (e.g. eIF4E, PABP) and interact with any given mRNA species in the cell; RBPs that interact and regulate a specific group of mRNAs (e.g. PTB1, Musashi1, HuR); and ribosomal proteins.

The advent of genomic methods to identify targets of RBPs constitute a major advance in the area of post-transcriptional regulation. The first set of experiments was conducted in Jack Keene's lab and the method of choice was termed RNA Immunoprecipitation (RIP) or ribonomics⁴⁴. At the time, the identification of RBP associated mRNAs was done with microarrays (RIP-Chip); whereas today most scientists have switched to deep sequencing to identify RIP targets (RIP-Seq). A second approach termed CLIP was developed in the Darnell lab⁴⁵. It differs from RIP in many ways including the use of UV crosslinking, RNase digestion and gel purification of RNA-protein complexes and permit a much more accurate mapping of RBP target sequences. Variations of the original protocol include PAR-Clip, iClip and HITS-Clip. For a more detailed discussion of the topic, see ref. 46–48.

Despite some claims that newer versions of above listed methods are very accurate and capture only true interactions, it is unrealistic to believe that all binding sites identified are “functional sites”. In the case of binding to ncRNA, determining functionality is even more challenging as the RBP binding can have diverse implications to ncRNA metabolism, transport/localization and function. RIP or CLIP data by itself is insufficient to determine the direction of the putative regulation (repression or activation) or its intensity. Therefore, it is absolutely necessary that these various methods are used in parallel with ribosome or polysome profiling and even complemented by proteomics. In a perfect scenario, ribosome profiling would be conducted upon knockdown or transgenic expression of the RBP under study and combined with CLIP and RIP data for the same RBP. These combined analyses would deliver a comprehensive list of “functional sites” that could be used to understand the mode of action of each pertinent RBP in translation. An example of such study was performed with Muscleblind-like proteins, which are implicated in eye and muscle development and myotonic dystrophy⁴⁹. Mapping of binding sites for Mbnl1 was performed by CLIP-seq in brain, heart and muscle and indicated enrichment for UGC- and GCU-containing motifs. To determine if Muscleblind proteins have an impact on translation, ribosomal profiling was performed with control and Mbnl1 and/or Mbnl2 depleted myoblasts. Suggesting a function in translation activation, subsets of Mbnl1 3' UTR targets in myoblasts depleted of both Mbnl1 and Mbnl2 showed substantially reduced RPKMs (ribosomal reads per kilobase of mRNA) in comparison to controls or myoblasts depleted of only one RBP⁴⁹.

The ribosome is a highly dynamic RNP complex and its composition can vary considerably as to the content of associated ribosomal proteins. An interesting question is whether ribosomal protein composition can induce translation bias, increasing or decreasing the expression of specific mRNA species. Indeed, work done in yeast by the Silver lab suggests the existence of a ribosome code⁵⁰ where combinations of RNA binding proteins, different forms of ribosomal proteins or rRNAs functionally fine-tune the translation of specific mRNAs to permit a new level of regulation of gene expression. Interestingly ribosomal proteins have been observed to be aberrantly expressed in tumors⁵¹. Moreover, changes in their expression have been connected to tumorigenic activity; as in the case of ribosomal protein L26 and L29 whose knockdown inhibited proliferation of human pancreatic cancer cells⁵². Therefore, it is possible that specific ribosomal proteins produce an increase (or decrease) of specific oncogenic proteins or tumor suppressor genes, thus having an important impact on tumorigenesis and potentially other diseases. Use of a potent technique such as ribosome profiling upon silencing of specific ribosomal proteins should help determine if this is indeed the case.

Direct precipitation of tagged ribosomal proteins has also been used to evaluate gene expression. The Heintz lab developed an approach called TRAP (translating ribosome affinity purification) in which a series of transgenic mice expressed EGFP-tagged ribosomal protein L10a are driven by promoters activated only in particular cell types of the central

nervous system. After lysis of the brain tissue in the presence of cycloheximide, polysome-associated RNAs were isolated via immuno-precipitation with anti-EGFP antibodies and subsequently identified with microarrays. This strategy produced an elaborate gene expression map of the central nervous system with gene expression profiles for more than twenty neuronal types^{53,54}. A second strategy called RiboTag developed by the Amieux lab took advantage of a mouse line where HA tags were inserted into the Rpl22 locus. As the strategy employed Cre recombinase, it is possible to drive the expression of HA-Rpl22 in specific cell types. Similar to the previous strategy, the use of RiboTag mice produced gene expression profiles for a number of neuronal types⁵⁵. Unfortunately, due to characteristics of both experimental designs, it is hard to say if immunoprecipitation of RNA-binding protein captures translation efficiency in an objective manner. If it does, such approaches could become interesting tools to study cell-type specific translation regulation. However, it remains to be determined if levels of immuno-precipitated mRNA are directly correlated with the number of ribosomes associated with a given mRNA.

