TRANSLATION and TRANSLATION HETEROGENEITY

The flow of information from DNA to messenger RNA and protein, is described by the central dogma of molecular biology. In 1957 Francis Crick, James Watson, and Maurice Wilkins were awarded the Nobel Prize in Physiology or Medicine for their discoveries concerning the molecular structure of DNA and RNA and the significance of the information transfer from DNA to proteins. Since then, there have been many discoveries and technological advances that led to a better understanding of the molecular mechanisms of gene expression. At the very end, the fate and function of each cell depends on its protein composition. Therefore, accurate gene expression control is critical for proper cell functioning.

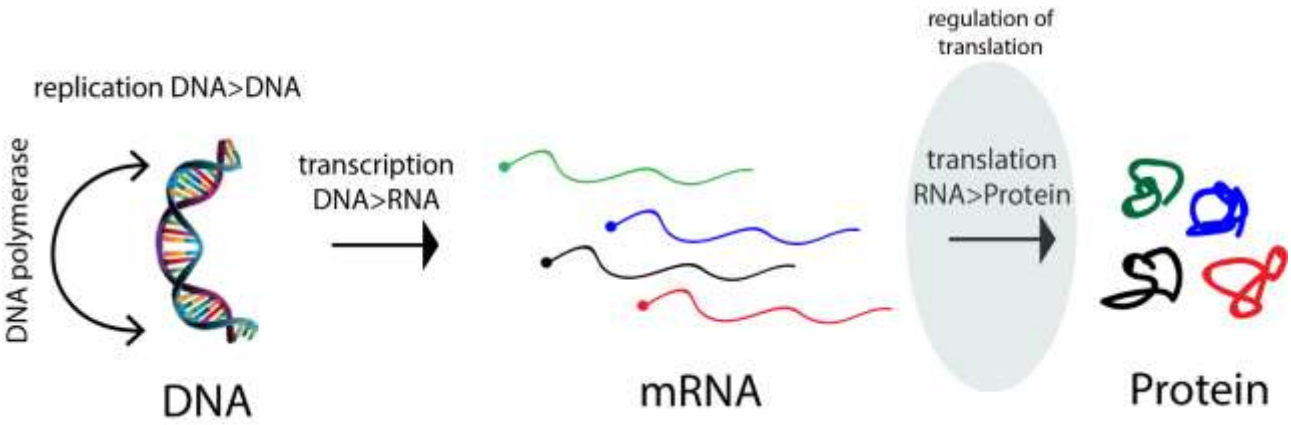

Figure 1. The central dogma of molecular biology

A key step in gene expression is mRNA translation: all organisms respond to environmental or physiological changes by altering the amounts and activities of specific proteins that are necessary for their adaptation and survival. The regulation of **mRNA translation** is fast, reversible and provides spatial control, making it a unique regulatory mechanism of gene expression. **In a simplistic view, one mRNA molecule encodes for a single protein. However, it is now clear that translation is far more heterogeneous.**

Gene expression is heavily controlled by a network of highly interconnected posttranscriptional regulatory factors, such as RNA-binding proteins and noncoding RNAs. The amplitude of translational regulation exceeds the sum of transcription, mRNA degradation and protein degradation. Therefore, it is essential to investigate translation in a global scale and the totality of the components in the translation process, including but not limited to translating mRNAs, ribosomes, tRNAs, regulatory RNAs and nascent polypeptide chains. **The variability in translation can be at four different levels:**

- mRNA molecules from different genes can be differentially translated in a single cell
- mRNA molecules originating from a single gene in a single cell can be translated differentially
- single mRNA can be translated differentially in different cell types
- a single mRNA molecule can be translated differentially in space and time in a single cell

Recent technical advances have brought breakthroughs in the investigation of composition and dynamics of many components involved in translation. All methodologies listed here below are commonly used for measuring different aspects of protein synthesis, and most of the time are combined together for a comprehensive view of the biological problem. Among them, ribosome profiling (*1*) is one of the main methods used to shed light into many aspects of protein synthesis.

