Dynamic transcriptome analysis (DTA): Kinetic modeling of synthesis and decay of mRNA transcripts upon perturbation in *S.cerevisiae*, *S.pombe* and *D.melanogaster*

Björn Schwalb

aus

Heilbronn, Deutschland

2012

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Patrick Cramer betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 17.05.2012

Björn Schwalb

Dissertation eingereicht am 21.05.2012

1. Gutachter: Prof. Dr. Patrick Cramer
2. Gutachter: Prof. Dr. Achim Tresch

Mündliche Prüfung am 12.07.2012

Acknowledgments

Abstract

So far, much attention has been paid to regulation of transcription. However, it has been realized that controlled mRNA decay is an equally important process. To understand the contributions of mRNA synthesis and mRNA degradation to gene regulation, we developed Dynamic Transcriptome Analysis (DTA). DTA allows to monitor these contributions for both processes and for all mRNAs in the cell without perturbation of the cellular system. DTA works by non-perturbing metabolic RNA labeling that supersedes conventional methods for mRNA turnover analysis. It is accomplished with dynamic kinetic modeling to derive the gene-specific synthesis and decay parameters.

DTA reveals that most mRNA synthesis rates result in several transcripts per cell and cell cycle, and most mRNA half-lives range around a median of 11 min. DTA can monitor the cellular response to osmotic stress with higher sensitivity and temporal resolution than standard transcriptomics. In contrast to monotonically increasing total mRNA levels, DTA reveals three phases of the stress response. In the initial shock phase, mRNA synthesis and decay rates decrease globally, resulting in mRNA storage. During the subsequent induction phase, both rates increase for a subset of genes, resulting in production and rapid removal of stress-responsive mRNAs. In the following recovery phase, decay rates are largely restored, whereas synthesis rates remain altered, apparently enabling growth at high salt concentration. Stress-induced changes in mRNA synthesis rates are predicted from gene occupancy with RNA polymerase II. Thus, DTA realistically monitors the dynamics in mRNA metabolism that underlie gene regulatory systems.

One of the technical obstacles of standard transcriptomics is the unknown normalization factor between samples, i.e. wild-type and mutant cells. Variations in RNA extraction efficiencies, amplification steps and scanner calibration introduce differences in the global intensity levels. The required normalization limits the precision of DTA. We have extended DTA to comparative DTA (cDTA), to eliminate this obstacle. cDTA provides absolute rates of mRNA synthesis and decay in *Saccharomyces cerevisiae* (*Sc*) cells with the use of *Schizosaccharomyces pombe* (*Sp*) as an internal standard. It therefore allows for direct comparison of RNA synthesis and decay rates between samples.

cDTA reveals that *Sc* and *Sp* transcripts that encode orthologous proteins have similar synthesis rates, whereas decay rates are five fold lower in *Sp*, resulting in similar mRNA concentrations despite the larger *Sp* cell volume. cDTA of *Sc* mutants reveals that a eukaryote can buffer mRNA levels. Impairing transcription with a point mutation in RNA polymerase (Pol) II causes decreased mRNA synthesis rates as expected, but also decreased decay rates. Impairing mRNA degradation by deleting deadenylase subunits of the Ccr4–Not complex causes decreased decay rates as expected, but also decreased synthesis rates.

In this thesis, we provide a novel tool to estimate RNA synthesis and decay rates: a quantitative dynamic model to describe mRNA metabolism in growing cells to complement the biochemical protocol of DTA/cDTA. It can be applied to reveal rate changes for all kinds of perturbations, e.g. in knock-out or point mutation strains, in responses to stress stimuli or in small molecule interfering assays like treatments with miRNA or siRNA inhibitors. In doing so, we show that DTA is a valuable tool for miRNA target validation. The DTA/cDTA approach is in principle applicable to virtually every organism. The bioinformatic workflow of DTA/cDTA is implemented in the open source R/Bioconductor package DTA.

Publications

Parts of this work have been published or are in the process of publication :

2011 **Drosophila miR-277 controls TOR activity and lifespan via tuning of branched-chain amino acid metabolism**
S. Esslinger, **B. Schwalb**, S. Helfer, H. Witte, K. Maier, D. Martin, B. Michalke, A. Tresch, P. Cramer, K. Förstemann
Manuscript in preparation.

2011 **Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA)**
B. Schwalb, D. Schulz, M. Sun, B. Zacher, S. Dümcke, D. Martin, P. Cramer, A. Tresch
Bioinformatics, Vol. 28, No. 6, pp. 884-885.
http://bioinformatics.oxfordjournals.org/content/28/6/884

Author contributions: B.S. and A.T. designed and carried out computational analyses. D.S. conducted the experiments. B.S., B.Z. and A.T. wrote the manuscript, with input from all authors.

2011 **Comparative Dynamic Transcriptome Analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation**
M. Sun*, **B. Schwalb***, D. Schulz*, N. Pirkl, S. Etzold, L. Larivière, K. Maier, M. Seizl, A. Tresch, P. Cramer (* joint first authorship)
Genome Research.
http://genome.cshlp.org/content/early/2012/03/30/gr.130161.111

Author contributions: M. Sun, B.S., L.L., and P.C. conceived and designed the study. M. Sun and D.S. implemented and evaluated experimental methods. S.E. und M. Seizl conducted the *rpb1-1* experiments. B.S. and A.T. designed and carried out computational analyses. M.S., D.S., N.P., K.M., S.E., and M. Seizl designed and performed experiments. P.C. and A.T. wrote the manuscript, with input from all authors. M.Sun, B.S., and D.S. contributed equally.

2010 **Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast**
C. Miller*, **B. Schwalb***, K. Maier*, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. Martin, A. Tresch, P. Cramer (* joint first authorship)
Molecular Systems Biology 7, 458.
http://www.nature.com/msb/journal/v7/n1/full/msb2010112.html

Author contributions: CM and KM conducted DTA experiments, BS, SD, BZ, and AT developed the statistical workflow and carried out computational analysis, DS and AM conducted ChIP experiments, JS conducted enzymatic assays, LM, LD, and DEM conducted initial experiments, LD and PC initiated the study, DEM supervised experimental work, and AT and PC supervised research and wrote the manuscript.

Contents

Part I
Introduction

The central dogma of molecular biology, first formulated by Francis Crick [46] (Figure 1), describes the transfer of genetic information along the three important classes of biopolymers: deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins. All three carry the important sequential information whose heredity and constant availability are crucial for development and function of all living organisms. The underlying biological processes, namely DNA replication, transcription into RNA and translation into proteins, are the essential bridges between those three building blocks. These processes are committed to transfer the genetic information in a faithful and deterministic manner to ensure the precise and accurate function of the resulting biopolymer.
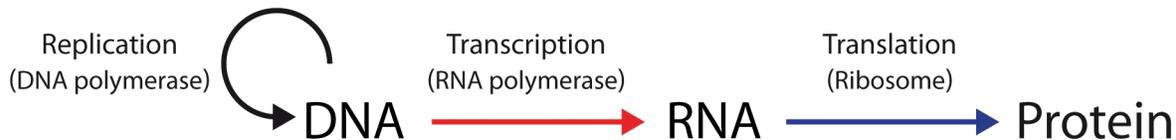
Figure 1: The central dogma of molecular biology describes the flow of genetic information along the three important classes of biopolymers: DNA, RNA and protein. Adapted from [46].

RNA, one of the three important macromolecules, is a highly versatile compound. It functions in the process of translation as rRNA and tRNA, contributes to the regulation of gene expression as miRNA, siRNA or other non-coding RNA, or acts as a pure messenger: the mRNA. The underlying mechanisms, with the eclectic and central roles they play in cell metabolism, are of high medical interest, as they are associated with several diseases and defects [42, 69, 58].

1   The mRNA life cycle

The mRNA is a transient message of genetic information subject to nuclear synthesis and cytoplasmic degradation. These two opposing mechanisms control mRNA levels and consequently cellular gene expression. The global and specific regulation of mRNA abundance can therefore either be accomplished by alterations of synthesis or decay rates. Despite the spatial separation of mRNA synthesis and degradation, there is evidence that these processes are coordinated [122, 123, 82]. mRNA turnover as well as its underlying mechanisms are a useful indicator for gene activity. High-throughput methods for quantification of protein abundance [30, 150, 10] are not as feasible as methods for mRNA quantification (Section 3.3). Nevertheless, changes in mRNA abundance were shown to be correlated to changes in protein abundance [171, 21], which play a major role in cell metabolism.

1.1   Synthesis of mRNA (The RNA polymerase II transcription cycle)

The most crucial enzyme in mRNA synthesis is the DNA-dependent RNA polymerase II (Pol II). Pol II is a multi-subunit complex composed of 12 subunits, termed Rpb1p-12p. It is assembled in the cytoplasm and then imported into the nucleus [210, 47], where it polymerizes mRNAs of all protein coding genes against the corresponding DNA template. Eukaryotic transcription by Pol II

is orchestrated in four major phases, namely initiation, elongation, termination and re-initiation (Figure 2). This circular arrangement and its underlying phases are highly regulated and require a variety of specific proteins, including general transcription factors (GTF), negative and positive cofactors, elongation factors, RNA processing factors and termination factors [79, 188]. Rpb1p, the largest subunit of Pol II contains the unique carboxyl-terminal domain (CTD). It consists of multiple, highly conserved heptapeptide tandem repeats (26 in yeast) [2, 43], which are phosphorylated and dephosphorylated during transcription [29]. The CTD functions as a general platform to recruit many regulatory factors involved in transcription, mRNA processing and histone modification [28]. Different CTD phosphorylation patterns were shown to be associated with different sets of regulatory factors required co-transcriptionally during different phases [156]. Control of Pol II activity is specifically regulated at individual genes and is absolutely crucial for homeostasis of the cell [66].
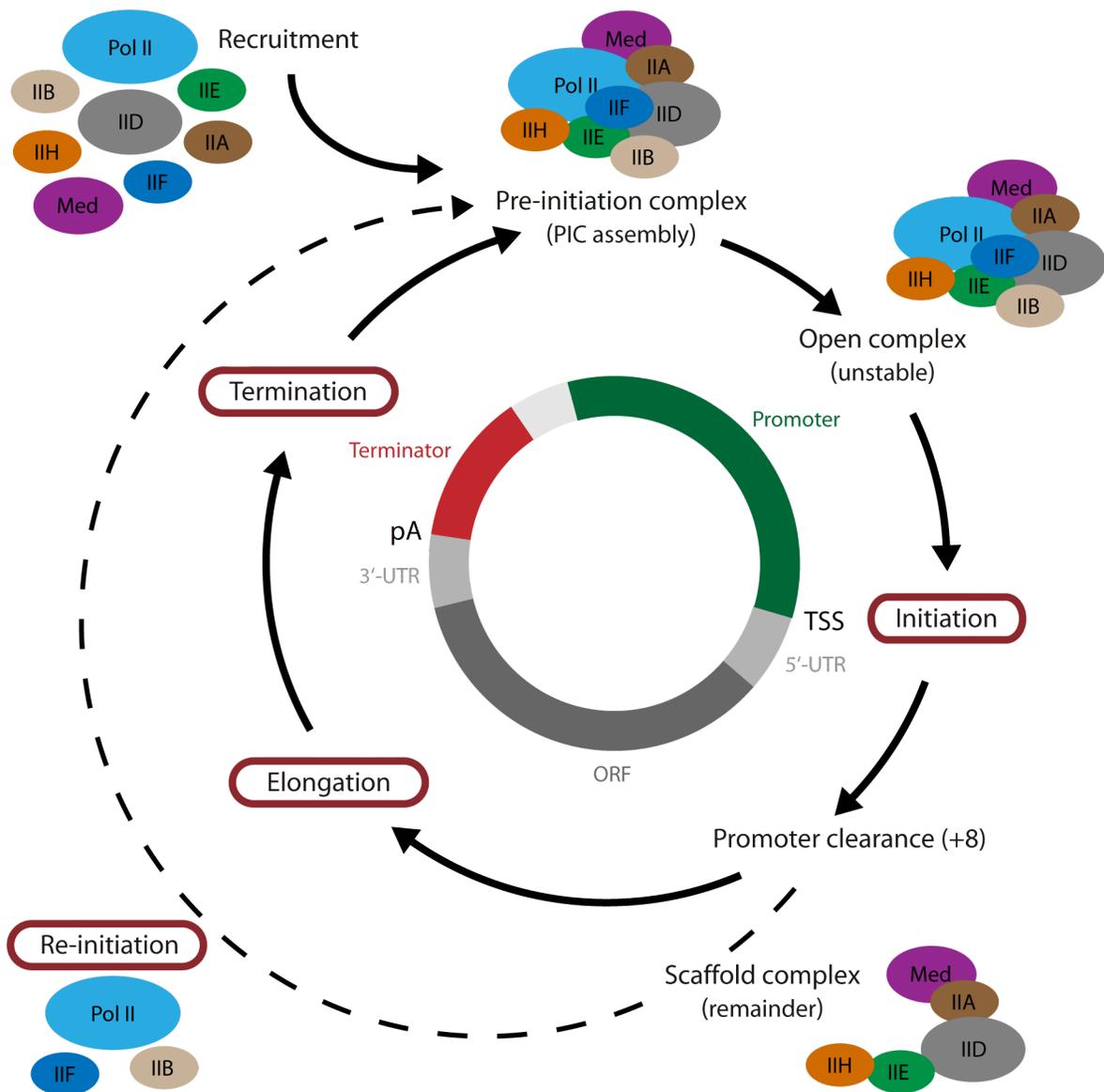
Figure 2: The RNA polymerase II transcription cycle. Eukaryotic transcription by Pol II is orchestrated in four major phases, initiation, elongation, termination and re-initiation. Adapted from [79, 188].

1.1.1 Initiation and the general transcription factors

Transcription initiation by Pol II requires the general transcription initiation factors TFIIA, -B, -D, -E, -F and -H. [179, 192, 203, 79]. Primarily, gene-specific factors (transcription activators) recognize unique DNA elements, like TATA, BRE, Inr and DPE, located in the core promoter. Upstream activation sequences (UASs) can increase the expression of a gene. UASs are promoter proximal regions and enhancer regions typically upstream of the transcription start site (TSS) of the gene. Subsequently, nucleosome remodellers, chromatin-modifying enzymes and coactivators, like SWI/SNF, INO80/SWR1, ISWI, NuA4, SAGA and Mediator, are recruited to remodel the chromatin architecture to ensure accessibility and promote the preinitiation complex (PIC) assembly at the core promoter [35, 147, 202]. The PIC consists of the RNA Pol II and the factors TFIIA, -B, -D, -E, -F and -H. These factors are required for transcription bubble formation, TSS scanning [106], and initial synthesis of the nascent RNA molecule up to the +8 position after the DNA is unwound [79]. Thus, the basic state of a promoter is inactive and transcription must be initiated by several TFs [186]. Gene-specific factors (transcription repressors) and chromatin-modifying enzymes can also cause inaccessibility of regulatory sequences [202]. Promoters in eukaryotes can be classified into two opposing categories regarding nucleosome occupancy, namely 'open' and 'covered', modulating constitutive and highly regulated genes, respectively [62, 193]. Thus, regulation of chromatin structure is substantially linked to the regulation of Pol II transcription initiation. All these important steps could, in principle, be rate limiting, adding an important layer that is prone to regulation [66].

1.1.2 Elongation and CTD phosphorylation

The transition from transcription initiation to early elongation happens at the +8 position when Pol II gets stable on both the DNA and the growing RNA chain and leaves the promoter region. This is known as promoter clearance or promoter escape and can also be rate limiting [66]. If not terminated by abortive initiation, Pol II enters into the productive elongation phase. Then, the capping enzymes are recruited and the nascent mRNA receives a 5'- cap when it reaches a chain length of 25-30 nucleotides [207]. The nascent RNA is then synthesized by a stable elongation complex resulting in a full-length RNA transcript [144, 128].

The phosphorylation level on the Ser2 residue of the Pol II CTD increases during the transcription cycle. Simultaneously, it decreases on the Ser5 and Ser7 residues [28, 129]. This is referred to as the CTD phosphorylation cycle which is associated with the transcription cycle. Different CTD phospho-isoforms (CTD code) determine at each stage of the transcription cycle which particular set of transcription, mRNA processing and chromatin-modifying factors have to be recruited [26, 55, 28, 131, 156, 159].

1.1.3 Termination and re-initiation

The final phase of the transcription cycle is reached when Pol II reaches the poly (A) site (pA site) of the transcribed gene. The elongation factors are then replaced by termination factors [164]. At the pA site the transcript is cleaved off, Pol II transcribes a little further downstream and is finally terminated [109, 104, 49]. During these steps, the mRNA receives a 3'-poly(A) tail. After termination, transcription can be initiated again either by recruitment of the complete transcription machinery to the promoter region or by facilitated re-initiation due to the promoter bound scaffold complex [216]. This remainder of the initial transcription machinery, including TFIIA, -D, -E, -H and Mediator, enables rapid PIC formation and thus efficient successive initiation. The underlying processes of termination are highly regulated and coupled to RNA processing events, which often occur far downstream of the poly (A) site [158, 27].

1.2 Decay of mRNA (The cytoplasmic mRNA decay machinery)

The highly complex and elaborate mechanisms of mRNA transcription are of limited use without equally precise control of mRNA degradation, which also plays a key role in gene regulation. After transcription by Pol II accompanied by co-transcriptional 5'-end capping and 3'-end polyadenylation, the mature mRNA is packaged and exported to the cytoplasm. In the cytoplasm mRNAs undergo a primary round of translation and progressive deadenylation. Surveillance pathways can identify and initiate exonucleolytic digestion of defective RNAs and ribonucleoprotein particles (RNPs): non-stop decay (NSD) and no-go decay (NGD) deal with stalled ribosomes, and nonsense mediated decay (NMD) handles mRNAs with premature translation termination codons (PTCs) [98, 130].
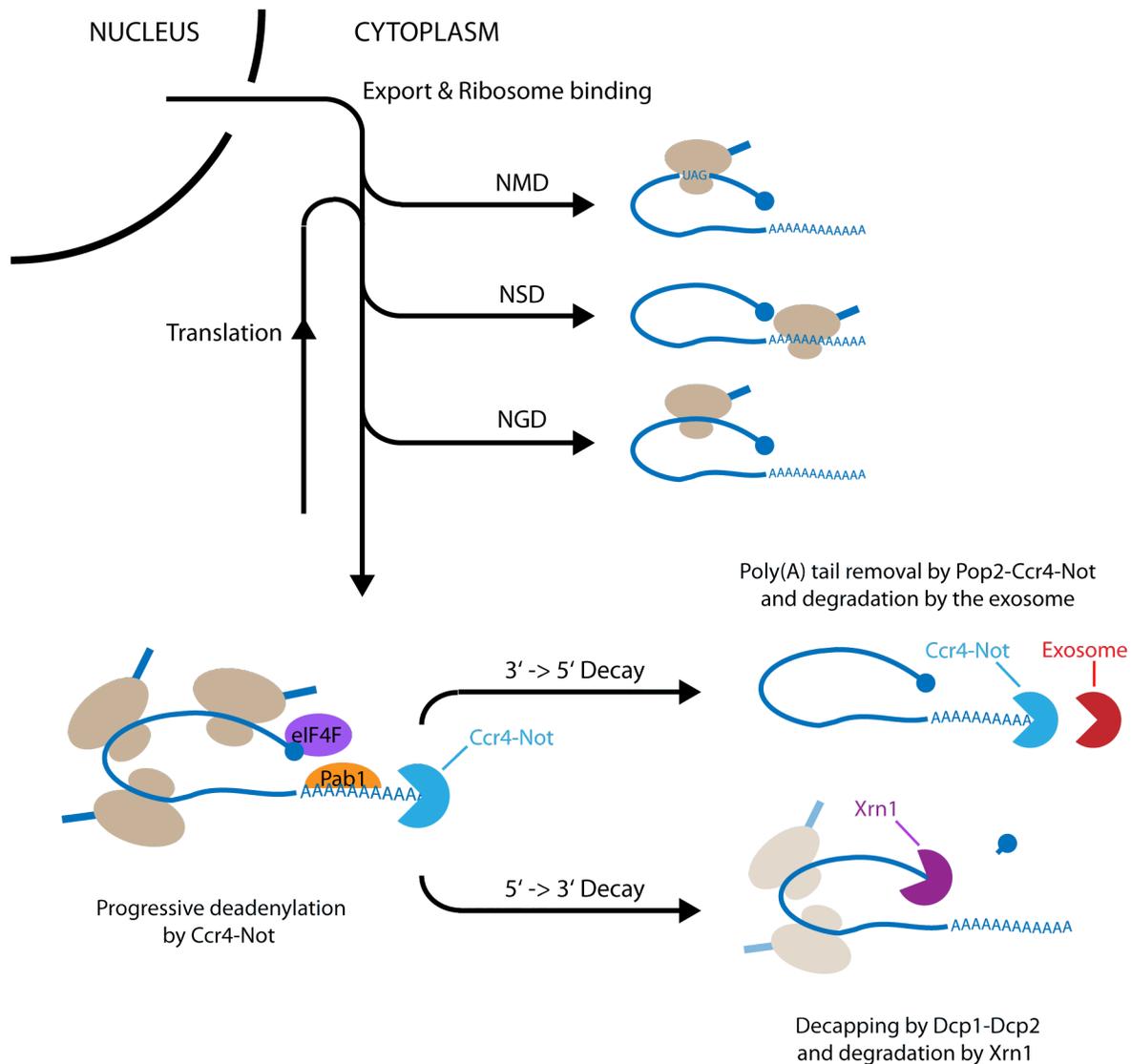
Figure 3: Cytoplasmic mRNA decay pathways. Surveillance pathways can identify and initiate degradation of defective RNAs and RNPs, including non-stop decay (NSD), no-go decay (NGD) and nonsense mediated decay (NMD). General cytoplasmic mRNA decay includes progressive deadenylation, decapping and degradation by exonucleases from both ends. Adapted from [91, 209].

General cytoplasmic mRNA degradation begins with shortening or removal of the poly(A) tail

[31, 153]. Then, the mRNA is decapped and degraded by exonucleases from both ends (Figure 3). Three major classes of RNA-degrading enzymes (ribonucleases or RNases) have been characterized: endonucleases that cleave RNA internally, 5'-exonucleases and 3'-exonucleases that hydrolyze RNA from the 5'- and 3'-end, respectively. Based on the large variety of RNases, which often possess overlapping activities, redundancy is a general feature of the RNA degradation system, presumably increasing the overall efficiency and robustness of degradation pathways.[91, 59, 152, 209].

1.2.1 Deadenylation and mRNA stability

The poly(A) tail of an mRNA is protected from exonucleases by poly(A)-binding proteins (PABPs/ Pab1p) which are recruited to the mRNA immediately after 3'-end polyadenylation during transcription termination and cover the entire poly(A) stretch [126, 68]. PABPs in turn can interact with the eukaryotic translation initiation factor 4F (eIF4F) that assembles on the 5'-cap of the mRNA, facilitates initiation of translation [101] and effectively causes circularization of the mRNA [209]. Progressive mRNA deadenylation occurs simultaneous to multiple rounds of translation. It is accomplished by the Ccr4-Not complex, the major deadenylase in eukaryotes [209, 91]. Ccr4-Not is composed of nine different subunits including Ccr4p and Pop2p (Caf1p), both with catalytic enzyme activity. It can be recruited by mRNA-binding proteins, like Puf (Pumilio/FBF) proteins bound to specific sequences in the 3-UTR, that control mRNA-specific rates of deadenylation [208, 75, 73, 194]. Cytoplasmic mRNA deadenylation can also be conducted by a conserved complex (PAN) consisting of the Pan2p and Pan3p proteins. This was demonstrated in the absence of the predominant deadenylase Ccr4, where residual deadenylation is dependent on Pan2p [195]. PAN is also involved in an initial phase of mRNA deadenylation, an early step in poly(A) metabolism that is required for poly(A) tail length control. The poly(A) tail is shortened to $55 - 75$ nucleotides depending on the individual mRNA species [24, 153]. Hence, eukaryotes own a highly sophisticated timing mechanism to influence the dynamics of gene expression [209, 91]. Enhanced mRNA stability to allow efficient translation versus destabilization due to shortening of the poly(A) tail inducing mRNA decay. In yeast, cell-cycle-regulated genes possess generally short poly(A) tails, suggesting that deadenylation also plays a crucial role in the regulation of their periodic expression [16, 149].

1.2.2 Decapping enzymes and degradation control

Early during elongation, eukaryotic mRNAs receive a protective 5'-cap structure. This 5'-cap must be removed to render the 5' end accessible for exonuclease degradation. Two types of decapping enzymes were identified in eukaryotic cells: The scavenger decapping enzyme (DcpS) decaps short 5'-cap oligonucleotides to release m7GMP [118]. These capped oligonucleotides are probably decay intermediates of 3' to 5' degradation by the cytoplasmic exosome, a large protein complex containing multiple 3' to 5' exonucleases [199, 138]. The second decapping enzyme consists of two subunits, Dcp1p and Dcp2p, with the latter as the catalytic subunit. Decapping by Dcp1p/Dcp2p follows deadenylation and simultaneously triggers 5' to 3' exonucleolytic digestion by Xrn1p [140, 92], a conserved 5' to 3' exonuclease. In summary, deadenylation is followed by one of two alternative degradation pathways: 3' to 5' degradation by the cytoplasmic exosome, or decapping and 5' to 3' degradation by Xrn1 [209, 91, 153]. This process occurs co-translationally, enabling the ribosome to complete translation before the mRNA is degraded [93]. Degradation is consequently controlled by the competition of translation and decapping due to deadenylation regulation [153].

1.2.3 P-bodies

Recent evidence indicates that deadenylated, or translationally repressed mRNAs, deadenylation factors, decapping factors, decapping activators and the 5' to 3' exonuclease Xrn1 colocalize within cytoplasmic domains referred to as processing bodies (P bodies) [63, 177, 59]. mRNA transcripts

associated with P bodies can either be subject to decay or return to translation. This suggests a post-transcriptional regulation mechanism able to store mRNAs. The rates of translation and degradation are thus influenced by the equilibrium between P bodies and translating polysomes. This might be dependent on individual mRNA species, although the underlying structure and organization remains still unclear [152].

1.3 miRNA induced silencing of mRNA targets

MicroRNA (miRNAs) are a class of small noncoding RNAs. Identified species of small RNAs, including also small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs), are distinguished via the differences in biogenesis, types of target regulation and RNA silencing pathways. RNA silencing pathways are highly conserved in all eukaryotes from *S.pombe* to plants and animals and are probably linked competitively and collaboratively due to gene regulation [74, 212]. miRNAs mostly attracted attention as negative post-transcriptional regulators of gene expression [4, 14, 197], e.g. thresholding in negative feed-back loops [37] and mediation of robustness against environmental fluctuations or extrinsic and/or intrinsic perturbations [117]. On the contrary, recent findings indicate, that miRNAs are able to up-regulate translation [200]. At least 30% of all protein coding genes in animals are predicted to be regulated by miRNAs [23, 107, 115, 214]. miRNAs, as well as other small RNAs, are characterized by their short length ($\sim 21 - 23$ nucleotides for miRNAs) and their association with the RNA-induced silencing complex (RISC). RISC complexes can be recruited sequence-specific to its corresponding target mRNAs, predominantly to repress their translation or stimulate their degradation. This is either accomplished by deadenylation and decapping resulting in exonucleolytic degradation from both ends (partial sequence complementarity) or directed site specific cleavage prior to digestion from the resulting accessable ends (perfect sequence complementarity) [74, 91, 212].

1.3.1 miRNA biogenesis

miRNAs originate from long, hairpin-containing precursor transcripts, referred to as primary miRNAs (pri-miRNAs), which are in general transcribed by RNA Pol II polycistronically [114, 113]. Biogenesis proceeds in the following sequential steps: Primarily, the pri-miRNAs are cleaved in the nucleus by the RNase III endonuclease Drosha, together with its double-stranded RNA-binding domain (dsRBD) partner Pasha [112, 50], into $\sim 60 - 70$ nucleotide long hairpin pre-miRNAs. Secondly, the Ran-GTP dependent nucleo/cytoplasm transporter Exportin-5 carries the pre-miRNAs to the cytoplasm. In the cytoplasm, the pre-miRNAs are cleaved by a second RNase III endonuclease, Dicer [95, 103], and its dsRBD partner Loqs, into $\sim 21 - 23$ nucleotide long double-stranded duplexes, containing both the mature miRNAs and their complementary strands [83, 12, 74].

1.3.2 Post-transcriptional repression by miRNAs

After processing, the double-stranded duplex is loaded into the RNA-induced silencing complex (RISC). RISC is a multi-protein complex including catalytic components of the Argonaut (Ago) protein family [51]. RISC incorporates the mature miRNA strand, while the complementary strand (often referred to as miRNA*) is expelled from the complex. The active RISC complex can subsequently target mRNAs with sequence complementarity to the seed region, a limited region at the 5' end of the miRNA, contributing the majority of the binding energy [5, 81]. miRNAs bind in general to the 3'-untranslated regions (3'-UTRs) of their corresponding target mRNAs. Consequently, single-stranded miRNAs function as a guide for Ago proteins to localize their mRNA targets, repress translation and initiate degradation [74, 83, 23, 116, 115, 110]. The miRNA loaded RISC is recruited to target mRNAs in two of the following ways: complete sequence complementarity of the binding regions leads to site-specific endonucleolytic cleavage by Ago. The resulting RNA fragments

are subsequently degraded by Xrn1 and the exosome from the accessable 5' and 3' ends. In most cases, however, miRNAs target mRNAs with only partial sequence complementarity. The short seed region can in turn lead to a large number of different mRNA targets regulated by a single miRNA [173, 9]. Translation is repressed and the mRNA is deadenylated and decapped by Ccr4-NOT [61] and Dcp1/Dcp2, respectively. The remaining mRNA is subsequently degraded by Xrn1 and the exosome from the 5' and 3' ends [60, 212]. The mechanism of mRNA target regulation by miRNAs consequently depends on the extent of complementarity to the binding regions [91, 74, 83].
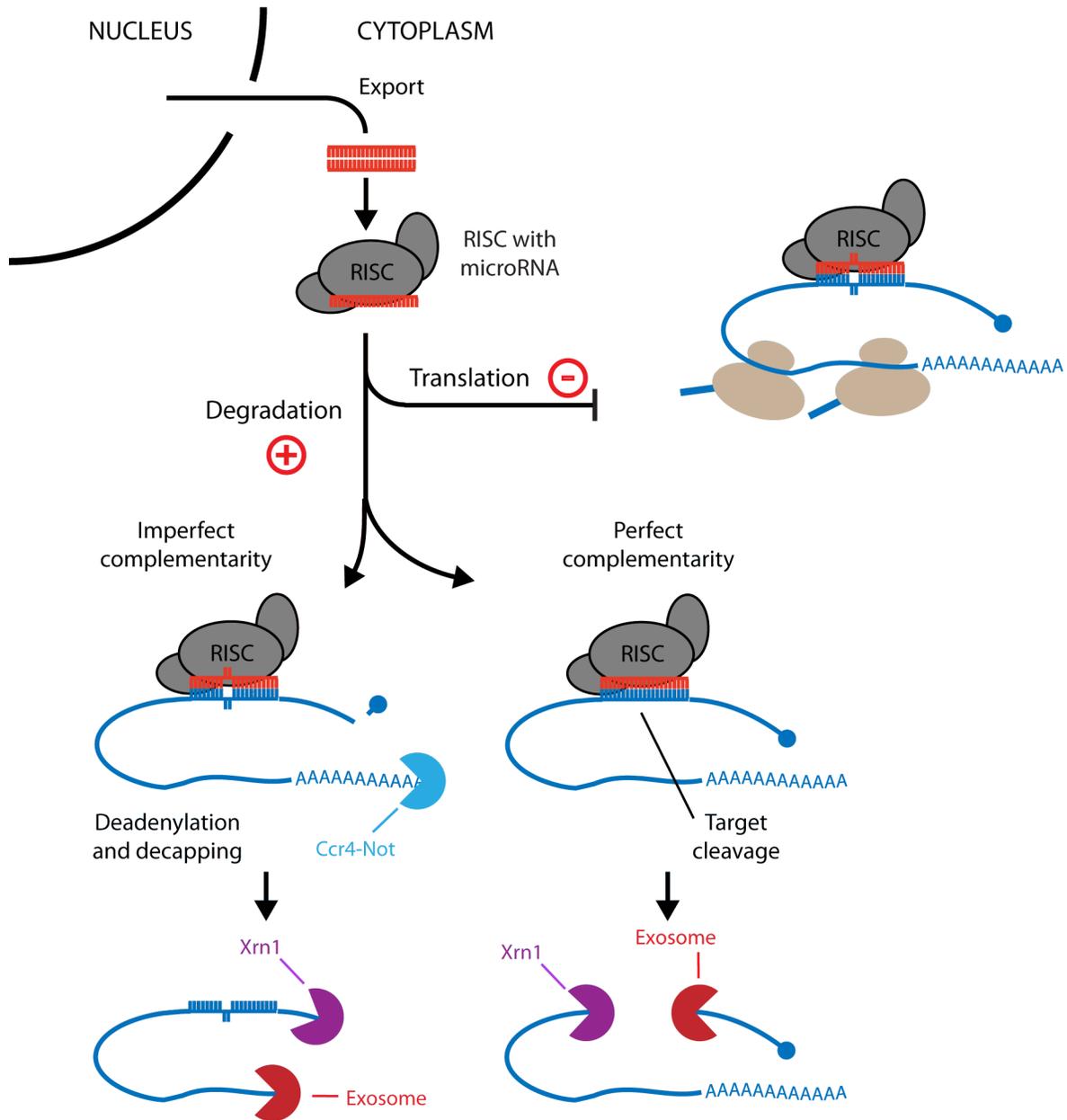
Figure 4: miRNA induced post-transcriptional silencing of mRNA targets (blue). The RNA-induced silencing complex (RISC) can be recruited to a corresponding targeted mRNA to either repress its translation or stimulate its degradation. Adapted from [57, 91]

2 Exponential decay and half-life of mRNA populations

It is a reasonable assumption to model cytoplasmic mRNA decay as a first order reaction, i.e. the rate of the reaction depends on the concentration of a reactant. This is presumably not entirely true as the enzyme-catalyzed reactions of degradation are thought to be tightly regulated with the adaption of gene expression programs to certain stimuli, but it remains to be a rational choice for a mixture of cells [111, 90, 205, 77, 176, 171, 52].
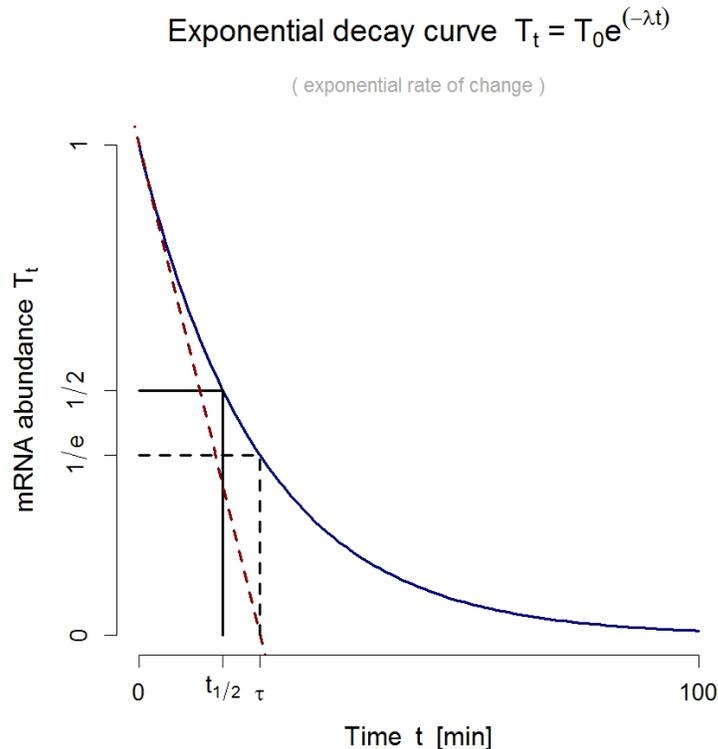
Figure 5: Exponential decay and half-life of mRNA populations. Exemplary illustration showing the relation of the total mRNA amount T_t at timepoint t, the decay rate λ, the half-life $t_{1/2}$ and the mean life time τ.