Novel transcript and ORF discovery

The emergence of polysome analysis and ribosome profiling coupled to RNAseq as a tool for discovery of novel transcripts has added new layers to RNA biology and gene regulation. New transcript models are being discovered by their association with active translation. For example, the Grimmond lab described the isolation of both membrane-bound and free polysomal RNA fractions from human embryonic stem cells and compared these sub-fractions to total cellular RNA from the same source⁵⁶. Not only did this approach allow the authors to extensively profile the RNAs encoding secreted proteins from their membrane-bound fractions, but also gained a significant appreciation for the inherent complexity of hESC pluripotency. Indeed, more than half of all detected and actively translated RNAs differed from typical gene models. Perhaps one of the more striking discoveries was the observation of >1,000 genes with extended 5' or 3' UTRs and transcripts with retained introns; gene models that are otherwise unannotated in current reference databases. These novel isoforms were typically rare, but were enriched in polysomal fractions indicating they are actively and perhaps even preferentially translated. The extended UTR structures of these novel transcripts could also supply additional regulatory sequences or possible structural features involved with regulatory RNA binding proteins involved in stability, transport or translational control.

It is clear from these results that novel transcripts associated with active translation provide an entirely new layer of thinking about genes and regulatory mechanisms. Indeed, integrating selective drug treatments with ribosome footprinting can help to define new aspects of translation dynamics. For example, use of the drug harringtonine results in the strong accumulation of ribosome densities exactly at the sites of translation initiation⁷ thus providing a convenient method to mark and profile individual ORFs, both known and unknown. This effect has been exploited to provide a more defined view of non-cognate (non-AUG) start codons and aided the identification of numerous novel upstream open reading frames and even characterized any potential translation activity of predicted non-coding RNAs. This resulted in the observation that long intergenic non-coding RNAs (lincRNAs) actually contain high densities of ribosome footprints suggestive of active translation on non-standard ORFs⁷. Similarly, two other groups made use of either puromycin⁵⁷ or lactimidomycin⁵⁸ in combination with cycloheximide to enrich for translation initiation sites at both cognate and near-cognate translation initiation sites in human ORFs. These experiments largely confirm the wide-spread use of novel ORFs, upstream ORFs (uORFs), non-cognate ORFs and even internal ORFs in eukaryotes. In yeast meiosis a number of unannotated transcripts containing short ORFs (sORFs) as well as known transcripts containing uORFs were identified and even provide potential negative

regulatory networks associated with extended 5' UTRs⁵⁹. The oxidative stress response in yeast was also found to be associated with examples of N-terminally extended proteins, use of alternative start codons, increased use of short uORFs and even activity of non-cognate translation start sites all within minutes after exposure to hydrogen peroxide⁶⁰. This basic approach to ORF identification has also been put to use in the context of Human Cytomegalovirus infection⁶¹, where the authors describe and clarify up to 751 translated and temporally expressed ORFs within the viral genome; more than double the previously predicted protein coding potential of this virus. Furthermore, over half of all the ORFs identified were shorter than 80 amino acids, suggesting an extensive use of short, almost peptide-length molecules for viral infection.

Collectively, these results are reminiscent of how quickly RNAseq has greatly expanded our view of transcriptome complexity. Polysome analysis and ribosome profiling are quickly providing a similar global view of the intricate world of translation. Ribosome profiling approaches has even been applied to understanding how new genes are created. Carvunis et al⁶² focused on short, species-specific ORF-like signatures from ribosome profiling data that specifically lie outside of known genes in yeast and establish a model for how new genes can evolve independent of gene duplication mechanisms. These novel peptide encoding 'proto' genes are not well conserved, yet are differentially regulated, suggesting they may be under selective pressure or at least may be beneficial. These results provide an evolutionary link to the potential function of intergenic transcription units. Much of the transcribed regions outside of annotated genes that are not predicted to contain coding potential are typically viewed as some sort of non-coding RNA that may or may not play a regulatory role in transcription or RNA stability. However, the use of ribosome profiling has elucidated that some of this RNA may indeed have translation activity to create peptides with novel activities and simultaneously providing a snapshot into the creation of new genes.

