Review Heterogeneity in mRNA Translation

Stijn Sonneveld,^{1,2} Bram M.P. Verhagen,^{1,2} and Marvin E. Tanenbaum^{1,*}

During mRNA translation, the genetic information stored in mRNA is translated into a protein sequence. It is imperative that the genetic information is translated with high precision. Surprisingly, however, recent experimental evidence has demonstrated that translation can be highly heterogeneous, even among different mRNA molecules derived from a single gene in an individual cell; multiple different polypeptides can be produced from a single mRNA molecule and the rate of translation can vary in both space and time. However, whether translational heterogeneity serves an important cellular function, or rather predominantly represents gene expression 'noise' remains an open question. In this review, we discuss the molecular basis and potential functions of such translational heterogeneity.

Mechanisms of Translational Control

The fate and function of each cell depends on its protein composition. Therefore, accurate gene expression control is critical for proper cell functioning. A key step in gene expression is mRNA translation (see Figure I in Box 1), during which the genetic information stored in mRNA is decoded. Regulation of translation can affect both protein sequence and abundance [1]. Moreover, translational regulation is fast, reversible and provides spatial control, making it a unique regulatory mechanism of gene expression. In a simplistic view, each mRNA molecule encodes a single protein and mRNAs derived from the same gene are decoded in the same way. However, it is now clear that translation is far more heterogeneous. Different types of translational heterogeneity can be distinguished. First, a single mRNA species can be translated differentially in different cell types, resulting in 'cell-to-cell' translational heterogeneity, which is important during various cellular processes such as differentiation [2]. Second, mRNA molecules originating from different genes can be differentially translated in a single cell, resulting in 'intergenic' translational heterogeneity. Third, mRNA molecules originating from a single gene in a single cell can also display translational heterogeneity, which we refer to as 'intragenic' heterogeneity. Finally, a single mRNA molecule can also be translated differentially over time, representing a special case of intragenic heterogeneity. While regulatory functions for cell-to-cell and intergenic translational heterogeneity are well known, potential functions and mechanisms of regulation for intragenic translational heterogeneity are less evident.

In this review, we focus on intragenic translational heterogeneity, a field that has rapidly emerged in recent years in part due to the development of new methods that provide sufficient detection sensitivity to study the translation of single mRNA molecules. First, we describe experimental evidence supporting the existence of intragenic translational heterogeneity as well as its potential functions. Next, we discuss the molecular mechanisms underlying intragenic translational heterogeneity. We also discuss whether intragenic translational heterogeneity represents (functional) translational regulation or whether it is a consequence of the variability inherent to translational regulation and mainly reflects 'noise' in the system.

Experimental Evidence for Heterogeneity in Translation

Two types of intragenic translational heterogeneity can be distinguished: heterogeneity in the amino acid sequence of newly synthesized proteins (Figure 1A) and heterogeneity in the protein

Highlights

Different types of translational heterogeneity can be distinguished: cell or tissue heterogeneity (translation of a gene differs in distinct cell or tissue types), 'intergenic' heterogeneity (mRNAs derived from different genes are translated differentially), and 'intragenic' heterogeneity (different mRNAs derived from one gene in a single cell show translational heterogeneity).

Recent studies using single-molecule imaging have revealed the widespread occurrence of intragenic translational heterogeneity.

Intragenic translational heterogeneity can have multiple origins, including heterogeneity in primary mRNA sequence, RNAbinding proteins, RNA modifications, mRNA structure, and ribosome composition.

Translational heterogeneity is likely to represent 'noise' as well as spatiotemporal regulation of translation.

¹Oncode Institute, Hubrecht Institute – KNAW and University Medical Center Utrecht, Utrecht, The Netherlands ²These authors contributed equally to this work

*Correspondence: m.tanenbaum@hubrecht.eu (M.E. Tanenbaum).





synthesis rate (Figure 1B). In this section, we discuss the experimental evidence and function of both types of translational heterogeneity.

Heterogeneity in Protein Sequence

Heterogeneity in protein sequence can be caused by alternative translation start site selection, ribosome frameshifting during translation elongation, or stop codon readthrough (RT) (Figure 1A). Here, we focus on translation start site heterogeneity and stop codon RT, as these processes are best described and likely to represent the predominant mechanisms underlying translational heterogeneity in eukaryotic cells.

5'-to-3' directed scanning along the mRNA by the 43S preinitiation complex (PIC) generally results in translation initiation at the first (i.e., most 5') AUG codon (see Figure IB in Box 1). The detection efficiency of AUG codons depends on the surrounding sequences, and AUG codons are most effectively recognized as a translation initiation site (TIS) when surrounded by the Kozak consensus sequence (GCCACCAUGG) [3]. However, translation initiation can also occur at non-AUG codons, such as GUG or CUG [4]. Moreover, a PIC can fail to initiate at an AUG codon when encountering it during scanning (a process termed leaky scanning). Several lines of evidence have demonstrated that heterogeneity in start site usage is indeed apparent in cells (see Box 2 for further information on the type of methods used to assess translational heterogeneity). Ribosomal profiling revealed that at least half of the human mRNA transcripts contain more than one TIS [4,5] and in a small number of cases proteins synthesized from an alternative TIS have been detected by mass spectrometry [6–10].

