Monitoring mRNA metabolism

Margaux Michel, Carlo Bäjen, Carina Demel, Michael Lidschreiber, Kerstin Maier, Björn Schwalb, Phillipp Torkler, Patrick Cramer
Department of Molecular Biology

Messenger-RNAs (mRNAs) are central molecules in cells that dictate which protein is synthesized at which time and to what extent. Understanding the dynamics of mRNA synthesis and degradation is thus of central importance. Here, we describe recent developments in our laboratory that enable us to globally monitor mRNA metabolism in a quantitative manner.

The concept of mRNA metabolism

The life cycle of mRNAs in eukaryotic cells starts with synthesis of pre-mRNA molecules in the nucleus during the transcription of protein-coding genes (Fig. 1). These pre-mRNAs are processed co-transcriptionally by capping of the RNA 5′ end, splicing of introns, and formation of the 3′ poly-A tail. Mature mRNAs are then exported to the cytoplasm and used as templates for protein synthesis during translation. Finally, mRNAs are degraded and their building blocks are recycled.

Over the last decade it has become clear that the different phases in the mRNA life cycle are intimately coupled. For example, 5′-capping occurs as soon as the nascent pre-mRNA first reaches

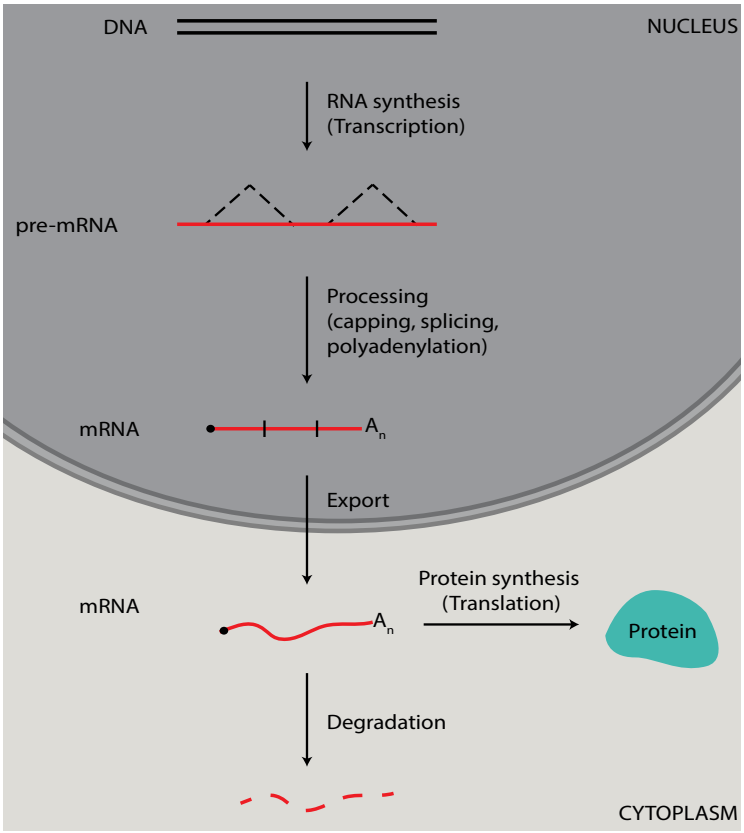

Fig. 1. Overview of cellular mRNA metabolism.

Abb. 1. Übersicht über den zellulären mRNA-Metabolismus.

