Post-transcriptional regulation

Guest lecture by:
Eva Maria Novoa

Outline

1. Is post-transcriptional regulation important?
2. Basic concepts on translation
3. How can we measure translation?
   3.1. mRNA levels VS protein levels
   3.2. Ribosome profiling
4. Codon evolution
   4.1. Basic concepts: codon usage bias, tRNAisoacceptor bias
   4.2. Measures to quantify codon bias
   4.3. The story is more complex: mRNA and tRNA modifications
   4.4. The importance of understanding codon evolution. Examples of applications.
5. Translational regulation. Tuning the protein levels.
   5.1. Regulation at tRNA level
   5.2. Regulation at mRNA level
   5.3. Regulation through cis- and trans-regulatory elements

mRNA levels tend to be used as a proxy for protein levels

There is a clear correlation between mRNA levels and protein abundance BUT is very FAR from being perfect.
Post-transcriptional regulation

There is a clear correlation between mRNA levels and protein abundance BUT is very FAR from being perfect.

Differential translation elongation rates
- codon usage bias
- mRNA adaptation
- RNA editing

Differential translation initiation rates
- AUG frequency
- TOP presence
- Type of initiation
  - cap-dependent/ IRES
- 3' end structures

Differential translation termination rates
- Termination codon identity

mRNA degradation rates
- polyA length
- capping
- mRNA editing
- 3' end structures

Protein degradation rates
- PEST sequences
- protein stability
- unstructured regions
- Presence of polar AA
- cis- and trans-regulatory elements
- AU-rich elements
- miRNAs and other small RNAs
- RNA binding proteins

2. Basic concepts protein translation

Transcribed RNA is not yet ready to be translated

One pre-mRNA can generate many different mRNAs: ALTERNATIVE SPLICING

tRNAs: the "adaptor" molecules

Codon-anticodon pairing

Adapted from Roy and Ibba, Nature 2006
Basic concepts on translation

Adapted from Roy and Ibba, Nature 2006

The Genetic Code:

3. How can we measure translation?

Measuring translation

Translation efficiency = [mRNA]/[Protein]

Methods:

1. Measuring mRNA and protein levels.
   BUT: not regarding synthesis/degradation rates

2. Drugs inhibiting transcription/translation
   BUT: drugs have side effects altering translation

3. Artificial fusion of proteins with tags
   BUT: protein tagging can affect protein stability
Measuring translation

Translation efficiency = [mRNA]/[Protein]

Methods:

1. Measuring mRNA and protein levels.
   BUT: not regarding synthesis/degradation rates

2. Drugs inhibiting transcription/translation
   BUT: drugs have side effects altering translation

3. Artificial fusion of proteins with tags
   BUT: protein tagging can affect protein stability

4. Pulse labelling with radioactive nucleosides or amino acids (SILAC)
   BUT: no information on dynamic changes (snapshot of the resulting mRNA and protein levels after X hours)

Ribosome profiling: method

Ribosome profiling: measuring protein translation at subcodon resolution

Why ribosome profiling?

1. Better approach to protein abundance

2. It is independent of protein degradation so better for measuring translation efficiency (compared to the protein abundance/mRNA ratio)

3. Allows to measure codon-specific translation rates
How can we know which codon is it decoding?

Ingolia et al. Science 2009

Mapped ribosome footprints

Decipher translating codon (depends on footprint length)

Mapped translating codon profile

Ingolia et al. Science 2009

Combining drugs with ribosome profiling

Michel and Baranov, 2013

Cyclohexamide

Harringtonine

Ribosome profiling: more than just quantifying translation

A. Prediction of alternative isoforms

Harringtonine

No Drug
Ribosome profiling: more than just quantifying translation

A. Prediction of alternative isoforms

B. Prediction of short non-annotated ORFs

C. Comparing translation across different conditions

D. Comparing translation across life cycle / stages

4. Codon evolution

* 4.1. Basic concepts*

Codons are not used with equal frequencies

Within species: Codons are not used with equal frequencies throughout different conditions.

Across species: Codons are not used with equal frequencies among different species.

