Ribosome profiling

Candidate of bio. Sci. Meldebekova Aliya A.

- Introduction
- What is Ribosome Profiling (RP)?
- An overview of ribosome profiling
- **Ribosome profiling** in determination of short **Open Reading Frames (sORFs)**
- **RP in** determination **translational pauses (ribosome pauses)**
- **RP** in in determination of **upstream ORFs (uORFs),**

Introduction

- Efforts to globally monitor gene expression have historically focused on measuring mRNA levels (for example, using microarrays or RNA-seq), although we know that translational control is an essential and regulated step in determining protein expression.
- Until recently, precisely monitoring protein translation was far more challenging than measuring mRNA levels.
- This has changed with the development of the ribosome profiling approach, first published in 2009.

What is Ribosome Profiling?

- A ribosome profiling is a tool to interrogate what is being translated, how this translation is regulated, and where in the cell the translation of specific sets of proteins occurs.

- Ribosome profiling is a deep sequencing-based tool that enables the detailed measurement of translation globally and in vivo. At the core of this approach is the observation that a translating ribosome strongly protects about 30 nucleotides of an mRNA from nuclease activity.

https://www.youtube.com/watch?v=mnuSkhRyBCg

Sequencing of **these ribosome-protected fragments** (**RFPs**) thus provides a precise record of the position of the ribosome at the time when translation was halted.

Measuring the density of protected fragments on a given transcript provides a proxy for the rate of protein synthesis.

Also, determining the position of the protected fragments makes it possible to empirically measure the identity of translation products (for example, **where they begin and end**, and **even the frame being read**).

This has led to the discovery of a number of novel or alternative protein products.

- Ribosome profiling exploits the classical molecular method of ribosome footprinting, in which in vitro translated mRNAs are nuclease-treated to destroy the regions that are not protected by the ribosome.
- Such treatment leaves ‘footprints’ of approximately 30 nucleotides, which can be mapped back to the original mRNA to define the exact location of the translating ribosome.

- Ribosome profiling requires:
- collection of a physiological sample,
- inhibition of translation to freeze ribosomes in the act of translation,
- nuclease digestion to produce ribosome-protected fragments,
- isolation of ribosomes and, subsequently, of ribosome footprints,
- Ribosome footprints are converted to a strand-specific library and subjected to next generation sequencing,
- the fragments are then mapped to the appropriate reference genome. Ribosome profiling is typically carried out on a split sample, with parallel libraries constructed for measuring mRNA abundance by mRNA-Seq. Comparison between the rates of protein synthesis and

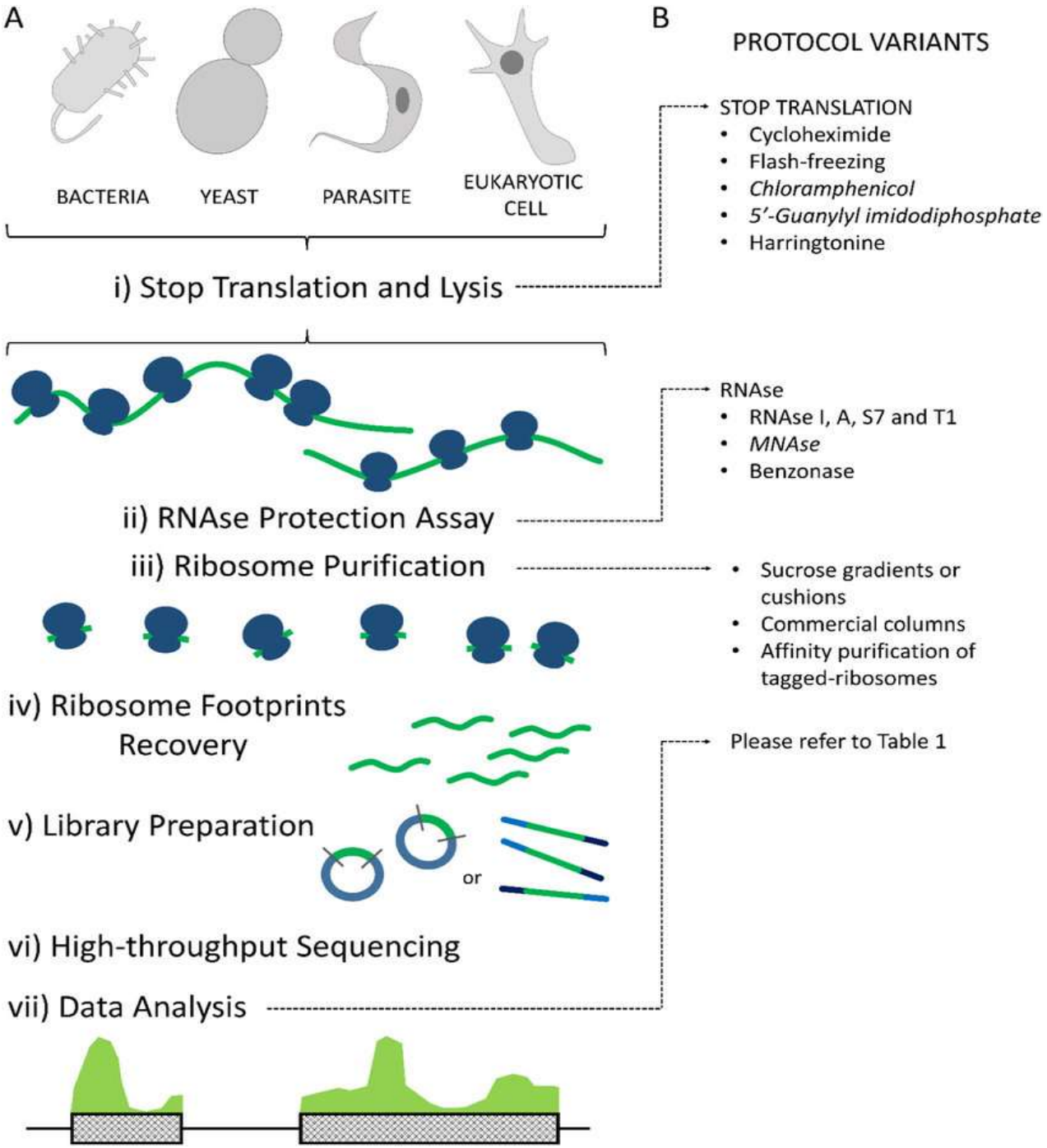


Figure.
Ribosome Profiling protocol description.
A general description of RP protocol is shown in A, representing the main steps described in the text.
The protocol variants discussed are summarized in B, linked to the corresponding step where would be applied.
Variants that correspond to prokaryotes are marked in italic.