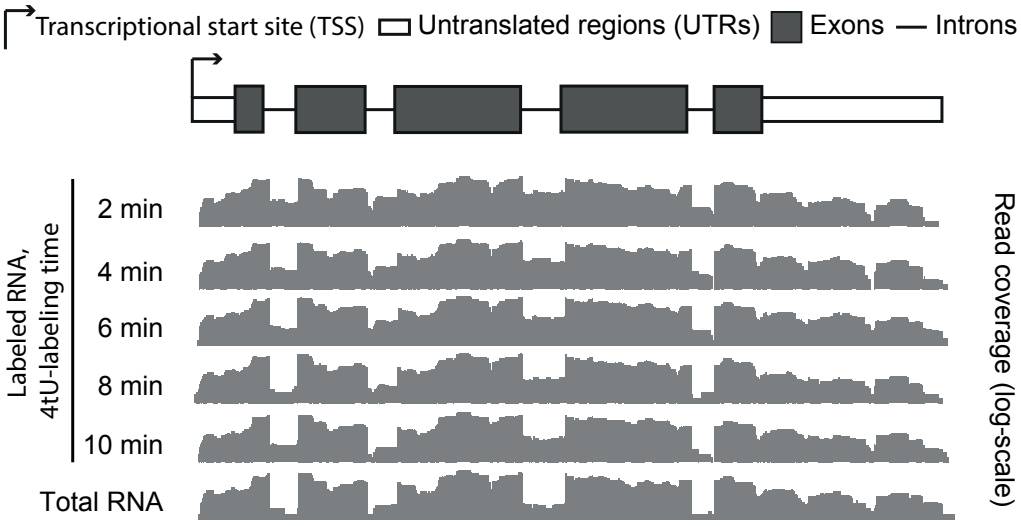

Fig. 2. Metabolic RNA labeling pulses reveal short-lived RNAs including intronic RNA regions. The vertical lines represent accumulated reads derived from sequencing newly synthesized labeled RNA in fission yeast cells. Only one example gene is depicted, although data extend to the entire genome.

Abb. 2. Die metabolische Markierung der RNA in Zellen ermöglicht es, kurzlebige RNA-Regionen wie etwa Introns zu detektieren. Senkrechte Linien stellen akkumulierte Sequenz-Reads aus der Spalthefe dar. Je kürzer die Zellen metabolisch markiert wurden, desto mehr kurzlebige RNAs konnten detektiert werden. Es ist nur ein Gen als Beispiel dargestellt, die Daten erstrecken sich aber über das ganze Genom.

the surface of the transcribing enzyme RNA polymerase II (Martinez-Rucobo et al., 2015). Splicing also occurs co-transcriptionally, although the underlying mechanisms remain unclear.

RNA 3′-processing is tightly linked to transcription of the polyadenylation site at the end of genes (Schreieck et al., 2014), and to the export of mRNAs. The degradation of mRNAs is generally coupled to translation. To reflect the observation that the different steps in mRNA life cycle are interconnected, we use the term mRNA metabolism.

Limitations of traditional methods to monitor mRNA

Traditionally, the levels of cellular mRNAs were measured by standard transcriptomics. RNA was isolated from cells and the amount of different mRNA species was measured using microarrays from companies such as Affymetrix or NimbleGen or later with the use of next generation sequencing technologies from Illumina, Roche, or Agilent. The latter method is known as RNA-Seq and is widely used to determine the gene expression patterns in cells. An advantage of RNA-Seq is its unbiased nature, that is all RNA molecules can in principle be detected, whereas on microarrays only those RNAs are observed for which complementary probes are present.

A disadvantage of such standard transcriptomic methods is that they only measure the total levels of RNAs, and thus do not allow to determine what the contributions of RNA synthesis and RNA degradation rates are to achieve these RNA levels in cells. Therefore, standard methods cannot determine whether changes in gene expression involve changes in RNA synthesis or RNA degradation or both, i.e. whether gene regulation occurs during transcription or on the level of RNA turnover.

Metabolic RNA labeling and dynamic transcriptome analysis

To overcome these limitations, we use metabolic RNA labeling, which directly measures the rates of RNA synthesis in cells (Miller et al., 2011). Cells are grown in media containing 4-thiouracil or 4-thiouridine, which is rapidly converted to 4-thio-UTP. RNA polymerases readily incorporate 4-thio-UTP instead of normal UTP into nascent RNA. Cellular RNA that has been labeled this way (typically for 5 minutes) is then purified based on its containing thiol groups using an affinity capture method, and is monitored by microarray analysis or next-generation sequencing. Metabolic RNA labeling enables detection of short-lived RNAs, including introns (Fig. 2).

The amount of newly synthesized, labeled mRNA provides mRNA synthesis rates. If total mRNA levels are also measured, it is then possible to estimate mRNA degradation rates. This requires fitting of the data with a kinetic model that describes the amount of total RNA as a function of RNA synthesis and degradation rates. For example, a particular mRNA that is more abundant but synthesized at the same rate than another mRNA must have a longer half-life, and thus a lower degradation rate. The required kinetic modeling was pioneered by our collaborator Achim Tresch (MPI for Plant Breeding Research, Cologne) and his research group and later adopted and further developed by Julien Gagneur (Gene Center Munich) and his team. We called the resulting method dynamic transcriptome analysis or DTA.