RNA stability
Strand-specific mutational bias
Transcriptional efficiency
**Synonymous codons... but NOT equivalent!**

Codons are not used with equal frequencies

- within species
- across species

---

**A “Silent” Polymorphism in the MDR1 Gene Changes Substrate Specificity**

Chuwa Kimchi-Sarfati, Jung Mi Oh, Jiwon Kim, Zhun B. Sauro, Anna Maria Caloggero, Suneel V. Anadull, and Michael G. Mottner

Science, 2007

---

**tRNA gene copy numbers**

- What about tRNAs?
- tRNA isoacceptors are not used with equal frequencies

---

**tRNA gene copy numbers VS tRNA abundance**

- Unicellular organisms (e.g., yeast)
- Multicellular organisms (e.g., human)

---

**Comparing tRNA usage and codon usage**

<table>
<thead>
<tr>
<th>tRNAs: Relative gene frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGF = CN / MCu</td>
</tr>
<tr>
<td>CN = tRNA Copy Number of the i-th isoacceptor</td>
</tr>
<tr>
<td>MCu = average Copy Number of the j-th amino acid</td>
</tr>
</tbody>
</table>

**Codons: Relative synonymous codon usage**

\[ \text{RSCU}_i = \frac{CN_i}{MCu} \]

- CN = tRNA Codon Usage of the i-th codon
- MCu = average Codon Usage of the j-th amino acid

- H. sapiens

<table>
<thead>
<tr>
<th>tRNA</th>
<th>N</th>
<th>BGF</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGC</td>
<td>1</td>
<td>0.53</td>
</tr>
<tr>
<td>GGC</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GGC</td>
<td>0</td>
<td>0.83</td>
</tr>
</tbody>
</table>

- E. coli

<table>
<thead>
<tr>
<th>Codon</th>
<th>N</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCU</td>
<td>750096</td>
<td>1.04</td>
</tr>
<tr>
<td>GCC</td>
<td>1127679</td>
<td>1.60</td>
</tr>
<tr>
<td>GCA</td>
<td>654371</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Example: Homo sapiens

Source: www.kazusa.or.jp

---

**Matching the codon with its corresponding anticodon**

Example:

<table>
<thead>
<tr>
<th>aa codon</th>
<th>anticodon</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>TTC</td>
</tr>
<tr>
<td>F</td>
<td>GTA</td>
</tr>
<tr>
<td>L</td>
<td>CTA</td>
</tr>
<tr>
<td>T</td>
<td>AUA</td>
</tr>
<tr>
<td>S</td>
<td>UCA</td>
</tr>
<tr>
<td>Y</td>
<td>UTA</td>
</tr>
<tr>
<td>T</td>
<td>ACA</td>
</tr>
<tr>
<td>W</td>
<td>AGA</td>
</tr>
<tr>
<td>Stop</td>
<td>TGA</td>
</tr>
</tbody>
</table>

Example: Homo sapiens

Source: www.kazusa.or.jp

But... still not known (and may be different for each organism)
Matching the codon with its corresponding anticodon

Toy example: decoding Ala in *H. sapiens*

<table>
<thead>
<tr>
<th>Codon</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGC</td>
<td>29</td>
</tr>
<tr>
<td>GGC</td>
<td>0</td>
</tr>
<tr>
<td>GCG</td>
<td>5</td>
</tr>
<tr>
<td>UGC</td>
<td>9</td>
</tr>
</tbody>
</table>

Direct matching RGF: RSCU

\[
\begin{align*}
\text{AGC} & \rightarrow \text{GCU} \\
\text{UGC} & \rightarrow \text{GCA} \\
\text{GGC} & \rightarrow \text{GCC} \\
\text{CGC} & \rightarrow \text{GCG}
\end{align*}
\]

4. Codon evolution

*4.2. Measures of codon bias*

\[ \text{CBI} = \frac{o_{\text{opt}} - o_{\text{rand}}}{o_{\text{exp}} - o_{\text{rand}}} \]

\[ \text{CBI} \rightarrow (-1, 1) \]

What will be the CBI of a gene with random codon usage?

Limitations: As the CBI, it requires the definition of a “REFERENCE SET” of proteins (highly expressed proteins, such as ribosomal proteins, are generally used).

b) CBI: Codon Bias Index

c) CAI: Codon Adaptation Index

\[ \text{CAI} = \exp \left( \frac{1}{L} \sum_{i,j} \log \left( \frac{w_i}{f_{ij}} \right) \right) \]

\[ w_{ij} = \frac{f_i}{\max(f)} \quad i, j \in \text{[synonymous codons for amino acid]} \]

Limitations: Requires the definition of a “REFERENCE SET” of proteins (highly expressed proteins, such as ribosomal proteins, are generally used).
d) \( N_c \): Effective number of codons

- Measures the total number of different codons used in a sequence, i.e., measures the bias toward the use of a smaller subset of codons, away from equal use of synonymous codons.

- \( N_c \rightarrow [20,61] \), where \( N_c=20 \) when only one codon is used per amino acid and \( N_c=61 \) when all possible synonymous codons are used with equal frequency.

Steps:
1. Compute homozygosity for each amino acid (estimated from the squared codon frequencies).
2. Obtain effective number of codons per amino acid \( N_c \rightarrow (1, k) \), where \( k \) is the number of synonymous codons.
3. Overall number of effective codons \( N_c = \sum_k n_k N_{c,k} \).