Alternative TISs can be used by the cells to drive isoform-specific expression of a protein. In-frame upstream and downstream initiation from alternative start sites leads to protein N-terminal extensions and N-terminal truncations, respectively, which may affect protein localization or function in the cell [9,11]. The N-terminal extended protein isoform of the c-myc gene, for example, that stems from an upstream alternative initiation event has enhanced DNA-binding capacity that drives the expression of cell cycle repressors [11]. While the example of c-myc clearly demonstrates the

Box 1. Mechanism of Translation Initiation, Elongation, and Termination

(1) Initiation

The 43S PIC comprises the small (40S) ribosomal subunit, GTP-bound eIF2, and the initiator methionine tRNA (Met-tRNA) along with several other translation initiation factors (Figure IB). The PIC is recruited to the 7-methylguanosine cap of an mRNA, a process that is mediated by an interaction of cap-bound eIF4E, eIF4G, and eIF4A translation initiation factors and the PIC. After recruitment to the cap, the PIC scans the mRNA from 5' to 3' in search of a start codon (Figure IB). Upon base pairing between the start codon and the Met-tRNA, the PIC undergoes a conformational change, resulting in the recruitment of the large (60S) ribosomal subunit and the start of the translation elongation phase (Figure IB).

(2) Elongation

During elongation, the ORF is decoded in steps of three nucleotides (one codon) to synthesize a polypeptide chain. The elongation cycle starts with a peptidyl-tRNA in the P site and an empty aminoacyl and exit site (A and E site, respectively) (Figure IC, step i). First, an aminoacyl-tRNA is selected at the empty A site by base pairing of the tRNA anticodon and the mRNA codon (Figure IC, step ii). Next, a peptide bond is formed in the peptidyl transferase center between the nascent peptide and the aminoacyl-tRNA, resulting in a deacylated-tRNA and a new peptidyl-tRNA, respectively (Figure IC, step iii). Finally, the tRNAs adopt a hybrid state (Figure IC, step iv), followed by eEF2-mediated translocation of the deacylated-tRNA and the new peptidyl-tRNA to the E and P site, respectively (Figure IC, step v). The elongation cycle is repeated until a stop codon is encountered.

(3) Termination and Recycling

Translation is terminated upon the recognition of a stop codon in the A site by the release factors eRF1 and eRF3 (Figure ID). After binding to the stop codon in the A site, eRF1 induces release of the polypeptide and stimulates recruitment of the ribosome recycling factor ABCE1, which in turn results in dissociation of the ribosome subunits from the mRNA.





Figure I. Mechanism of Translation Initiation, Elongation, and Termination. (A) Overview of the different phases of translation. (B–D) Schematic of translation initiation (B), translation elongation (C), and translation termination and recycling (D) (see Box 1 text for details).

possibility of precise regulation of gene expression through alternative TIS usage, it is unclear whether the widespread occurrence of alternative TIS usage contributes to functional proteome diversification. Alternative TIS usage could also act as part of a regulatory mechanism to deflect ribosomes away from the main open reading frame (ORF). Translation from alternative TISs may result in 'junk' polypeptides that are produced as a result of infidelity of the translation initiation machinery. This is especially the case for alternative TISs that are out of frame with respect to the main protein-coding ORF, which generally result in the synthesis of short peptides with an amino acid sequence unrelated to the main protein-coding ORF. Additional work is required to catalog the expression of such alternative protein products and identify functions for these (poly)peptides. In this context, it is





Figure 1. Intragenic Translational Heterogeneity. Intragenic translational heterogeneity can be divided into (A) heterogeneity in protein sequence and (B) heterogeneity in protein synthesis rate. (A) Different mRNA molecules are shown that produce proteins with different amino acid sequences due to distinct mechanisms: (i) canonical translation; (ii) alternative translation initiation; (iii) ribosome frame shifting; and (iv) stop codon readthrough. (B) Two mRNAs are depicted that have distinct synthesis rates of the protein encoded in the canonical open reading frame (ORF), resulting in translational heterogeneity between the two mRNA molecules (high and low synthesis rates for top and bottom mRNAs, respectively). Protein synthesis rates of the canonical ORF are affected by: (i) 43S ribosome recruitment rates; (ii) the presence of upstream ORFs (uORFs); and (iii) ribosomal stalling in the transcript.



Box 2. Methods for Measuring Translational Heterogeneity

Mass Spectrometry

Mass spectrometry allows direct detection of distinct protein isoforms; for example, those arising from alternative TISs, stop codon RT, or the translation of alternatively spliced isoforms. Identification of distinct protein isoforms using mass spectrometry is limited by the detection sensitivity of the instrument, which can be challenging for the detection of rare and/or unstable translation products. To enhance the detection of alternative translation products, multiple enrichment strategies have been developed, such as enrichment of protein N termini for the identification of alternative TISs [96,97]. While distinct translation products can be identified by mass spectrometry, single-cell analysis has not yet been achieved, making it challenging to study intragenic translational heterogeneity in single cells using this method.

Ribosome Profiling

Ribosome profiling is a sequencing-based method used to identify the precise position of ribosomes along mRNA by sequencing the ribosome-protected mRNA fragments [19]. Ribosome profiling is well suited to study intergenic translational heterogeneity in a genome-wide fashion, as it enables the identification of TISs, translation of the noncoding regions of the transcript, and measurements of translation efficiency. Since ribosome profiling is an ensemble method, it requires averaging of thousands of mRNAs and cells, so measurements of the translation of different mRNA molecules in single cells are challenging.