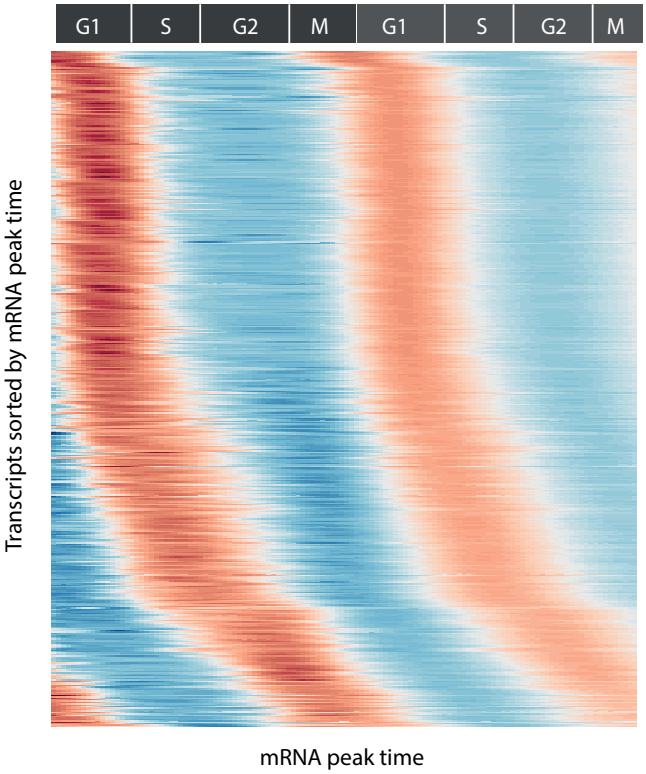

Fig. 3. Comparative dynamic transcriptome analysis (cDTA) monitors periodic mRNA synthesis and degradation during the cell cycle in budding yeast. Red and blue horizontal lines represent high and low levels of mRNA synthesis for hundreds of periodically expressed genes at different times during the cell cycle.

Abb. 3. Die vergleichende dynamische Transkriptomanalyse (cDTA) ermöglicht es, die periodisch wiederkehrende Synthese und den Abbau von mRNAs während des Zellzyklus der Bäckerhefe zu verfolgen. Hohe und niedrige mRNA-Synthese ist in rot und blau für im Zellzyklus periodisch exprimierte Gene gezeigt.

Application of DTA to yeast cells revealed that certain classes of genes such as genes encoding for cell cycle regulators or transcription factors produce only short-lived mRNAs (Eser et al., 2014). Also it was found that mRNAs required at high levels in cells are both synthesized at high rates and show low degradation rates. DTA allowed us to monitor changes in mRNA metabolism over time and showed that both changes in mRNA synthesis and degradation occur in a coordinated fashion during a cellular stress response. This method is more sensitive and has a higher temporal resolution than standard transcriptomics.

Buffering of mRNA levels

Another complication of standard transcriptomics is the inability to put two experiments on the same scale, that is to normalize data between samples. Such normalization is not required if one wishes to only see relative changes in mRNA metabolism between different mRNAs, but it is absolutely necessary if one wants to detect global effects that relate to all mRNAs.

We established a procedure that allows for the required normalization by introducing an internal standard to experiments (Sun et al., 2012). When we applied the resulting comparative DTA (cDTA) protocol to mutant yeast cells, we observed a very interesting phenomenon that we called RNA synthesis-degradation compensation or mRNA level buffering (Sun et al., 2013). Briefly, when RNA synthesis was perturbed by mutation of the polymerase, RNA degradation rates changed in a compensatory manner, resulting in similar mRNA levels in the mutant cells. Also, when a mutation changed cellular mRNA degradation rates, mRNA synthesis compensated for this defect, again leading to relatively stable mRNA levels.

The mechanisms underlying such buffering of mRNA levels remain poorly understood. By systematically testing many different mutant yeast strains, however, we could show that the mRNA degradation enzyme Xrn1 is important for mRNA level buffering (Sun et al., 2013). When the Xrn1 RNA exonuclease is deleted, cells are growing very poorly and accumulate a lot of mRNA, because synthesis still occurs at a high rate whereas mRNA degradation is defective. We proposed a simple mechanism underlying these observations. The mRNA encoding Xrn1 is under the control of its own product, the Xrn1 nuclease. This forms a feedback loop that allows Xrn1 to control not only its own mRNA level, but rather all cellular RNA levels, because it can act on all mRNAs.