A quantity T decays exponentially if it declines at a rate proportional to its value [145]. This can be stated as in the following differential equation. Units of concentration per units of time

$$\frac{dT}{dt} = -\lambda T \tag{1}$$

with time t and decay rate λ. The decay rate hence gives the proportion of the number of total mRNA transcripts that are decayed in the first unit of time. The solution to this equation yields

$$T_t = T_0 e^{-\lambda t} \tag{2}$$

with the initial value $T_0 = T(0)$ (Figure 5). As a consequence, Equation (2) defines a relation between the initial value T_0, the current value T_t at time t and the decay rate λ. Equation (2) can be normalized to give a probability density. By this means, we obtain the exponential distribution function, which is characterized by $\frac{1}{\lambda}$ as its expected value [19]. This subsequently enables us to derive the average time an mRNA remains stable via

$$\tau = \frac{1}{\lambda}. \tag{3}$$

τ is called the mean life time or exponential time constant. It is often referred to as scaling time, which can be seen by rearranging Equation (2) with respect to (3):

$$T_t = T_0 e^{-t/\tau} \tag{4}$$

Consequently, τ is the time at which the initial amount of mRNAs is reduced to $e^{-1} = 1/e$ times its initial value. The most common characteristic of exponential decay is the half-life $t_{1/2}$ defined as

$$t_{1/2} = \frac{\log 2}{\lambda} = \tau \log 2 \,. \tag{5}$$

Given this equation, we can restate Equation (2) to yield an alternative formula

$$T_t = T_0 2^{-t/t_{1/2}} \tag{6}$$

This intuitive equation allows the interpretation, that after a period of $n \cdot t_{1/2}$ minutes the amount of mRNAs is reduced to $1/2^n$ times its initial value.

3   Conventional methods for mRNA turnover analysis

mRNA decay rates have previously been determined using transcriptional arrest [90, 205, 77, 176, 18, 78, 162, 174, 215, 7, 143] and then monitoring genome-wide mRNA degradation over time. The intensity of each mRNA species relative to that observed in a wild-type cell can be used to derive a measure for mRNA decay, applying the usual first-order exponential decay model (Section 2). The assumption, however, that mRNA decay is unaffected after inhibition of the transcription machinery, is invalid. Transcriptional arrest can be induced with a temperature-sensitive mutation in RNA polymerase II (Section 3.1) or via treatment with chemical inhibitors (Section 3.2). This requires a perturbing heat shock, i.e. cells have to be shifted to the nonpermissive temperature of 37°C, or is inherently cell invasive and also shows induced heat shock response genes [1, 33, 70, 178]. Both methods induce a general stress response of the cell which leads to a strong stabilization of all transcripts probably due to mRNA decay compensation of the decreased mRNA synthesis (Sections 15.5, 16). Furthermore, it can influence the regulatory mechanisms of mRNA degradation, which lead to different decay profiles, exhibiting differential internal degrees of stabilization. (Section 16.3). Additionally, both approaches are not capable of measuring newly transcribed mRNA as an estimator for transcriptional activity, i.e. synthesis rates.

3.1   Blocking transcription: the *rpb1-1* mutant strain

The *rpb1-1* mutant strain was first isolated by Nonet *et al.* [148] as a yeast conditional mutant. It carries a temperature-sensitive mutation in the largest subunit Rpb1p of RNA polymerase II. It was shown that this functionally defective RNA polymerase II mutant rapidly ceases mRNA synthesis when shifted to the nonpermissive temperature of 37°C [148]. As a consequence, the *rpb1-1* mutant strain can be used to measure mRNA decay rates via monitoring the remaining mRNA populations after the temperature shift. The intensity of each mRNA species relative to the initial amount is then used to derive a rate of degradation, assuming first-order exponential decay (Section 2).

3.2   Chemical inhibition of transcription

There are several chemicals that allow transcriptional arrest: e.g. 1,10-phenanthroline, 6-azauracil, thiolutin, ethidium bromide, cordycepin and actinomycin-D. These treatments indeed, are applicable to all kinds of organisms [77, 18, 78, 162, 174, 215, 7, 143] in comparison to the *rpb1-1* mutant strain (Section 3.1). The calculation of the decay rate, however, is analogous (Sections 3.1, 2).

3.3 Gene-expression profiling

The analysis of mRNA transcript abundance has been preferably accomplished with high resolution profiling on microarrays over the last decades [105]. These methods allow genome-wide molecular read-outs to analyze gene expression. Genes are represented by short oligonucleotide sequences (probes, 25-mers) identical to parts of the corresponding DNA sequence (Affymetrix® GeneChip® Yeast Genome 2.0 Array, [120, 211]). Several probes distributed along the DNA sequence of a certain gene are combined to a probeset. mRNA samples are amplified via reverse transcription into cDNA, fluorescently labeled during transcription into cRNA and subsequently hybridized to the corresponding probes that are immobilized on a solid surface on the microarray. Afterwards, the fluorescence of the bound DNA:cRNA hybrid is detected in a scanning procedure to yield a numerical read-out that is assumed to be proportional to the initial mRNA concentration in the sample [84, 85, 86, 180] (Section 16.1, Figure 44). In order to quantify non-specific hybridization, each probe (perfect match probe) is complemented with a mismatch probe with a single nucleotide mismatch at position 13. Mismatch probes can be used for background correction of the perfect match fluorescence signal, particularly its numerical value [105].

4 Dynamic Transcriptome Analysis (DTA)

mRNA levels in a cell are the consequence of two opposing mechanisms, namely nuclear mRNA synthesis and cytoplasmic mRNA degradation. These individual contributions keep mRNA levels in a dynamic equilibrium and can be monitored by Dynamic Transcriptome Analysis (DTA) [52, 64, 6, 136, 187, 168]. A technique that allows us to understand the contributions of both processes to gene regulation in a non-disruptive way.
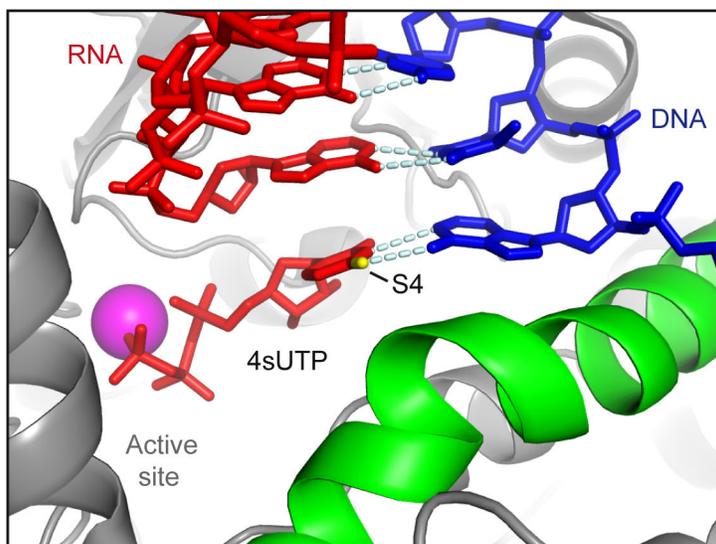
Figure 6: 4sUTP is modeled into the crystal structure of a Pol II transcribing complex (PDB code: 1I6H). The thiol group at position 4 can form a hydrogen bond with the DNA template strand (blue). Nascent mRNA is in red. Taken from [136].

4.1 Non-perturbing RNA labeling (in yeast)

The nucleoside analog 4-thiouridine (4sU) is taken up by eukaryotic cells and incorporated into mRNA during Pol II transcription [133] (Figure 6). The thiol-labeled newly transcribed RNA can

then be isolated by biotinylation and purification with streptavidin-coated magnetic beads [36, 52]. Although this approach is generally applicable to mammalian, insect, and plant cells, RNA labeling is not directly applicable to yeast. The nucleoside analog 4sU is readily taken up by cells of a broad range of eukaryotic organisms and is efficiently incorporated into their newly transcribed RNA. This can be used to metabolically label and isolate newly transcribed RNA from total cellular RNA with high specificity [102, 52]. In the fission yeast *S.pombe*, expression of the human equilibriative nucleoside transporter (hENT1) enables cellular uptake of the nucleoside analog 5-bromo-2'-deoxyuridine, resulting in labeling of DNA during replication [88]. This transporter can also mediate efficient uptake of 4sU in the budding yeast *S.cerevisiae*, and thus allow efficient RNA labeling at a level similar to that generally achieved in mammalian cells. This subsequently facilitates efficient separation of total cellular RNA into newly transcribed and pre-existing RNA. It is however also possible to use 4-thiouracil (4tU) instead of 4sU for RNA labeling in *S.cerevisiae* [141] and *S. pombe*, because it is taken up by the cells without expression of a nucleoside transporter [136]. 4tU labeling does also not affect normal cell physiology [136] and even allows growth of yeast in complete medium (YPD) instead of selective medium (SD), which is needed for selection due to the nucleoside transporter (hENT1).

4.2   DTA

The experimental setup for DTA requires culturing cells in the presence of a labeling substrate (e.g. 4-thiouridine (4sU) or 4-thiouracil (4tU)) for a certain amount of time (Section 4.1). During a short, non-perturbing RNA labeling pulse, cells incorporate 4sUTP into newly transcribed RNA instead of UTP. The cells are subsequently lysed and total cellular RNA is extracted. The newly transcribed labeled RNA is biotinylated and purified with streptavidin-coated magnetic beads [36, 52]. This setup yields three types of RNA fractions: total cellular RNA, pre-existing unlabeled RNA (flow-through) and newly transcribed labeled RNA (Figure 7). The quantification of these fractions through gene expression profiling on microarrays (Section 3.3) or RNAseq is used to estimate mRNA synthesis and decay rates with our novel statistical approach on a genome-wide scale, assuming exponential decay (Sections 2, II, 15). Monitoring mRNA degradation without transcriptional arrest and additionally measuring mRNA synthesis simultaneously in a single experimental setting is the outstanding advantage of DTA.
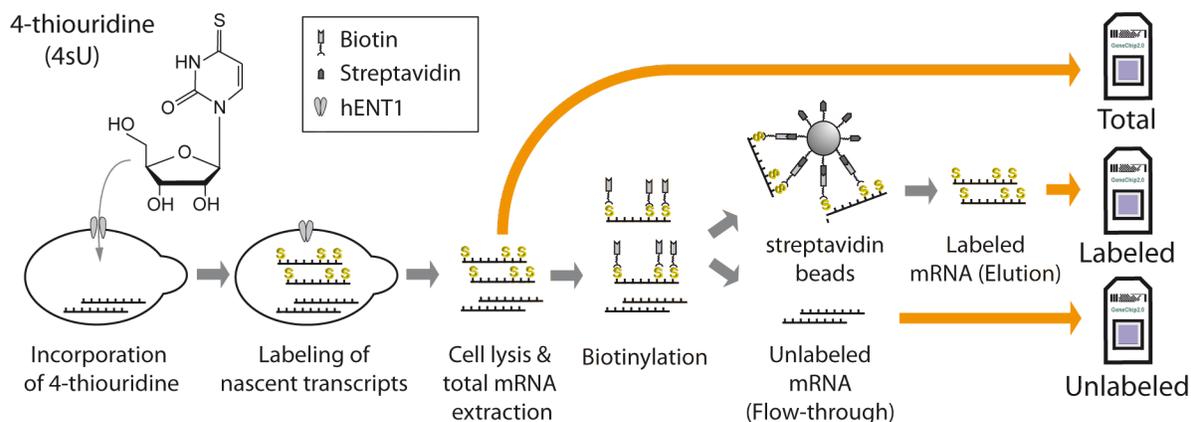
Figure 7: Scheme of metabolic mRNA labeling in yeast. Nascent mRNA is labeled with 4-thiouridine (4sU) and thiol-specifically biotinylated following cell lysis and preparation of total cellular mRNA. Fractionation into pre-existing and nascent mRNA is achieved with streptavidin beads. Adapted from [136].

4.3 cDTA (comparative DTA)

One of the major drawbacks in conventional transcriptomics is the unknown normalization factor between samples, i.e. wild-type and mutant cells. Variations in lysis and RNA extraction efficiencies, amplification steps in the protocol and the mandatory scanner calibration introduce differences in the global intensity levels of the measurements (Section 18.4). The estimation of the normalization factors limits the precision of DTA. We have extended DTA to comparative DTA (cDTA) to eliminate this obstacle (Section 16). This is accomplished by mixing an internal normalization standard of the distantly related fission yeast *S. pombe* (*Sp*) cells to samples of wild-type or perturbed *S. cerevisiae* (*Sc*) cells in the defined ratio of 3:1 (*Sc*:*Sp*) after cell counting (Figure 8). The resulting cell mixture is subsequently lysed, total mRNA extracted, labeled RNA purified, and the RNA mixture is quantified on a microarray that contains probes for both *Sc* and *Sp* transcripts (Affymetrix® GeneChip® Yeast Genome 2.0 Array, Section 3.3). A large culture of *Sp* cells was labeled to generate a stock and eliminate errors due to variations in the standard. cDTA provides absolute rates of mRNA synthesis and decay in *S. cerevisiae* (*Sc*) cells with the use of *S. pombe* (*Sp*) as an internal standard. It therefore allows for direct comparison of RNA synthesis and decay rates between samples.
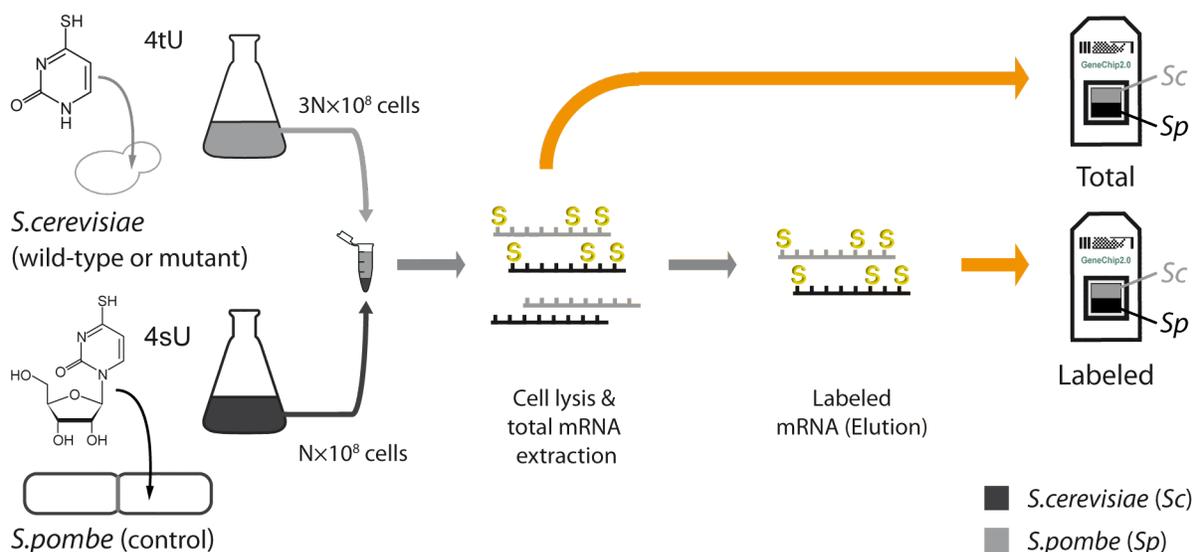
Figure 8: The *Sc* cells are labeled by adding 4tU into the media whereas *Sp* cells are labeled by adding 4sU. The cells are then counted. *Sc* cells from different experiments are mixed with always the same amount of labeled *Sp* cells from a single batch. Cells are then lysed, mRNA is extracted, biotinylated, and labeled mRNA separated. Microarrays containing probes against both *Sc* and *Sp* transcripts are then used to quantify both total and labeled mRNA. Adapted from [187].

5 Aims and scope of this thesis

The regulation of transcription has been intensively studied for a considerable time. On the other hand, mRNA decay has only recently been noticed to be an equally important process of gene regulation. To understand the contributions of mRNA transcription and stability to gene regulatory processes, methods must be developed to measure the rates for both processes and for all mRNAs in the cell. Furthermore, the rates of mRNA synthesis and decay must be measured without perturbation of the cellular system, and changes in these rates must be monitored to follow a cellular response. This cannot be achieved by standard transcriptomics, which only measures mRNA abundance. Synthesis rates can be measured by genomic run-on (GRO) [67, 155], but this requires

sarkosyl treatment that inhibits cellular processes. Decay rates can be measured after blocking transcription with inhibitors [111, 77, 18, 78, 162, 174, 215, 7, 143] but this is inherently cell invasive (Section 3.2). In addition, decay rates can be measured with a temperature-sensitive yeast strain [90, 205, 77, 176], but this requires a perturbing heat shock (Section 3). Unperturbed RNA synthesis and decay rates can be obtained via metabolic RNA labeling and kinetic modeling [36, 102, 52, 64] (Section 4). An additional limitation of standard transcriptomics is the unknown normalization factor between samples, i.e. wild-type and mutant cells. Variations in RNA extraction efficiencies, amplification steps and scanner calibration introduce differences in the global intensity levels.

The aim of this thesis was the development of a quantitative, dynamic model that describes mRNA metabolism in growing cells to supplement the biochemical part of the DTA approach (Section 4), which is based on genetically facilitated uptake of 4sU resp. 4tU, metabolic RNA labeling and microarray measurements which eliminate the obstacle of incomparability between different samples (Section 4.3) and allow the estimation of synthesis and decay rates simultaneously in a single experimental setting. The whole DTA/cDTA approach can monitor changes in synthesis and decay rates at unprecedented sensitivity and temporal resolution. It is a highly valuable tool for the analysis of dynamic changes in mRNA metabolism and a method that can provide quantitative data for modeling complex gene-regulatory systems, which is needed to provide new insights and to uncover misleading drawbacks of traditional methods. Furthermore, DTA/cDTA can be applied to reveal rate changes for all kinds of perturbations, e.g. knock-out or point mutation strains, responses to stress stimuli or small molecule interfering assays like treatments with miRNA or siRNA inhibitors. In doing so, we show that DTA is a valuable tool for miRNA target validation. The DTA/cDTA approach is in principle applicable to virtually every organism. The bioinformatic workflow of DTA/cDTA is implemented in the open source R/Bioconductor package *DTA* [170] as a part of this thesis.

6   Software

6.1   Statistical computing and graphics

The R programming environment [160] was used to carry out data handling, parameter estimation, statistical calculation and graphical display of this thesis. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. It runs on a wide variety of UNIX platforms, Windows and MacOS.

6.2   Implementation and availability

A major part of the workflow of this thesis is implemented in the open source R/Bioconductor package *DTA* [170]. The *DTA* package delivers straightforward methods to estimate mRNA synthesis and decay rates from pre-processed microarray or RNA-Seq measurements that are obtained via the DTA/cDTA protocol. It is capable of a thorough bias correction, detailed visualisation of quality control aspects and proper handling of comparative or time-series DTA data. The *DTA* package also allows to simulate all kinds of gene regulatory events that underlie the control of synthesis or decay of a gene or geneset. The implemented simulation methods are also quite useful to explore the limits of the model that builds the basis of the interplay between synthesis and decay. Additionally it can be used to assess minimum data quality requirements and is equipped with a proper error model able to give confidence regions of the estimated rates. The *DTA* package satisfies the high standard of the Bioconductor framework, regarding documentation and usability. It can therefore be easily incorporated in R workflows for pre-processing. Further statistical analysis of the results can readily be carried out by other methods within the R/Bioconductor programming environment. *DTA* is implemented in R version 2.15 [160] and part of Bioconductor version 2.10 [72]. To download and

install *DTA* and all its dependencies refer to

http://www.bioconductor.org/packages/release/bioc/html/DTA.html.

DTA builds upon the R package *LSD* [169]. *LSD* is a high-level plotting package, which provides the ability to depict all kinds of data in a suitable fashion and offers a plethora of variations to unveil its underlying structure. To download and install *LSD* and all its dependencies refer to

http://cran.r-project.org/web/packages/LSD/index.html.

6.3   Usage and application

For a more detailed explanation and example R code of the *DTA* package [170], refer to [168] and the package vignette

http://www.bioconductor.org/packages/release/bioc/vignettes/DTA/inst/doc/DTA.pdf.

For plotting examples of the *LSD* package [169], we refer to the `LSD.demo.tour()` and the reference manual

http://cran.r-project.org/web/packages/LSD/LSD.pdf.

Part II
Methods

7 A model for mRNA synthesis and degradation

In the following, we propose a model that complements the experimental DTA approach (Section 4). It accounts for exponential cell growth and for variations in RNA extraction efficiencies. It also corrects for differences in the fraction of newly synthesized mRNAs that escape labeling:

Let $r \in R$ be a sample. At time $t = 0$, we start mRNA labeling. During the labeling period, 4sUTP is integrated into newly synthesized mRNA (Section 4.1). At the timepoint t_r, when the mRNA is extracted, the total mRNA amount $C_{gr}(t_r)$ of gene g in the sample r is composed of the amount $B_{gr}(t_r)$ of (pre-existing) mRNA that has been synthesized *before* $t = 0$ and the amount $A_{gr}(t_r)$ of mRNA that has been newly synthesized *after* $t = 0$,

$$\underbrace{C_{gr}(t_r)}_{\text{total RNA}} = \underbrace{A_{gr}(t_r)}_{\text{newly synthesized mRNA}} + \underbrace{B_{gr}(t_r)}_{\text{pre-existing mRNA}} \tag{7}$$

Let $N_r(t_r)$ denote the number of cells in the sample r at time t_r. The cells are grown and harvested during mid-log phase, i.e. the cell number follows an exponential law with growth rate,

$$\alpha = \frac{log(2)}{CCL} \geq 0 \tag{8}$$

with the cell cycle length CCL. This means

$$\frac{dN_r(t)}{dt} = \alpha N_r(t) \tag{9}$$

and therefore

$$N_r(t_r) = N_r(0)e^{\alpha t_r}. \tag{10}$$

α is often refered to as the dilution rate, i.e. the reduction of concentration due to the increase of cell volume during growth. Say we have a cellular expression level $m_g = m_g(t)$ (transcripts of gene g per cell), then

$$C_{gr}(t) = m_g(t)N_r(t) \tag{11}$$

and hence

$$\begin{aligned} \frac{dC_{gr}(t)}{dt} &= m_g(t)\frac{dN_r(t)}{dt} + \frac{dm_g(t)}{dt}N_r(t) \\ &= m_g(t)\alpha N_r(t) + \frac{dm_g(t)}{dt}N_r(t) \\ &= \left(m_g(t)\alpha + \frac{dm_g(t)}{dt}\right)N_r(t). \end{aligned} \tag{12}$$

We assume that the mRNA population of a gene g decays exponentially at a (relative) rate given by $\lambda_g = \lambda_g(t)$ (Section 2). This means that the amount of $\lambda_g(t)C_{gr}(t)$ molecules is degraded during the interval of time dt. We further assume that the mRNA population of a gene g is synthesized at an absolute rate per cell given by $\mu_g = \mu_g(t)$. Hence the amount of $\mu_g(t)N_r(t)$ molecules is synthesized

during dt. In summary, the total mRNA amount follows the following differential equation:

$$
\begin{aligned}
\frac{dC_{gr}(t)}{dt} &= \mu_g(t)N_r(t) - \lambda_g(t)C_{gr}(t) \\
&= \mu_g(t)N_r(t) - \lambda_g(t)m_g(t)N_r(t) \\
&= \left(\mu_g(t) - \lambda_g(t)m_g(t)\right)N_r(t).
\end{aligned}
\tag{13}
$$

Using Equations (12) and (13), we get

$$
\left(\mu_g(t) - \lambda_g(t)m_g(t)\right)N_r(t) = \left(m_g(t)\alpha + \frac{dm_g(t)}{dt}\right)N_r(t)
\tag{14}
$$

and hence

$$
\mu_g(t) = m_g(t)\left(\alpha + \lambda_g(t)\right) + \frac{dm_g(t)}{dt}.
\tag{15}
$$

Additionally, we assume that the pre-existing mRNA population of a gene g decays exponentially at the same (relative) rate $\lambda_g = \lambda_g(t)$. This means for the pre-existing mRNA fraction that

$$
\frac{dB_{gr}(t)}{dt} = -\lambda_g(t)B_{gr}(t).
\tag{16}
$$

Consequently, the newly synthesized mRNA fraction can be given by

$$
\frac{dA_{gr}(t)}{dt} = \frac{dC_{gr}(t)}{dt} - \frac{dB_{gr}(t)}{dt} = \mu_g(t)N_r(t) - \lambda_g(t)A_{gr}(t).
\tag{17}
$$

with respect to Equation (7).

8 A model for constant synthesis and decay rates

The differential equations proposed in Section 7 can be solved without loss of generality (see Section 12). The DTA derived measurements (Sections 15, 16), however, suggest to simplify our model in order to increase the feasibility of rate extraction. To that end, we assume genes to have a (time averaged) constant synthesis rate μ_g and decay rate λ_g during 4sU/4tU labeling, i.e. $\mu_g(t) = \mu_g$ and $\lambda_g(t) = \lambda_g$ are given as constants. Nevertheless, it allows for changes in synthesis and decay rates in distinct labeling time windows, e.g. time course measurements. For the newly synthesized mRNA, this means

$$
\frac{dA_{gr}(t_r)}{dt} = \mu_g N_r(t_r) - \lambda_g A_{gr}(t_r) = \mu_g N_r(0)e^{\alpha t_r} - \lambda_g A_{gr}(t_r).
\tag{18}
$$

The solution of this differential equation yields

$$
A_{gr}(t_r) = ce^{-\lambda_g t_r} + \frac{\mu_g N_r(0)e^{\alpha t_r}}{\alpha + \lambda_g}
\tag{19}
$$

with an initial value $A_{gr}(0) = 0$, and so

$$
0 = c + \frac{\mu_g N_r(0)}{\alpha + \lambda_g}.
\tag{20}
$$

This finally leads to

$$
A_{gr}(t_r) = \frac{\mu_g N_r(0)}{\alpha + \lambda_g}\left[e^{\alpha t_r} - e^{-\lambda_g t_r}\right]
\tag{21}
$$

and therefore

$$
\mu_g = \frac{A_{gr}(t_r)(\alpha + \lambda_g)}{N_r(0)\left[e^{\alpha t_r} - e^{-\lambda_g t_r}\right]}.
\tag{22}
$$

For the total mRNA, we get

$$\frac{dC_{gr}(t_r)}{dt} = \mu_g N_r(t_r) - \lambda_g C_{gr}(t_r) = \mu_g N_r(0)e^{\alpha t_r} - \lambda_g C_{gr}(t_r). \tag{23}$$

The solution of this differential equation yields

$$C_{gr}(t_r) = ce^{-\lambda_g t_r} + \frac{\mu_g N_r(0)e^{\alpha t_r}}{\alpha + \lambda_g} \tag{24}$$

with

$$c = C_{gr}(0) - \frac{\mu_g N_r(0)}{\alpha + \lambda_g}. \tag{25}$$

This in turn leads to

$$C_{gr}(t_r) = C_{gr}(0)e^{-\lambda_g t_r} + \frac{\mu_g N_r(0)}{\alpha + \lambda_g}\left[e^{\alpha t_r} - e^{-\lambda_g t_r}\right] \tag{26}$$

and finally

$$B_{gr}(t_r) = C_{gr}(t_r) - A_{gr}(t_r) = B_{gr}(0)e^{-\lambda_g t_r} \tag{27}$$

with $B_{gr}(0) = C_{gr}(0)$. From (22) and (26) we can deduce

$$C_{gr}(t_r) = C_{gr}(0)e^{-\lambda_g t_r} + A_{gr}(t_r) \tag{28}$$

and rearrange it as follows

$$C_{gr}(t_r) - A_{gr}(t_r) = C_{gr}(0)e^{-\lambda_g t_r}. \tag{29}$$

Consequently

$$e^{-\lambda_g t_r} = \frac{C_{gr}(t_r) - A_{gr}(t_r)}{C_{gr}(0)} \tag{30}$$

and hence

$$\lambda_g = -\frac{1}{t_r}log\left[\frac{C_{gr}(t_r) - A_{gr}(t_r)}{C_{gr}(0)}\right] \tag{31}$$

or likewise

$$\lambda_g = -\frac{1}{t_r}log\left[\frac{B_{gr}(t_r)}{C_{gr}(0)}\right]. \tag{32}$$

As λ_g is assumed to remain constant in the above derivation, its value will be an average value along the labeling period.

9 mRNA turnover in steady state

In the following, we assume that the cells exhibit constant growth under constant environmental conditions. In particular, this implies that the amount of each mRNA population is constantly diluted by cell growth over time, being the result of a dynamic equilibrium of a constant mRNA synthesis and decay, i.e. steady state. In this case we assume genes to have a (time averaged) constant cellular expression level m_g during 4sU/4tU labeling. Thus, Equation (11) simplifies to

$$C_{gr}(t_r) = m_g N(t_r). \tag{33}$$

This leads to

$$C_{gr}(t_r) = m_g(0)N_r(0)e^{\alpha t_r} \tag{34}$$

as $m_g(t) = const. = m_g(0) = m_g$ and hence

$$C_{gr}(t_r) = C_{gr}(0)e^{\alpha t_r}. \tag{35}$$

Equation (31) can therefore be written as

$$\lambda_g = -\frac{1}{t}log\left[\frac{C_{gr}(t_r) - A_{gr}(t_r)}{C_{gr}(t_r)e^{-\alpha t}}\right]. \tag{36}$$

As a consequence

$$\lambda_g = -\alpha - \frac{1}{t}log\left[1 - \frac{A_{gr}(t_r)}{C_{gr}(t_r)}\right] \tag{37}$$

or

$$\lambda_g = -\alpha - \frac{1}{t}log\left[\frac{B_{gr}(t_r)}{C_{gr}(t_r)}\right]. \tag{38}$$

Steady state mRNA levels can now be derived from Equations (13) and (12):

$$\frac{dC_{gr}(t_r)}{dt} = \mu_g N_r(t_r) - \lambda_g C_{gr}(t_r) = m_g(t_r)\alpha N_r(t_r)$$

yields

$$m_g(0) = \frac{\mu_g}{\alpha + \lambda_g}.$$

As a result

$$C_{gr}(0) = \frac{\mu_g N_r(0)}{\alpha + \lambda_g} \tag{39}$$

gives the total mRNA level achieved by a dynamic equilibrium of a constant mRNA synthesis and decay. In this special case μ_g can be stated as

$$\mu_g = m_g(0)(\alpha + \lambda_g) \tag{40}$$

by implication.