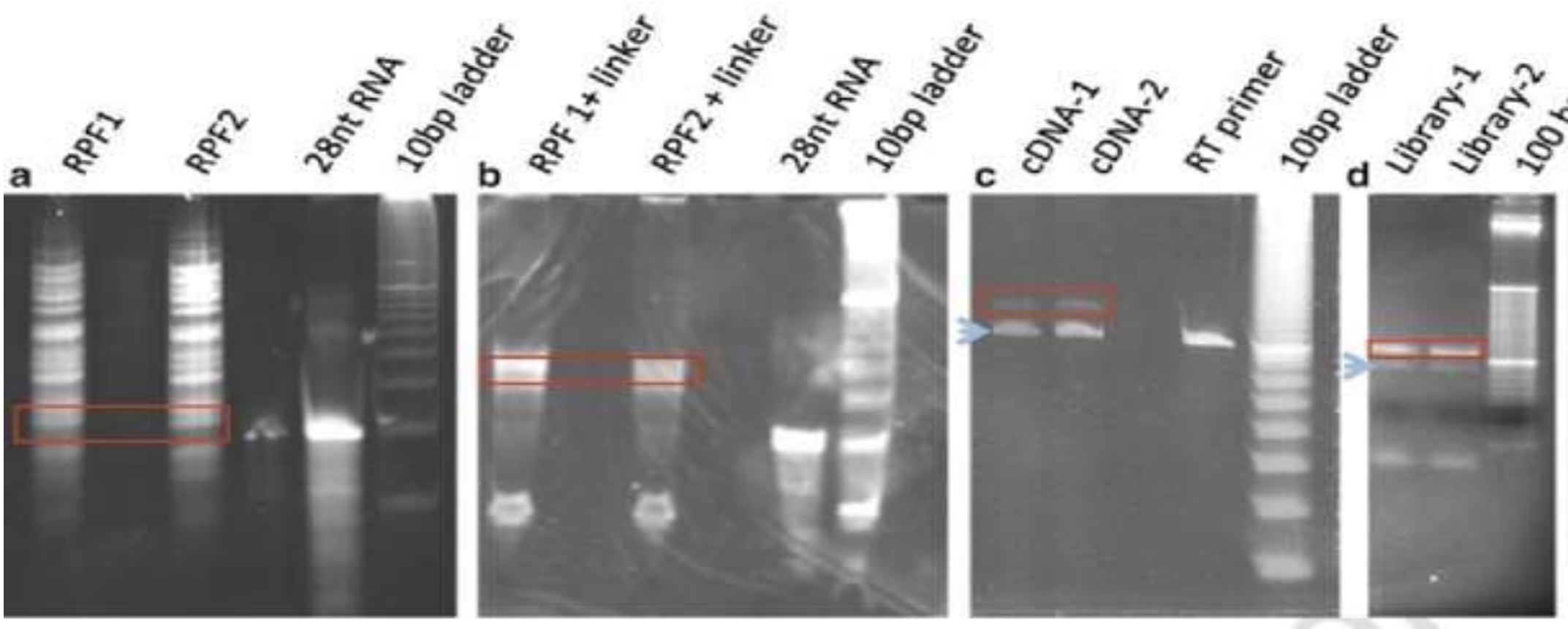


Figure. Representative gel images from ribosome-protected fragment (RPF) library construction.

(a) Size selection of RPFs. The ribosome footprint RNA fragments are derived from ribosomes obtained by TRAP from two Arabidopsis rosette leaf samples expressing the transgene 35S:His6-FLAG-RPL18

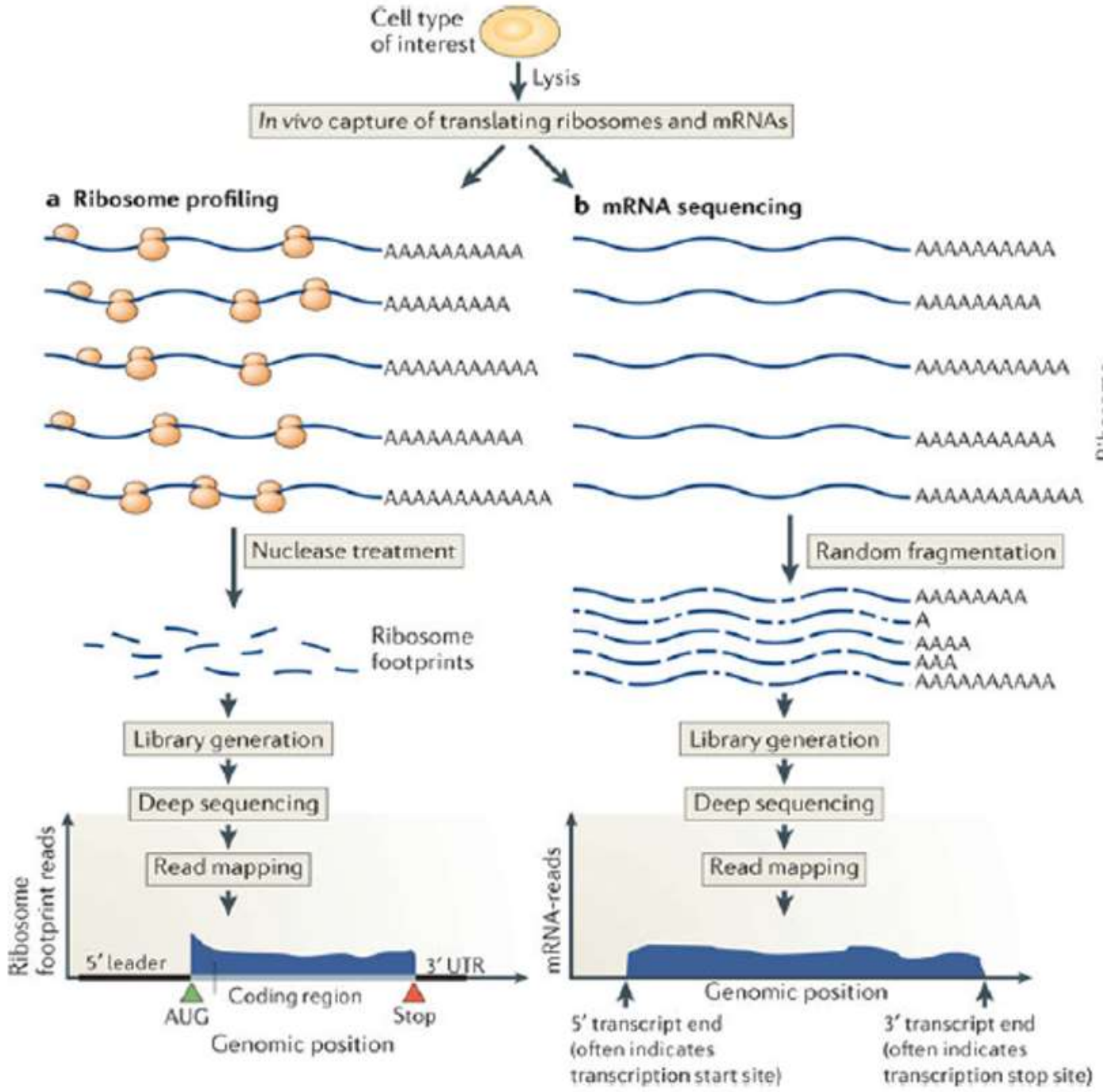
(b) Purification of ligation products (Subheading 3.5.3). Ligated RPFs are 20 nt larger than the 28 nt RNA marker. The lower band represents free linker. The red rectangle indicates the gel region to be extracted.

(c) (c) Purification of reverse transcription products.

(d) (d) Purification of PCR products.

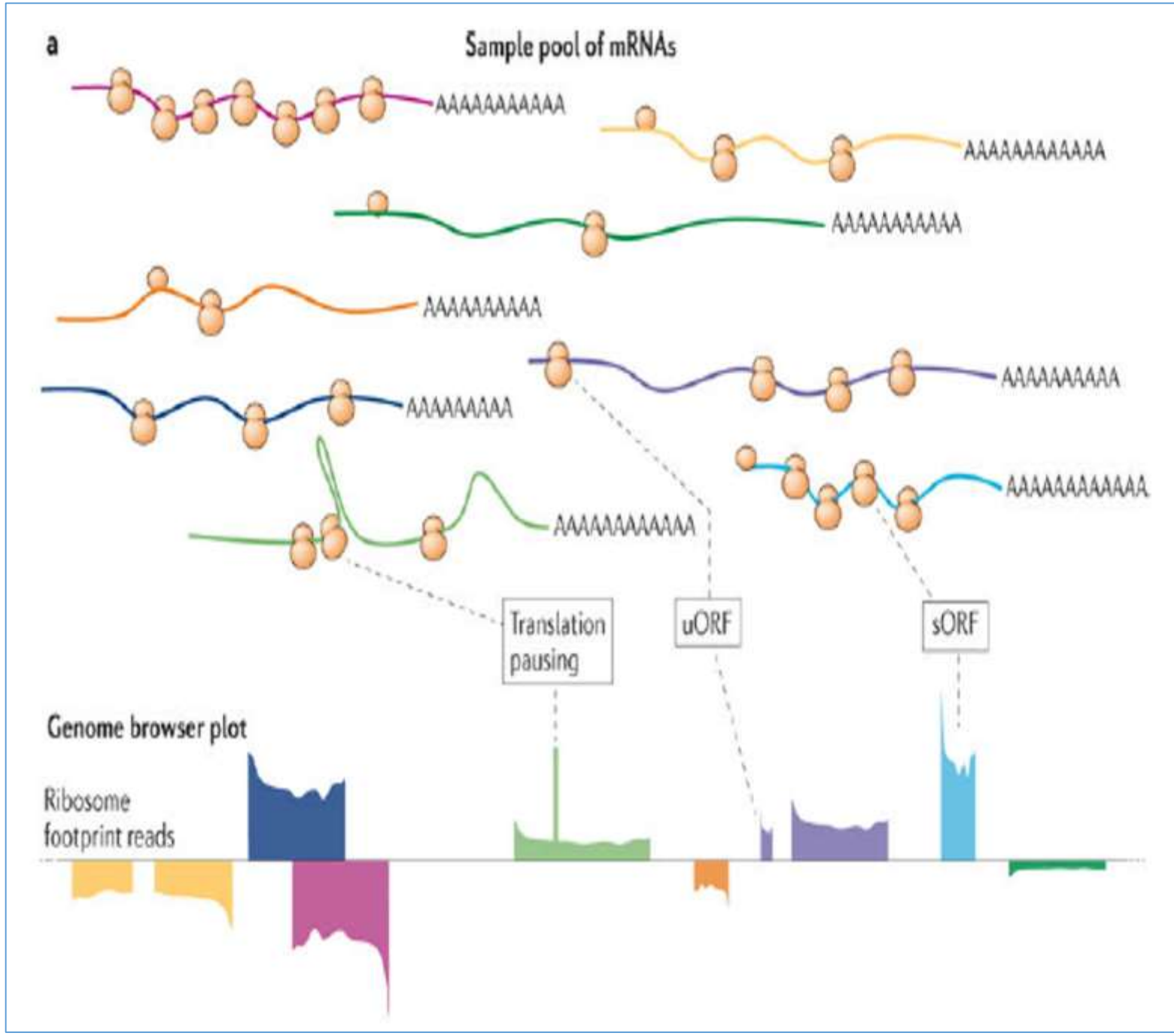
Ribosome Profiling: A Tool for Quantitative Evaluation of Dynamics in mRNA Translation. March 2015 Methods in molecular biology (Clifton, N.J.) 1284:139-73, •DOI: 10.1007/978-1-4939-2444-8_7