Advantages:
- Does not require knowledge of tRNA-codon pairing
- Does not require "REFERENCE SET"

Limitations:
- Does not take into account tRNA pool

Wright, 1990

---

Example

Sequence = GCC GCC GCC GCC GCC GCC GCC

\begin{tabular}{lcc}
\hline
Ala & Codon & RSCU & Freq \\
\hline
GCU & 1.04 & 0.26 \\
GCC & 1.60 & 0.40 \\
GCG & 0.44 & 0.11 \\
GCA & 0.92 & 0.23 \\
\hline
\end{tabular}

Data from human

"High expressed genes"

---

e) \( tAI \): tRNA adaptation index

- Assumption: tRNA gene copy number has a high positive correlation with tRNA abundance within the cell.

- Measures "how well a gene is adapted to the tRNA pool" (\( W \) = relative adaptiveness).

\[ W_i = \begin{cases} \frac{\sum_j (1 - x_{ij}) GCN_{ij}}{x_{ij}} & \text{if } W_{ij} \neq 0 \\ \text{otherwise} & \end{cases} \]

\[ tAI_p = \left( \frac{\prod_i w_i}{n_i} \right)^{1/\gamma} \]

dos Reis et al., 2004

---

Example

Sequence = GCC GCC GCC GCC GCC GCC GCC

\begin{tabular}{lcc}
\hline
Ala & Codon & RSCU & Freq \\
\hline
GCU & 1.04 & 0.26 \\
GCC & 1.60 & 0.40 \\
GCG & 0.44 & 0.11 \\
GCA & 0.92 & 0.23 \\
\hline
\end{tabular}

Data from human

"High expressed genes"

---

Comparison of the different measures

Do these indexes capture any relevant information on experimental data?

Ex. \( S. \) cerevisiae
**Why choosing the right index is important. Example**

If our aim is to compare codons from diverse amino acids between them, codon frequency (amongst others) might be better than RSCU.

\[
\text{RSCU} = \frac{\text{Observed frequency of codon}}{\text{Expected random frequency of codon}}
\]

Values $(0, N)$

\[
\text{Codon frequency} = \frac{\text{Number of observed codons}}{\text{Number of isoacceptors for that AA}}
\]

Values $(0, 1)$

**4. Codon evolution**

*4.3. The story gets complicated: tRNA modifications*

Some modifications can EXPAND or RESTRICT the wobbling capacity of the tRNA.

**tRNA modifications**

Some modifications can EXPAND or RESTRICT the wobbling capacity of the tRNA.

**tRNA modifications. Example**

a) **INOSINE MODIFICATIONS**

b) **xo5U MODIFICATIONS**

- ADAT: tRNA-dependent ADENOSINE DEAMINASES
- UN: tRNA-dependent URIDINE METHYLTRANSFERASES
**tRNA modifications. Example**

**INOSINE MODIFICATIONS**

\[
\begin{array}{c}
\text{tRNA wobble base} \\
G \\
C \\
A \\
U \\
\text{3rd codon position}
\end{array}
\]

ADAT: tRNA-dependent ADENOSINE DECARboxYlASE

**xo5U MODIFICATIONS**

\[
\begin{array}{c}
\text{tRNA wobble base} \\
A \\
G \\
C \\
U \\
\text{3rd codon position}
\end{array}
\]

UN: tRNA-dependent URIDINE METHYLTRANSFERASE

**tRNA modifications. Consequence**

Modified tRNAs will now be able to decode a codon that before they couldn’t read.

**mRNA modifications**

Some mRNA modifications can ALSO EXPAND or RESTRICT the codon:anticodon pairings, and even change the amino acid sequence.

A-to-I editing also exists on mRNA.

**4. Codon evolution**

* 4.4. The importance of understanding codon evolution*  
Examples of applications

**Applications. 1: Codon optimization for heterologous protein expression**

Heterologous expression of proteins from *Plasmodium falciparum: Results from 1000 genes*

As part of a structural genomics initiative, 1000 open reading frames from *Plasmodium falciparum*, the causative agent of the most deadly form of malaria, were tested in an E. coli protein expression system. Three hundred and thirty-seven of these targets were observed to express, although typically the protein was insoluble. Sixty-three of the targets provided soluble protein in yields ranging from 0.9 to 406.6 mg from one liter of rich synthetic *Plasmodium* genes.
Applications. 2: Predicting coding and non-coding regions of a genome