1. Ribosome profiling
2. Mass spectrometry
3. Polysome profiling/sucrose sedimentation
4. RNAseq
5. Chemical RNA probing
6. Single-Molecule Imaging
7. Ribosome affinity purification

At the level of protein synthesis, different factors can cause heterogeneity. Here below, we briefly list the most important elements underling molecular mechanisms of spatial and temporal regulation of translation.

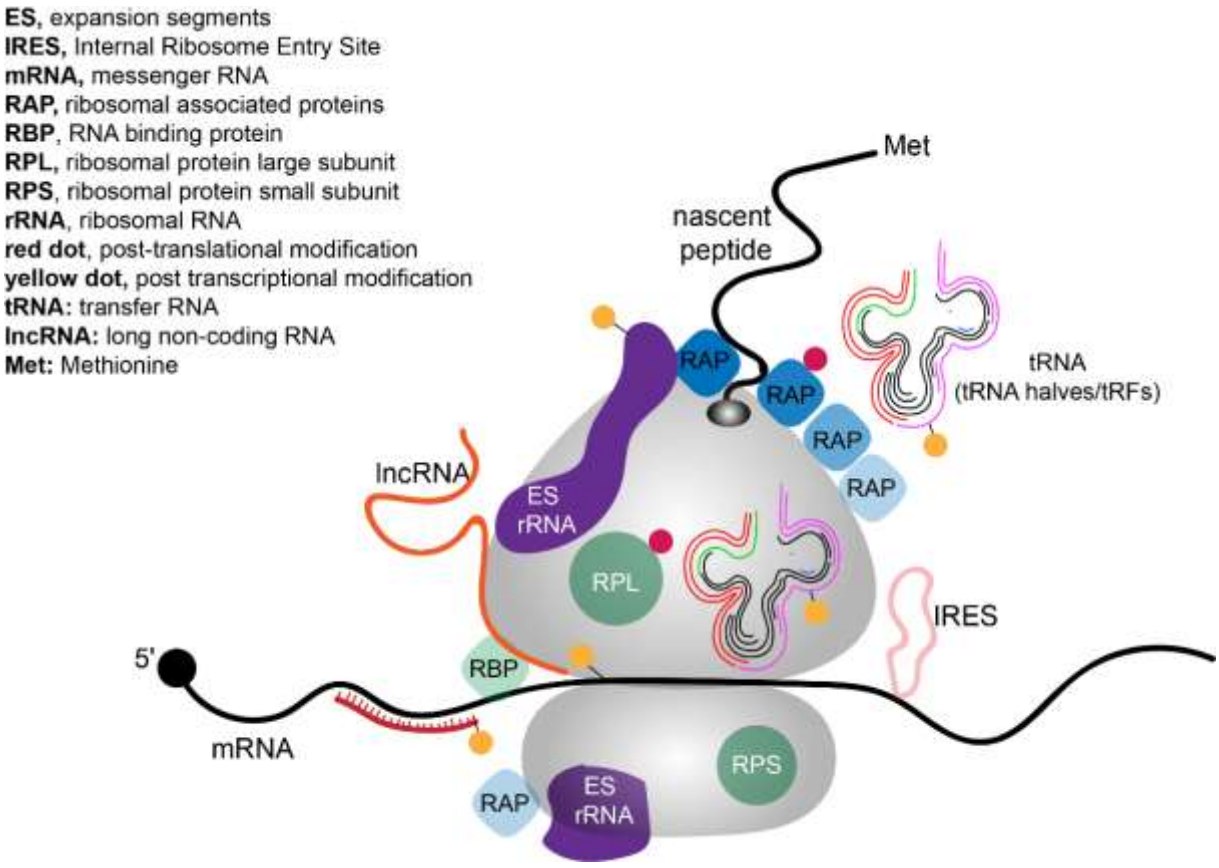

Figure 2. Heterogeneity at the ribosomal level

Ribosome Heterogeneity

Not all ribosomes are the same: while previously viewed as a homogenous machine, ribosome diversity play a central role in the regulation of gene expression. Several studies have suggested