Conclusions

This review discusses the use of next-generation mRNA sequencing and its applications in methods such as ribosome footprinting or RIP-seq, to study translation regulation at a system-wide scale. Ribosome profiling is a technique that provides position-specific interactions between ribosomes and translating mRNA, thus allowing measurements of translational efficiency under various experimental conditions. This approach is useful to describe, in great detail, critical stages of translation such as initiation, elongation, termination and the extensive control mechanisms that shape and guide gene expression and expression changes. As recent studies have shown, translation (and protein degradation) are just as important in regulating gene expression as are transcription (and transcript degradation)^{2,27}. In mammalian systems, perhaps as much as ~40% of gene expression variation may be accounted for by translational control highlighting the importance of novel techniques such as ribosome profiling to study the details of translation regulation. Indeed, this review highlighted some of the significant contributions of ribosome profiling to this area of research, e.g. the unexpectedly large number of uORFs that are translated. Future work will have to address questions building on these observations, such as how many of the short peptides translated from uORFs are stable and functional in the cell, how translation initiation may impact overall translation efficiency, or how the mRNA pool may consist of translationally functional and non-functional transcripts that affect their correlation with the observed protein concentrations.

Further coupled to specific and selective purification tools like RIP or CLIP assays, ribosome profiling techniques will be ever more useful to dissect the details of gene regulatory networks and RNA-mediated expression control. The expectation is that these tools will only continue to get more accurate and sensitive in the future. However, as these

techniques evolve into more mainstream utility, the limitations to discovery will increasingly become clogged at the level of informatics and data mining. Thus, the continued development of simple and user-friendly informatics tools for processing deep sequencing data is welcomed and encouraged.

In the last few years RNAseq applications have rapidly supplanted many other applications designed to profile gene expression and regulation^{63–65}. Reasons for this shift include the speedy improvements in sequencing technology and tools available for data analysis, resulting in a dramatic decrease in the cost involved in performing RNAseq experiments. Beyond the technological reasons are also the scientific benefits; RNAseq is a relatively neutral approach to RNA profiling. This means the user does not need to know ahead of time what transcripts are the most important to profile and monitor. This contrasts with more established microarray or even qRT-PCR approaches where the desired sequences for detection must be already known to the user to aid in probe or primer design algorithms. Indeed, the relative number and complexity of newly discovered RNA types and transcripts have grown exponentially in the last five years and contributed to a new appreciation for a ‘whole transcriptome’ or system-wide view of the importance of RNA as a regulatory entity. Indeed, the recent publication of over 30 papers involved with the ENCODE project highlights the significance of these new discoveries⁶⁶. Ribosome footprinting and its ‘cousins’ are showing the path to a system-wide understanding of translation regulation in prokaryotic and eukaryotic cells.

Acknowledgments

Work in the Penalva lab is supported by The Max and Minnie Tomerlin Voelcker Fund, and NIH grant R01HG006015.

References

- Costa FF. Non-coding RNAs: Meet thy masters. *Bioessays*. 2010 Jul; 32(7):599–608. Review. [PubMed: 20544733]
- Vogel C, Abreu Rde S, Ko D, Le SY, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO. Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol*. 2010 Aug 24; 6:400. [PubMed: 20739923]
- Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. *Trends Genet*. 2013 Feb 14.
- Melamed D, Arava Y. Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. *Methods Enzymol*. 2007; 431:177–201. Review. [PubMed: 17923236]
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009 Apr 10; 324(5924): 218–223. [PubMed: 19213877]
- Ingolia NT. Genome-wide translational profiling by ribosome footprinting. *Methods Enzymol*. 2010; 470:119–142. [PubMed: 20946809]
- Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011 Nov 11; 147(4):789–802. [PubMed: 22056041]
- Kolle G, Shepherd JL, Gardiner B, Kassahn KS, Cloonan N, Wood DL, Nourbakhsh E, Taylor DF, Wani S, Chy HS, Zhou Q, McKernan K, Kuersten S, Laslett AL, Grimmond SM. Deep-transcriptome and ribosome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. *Genome Res*. 2011 Dec; 21(12):2014–2025. [PubMed: 22042643]
- Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc*. 2012 Jul 26; 7(8):1534–1550. [PubMed: 22836135]

10. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*. 2010 Apr 16; 141(2):344–354. [PubMed: 20403328]
11. Cannarozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, Roth AC, Gonnet P, Gonnet G, Barral Y. A role for codon order in translation dynamics. *Cell*. 2010 Apr 16; 141(2):355–367. Erratum in: *Cell*. 2010 May 14;141(4):728. [PubMed: 20403329]
12. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*. 2011 Jan; 12(1):32–42. Review. [PubMed: 21102527]
13. Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. 2012 Mar 28; 484(7395):538–541. [PubMed: 22456704]
14. Efeyan A, Zoncu R, Sabatini DM. Amino acids and mTORC1: from lysosomes to disease. *Trends Mol Med*. 2012 Sep; 18(9):524–533. Epub 2012 Jun 28. Review. [PubMed: 22749019]
15. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010 Jul 13; 18(1):11–22. [PubMed: 20579941]
16. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, Shi EY, Stumpf CR, Christensen C, Bonham MJ, Wang S, Ren P, Martin M, Jessen K, Feldman ME, Weissman JS, Shokat KM, Rommel C, Ruggero D. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature*. 2012 Feb 22; 485(7396):55–61. [PubMed: 22367541]
17. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012 Mar 13; 13(4):227–232. [PubMed: 22411467]
18. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Paabo S, Mann M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*. 2011; 7:548. [PubMed: 22068331]
19. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002; 1(5):376–386. [PubMed: 12118079]
20. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*. 2001; 19(3):242–247. [PubMed: 11231557]
21. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics*. 2006; 5(1):144–156. [PubMed: 16219938]
22. Brewis IA, Brennan P. Proteomics technologies for the global identification and quantification of proteins. *Adv Protein Chem Struct Biol*. 2010; 80:1–44. [PubMed: 21109216]
23. Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol*. 2007; 25(1):117–124. [PubMed: 17187058]
24. de Sousa Abreu R, Penalva LO, Marcotte E, Vogel C. Global signatures of protein and mRNA expression levels. *Mol Biosyst*. 2009; 5(12):1512. [PubMed: 20023718]
25. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012; 13(4):227–232. [PubMed: 22411467]
26. Schwanhaussier B, Gossen M, Dittmar G, Selbach M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics*. 2009; 9(1):205–209. [PubMed: 19053139]
27. Schwanhaussier B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature*. 2011 May 19; 473(7347):337–342. [PubMed: 21593866]
28. Lackner DH, Schmidt MW, Wu S, Wolf DA, Bahler J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biol*. 2012; 13(4):R25. [PubMed: 22512868]

29. Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, Gasch AP. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol Syst Biol.* 2011; 7:514. [PubMed: 21772262]
30. Olsen PH, Ambros V. The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol.* 1999 Dec 15; 216(2):671–680. [PubMed: 10642801]
31. Seggerson K, Tang L, Moss EG. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene *lin-28* after translation initiation. *Dev Biol.* 2002 Mar 15; 243(2):215–225. [PubMed: 11884032]
32. Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S, Barbacioru C, Steptoe AL, Martin HC, Nourbakhsh E, Krishnan K, Gardiner B, Wang X, Nones K, Steen JA, Matigian NA, Wood DL, Kassahn KS, Waddell N, Shepherd J, Lee C, Ichikawa J, McKernan K, Bramlett K, Kuersten S, Grimmond SM. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* 2011 Dec 30.12(12):R126. [PubMed: 22208850]
33. Huntzinger E, Izaurralde E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet.* 2011 Feb; 12(2):99–110. [PubMed: 21245828]
34. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 2006 Mar 1; 20(5):515–524. [PubMed: 16510870]
35. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature.* 2010 Aug 12; 466(7308):835–840. [PubMed: 20703300]
36. Bazzini AA, Lee MT, Giraldez AJ. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science.* 2012 Apr 13; 336(6078):233–237. [PubMed: 22422859]
37. Djuranovic S, Nahvi A, Green R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. *Science.* 2012 Apr 13; 336(6078):237–240. [PubMed: 22499947]
38. Nottrott S, Simard MJ, Richter JD. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nat Struct Mol Biol.* 2006 Dec; 13(12):1108–1114. [PubMed: 17128272]
39. Stadler M, Artiles K, Pak J, Fire A. Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of *C. elegans* heterochronic miRNA targets. *Genome Res.* 2012 Oct 25.
40. Thermann R, Hentze MW. *Drosophila* miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature.* 2007 Jun 14; 447(7146):875–878. [PubMed: 17507927]
41. Ankö ML, Neugebauer KM. RNA-protein interactions in vivo: global gets specific. *Trends Biochem Sci.* 2012 Jul; 37(7):255–262. [PubMed: 22425269]
42. Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsvelde J, Hentze MW. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell.* 2012 Jun 8; 149(6):1393–1406. [PubMed: 22658674]
43. Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell.* 2012 Jun 8; 46(5):674–690. [PubMed: 22681889]
44. Tenenbaum SA, Carson CC, Lager PJ, Keene JD. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A.* 2000 Dec 19; 97(26):14085–14090. [PubMed: 11121017]
45. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP identifies Nova-regulated RNA networks in the brain. *Science.* 2003 Nov 14; 302(5648):1212–1215. [PubMed: 14615540]
46. Morris AR, Mukherjee N, Keene JD. Systematic analysis of posttranscriptional gene expression. *Wiley Interdiscip Rev Syst Biol Med.* 2010 Mar-Apr; 2(2):162–180. [PubMed: 20836020]
47. König J, Zarnack K, Luscombe NM, Ule J. Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet.* 2012 Jan 18; 13(2):77–83. [PubMed: 22251872]