Changes in mRNA metabolism during the cell cycle

The cDTA method also enabled us to monitor changes in mRNA synthesis and degradation rates over the cell cycle (Eser et al., 2014). When cell growth is synchronized in a population and all mRNA synthesis and degradation rates were estimated every five minutes, we found that about 10% of all yeast genes change their expression during the cell cycle (Fig. 3). For those cycling genes, subsequent peaks of RNA synthesis and degradation generate high and sharp peaks of mRNAs. If there were only peaks of mRNA synthesis without changes in mRNA degradation, mRNA levels would rise, but would remain high for extended periods of time. As a consequence, gene expression would not

be restricted to defined times in the cell cycle. Thus, both mRNA synthesis and mRNA degradation co-operate during cell cycle gene expression.

Transcriptome surveillance in yeast cells

A very surprising finding from the first RNA-Seq experiments was that in addition to mRNAs and well-known non-coding RNAs (ncRNAs) such as small nuclear RNAs, many new ncRNAs were detected in regions of the genome that were not known to be transcribed. As sequencing methods became more sensitive, it was found that extended regions of genomes in various organisms give rise to such cryptic ncRNA species. Whereas some of these new ncRNAs may have functions, it is commonly believed that many are aberrant RNAs that stem from spurious transcription and must be degraded. This requires cellular surveillance mechanisms that detect aberrant ncRNAs and govern their removal.

We could characterize a mechanism for transcriptome surveillance in yeast cells. To this end we sequenced newly synthesized cellular RNA using a combination of RNA metabolic labeling and next generation sequencing (4tU-Seq). We found that the factor Nrd1, which was known to terminate RNA synthesis of small nuclear RNAs, acts globally to terminate synthesis of ncRNAs (Schulz et al., 2013). In collaboration with the group of Johannes Söding (MPI-BPC), we also found that Nrd1 preferentially binds to nascent RNAs that contain certain tetramer motifs that are depleted in mRNAs, but found in ncRNAs, explaining how the factor is recruited to its target RNAs. The same factor can also stimulate subsequent degradation of the RNAs by the exosome (Tudek et al., 2014). Many of these aberrant ncRNA arise from bidirectional promoter regions, and we are now trying to find out how such bidirectional transcription is controlled in cells.

Globally monitoring pre-mRNA splicing

We have recently adopted our 4tU-Seq protocol to cells from the fission yeast *Schizosaccharomyces pombe* (Eser *et al.*, unpublished). Our data show that short-lived RNA species can be monitored accurately, including pre-mRNA introns that are rapidly spliced out and degraded (Fig. 2). In collaboration with the laboratory of Julien Gagneur we could extend the kinetic modeling to include rates of pre-mRNA splicing, in addition to rates of mRNA synthesis and degradation. We could see that rates of splicing vary for different genes and introns, and that this variation is both a function of the mRNA synthesis rate and of the intron sequence. When introns contain optimal splicing sequences, they are more rapidly removed and degraded *in vivo*. We also found new RNA motifs located in the untranslated regions (UTRs) that are associated with rapid turnover of mRNAs.

The transient transcriptome of human cells

Other unpublished results from our group show the value of the metabolic labeling method for monitoring mRNA metabolism in human cells (Schwalb, Michel, unpublished). Due to the presence of alternative splicing in human cells, the splicing patterns are currently too complicated to be interpreted in a systematic manner. However, due to the sensitivity of the method we are able to detect many new ncRNAs and to observe short-lived ncRNAs that were thus far only observed after components of the RNA degradation machinery were depleted. For example, the synthesis of short ncRNAs arising from bidirectional transcription is directly observed, and RNAs synthesized from regulatory regions of the genome, so-called enhancer-RNAs, are visualized directly. We also observe very short-lived RNA portions at the end of genes that are produced just before transcription terminates. We refer to the entirety of these short-lived RNAs and RNA regions as the transient transcriptome. We could also assign degradation rates to different classes of RNAs. Whereas mRNAs on average are stable for about one hour, enhancer-RNAs have a half-life of only around one minute.