Single-Molecule Imaging Methods

Recently, several methods have been developed that enable visualization of translation of single mRNA molecules [21–26,98]. Most of these methods rely on the introduction of multiple copies of a short peptide sequence (e.g., the SunTag [99]) into the coding sequence of a gene of interest. Upon translation, the short epitope sequence is synthesized and bound by a fluorescently labeled antibody that is stably expressed in the cell, resulting in a bright fluorescence signal that reports on the translation efficiency of individual mRNA molecules [100]. Through direct visualization of translation, heterogeneity has been observed in many aspects of mRNA translation and decay [16,21–26,101–104]. Furthermore, use of multiple tags (e.g., SunTag and MoonTag or Frankenbody [16,103]) allows quantification of translation of multiple ORFs at a single molecule resolution. Therefore, single-molecule imaging methods are uniquely suited to study intragenic translational heterogeneity in single cells. A drawback of these methods is the low throughput (only one or a couple of genes can be analyzed at a time) and the need to introduce bulky tags into the gene of interest, which has so far mostly limited analysis to reporter genes, although endogenously tagged genes are becoming available [23].

interesting to note that prokaryotes use the Shine–Dalgarno (SD) sequence, which is located directly upstream of the main TIS, to guide ribosomes directly to the correct site of initiation, rather than using a scanning mechanism for TIS identification (reviewed in [12]). It is possible that TIS identification by scanning provides more flexibility for alternative TIS selection and, if so, would have evolved only if alternative TIS selection is functionally important (and thus not only noise).

A second process that can cause protein sequence heterogeneity is translation stop codon RT. Stop codon recognition is generally efficient and results in translation termination. However, in rare cases the stop codon can be decoded as a sense codon, resulting in a Cterminally extended protein. Estimates of stop codon RT vary from 0.01% to 0.1% in mammalian cells for most genes (with the UGA stop codon being the most prone to induce RT [13]), although in some examples the RT frequency can exceed 30% of translation termination events [14]. Recent studies in both mammalian cells and bacteria used live-cell imaging to directly visualize translation of the 3' untranslated region (UTR). Surprisingly, both studies found that 3' UTR translation was highly heterogeneous between [15] and even within [16] individual cells. Furthermore, a genome-wide study revealed extensive changes in RT for many mRNAs during Drosophila development [17]. While there is limited understanding of the mechanisms controlling RT, one recent paper on the RT of the AGO1 mRNA stop codon revealed that RT is enhanced by miRNA binding downstream of the stop codon [18]. It will be interesting to learn whether this or a similar mechanism also acts on other genes. Together, these studies show that RT is highly heterogeneous and is likely to be a regulated process.



Heterogeneity in Protein Synthesis Rate

A second type of translational heterogeneity is heterogeneity in the translation rate (i.e., the number of proteins synthesized from a single mRNA molecule per unit time) (Figure 1B). Transcriptome-wide methods, such as ribosome profiling [19], are widely used to study the translation rate of individual genes and can readily uncover cell-to-cell and intergenic heterogeneity in the translation rate (Box 2). Furthermore, a recent genome-wide study using polysome profiling demonstrated that different transcript isoforms can often have distinct translation rates [20], suggestive of intragenic translational heterogeneity. More recently, new single-molecule imaging approaches (Box 2) have revealed intragenic translation rate heterogeneity for individual mRNAs as well as temporal fluctuations in the translation rate of single mRNAs [21–26], further confirming the existence of intragenic translation rate heterogeneity.

Since translation initiation is generally the rate-limiting step for protein synthesis, heterogeneity in the translation rate is likely to originate predominantly at the initiation step. Differential recruitment rates of the PIC to the 5' cap can cause translation rate heterogeneity. In addition, the translation rate of the main ORF can be affected by the usage of alternative TISs. For example, (short) upstream ORFs (uORFs) often reduce translation of the main ORF by promoting translation initiation of the uORF followed by ribosome recycling after termination at the uORF stop codon, preventing initiation at the main ORF TIS [27]. Importantly, differential usage of uORFs between distinct mRNA molecules may result in intragenic translation rate heterogeneity, providing a direct link between translation start site heterogeneity and translation rate heterogeneity [16]. In addition to the translation initiation step, variability in the translation elongation rate (e.g., due to ribosome pausing) may contribute to intragenic translation rate heterogeneity altors pausing) may contribute to intragenic translation rate heterogeneity and translation rate (e.g., due to ribosome pausing) may contribute to intragenic translation rate heterogeneity altors pausing) may contribute to intra-

Intragenic translation rate heterogeneity can represent spatial control over the translation rate. For example, in neuronal cells certain transcripts are translated only locally in axons (see [28] for a review). In addition, translation rate heterogeneity could be a consequence of cell-wide translation rate regulation; to achieve a cell-wide reduction in translation rate, either the translation rate of each mRNA molecule is reduced partially or the translation of a subset of mRNAs is reduced severely while other mRNAs are translated normally. Experimental evidence suggests that the latter mechanism may occur, at least under some conditions. For example, on inhibition of mTOR signaling, the translational repressor 4E-BP binds to elF4E associated with individual mRNAs [29], which is likely to result in complete inhibition of translation of those mRNAs, while mRNAs that are not bound by 4E-BP are likely to be unaffected (at least at short timescales). In these examples, intragenic translation rate heterogeneity is simply a consequence of the stochastic nature of translation initiation. For example, the translation rate may fluctuate as translation factors bind and release stochastically from single mRNAs. Dissecting the contributions of active regulation and stochastic events to translation rate heterogeneity is an important future goal.

Molecular Mechanisms of Translational Heterogeneity

In the previous sections we discussed emerging evidence of intragenic translational heterogeneity. In the following sections, we provide an overview of different possible mechanisms that could cause translational heterogeneity (Figure 2).

mRNA Primary Sequence

Many eukaryotic genes encode multiple mRNA isoforms that differ in their primary nucleotide sequence. A median of six transcript isoforms has been detected per human protein-coding gene (GENCODE, Release 33) and different transcript isoforms can coexist in a single cell [30–33]. Primary mRNA sequence variations can originate from variable transcription start site (TSS) usage,





Figure 2. Molecular Mechanisms of Translational Heterogeneity. Overview of the potential origins of intragenic translational heterogeneity. Two mRNA molecules are shown that differ in: (i) the composition of the translating ribosome; (ii) mRNA structure; (iii) mRNA sequence; (iv) nucleotide modifications; and (v) RNA-binding proteins (RBPs) bound to the mRNA molecule. Together these different features drive translational heterogeneity.

alternative usage of polyadenylation sites, or through alternative splicing. The possible contributions of these types of sequence variability to intragenic translational heterogeneity are discussed below.