An overview of ribosome profiling.
a) Ribosome-bound mRNAs are isolated by size and treated with nonspecific nuclease (typically RNAse I or micrococcal nuclease), which results in protected mRNA fragments or ‘footprints’. These ribosome footprints are isolated and converted to a library for deep sequencing. The ribosome footprints typically show precise positioning between the start and stop codon of a gene, which enables global and experimental genomic coding region identification.
b) By comparison, mRNA sequencing captures random fragments covering the entire mRNA transcript. The positional information determined by standard mRNA sequencing enables approximate determination of transcript boundaries, but is less precise than ribosome profiling due to the loss of 5′ and 3′ ends during the fragment generation method typically used.



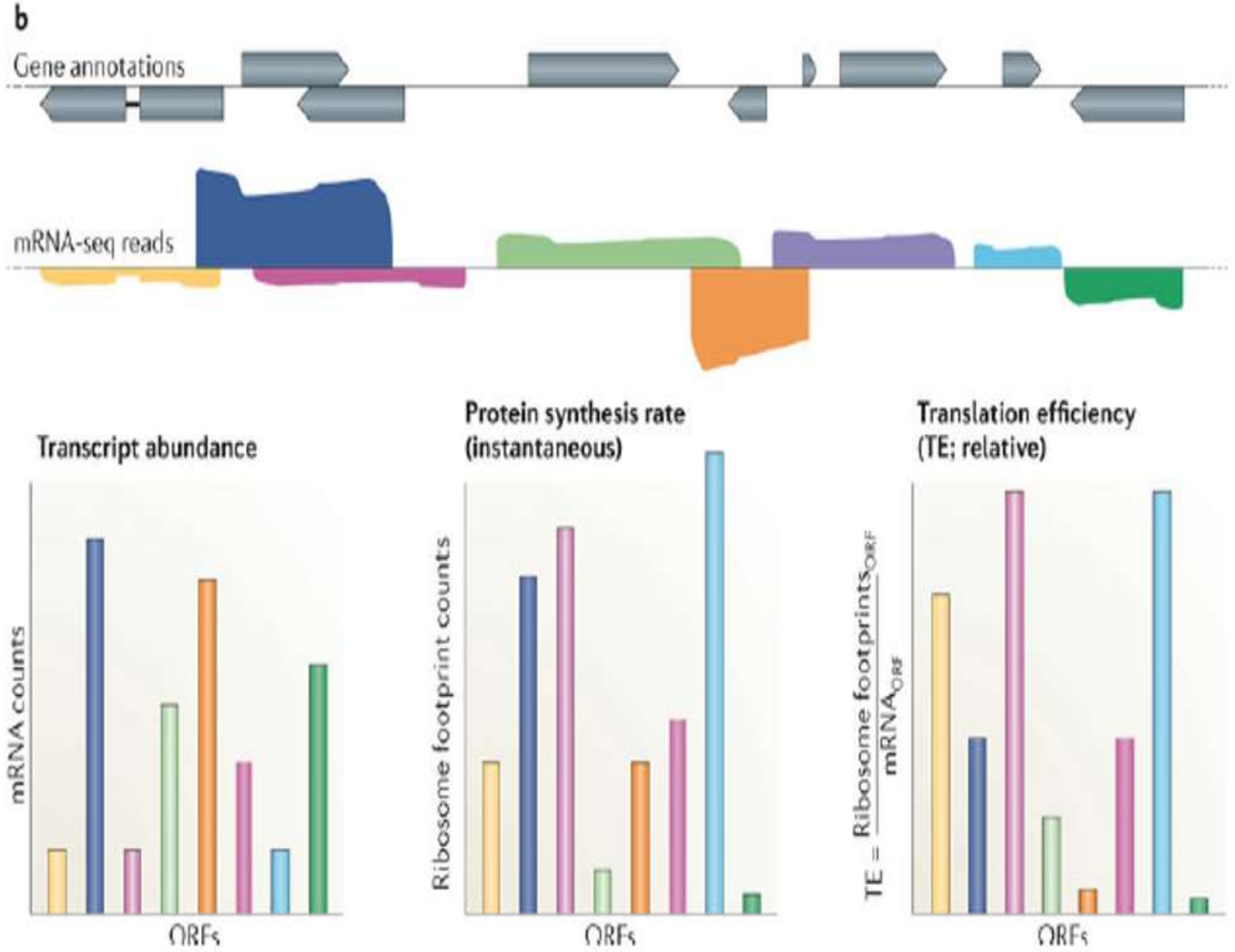
Comparing **the rates of protein synthesis** and **the mRNA abundance** enables us to determine the **transational efficiency (TE)** for each mRNA.

Qualitative and quantitative data provided by ribosome profiling. **a)** A diverse sample pool of mRNAs, distinguished by color, are shown together with a corresponding representative genome browser plot of ribosome profiling data derived from this pool.
Note that ribosome profiling enables experimental determination of translated regions, including short **Open Reading Frames (sORFs)**, which may be an important new source of cellular peptides, and **upstream ORFs (uORFs),** which are thought to be largely regulatory.
Pausing during translation elongation may result in peaks in ribosome footprints within ORFs.