The future is bright

Based on these observations it is clear that metabolic RNA labeling coupled to deep sequencing of the newly synthesized RNAs in cells holds great promise for globally monitoring mRNA metabolism in a very sensitive and non-perturbing way during key cellular processes such as differentiation or the response to stimuli. The method also nicely complements other methods of functional genomics that provide occupancy maps of protein factors over the genome or the transcriptome. Together, these methods hold the promise of gaining a better understanding of the regulatory circuits that underlie genome expression and its regulation. Future projects will make use of these methods to systematically identify active enhancers, to understand how mammalian cells repress spurious transcription, and to monitor mRNA metabolism during the reprogramming of human cells. From such studies we hope to unravel gene expression programs and their modulation by intrinsic and extrinsic perturbations. Ultimately, the regulatory code of the genome may emerge.

Before sequencing, DNA must be fragmented with ultra-focused acoustics.

Vor der Sequenzierung wird die DNA mit ultra-fokussierten akustischen Wellen fragmentiert.

Zusammenfassung

Die im Erbgut eukaryotischer Zellen gespeicherte Information wird exprimiert, indem bestimmte Regionen in RNA-Moleküle umgeschrieben werden (Transkription). Um zu verstehen, wie das Erbgut zur Steuerung der Lebensfunktionen verwendet wird, müssen Methoden entwickelt werden, um die Gesamtheit der RNA-Kopien in Zellen zu erfassen und deren Synthese, Prozessierung und Abbau global zu verfolgen. Hier beschreiben wir unsere Arbeiten der letzten Jahre und die aktuellen Projekte, die darauf abzielen, den RNA-Metabolismus in Zellen zu verfolgen und so die Grundlagen für verschiedene zelluläre Prozesse zu studieren. Dabei verwenden wir unter anderem die sogenannte 4tU-Seq-Methode. Sie beruht darauf, dass man Zellen mit einem Analogon von Uracil füttert, um so während der Transkription neu synthetisierte RNA-Moleküle zu markieren.

Die markierte, neu synthetisierte RNA lässt sich dann isolieren und mithilfe der Sequenzierung analysieren. Gemeinsam mit einer mathematischen Modellierung ermöglicht diese Methode Einblicke in die Dynamik des RNA-Metabolismus. Wie schnell werden welche RNAs synthetisiert? Wie lange leben diese RNAs in der Zelle, bevor sie abgebaut werden? Diese Methode ermöglicht es auch, biologische Prozesse wie zum Beispiel den Zellzyklus oder die Antwort auf einen Hitzestress viel sensitiver und genauer zu untersuchen. Unser Labor hat nach Entwicklung dieser Methoden zuerst grundlegende Mechanismen in der Hefe aufgeklärt, nun wenden wir diese Technik auf menschliche Zellen an. Die Herausforderung der nächsten Jahre wird es sein, die Komplexität der menschlichen Genregulation zu verstehen, etwa während Zellen sich differenzieren.

References

Eser P, Demel C, Maier KC, Schwalb B, Pirkl N, Martin DE, Cramer P, Tresch A: Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Mol Syst Biol* **10**, 717 (2014).

Martinez-Rucobo FW, Kohler R, Van De Waterbeemd M, Heck AJ, Hemann M, Herzog F, Stark H, Cramer P: Molecular basis of transcription-coupled pre-mRNA capping. *Mol Cell* **58**, 1079-89 (2015).

Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L, Martin DE, Tresch A, Cramer P: Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**, 458 (2011).

Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA: RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* **21**, 175-179 (2014).

Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P: Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155**, 1075-1087 (2013).

Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, Cramer P: Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell* **52**, 52-62 (2013).

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Lariviere L, Maier KC, Seizl M, Tresch A, Cramer P: Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22**, 1350-1359 (2012).

Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R, Libri D: Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell* **55**, 467-481 (2014).

Cover picture / Titelbild

Monitoring global changes in RNA metabolism. Shown here are differences in the rates of RNA synthesis from yeast transcripts after depletion of the factor Nrd1, a nuclear RNA-binding protein. Transcripts above or below the dashed line are significantly up- or down-regulated. RNAs from different classes are shown as dots in different colors. These data establish Nrd1 as a surveillance factor that suppresses the erroneous synthesis of aberrant RNAs.

Globale Änderung der Gentranskription. Die Grafik zeigt Änderungen in der Syntheserate von RNAs, wenn der RNA-bindende Faktor Nrd1 aus Hefezellen depletiert wird. RNAs aus verschiedenen Klassen sind in unterschiedlichen Farben dargestellt. Diese Daten etablieren Nrd1 als einen Faktor, der die fehlerhafte Synthese von funktionslosen RNAs unterdrückt.