TSS mapping using CAGE in over 900 human cell lines and tissues uncovered more than 1 million TSSs [34]. An alternative TSS creates an alternative 5' UTR and therefore may affect the translation rate of the transcript isoform [35]. Indeed, several studies have identified hundreds of genes for which transcripts with alternative 5' UTRs showed distinct translation rates [20,36]. How differences in 5' UTRs result in different translation rates is not completely understood. Differential inclusion of a uORF or alternative TISs in a subset of transcript isoforms could contribute to the observed differences in translation rate. However, additional mechanisms may also exist; one study combined a massively parallel reporter assay and advanced computing and identified multiple new sequence motifs that regulate the translation initiation efficiency of 5' UTRs [37], although the mechanism by which these motifs affect translation remains to be explored. Alternative TSS usage can also affect protein sequence. For example, if the TSS is located downstream of the TIS, the mRNA will generate a truncated protein. Similarly, an upstream TSS could introduce additional TISs in the 5' UTR that result in either short peptides or N-terminally extended proteins [36]. Thus, 5' UTR heterogeneity occurs for many, if not most, genes due to alternative TSS usage and may be an important driver of intragenic translational heterogeneity.

Sequence analysis of exon–exon boundaries in human cells revealed that most, if not all, multiexon genes undergo alternative splicing [38], and in many cases differentially spliced transcripts appear to coexist in individual cells [30]. Differential splicing in a single cell may be caused by imperfect recognition of the splice sites by the spliceosome or weak/partial regulation of alternative splicing, such that a subset of mRNA molecules escapes the regulation. Most events of alternative splicing occur in the coding sequence and thus mainly cause protein sequence heterogeneity rather than translation rate heterogeneity. Nonetheless, a subset of splicing events, especially those in the 5' UTR that impact the 5' UTR sequence, can also cause translation rate heterogeneity [20].



During transcription, the 3' end of mRNAs is processed through cleavage of the pre-mRNA followed by the addition of a polyA tail. pre-mRNA cleavage occurs downstream of a cleavage and polyadenylation sequence (e.g., AAUAAA) and is regulated by a number of RNA-binding proteins (RBPs) [39]. Interestingly, over half of the human and mouse genes express multiple transcript isoforms that differ in their 3' UTR sequence due to cleavage and polyadenylation at alternative sites [40]. Even within single cells, multiple transcript isoforms are often detected that differ in the site of polyadenylation [32]. Why different pre-mRNA molecules originating from a single gene are processed differently is largely unclear. It is possible that differentially polyadenylated mRNAs were produced at different times in the cell's life cycle when distinct regulatory processes were active. Alternatively, determination of the polyadenylation site may be stochastic, with different probabilities of usage for different polyadenylation sites. Alternative usage of polyadenylation sites generally alters the sequence of the 3' UTR of mRNAs. Since many RBPs and miRNAs, both of which can affect mRNA fate, bind to the 3' UTR, an alternative 3' UTR sequence may lead to an altered translation rate. Moreover, interactions between RBPs and microtubule motor proteins can control the localization of an mRNA in the cell [40], and a recent study found that hundreds of transcript isoform pairs with different polyadenylation site usage showed distinct localizations in the brain [41]. Since the localization of a transcript can affect its translation (especially in neurons), alternative localization of specific mRNA isoforms may also affect their translation.

In summary, most genes express multiple transcript isoforms with distinct primary sequences, often within a single cell. Distinct transcript isoforms frequently show different translation rates due to inclusion or exclusion of specific regulatory sequences or alternative translation start sites, and, when the ORF sequence is affected, protein sequence is likely to be altered as well.

mRNA Modifications

Information in mRNA is stored not only in the primary sequence but also through posttranscriptional modifications of RNA nucleotides. Eleven distinct nucleotide modifications have been described for mRNAs, which collectively have been termed the 'epitranscriptome' [42]. The most prevalent modification is methylation of the adenosine base at the nitrogen-6 position (m⁶A), with an average occurrence of one to three modified adenosines per mRNA [42]. While other modifications are involved in translational regulation as well, we focus our discussion on the m⁶A modification because it is the most abundant and best characterized modification on mRNAs. A plethora of studies has shown that m⁶A modifications affect mRNA translation. First, m⁶A modifications can increase the translation initiation rate by directly recruiting eIF3, METTL3, or YTHDF1 to the mRNA [43–46]. Second, it has been reported that m⁶A present in the coding sequence inhibits translation [47–49], possibly by slowing down translation elongation [50]. Thus, m⁶A plays an important regulatory role during translation.