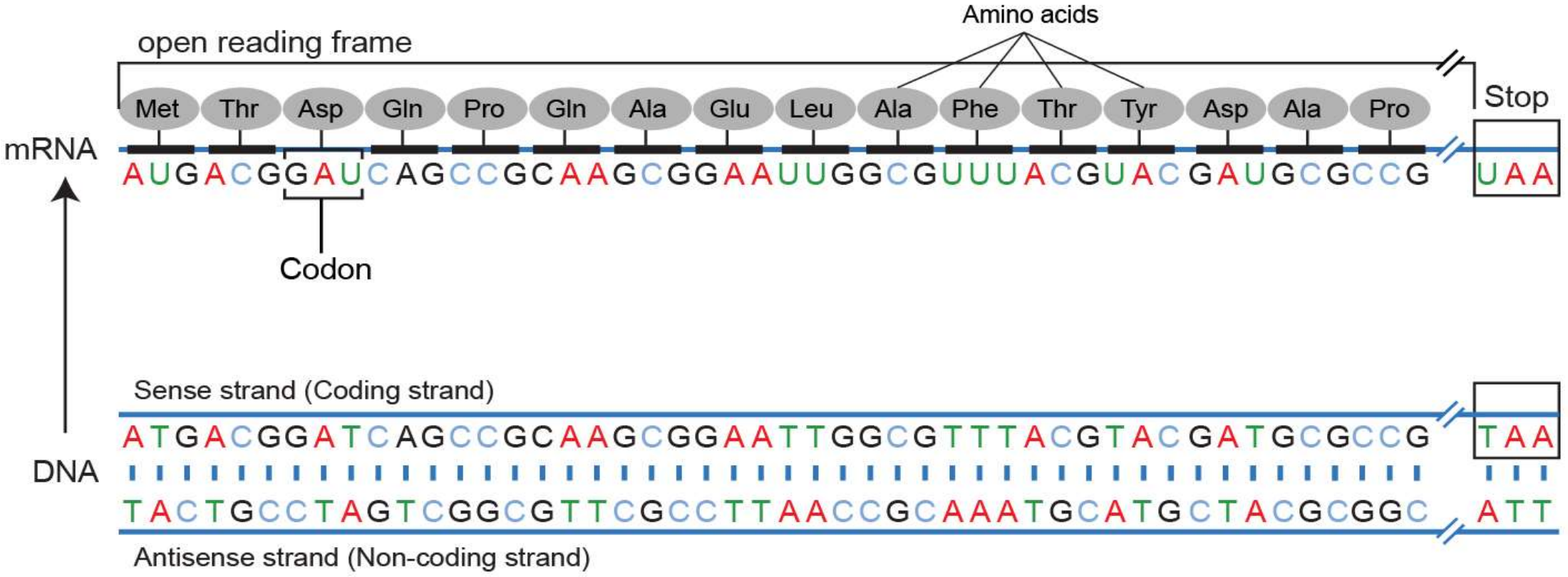
b) Overlaid gene annotations and mRNA-seq data for the examples shown in (a). c) Examples of quantitative data derived from b). Note that transcript abundances may not correlate closely with the instantaneous protein synthesis rates.

The collection of quantitative data for both transcript abundances and their protein synthesis rates enables inference of the relative translation efficiencies. These can vary over several orders of magnitude within a given organism in a given state. **The translation efficiency can also change over time for a given mRNA**, reflecting dynamic regulation at the level of translation.



Translation and Open Reading Frame (ORF)

- In translation codons of three nucleotides determine which amino acid will be added next in the growing protein chain.
- It is important then to decide which nucleotide to start translation, and when to stop, this is called an open reading frame.



- An open reading frame is a portion of a DNA molecule that, when translated into amino acids, contains no stop codons.
- The genetic code reads DNA sequences in groups of three base pairs (codons), which means that a double-stranded DNA molecule can read in any of six possible reading frames (three in the forward direction and three in the reverse). A long open reading frame is likely part of a gene.

Once a gene has been sequenced it is important to determine the correct **open reading frame** (ORF). Every region of DNA has six possible **reading frames**, three in each direction.

The reading frame that is used determines which amino acids will be encoded by a gene. Typically only one reading frame is used in translating a gene (in eukaryotes), and this is often the longest open reading frame.

Once the open reading frame is known the DNA sequence can be translated into its corresponding amino acid sequence. An open reading frame starts with an atg (Met) in most species and ends with a stop codon (taa, tag or tga).

- For example, the following sequence of DNA can be read in six reading frames. Three in the forward and three in the reverse direction.
- **The three reading frames** in the **forward direction** are shown with the translated amino acids below each DNA seqeunce.
- Frame **1** starts with the "**a**", Frame **2** with the "**t**" and Frame **3** with the "**g**". Stop codons are indicated by an "*" in the protein sequence.
- The longest ORF is in Frame **1**.

5' **3'**

atgcccaagctgaatagcgtagaggggttttcatcatttgaggacgatgtataa

1) **a**tg ccc aag ctg aat agc gta gag ggg ttt tca tca ttt gag gac gat gta **taa**
M P K L N S V E G F S S F E D D V *

2) **t**gc cca agc **tga** ata gcg **tag** agg ggt ttt cat cat ttg agg acg atg tat
C P S * I A * R G F H H L R T M Y

3) **g**cc caa gct gaa **tag** cgt aga ggg gtt ttc atc att **tga** gga cga tgt ata
A Q A E * R R G V F I I * G R C I

Small open reading frames (sORFs)

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3203329/
Article is Dated 2011

- An increasing body of evidence shows that proteins translated from small open reading frames (sORFs; <100 codons) are involved in a variety of important functional classes. These biological functionality includes but not limited to mating pheromones, energy metabolism, proteolipids, chaperonins, stress proteins, transporters, transcriptional regulators, nucleases, ribosomal proteins, thioredoxins, metal ion chelators and transmembrane proteins [1].
- For example, *tarsal-less* (*tal*) gene, a 33-nucleotide-long ORF, is translated into 11-amino-acid-long peptide and controls gene expression and tissue folding in the *Drosophila* [2]. Another sORF gene, *polished rice* (*pri*), which is of 11-32 amino acids long, controls epidermal differentiation in *Drosophila* by modifying the transcription factor Shavenbaby [3, 4].
- In *Bacillus subtilis*, the 46-amino-acid-long Sda protein inhibits the onset of sporulation by preventing activation of a transcription factor required for sporulation [5]. In plants, the products of sORF protein coding genes are important components of photosynthetic supracomplex [6].
- These identified and characterized examples indicate that **sORFs are ubiquitous and play significant roles in various biological processes**.

- While some sORFs have important function, the majority of sORFs are probably meaningless and arisen by chance.
- Thus, identifying **functional sORFs** from huge number of **fake sORFs** is a daunting task for genome annotation projects.
- In practice, **sORFs are generally eliminated from any genome annotations** (i.e., **only ORFs with at least 100 codons are annotated**) [7] and the functions of proteins encoded by sORFs are largely un-explored.
- Hence, it is critical to employ effective computational methods to pinpoint potentially genuine sORFs "buried" under piles of meaningless ones [1].
- Such **an accurate sORF list** (narrowed down by computational methods to a practical scale) **is highly desirable for investigators** to perform follow-up experimental characterization.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3203329/
Article is Dated 2011

Recent articles. Example.

- Whole transcriptome analyses have revealed that mammalian genomes are massively transcribed, resulting in the production of huge numbers of **transcripts with unknown functions (TUFs)**.
- Previous research has categorized **most TUFs** as noncoding RNAs (ncRNAs) because most previously studied TUFs do not encode open reading frames (ORFs) with biologically significant lengths [>100 amino acids (AAs)].
- Recent studies, however, have reported that several transcripts harbouring small ORFs **that encode peptides shorter than 100 AAs are translated** and play important **biological functions**.
- Here, we examined the translational capacity of transcripts annotated as ncRNAs in human cells, and identified several hundreds of ribosome-associated transcripts previously annotated as ncRNAs. **Ribosome footprinting and polysome profiling analyses revealed that 61 of them are potentially translatable**.