Species

Alignments of coding & non-coding regions

Codon substitution matrices

Predictions

Looks as coding / exonic

Looks as non-coding / intronic

Lin et al., Bioinformatics 2011

Applications. 2: Predicting coding and non-coding regions of a genome

Species

Alignments of coding & non-coding regions

Codon substitution matrices

Predictions

Alternative readthrough

Stop codon

Jungreis et al., Genome Res 2011

Applications. 3: Predicting codonreadthrough

mRNA

ATG

Stop codon

TGA

TAA

TAG

Start codon

Protein-coding sequence

Alternative readthrough protein-coding sequence

mRNA

ATG

Stop codon

TGA

TAA

TAG

Start codon

Protein-coding sequence

Alternative readthrough stop codon

Jungreis et al., Genome Res 2011
Applications. 3: Predicting codon readthrough

Alternative readthrough protein-coding sequence

mRNA

Start codon

Stop codon

Alternative readthrough stop codon

ATG TGA

TAA

TAG

Typical gene
Codon conservation ends after stop codon

Candidate gene with stop-codon readthrough
Codon conservation continues until next stop codon

Jungreis et al., Genome Res 2011

Applications. 4: Understanding how genes are decoded

Studying patterns of codon usage bias ALONG genes

5. Translational regulation

Why translation regulation?

Steady-state

Regulation through transcription

Regulation through translation

Types of post-transcriptional regulation

5.1. Modulation of tRNA availability
  5.1.1. Changes in levels of tRNA isoacceptors
  5.1.2. Changes in tRNA modifications
  5.1.3. Changes in tRNA aminoacylation levels

5.2. Changes in mRNA
  5.2.1. Changes in mRNA modifications (RNA editing)
  5.2.2. Changes in polyA tail
  5.2.3. Changes in splicing
  5.2.4. Changes in capping
  5.2.5. Changes in the localization of mRNA (import/export to nucleus)

5.3. Cis- and trans-regulatory elements
  5.3.1. RNA interference: miRNAs and siRNAs
  5.3.2. Frameshift events
  5.3.3. Riboswitches

5.4. Many more.... And many still unknown!
Types of post-transcriptional regulation

5.1. Modulation of tRNA availability
5.1.1. Changes in levels of tRNA isoacceptors
5.1.2. Changes in tRNA modifications
5.1.3. Changes in tRNA aminoacylation levels

5.2. Changes in mRNA
5.2.1. Changes in mRNA modifications (RNA editing)
5.2.2. Changes in polyA tail
5.2.3. Changes in splicing
5.2.4. Changes in capping
5.2.5. Changes in the localization of mRNA (import/export to nucleus)

5.3. Cis- and trans-regulatory elements
5.3.1. RNA interference: miRNAs and siRNAs
5.3.2. Frameshift events
5.3.3. Riboswitches

5.4. Many more…. And many still unknown!

Example: Cell cycle-dependent changes in tRNA modification levels in yeast

Mcm5 modification --> preference in decoding AGA codons (Rnr1 is enriched in AGA)

Types of post-transcriptional regulation
5.2. Changes in mRNA

5.2.1. Changes in mRNA modifications

Example: m6A modifications

5.1. Modulation of tRNA availability

5.1.1. Changes in levels of tRNA isoacceptors
5.1.2. Changes in tRNA modifications
5.1.3. Changes in tRNA aminoacylation levels

5.2. Changes in mRNA

5.2.1. Changes in mRNA modifications (RNA editing)
5.2.2. Changes in polyA tail
5.2.3. Changes in splicing
5.2.4. Changes in capping
5.2.5. Changes in the localization of mRNA (import/export to nucleus)

5.3. Cis- and trans-regulatory elements

5.3.1. RNA interference: miRNAs and siRNAs
5.3.2. Frameshift events
5.3.3. Riboswitches

5.4. Many more... And many still unknown!

5.4. Many more... And many still unknown!

Types of post-transcriptional regulation

5.3. Regulation through cis- and trans-regulatory elements

5.3.1. RNA interference: siRNA and miRNA

TAKE HOME MESSAGES

1. Is post-transcriptional regulation important?
   Yes! mRNA levels ≠ protein levels
1. Is post-transcriptional regulation important?
   - Yes! mRNA levels ≠ protein levels

2. Basic concepts on translation
   - The genetic code is NOT universal
   - ‘Preferred’ tRNA-codon pairs are DYNAMIC
   - Synonymous mutations are NOT equivalent

3. How can we measure translation?
   - Ribosome profiling allows to measure translation with subcodon resolution

4. Codon evolution
   - Albeit all this, you CAN model translation/ codon evolution, and build tools that help to:
     - Increase translation efficiency/folding of proteins in heterologous systems
     - Predict coding regions
     - Understand cell type-specific translation patterns
     - Compare translation between healthy and disease states

5. Translational regulation. Tuning the protein levels.
   - Many different ways to achieve post-transcriptional regulation
   - Interconnection between the different strategies?

---

Doubts or questions? Please email me: enovoa@mit.edu