Several studies have performed transcriptome-wide mapping of m⁶A sites to uncover the positions of m⁶A modifications in all mRNAs [51–56]. Interestingly, detailed quantitative analysis of individual m⁶A sites revealed that, for most individual m⁶A sites, only a subset of mRNA molecules contain the modification [57]. This 'm⁶A stoichiometry' (i.e., the fraction of mRNA molecules that contains a m⁶A modification at a specific site) can range from 7% to 77% depending on the modification site and cell line [57]. Recently, a new technique, MAZTER-seq, has further expanded the analysis of m⁶A stoichiometry to allow the measurement of m⁶A stoichiometry for a substantial fraction of all m⁶A sites in the transcriptome. This study confirmed that, in most cases, only a subset of mRNA molecules is modified at any given site ('intrasite heterogeneity'). In addition, substantial variability exists in m⁶A stoichiometry for different sites ('intersite



heterogeneity') [58]. Intersite m⁶A heterogeneity (i.e., different stoichiometries for different m⁶A sites) depends on local sequence context and RNA structure, suggesting that the specific stoichiometry of a m⁶A site is, at least in part, 'hard coded' in the mRNA sequence [58]. By contrast, intrasite m⁶A heterogeneity is likely to be caused by variability in m⁶A deposition and/or removal. For example, some mRNAs show decreased m⁶A methylation levels after transport to axons, where these mRNA molecules are locally translated [49]. Taken together, these studies reveal that m⁶A modifications are highly heterogeneous and are a potential source of intragenic translational heterogeneity.

mRNA Structure

RNA can form intricate higher-order structures, adding another layer of information to RNA molecules. RNA structure is formed by Watson–Crick base pairing and can be further stabilized through other RNA interactions, such as sugar-backbone interactions. Structure in mRNAs can affect translation in multiple ways. First, mRNA structures can inhibit translation initiation by physically blocking the PIC during scanning [59]. Second, mRNA structure can also promote translation initiation; for example, through recruitment of the translation initiation factor eIF3 to specific RNA hairpin structures [60]. Moreover, mRNA structure can stimulate translation initiation from noncanonical start codons through stalling of the PIC during scanning [61,62] or induce noncanonical translation initiation through direct recruitment of ribosomes to internal ribosome entry sites (IRESs) [63]. Finally, RNA structures in the coding sequence can stall ribosomes during translation elongation [64,65]. Collectively, these examples illustrate that mRNA structures have a profound impact on translation.

In recent years, many methods have been developed that use chemical probing combined with deep sequencing to investigate the structure of mRNA molecules *in vivo* (see [66] for a comprehensive review). While such methods have provided a wealth of new information on mRNA structures *in vivo*, these methods often provide only an ensemble mRNA structure. To understand whether mRNA structural heterogeneity contributes to translational heterogeneity, methods are required that resolve all the different structures of a (full length) mRNA, preferably even in a single cell. Although this is challenging, new computational methods, analogous to methods used for NMR data analysis, have been developed to determine individual mRNA structures in chemical probing data sets [67–69]. Through these approaches, a recent study suggested that *ACTB* mRNA adopts multiple structural conformations *in vivo*, affecting the accessibility of the binding site of a protein [70]. Thus, mRNA molecules are likely to adopt different mRNA structures *in vivo*, which may be an important contributor to translational heterogeneity.

RNA structure is likely to be highly dynamic, such that a single mRNA molecule can adopt multiple different conformations over time *in vivo*, which may further contribute to structural heterogeneity. Several studies have shown that mRNA molecules can adopt different conformations during their lifetime; for example, as mRNAs translocate from the nucleus to the cytoplasm or during their translation by ribosomes [71–74]. Moreover, our recent work using single-molecule imaging uncovered substantial structural dynamics at short timescales (s to min) as well [75]. Such dynamic changes in mRNA structure can affect binding site accessibility of regulatory proteins with the mRNA, possibly affecting translation. Similarly, an *in vitro* study found that bacterial mRNA molecules constantly refold into different conformations, affecting the accessibility of the SD sequence over time, which, *in vivo*, may result in temporal translational heterogeneity [76]. Thus, to understand the contribution of mRNA structure to translational heterogeneity, it is necessary to assess both the suite of mRNA structures that are adopted by different mRNA molecules and the dynamics of these mRNA structures.

Trends in Cell Biology



RBPs

RBPs regulate the fate of mRNAs in multiple ways, including the regulation of mRNA stability, localization, and translation. Most RBPs associate with the mRNA through one or multiple welldefined RNA-binding domains (RBDs) to form ribonucleoprotein complexes (RNPs). RBDs often bind a relatively short sequence motif of around two to eight nucleotides. In a number of cases, the binding affinities of a RBP and RNA have been measured in vitro, which revealed dissociation constants ranging from low nanomolar (strong binder) to micromolar [77-80]. Affinities in the nanomolar to micromolar range will generally result in interaction half-lives in the (m)s to min range [78,81,82]. Since a typical mRNA molecule has a half-life of several hours [1], the duration of these interaction half-lives suggest that the protein composition of every mRNP is constantly changing, resulting in a high intragenic mRNP heterogeneity. The degree of mRNP heterogeneity depends on RBP concentration as well, as at high RBP concentrations binding sites may become saturated. Additional intragenic heterogeneity in mRNP composition is expected as a result of competition between RBPs that bind the same, or overlapping, binding sites on an mRNA. While experimental evidence for intragenic heterogeneity in mRNP composition is limited due to technical limitations in detecting interactions between a single mRNA and protein molecules, the existence of mRNP compositional heterogeneity can be deduced from RBP-mRNA interaction durations and the mRNA half-life, and may have a major effect on translational heterogeneity.

While heterogeneity in mRNP composition is likely to exist for all mRNAs due to stochastic binding and unbinding of RBPs to individual mRNAs, mRNA compositional heterogeneity can also be due to active regulation. For example, during early development *caudal* mRNA is translationally active only at the posterior side of the embryo due to the asymmetric distribution of the translational repressor bicoid [83]. Even in nonpolarized somatic cells, spatial distribution of mRNAs can cause translational heterogeneity; for example, a recent study identified a new membraneless organelle, called a TIS granule (assemblies of the protein TIS11B), which resides adjacent to the endoplasmic reticulum (ER) [84]. The localization of mRNAs to TIS granules depends on RBPs, including HuR, that selectively bind specific mRNA transcripts [85]. Thus, translational heterogeneity can arise through heterogeneous localization of the mRNA transcripts, which can be mediated by RBPs.