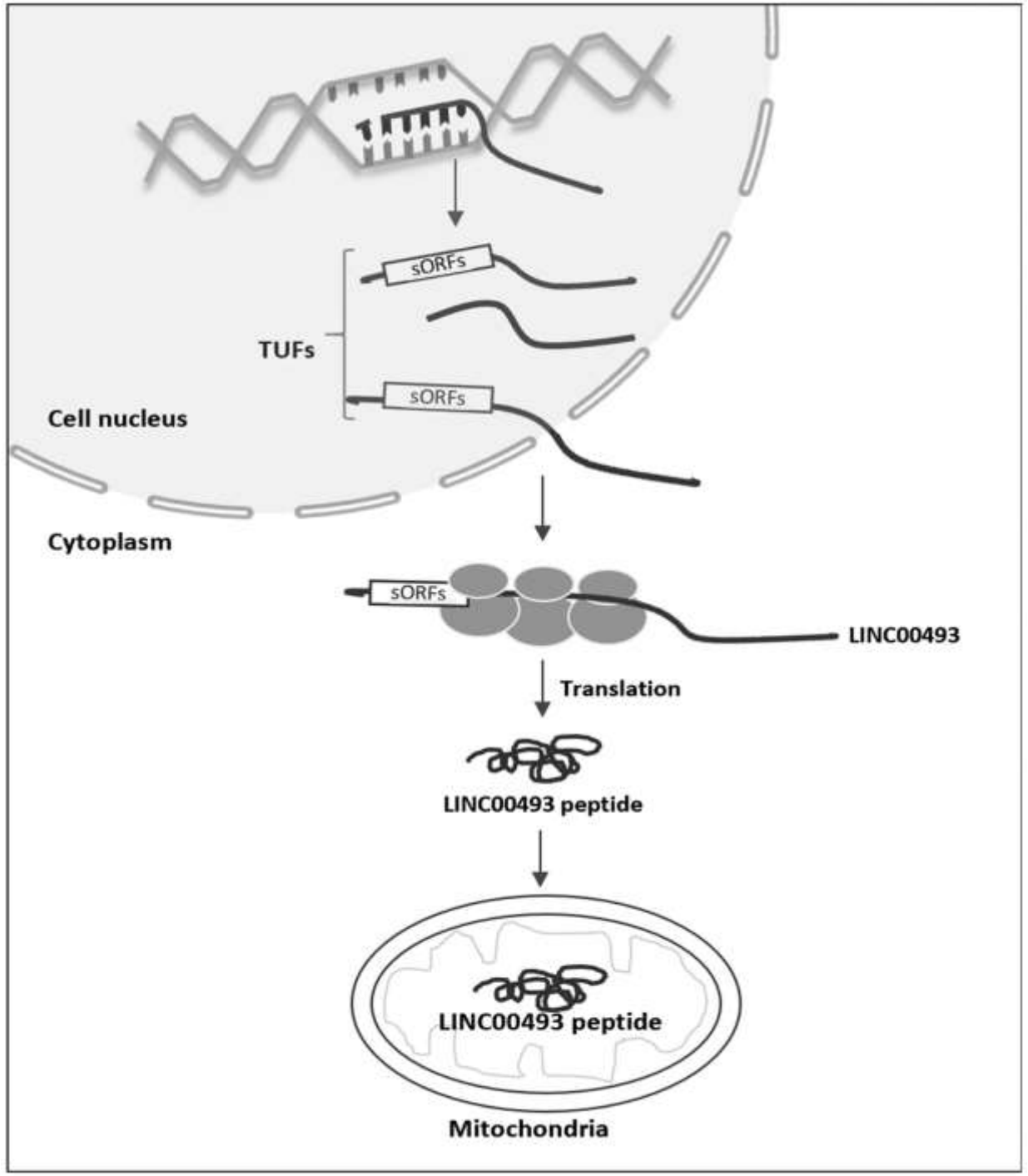
The Journal of Biochemistry, Volume 169, Issue 4, **April 2021**, Pages 421–434, https://doi.org/10.1093/jb/mvaa143

Identification and analysis of short open reading frames (sORFs) in the initially annotated noncoding RNA LINC00493 from human cells

- **We confirmed the translation of one ncRNA, LINC00493**, by luciferase reporter assaying and western blotting of a FLAG-tagged LINC00493 peptide.
- While proteomic analysis revealed that **the LINC00493 peptide interacts with many mitochondrial proteins**, immunofluorescence assays showed that its peptide **is mitochondrially localized**.
- Our findings indicate that **some transcripts annotated as ncRNAs encode peptides and that unannotated peptides may perform important roles in cells**.

The Journal of Biochemistry, Volume 169, Issue 4, **April 2021**, Pages 421–434, https://doi.org/10.1093/jb/mvaa143

Identification and analysis of short open reading frames (sORFs) in the initially annotated noncoding RNA LINC00493 from human cells



While proteomic analysis revealed that **the LINC00493 peptide interacts with many mitochondrial proteins**, immunofluorescence assays showed that its peptide **is mitochondrially localized**. Our findings indicate that **some transcripts annotated as ncRNAs encode peptides and that unannotated peptides may perform important roles in cells**.

J Biochem, Volume 169, Issue 4, April 2021, Pages 421–434, https://doi.org/10.1093/jb/mvaa143

- The distribution of ribosome footprints can provide insights into the mechanism of translational control (for example, it can be used to identify:
- **a) regulatory translational pauses** and
- **b)** translated **upstream open reading frames (uORFs)**

A) Regulatory translational pauses (Ribosome pauses)

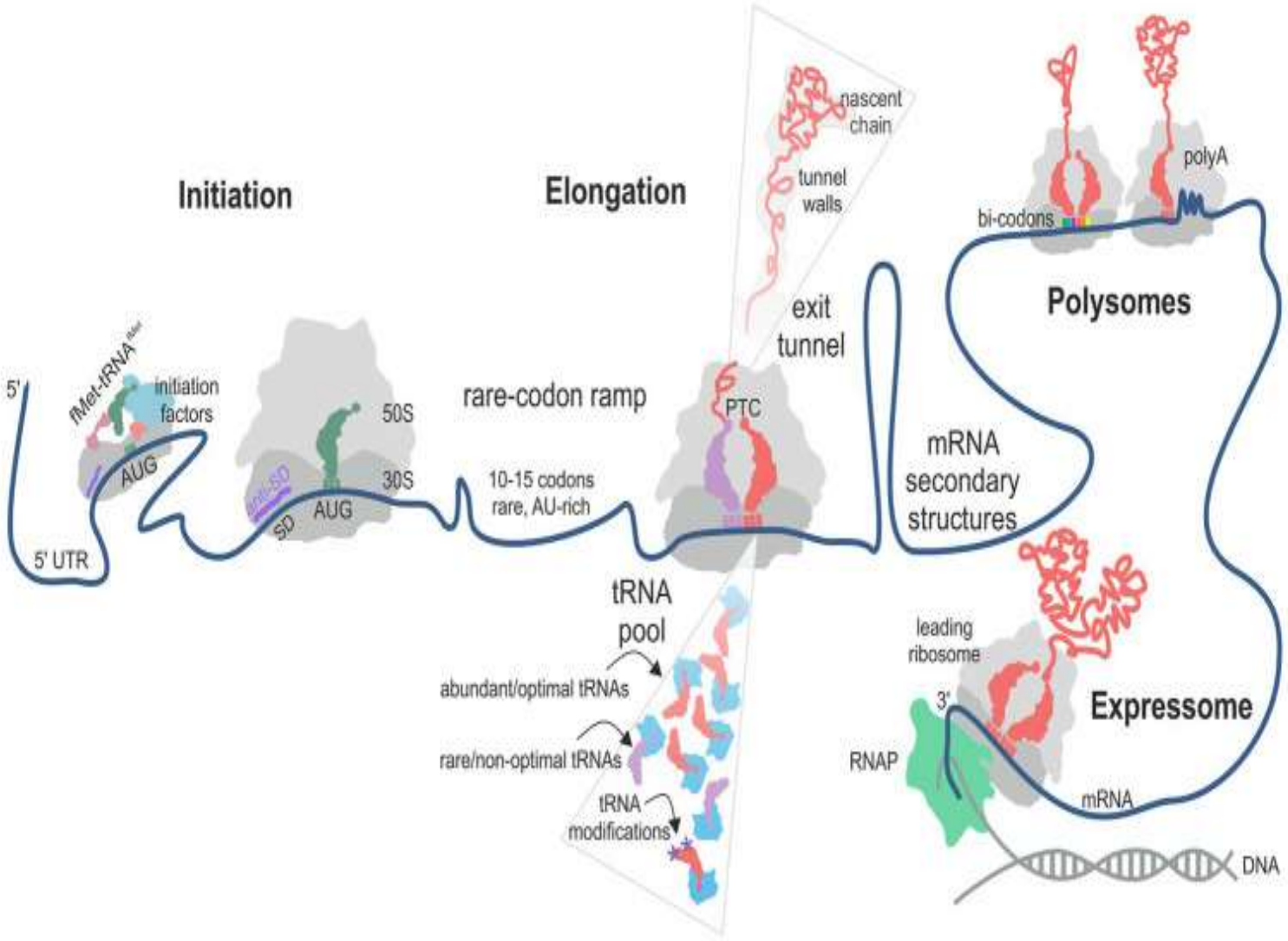
- **Translational efficiency (TE)** of a given mRNA depends on **how often the ribosomes initiate synthesis of a new polypeptide** and **how quickly they read the coding sequence** to produce a full-length protein.
- The pace of ribosomes along the mRNA is **not uniform**: **periods of rapid synthesis are separated by pauses**.
- Ribosome pausing affects translational efficiency and protein folding.