Note: Without any disturbance of the cells, the synthesis rates μ_g and decay rates λ_g pertaining to the mRNA level of a transcript g are considered constant if averaged over a cell cycle period. Although there are transcripts whose synthesis rates vary considerably during the cell cycle, this assumption is valid as we measure a large, unsynchronized population of cells. The same argument holds if transcription happens in bursts, i.e. in each cell, periods of high transcriptional activity (many mRNA copies in a few minutes) are followed by periods of complete inactivity.

10 Adaption to the DTA measurement process

Variations in lysis and RNA extraction efficiencies, amplification steps in the protocol and the mandatory scanner calibration introduce discrepancies between the real concentrations of the mRNA fractions $A_{gr}(t_r)$, $B_{gr}(t_r)$ and $C_{gr}(t_r)$ and the measured intensity levels of the labeled mRNA $L_{gr}(t_r)$, unlabeled mRNA $U_{gr}(t_r)$ and total mRNA $T_{gr}(t_r)$. Ideally, these fractions would respectively equal each other. The amount $L_{gr}(t_r)$ of labeled mRNA for instance is proportional to the amount of labeled mRNA $A_{gr}(t_r)$ at the time t_r of sampling,

$$L_{gr}(t_r) = a_r A_{gr}(t_r) , \tag{41}$$

with an unknown array-specific constant a_r. Analogously, the measured amounts $T_{gr}(t_r)$ and $U_{gr}(t_r)$ depend on the actual amounts $C_{gr}(t_r)$ and $B_{gr}(t_r)$ respectively via

$$T_{gr}(t_r) = c_r C_{gr}(t_r) , \tag{42}$$

and

$$U_{gr}(t_r) = b_r B_{gr}(t_r) = b_r \left(C_{gr}(t_r) - A_{gr}(t_r) \right) \tag{43}$$

with unknown array-specific constants c_r and b_r.

There might also be differences in the fraction of newly synthesized mRNAs that escape labeling. This fraction is larger for shorter RNAs, and depends on the uracil content of the mRNA and the 4sU/4tU labeling efficiency (Section 11.1, Figure 9). Let p_r^{lab} be the probability that during the labeling process of sample r, a UTP is replaced by 4sUTP and afterwards attached to a biotin molecule. Let l_{gr} represent the fraction of newly synthesized mRNAs of gene g in sample r that are biotinylated. We assume that all biotinylated mRNAs are captured by the streptavidin-coated magnetic beads. Denote by $\#u_g$ the number of uracil residues present in the mRNA corresponding to gene g. The probability that the mRNA of gene g containing $\#u_g$ uracils is *not* captured at all (assuming independence of events for each uracil position) is $(1 - p_r^{lab})^{\#u_g}$. Conversely, the probability that a newly transcribed mRNA with $\#u_g$ uracils is captured in the labeled fraction is

$$l_{gr} = l(p_r^{lab}, \#u_g) = 1 - (1 - p_r^{lab})^{\#u_g} \tag{44}$$

l_{gr} is thus the probability that at least one UTP is replaced by 4sUTP and afterwards attached to a biotin molecule. Accordingly, we have to correct Equations (41),(42) and (43) to that effect. As a consequence we have the dependencies:

$$L_{gr}(t_r) = l_{gr} a_r A_{gr}(t_r) \,, \tag{45}$$

$$T_{gr}(t_r) = c_r C_{gr}(t_r) \tag{46}$$

and

$$U_{gr}(t_r) = b_r \left(C_{gr}(t_r) - l_{gr} A_{gr}(t_r) \right) . \tag{47}$$

(48) can also be stated as

$$U_{gr}(t_r) = b_r u_{gr} B_{gr}(t_r) \tag{48}$$

where u_{gr} gives the increase of pre-existing mRNAs by newly synthesized mRNAs of gene g in sample r that are not biotinylated. With Equation (47) and (48) we have

$$C_{gr}(t_r) = l_{gr} A_{gr}(t_r) + u_{gr} B_{gr}(t_r) \,, \tag{49}$$

and therefore

$$u_{gr} B_{gr}(t_r) - B_{gr}(t_r) = A_{gr}(t_r) - l_{gr} A_{gr}(t_r) \tag{50}$$

with respect to Equation (7). Solving for u_{gr} yields

$$u_{gr} = 1 + \frac{A_{gr}(t_r)}{B_{gr}(t_r)} (1 - p_r^{lab})^{\#u_g} \tag{51}$$

and thus

$$u_{gr} \left(p_r^{lab}, \frac{b_r}{a_r}, u_g \right) = 1 + \frac{b_r}{a_r} (1 - p_r^{lab})^{\#u_g} \quad \text{if} \quad \frac{u_{gr} L_{gr}(t_r)}{l_{gr} U_{gr}(t_r)} = 1 \, . \tag{52}$$

The proportional constants then relate as follows

$$\frac{c_r}{a_r} = \frac{\frac{b_r}{a_r}}{1 + \frac{b_r}{a_r}} \quad \text{if also} \quad \frac{L_{gr}(t_r)}{l_{gr} T_{gr}(t_r)} = 1 \, . \tag{53}$$

11 Parameter Estimation

In summary, our model contains the parameters

$$\Theta = \left\{ \alpha,\ p_r^{lab},\ a_r,\ b_r,\ c_r,\ l_{gr},\ u_{gr},\ \lambda_{gr},\ \mu_{gr} \mid r \in R,\ g \in G \right\}.$$

The two sources of experimental bias, namely the parameters a_r, b_r, c_r that account for multiplicative bias introduced via sample preparation, and the parameter p_r^{lab} that models the labeling bias (Figure 9), are part of the model and thus allow for an implicit normalization.

Methods for RNA quantification are often prone to measurement errors. Genes that exhibit residual expression, for instance, lead to noisy measurement signals. On the other hand, genes that show very high expression tend to saturate their corresponding probes on microarrays (see Section 3.3 and Figure 44). All these measurements could lead to flawed parameters, and hence all genes that are considered valid for the parameter estimation are aggregated in $G^{reliable}$.

Since the doubling times of the cells are usually known or can be measured accurately, α is given by $\alpha = \log 2/CCL$. We propose a 4-step procedure for the identification of the remaining parameters in Θ:

11.1 Estimation of the labeling probability p_r

The estimation of the sample-related parameters $\{p_r^{lab}, a_r, b_r, c_r \mid r \in R\}$ is done on the basis of the reliable genes $G^{reliable}$. The quotient of observed total and labeled mRNA levels can be written as

$$\frac{L_{gr}}{T_{gr}} = \frac{l_{gr} a_r A_{gr}(t_r)}{c_r C_{gr}(t_r)} = l_{gr} \frac{a_r}{c_r} \left[1 - e^{-t_r(\alpha + \lambda_g)} \right]. \tag{54}$$

The first equation follows by (45) and (46), the second by (26) and (21). We can visualize this dependence conveniently by plotting $\#u_g$ versus $\log \frac{L_{gr}}{T_{gr}}$ (see Figure 9). If all decay rates were equal, all points would lie on the graph given by the relationship of $\#u_g$ versus $\log l_{gr} + \log \frac{a_r}{c_r}$. The scatter around this graph is caused by measurement errors and differences in decay rates. We can also calculate the quotient

$$\frac{U_{gr}}{T_{gr}} = \frac{b_r u_{gr} B_{gr}(t_r)}{c_r C_{gr}(t_r)} = u_{gr} \frac{b_r}{c_r} \left[e^{-t_r(\alpha + \lambda_g)} \right]. \tag{55}$$

This equation follows by (46) and (47). We will predominantly use Equation (54) for the estimation of p_r^{lab} and the synthesis and decay rate estimates (see Section 13, Figure 28) due to our short labeling duration. Taking logs in Equation (54) and rearranging terms, we obtain

$$\log \frac{L_{gr}}{T_{gr}} = \log \frac{a_r}{c_r} + \log l(p_r^{lab}, \#u_g) + \log \left[1 - e^{-t_r(\alpha + \lambda_g)} \right]. \tag{56}$$

This equation holds for all genes. $p_r^{lab} < 0.95$ implies that for $\#u_g > 700$ say, the approximation $\log(l_{gr}) \approx 0$ is almost exact, as $(1 - p_r^{lab})^{\#u_g}$ is a monotonic sequence with

$$\lim_{\#u_g \to \infty} (1 - p_r^{lab})^{\#u_g} = 0$$

and $(1 - p_r^{lab})^{700} < 2.5 \cdot 10^{-16}$. Hence Equation (54) simplifies to

$$\log \frac{L_{gr}}{T_{gr}} = \log \frac{a_r}{c_r} + \log \left[1 - e^{-t_r(\alpha + \lambda_g)} \right] \quad \text{for} \quad \#u_g > 700. \tag{57}$$

If we furthermore assume that there is no systematic dependence of the decay rate λ_g on the uracil content of g, we can model the L_g/T_g ratio with respect to Equation (57) as

$$\log \frac{L_g}{T_g} = \log \left[1 - (1 - p^{lab})^{\#u_g} \right] + \epsilon_g$$

Figure 9: The number of uracils $\#u_g$ is plotted versus the log-ratio of L_{gr} and T_{gr} for two replicates of *S.cerevisiae* at a labeling time of 6 minutes. The points of the scatterplot are colored according to the (estimated) point density in that region [169]. The labeling bias parameter $p_r^{est} = 0.0048$ and $p_r^{est} = 0.0055$ imply that approximately every 208th resp. 182th uracil residue is replaced by 4sU. mRNAs which contain less than 500 uracil residues (approx. 72% of all *S.cerevisiae* mRNAs) are not captured efficiently.

with i.i.d errors ϵ_g, $g \in G^{reliable}$. Thus, we estimate $asymptote_r^{L/T}$ by setting

$$asymptote_r^{L/T} = \text{median} \left\{ \log \frac{L_{gr}}{T_{gr}} \mid g \in G^{reliable}, \ \#u_g > 700 \right\} . \tag{58}$$

Given Equation (58), it follows that a robust estimate of p_r^{lab} can be obtained by finding an optimal fit to (56) (see Figure(9)), for all $g \in G^{reliable}$ with $\#u_g < 500$. So we optimize the value of p_r^{lab}, $r \in R$, by minimizing the l_1-loss function

$$p_r^{est} = \underset{q \in (0,1)}{\text{argmin}} \, \text{loss}(q) \quad \text{with} \quad \text{loss}(q) = \sum_{g \in G, \#u_g < 500} \left| \log \frac{L_{gr}}{T_{gr}} - \log l_{gr}(q, u_g) - asymptote_r^{L/T} \right| \tag{59}$$

Here, 500 is an upper bound that ensures that the measurements are still responsive to changes in $\#u_g$.

11.2 Estimation of the ratio of labeled to unlabeled mRNA $\frac{b_r}{a_r}$

Notice that for the purpose of synthesis and decay rate estimation it is sufficient to merely determine the quotients $\frac{b_r}{a_r}$, $\frac{a_r}{c_r}$ or $\frac{b_r}{c_r}$ instead of the individual constants a_r, b_r and c_r. As described in Section (11.1), we can use Equation (55) to estimate

$$asymptote_r^{U/T} = \text{median} \left\{ \log \frac{U_{gr}}{T_{gr}} \mid g \in G^{reliable}, \ \#u_g > 700 \right\} \tag{60}$$

and consequently optimize the value of $\frac{b_r}{a_r}$, $r \in R$, by minimizing the l_1-loss function

$$\frac{b_r}{a_r}^{est} = \underset{q \in (0,5)}{\text{argmin}} \, \text{loss}(q) \quad \text{with} \quad \text{loss}(q) = \sum_{g \in G, \#u_g < 500} \left| \log \frac{U_{gr}}{T_{gr}} - \log u_{gr}(p_r^{lab}, q, \#u_g) - asymptote_r^{U/T} \right| . \tag{61}$$
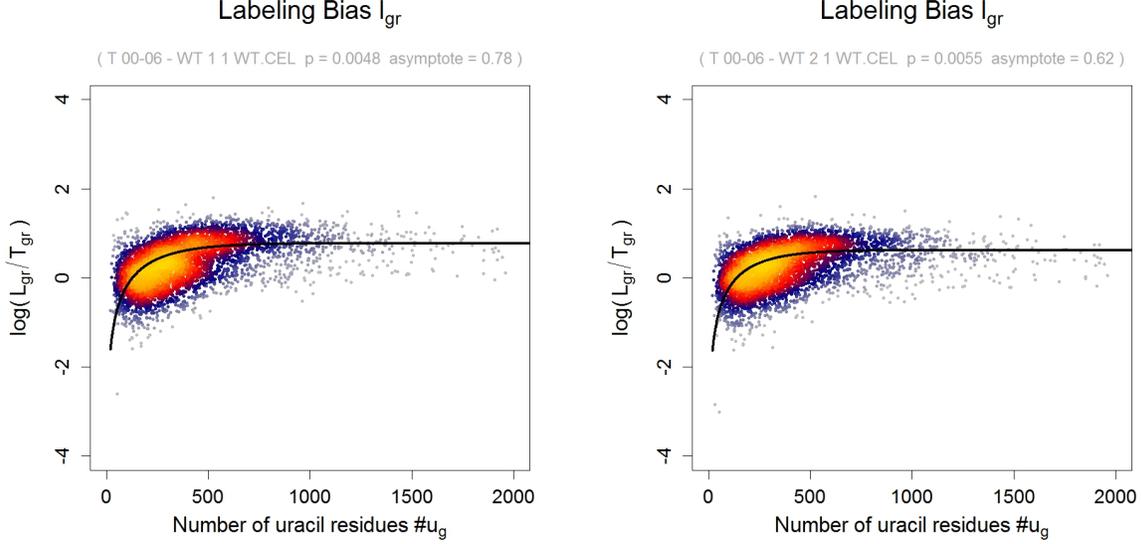
This dependence is shown in Figure 10.
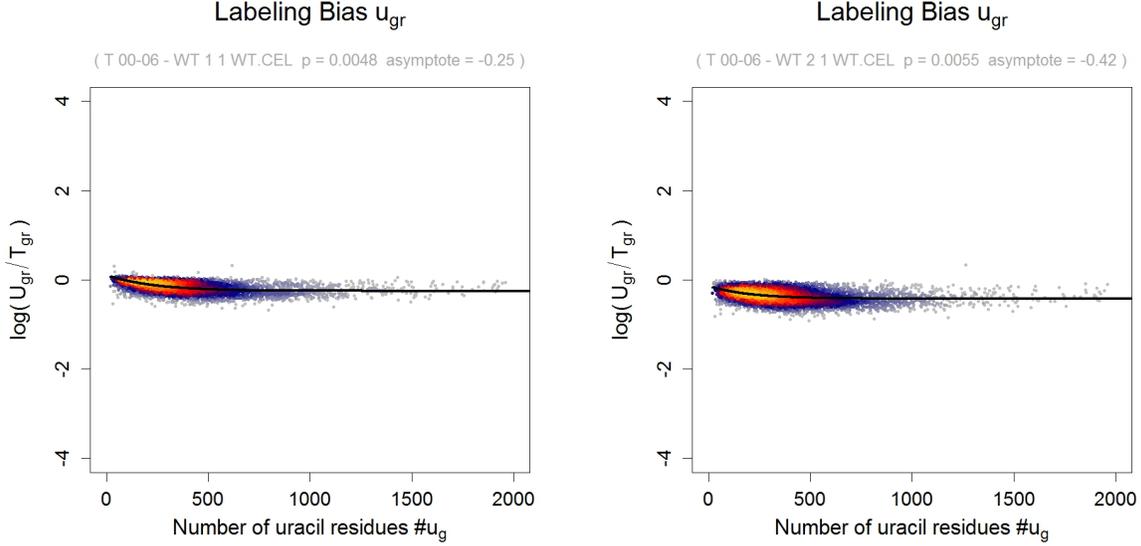
Figure 10: The number of uracils $\#u_g$ is plotted versus the log-ratio of U_{gr} and T_{gr} for two replicates of *S.cerevisiae* at a labeling time of 6 minutes. The labeling bias parameter $p_r^{est} = 0.0048$ and $p_r^{est} = 0.0055$ estimated via the labeled mRNA fraction L_{gr} as described in Section (11.1) reveal the ratio of newly synthesized to pre-existing mRNA given by $\frac{b_r}{a_r}^{est}$.

11.3 Estimation of the ratio of labeled to total mRNA $\frac{a_r}{c_r}$ and unlabeled to total mRNA $\frac{b_r}{c_r}$

In order to determine the quotients $\frac{a_r}{c_r}$ or $\frac{b_r}{c_r}$, we simply multiply Equations (54) and (55) by the inverse of those quotients and add them up to obtain

$$\frac{c_r}{a_r}\frac{L_{gr}}{T_{gr}} + \frac{c_r}{b_r}\frac{U_{gr}}{T_{gr}} = 1 \quad \text{or} \quad T_{gr}(t_r) = \frac{c_r}{a_r}L_{gr} + \frac{c_r}{b_r}U_{gr} \tag{62}$$

Equation (62) describes a plane $\{(T_{gr}, L_{gr}, U_{gr}) \mid T_{gr} = \frac{c_r}{a_r}L_{gr} + \frac{c_r}{b_r}U_{gr}\}$ in the 3-dimensional Euclidean space. As the observed variables have measurement errors on both sides of Equation (62), we perform a total least squares regression [76] of T_{gr} versus L_{gr} and U_{gr}, which accounts for a Gaussian error in the dependent variable T_{gr} and, in contrast to ordinary linear regression, also in the independent variables L_{gr}, U_{gr}. The total least squares regression minimizes the orthogonal distance of the data points to the inferred plane as opposed to a linear regression, which minimizes the distance of T_{gr} to the inferred linear function of L_{gr} and U_{gr}. We use a robust version of total least squares regression. After the first run, we remove the data points with the 5% largest residuals to avoid the potentially detrimental influence of outlier values on the parameter estimation process (see Figure (11) and (12)). Note that the estimation steps proposed in Section 11.2 and 11.3 are separately sufficient to calculate the synthesis and decay rate parameters (see Section 11.4), as the proportional constants $\frac{b_r}{a_r}$, $\frac{a_r}{c_r}$ or $\frac{b_r}{c_r}$ relate to each other according to Equation (53).
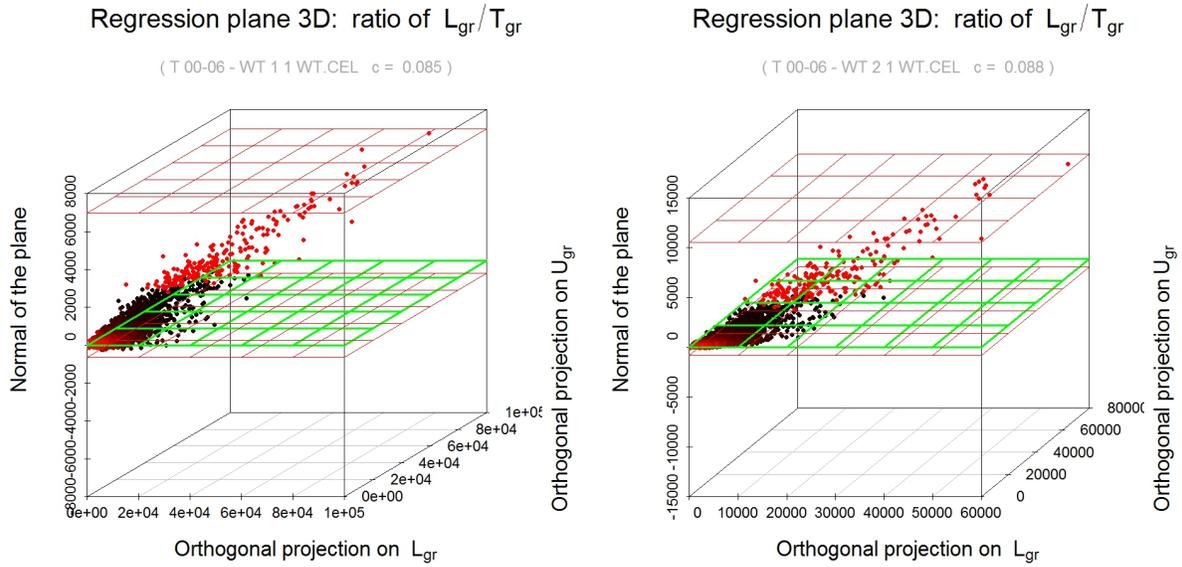
Figure 11: Two rounds of total least squares regression (*tls*) for two replicates of *S.cerevisiae* at a labeling time of 6 minutes. The resulting plane is colored green. The x-axis is chosen as the orthogonal projection on L_{gr}. The y-axis is the normal of the plane. The second round of *tls* is performed without the 5% largest residuals of the 1st round, depicted in red. Red planes indicate maximal residuals.
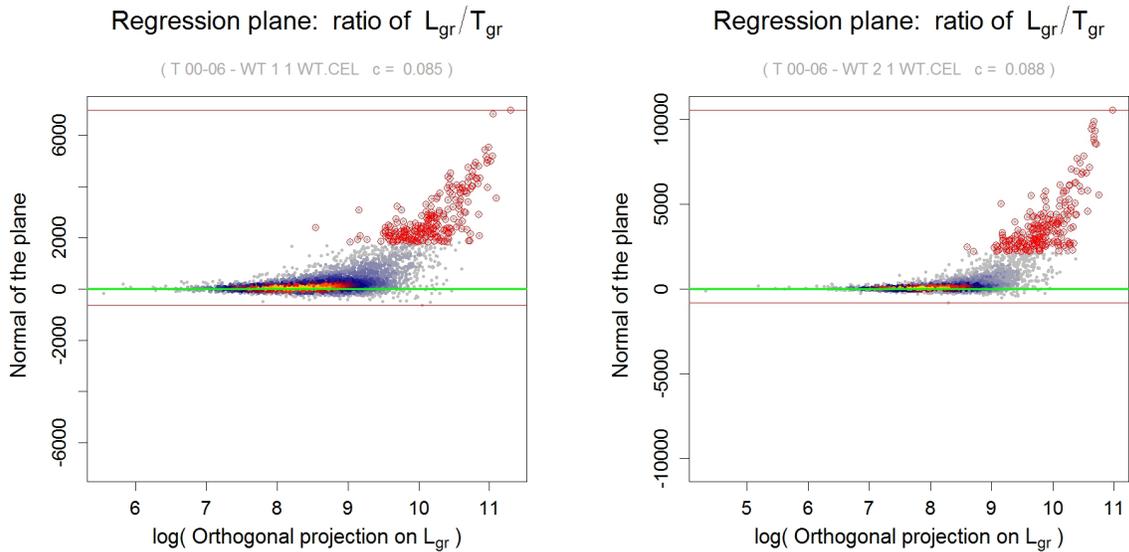
Figure 12: Two rounds of total least squares regression (*tls*) for two replicates of *S.cerevisiae* at a labeling time of 6 minutes in a different representation. The resulting plane is shown exactly from the side and is indicated by the green line. As in Figure (11), the x-axis is chosen as the orthogonal projection on L_{gr}, but in logarithmic scale. The y-axis is the normal of the plane. The 5% largest residuals of the 1st round are depicted in red. Red lines indicate maximal residuals.
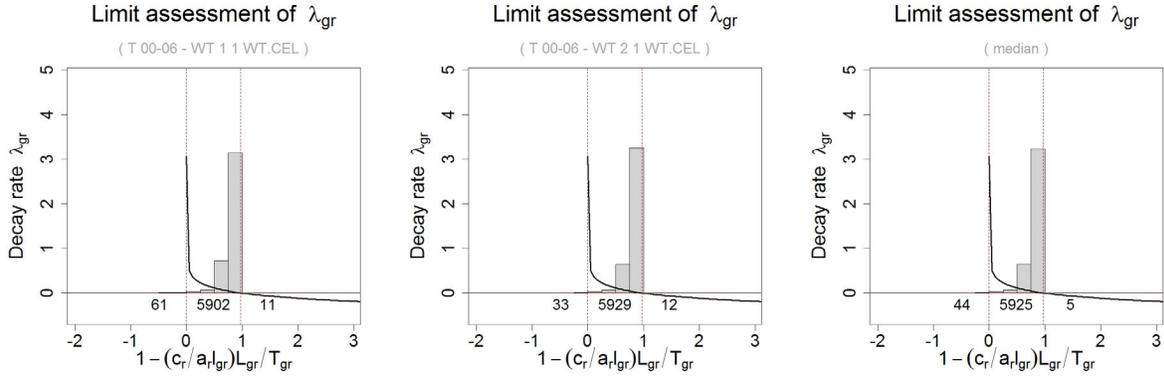
Figure 13: Dependency of the estimated decay rate on the term $1 - \frac{c_r}{a_r l_{gr}} L_{gr}/T_{gr}$ for two replicates of *S.cerevisiae* at a labeling time of 6 minutes. Reasonable decay rates can only be obtained for $1 - \frac{c_r}{a_r} L_{gr}/T_{gr}$ values between the two dashed lines. Right plot shows the median of both replicates.

11.4 Estimation of the decay rate λ_{gr}

To estimate the decay rate λ_{gr}, we use Equation (31) to yield

$$\lambda_{gr} = -\frac{1}{t_r} \log \left[\frac{T_{gr}(t_r) - \frac{c_r}{l_{gr} a_r} L_{gr}(t_r)}{T_{gr}(0)} \right] \tag{63}$$

for the dynamic case, and Equation (37) to deduce

$$\lambda_{gr} = -\alpha - \frac{1}{t_r} \log \left[1 - \frac{\frac{c_r}{l_{gr} a_r} L_{gr}(t_r)}{T_{gr}(t_r)} \right] \tag{64}$$

for the steady state case. It is also possible to estimate the decay rate λ_g in an alternate way according to Equation (38) as

$$\lambda_{gr} = -\alpha - \frac{1}{t_r} \log \left[\frac{\frac{c_r}{u_{gr} b_r} U_{gr}(t_r)}{T_{gr}(t_r)} \right] . \tag{65}$$

All three Equations (63), (64) and (65) are subject to logarithmic calculation. The decay rate λ_g can not be calculated in cases where the quotient in the logarithm is ≤ 0. If this quotient reaches a certain size, the Equations (63), (64) and (65) will yield negative decay rates which are then discarded. This circumstance is assessed in Figure (13). Each reasonable λ_g estimate can be used to calculate the half-life estimate

$$t_{1/2_{gr}} = \frac{\log(2)}{\lambda_{gr}} . \tag{66}$$

All measured samples r are combined to yield estimates

$$\lambda_g^{est} = \text{median} \left\{ \lambda_{gr} \mid r \in R \right\} , \tag{67}$$

and

$$t_{1/2_{gr}}^{est} = \text{median} \left\{ t_{1/2_{gr}} \mid r \in R \right\} . \tag{68}$$

The reproducibility of replicate measurements can be investigated by comparison of the quotient in the logarithm of Equations (63), (64) and (65), see Figure (14).

Figure 14: Pairwise heatscatter plots of the ranks of the respective $1 - \frac{c_r}{a_r l_{gr}} L_{gr}/T_{gr}$ value distributions for two replicates of *S.cerevisiae* at a labeling time of 6 minutes are shown in the upper panel. The lower panel shows the respective Spearman correlations.

11.5 Estimation of the synthesis rate μ_{gr}

To estimate the synthesis rate μ_{gr}, we use Equation (22) to yield

$$\mu_{gr} = \frac{\frac{c_r}{l_{gr} a_r} L_{gr}(t_r)(\alpha + \lambda_{gr})}{\left[e^{\alpha t_r} - e^{-\lambda_{gr} t_r} \right]} \ . \tag{69}$$

since $N_r(0)$ is set to 1, these values are on an arbitrary scale. For each replicate experiment r, the number m_{gr} of mRNA transcripts per cell of gene g is proportional to its total mRNA intensity value T_{gr}, $m_{gr} = d_r T_{gr}$. Assuming a total number of $\#mRNAs$ mRNAs per cell, this means that

$$\#mRNAs = \sum_{g \in G} m_{gr} = d_r \cdot \sum_{g \in G} T_{gr} \ , \tag{70}$$

and therefore

$$d_r = \frac{\#mRNAs}{\sum_{g \in G} T_{gr}} \ . \tag{71}$$

Together with Equation (40), we may estimate μ_{gr} as

$$\mu_{gr}^{rescaled} = \frac{m_{gr}}{T_{gr}} \mu_{gr} \cdot CCL = d_r \mu_{gr} \cdot CCL \tag{72}$$

in molecules per cell and cell cycle. All measured samples r are combined to yield estimates

$$\mu_g^{est} = \text{median} \left\{ \mu_{gr} \ \mid \ r \in R \right\} \ . \tag{73}$$

To get an overview of the correlations of the measured and derived value distributions, see Figure (15). As expected, the synthesis rate is in good agreement to the labeled mRNA fraction. The labeling bias was properly corrected: According to the assumption that the distribution of decay rates does not depend on the mRNA length, there is no dependence of any estimate to the number of uracil residues. Most surprisingly, the gene expression levels and mRNA half-lives correlate only weakly. We did extensive checks to ensure that this observation is not an artefact of the estimation procedure (Supplementary Section 13.5 [136]).
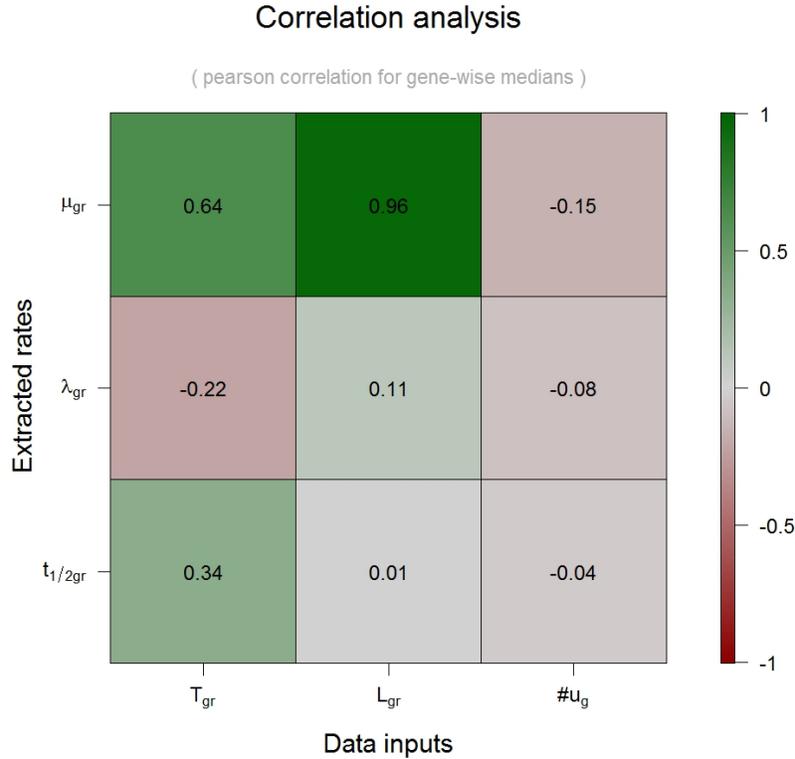
Figure 15: The pairwise correlations between the labeled expression values L_{gr}, total expression values T_{gr}, the number of uracils per transcript $\#u_g$ and the estimated synthesis rate μ_{gr}, decay rate λ_{gr} and half-life $t_{1/2_{gr}}$ is given in a color-coded image plot for the gene-wise median of two replicates of *S.cerevisiae* at a labeling time of 6 minutes .