Ribosome Heterogeneity

The protein composition of the ribosome is generally considered invariant. However, several studies have suggested that some degree of heterogeneity may exist in the composition of the ribosome and that distinct ribosomes may be capable of performing unique functions (reviewed in [86,87]).

Heterogeneity in ribosome composition is supported by gene expression analysis of ribosomal genes across various tissues and cell lines [88,89], which revealed that up to 25% of the ribosomal genes are differentially expressed [89]. Moreover, several paralogs of ribosomal proteins (RPs) are exclusively expressed in one tissue or cell type [88]. Given the structural similarity between RPs and their paralogs, RP paralogs may substitute canonical RPs during ribosome assembly [90,91]. Exclusion of canonical RPs or incorporation of an RP paralog may result in an 'alternative ribosome', potentially capable of performing specific functions (e.g., translating a subset of mRNA molecules differently). Indeed, four RPs were recently identified to be substoichiometric in mouse embryonic stem cells, suggesting that even within single cells some ribosomes contain these RPs while others do not. Ribosome profiling using ribosomes containing these RPs [92]. In another example, haploinsufficiency of RPL38 causes severe developmental defects, originating from impaired IRES-mediated translation of a specific set of HOX genes [92,93]. While alternative ribosomes may preferentially translate specific transcripts, some genes may also be more sensitive than



others to the availability of ribosomes (i.e., when the ribosome number is reduced), resulting in gene-specific changes in translation efficiency [94]. Ribosomes were also found to associate with hundreds of additional proteins [95], providing another source of potential heterogeneity in ribosome function. One of these ribosome-associated proteins, PKM, was identified as a factor specifically involved in the translation of ER-associated mRNAs, suggesting that the compositional heterogeneity of ribosome-associated proteins can also contribute to spatial heterogeneity in translation [95].

Concluding Remarks

The main function of translation is to decode the genetic information that is stored in the DNA. In this light, translation might be expected to be highly homogeneous, as the information stored in mRNA should be faithfully translated to protein. However, recent experimental evidence has demonstrated that mRNA molecules originating from the same gene can produce different amounts of protein and can synthesize entirely different polypeptides, suggesting that translation is highly heterogeneous. The development of new tools to visualize mRNA translation of individual mRNA molecules in space and time provides exciting new opportunities to study translational heterogeneity and will hopefully help to address the many questions that remain unanswered (see Outstanding Questions).

To what extent translational heterogeneity is functionally important remains a central question. Translational heterogeneity may be a consequence of temporal or spatial regulation of translation. Additionally, translational heterogeneity may contribute to functional proteome diversification [9]. On the other hand, translational heterogeneity could be a side-effect of the stochastic nature of the processes underlying mRNA translation. While translational control and functional proteome diversification are important for cellular function, stochasticity in translation may result in 'unwanted' heterogeneity that can result in the production of aberrant or even toxic proteins. We speculate that the flexibility required for complex translational regulation, for example, the synthesis of multiple functional protein isoforms from a single mRNA, may have inadvertently resulted in unwanted translational heterogeneity as well. For example, regulation of the translation start site requires 'flexible' rules for translation initiation, which may also induce stochastic translation initiation at nonfunctional sites and thus the production of aberrant proteins. It will be of great interest to further study how maximal regulatability is achieved while unwanted heterogeneity is minimized.

As discussed in this review, many different mechanisms can underlie intragenic translational heterogeneity and it is currently unclear what the contribution of individual processes (e.g., RNA structural dynamics, m⁶A stoichiometry) is to the overall levels of heterogeneity. Moreover, it is unclear whether the heterogeneity caused by each of these processes individually mostly results in functional heterogeneity or mostly represents unwanted noise. Finally, a direct, causal link between the observed intragenic translational heterogeneity and its potential molecular origins requires new methods that can quantify translation of single mRNA molecules and simultaneously measure different features (e.g., m⁶A modification, primary sequence) of the same mRNA molecule. A better understanding of the origins and potential functions of translational heterogeneity will provide a full picture of the mechanisms underlying the decoding process of genetic information, a central process in life.

Acknowledgments

We thank members of the Tanenbaum laboratory for helpful discussions. This work was financially supported by an ERC Starting Grant (ERC-STG 677936-RNAREG) and by the Howard Hughes Medical Institute through an International Research Scholar grant to M.E.T. (HHMI/IRS 55008747). M.E.T. was also financially supported by the Oncode Institute.

Outstanding Questions

While different mechanisms can contribute to translational heterogeneity, the relative contributions of each mechanism to overall translational heterogeneity remain to be determined.

How widespread is intragenic translational heterogeneity on endogenous genes? Several studies have found extensive translational heterogeneity on reporter genes. However, the extent of translational heterogeneity on native mRNAs has not been explored. It will be interesting to determine whether different mRNA species show distinct levels of intragenic translational heterogeneity.

mRNA molecules can have differing mRNA sequences, RBP composition, nucleotide modification status, or mRNA structure. Does the large number of heterogeneous parameters mean that (almost) every mRNA molecule is unique or do subgroups of mRNA molecules exist due to coregulation of different parameters?

What are the timescales on which the translation of individual mRNA molecules changes? The answer to this question is likely to vary for each type of translational heterogeneity.

To what extent are heterogeneous mRNA features actively regulated? Regulation of mRNA features may result in spatiotemporal regulation of translation, while unregulated mRNA features may contribute to translation 'noise'.