Front Microbiol. 2020; 11: 619430. Published online 2021 Jan 11. doi: 10.3389/fmicb.2020.619430

- In principle, translation can be regulated at any of its steps, i.e., initiation, elongation, termination, or ribosome recycling. The initiation step, at which the ribosome selects the mRNA and finds the open reading frame (ORF), to a large extent, controls the frequency at which a given mRNA is translated (Figure).
- **The basal translation level is determined by** the accessibility of the ribosome binding site on the mRNA, the nature of the start codon, the **position of the Shine-Dalgarno (SD) sequence** relative to the start codon and its complementarity to the anti-SD sequence in the 16S rRNA, and **the presence of A/U rich sequences** that may be specifically recognized by ribosomal protein bS1.

- **The accessibility of the ribosome binding site can change depending on the environment conditions due to ligand- or temperature-induced re-folding of the mRNA** or its interactions with proteins,

Front Microbiol. 2020; 11: 619430. Published online 2021 Jan 11. doi: 10.3389/fmicb.2020.619430



Front Microbiol. 2020; 11: 619430. Published online 2021 Jan 11. doi: 10.3389/fmicb.2020.619430

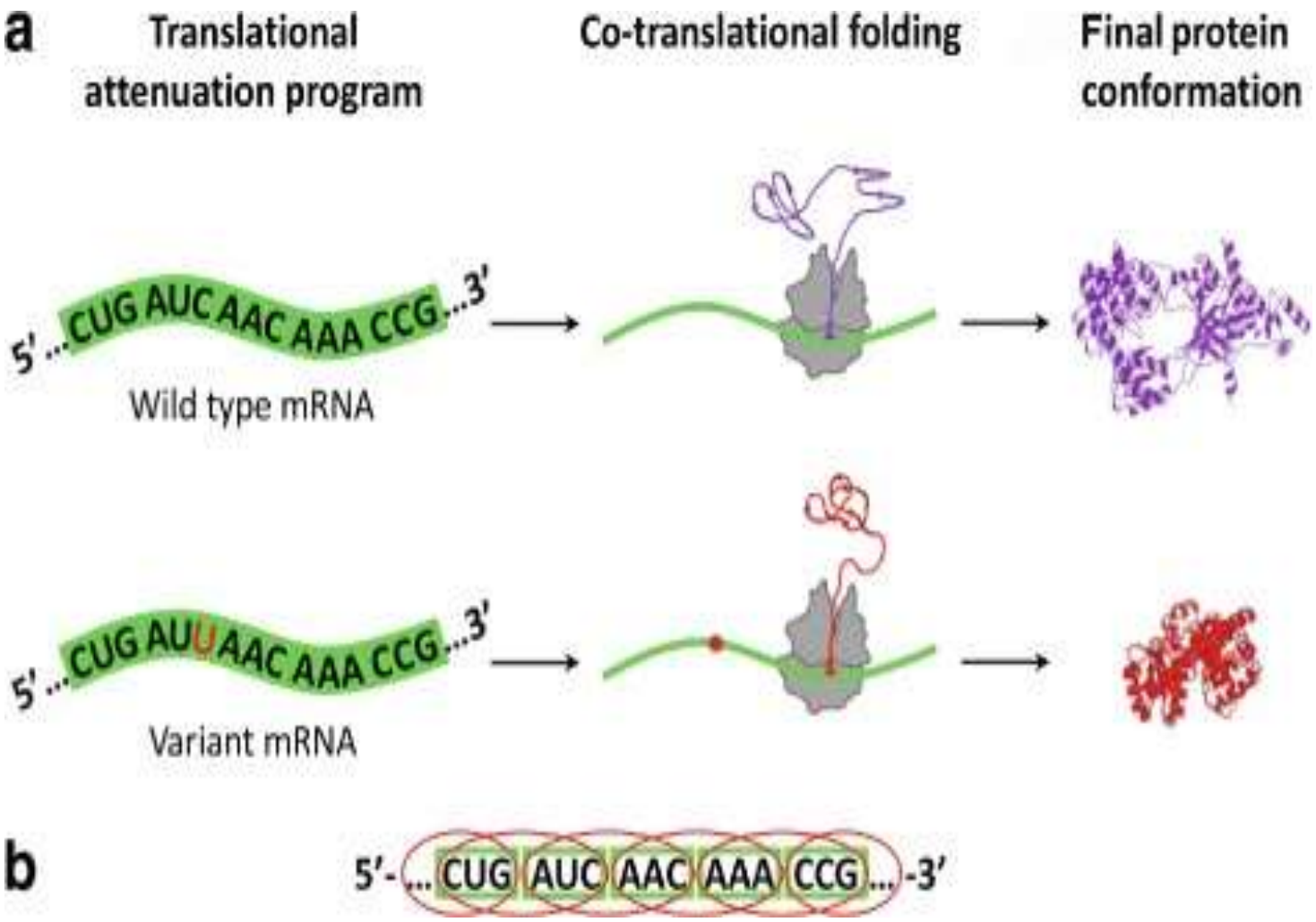
Factors contributing to translational efficiency (TE) and protein folding in bacterial translation.
The TE of an mRNA is largely determined at the **translation initiation step** **when the 30S ribosomal subunit is recruited to the start codon on the mRNA**. In some cases, next 30S subunit can be recruited to a stand-by site upstream of the initiation site. **A rare-codon ramp of 10-15 A/U-rich codons at the beginning of the coding region can increase TE** by disfavoring mRNA secondary structures at the start codon.
During translation, rare codons, low-abundance aa-tRNAs, lack of tRNA modifications and the interactions of the nascent chain with the polypeptide exit tunnel of the ribosome may cause ribosome pausing. mRNA secondary structures can regulate ribosome occupancy at the upstream sequences. Some mRNA contexts, such as particular bi-codons, poly(Pro) and poly(Lys) sequences, cause rearrangements in the **peptidyl transferase center (PTC)** and **promote formation of unusual structures in the A-site**, thereby promoting ribosome stalling (задержка рибосомы).

Bicodons

- A large number of human diseases have been associated with single nucleotide polymorphisms (SNPs). These small genetic variants can occur in gene-coding regions or in non-coding regions (introns or intergenic regions).
- SNPs in a coding region can result in a codon that encodes a different amino acid (missense mutation) or in a premature stop signal (nonsense mutation).
- Depending on the position and/or the chemical properties of the replaced amino acid, these mutations can lead to nonfunctional proteins, causing human genetic diseases such as ***epidermolysis bullosa, sickle-cell anemia, mandibuloacral dysplasia , SOD1-mediated amyotrophic lateral sclerosis*** and ***cancer***, among others.
- Due to **the degeneracy of the genetic code** (i.e., the existence of more codons than the number of different amino acids and stop signals), a SNP does not necessarily change the amino acid sequence of a translated protein. These types of substitutions are known as **synonymous SNPs (sSNPs)** and, for a long time, they were considered as silent mutations because it was assumed that they had no phenotypic consequences.
- However, **strong evidence currently supports** the fact that synonymous codons have phenotypic consequences, and the notion that sSNPs are innocuous has dramatically changed in the last decade. Synonymous mutations can lead to disease.