11.6   Estimation of the relative 4sUTP incorporation efficiency

We define three probabilities associated with the mRNA labeling process: The *incorporation efficiency* p^{inc} is the probability that a 4sUTP nucleotide is incorporated into a nascent mRNA instead

of a UTP. It is the product of the relative 4sUTP concentration (relative to the UTP concentration) in the nucleus and the relative affinity of Pol II for 4sUTP (compared to UTP). Note that the relative affinity does not change if the transcription machinery is not impaired. Since we have no handle to separate these two quantities, we merge them into p^{inc}. Secondly, the *capture efficiency* p^{cap} is the probability for a 4sUTP which is included into a mRNA of being biotinylated, captured and recovered from the streptavidin-coated magnetic beads. Third, the probability that both events occur, assuming independence, is $p^{lab} = p^{inc} \cdot p^{cap}$, the *labeling efficiency* (Section 10, Equation (44)). A *labeling efficiency* substantially below 1 introduces a uracil-dependent labeling bias by letting newly transcribed, uracil-poor mRNA have a higher probability to escape labeling (Section 11.1). All three probabilities are sample- and strain-specific. The *labeling efficiency* can be estimated directly from cDTA data (Section 4.3). We can use cDTA to conclude from p^{lab} to the relative *incorporation efficiencies*. Note that for a sample x,

$$p_x^{cap}(Sc) = p_x^{cap}(Sp) \tag{74}$$

since the labeled Sc and Sp mRNA from one sample are processed simultaneously. Moreover,

$$p_x^{inc}(Sp) = p_y^{inc}(Sp) \tag{75}$$

for two experiments x, y, since we use a common Sp standard for all experiments. Then,

$$\frac{p_x^{lab}(Sc)}{p_x^{lab}(Sp)} = \frac{p_x^{inc}(Sc) \cdot p_x^{cap}(Sc)}{p_x^{inc}(Sp) \cdot p_x^{cap}(Sp)} \overset{(74)}{=} \frac{p_x^{inc}(Sc)}{p_x^{inc}(Sp)} . \tag{76}$$

Consequently, the relative *incorporation efficiencies* of two samples x and y is

$$\frac{p_x^{inc}(Sc)}{p_y^{inc}(Sc)} \overset{(75)}{=} \frac{p_x^{inc}(Sc)}{p_x^{inc}(Sp)} \cdot \frac{p_y^{inc}(Sp)}{p_y^{inc}(Sc)} \overset{(76)}{=} \frac{p_x^{lab}(Sc)}{p_x^{lab}(Sp)} \cdot \left(\frac{p_y^{lab}(Sc)}{p_y^{lab}(Sp)} \right)^{-1} . \tag{77}$$

Equation (77) can be used to estimate the variation in the relative *incorporation efficiency*, a variation estimate is given by

$$v = \text{std.dev.} \left(\frac{p_{x_1}^{inc}(Sc)}{p_{x_2}^{inc}(Sc)}, \ x_j \in \text{replicates of group } j \right) . \tag{78}$$

12 Model properties and limitations

A simulation environment can help to assess the properties of our model proposed in Section (7). We can simulate the impact of altered synthesis and decay rates on the resulting total mRNA level (Figure 16). For that purpose, we can solve the stated differential equations (13),(17) and (16) without loss of generality as

$$C_{gr}(t) = e^{\int_0^t -\lambda_g(\xi_1)d\xi_1} \left[C_{gr}(0) + \int_0^t N_r(\xi_2)\mu_g(\xi_2)e^{\int_0^{\xi_2} \lambda_g(\xi_1)d\xi_1} d\xi_2 \right] \tag{79}$$

for the total mRNA,

$$A_{gr}(t) = e^{\int_0^t -\lambda_g(\xi_1)d\xi_1} \left[A_{gr}(0) + \int_0^t N_r(\xi_2)\mu_g(\xi_2)e^{\int_0^{\xi_2} \lambda_g(\xi_1)d\xi_1} d\xi_2 \right] \tag{80}$$

for the newly synthesized mRNA and

$$B_{gr}(t) = B_{gr}(0)e^{\int_0^t -\lambda_g(\xi_1)d\xi_1} \tag{81}$$

for the pre-existing mRNA. Say we start labeling at time t_0, the amount of pre-existing mRNA at time t can be calculated as

$$B_{gr}(t, t_0) = C_{gr}(t_0)e^{\int_{t_0}^{t} -\lambda_g(\xi_1)d\xi_1} \tag{82}$$

with $B_{gr}(t_0) = C_{gr}(t_0)$. Equation (82) allows a general notation that emphasizes the start of the labeling process. For the newly synthesized mRNA at time t we consequently have

$$A_{gr}(t, t_0) = C_{gr}(t) - B_{gr}(t, t_0) , \tag{83}$$

as $A_{gr}(t) = C_{gr}(t) - B_{gr}(t)$, see Equation (7).

Given the steady state level of a certain mRNA population (39), it is possible to regulate the total mRNA amount to a new level either by alteration of the synthesis rate or the decay rate. According to our model (Section 7), it is not feasible for the cell to lower the total mRNA amount by reducing synthesis (Figure 17) nor induce it by reduction of decay (Figure 18). It is further much more efficient to accomplish mRNA level alterations directly. Up-regulation through synthesis and down-regulation through decay (Figure 17 and 18). As expected, regulation of mRNA levels by the opposing manipulation of both mechanisms seems also not to be required. In contrast to a very efficient way to reduce levels by means of decay which is mediated almost directly, up-regulation by induction of synthesis needs to be stimulated in a much higher fold to efficiently boost total mRNA levels (Figure 19).

The assumptions met in the proposed model (Section 7) might not be entirely correct. Nevertheless it is reasonable to assume that the synthesis rate affects the total mRNA level linearly (zero order reaction). It is also rational to model decay as an exponential process (first order reaction). However this does not take into account, that the mechanism of decay might also be regulated by a specific recruitment of the decay machinery. Another drawback of the previous discussed model is, that decay can affect newly synthesized transcripts as soon as they are produced. This however does not account for the delay given by the time a transcript needs to be exported to the cytoplasm, nor does it include the possibility of RNA binding proteins with stabilizing or destabilizing effect.
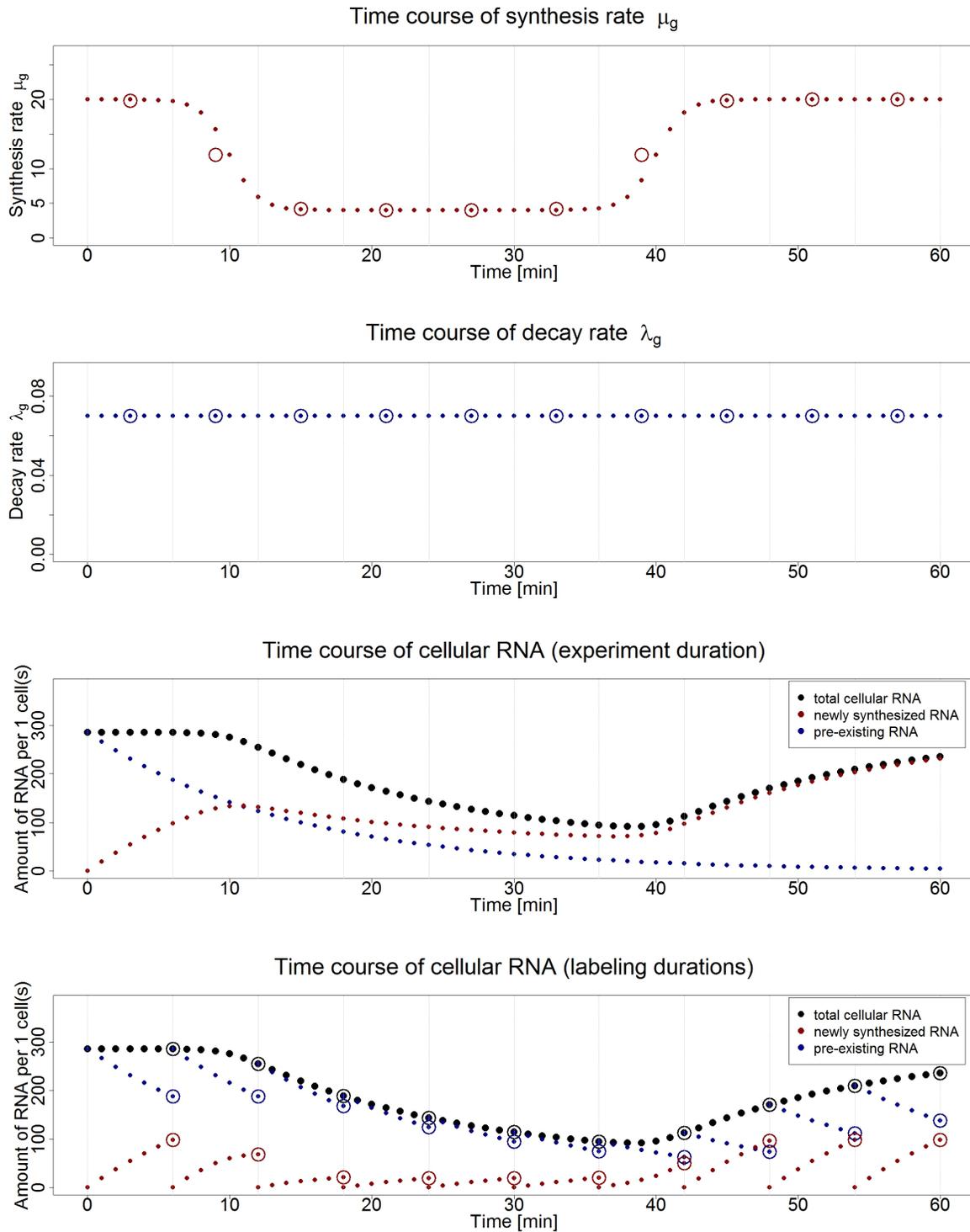
Figure 16: Plot shows a time course example of a single simulated gene given the course of the decay rate $\lambda_g(t) = const. = 0.07$ (shown in the 2nd panel in blue) and the synthesis rate $\mu_g = \mu_g(t)$ (shown in the 1st panel in red) which is down-regulated from 20 to 4 in a sigmoidal manner at 6 min and subsequently up-regulated to the initial level at 36 min. The 3rd and 4th panel show the development of the total, newly synthesized and pre-existing mRNA in black, red and blue respectively. The 4th panel shows the development of all three fractions for labeling durations of each 6 min. The respective expression readouts are indicated by colored circles.

Time course of synthesis rate μ_g

Time course of decay rate λ_g

Time course of total RNA (experiment duration)
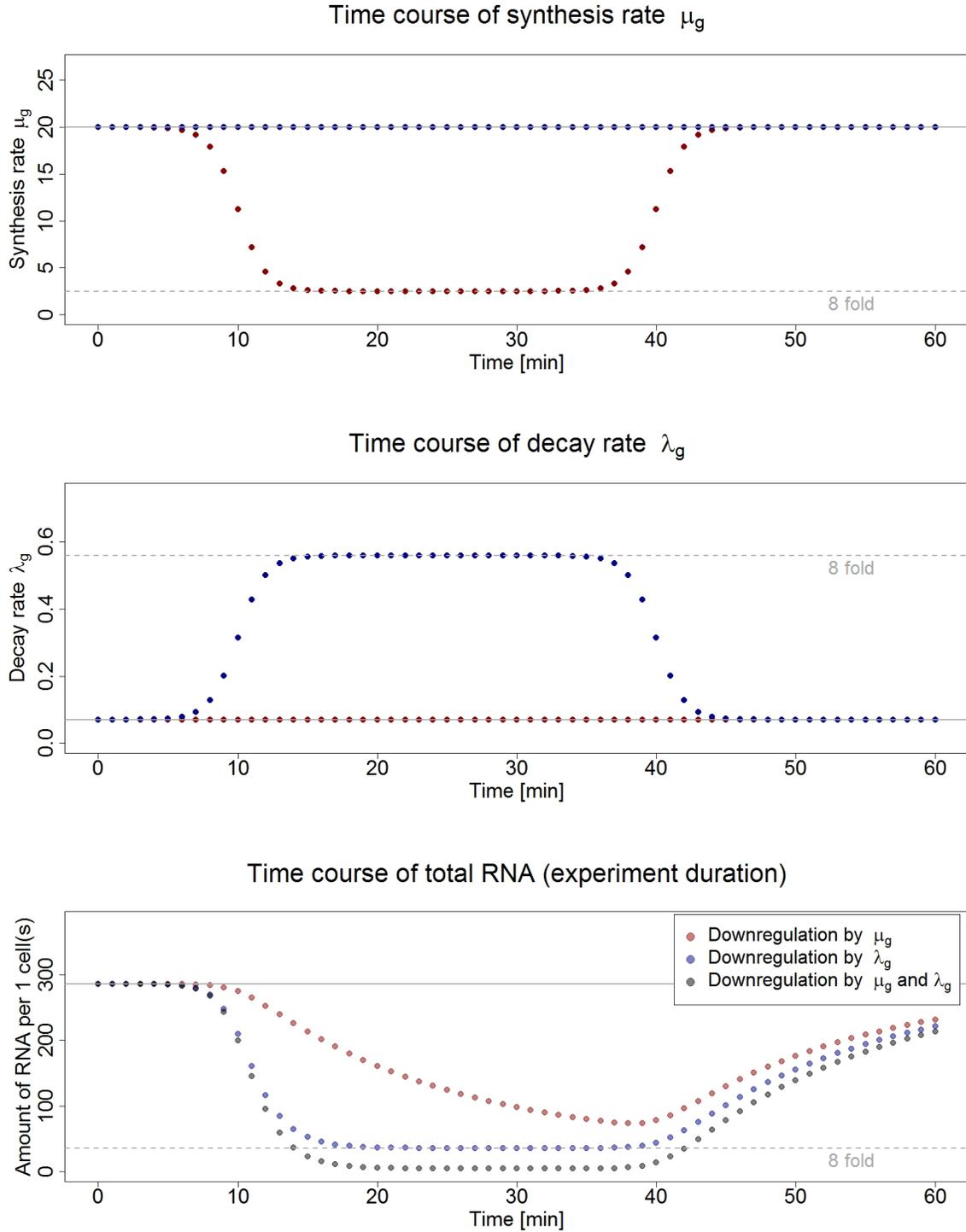
Figure 17: Plot shows a time course example of two simulated genes given the course of the synthesis rate $\mu_g = \mu_g(t)$ (shown in the 1st panel) and the decay rate $\lambda_g = \lambda_g(t)$ (shown in the 2nd panel). One gene is 8-fold down-regulated in the synthesis rate for 40 min with a constant decay rate (red). The other one is 8-fold up-regulated in the decay rate for 40 min with a constant synthesis rate (blue). The 3rd panel shows the development of the total mRNA for both down-regulation types in the respective color. The black dots indicate total mRNA resulting from a 8-fold down-regulated synthesis rate and a 8-fold up-regulated decay rate at the same time.
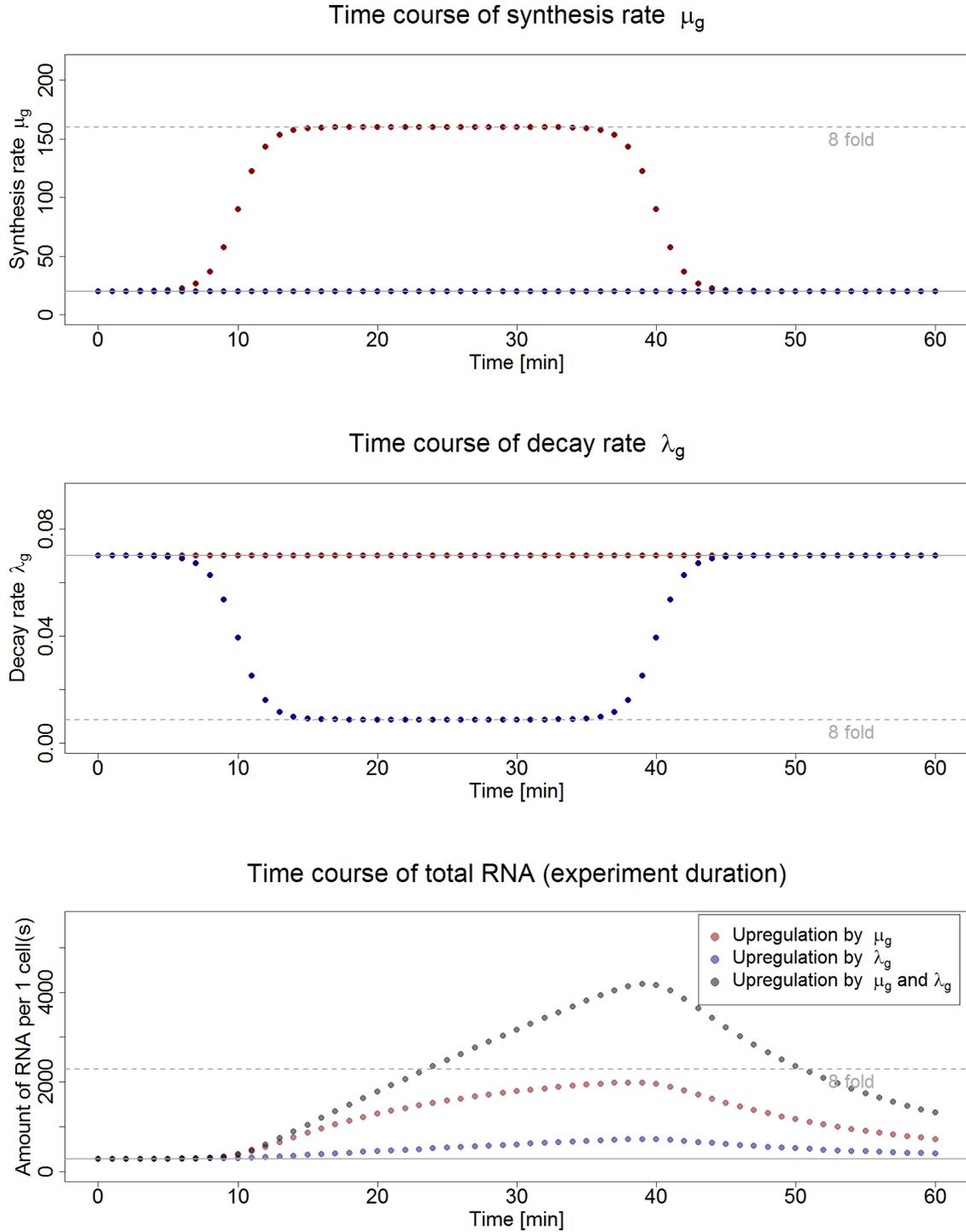
Figure 18: Plot shows a time course example of two simulated genes given the course of the synthesis rate $\mu_g = \mu_g(t)$ (shown in the 1st panel) and the decay rate $\lambda_g = \lambda_g(t)$ (shown in the 2nd panel). One gene is 8-fold up-regulated in the synthesis rate for 40 min with a constant decay rate (red). The other one is 8-fold down-regulated in the decay rate for 40 min with a constant synthesis rate (blue). The 3rd panel shows the development of the total mRNA for both up-regulation types in the respective color. The black dots indicate total mRNA resulting from a 8-fold up-regulated synthesis rate and a 8-fold down-regulated decay rate at the same time.
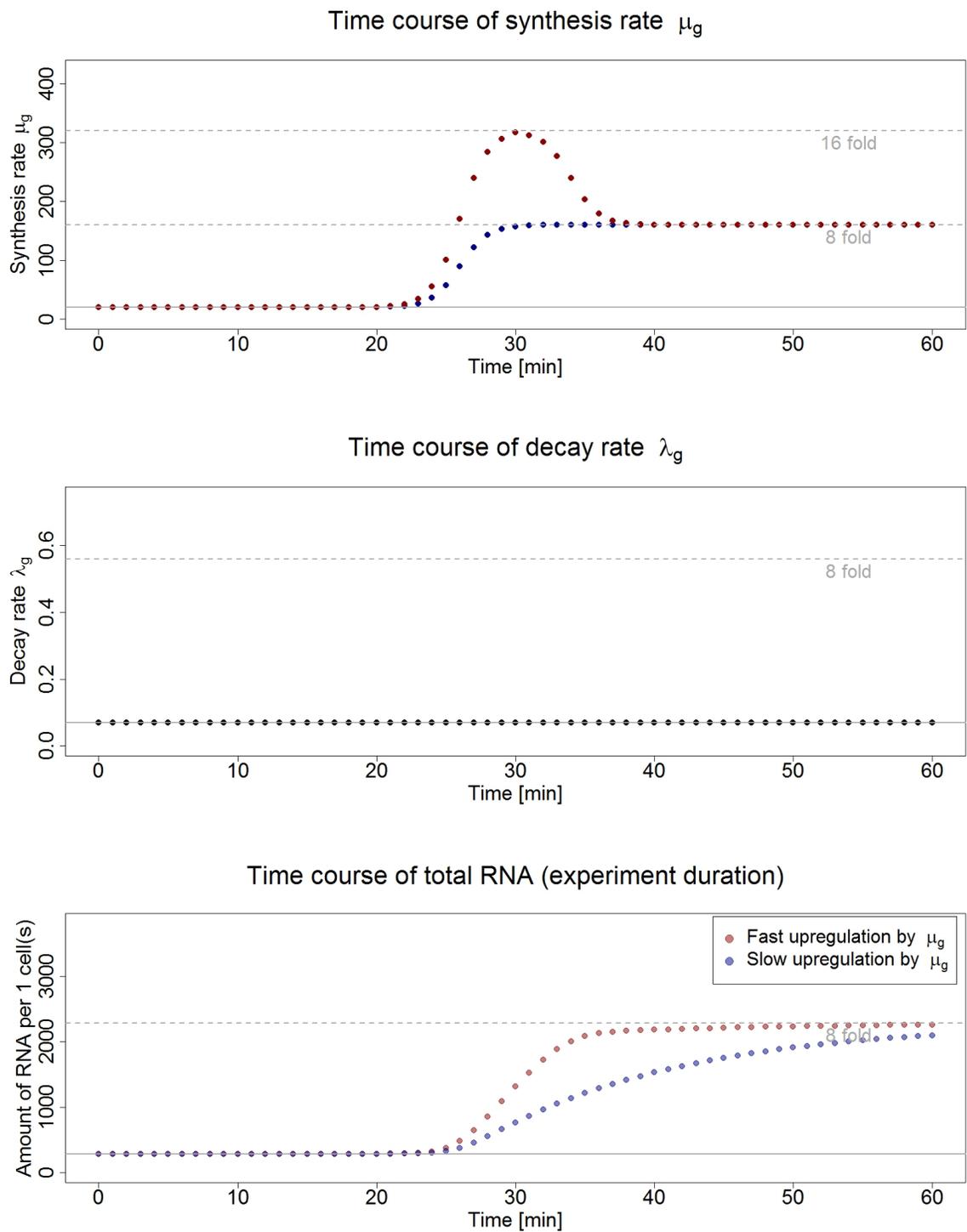
Figure 19: Plot shows a time course example of two simulated genes given the course of the decay rate $\lambda_g(t) = const. = 0.09$ (shown in the 2nd panel in blue) and the synthesis rate $\mu_g = \mu_g(t)$ (shown in the 1st panel in red) which are both up-regulated 8-fold (blue, red) and with a short additional 16-fold induction at the very beginning. The 3rd panel shows the development of the total mRNA for both up-regulation types in the respective color.

13   Error assessment and propagation

13.1   Regularized standard deviation

In the following we derive an error model which intends to assign confidence regions to the estimated parameters from Section 11. The parameters are estimated from microarray measurements (Section 3.3) and are thus prone to errors. Given that methods should be cost- and time-saving, the number of replicate measurements is often dramatically reduced. A low number of replicate measurements, however, is not sufficient to derive a good estimate for the sample variance. This typically requires at least 10 samples, which would make experiments prohibitively expensive and time-consuming. The following error model is intended to regularize the gene-wise empirical standard deviation according to the overall distribution of standard deviations on a microarray. It appears for typical microarray measurements that the variance of measured RNA intensities is a function of the intensity (Figure 20). This dependence can be modeled via a LOESS fit (see Section 18.1) to give the *gene-specific mean deviation*. The empirical standard deviation can then be regularized given the distribution of all gene-wise sample deviations for replicate measurements, i.e. the empirical standard deviation of a gene is corrected by its expected deviation from the *gene-specific mean deviation*. It is reasonable to assume that the error for the expression values on the log scale is normally distributed [94]:

$$\log x_g \sim \mathcal{N}(\mu_{x_g}, \sigma^2_{x_g}) \tag{84}$$

with gene-specific parameters μ_{x_g}, the sample mean and σ_{x_g}, the standard deviation. We further assume that the gene-specific standard deviation σ_{x_g} follows a gamma distribution (Figure 20)

$$\sigma_{x_g} \sim \Gamma(\frac{f(\mu_{x_g})^2}{\nu}, \frac{\nu}{f(\mu_{x_g})}) \tag{85}$$

with scale

$$\theta = \nu/f(\mu_{x_g}) \tag{86}$$

and shape

$$k = f(\mu_{x_g})^2/\nu \tag{87}$$

where $f(\mu_{x_g})$ is the *gene-specific mean deviation* obtained by the LOESS fit and ν is the variance of the gamma distribution derived from the LOESS residual distribution. ν is obtained via the method of moments (Section 18.2). To that end, the LOESS residual distribution is shifted, so that the smallest residual resembles 0. The gamma distribution in Equation (85) with its associated parameters, i.e. scale $\theta = \nu/f(\mu_{x_g})$ and shape $k = f(\mu_{x_g})^2/\nu$, then possesses the *gene-specific mean deviation* as expected value and the array-specific deviation ν as its variance (Histogram and red distribution in Figure 20).

We can further state the likelihood of the log(expression) values $\{x_{gi} \mid i = 1, \ldots, n\}$ (n gives the number of replicates), to be explained by the parameters μ_{x_g} and σ_{x_g}, and simultaneously that the parameter σ_{x_g} itself originates from a gamma distribution with parameters scale $\theta = \nu/f(\mu_{x_g})$ and shape $k = f(\mu_{x_g})^2/\nu$:

$$
\begin{aligned}
P(x_{gi}; \mu_{x_g}, \sigma_{x_g}, \nu) &= \prod_{i=1}^{n} \mathcal{N}(x_{gi}; \mu_{x_g}, \sigma^2_{x_g}) \Gamma(\sigma_{x_g}; \frac{f(\mu_{x_g})^2}{\nu}, \frac{\nu}{f(\mu_{x_g})}) \tag{88} \\
&= \prod_{i=1}^{n} \frac{1}{\sigma_{x_g}\sqrt{2\pi}} e^{-\frac{(x_{gi}-\mu_{x_g})^2}{2\sigma^2_{x_{gi}}}} \frac{1}{\theta^k} \frac{1}{\Gamma(k)} (\sigma_{x_g})^{k-1} e^{-\frac{\sigma_{x_g}}{\theta}})
\end{aligned}
$$

46

The parameter μ_{x_g} can be obtained from the sample population as the sample mean. As the parameter ν can be derived from the LOESS residual distribution via the method of moments (Section 18.2), we can derive a regularized standard deviation by finding

$$\sigma_{x_g}^{regularized} = \underset{\sigma_{x_g}}{\operatorname{argmax}} \left[P(x_{gn}; \mu_{x_g}, \sigma_{x_g}, \nu) \right] . \tag{89}$$
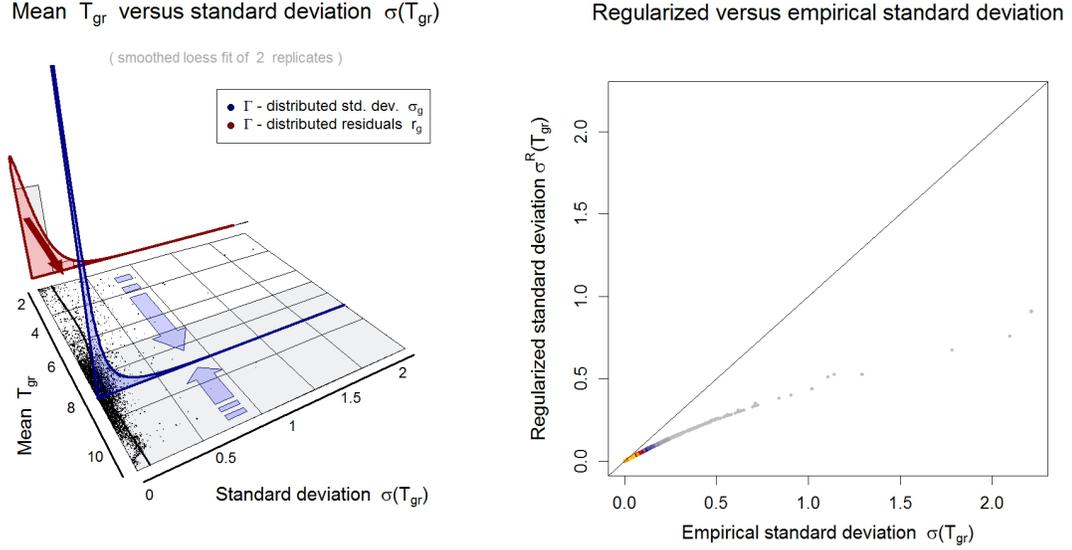
Figure 20: Left: Figure shows the gene-wise mean $\log(\text{expression})$ μ_{T_g} (x-axis) versus its standard deviation σ_{T_g} (y-axis), and the corresponding LOESS curve (black line) (Section 18.1). The red density (resp. histogram) shows the gamma distributed residuals (LOESS). Based on the variance of the red distribution the gene-specific standard deviation can be regularized if the number of replicate measurements is low. Right: The empirical versus the regularized standard deviation.

13.2   Sampling and confidence regions

The previously obtained regularized standard deviation can subsequently be used to calculate confidence regions of the extracted rates with a sampling approach. It is inevitable to investigate how the initial measurement error propagates numerically through the formulae of our model (Section 7). The regularized standard deviation is used instead of the empirical standard deviation in order to derive robust estimates of confidence intervals of the estimated rates. We can asses the reliability of half-life, synthesis and decay rate estimates, given a set of parameters: The mean expression μ_{L_g} and μ_{T_g} for the labeled and total fraction (bias corrected via l_{gr} and rescaled via $\frac{a_r}{c_r}$), the respective regularized standard deviations $\sigma_{L_g}^{regularized}$ and $\sigma_{T_g}^{regularized}$, the underlying labeling duration t and the cell cycle length CCL. Given the assumption of normal distributed errors for the total expression on the log scale, we construct

$$T_1, ..., T_N \underset{iid}{\sim} \mathcal{N}(\mu_{T_g}, \sigma_{T_g}^{regularized}) \tag{90}$$

with N realizations of the random variable T_i, giving the $\log(\text{expression})$ $\mu_{T_g} + \epsilon^T$ with $\epsilon^T \underset{iid}{\sim}$ $\mathcal{N}(0, \sigma_{T_g}^{regularized})$. Analogously to Equation (90), we construct

$$L_1, ..., L_N \underset{iid}{\sim} \mathcal{N}(\mu_{L_g}, \sigma_{L_g}^{regularized}) \tag{91}$$

with N realizations of the random variable L_i, giving the log(expression) $L_\mu + \epsilon^L$ with $\epsilon^L \underset{iid}{\sim} \mathcal{N}(0, \sigma_{L_g}^{regularized})$. We further set $\alpha = \log(2)/CCL$ according to our model in Section 7. The rate calculation then follows Equations (64), (68) and (69): The decay rate

$$\lambda_i = -\alpha - \frac{1}{t} \log \left[1 - \frac{L_i}{T_i} \right] \ , \tag{92}$$

the half-life

$$t_{1/2_i} = \frac{\log(2)}{\lambda_i} \tag{93}$$

and the synthesis rate

$$\mu_i = \frac{L_i(\alpha + \lambda_i)}{[e^{\alpha t} - e^{-\lambda_i t}]} \ . \tag{94}$$

Given the sample populations

$$\lambda_g^P = (\lambda_i)_{i=1,\dots,N} \ , \quad t_{1/2_g}^P = \left(t_{1/2_i}\right)_{i=1,\dots,N} \quad \text{and} \quad \mu_g^P = (\mu_i)_{i=1,\dots,N} \ , \tag{95}$$

we are able to assign confidence regions to the underlying rate estimates given the interquantile ranges

$$\mathcal{C}_{\lambda_g} = \left[Q_{0.025}(\lambda_g^P), Q_{0.975}(\lambda_g^P) \right] \ ,$$

$$\mathcal{C}_{t_{1/2_g}} = \left[Q_{0.025}(t_{1/2_g}^P), Q_{0.975}(t_{1/2_g}^P) \right] \ ,$$

and

$$\mathcal{C}_{\mu_g} = \left[Q_{0.025}(\mu_g^P), Q_{0.975}(\mu_g^P) \right] \ .$$

They are derived by using the 2.5 and 97.5 percentiles of the respective sample populations. Hence, these limits yield 95% confidence regions.

13.3 Parameter dependent error propagation

We are able to assess how a given error is propagated by the rate estimation under varying experimental and gene-specific parameters, given the framework in Section 13.2. For feasibility reasons, we will only assess how the *gene-specific mean deviations* $f(\mu_{T_g})$ and $f(\mu_{L_g})$ propagate according to changes in the underlying parameter set. The *gene-specific mean deviations* are modeled as the function giving the dependence of the log(expression) deviation among replicates and the log(expression) mean (Section 13.1). The *gene-specific mean deviation* observed for 12 wild-type replicates of *S.cerevisiae* is shown in Figure 21.

Mean T_g, L_g versus standard deviation $\sigma(T_g)$, $\sigma(L_g)$

(smoothed loess fit of 12 replicates r of *S.cerevisiae*)
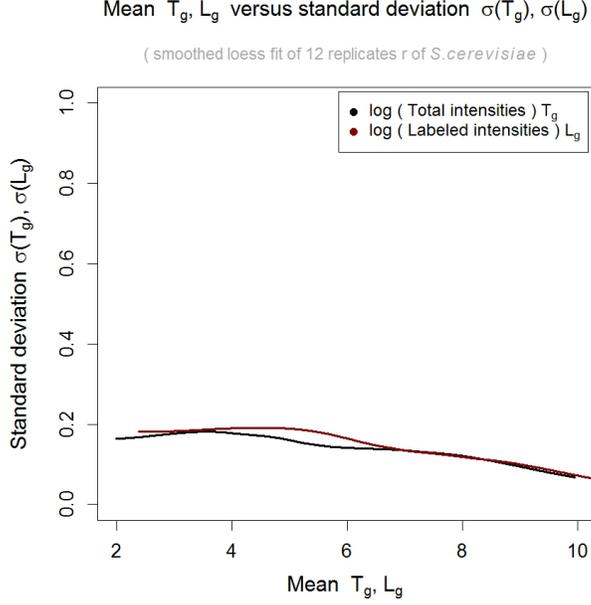
Figure 21: Plot shows a smoothed LOESS fit curve giving the relationship of the mean log(expression) value T_μ, L_μ on the x-axis and the standard deviation $\sigma(T_\mu)$, $\sigma(L_\mu)$ on the y-axis for the total mRNA fraction (black) and the labeled mRNA fraction (red). 12 replicate measurements of wild type *S.cerevisiae* after a 6 min labeling period were used to calculate a standard deviation (sd) and the mean intensities for each gene.