References

- Schwanhausser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
 Buszczak, M. *et al.* (2014) Cellular differences in protein syn-
- Buszczak, M. et al. (2014) Cellular offerences in protein synthesis regulate tissue homeostasis. Cell 159, 242–251
- Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292
- Ingolia, N.T. et al. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147, 789–802
- Lee, S. et al. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proc. Natl. Acad. Sci. U. S. A. 109, 2424–2432
- Na, C.H. et al. (2018) Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Res. 28, 25–36
- Van Damme, P. et al. (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol. Cell. Proteomics* 13, 1245–1261
- Yeom, J. et al. (2017) Comprehensive analysis of human protein N-termini enables assessment of various protein forms. Sci. Rep. 7, 6599
- Chen, J. et al. (2020) Pervasive functional translation of noncanonical human open reading frames. Science 367, 1140–1146
- Martinez, T.F. *et al.* (2020) Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* 16, 458–468
- Hann, S.R. et al. (1992) Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. Genes Dev. 6, 1229–1240
- 12. Rodnina, M.V. (2018) Translation in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 10, a032664
- Dabrowski, M. *et al.* (2015) Translational readthrough potential of natural termination codons in eucaryotes – the impact of RNA sequence. *RNA Biol.* 12, 950–958
- Loughran, G. et al. (2014) Evidence of efficient stop codon readthrough in four mammalian genes. Nucleic Acids Res. 42, 8928–8938
- Fan, Y. *et al.* (2017) Heterogeneity of stop codon readthrough in single bacterial cells and implications for population fitness. *Mol. Cell* 67, 826–836
- Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472
- Dunn, J.G. *et al.* (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster. Elife* 2, e01179
- Singh, A. et al. (2019) Let-7a-regulated translational readthrough of mammalian AGO1 generates a microRNA pathway inhibitor. EMBO J. 38, e100727
- Ingolia, N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* 15, 205–213
- Floor, S.N. and Doudna, J.A. (2016) Tunable protein synthesis by transcript isoforms in human cells. *Elife* 5, e10921
- 21. Yan, X. et al. (2016) Dynamics of translation of single mRNA molecules in vivo. Cell 165, 976–989
- Wu, B. *et al.* (2016) Translation dynamics of single mRNAs in live cells and neurons. *Science* 352, 1430–1435
- Pichon, X. et al. (2016) Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. J. Cell Biol. 214, 769–781
- Morisaki, T. *et al.* (2016) Real-time quantification of single RNA translation dynamics in living cells. *Science* 352, 1425–1429
- Wang, C. *et al.* (2016) Real-time imaging of translation on single mRNA transcripts in live cells. *Cell* 165, 990–1001
- Halstead, J.M. et al. (2015) Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science 347, 1367–1671
- Johnstone, T.G. et al. (2016) Upstream ORFs are prevalent translational repressors in vertebrates. EMBO J. 35, 706–723
- Glock, C. et al. (2017) mRNA transport & local translation in neurons. Curr. Opin. Neurobiol. 45, 169–177

- Peter, D. et al. (2015) Molecular architecture of 4E-BP translational inhibitors bound to eIF4E. Mol. Cell 57, 1074–1087
- Song, Y. et al. (2017) Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. Mol. Cell 67, 148–161
- Shalek, A.K. et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240
- Velten, L. et al. (2015) Single-cell polyadenylation site mapping reveals 3' isoform choice variability. Mol. Syst. Biol. 11, 812
- Waks, Z. et al. (2011) Cell-to-cell variability of alternative RNA splicing. Mol. Syst. Biol. 7, 506
- Kanamori-Katayama, M. *et al.* (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* 21, 1150–1159
- Hinnebusch, A.G. *et al.* (2016) Translational control by 5'untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416
- Wang, X. et al. (2016) Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol. Syst. Biol.* 12, 875
- Sample, P.J. et al. (2019) Human 5' UTR design and variant effect prediction from a massively parallel translation assay. Nat. Biotechnol. 37, 803–809
- Wang, E.T. *et al.* (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
- Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30
- Mayr, C. (2016) Evolution and biological roles of alternative 3' UTRs. Trends Cell Biol. 26, 227–237
- Tushev, G. et al. (2018) Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* 98, 495–511
- 42. Zaccara, S. et al. (2019) Reading, writing and erasing mRNA methylation. Nat. Rev. Mol. Cell Biol. 20, 608–624
- Meyer, K.D. et al. (2015) 5' UTR m⁶A promotes capindependent translation. Cell 163, 999–1010
- Zhou, J. et al. (2015) Dynamic m⁶A mRNA methylation directs translational control of heat shock response. *Nature* 526, 591–594
- Choe, J. et al. (2018) mRNA circularization by METTL3-elF3h enhances translation and promotes oncogenesis. *Nature* 561, 556–560
- 46. Wang, X. et al. (2015) N^6 -Methyladenosine modulates messenger RNA translation efficiency. Cell 161, 1388–1399
- Slobodin, B. et al. (2017) Transcription impacts the efficiency of mRNA translation via co-transcriptional N⁶-adenosine methylation. *Cell* 169, 326–337
- Qi, S.T. et al. (2016) N⁶-Methyladenosine sequencing highlights the involvement of mRNA methylation in occyte meiotic maturation and embryo development by regulating translation in Xenopus laevis. J. Biol. Chem. 291, 23020–23026
- Yu, J. *et al.* (2018) Dynamic m⁶A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* 46, 1412–1423
- Choi, J. et al. (2016) N⁶-Methyladenosine in mRNA disrupts tRNA selection and translation–elongation dynamics. *Nat. Struct. Mol. Biol.* 23, 110–115
- Meyer, K.D. et al. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell 149, 1635–1646
- Dominissini, D. et al. (2012) Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. Nature 485, 201–206
- Linder, B. et al. (2015) Single-nucleotide-resolution mapping of m⁶A and m⁶Am throughout the transcriptome. *Nat. Methods* 12, 767–772
- Ke, S. *et al.* (2015) A majority of m⁶A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* 29, 2037–2053
- Meyer, K.D. (2019) DART-seq: an antibody-free method for global m⁶A detection. *Nat. Methods* 16, 1275–1280
- Zhang, Z. et al. (2019) Single-base mapping of m⁶A by an antibody-independent method. Sci. Adv. 5, eaax0250