that some degree of heterogeneity may exist in the composition of the ribosome in term of rRNA, ribosomal protein and ribosome associated proteins (*2–4*). This variability generates ribosome with preferential regulation for any aspect of translational control (called "specialized ribosomes") (*5*, *6*). Gene expression analysis of ribosomal genes revealed that up to 25% of the ribosomal genes are differentially expressed across various tissues and cell lines and several paralogs of ribosomal proteins are exclusively expressed in one tissue or cell type. Additionally, ribosomes have been shown to be associated with many ribosome-associated proteins (*7*). These proteins, as well as ribosomal proteins, can be posttranslational modified; adding another sources of heterogeneity in ribosomes (*3*). Defects in ribosome biogenesis/composition can generate ribosomopathies (*8*).

mRNA sequence

Eukaryotic genes encode multiple mRNA isoforms that differ in their primary nucleotide sequence (a median of six transcript isoforms has been detected for each protein coding gene) (*9*, *10*). The different nucleotide sequence can originate from different polyadenylation sites (*11*), alternative splicing (*12*) and variable transcription start site (TSS) usage (> 1 million TTSs are identified) (*13*).

RNA type

An incredibly small fraction of the mammalian genome is annotated as protein-coding (< 3 %), while the number of potentially functional non-coding genes remains unclear (*14*). It has been discovered that a large portion of noncoding RNAs are partially translated in short peptides or is associated to ribosomes. Among them, long non-coding RNAs (lncRNAs) have been shown to interact with ribosomes, encode peptides, or act as ribosome sponges (*15–17*).

RNA modifications

Posttranscriptional modifications of RNA nucleotides play a key role in the storage and transfer of the genetic information. Distinct nucleotide modifications have been described for mRNAs, which have been termed the 'epitranscriptome'. Many other modifications are described on tRNA and rRNA, with the ribosome turned out to be a platform of RNA modifications (*18*). The most prevalent modification is methylation of the adenosine base at the nitrogen-6 position (m6A) (average occurrence: one to three modified adenosines per mRNA). Our understanding of the effect of these modification in translation and diseases is only at the beginning (*19*).

mRNA structure

An additional layer of information comes from the complex and high-ordered RNA structures. These structure could affect translation in multiple ways: inhibit translation initiation, promote translation initiation or stall ribosomes during translation elongation (*20*, *21*). A subset of mRNAs contains internal ribosomal entry sites (IRESs), usually in the 5' UTR, enabling cap-independent initiation (*22*).

RNA binding proteins

The majority of the RNA binding proteins associate with mRNAs through one or multiple well-defined RNA-binding domains to form ribonucleoprotein complexes (RNPs). RNA binding domains often bind a relatively short sequence motif of around two to eight nucleotides (*23*). The interaction of the RBP with the target mRNA affects stability and localization of both protein and the target RNA in a complex network.

All the elements listed above could cause differences in protein synthesis inside each single cell. This can result in:

1. **Production of proteins with different amino acid composition, with/without differences in length.**
   This effect can be caused by (i) ribosome frameshifting during translation elongation, (ii) alternative translation start site selection or (iii) stop codon readthrough:

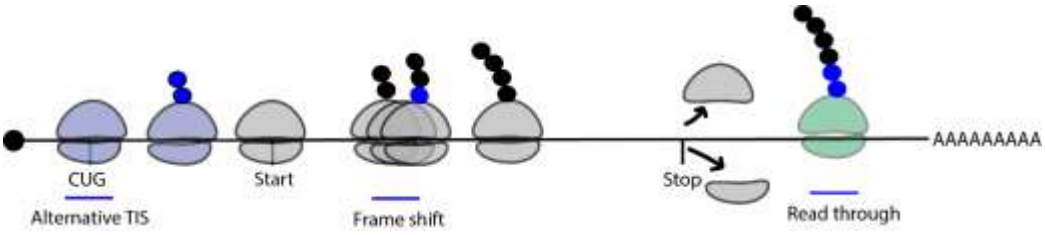

Figure 3. Heterogeneity in protein composition

- Frameshifting is caused mainly by sequence elements in mRNA/tRNA that stochastically redirect translating ribosomes into a new reading frame (i.e., by + 1 or − 1 nucleotide). It is typical in viral gene expression in the framework of host-pathogen interactions (*24*, *25*).