As initialization, we set the intensity dependent standard deviation according to the mean total log(expression) values $T_\mu = 8$ by evaluating the respective LOESS function (Section 18.1), that is

$$\sigma(T_\mu) = f(T_\mu) = 0.12 \ . \tag{96}$$

Likewise, we choose the following fixed parameters: The median decay rate $\lambda_\mu = \log(2)/15$, a labeling duration of $t_\mu = 6$ minutes and a corresponding ratio of labeled to total mRNA $\frac{a_\mu}{c_\mu}(t_\mu) = 0.15$. Additionally, we set the cell cycle length to $CCL = 90$ minutes, the labeling probability $p_\mu^{lab} = 0.005$ and the number of uracils to $\#u_\mu = 1000$.

In order to obtain comparability among different scales we build the coefficient of variation, which is defined as cv = standard deviation/mean. This transformation reveals that the relative error given by the coefficient of variation decreases with the mean expression value (Figure 20).

We can slightly alter the proposed sampling approach from Section 13.2 according to the following modifications: In order to obtain the given decay rate λ_μ, we set

$$offset_\mu = L_\mu - T_\mu = \log\left[\frac{l_\mu(p_\mu, \#u_\mu)\, a_\mu}{c_\mu}\left(1 - e^{-t_\mu(\alpha + \lambda_\mu)}\right)\right] \tag{97}$$

to define

$$L_\mu = T_\mu + offset_\mu \ . \tag{98}$$

This determines L_μ and all remaining parameters. We can use T_μ, L_μ and $\sigma(T_\mu)$, $\sigma(L_\mu)$ as the mean and standard deviation for Equations (90) and (91). The rate calculation then follows Equations (64), (68) and (69): The decay rate

$$\lambda_i = -\alpha - \frac{1}{t_\mu}\log\left[1 - \frac{\frac{c_\mu}{l_\mu a_\mu}L_i}{T_i}\right] \ , \tag{99}$$

49

the half-life

$$t_{1/2_i} = \frac{\log(2)}{\lambda_i} \tag{100}$$

and the synthesis rate

$$\mu_i = \frac{\frac{c_\mu}{l_\mu a_\mu} L_i(\alpha + \lambda_i)}{\left[e^{\alpha t_\mu} - e^{-\lambda_i t_\mu}\right]} \ . \tag{101}$$

As a measure for the relative error, we build the coefficient of variation for the decay rate

$$cv^\lambda = \frac{\text{sd}\left\{\lambda_i \,|\, i = 1, \dots, N\right\}}{\text{mean}\left\{\lambda_i \,|\, i = 1, \dots, N\right\}} \ , \tag{102}$$

the half-life

$$cv^{t_{1/2}} = \frac{\text{sd}\left\{t_{1/2 i} \,|\, i = 1, \dots, N\right\}}{\text{mean}\left\{t_{1/2 i} \,|\, i = 1, \dots, N\right\}} \tag{103}$$

and the synthesis rate

$$cv^\mu = \frac{\text{sd}\left\{\mu_i \,|\, i = 1, \dots, N\right\}}{\text{mean}\left\{\mu_i \,|\, i = 1, \dots, N\right\}} \ . \tag{104}$$

Note that, the sample mean and standard deviation - though they are extremely common measures for location and dispersion - do not reflect the skewness and kurtosis of the underlying distribution. Nevertheless, it is a sufficient measure in this case as the resulting distributions are approximately Gaussian (validation not shown).

In the following, we assessed the influence of one or two varying parameters on the relative error of the estimators synthesis rate, decay rate and half-life, leaving the other parameters fixed. As expected, the relative error observed in half-life, synthesis and decay rate estimates also decreases with the magnitude of the underlying expression value (Figures 22, 24 and 25). This motivates a lower intensity cut off (Section 11). This behavior, however, is not the case among a typical range of half-lives. As it turns out, the relative error of decay rates and half-lives increases almost linearly with the magnitude of the given half-life $t_{1/2_\mu}$, except for relative strong errors for very short half-lives (Figures 23 and 24). Reliable half-life estimates can therefore predominantly be obtained between 5 and 70 minutes. This phenomenon is explained by the increasing labeled to total ratio L_μ/T_μ as the decay rate rises, giving the error in the labeled fraction a stronger influence, see Equation (99). Surprisingly, the relative error in the synthesis rate almost vanishes, downsized by the decreasing given decay rate λ_μ. (Figure 23 and 25, Equation (101)).

It is also possible to estimate the decay rate according to Equation (65). This allows to compare both alternate estimators for the decay rate (Section 11.4). It turns out that the decay rate estimator based on the ratio L_μ/T_μ tends to exhibit smaller relative errors for short labeling durations t_μ, whereas the decay rate estimator based on U_μ/T_μ is prone to smaller relative errors for long labeling durations t_μ. The relative error increases as the ratio L_μ/T_μ or U_μ/T_μ approaches 1 (Figure 28). This justifies the use of Equation (54) instead of Equation (55), as we predominantly use a labeling time of 6 minutes.

Coefficient of variation σ/μ assessment (log(total expression))
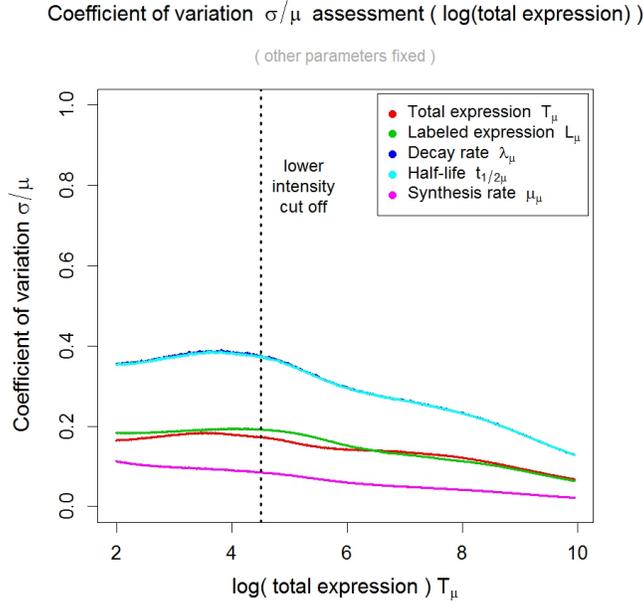
Figure 22: Plot shows the coefficient of variation for total expression T_μ, labeled expression L_μ, the decay rate λ_μ, the half-life $t_{1/2_\mu}$ and the synthesis rate μ_μ on the y-axis over the typical range of expression values (log-scale) on the x-axis leaving the other parameters fixed, i.e. the decay rate λ_μ, the ratio of labeled to total $\frac{a_\mu}{c_\mu}$, the labeling duration t_μ, the cell cycle length CCL, the labeling probability p_μ and the number of uracil residues $\#u_\mu$. Dashed line indicates the lower intensity cut off (Section 11).

Coefficient of variation σ/μ assessment (half-life)
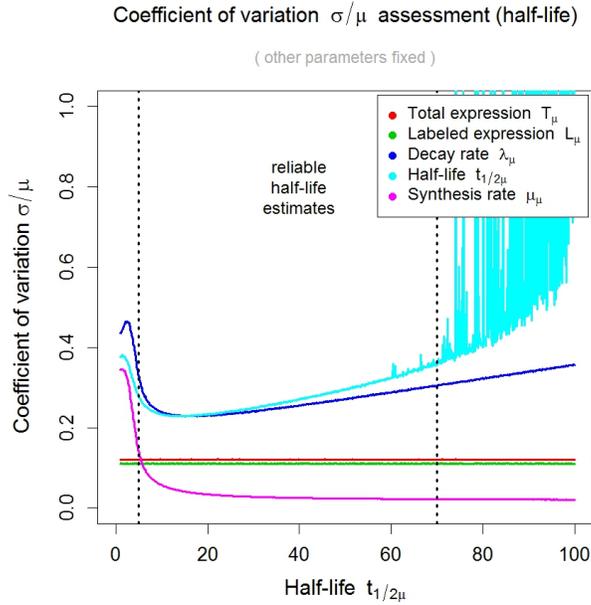
Figure 23: Plot shows the coefficient of variation for total expression T_μ, labeled expression L_μ, the decay rate λ_μ, the half-life $t_{1/2_\mu}$ and the synthesis rate μ_μ on the y-axis over the typical range of half-lives on the x-axis leaving the other parameters fixed. Dashed lines indicate that reliable half-life estimates range from 5 to 70 minutes.
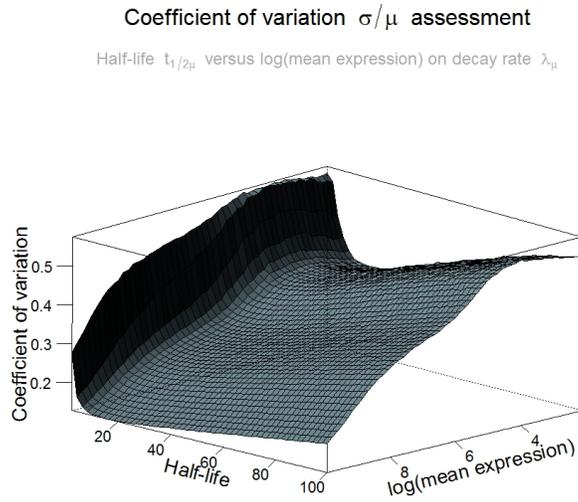
Figure 24: Plot shows the coefficient of variation for the decay rate λ_μ on the z-axis over the typical range of expression values (log-scale) on the y-axis and the typical range of half-lives on the x-axis leaving the other parameters fixed.
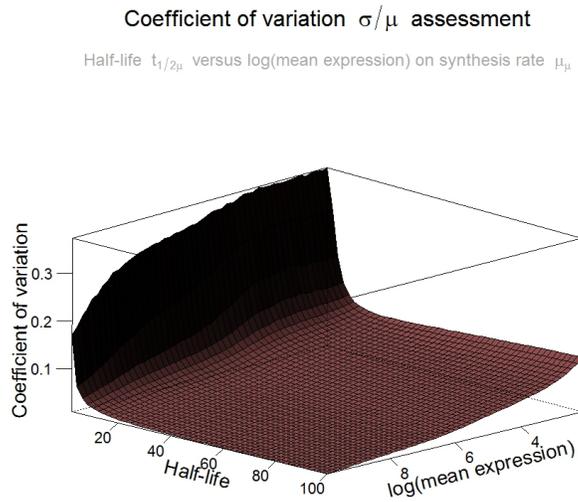
Figure 25: Plot shows the coefficient of variation for the synthesis rate μ_μ on the z-axis over the typical range of expression values (log-scale) on the y-axis and the typical range of half-lives on the x-axis leaving the other parameters fixed.
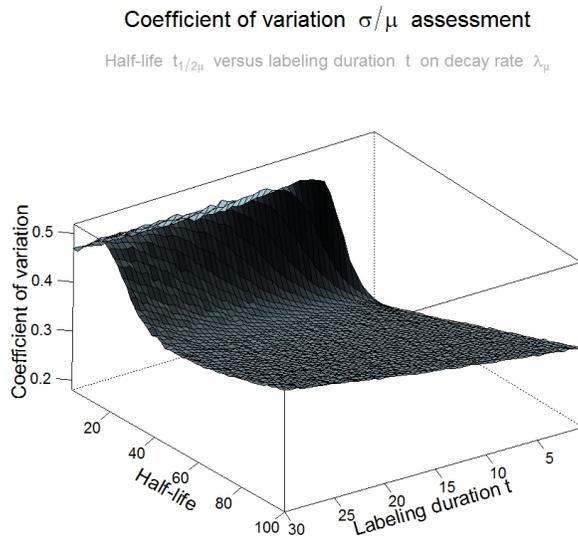
Figure 26: Plot shows the coefficient of variation for the decay rate λ_μ on the z-axis over the labeling duration on the y-axis and the typical range of half-lives on the x-axis leaving the other parameters fixed.
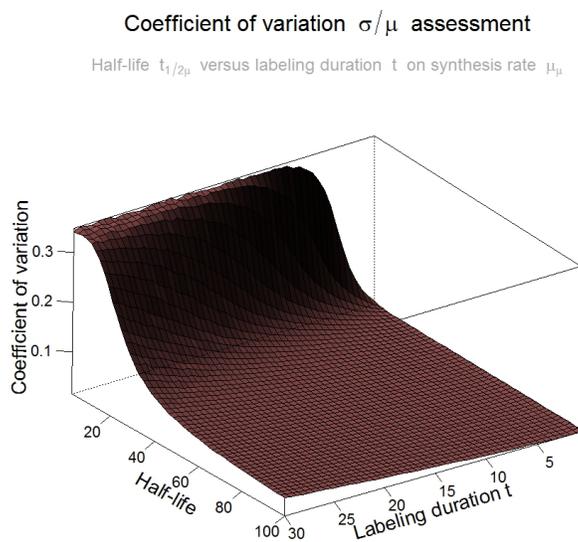
Figure 27: Plot shows the coefficient of variation for the synthesis rate μ_μ on the z-axis over the labeling duration on the y-axis and the typical range of half-lives on the x-axis leaving the other parameters fixed.

Coefficient of variation σ/μ assessment (labeling duration)

(other parameters fixed)
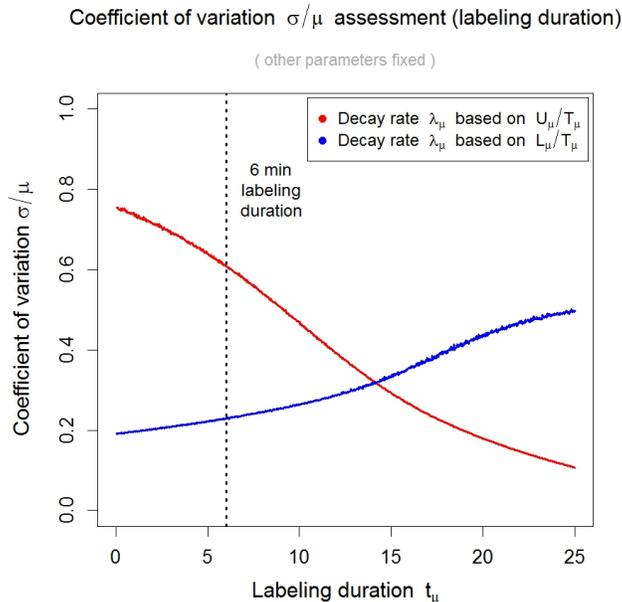
Figure 28: Plot shows the coefficient of variation for the decay rate λ_μ calculation based on the ratio L_μ/T_μ (blue) and based on the ratio U_μ/T_μ (red) on the y-axis over the labeling duration t_μ on the x-axis leaving the other parameters fixed. Dashed line indicates the labeling time predominantly used in our experiments. This justifies the use of Equation (54) instead of Equation (55).

14   *In silico* simulation

Simulations, based on the proposed model, are of major importance. They can be used to assess if the designed parameter estimation process works properly and does not introduce further biases. In general, a realistic *true* parameter set is chosen or created as a gold standard. Noise can subsequently be added to the *true* data set. Bias and variance can then be assessed by regaining the *true* parameter set in order to anticipate the prediction quality of the chosen model.

14.1   Steady state case

To examine the estimation procedure described in Section 11, we can simulate measurements of the labeled L_{gr}, unlabeled U_{gr} and total fraction T_{gr} for a given number of genes $g \in G$ and replicates $r \in R$ by providing a random half-life distribution, a random total mRNA distribution and a random sample of corresponding size giving the number of uracil residues. The latter is drawn randomly from an F distribution

$$u_g \underset{iid}{\sim} \mathcal{F}(5, 10) \cdot 350 \tag{105}$$

with degrees of freedom 5 and 10 and subsequently scaled by 350 to resemble a real distribution of uracil numbers observed in *S.cerevisiae*, *S.pombe* or *D.melanogaster* (data not shown). Half-lives are drawn randomly from an F distribution

$$t_{1/2_{gr}} \underset{iid}{\sim} \mathcal{F}(15, 15) \cdot 12 \tag{106}$$

with degrees of freedom 15 and 15 and subsequently scaled by 12 to resemble our right skewed observed distribution, see Section 15. This can in turn be used to yield the decay rate

$$\lambda_{gr} = \log(2)/t_{1/2_{gr}} \tag{107}$$

54

The *true* amount of log(total mRNA) at timepoint $t = 0$ is drawn randomly from a normal distribution with mean equal to 8 and standard deviation equal to 1. This gives a microarray-typical intensity distribution in log-scale

$$\gamma_{gr} \sim \mathcal{N}(8, 1) \tag{108}$$

and can subsequently be used to get absolute intensity levels

$$C_{gr}(0) = e^{\gamma_{gr}}. \tag{109}$$

The following steps are now analogous to our model (Section 7):

$$C_{gr}(t_r) = C_{gr}(0)e^{\alpha t_r} \tag{110}$$

where α is given by $\alpha = \log(2)/CCL$. t_r as well as CCL can be chosen arbitrarily in a reasonable manner. $T_{gr}(t)$ can be build from $C_{gr}(t)$ by adding noise:

$$T_{gr}(t_r) = C_{gr}(t_r) + \epsilon_{gr}^{total} \quad \text{with} \quad \epsilon_{gr}^{total} \underset{iid}{\sim} \mathcal{N}(0, 0.125 \cdot C_{gr}(t_r)) \tag{111}$$

where $\mathcal{N}(0, 0.125 \cdot C_{gr}(t_r))$ denotes a normal distribution with mean equal to 0 and standard deviation equal to $0.125 \cdot C_{gr}(t_r)$. This is consistent with the fact that the variance of measured data increases with intensity (data not shown). We set

$$c_r := \frac{\text{median} \{T_{gr}(t_r) \mid g \in G\}}{\text{median} \{C_{gr}(t_r) \mid g \in G\}} \tag{112}$$

which can be derived from Equation (42). The *true* amount of newly synthesized mRNA, called $A_{gr}(t)$ in our model, is constructed as follows:

$$A_{gr}(t_r) = C_{gr}(t_r)(e^{\alpha t_r} - e^{-\lambda_{gr} t_r}) \tag{113}$$

To obtain L_{gr} we include a labeling bias l_{gr} as in (44), and add noise:

$$L_{gr}(t_r) = l_{gr} A_{gr}(t_r) + \epsilon_{gr}^{labeled} \quad \text{with} \quad \epsilon_{gr}^{labeled} \underset{iid}{\sim} \mathcal{N}(0, 0.125 \cdot l_{gr} A_{gr}(t_r)) \tag{114}$$

We set

$$a_r := \frac{\text{median} \{L_{gr}(t_r) \mid g \in G\}}{\text{median} \{l_{gr} A_{gr}(t_r) \mid g \in G\}} \tag{115}$$

which can be seen from Equation (41). Further we rescale L_{gr} to the same range as T_{gr}:

$$L_{gr} = \frac{L_{gr} \cdot \text{median} \{T_{gr}(t_r) \mid g \in G\}}{\text{median} \{L_{gr}(t_r) \mid g \in G\}} \tag{116}$$

This strategy simulates amplification steps in the biochemical protocol and scanner calibration. We further need the *true* amount of unlabeled mRNA:

$$\widetilde{B}_{gr}(t_r) = C_{gr}(t_r) - l_{gr} A_{gr}(t_r) \tag{117}$$

and again we add noise

$$U_{gr}(t_r) = \widetilde{B}_{gr}(t_r) + \epsilon_{gr}^{unlabeled} \quad \text{with} \quad \epsilon_{gr}^{unlabeled} \underset{iid}{\sim} \mathcal{N}(0, 0.125 \cdot \widetilde{B}_{gr}(t_r)) \tag{118}$$

and also rescale U_{gr} to the same range as T_{gr}:

$$U_{gr} = \frac{U_{gr} \cdot \text{median} \{T_{gr}(t_r) \mid g \in G\}}{\text{median} \{U_{gr}(t_r) \mid g \in G\}} \tag{119}$$

Finally, we set

$$b_r := \frac{\text{median}\,\{U_{gr}(t_r)\mid g\in G\}}{\text{median}\,\left\{\widetilde{B}_{gr}(t_r)\mid g\in G\right\}} \tag{120}$$

We generated an artificial data set consisting of duplicate measurements for 5000 genes with a labeling duration of $t = 6$ and 12 minutes each. The cell cycle length CCL was set to 150 minutes, and the labeling probability p_r^{lab} was set to 0.005.
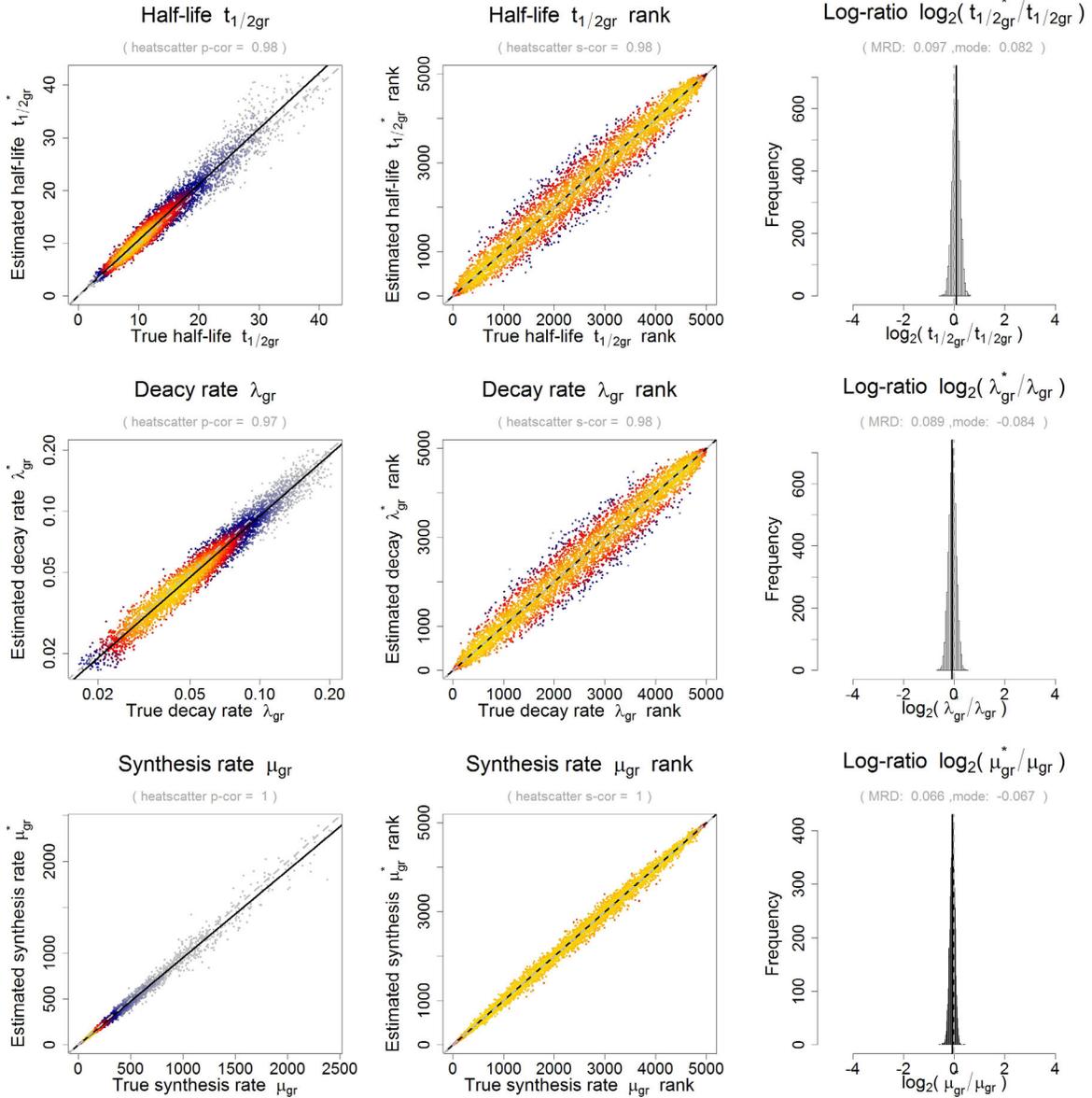
Figure 29: Comparison between *true* (= simulated) and estimated parameters in an absolute manner and on the basis of their ranks: synthesis rate μ_g, decay rate λ_g and half-life $t_{1/2_g}$. The rightmost plots shows the log-ratio of the estimated vs. the true parameters in a histogram. The mode is the maximum of the corresponding density indicated by the blue line. As a measure for a systematic deviation from zero (depicted in green) we calculated the mean relative deviation (MRD). The MRD is defined by mean $\left(\frac{|\tau_g^{est}-\tau_g^{true}|}{\tau_g^{true}}\right)$ with $\tau = \mu$, λ or $t_{1/2}$.

The *true* coefficients, that are to be recovered in (62) by the total least squares regression, are

now just the quotients of the individual constants a_r, b_r and c_r. We then applied our estimation procedure (Section 11) to fit the parameters $\frac{a_r}{c_r}$, $\frac{b_r}{c_r}$, $\frac{b_r}{a_r}$, p_r, λ_{gr}, $t_{1/2_{gr}}$ and μ_{gr} see (Figure 29).
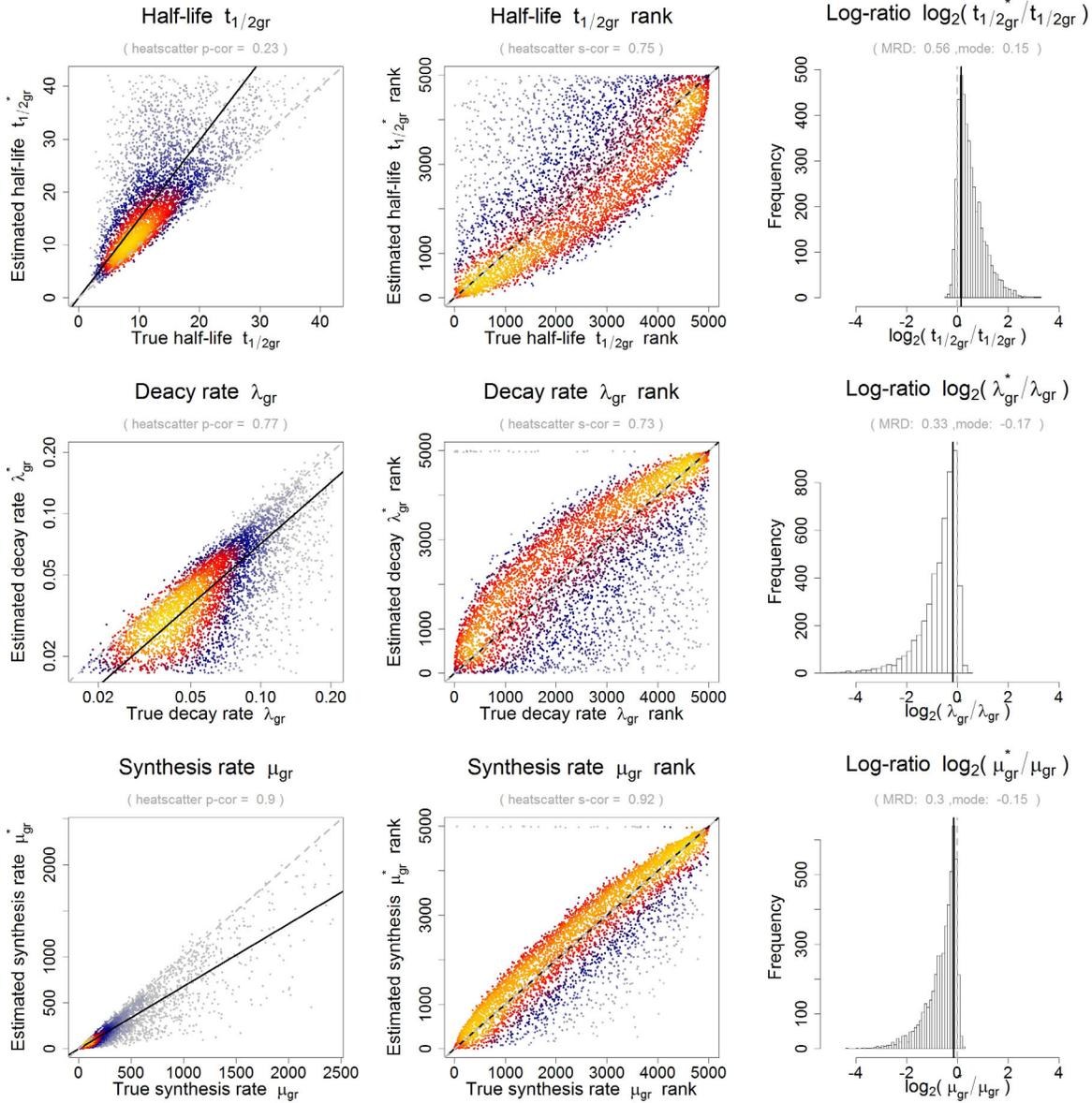
Figure 30: Comparison between *true* (= simulated) and estimated parameters obtained without labeling bias correction: synthesis rate μ_g, decay rate λ_g and half-life $t_{1/2_g}$. The rightmost plots shows the log-ratio of the estimated vs. the true parameters in a histogram. The mode is the maximum of the corresponding density indicated by the blue line. As a measure for a systematic deviation from zero (depicted in green) we calculated the mean relative deviation (MRD). The MRD is defined by mean $\left(\frac{|\tau_g^{est} - \tau_g^{true}|}{\tau_g^{true}} \right)$ with $\tau = \mu$, λ or $t_{1/2}$. The points in the scatterplots should be located directly and in a symmetric manner around the blue line to represent a good fit. The histogram is broader than in the results of our procedure, and it additionally shows a systematic bias towards longer half-lives and lower synthesis rates.

Omission of the labeling bias correction can lead to skewed estimates. To demonstrate this, we also used our simulated data to recompute the *true* coefficients without correcting for the labeling bias. There is a clear tendency to underestimate the decay rates as the transcripts with a uracil number of less than 500 make up almost 2/3 of all mRNAs. Assuming a labeling efficiency of $p^{lab} = 0.005$ as has

been estimated in our data. Figure 30 shows that omitting labeling bias correction results in half-life estimates that are severely systematically biased towards longer half-lives and lower synthesis rates. 20% of the mRNA half-lives in our experiment would have been biased more than twofold (Figure 31). The labeling bias is smaller for higher labeling efficiencies.
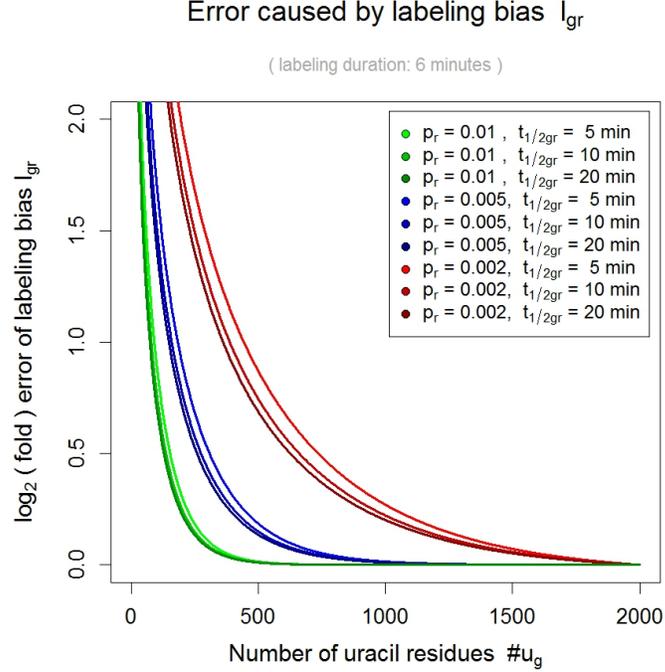
Figure 31: Plot shows $\#u_g$ vs. bias ($\log_2$(estimated decay/true decay)) curves, each curve corresponding to a given half-life (5, 10 or 20 min) and labeling efficiency ($p_r = 0.002, 0,005$ or 0.1).