- Liu, N. *et al.* (2013) Probing N⁶-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856
- Garcia-Campos, M.A. et al. (2019) Deciphering the "m⁶A code" via antibody-independent quantitative profiling. Cell 178, 731–747
- Mortimer, S.A. et al. (2014) Insights into RNA structure and function from genome-wide studies. Nat. Rev. Genet. 15, 469–479
- Lee, A.S. et al. (2015) eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* 522, 111–114
- Guenther, U.P. et al. (2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. Nature 559, 130–134
- Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8301–8305
- Weingarten-Gabbay, S. *et al.* (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* 351, aad4939
- Jungfleisch, J. et al. (2017) A novel translational control mechanism involving RNA structures within coding sequences. *Genome Res.* 27, 95–106
- Burkhardt, D.H. et al. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife* 6, e22037
- Strobel, E.J. et al. (2018) High-throughput determination of RNA structures. Nat. Rev. Genet. 19, 615–634
- Spasic, A. *et al.* (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* 46, 314–323
- Li, H. and Aviran, S. (2018) Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* 9, 606
- Frank, A.T. *et al.* (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.* 37, 3670–3679
- Woods, C.T. *et al.* (2017) Comparative visualization of the RNA suboptimal conformational ensemble *in vivo. Biophys. J.* 113, 290–301
- Beaudoin, J.D. *et al.* (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* 25, 677–686
- Adivarahan, S. et al. (2018) Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell* 72, 727–738
- Mustoe, A.M. et al. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173, 181–195
- Mizrahi, O. *et al.* (2018) Virus-induced changes in mRNA secondary structure uncover *cis*-regulatory elements that directly control gene expression. *Mol. Cell* 72, 862–874
- Ruijtenberg, S. *et al.* (2019) mRNA structural dynamics shape Argonaute-target interactions. *bioRxiv*. Published online October 30, 2019. https://doi.org/10.1101/822452
- Rinaldi, A.J. et al. (2016) The Shine–Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. Nat. Commun. 7, 8976
- Jens, M. and Rajewsky, N. (2015) Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* 16, 113–126
- Buenrostro, J.D. et al. (2014) Quantitative analysis of RNAprotein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* 32, 562–568
- Webster, M.W. *et al.* (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* 8, e40670

- Benhalevy, D. et al. (2017) The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.* 18, 2979–2990
- D'Agostino, V.G. et al. (2013) A novel high throughput biochemical assay to evaluate the HuR protein–RNA complex formation. PLoS One 8, e72426
- Nicastro, G. *et al.* (2017) Mechanism of beta-actin mRNA Recognition by ZBP1. *Cell Rep.* 18, 1187–1199
- Niessing, D. et al. (2002) Bicoid associates with the 5'-capbound complex of caudal mRNA and represses translation. *Genes Dev.* 16, 2576–2582
- Ma, W. and Mayr, C. (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3' UTR-mediated protein– protein interactions. *Cell* 175, 1492–1506
- Berkovits, B.D. and Mayr, C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522, 363–367
- Gerst, J.E. (2018) Pimp my ribosome: ribosomal protein paralogs specify translational control. *Trends Genet.* 34, 832–845
- Genuth, N.R. and Barna, M. (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell* 71, 364–374
- Gupta, V. and Warner, J.R. (2014) Ribosome-omics of the human ribosome. *RNA* 20, 1004–1013
- Guimaraes, J.C. and Zavolan, M. (2016) Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* 17, 236
- Jiang, L. et al. (2017) RPL10L is required for male meiotic division by compensating for RPL10 during meiotic sex chromosome inactivation in mice. *Curr. Biol.* 27, 1498–1505
- O'Leary, M.N. et al. (2013) The ribosomal protein Rpl22 controls ribosome composition by directly repressing expression of its own paralog, Rpl2211. PLoS Genet. 9, e1003708
- Shi, Z. et al. (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* 67, 71–83
- Xue, S. et al. (2015) RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. Nature 517, 33–38
- Mills, E.W. and Green, R. (2017) Ribosomopathies: there's strength in numbers. *Science* 358, eaan2755
- Simsek, D. *et al.* (2017) The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell* 169, 1051–1065
- Staes, A. et al. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8, 1362–1370
- Kleifeld, O. *et al.* (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* 6, 1578–1611
- Na, Y. et al. (2016) Real-time imaging reveals properties of glutamate-induced Arc/Arg 3.1 translation in neuronal dendrites. *Neuron* 91, 561–573
- Tanenbaum, M.E. et al. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
- Khuperkar, D. et al. (2020) Quantification of mRNA translation in live cells using single-molecule imaging. *Nat. Protoc.* 15, 1371–1398
- Hoek, T.A. et al. (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. Mol. Cell 75, 324–339
- Moon, S.L. et al. (2019) Multicolour single-molecule tracking of mRNA interactions with RNP granules. Nat. Cell Biol. 21, 162–168
- Lyon, K. et al. (2019) Live-cell single RNA imaging reveals bursts of translational frameshifting. *Mol. Cell* 75, 172–183
- Wilbertz, J.H. et al. (2019) Single-molecule imaging of mRNA localization and regulation during the integrated stress response. *Mol. Cell* 73, 946–958