- Start codon (AUG) are most effectively recognized as a translation initiation site (TIS) when surrounded by specific sequences (e.g Kozak consensus sequence (GCCACCAUGG) (*26*). However, translation initiation can also occur at non-AUG codons, such as GUG or CUG, and it can fail to initiate at an start site caused by leaky scanning. Ribosomal profiling revealed that at least half of the human mRNA transcripts contain more than one TIS (*27*–*29*). Alternative TIS usage could also act as part of a regulatory mechanism to bring ribosomes away from the main open reading frame (ORF), resulting (at least in eukaryotes) in 'junk' polypeptides; mainly for TISs that are out of frame with respect to the main protein coding ORF. Finally, translation form unconventional TISs is known to drive tumorigenesis (*30*).

- Stop codon recognition is generally efficient and results in translation termination. However, with a frequency from 0.01% to 0.1% the stop codon (mainly UGA) can be decoded as a sense codon, resulting in a C terminally extended protein (*31*, *32*), at least in mammalian cells.

2. **Protein produced with a different synthesis rate**

   Heterogeneity in the translation rate is likely to originate predominantly at the initiation step, mainly caused by differential recruitment rates of the pre-initiation complex to the 5' cap and usage of alternative TISs (*33*). In addition, variability in the translation elongation rate (e.g., due to ribosome stalling) may contribute to generate additional variability and used by the cells for spatial control of gene expression (e.g. in neurons) (*34*).