14.2 Dynamic case

To examine the estimation procedure for the dynamic case described in Section 11, we can simulate measurements of the labeled L_{gr}, unlabeled U_{gr} and total fraction T_{gr} for a given number of genes $g \in G$ and replicates $r \in R$. By applying the formulae from Section 12 we can yield the values for A_{gr}, B_{gr} and C_{gr} and subsequently adopt the procedure given in Section 14.1 to finally derive L_{gr}, U_{gr} and T_{gr}. To improve over the fact that Equations (79) and (80) are subject to a long run time due to double integrals that need to be solved numerically, we propose the following restriction: The decay rate λ_g is assumed to be piecewise constant (Figure 32), i.e.

$$\lambda_g(t) = \lambda_{g,i} \quad h_{g,i-1} \le t < h_{g,i}, \quad i = 0, ..., n \quad h_{g,0} = 0, \quad h_{g,n} = \infty, \quad h_{g,i} \in \mathbb{N}_0^+, \quad \lambda_{g,i} \in \mathbb{R}^+. \tag{121}$$

Its integral can hence be written as sum

$$\Theta_g^*(t) := \int_0^t \lambda_g(\xi)d\xi = \sum_{\{i\,|\,t \le h_{g,i}\}} (\lambda_g \cdot min(t, h_{g,i}) - h_{g,i-1}), \tag{122}$$

see Figures 32 and 33.
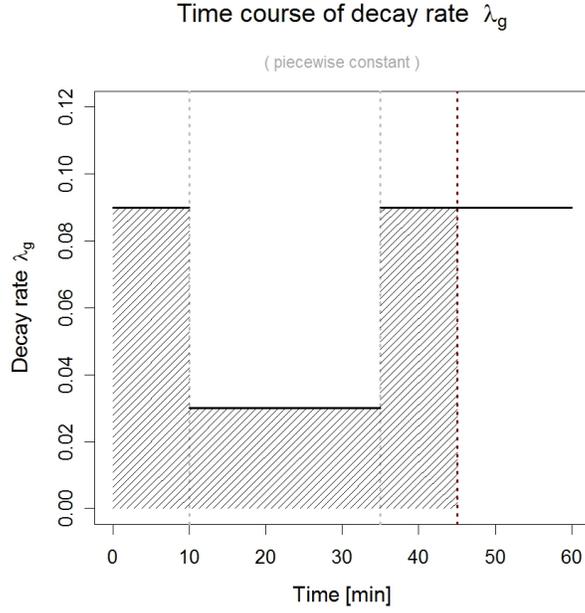
Figure 32: Plot shows an example of a piecewise constant decay rate. The striped area under the curve gives the value of its antiderivative at timepoint 45.
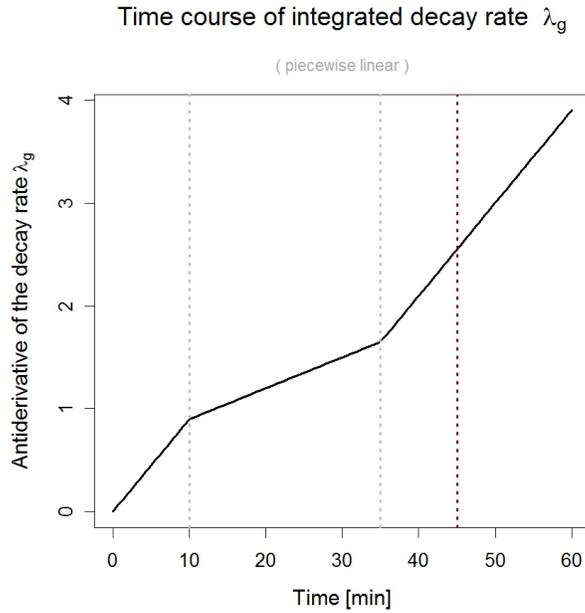
Figure 33: Plot shows the antiderivative of the decay rate function given in Figure 32. The resulting function can be expressed piecewise linear.

We set

$$\Theta_g(t, t_0) := \int_{t_0}^{t} \lambda_g(\xi)d\xi = \int_{0}^{t} \lambda_g(\xi)d\xi - \int_{0}^{t_0} \lambda_g(\xi)d\xi = \Theta_g^*(t) - \Theta_g^*(t_0) \tag{123}$$

and in this way

$$C_{gr}(t) = e^{-\Theta_g(t,0)} \left[C_{gr}(0) + \int_0^t N_r(0)\mu_g(\xi)e^{\alpha\xi+\Theta_g(\xi,0)}d\xi \right]. \tag{124}$$

We also set

$$B_{gr}(t,t_0) = C_{gr}(t_0)e^{-\Theta_g(t,t_0)} \tag{125}$$

with

$$B_{gr}(t_0) = C_{gr}(t_0) \tag{126}$$

and hence

$$A_{gr}(t,t_0) = C_{gr}(t) - B_{gr}(t,t_0). \tag{127}$$

We further assume the synthesis rate μ_g to be a piecewise constant function

$$\mu_g(t) = \mu_{g,i} \quad h_{g,i-1} \le t < h_{g,i}, \quad \mu_{g,i} \in \mathbb{R}^+. \tag{128}$$

Note: We can always achieve that the synthesis rate $\mu_{g,i}$ and the decay rate $\lambda_{g,i}$ have identical breakpoints $(h_{g,i})_{i=0,\dots,n}$ by refining the respective intervals.
As a consequence, Equation (124) can be formulated as

$$C_{gr}(t) = e^{-\Theta_g(t,0)} \left[C_{gr}(0) + N_r(0)\sum_i \phi(min(t,h_{g,i-1}), \, min(t,h_{g,i})) \right] \tag{129}$$

with

$$C_{gr}(0) = \frac{\mu_{g,0}}{\lambda_{g,0}+\alpha} \tag{130}$$

and

$$\phi(h_{g,i-1}, h_{g,i}) = \int_{h_{g,i-1}}^{h_{g,i}} \mu_{g,i}e^{\alpha\xi+\Theta_g(\xi,0)}d\xi. \tag{131}$$

Equation (131) can also be formulated as

$$\phi(h_{g,i-1}, h_{g,i}) = \int_{h_{g,i-1}}^{h_{g,i}} \mu_{g,i}e^{\alpha\xi+\lambda_{g,i}\cdot(\xi-h_{g,i-1})+\Theta_g(h_{g,i-1},0)}d\xi \tag{132}$$

and be solved as

$$\phi(h_{g,i-1}, h_{g,i}) = \left[\frac{\mu_{g,i}e^{\alpha\xi+\lambda_{g,i}\cdot(\xi-h_{g,i-1})+\Theta_g(h_{g,i-1},0)}}{\alpha+\lambda_{g,i}} \right]_{h_{g,i-1}}^{h_{g,i}} \tag{133}$$

So we finally get

$$\phi(h_{g,i-1}, h_{g,i}) = \frac{\mu_{g,i}e^{\Theta_g(h_{g,i-1},0)} \left[e^{\alpha h_{g,i}+\lambda_{g,i}\cdot(h_{g,i}-h_{g,i-1})} - e^{\alpha h_{g,i-1}} \right]}{\alpha+\lambda_{g,i}} \tag{134}$$

We simulated profiles for 5000 genes g in the following way: The initial synthesis rate and decay rate distributions are drawn randomly from F distributions.

$$t_{1/2_g} \underset{iid}{\sim} \mathcal{F}(15,15) \cdot 12 \tag{135}$$

to yield

$$\lambda_g^{initial} = \log(2)/t_{1/2_g} \tag{136}$$

and

$$\mu_g^{initial} \underset{iid}{\sim} \mathcal{F}(5,5) \cdot 18. \tag{137}$$

We further set

$$shock_g \underset{iid}{\sim} 1/\max(\eta_g, 1), \quad \eta_g \underset{iid}{\sim} \mathcal{N}(8,4) \tag{138}$$

and

$$induction_g \underset{iid}{\sim} \max(\iota_g, 1), \quad \iota_g \underset{iid}{\sim} \mathcal{N}(8,4). \tag{139}$$

For the synthesis rate we set

$$\mu_g(t) = \begin{cases} \mu_g^{initial} & t < 12 \\ \mu_g^{initial} \cdot shock_g & 12 \leq t < 18 \\ \mu_g^{initial} \cdot shock_g \cdot induction_g & t \geq 18 \end{cases} \tag{140}$$

and for the decay rate

$$\lambda_g(t) = \begin{cases} \lambda_g^{initial} & t < 12 \\ \lambda_g^{initial} \cdot shock_g & 12 \leq t < 18 \\ \lambda_g^{initial} \cdot shock_g \cdot induction_g & 18 \leq t < 2 \\ \lambda_g^{initial} & t \geq 27 \end{cases} \tag{141}$$
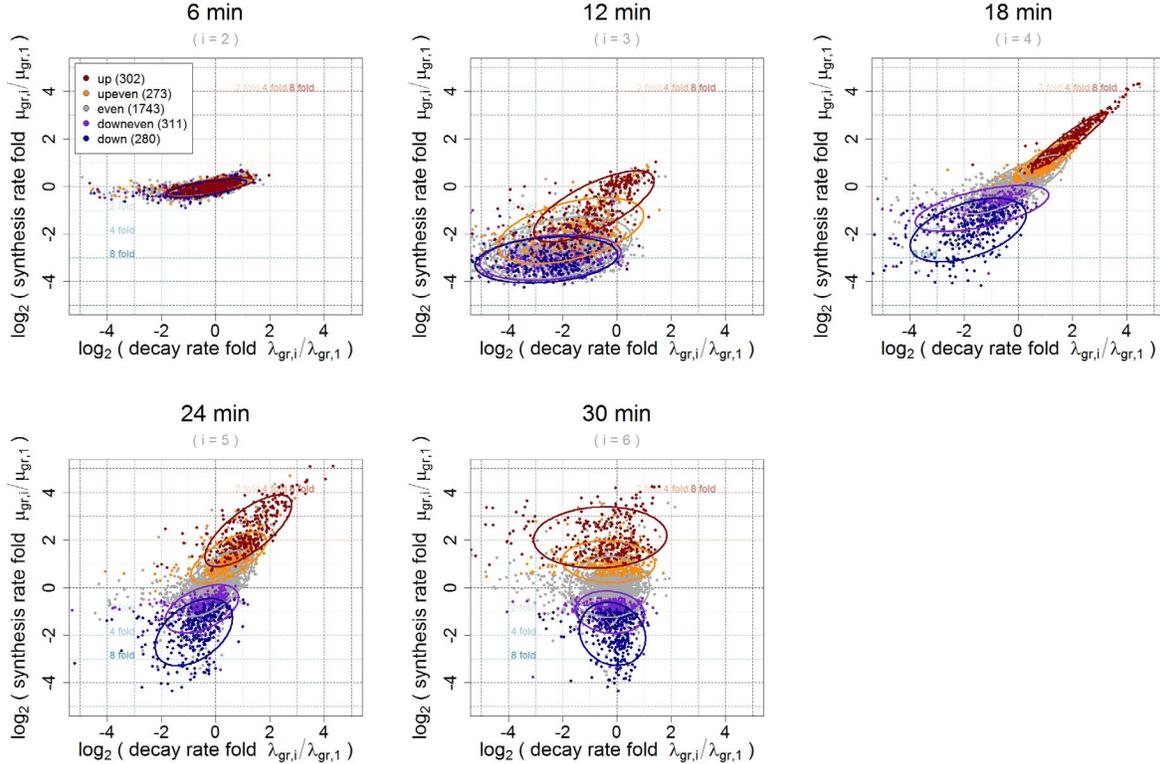
Figure 34: Each plot corresponds to one timepoint. Log_2(decay rate fold $\lambda_{g,i}/\lambda_{g,1}$) for $i = 2, \dots, 6$ versus $\log_2$(synthesis rate fold $\mu_{g,i}/\mu_{g,1}$) for $i = 2, \dots, 6$ for the last timepoint compared to the first timepoint. Each point corresponds to one gene, which is colored according to its affiliation with one of 5 clusters defined in a normalization-independent manner [170]. Ellipses show the 75% regions of highest density within each cluster, assuming Gaussian distributions. The shape of the ellipses indicates the correlation structure within a cluster.

As a result, Equations (140) and (141) give a time course of 36 minutes in total (Figure 34). The labeling durations were each set to 6 minutes. We then applied our estimation procedure (Section 11) to fit the parameters $\frac{a_r}{c_r}$, $\frac{b_r}{c_r}$, $\frac{b_r}{a_r}$, p_r, λ_{gr}, $t_{1/2_{gr}}$ and μ_{gr} see (Figure 35).
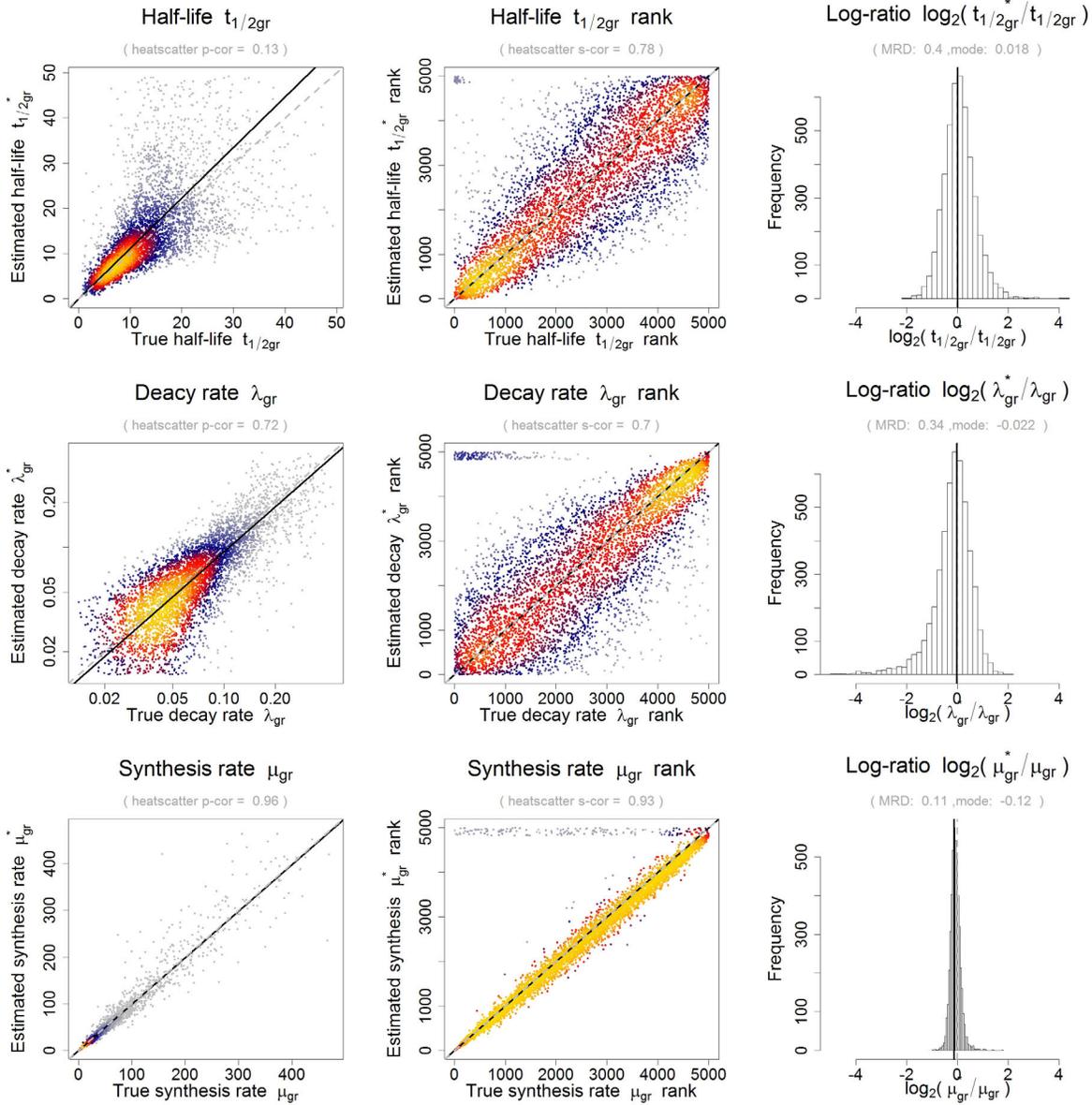
Figure 35: Comparison between *true* (= simulated) and estimated parameters in an absolute manner and on the basis of their ranks: synthesis rate μ_g, decay rate λ_g and half-life $t_{1/2_g}$ of the time window $18-24$. The rightmost plots shows the log-ratio of the estimated vs. the true parameters in a histogram. The mode is the maximum of the corresponding density indicated by the blue line. As a measure for a systematic deviation from zero (depicted in green) we calculated the mean relative deviation (MRD). The MRD is defined by $\text{mean}\left(\frac{|\tau_g^{est} - \tau_g^{true}|}{\tau_g^{true}}\right)$ with $\tau = \mu$, λ or $t_{1/2}$.

62

Part III
Results & Discussion

15   DTA measures rates of mRNA synthesis and decay in *S.cerevisiae*

Cellular growth and stress response require gene regulation at the level of mRNA transcription and stability. To study gene regulation in a eukaryotic cell, the rates of mRNA synthesis and decay must be measured without perturbation of the cellular system, and changes in these rates must be monitored to follow a cellular response. This cannot be achieved by standard transcriptomics, which only measures mRNA abundance. The yeast *Saccharomyces cerevisiae* is an ideal model eukaryote for systemic analysis, but mRNA synthesis and decay rates can currently not be measured without cellular perturbation. Synthesis rates can be measured by nuclear run on [67], but this requires sarkosyl treatment that inhibits cellular processes. Decay rates can be measured after blocking transcription with inhibitors (Section 3.2) [77, 18, 78, 162, 174, 215, 7, 143] but this is inherently cell invasive. In addition, decay rates can be measured with a temperature-sensitive yeast strain (Section 3.1) [90, 205, 77, 176], but this requires a perturbing heat shock.

Unperturbed RNA synthesis and decay rates can be obtained via metabolic RNA labeling and kinetic modeling (Section 4) [36, 102, 52, 64, 137]. The nucleoside analog 4-thiouridine (4sU) is taken up by eukaryotic cells and incorporated into mRNA during Pol II transcription (Section 4.1) [133]. The thiol-labeled newly transcribed RNA can then be isolated by affinity chromatography [102] or by biotinylation and purification with streptavidin-coated magnetic beads [36, 52]. Although this approach is generally applicable to mammalian, insect, and plant cells, RNA labeling is not directly applicable to yeast.

In this study, we report on the development of an easy-to-use, non-perturbing method to measure mRNA synthesis and decay rates in yeast, referred to as DTA (Section (4.2)). Thereby, changes in synthesis and decay rates can be monitored in yeast at unprecedented sensitivity and temporal resolution. This is exemplified by analysis of the osmotic stress response, a conserved stress response pathway and one of the best studied gene-regulatory systems in yeast [89, 48].

Osmotic stress response is induced in yeast upon exposure to high concentrations of salt. The stress response involves activation of the conserved MAP kinase Hog1, which induces an altered activity of ion membrane transporters [157], cell cycle arrest [56], dissociation of many chromatin-bound proteins [157], translation inhibition [196], and reprogramming of transcription [124]. The osmotic stress response was analyzed globally by transcriptomics [70, 33, 132, 124], run-on analysis [166], and transcription inhibition [139]. These studies revealed changes in mRNA synthesis and decay and suggested three phases of the stress response, referred to as shock, induction, and recovery phase.

We report that DTA recaptures many known features of the stress response, but that it also provides new insights and uncovers misleading drawbacks of traditional methods. DTA reveals new salt stress genes, and a temporary interdependence of mRNA synthesis and decay. The results are validated by genomic occupancy profiling of Pol II before and after stress. This confirmed that redistribution of Pol II over the genome predicts global changes in mRNA synthesis rates. These results establish DTA as a highly valuable tool for the analysis of dynamic changes in mRNA metabolism and as a method that can provide quantitative data for modeling complex gene-regulatory systems.

15.1   Non-perturbing RNA labeling in yeast

The nucleoside analog 4sU is readily taken up by cells of a broad range of eukaryotic organisms and is efficiently incorporated into their newly transcribed RNA (Section (4.1)). This can be used to metabolically label and isolate newly transcribed RNA from total cellular RNA with high specificity [102, 52]. To establish 4sU labeling in the budding yeast *S.cerevisiae*, we cultured cells in the presence

of $100\mu M - 5mM$ 4sU. Although we observed concentration dependent, specific incorporation of 4sU, the efficiency of incorporation was low and the amount of recovered newly transcribed RNA was very small (not shown). This hinted at inefficient uptake of 4sU into yeast cells rather than an intracellular block in activation or incorporation by RNA polymerases.

In the fission yeast *Schizosaccharomyces pombe*, expression of the human equilibriative nucleoside transporter (hENT1) enables cellular uptake of the nucleoside analog 5-bromo-2'-deoxyuridine, resulting in labeling of DNA during replication [88]. To test whether this transporter could also mediate efficient uptake of 4sU in the budding yeast *S.cerevisiae*, and thus install efficient RNA labeling, we grew a BY4741 strain expressing hENT1 up to a logarithmic phase, added 4sU, and isolated RNA at several time points (Figure 7, Section 19.1). This significantly enhanced 4sU incorporation to a level similar to that generally achieved in mammalian cells, thereby facilitating efficient separation of total cellular RNA into newly transcribed and pre-existing RNA.

We next tested whether Pol II incorporates the thionucleotide normally into RNA *in vitro* [25, 189]. Pol II used the substrates UTP and 4sU-triposphate (4sUTP) with very similar kinetics (Supplementary Figure S1 [136]). Whereas k_{cat} was unchanged, K_M increased from 3 nM for UTP to 13 nM for 4sUTP, indicating a slightly decreased substrate affinity that may result from weaker base pairing between 4sUTP and the template (Figure 6). This minor difference is likely irrelevant *in vivo*, where substrate concentration is higher by several orders of magnitude than these K_M values. To investigate whether RNA labeling perturbed gene expression *in vivo*, we compared RNA levels in 4sU-treated hENT1-expressing cells with untreated wild-type cells (Section 19.1). For a labeling period of 6 min, there were no significant changes in RNA levels as measured with Affymetrix expression arrays (Figure 36, Supplementary Section 12.6, Supplementary Table 4, Supplementary Figures S11, S12 [136]). Thus, RNA labeling perturbs neither transcription nor the yeast transcriptome. Although other cellular processes may be influenced by 4sU, their effect on mRNA metabolism is apparently not significant, as changes in the total mRNA levels were not observed.
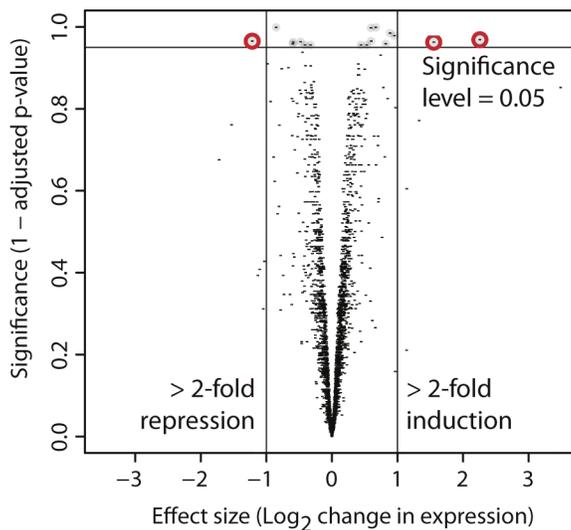
Figure 36: The yeast transcriptome is undisturbed by expression of the human nucleoside transporter hENT1. The volcano plot compares mRNA levels after 6 min labeling versus wild-type cells without labeling. Each dot corresponds to one gene, the x axis displays the $\log_2$(fold) of that gene, the y axis represents the multiple testing adjusted P-value (see Section 18.3). In all, 17 genes showed a significant change in mRNA levels (adjusted P-value 5%), only three of which were at least twofold. The results for the other tested labeling periods are given in the Supplementary Section 12.6.[136]

15.2 Dynamic transcriptome analysis

To determine the optimum labeling time, we purified total, newly transcribed (labeled), and pre-existing (unlabeled) RNA at 3, 6, 12, and 24 min after 4sU addition, and subjected these fractions to expression array analysis (Section 19.1). Replicate data always showed correlations above 0.9 for each RNA fraction at each time point (Supplementary Figure S2 [136]). To estimate mRNA synthesis and decay rates from individual time point measurements, we developed a new quantitative steady-state model (Section II).

Reproducibility assessment of the data (Supplementary Section 12.6 [136]) and simulation studies (Supplementary Section 13.4 [136]) suggested an optimum labeling time of 6 min, which was subsequently used in all experiments. This was short enough to meet the assumption of constant synthesis and decay rates during labeling, but sufficiently long to yield enough labeled RNA for robust measurements. The relative decay rates within an experiment can be estimated reliably, but the absolute values are more difficult to obtain (Supplementary Section 13.5 [136], Section 18.4). To investigate this, we conducted a classical decay time course experiment using transcription inhibition and RT–PCR, and obtained generally consistent decay rates (Supplementary Figure S3 [136]).
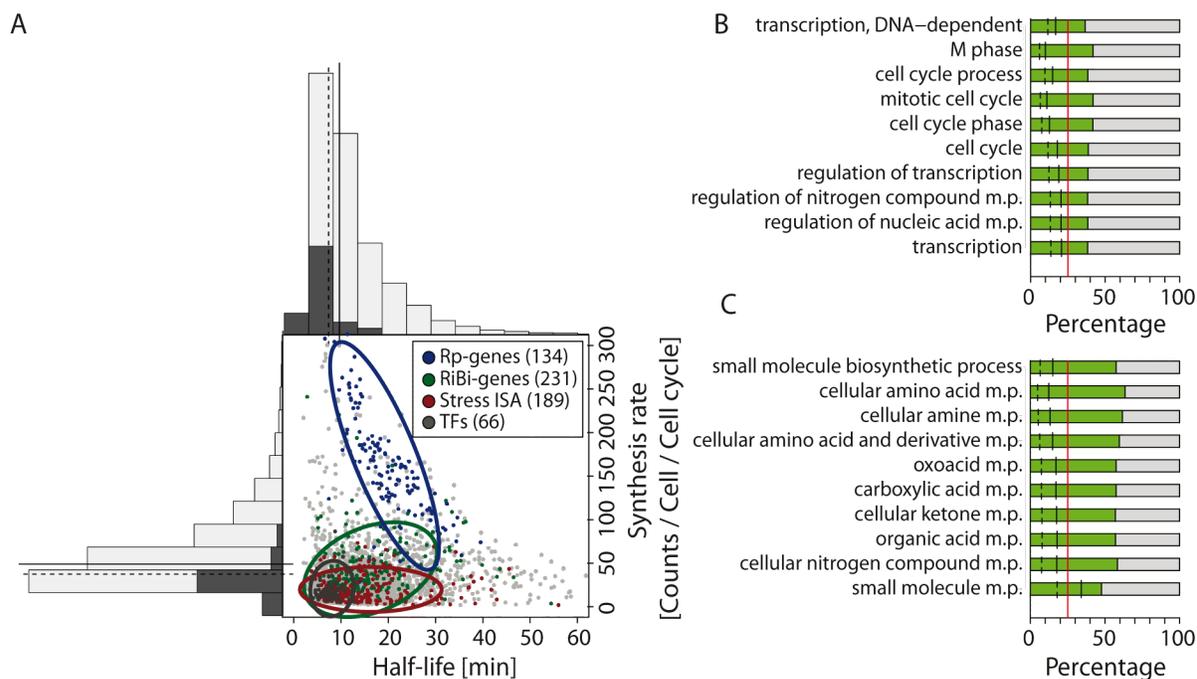
Figure 37: (A) Scatter plot of the mRNA half-lives $t_{1/2_g}$ and synthesis rates μ_g for exponentially growing yeast cells. Colored points belong to the indicated gene sets (green, ribosomal biogenesis genes; violet, ribosomal protein genes; red, stress genes; dark gray, transcription factors (TFs). Assuming Gaussian distributions, ellipses show the 75% regions of highest density for the respective sets. Overall half-lives and synthesis rates are uncorrelated (Spearman correlation 0.06), however some gene groups behave differently (correlations: Ribosomal protein genes (Rp) 0.79, Ribosomal biogenesis genes (RiBi) 0.35, ISA stress module genes 0, TFs 0.07). (B) Gene Ontology (GO) analysis of the short-lived mRNAs (lower 25% of the half-life distribution). The 10 most significant categories are displayed, sorted from bottom (most significant) to top (Materials and methods). Red line, proportion of short-lived transcripts in the whole population (25% by construction). The number of short-lived transcripts in the resp. GO category is given relative to the GO category size (green bar) and relative to the number of short-lived transcripts (black line). Dashed line, relative size of the GO set in the whole population. (C) Gene Ontology (GO) analysis of the long-lived mRNAs (upper 25% of the half-life distribution), analogous to (B)

15.3 Synthesis rates are low for most mRNAs

We used DTA to derive synthesis rates and decay rates (half-lives) for most (4508) of the yeast mRNAs (Section II). On the basis of a published rough estimate of 15000 mRNA transcripts per yeast cell [87], we calculated the synthesis rate as the number of mRNA molecules produced per cell per cell cycle time (150 min) (Figure 37). The obtained rates correlated with previously reported rates obtained by nuclear run-on [155] (Supplementary Figure S4 [136]). Synthesis rates ranged from 1 to ~ 600 mRNAs per cell per cell cycle time. The synthesis rate distribution is strongly right skewed (skewness 5), with a median synthesis rate of 18 RNAs per cell and cell cycle time (mean 31, 1st quartile 11, and 3rd quartile 33). This shows that only a few copies are made for most mRNAs (Figure 37A). This observation is generally consistent with single molecule live-cell imaging [151]. We observed that mRNAs with high synthesis rates encoded ribosomal protein genes and genes involved in ribosome biogenesis, whereas mRNAs with low synthesis rates stemmed from genes that are silenced during normal growth, including most TFs (Figure 37).

15.4 mRNA decay is not correlated with synthesis

DTA measured a median mRNA half-life of 11 min (mean 14, 1st quartile 9, and 3rd quartile 17 min, Figure 37). The half-life distribution is strongly right skewed (skewness 8). Thus, most mRNAs in yeast are synthesized and degraded several times during a cell cycle time. Gene ontology (GO) analysis revealed that mRNAs with the shortest half-lives are involved in the regulation of transcription and the cell cycle, and mRNA processing (Figure 37B). In contrast, mRNAs with long half-lives are involved in carbon and nitrogen metabolism and include many transcripts encoding housekeeping enzymes (Figure 37C). The decay rates did not correlate with published rates [90, 205, 77, 176, 53] (see also Section 16.3), which were obtained with protocols that perturb mRNA metabolism (Section 3). The decay rates did not correlate with mRNA length (Supplementary Figure S6 [136]), inconsistent with models that assume stochastic degradation, but consistent with degradation control at the level of mRNA deadenylation and decapping. Many mRNAs with long half-lives contained AU-rich elements in their 3'-untranslated region, consistent with a stabilizing role of these elements [11]. Decay rates correlated weakly with mRNA levels (Spearman correlation 0.59), but synthesis rates correlated well with mRNA levels (Spearman correlation 0.84). However, decay rates did not correlate with synthesis rates (Spearman correlation 0.15). This indicates that mRNA decay and synthesis are functionally independent during normal growth, and that both processes contribute to setting cellular mRNA levels.

15.5 DTA monitors rate changes during osmotic stress

The above analysis and published studies estimate mRNA synthesis and decay rates only in the steady state [52, 6]. To monitor rate changes, and thus the dynamics in mRNA metabolism, we extended DTA to a time-resolved analysis of the osmotic stress response (Section II). Cells were grown up to a logarithmic phase and split into control and sample cultures (Figure 38A). Sodium chloride was added to the sample culture and made up to a concentration of 0.8 M. Control and sample cultures were divided in aliquots, and 4sU was added at 0, 6, 12, 18, 24, and 30 min after salt addition. After labeling for 6 min, total, labeled, and unlabeled RNA was purified and analyzed with gene expression arrays. The rates were estimated using DTA within the time windows $0 - 6$, $6 - 12$, $12 - 18$, $18 - 24$, $24 - 30$, and $30 - 36$ min after stress induction. The results were confirmed for selected genes by quantitative RT–PCR at 12, 30 and 36 min after salt stress, wherein strong rate changes were observed (Supplementary Figure S3 [136]).
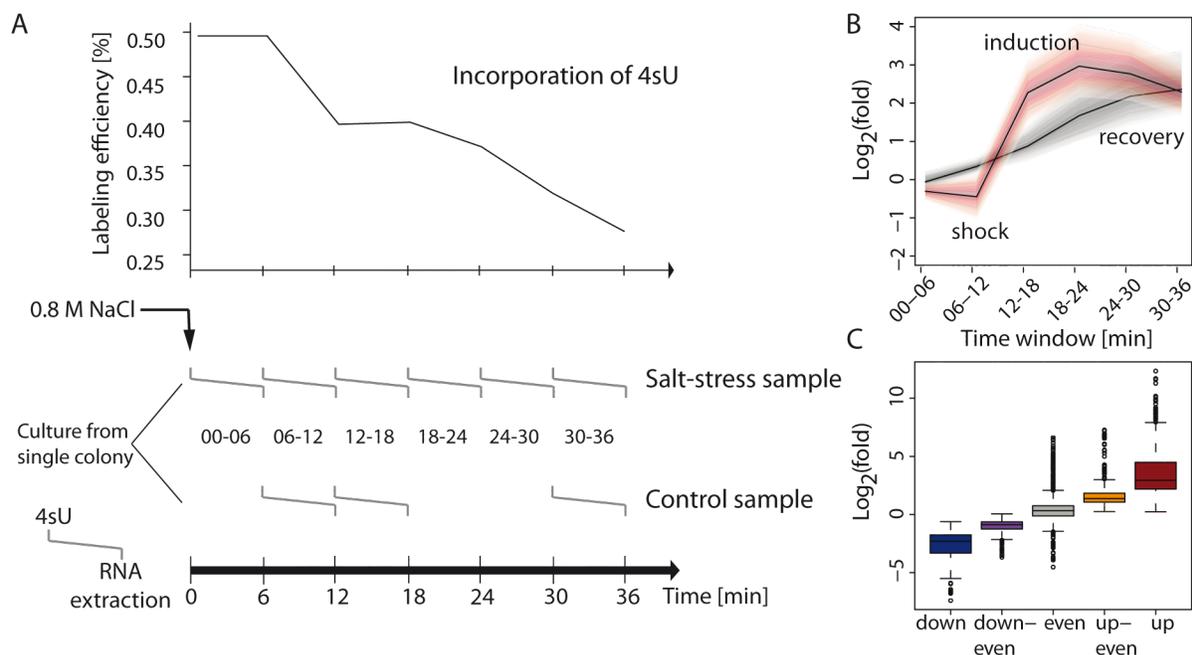
Figure 38: (A) Design of the time series experiment. Each time window corresponds to one sample, left end marks the start of the 4sU labeling, right end marks the time of mRNA extraction. Upper panel shows the drop in labeling efficiency from roughly one 4sU in 200 uracils to one 4sU in 400 uracils during the osmotic stress response. (B) Increased sensitivity and temporal resolution of DTA compared with standard transcriptomics. Gray, time course of the total mRNA fraction of the Hog1-induced genes [32]. Red, time course of the synthesis rates of the same gene set. The solid lines represent the time course of the median, the shaded bands are the central 95% regions. In contrast to the monotonically increasing total mRNA time course, the synthesis rates clearly show three response phases. (C) Expression changes of the five clusters (up, up-even, even, down-even, down see Materials and methods) that were defined in a normalization-independent manner. The box plot shows synthesis rate folds (30 versus 0 min).