48. Kapeli K, Yeo GW. Genome-wide approaches to dissect the roles of RNA binding proteins in translational control: implications for neurological diseases. *Front Neurosci.* 2012; 6:144. [PubMed: 23060744]
49. Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, Lécuyer E, Burge CB. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell.* 2012 Aug 17; 150(4):710–724. [PubMed: 22901804]
50. Komili S, Farny NG, Roth FP, Silver PA. Functional specificity among ribosomal proteins regulates gene expression. *Cell.* 2007 Nov 2; 131(3):557–571. [PubMed: 17981122]
51. Galante PA, Sandhu D, de Sousa Abreu R, Gradassi M, Slager N, Vogel C, de Souza SJ, Penalva LO. A comprehensive in silico expression analysis of RNA binding proteins in normal and tumor tissue: Identification of potential players in tumor formation. *RNA Biol.* 2009 Sep-Oct;6(4):426–433. [PubMed: 19458496]
52. Li C, Ge M, Yin Y, Luo M, Chen D. Silencing expression of ribosomal protein L26 and L29 by RNA interfering inhibits proliferation of human pancreatic cancer PANC-1 cells. *Mol Cell Biochem.* 2012 Nov; 370(1–2):127–139. [PubMed: 22868929]
53. Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ, Greengard P, Heintz N. A translational profiling approach for the molecular characterization of CNS cell types. *Cell.* 2008 Nov 14; 135(4):738–748. [PubMed: 19013281]
54. Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N. Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell.* 2008 Nov 14; 135(4):749–762. Erratum in: *Cell.* 2009 Nov 25;139(5):1022. [PubMed: 19013282]
55. Sanz E, Yang L, Su T, Morris DR, McKnight GS, Amieux PS. Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proc Natl Acad Sci U S A.* 2009 Aug 18; 106(33):13939–13944. [PubMed: 19666516]
56. Kolle G, Shepherd JL, Gardiner B, Kassahn KS, Cloonan N, Wood DL, Nourbakhsh E, Taylor DF, Wani S, Chy HS, Zhou Q, McKernan K, Kuersten S, Laslett AL, Grimmond SM. Deep-transcriptome and ribonome sequencing redefines the molecular networks of pluripotency and the extracellular space in human embryonic stem cells. *Genome Res.* 2011 Dec; 21(12):2014–2025. [PubMed: 22042643]
57. Fritsch C, Herrmann A, Nothnagel M, Szafranski K, Huse K, Schumann F, Schreiber S, Platzer M, Krawczak M, Hampe J, Brosch M. Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 2012 Nov; 22(11):2208–2218. [PubMed: 22879431]
58. Lee S, Liu B, Lee S, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A.* 2012 Sep 11; 109(37):E2424–E2432. [PubMed: 22927429]
59. Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science.* 2012 Feb 3; 335(6068):552–557. [PubMed: 22194413]
60. Gerashchenko MV, Lobanov AV, Gladyshev VN. Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci U S A.* 2012 Oct 23; 109(43):17394–17399. [PubMed: 23045643]
61. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, Mann M, Ingolia NT, Weissman JS. Decoding human cytomegalovirus. *Science.* 2012 Nov 23; 338(6110):1088–1093. [PubMed: 23180859]
62. Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA, Simonis N, Charlotiaux B, Hidalgo CA, Barbette J, Santhanam B, Brar GA, Weissman JS, Regev A, Thierry-Mieg N, Cusick ME, Vidal M. Proto-genes and de novo gene birth. *Nature.* 2012 Jul 19; 487(7407):370–374. [PubMed: 22722833]
63. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 2011 May 31;9:34. [PubMed: 21627854]