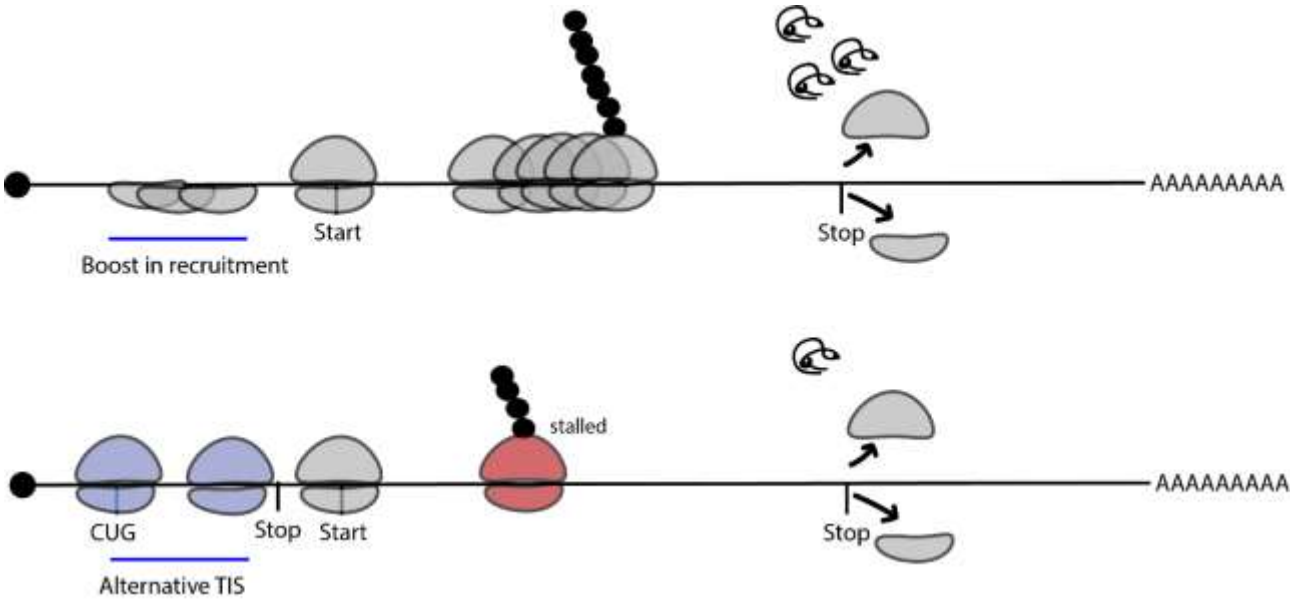

Figure 4. Heterogeneity in protein synthesis rate. Top, enhancement of the rate of protein synthesis. Bottom, depression of protein synthesis rate.

Overall, there are many steps involved that contribute to gene expression and in the past years the mRNA levels (RNA-seq) was considered to be the main read out of gene expression. However, it is just one of many indicators. **Gene expression can be highly regulated during protein synthesis. Immagina BioTechnology s.r.l. provides uniquely enabling technologies to break down walls on translation and ribosome profiling studies.**

References

1. M. Clamer, T. Tebaldi, F. Lauria, P. Bernabò, R. F. Gómez-Biagi, M. Marchioretto, D. T. Kandala, L. Minati, E. Perenthaler, D. Gubert, L. Pasquardini, G. Guella, E. J. N. Groen, T. H. Gillingwater, A. Quattrone, G. Viero, Active Ribosome Profiling with RiboLace. *Cell Rep.* **25**, 1097-1108.e5 (2018).
2. J. C. Guimaraes, M. Zavolan, Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* **17**, 236 (2016).
3. K. Wild, M. Aleksić, K. Lapouge, K. D. Juaire, D. Flemming, S. Pfeffer, I. Sinning, MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat. Commun.* **11**, 776 (2020).
4. D. Simsek, G. C. Tiu, R. A. Flynn, G. W. Byeon, K. Leppek, A. F. Xu, H. Y. Chang, M. Barna, The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell*. **169**, 1051-1065.e18 (2017).
5. N. R. Genuth, M. Barna, The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell*. **71**, 364–374 (2018).
6. Z. Shi, K. Fujii, K. M. Kovary, N. R. Genuth, H. L. Röst, M. N. Teruel, M. Barna, Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell*. **67**, 71-83.e7 (2017).
7. V. Gupta, J. R. Warner, Ribosome-omics of the human ribosome. *RNA*. **20**, 1004–1013 (2014).
8. E. W. Mills, R. Green, Ribosomopathies: There's strength in numbers. *Science (80-.).* **358**, eaan2755 (2017).
9. A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park, A. Regev,

Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. **498**, 236–240 (2013).
10. L. Velten, S. Anders, A. Pekowska, A. I. Järvelin, W. Huber, V. Pelechano, L. M. Steinmetz, Single-cell polyadenylation site mapping reveals 3' isoform choice variability. *Mol. Syst. Biol.* **11**, 812 (2015).
11. B. Tian, J. L. Manley, Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* **18**, 18–30 (2017).
12. E. T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, C. B. Burge, Alternative isoform regulation in human tissue transcriptomes. *Nature*. **456**, 470–476 (2008).
13. M. Kanamori-Katayama, M. Itoh, H. Kawaji, T. Lassmann, S. Katayama, M. Kojima, N. Bertin, A. Kaiho, N. Ninomiya, C. O. Daub, P. Carninci, A. R. R. Forrest, Y. Hayashizaki, Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* **21**, 1150–1159 (2011).
14. S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature*. **489**, 101–108 (2012).
15. J. Carlevaro-Fita, A. Rahim, R. Guigó, L. A. Vardy, R. Johnson, Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*. **22**, 867–82 (2016).
16. J. Ruiz-Orera, X. Messeguer, J. A. Subirana, M. M. Alba, Long non-coding RNAs as a source of new peptides. *Elife*. **3**, e03523 (2014).
17. G. V. and M. C. Luca Minati†1, Claudia Firrito†1, Alessia Del Piano†1, Alberto Peretti†1, Simone Sidoli2, Daniele Peroni3, Romina Belli3, Francesco Gandolfi4, Alessandro Romanel4, Paola Bernabò1, Jacopo Zasso5, Alessandro Quattrone5, Graziano Guella6, Fabio Lauria7, One-shot analysis of translated mammalian lncRNAs with AHARIBO, doi:10.1101/2020.04.20.050062.
18. S. K. Natchiar, A. G. Myasnikov, H. Kratzat, I. Hazemann, B. P. Klaholz, Visualization of chemical modifications in the human 80S ribosome structure. *Nature*. **551**, 472–477 (2017).
19. E. Destefanis, G. Avşar, P. Groza, A. Romitelli, S. Torrini, P. Pir, S. G. Conticello, F. Aguilo, E. Dassi, *RNA*, in press, doi:10.1261/rna.077271.120.
20. J. Jungfleisch, D. D. Nedialkova, I. Dotu, K. E. Sloan, N. Martinez-Bosch, L. Brüning, E. Raineri, P. Navarro, M. T. Bohnsack, S. A. Leidel, J. Díez, A novel translational control mechanism involving RNA structures within coding sequences. *Genome Res.* **27**, 95–106 (2017).
21. D. H. Burkhardt, S. Rouskin, Y. Zhang, G.-W. Li, J. S. Weissman, C. A. Gross, Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife*. **6** (2017), doi:10.7554/eLife.22037.
22. C. U. T. Hellen, Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.* **15**, 1593–1612 (2001).
23. J. D. Buenrostro, C. L. Araya, L. M. Chircus, C. J. Layton, H. Y. Chang, M. P. Snyder, W. J. Greenleaf, Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **32**, 562–568 (2014).
24. L. V. Bock, N. Caliskan, N. Korniy, F. Peske, M. V. Rodnina, H. Grubmüller, Thermodynamic control of −1 programmed ribosomal frameshifting. *Nat. Commun.* **10**, 4598 (2019).
25. M. Mikl, Y. Pilpel, E. Segal, High-throughput interrogation of programmed ribosomal frameshifting in human cells. *Nat. Commun.* **11**, 3061 (2020).
26. M. Kozak, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*. **44**, 283–292 (1986).
27. N. T. Ingolia, L. F. Lareau, J. S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. **147**, 789–802 (2011).
28. S. Lee, B. Liu, S. Lee, S.-X. Huang, B. Shen, S.-B. Qian, Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* **109**, E2424–E2432 (2012).
29. C. H. Na, M. A. Barbhuiya, M.-S. Kim, S. Verbruggen, S. M. Eacker, O. Pletnikova, J. C. Troncoso, M. K. Halushka, G. Menschaert, C. M. Overall, A. Pandey, Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res.* **28**, 25–36 (2018).
30. A. Sendoel, J. G. Dunn, E. H. Rodriguez, S. Naik, N. C. Gomez, B. Hurwitz, J. Levorse, B. D. Dill, D. Schramek, H. Molina, J. S. Weissman, E. Fuchs, Translation from unconventional 5' start sites drives tumour initiation. *Nature*. **541**, 494–499 (2017).

31. G. Loughran, M.-Y. Chou, I. P. Ivanov, I. Jungreis, M. Kellis, A. M. Kiran, P. V. Baranov, J. F. Atkins, Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res.* **42**, 8928–8938 (2014).
32. M. Dabrowski, Z. Bukowy-Bieryllo, E. Zietkiewicz, Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biol.* **12**, 950–958 (2015).
33. S. Boersma, D. Khuperkar, B. M. P. Verhagen, S. Sonneveld, J. B. Grimm, L. D. Lavis, M. E. Tanenbaum, Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell*. **178**, 458-472.e19 (2019).
34. C. Glock, M. Heumüller, E. M. Schuman, mRNA transport & local translation in neurons. *Curr. Opin. Neurobiol.* **45**, 169–177 (2017).