15.6 Three phases of the osmotic stress response

DTA resolved the three phases of the osmotic stress response with unprecedented clarity. In the first 12 min after salt addition (shock phase), essentially all synthesis and decay rates decreased, reflecting global transcription down-regulation and mRNA stabilization. Within $12 - 24$ min after salt addition (induction phase), synthesis rates strongly increased for a subset of mRNAs. These stress-induced mRNAs show increased decay rates, likely to ensure their rapid removal toward the end of the response. Finally, decay rates were mostly restored, whereas a fraction of the synthesis rates remained at levels distinct from the starting values (recovery phase). We could not monitor complete recovery, which takes about 2 h [124], but a fraction of synthesis rates apparently remains at values different from the starting values, to ensure continued expression of salt homeostasis genes, and lower expression of housekeeping genes. DTA also revealed a drop of labeling efficiency from 0.5% (1 4sU in ~ 200 nucleotides) to 0.27% (Figure 38A), reflecting the known inhibition of cellular uptake of small molecules during stress.

15.7 Temporary correlation of mRNA synthesis and decay rates

We transformed all rates to their ranks within the rate distributions, to circumvent an error-prone estimation of an unknown normalization factor between measurements at different time points. By comparing the ranks of synthesis rates in the datasets, 6 and 36 min after salt addition, five clusters

of genes were defined (Figure 38C): *'up'* (217 genes, rank gain > 2000),*'up-even'* (456 genes, rank gain $1000 - 2000$), *'down-even'* (498 genes, rank loss $1000 - 2000$), *'down'* (401 genes, rank loss > 2000), and *'even'* (all remaining 2936 genes). Although global mRNA synthesis and decay were not correlated before stress, some gene groups showed positive and negative correlations during stress (Figure 37). An analysis of the changes in synthesis and decay rates reveals a temporary interdependence of the rate changes of mRNA synthesis and decay during the first two stress phases (Figure 39, Supplementary movie [136]). During the shock phase, a decrease in synthesis rate generally goes along with a decrease in decay rate. During the induction phase, an increase in synthesis rates generally goes along with an increase in decay rate. They again become uncorrelated during recovery. The nature of a possible physical coupling underlying this temporary correlation of rates remains to be explored.
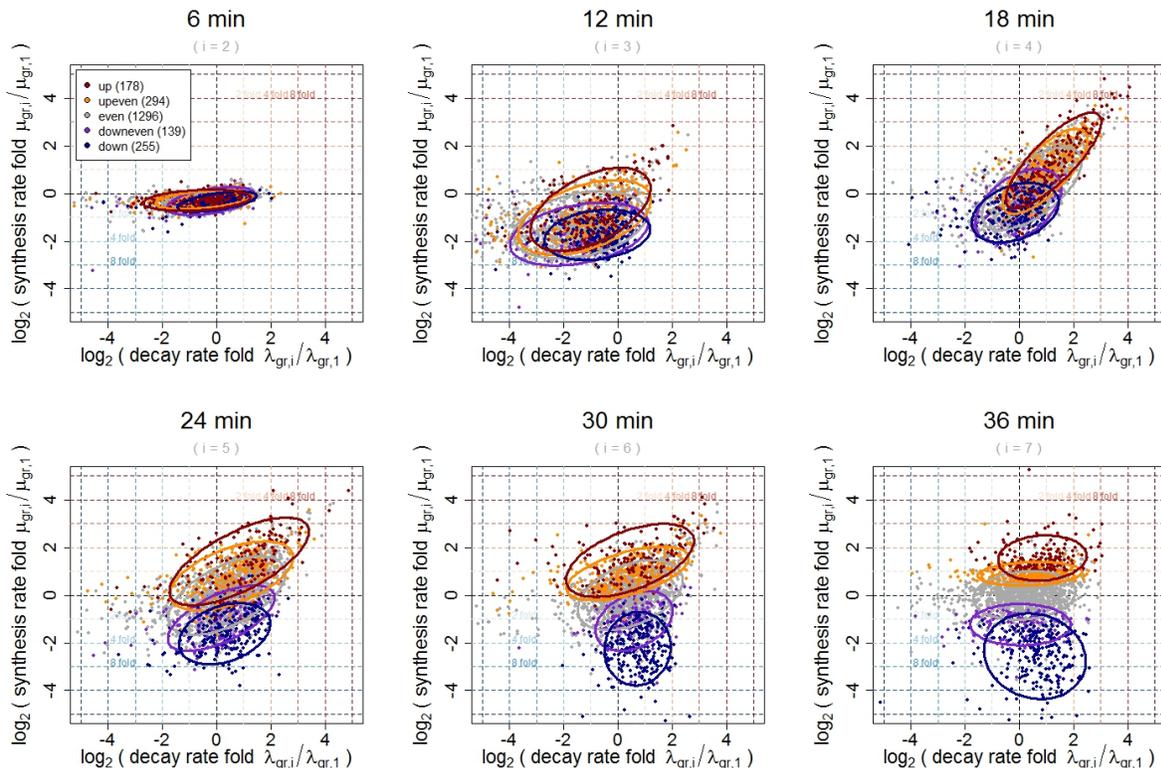
Figure 39: Dynamics of synthesis and decay rates in the osmotic stress time series. Each plot corresponds to one timepoint. Log$_2$(decay rate fold $\lambda_{g,i}/\lambda_{g,1}$) for $i = 2, ..., 7$ versus log$_2$(synthesis rate fold $\mu_{g,i}/\mu_{g,1}$) for $i = 2, ..., 7$ for the last timepoint compared to the first timepoint. Each point corresponds to one gene, which is colored according to its affiliation with one of the clusters in Figure 38. Ellipses show the 75% regions of highest density within each cluster, assuming Gaussian distributions. The shape of the ellipses indicates the correlation structure within a cluster.

15.8   High temporal resolution reveals mRNA dynamics

Resolution of the three phases of stress response depended on DTA and was not possible by measuring total mRNA levels alone (Figure 38B). To test the performance of DTA with an unbiased gene set, we monitored the previously described 305 Hog1-responsive genes [32]. DTA detected an initial decrease in synthesis rates during shock, whereas total RNA levels increased (Figure 38B). This was, however, not due to increased transcriptional activity, but rather due to residual transcription activity combined with mRNA stabilization (Figure 39). Also, the signal-to-noise ratio in detection of changes in synthesis rates was on average two times higher than that of measuring differences

in total mRNA levels (Supplementary Figure S7 [136]). Thus conventional transcriptomics fails to unveil the nature of the changes in mRNA metabolism upon stress, which are however monitored by DTA.

15.9   High sensitivity reveals new stress response genes

Owing to the increased sensitivity, DTA reveals many genes that are induced during stress. The up cluster contained genes associated with GO terms related to stress response. Of the stress module genes as defined by the Iterative Structure Algorithm [97], 74% showed a rank gain greater than 1000. The up cluster contained only three TFs, consistent with the pre-existence of TFs for stress response and their post-transcriptional activation ([157], and references therein). The up cluster contained 62% of the genes that were considered significantly up-regulated by a factor of at least two in a recent study of the osmotic stress response [32]. However, DTA also detected 58 new genes involved in the osmotic stress response (Supplementary Figure S8 [136]). These are mostly genes of unknown function, except *Ubc5*, which is known to mediate degradation of abnormal proteins during cellular stress. Of genes in the up cluster, 35% were uncharacterized, compared with only 16% over all yeast genes. Yeast strains with single knockouts of the newly revealed stress genes did not generally show growth defects under high salt conditions (Supplements, Part IV [136]), providing a possible explanation for why they were not discovered previously.
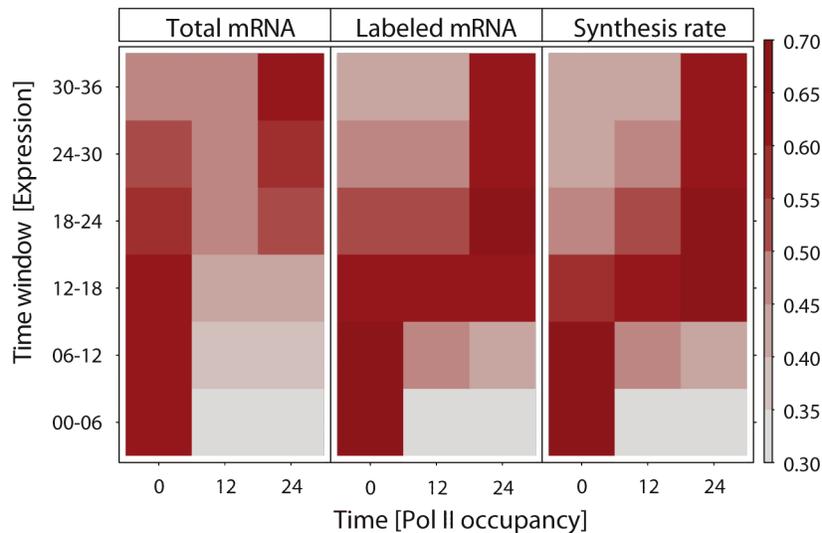
Figure 40: Pol II gene occupancy predicts mRNA synthesis. The vectors of mean Pol II occupancies on transcribed regions were calculated from ChIP-chip data at 0, 12, and 24 min after salt stress and compared with the vectors of total mRNA levels (left), labeled mRNA (middle), and synthesis rates (right) at each time point of the osmotic stress time course experiment. The pair-wise Spearman correlation values are represented by color-coded squares.

15.10   Genomic Pol II redistribution predicts mRNA synthesis rate changes

To investigate whether mRNA synthesis rates correlate with the presence of Pol II at transcribed genes, we determined occupancy profiles for the Pol II subunit Rpb3 by chromatin immunoprecipitation (ChIP) and tiling microarray (chip) analysis, and calculated the mean Pol II occupancy between the transcription start site (TSS) and the polyadenylation site (pA) for each gene [129]. We also measured ChIP-chip profiles 12 and 24 min after salt addition, to investigate whether Pol II is redistributed over the genome upon stress. At all three time points (0, 12, and 24 min), the mean
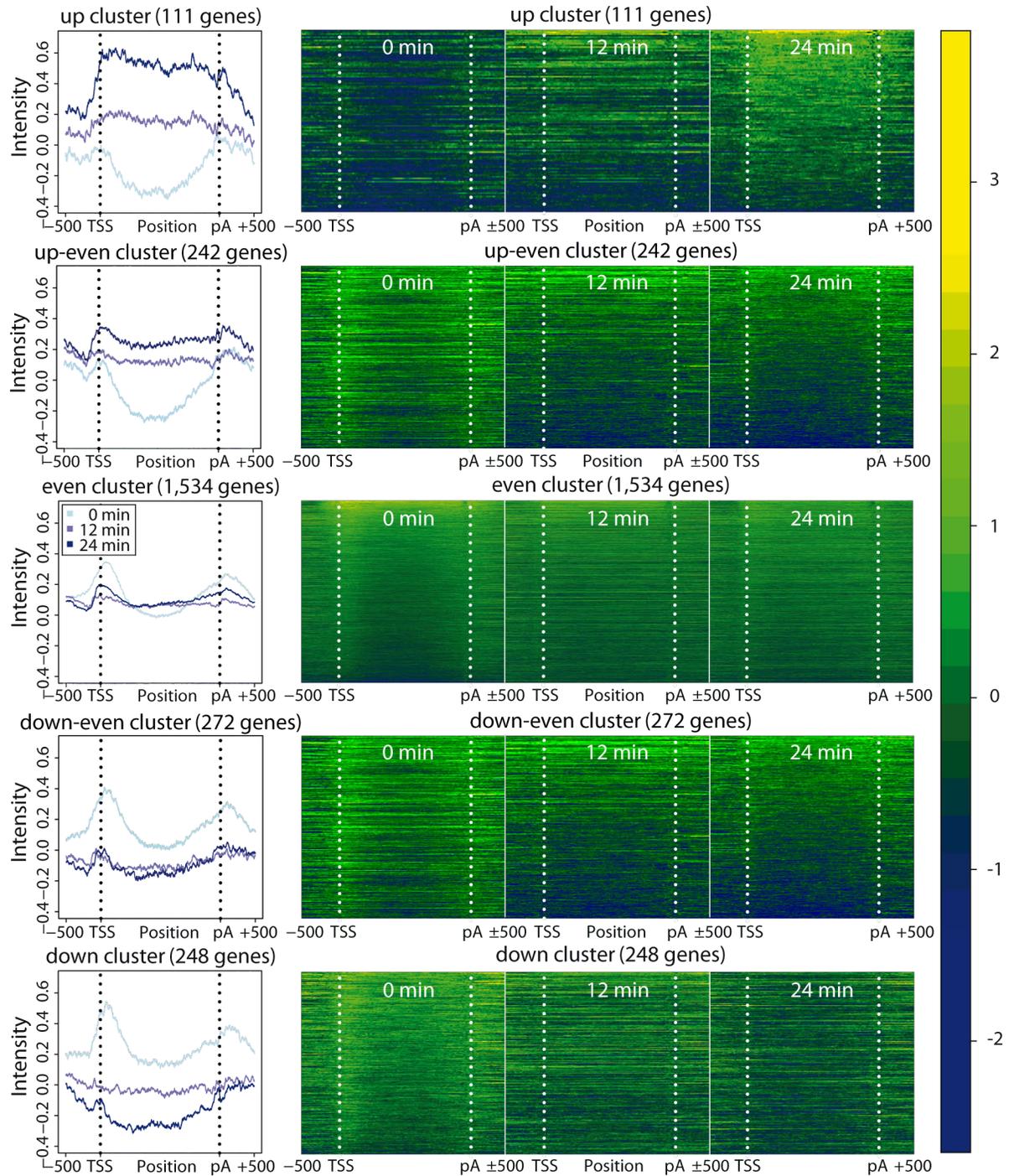
Figure 41: Dynamics of the mean Pol II occupancy profiles of all selected clusters (Figure 38, Supplementary Section 17 [136]). Only genes of at least 1000 base pairs (bps) are considered. Left: Profiles are taken after 0, 12, and 24 min of osmotic stress (light blue, blue, and dark blue lines). Vertical dotted lines are drawn at the transcription start site (TSS) and the polyadenylation site (pA) site. The plot shows the ±500 bp regions around the TSS and the pA sites, together with the intermediate part, which has been rescaled to a length of 400 (Supplementary Section 16 [136]). Right: Heatmaps of the Pol II profiles for all cluster at 0, 12, and 24 min. Each row corresponds to one gene. The vertical dotted lines mark TSS and pA of each gene. Pol II occupancy from low to high is coded with colors ranging from dark to bright.

Pol II gene occupancy was calculated. The three resulting Pol II occupancy vectors were compared with the vectors of total RNA, newly synthesized RNA, and synthesis rates at all six 6-min time windows of the osmotic stress (Figure 40).

Pol II gene occupancies at 0, 12 and 24 min correlated only weakly with mRNA levels, but very well with the levels of labeled mRNA and with the synthesis rates at the corresponding time points (Figure 40). The results also demonstrated the inferior temporal resolution of standard transcriptomics, as Pol II occupancy 12 and 24 min after stress induction correlated with mRNA levels at a later time point (Figure 40).We averaged Pol II occupancy profiles over genes belonging to the *even*, *down*, and *up* clusters (Figure 41). The *even* cluster showed a typical gene-averaged profile with elevated Pol II levels on the transcribed region and peaks around the TSS and pA site. This profile persisted during stress, although overall polymerase levels decreased. The *down* cluster genes apparently lost most if not all Pol II during stress. In contrast, the *up* cluster genes did not contain detectable amounts of Pol II before stress but gained Pol II during stress. The shape of the averaged profile of *up* cluster genes after 24 min of salt stress showed an even distribution of Pol II that was very different from the canonical profile (Figure 41), maybe because of a high density of Pol II on these stress-induced genes. Thus, Pol II occupancy predicted mRNA synthesis rates and Pol II redistribution upon stress predicted changes in synthesis rates. On the other hand, the observed correlations confirm that DTA realistically monitors changing transcriptional activity.

15.11   Summary & Outlook

In this study, we develop DTA as a tool for measuring mRNA synthesis and decay rates in yeast on a global scale and in a dynamic manner without system perturbation. DTA involves genetically facilitated cellular uptake of the nucleoside analog 4sU, metabolic RNA labeling, separation and microarray analysis of three different RNA fractions (total, newly transcribed, and pre-existing RNA). A quantitative dynamic model then enables extraction of synthesis and decay rates from the array data. As we select for polyadenylated RNA in the protocol, and the probes on the array are predominantly located in the 3' region of transcripts, only complete transcripts are detected and the obtained synthesis rates reflect the production of complete mRNAs. The application of DTA to growing yeast cells revealed that most genes produce only a few copies of mRNA per cell per cell cycle time, and that transcript turnover is generally rapid, with a median mRNA half-life of 11 min, as well as that synthesis and decay are generally not correlated.

DTA was used to follow dynamic changes in mRNA synthesis and decay rates, as exemplified by the osmotic stress response. This showed that DTA has higher sensitivity and temporal resolution than ordinary transcriptomics, and thus provides new biological insights. DTA identified 58 new genes induced by osmotic stress, including many genes of unknown function. DTA also revealed a correlated alteration of mRNA synthesis and decay during the first two phases of the stress response. In the initial shock phase, transcription is globally slowed down and mRNAs are stabilized, apparently to store them. During the subsequent induction phase, synthesis rates for a subset of mRNAs are strongly increased, and the resulting mRNAs are destabilized, thus allowing for their rapid removal after stress. During recovery, decay rates are restored whereas synthesis rates may remain at altered values.

More generally, altered gene expression programs resulting from cellular adaptation may involve persistently altered mRNA synthesis rates, whereas mRNA decay rates may be similar for different programs. Indeed, theoretical considerations show that changes in synthesis rate can more quickly change mRNA levels for transcripts with low abundance, which dominate in the cell [3]. Changes in decay rate can only efficiently change mRNA levels for highly abundant transcripts (Section 12). This concept is realized in the osmotic stress response, and likely in other gene regulatory systems. Many of our observations are consistent with published data on the osmotic stress response. Genomic run-on analysis showed that osmotic stress provokes mRNA stabilization and sequestration

into P-bodies [166]. P-bodies develop at 0.8 M NaCl within minutes [191]. Global transcription inhibition showed that repressed mRNAs are stabilized, whereas induced mRNAs are destabilized [139]. The rapid decrease of mRNA synthesis and decay rates during the shock phase is pre-transcriptional as suggested [157, 134]. This may go along with a transient depression of translationally active ribosomes [206]. Part of the shock response may stem from changing the post-translational modification of pre-existing proteins, as 15% of protein phosphorylation sites change within 5 min after stress [183]. Only during the induction phase, an increase in abundance of a subset of proteins occurs that correlates with an increase in mRNA levels (Supplementary Figure S27 [136]). A correlated response of mRNA levels and the translatome is also revealed by global analysis of ribosome associated mRNAs [80]. In the future, insights from DTA and Pol II profiling may be used to improve models of the osmotic stress response [32, 142].

The mRNA synthesis rates obtained by DTA are robust and realistic, as they correlate with Pol II occupancy of transcribed regions, as measured here by ChIP-chip profiling. DTA and Pol II ChIP-chip apparently monitor the same biological process, transcriptional activity, in contrast to conventional transcriptomics, which monitors mRNA levels. Thus, previously obtained correlations of Pol II occupancy with mRNA levels [8, 185, 154, 201, 129, 165] are only an indirect effect of the correlation of mRNA synthesis rates with mRNA levels as reported here. DTA-derived rates of mRNA synthesis are averaged over a cell population and a time period of 6 min, suggesting that they are independent of the nature of transcription, which may occur in bursts and discontinuously. DTA has a great potential for studying gene regulatory systems and mRNA metabolism. It may be used to characterize gene expression programs under defined conditions, and to analyze the influence of changes in environmental conditions or the effect of small molecules on mRNA turnover. It may also be used to study the global mechanisms of gene regulation by introducing mutations that interfere with particular aspects of transcription and its regulation. DTA may further be used to decipher the pathways of mRNA decay and the global regulation of mRNA half-lives by RNA-binding proteins. Such studies however require genetic perturbation and, therefore, new tools for the normalization of rates extracted from array measurements with mutant and wild-type yeast strains.

16 cDTA reveals mutual feedback between mRNA synthesis and degradation

Cellular gene expression is controlled by mRNA levels, which are governed by the rates of nuclear mRNA synthesis and cytoplasmic mRNA degradation. The rates of mRNA synthesis are regulated during RNA polymerase (Pol) II transcription in the nucleus (Section 1.1), whereas bulk mRNA degradation occurs in the cytoplasm (Section 1.2). During transcription, the mRNA receives a 5'-cap and a 3'-poly(A) tail. The mature mRNA is then exported to the cytoplasm, translated, and eventually degraded co-translationally [93]. Cytoplasmic mRNA degradation generally begins with shortening of the poly(A) tail by the Ccr4-Not complex that contains the deadenylases Ccr4 and Pop2 (also known as Caf1) [119, 195]. The mRNA is then decapped and degraded by exonucleases from both ends. Despite the spatial separation of mRNA synthesis and translation/degradation, there is evidence that these processes are coordinated [122, 123, 82].

To investigate coordinated RNA synthesis and degradation, absolute changes in synthesis and decay rates must be measured after introducing a genetic perturbation that impairs either synthesis or degradation. Rates of mRNA synthesis and degradation can be measured by Dynamic Transcriptome Analysis (DTA) in yeast (Section 15). Newly synthesized RNA is labeled with 4-thiouridine (4sU), which is taken up by cells that express a nucleoside transporter. After 6 minutes of labeling, total RNA is extracted and separated into newly synthesized (labeled) and pre-existing (unlabeled) fractions. Total, labeled, and unlabeled fractions are analyzed with microarrays and the data are fitted with a dynamic kinetic model to extract synthesis and decay rates (Section II). Whereas DTA accurately measures the relative rates for different RNAs within a single sample, it cannot compare rates from different samples, since the samples differ by an unknown global factor (Section 18.4). In standard transcriptomics, comparison between samples with different mRNA levels may be achieved by counting cells and spiking RNA standards into the samples [90, 205, 198]. However, such normalization does not take into account differences in cell lysis and RNA extraction efficiency, which can vary so strongly that no conclusions are possible.

To enable normalization between DTA measurements of different samples, we extended DTA to comparative DTA (cDTA, Section 4.3). In cDTA, a defined number of labeled fission yeast *Schizosaccharomyces pombe* (*Sp*) cells is added to the budding yeast *Saccharomyces cerevisiae* (*Sc*) sample before cell lysis and RNA preparation, and is used as an internal standard. Thereby, cDTA allows the absolute quantification and accurate comparison of mRNA synthesis and decay rates between samples. cDTA is a novel method that monitors absolute changes in eukaryotic mRNA metabolism upon genetic perturbation. We applied cDTA to *Sc* cells that are impaired in either mRNA synthesis or degradation. This revealed compensatory changes in degradation and synthesis, respectively, which indicates that a eukaryote can buffer mRNA levels to render gene expression robust. After our work was completed, an independent study appeared that postulates a similar compensation on an evolutionary scale [53].

16.1 Comparative Dynamic Transcriptome Analysis (cDTA)

To measure changes in mRNA synthesis and decay rates between different strains of budding yeast (*Sc*), we included the distantly related fission yeast (*Sp*) in our DTA protocol as an internal standard (Figure 8, Section 4.3). We counted *Sc* sample cells and *Sp* control cells and mixed them in a defined ratio (Section 20.2). The resulting cell mixture was lysed, total mRNA extracted, labeled RNA purified, and microarrays were hybridized as described in Section 19.1. The RNA mixture was quantified on a microarray that contains probes for both *Sc* and *Sp* transcripts (Affymetrix® GeneChip® Yeast Genome 2.0 Array) [136]. We used 4-thiouracil (4tU) instead of 4sU for *Sc* RNA labeling, because it is taken up by *Sc* [141] without expression of a nucleoside transporter [136]. 4tU labeling did not affect normal cell physiology (Supplementary Figures S1 [187]) and allowed growth

of yeast in YPD instead of selective medium. We quantified only labeled and total RNA, because the unlabeled fraction was not required for rate extraction.
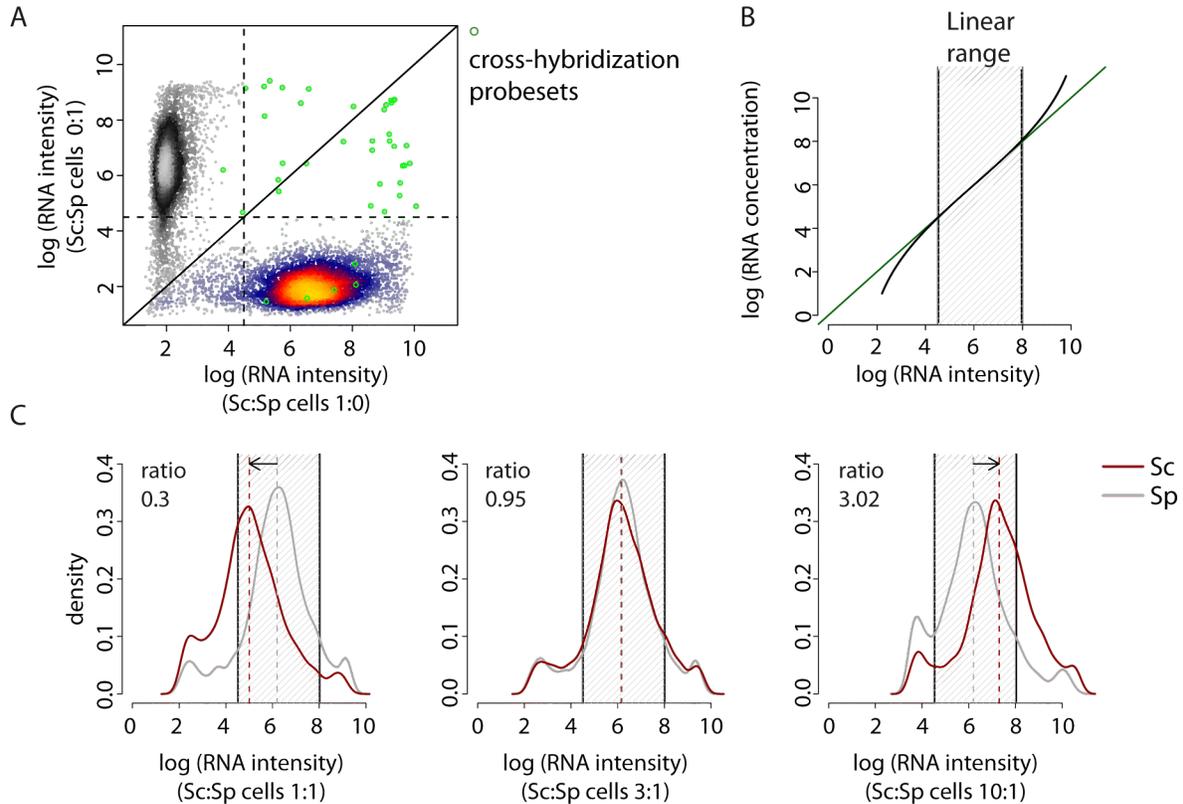

Figure 42: (A) Assessment of cross-hybridization. Scatterplot of log intensities of $10,928$ Affymetrix probe sets. The values on the x- resp. y-axis are obtained as the mean of two pure Sc resp. Sp replicate samples that were hybridized to the arrays. Sc and Sp probe sets (heat-colored and grey-scaled, respectively) can be separated almost perfectly. 23 out of $5,771$ Sc probe sets show intensities above a (log) background intensity threshold of 4.5 in the Sp sample, whereas 8 out of $5,028$ Sp probe sets were above background in the Sc sample. These 31 probe sets are regarded as affected by cross-hybridization (green circles). Of these, 16 probe sets were excluded from analysis because all probes were affected by cross-hybridization (Methods [187]). (B) Linear measurement range. Exemplary illustration showing that the relation of mRNA concentration (real amount) and mRNA intensity (fluorescent scanner readout) follows the Langmuir adsorption model [84, 85, 86, 180]. The green line indicates linearity. The black line shows sigmoidal behavior, resulting from noise at low hybridization levels and saturation effects at high hybridization levels. The grey stripe indicates the linear measurement range that we defined as an intensity range of $4.5-8$ (natural log basis) based on noise signals below 4.5, for example for probes that detect transcripts of genes that were knocked out, and based on observed saturation effects above 8. (C) Calibration of Sc:Sp cell mixture ratio. The optimal cell mixture ratio has been chosen to maximize the number of probes for both Sc and Sp that fall into the linear measurement range (B). Sc and Sp cells were mixed in Sc:Sp ratios of 1:1, 3:1, and 10:1. The respective median mRNA level ratios are 0.3, 0.95, and 3.02. Log(RNA intensity) distributions of Sc (red) and Sp (grey) are shown. The median intensity level of Sp is approximately three times higher than that of Sc. As a consequence, a Sc:Sp cell mixture ratio of 3:1 was used. (D) Comparison of the three different cell mixtures of (C) in pairwise log-log scatter plots. All arrays are normalized to a common median of $4,052$ Sp probe sets (grey-scaled). $4,475$ Sc probe sets (those in the linear measurement range) are shown in heat colors. The parallel offset of the Sc probe sets from the main diagonal measures the mRNA level differences of Sc in the three cell mixtures. The differences should be 3.3, 10, and 3 when we plot Sc:Sp ratios of 10:1 vs. 3:1, 10:1 vs. 1:1, and 3:1 vs. 1:1, respectively. The corresponding measured offsets are 3.14, 9.46, and 3.01, and thus in very good agreement.

We first tested whether the Sc sample showed cross-hybridization to Sp array probes and vice versa. When either a Sc or Sp sample was hybridized to the array, cross-hybridization occurred for a minor fraction of the probes (Section 20.3) when a conservative intensity cut-off of 4.5 (log(intensity) values after pre-processing) was used (Figure 42A). Cross-hybridizing probes were excluded from further analysis, leading to loss of only 16 out of $10,799$ probe sets (Section 20.3). The mixing ratio of Sc:Sp cells was tuned to 3:1, to maximize the overlap of the Sc and Sp expression intensity distributions (Figure 42B). This ensures that after calibration most Sc and Sp probe intensities are in the linear measurement range of the microarray, an important prerequisite for our calculations. We restricted our analysis to RNAs with log(intensity) signals above 4.5 and below 8 (Figure 42B).
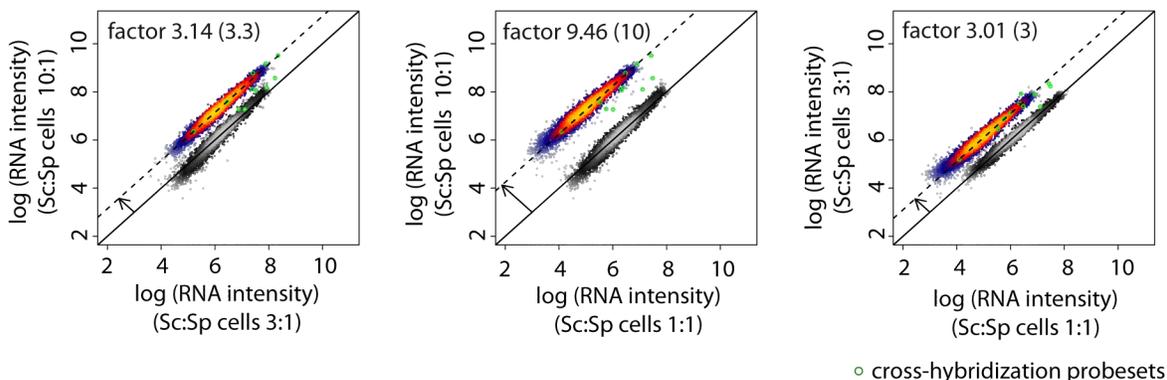

Figure 43: Comparison of the three different cell mixtures of Figure 42 in pairwise log-log scatter plots. All arrays are normalized to a common median of $4,052$ Sp probe sets (grey-scaled). $4,475$ Sc probe sets (those in the linear measurement range) are shown in heat colors. The parallel offset of the Sc probe sets from the main diagonal measures the mRNA level differences of Sc in the three cell mixtures. The differences should be 3.3, 10, and 3 when we plot Sc:Sp ratios of 10:1 vs. 3:1, 10:1 vs. 1:1, and 3:1 vs. 1:1, respectively. The corresponding measured offsets are 3.14, 9.46, and 3.01, and thus in very good agreement.