64. Mane SP, Evans C, Cooper KL, Crasta OR, Folkerts O, Hutchison SK, Harkins TT, Thierry-Mieg D, Thierry-Mieg J, Jensen RV. Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing. *BMC Genomics*. 2009 Jun 12;10:264. [PubMed: 19523228]
65. Liu S, Lin L, Jiang P, Wang D, Xing Y. A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Res*. 2011 Jan; 39(2):578–588. [PubMed: 20864445] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6; 489(7414):57–74. [PubMed: 22955616]
66. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6; 489(7414):57–74. [PubMed: 22955616]

Figure 1. Comparison of whole-transcriptome sequencing, polysome analysis, ribosome profiling and protein mass spectrometry methods to measure gene expression

Transcription of nuclear localized, non-coding and coding mRNA takes place in the nucleus of eukaryotic cells (orange box). The downward arrows represent the general path of events leading from transcription to translation of mRNA. Some RNA remain in the nucleus, but coding mRNA and some non-coding RNA are exported to the cytoplasm. Once cytoplasmic, RNA can be generally categorized into three main groups; actively translated mRNA (left), RNA subjected to some level of translational control (middle) or RNA that is targeted for degradation/decay (right). As discussed in the text, actively translated and regulated mRNA (e.g. via miRNA, localization or other types of RBPs) can be found associated with the polysome fraction of the cell. Polysomes are the cellular structures where mRNA is bound by multiple ribosomes and generally lead to protein synthesis. What complicates this view is that translational control events such as miRNA-mediated regulation also seem to be associated with polysomal structures. To help dissect these types of events, methods like polysome analysis and nuclease-treated ribosome profiling approaches can be used. The arrows pointing to the right suggest the general steps involved in performing gene 'expression' analysis by whole transcriptome RNAseq (beige box), polysomal analysis (red box), ribosome profiling (blue box) or protein mass spectrometry (green box). Whole transcriptome sequencing involves extracting all the RNA contained within the cell, regardless of location, compartmentalization, function or translational activity. Polysome

analysis involves selectively isolating the heavy fraction of mRNPs by sucrose gradient techniques or even immunoprecipitation approaches (RIP). Thus, polysome profiling is generally thought to detect the actively translating fraction of RNA in the cell. Ribosome profiling differs slightly in that nuclease is used to break apart the intact polysomal fraction into monosomes and the mRNA fragments that are protected from digestion (i.e. ribosome protect mRNA fragments) are isolated and sequenced. Protein mass spectrometry involves directly measuring the protein composition in the cell using a variety of techniques, potentially including protein labeling strategies. The results of these different approaches, despite their limitations, are all be very powerful methods to help understand and profile gene expression.

Figure 2. Example ribosome footprint

Comparison of read densities from the CWP2 gene in yeast showing a ribosome profiling sample (upper panel) and a polyA+ selected mRNAs sample (lower panel). The ribosome profiling sample results in coverage only within the open reading frame (ORF) of CWP2 and none in the 5' or 3' UTR regions. This reflects the primary regions of transcripts protected from nuclease digestion by bound ribosomes. In contrast, the mRNA sample results in coverage within both the ORF and the UTR regions, reflective of non-nuclease treated, intact RNA samples.

Figure 3. A model for miRNA-mediated gene regulation

To translate an mRNA, the 5' and 3' ends must establish communication through a network of proteins associated with the mRNP (messenger ribonucleoprotein complex)(see ref. 25 for more details). In normal translation (left side) the ribosome subunits (green) quickly initiate translation to produce nascent peptides. These transcripts will produce Ribosome Protected Fragments coincident with the density of associated ribosomes. In contrast, when a target mRNA is regulated by miRNA (right side) there is a disruption in the communication between the 5' and 3' ends which leads to an order of events, such as 1) Translational Repression; 2) Deadenylation and 3) eventual turnover or decay of the target mRNA. In these cases, the relative amounts of RPFs produced from miRNA regulated targets is lower than in the absence of the miRNA.