Bicodon bias can determine the role of synonymous SNPs in human diseases

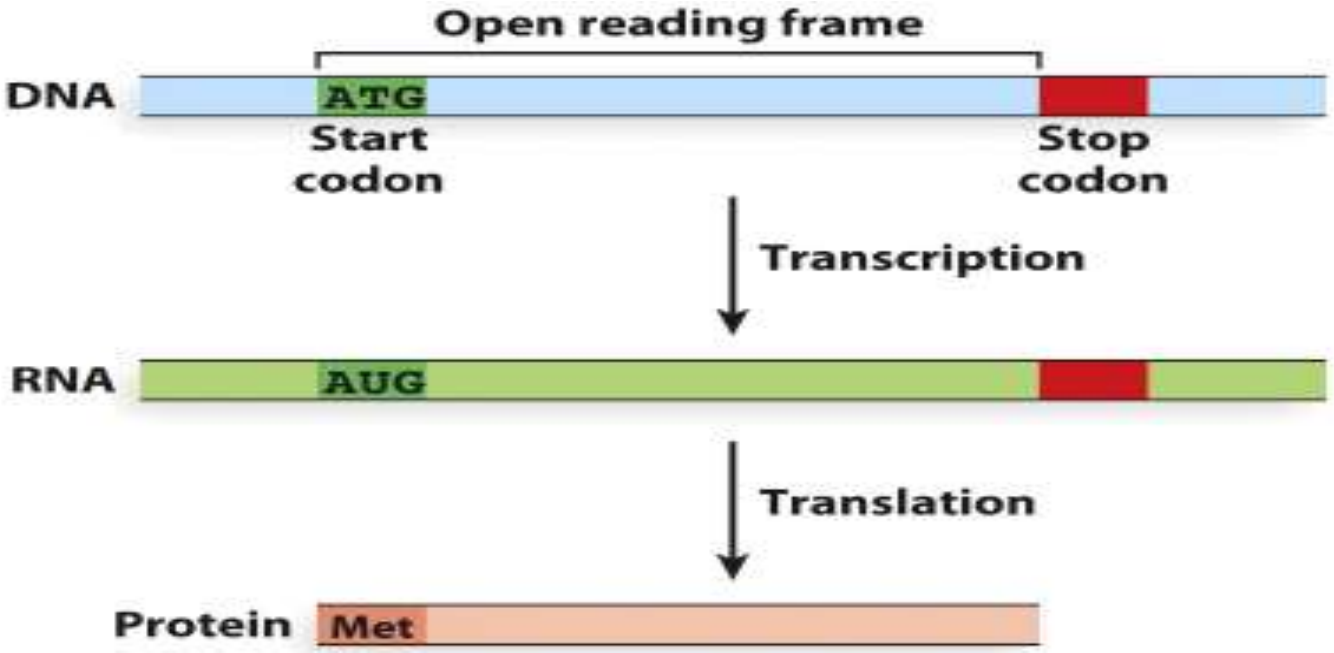
BMC Genomics. 2017; 18: 227. Published online 2017 Mar 13. doi: 10.1186/s12864-017-3609-6



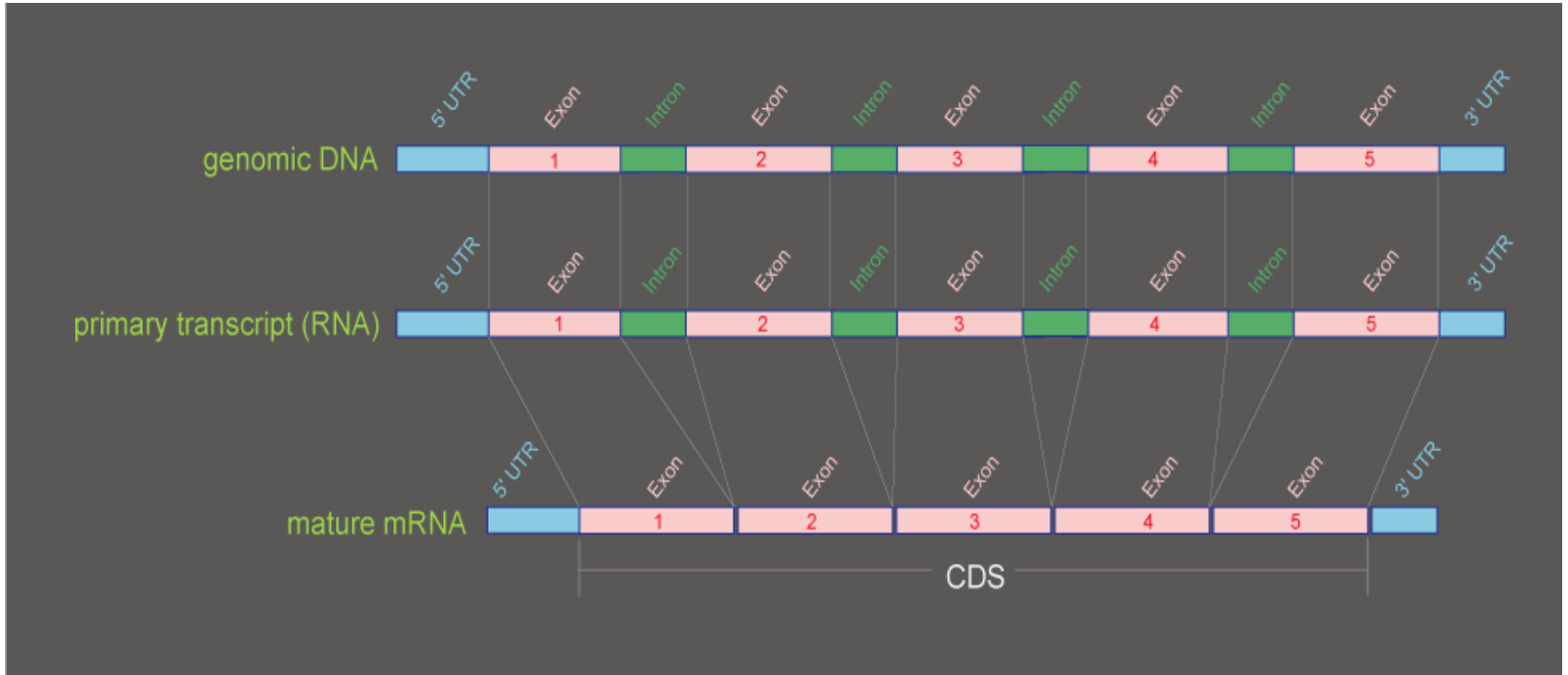
A Graphical representation of how a sSNP can alter the ribosome-mediated translational attenuation program and, ultimately, final protein conformation.
Consequently, this can affect protein function leading to pathological phenotypes. b Simplified graph showing codons (green boxes) and bicodons (red ellipses) within an mRNA sequence. Owing to the superposition of bicodons, each codon is part of two different bicodons in a given open reading frame. **Alterations of the ribosome-mediated translational attenuation program**, with the resulting alternative protein conformations and functions, **could be encoded by bicodons**, rather than by codons

The open reading frame (ORF)

Translated regions of mRNA



The 5′ leader (traditionally referred to as the 5′ untranslated region=**5'UTR**) of mRNAs are one well-studied source of protein synthesis regulation [6,7,8,9]. 5′ leaders can regulate the synthesis of the main coding sequence (CDS) product



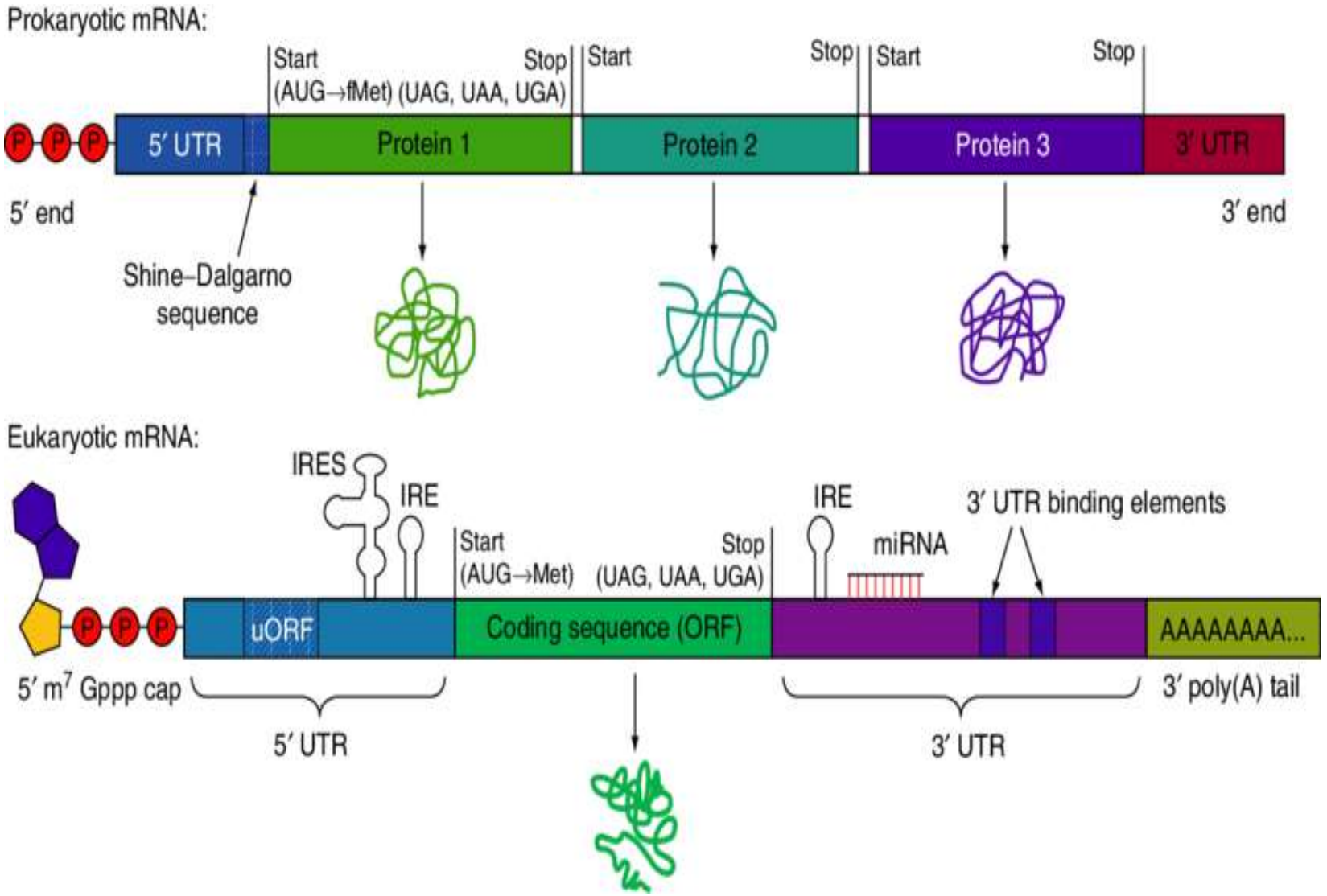


Figure 1 Schematic diagram of prokaryotic and eukaryotic mRNA

5′UTRThe 5′UTR, or leader sequence, begins at the 5′terminal end and ends one nucleotide before the AUG start site. In eukaryotes, this tends to be long, anywhere from 100 to several thousand nucleotides.

In warm-blooded eukaryotes, this is a GC rich region with a G+C percentage of about 60%. This long stretch of nucleotides contains features often used to regulate translation of the mRNA.

While termed the untranslated region, a portion of the 5′UTR is sometimes translated to make a protein product which can then regulate translation of the main coding region of the mRNA. These coding sequences, termed **upstream open reading frames (uORF) contain their own initiation codon** and are fairly common as a means to regulate translation of the mRNA, occurring in 35–50% of all human mRNA.

Messenger RNA (mRNA): The Link between DNA and Protein. Available from: ttps://www.researchgate.net/publication/305698915_Messenger_RNA_mRNA_The_Link_between_DNA_and_Protein

b) **Upstream open reading frames (uORFs)**

- uORFs are small ORFs (open reading frame) **located in the 5′ untranslated regions (5'UTRs) of many eukaryotic mRNAs** (Fig.). uORFs can regulate eukaryotic gene expression.
- Recent genome-wide analyses revealed that **10%-50% of eukaryotic genes contain at least one uORF**.
- Since the **ribosomes of eukaryotes scan mRNAs from their 5′ ends**, the presence of a uORF can negatively modulate the translational efficiency of the main ORF downstream.
- uORFs regulates processes such as stress responses and feedback regulation of biosynthesis.

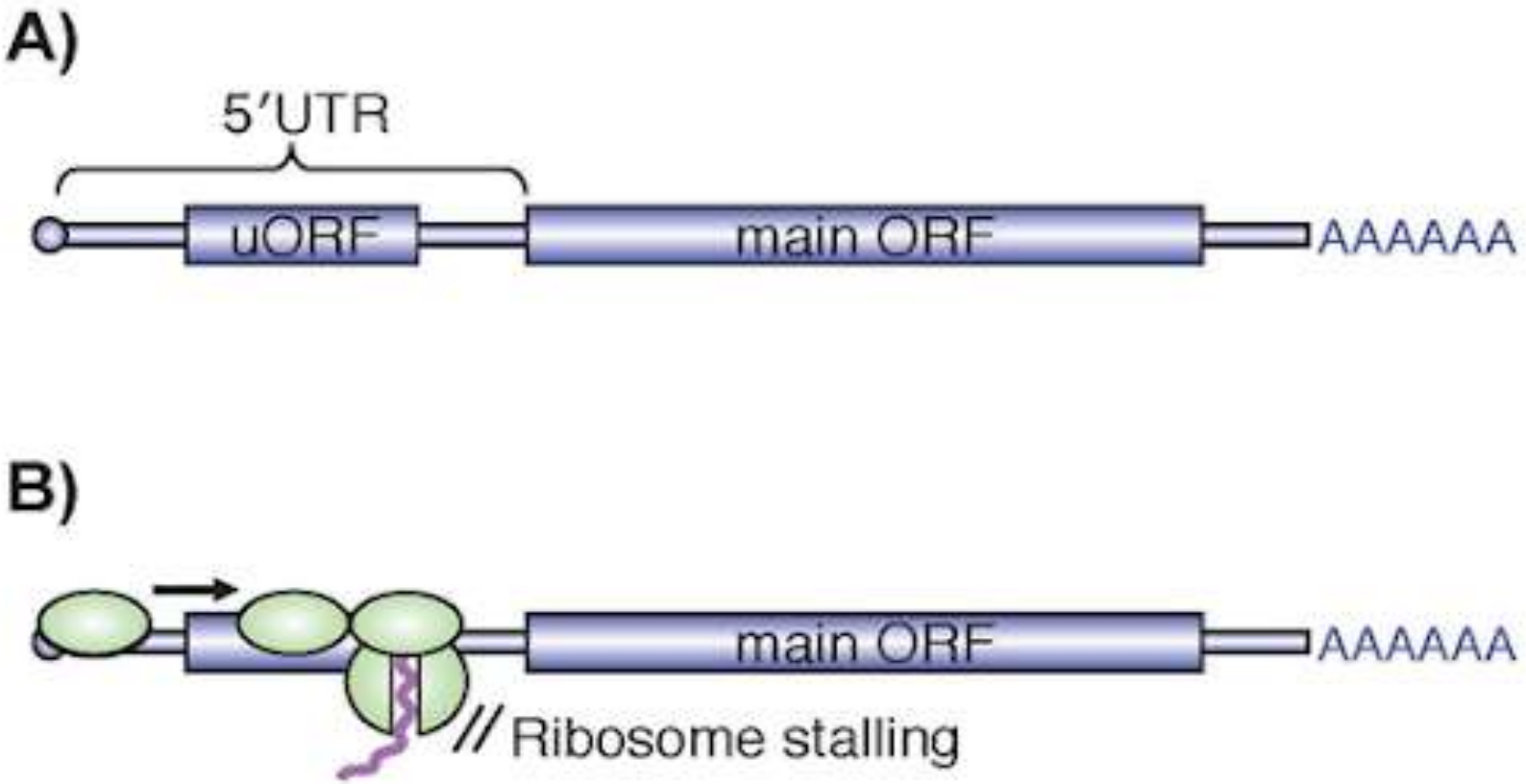


Fig. 1. A model for translational regulation mediated by a uORF-encoded peptide

(A) Schematic representation of a transcript containing a uORF.
(B) If a uORF-encoded nascent peptide causes ribosome stalling, the stalled ribosome blocks the progression of other scanning ribosomes, and therefore results in translational repression of the downstream main ORF.

Several uORFs affect the translation of the main ORF. Feedback regulation of biosynthesis:

For example, in translational feedback regulation of the ***Neurospora crassa arg-2*** gene, which encodes the small subunit of arginine-specific **carbamoyl phosphate synthetase**, **the 24-residue nascent peptide** encoded by the arg-2 uORF **causes ribosome stalling at the stop codon of the uORF** in response to arginine.
The stalled ribosome, in turn, prevents other scanning ribosomes from reaching the initiation codon of the main ORF, resulting in translational inhibition of the main ORF.