16.2   Rate extraction from cDTA data

To obtain absolute synthesis and decay rates for Sc and Sp, we derived the ratios of labeled to total RNA intensities c_{Sc} and c_{Sp} for Sc and Sp, respectively. These ratios set the global median level of synthesis and decay rates and rely on a robust previous estimate of the median Sc half-life [136] for which labeled, total, and unlabeled RNA fractions were available. Once c_{Sp} is known, the measured levels of the Sp standard can be used to calibrate the Sc data (Figure 44A). This new normalization method allows rate estimation from labeled and total quantities alone (Section 20.3). Our published median half-life for Sc mRNAs [136] enabled determination of the median Sp half-life relative to Sc (Supplementary Figures S2 [187]). We measured growth curves, and obtained a doubling time of 90 minutes for Sc in YPD medium at 30°C and 116 minutes for Sp in YES medium at 32°C (Supplementary Figures S3 [187]). These doubling times were used in kinetic modeling (Section II). We confirmed that the rates obtained by cDTA are essentially the same as the ones previously obtained by DTA (Table 1, Supplementary Figures S2 [187]). RNA half-lives that were recently determined by 4tU pulse-chase labeling in Sc are 1.5-fold longer [141], likely because a very long labeling time was used that allowed for thionucleotide re-incorporation after mRNA decay. We calculated mRNA synthesis rates as the number of complete transcripts made per cell and per 90 minutes (the cell cycle time for wild type Sc), using a new estimate of $60,000$ transcripts per yeast cell [219], instead of the previously used, older and four-fold lower estimate [87]. For Sp, we estimated the number of transcripts from the observed 2.51-fold cumulative total RNA level to be $150,801$. Our rate estimates are unaffected by the efficiency of 4tU labeling, which varies between
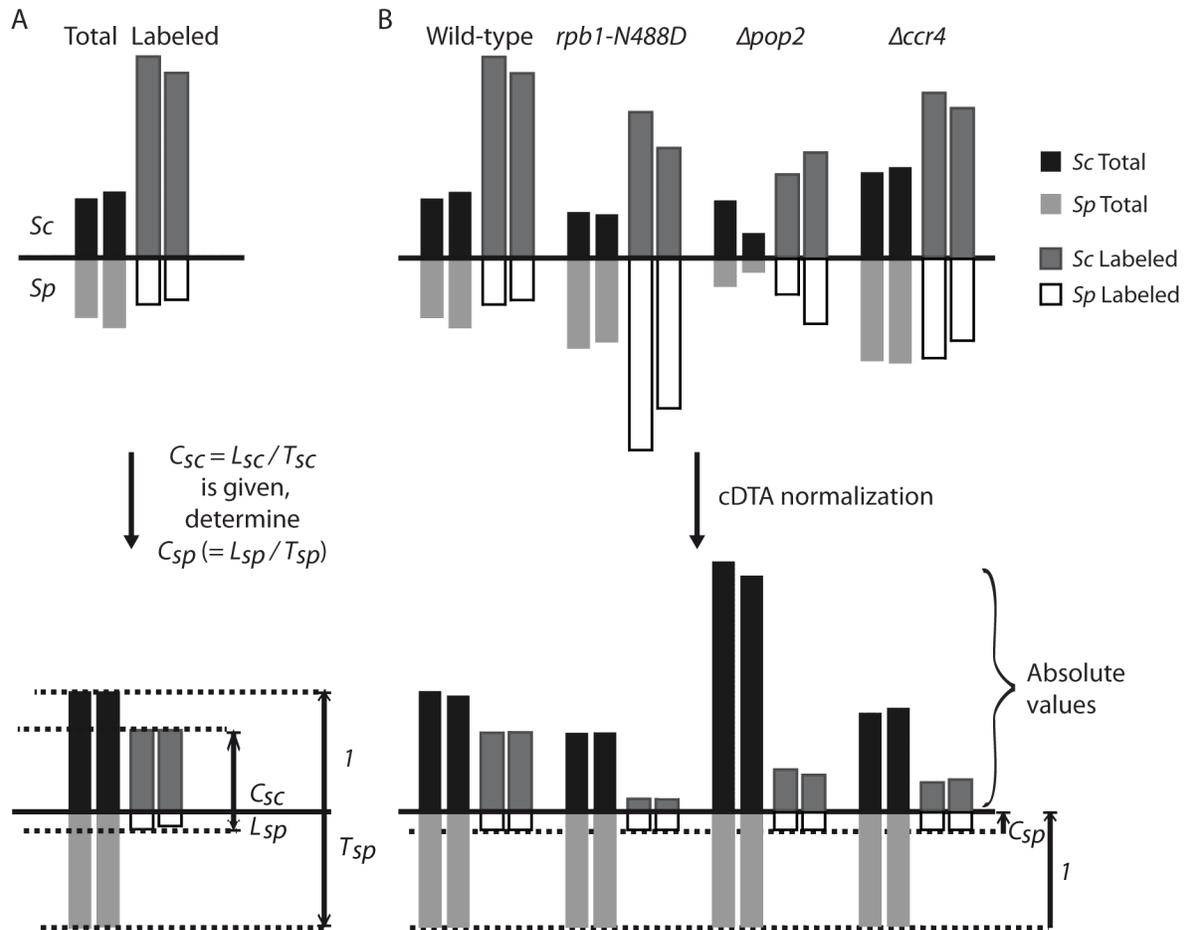
Figure 44: (A) Determination of c_{sp}, the ratio of labeled over total Sp mRNA. To obtain absolute synthesis and decay rates for Sc and Sp, we derived ratios of labeled to total RNA c_{sc} and c_{sp} for Sc and Sp, respectively. The c_{sc} ratio was obtained in our previous study (Section 15). To determine c_{sp}, L_{sc} and T_{sc} are set to c_{sc} and 1, respectively. L_{sp} and T_{sp} are then linearly rescaled. The resulting L_{sp}/T_{sp} is defined as c_{sp} and then used in the further experiments as global cDTA normalization factor. (B) cDTA normalization uses Sp signals as internal standard. The bars indicate the median intensities of the array probe sets. Due to our experimental design, the ratio of labeled to total Sp RNA ($c_{sp} = L_{sp}/T_{sp}$) must be the same in all experiments. To correct for differences in cell lysis, RNA extraction efficiency, and RNA purification efficiencies, the levels of Sp total and labeled mRNA are rescaled to the same values in all experiments. The Sc RNA levels are then corrected by median centering of Sp RNA levels. This normalization allows for a direct comparison of Sc data between experiments. Shown are both replicates for each of the four cDTA experiments.

strains and experiments (Supplementary Figures S1 [187]).

For normalization between different Sc samples, we linearly rescaled all array intensities such that the total and labeled Sp fractions have a median intensity of 1 or c_{Sp} (Figure 44B). We assessed the accuracy of the cDTA procedure by estimating the intensity ratios of Sc:Sp cells that were mixed at 1:1, 3:1 and 10:1. The correct values were recovered with an accuracy of 5% (Figure 42C, 43). Selected mRNA levels of the 1:1 and 10:1 ratio mixtures were additionally quantified by RT-qPCR (Section 20.4). The expected ratio of the four tested Sc transcripts was recovered within a relative error of 9% when normalized to two housekeeping Sp genes (not shown). In summary, cDTA normalization removes the major sources of experimental differences between samples in RNA labeling efficiency, cell lysis, RNA extraction, RNA biotinylation and labeled RNA purification, and array hybridization. cDTA detects global changes between Sc samples, in contrast to standard

normalization procedures that eliminate global changes because they assume constant median RNA levels.

	Species	cDTA	DTA
Median mRNA half-life [min]	Sc	12	11.5
	Sp	59	$N.A.$
Median mRNA synthesis rate [mRNAs/cell/cell cycle time] ^{a}	Sc	53^a	$18\,(72)^a$
	Sp	44	$N.A.$

Table 1: The cDTA contains the estimates obtained from using the labeled:total ratio of the complementary strain and the known total and labeled Sc:Sp ratios to calculate the missing labeled:total ratio, i.e. $L_{Sc}/T_{Sc} = L_{Sp}/T_{Sp} \cdot T_{Sp}/T_{Sc} \cdot L_{Sc}/L_{Sp}$. The DTA column shows the Sc half-life estimate obtained in Section 15. Note that the Sc estimates are virtually identical to ours, although we used a different labeling technique (4tU instead of 4sU) and had spiked-in Sp controls in the sample. ^{a}Please note we previously used in our calculations a total number of transcripts per cell of $15,000$, according to an old estimate [87], whereas we now used a recent estimate of $60,000$ [219]. If the same number of transcripts is used, the median synthesis rate obtained by DTA would be 72, comparable to our new estimate obtained by cDTA, despite the difference in media and cell cycle time (Section 19.1).

16.3   cDTA supersedes conventional methods

Conventional methods measure mRNA half-lives by inducing transcription arrest and following changes in mRNA levels over time. Transcription arrest has been achieved by adding the transcription inhibitor 1,10-phenanthroline [53], or by shifting the temperature-sensitive mutant strain *rpb1-1*, which carries a point mutations in the largest subunit of Pol II [148], to the restrictive temperature [90, 205, 77, 176] (Section 3). To investigate whether the latter method yields reliable data or whether it perturbs mRNA metabolism, we re-generated the *rpb1-1* strain and analyzed it with cDTA using published growth parameters [90] (Section 20.1). This revealed that mRNA synthesis rates were decreased globally by a factor of 2.7 already at the permissive temperature of 30°C (Figure 45). After 24 minutes at the restrictive temperature, mRNA synthesis rates had decreased further by a factor of 3.4, but recovered essentially to the rates measured at permissive temperature after 66 minutes (Figure 45).

These observations indicated that the mRNA metabolism in the *rpb1-1* strain is already perturbed at the permissive temperature, and that the temporary changes in mRNA metabolism observed at the restrictive temperature are mainly due to a heat shock response. To test this, we conducted a corresponding heat shock experiment on wild type cells. We analyzed the total mRNA from this experiment together with the data from the *rpb1-1* mutant by conventional decay time series analysis [90, 205, 77, 176] (Section 3). The obtained mRNA half-lives during heat shock correlated very well with data derived from the *rpb1-1* mutant strain, and with published half-lives obtained with this strain (Figure 46). The obtained half-lives were longer than the half-lives measured in unperturbed cells, likely because mRNA degradation was down-regulated during the stress response. There was also a good correlation with half-lives obtained after adding 1,10-phenanthroline, and even with our previous data obtained during the osmotic stress response (Section 15.5), if processed in the conventional way. This indicates that all these data are dominated by perturbations in mRNA metabolism that result from a general stress response. In contrast, published half-lives derived from metabolic RNA labeling [141] and our cDTA-derived half-lives do not correlate with data obtained by perturbing conventional methods. We conclude that conventional methods for estimating mRNA half-lives using the *rpb1-1* mutant strain or transcription inhibition cannot be used to obtain reliable half-life estimations.
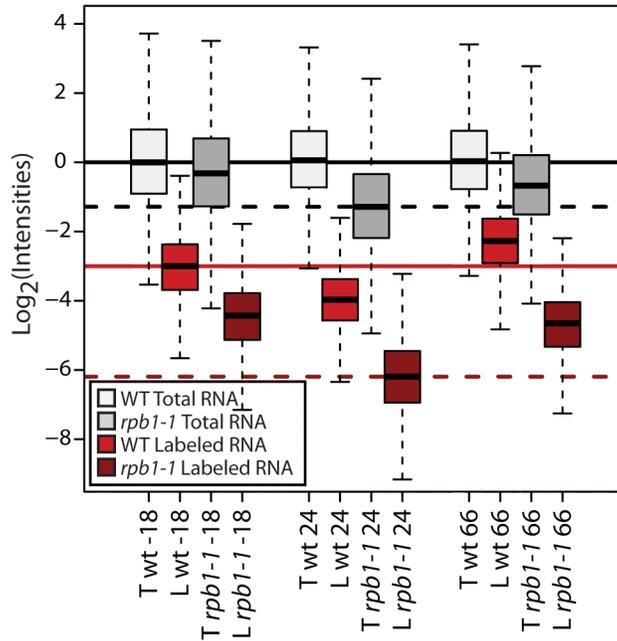
Figure 45: Box plots of the expression distributions of the total and the labeled (newly synthesized) mRNA after cDTA normalization, obtained from the wild type and the *rpb1-1* mutant before and 24 and 66 minutes after the shift to restrictive temperature. Transcriptional activity is roughly restored in both strains after 66 minutes. The global shifts in labeled expression are only slightly more pronounced in the *rpb1-1* mutant, indicating a dominant role of heat shock in the profiles of *rpb1-1*.

16.4 Comparison of mRNA metabolism in distant yeast species

As an immediate result, cDTA reveals similarities and differences in the mRNA metabolism of *Sc* and *Sp*. First, the median mRNA synthesis rates are very similar in *Sc* and *Sp* (Figure 47A). The median synthesis rate was 53 mRNAs per cell and 90 minutes for wild type *Sc* and 44 mRNAs per cell and 90 minutes for *Sp*. Second, *Sp* mRNAs have about five-fold longer half lives on average than *Sc* mRNAs, with a median of 59 minutes (Figure 47A, Supplementary Figures S4 [187]), compared to 12 minutes for *Sc*. As expected, the cDTA-derived *Sp* half-lives show a fair correlation with half-lives obtained by another non-perturbing metabolic labeling [6]. Furthermore, reprocessing the data of [6] with our DTA algorithm, which takes into account the labeling bias and an additional parameter to correct for cell growth, increases the correlation to our results and leads to a median half-live of 50 minutes, in good agreement with an estimate of 59 minutes in our study (Supplementary Figures S2 [187]). Third, the overall mRNA levels in *Sp* were about 3.1-fold higher than in *Sc*. Since the haploid *Sc* cells with a median volume of $42\,\mu\mathrm{m}^3$ are approximately two- to threefold smaller than *Sp* cells with a median cell volume of approximately $115\,\mu\mathrm{m}^3$ [100, 146], the higher mRNA levels apparently lead to similar cellular mRNA concentrations. The change in mRNA levels is mainly a global multiplicative change ($R^2 = 0.82$, Supplementary Figures S4 [187]). Taken together, these data suggest that *Sp* cells generally contain more stable mRNAs than *Sc* cells to reach similar mRNA concentrations at similar mRNA synthesis rates despite their larger volume.
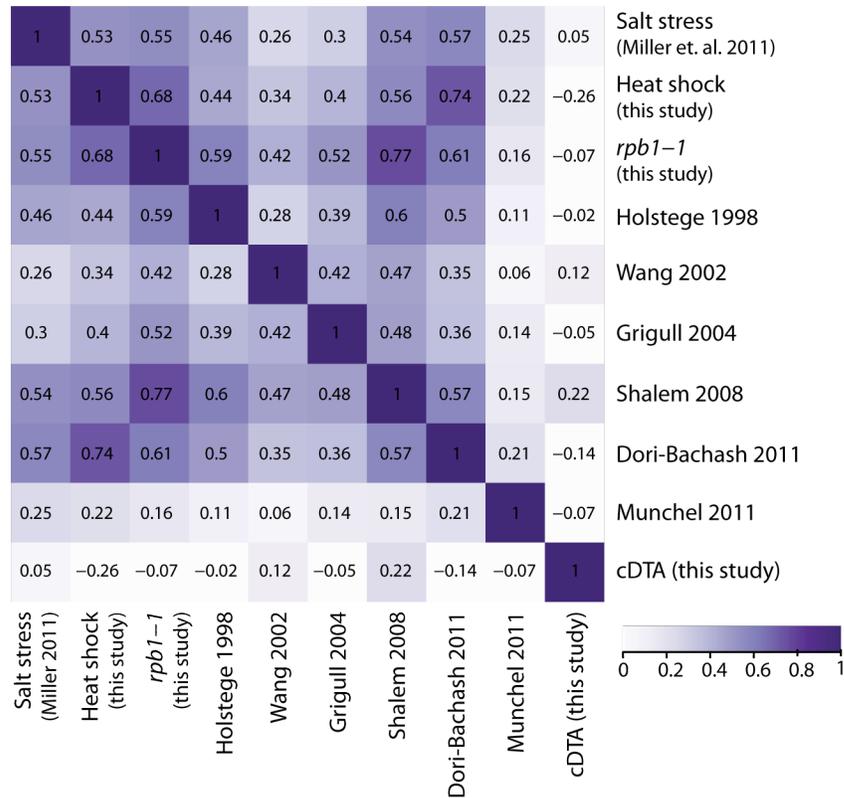
Figure 46: Correlation analysis of mRNA half-life measurements. The heatmap shows pair wise Spearman correlation coefficients of half-life measurements (white: negative or zero correlation; purple: perfect correlation). The published half-life estimates except for Munchel *et al.* [141] were obtained by experiments using transcriptional arrest. The estimates of Holstege *et al.* [90], Wang *et al.* [205], Grigull *et al.* [77] and Shalem *et al.* [176] were obtained using a yeast strain containing the Pol II temperature sensitive mutant *rpb1-1*. Dori-Bachash *et al.* [53] used the transcription inhibitor 1,10-phenanthroline.

We investigated whether mRNA sequence conservation correlates with a conservation of total RNA levels, synthesis rates, or decay rates (Figure 47B, Supplementary Figures S4 [187]). This analysis revealed a conservation of the relative total levels of mRNAs that encode orthologous proteins in *Sc* and *Sp*. The levels of mRNAs that encode proteins with an amino acid sequence identity of at least 25% (2568 mRNAs) show a high Spearman correlation of 0.69. Synthesis rates correlate well between both species (Spearman correlation 0.61), but the half-lives show only a fair correlation (Spearman correlation 0.4). Although the data suggest that *Sp* cells have globally shifted decay rates, to reach similar cellular mRNA concentrations, there is a minor fraction of transcripts that behave exceptionally. In particular, 93 *Sp* transcripts show almost unchanged mRNA levels (< 1.5 fold), but significantly higher synthesis and decay rates (> 1.5 fold), and are enriched for ribosomal protein genes (Figure 47A). More generally, transcripts that encode highly conserved proteins show similar levels, but a faster turnover in *Sp* (Figure 47B). We also assessed the correlation of synthesis rates with transcript lengths, and revealed a substantially higher Pol II drop-off rate in *Sp* (Supplementary Figures S5 [187]).
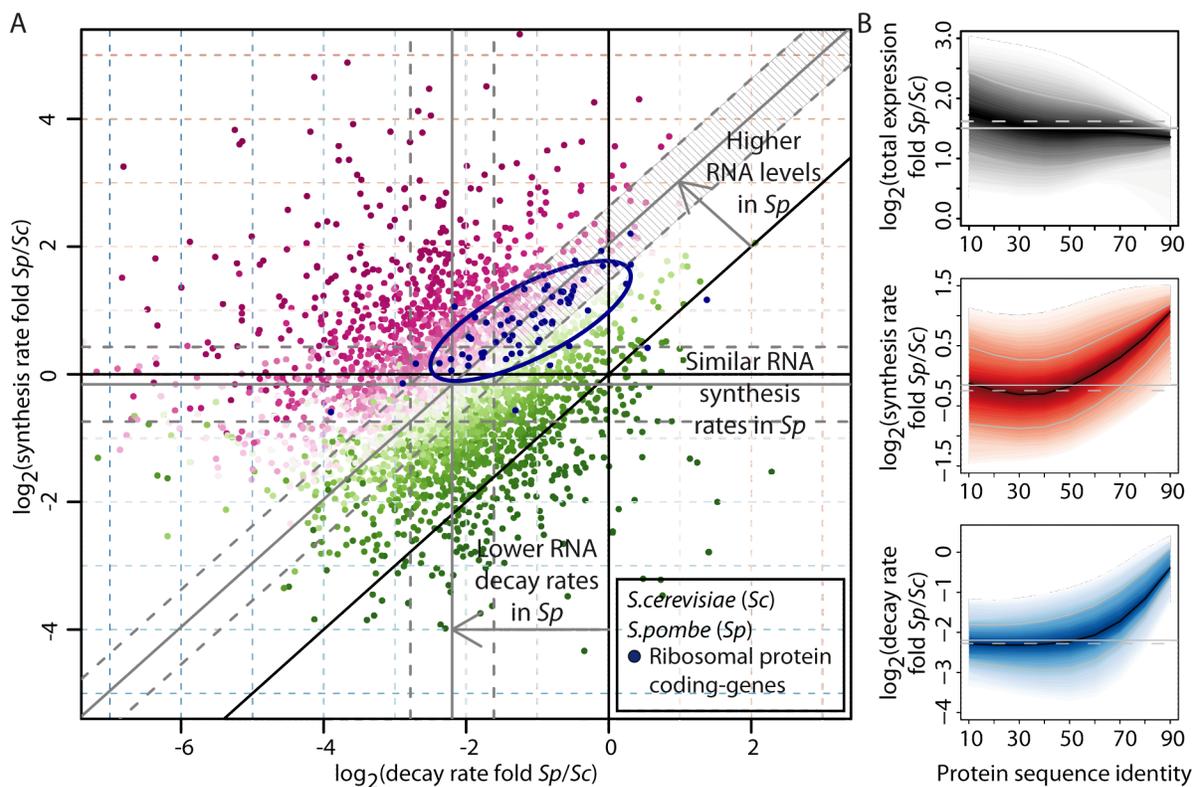
Figure 47: (A) Scatter plot comparing mRNA decay rate folds versus synthesis rate folds of *Sp* and *Sc* transcripts encoding protein orthologs ($> 25\%$ amino acid sequence identity). The offset of grey lines to parallel black lines indicate *Sp*:*Sc* ratios of median decay rates, synthesis rates, or total mRNA (0.20/0.83/2.72). Dashed grey lines indicate 1.5-fold changes from the median (grey lines). Color scheme corresponds to folds in total mRNA (magenta, positive $\log_2$ fold; green, negative $\log_2$ fold). A set of genes that show higher decay and synthesis rates (1.5-fold & adjusted P-value 0.5%) but almost unchanged (< 1.5-fold) total mRNA (93 transcripts, striped area) was selected and tested with a Bayesian network-based gene set analysis (MGSA) [15]. In this gene set, the ribosomal protein genes were enriched (blue dots; ellipse shows the 75% region of highest density). (B) Plots show $\log_2$ fold distributions of total mRNA (grey), synthesis rate (red) and decay rate (blue) of *Sp* versus *Sc* transcripts encoding orthologous proteins as a function of amino acid sequence identity (%). Transcripts encoding highly conserved proteins such as ribosomal proteins are located on the right. They show more rapid turnover (synthesis and decay) in *Sp*, resulting in similar mRNA levels. The solid black lines represent the median $\log_2$ fold, the shaded bands are the central 80% regions. The solid/dashed grey lines indicate the median $\log_2$ fold of all orthologs/all genes.

16.5 Impaired mRNA synthesis is compensated by decreased degradation

We applied cDTA to the question of whether the speed of Pol II is relevant for setting the cellular rates of mRNA synthesis. We used a yeast strain that carries the non-disruptive point mutation N488D in the largest Pol II subunit Rpo21p (also known as Rpb1p) (*rpb1-N488D*). This mutation slows down Pol II speed in RNA elongation assays *in vitro* [125] and is located near the active site [45]. We subjected this strain and an isogenic wild type strain to cDTA, and collected two biological replicates, which showed a Spearman correlation of 0.99 for total and labeled RNA (Supplementary Figures S6 [187]). We measured cell-doubling times, and used these in the kinetic modeling, to correct synthesis rates for a change in doubling time (Supplementary Figures S3 [187]). In the *rpb1-N488D* mutant strain, mRNA synthesis rates were globally decreased 3.9-fold (Figure 48A). This is consistent with the observed 2- to 4.5-fold decrease in Pol II speed measured *in vitro* [125]. We observed a Pol II drop-off rate similar to that described previously [99], but quantitative modeling excludes drop-off

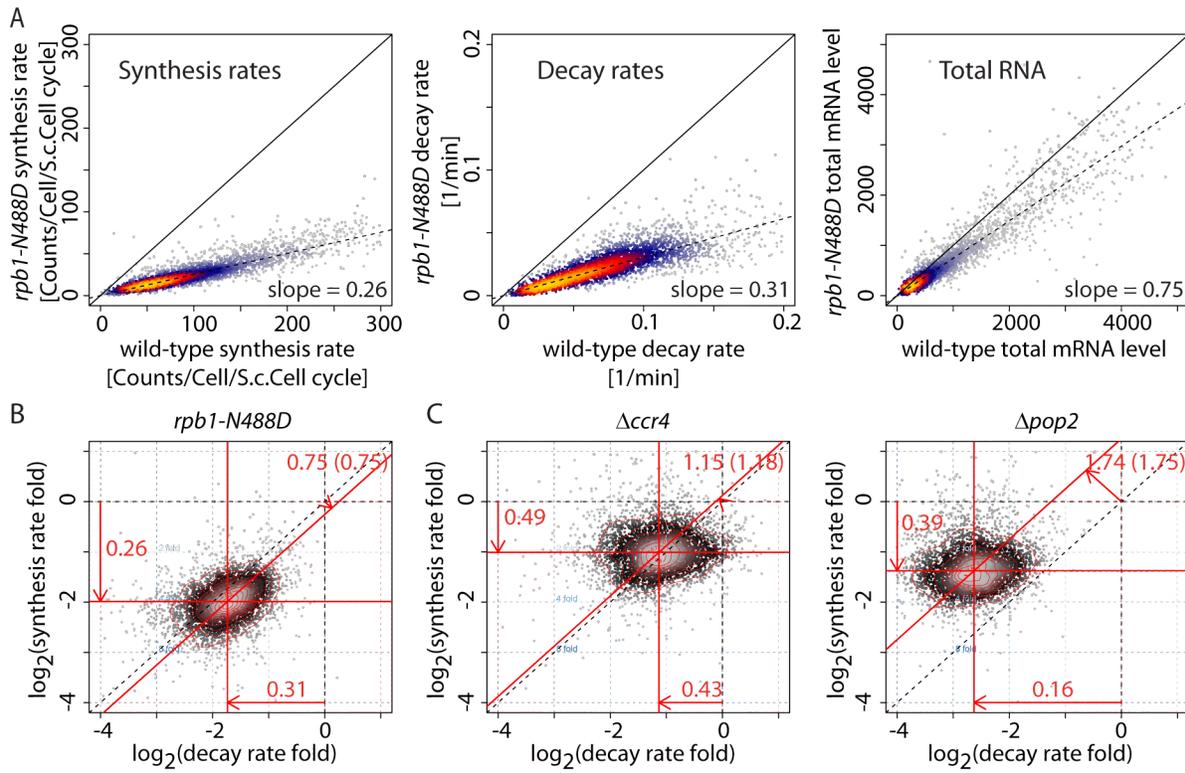of Pol II during elongation as the cause for the decreased synthesis rates (Supplementary Figures S7 [187]).

Figure 48: (A) Linear scatter plots (heat-colored) of mRNA synthesis rates, decay rates, and total mRNA levels in wild type and mutant *rpb1-N488D* yeast strains as measured by cDTA. Slopes indicate global shift ratios of median synthesis rates, decay rates, and total mRNA of the *rpb1-N488D* mutant strain compared to wild type (0.26/0.31/0.75). (B) Alternative representation of the data from panel A in a single scatter plot comparing the changes in mRNA synthesis rates ($\log_2$ folds, x-axis) and decay rates ($\log_2$ folds, y-axis) in the *rpb1-N488D* mutant strain compared to the wild type strain. Each point corresponds to one mRNA. The density of points is encoded by their brightness (grey scale). Contour lines define regions of equal density. The center of the distribution is located at $(-1.8, -1.6)$, indicating that there is a global shift in the median synthesis rate by a factor of 0.26 (shift of the horizontal red line relative to the dashed x-axis line), and a global shift in the median decay rate by a factor of 0.31 (shift of the vertical red line relative to the dashed y-axis line). The global change in total mRNA levels is predicted by the offset of the diagonal red line from the dashed main diagonal, which corresponds to a change by a factor of 0.75. The number in brackets following this number (0.75) is the global change as it has been observed in the total mRNA measurements, which agrees well with the predicted number. The changes in total RNA levels do not exactly equal the quotient of synthesis and decay rate changes, due to an additional parameter for cell growth. (C) Scatter plots as in (B) comparing synthesis rates, decay rates, and total mRNA levels of *Δccr4* and *Δpop2* mutant strains to wild type yeast. Ratios of median synthesis rates, decay rates, and total mRNA of the *Δccr4*/*Δpop2* mutant strain compared to wild type are 0.49/0.39, 0.43/0.16, and 1.15/1.74, respectively.

Despite the lower synthesis rates, global mRNA levels were not changed very much in the slow Pol II mutant strain (Figure 48A). This resulted from a strong decrease in mRNA decay rates of 3.2-fold on average. Synthesis and decay rates of all mRNAs were shifted by approximately the same factor, independent of their wild type expression level, synthesis rate, or decay rate. The globally increased mRNA half-lives apparently compensated for the decreased mRNA synthesis rates, to buffer cellular mRNA levels, which were decreased 1.3- fold only (Figure 48B). The measured total RNA levels agreed well with total mRNA levels calculated from the changed synthesis and decay rates (not

shown). These results show that cells with a strong defect in mRNA synthesis can maintain nearly normal mRNA levels by compensatory changes in mRNA decay rates.

16.6 Impaired degradation is compensated by decreased synthesis

The observed synthesis-decay compensation implies that cells buffer total mRNA levels. If true, cells should also be able to compensate for a mutation that impairs mRNA degradation with a change in mRNA synthesis rates. To investigate this, we applied cDTA to mutant yeast strains with global defects in mRNA degradation. The choice of mutant was difficult, since RNA degradation involves multiple enzymes in the nucleus and cytoplasm (Section 1.2) [91]. We decided to use mutant strains that lack either one of the two catalytic subunits of the Ccr4-Not complex, Ccr4 or Pop2, which show a defect in mRNA deadenylation, a rate-limiting step in mRNA degradation [194]. As predicted, mRNA decay rates were globally decreased in the $\Delta ccr4$ and $\Delta pop2$ strains, and changed on average by a factor of 0.43 and 0.16, respectively (Figure 48C). This suggests that Ccr4 and Pop2 mRNA degradation factors are used globally.

In both degradation-deficient knock-out strains, an unexpected decrease in mRNA synthesis rates was observed (Figure 48C). Synthesis rates were changed by a factor of 0.49 and 0.38 in the $\Delta ccr4$ and $\Delta pop2$ strains, respectively, limiting the increase in total mRNA levels due to highly defective degradation to a factor of only 1.18 and 1.75, respectively (Figure 48C). This effect could be observed directly in the labeled fractions of the $\Delta ccr4$ and $\Delta pop2$ strains. Only 62% or 46% of the RNA was labeled within the same labeling time, indicating lower synthesis rates. Thus the defects in RNA degradation in these strains are at least partially compensated by decreased mRNA synthesis rates, to buffer mRNA levels. This mutual compensation cannot be explained by measurement variance. A variation analysis for the estimation of the median synthesis and decay rates (Figure 49, Supplementary Methods S9 [187]) shows that the 95% confidence regions of the median synthesis and decay rate estimates are clearly disjoint.
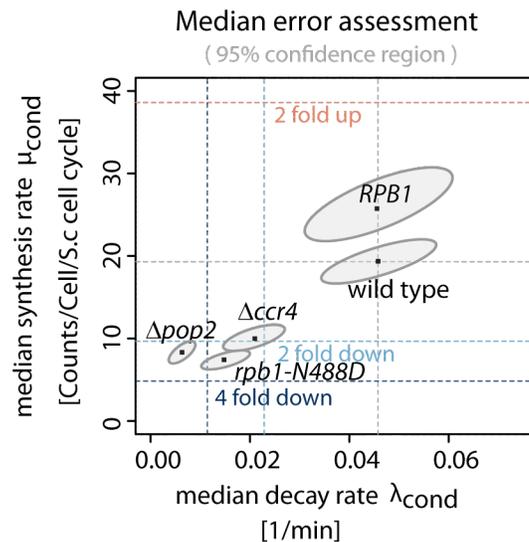
Figure 49: Coupling of synthesis and decay rates, on the absolute level. For each condition, the median synthesis rate (y-axis) and degradation rate (x-axis) is shown (dark dots). Dashed lines indicate fold induction/repression relative to wild-type. The dots lie approximately on a line with positive slope, indicating synthesis-decay compensation. A variation analysis for the estimation of the median synthesis and decay rates with cDTA has been performed. The ellipses show the 95% bootstrap confidence regions in each condition. The main axes of the ellipses reveal that the errors in the estimation of synthesis and decay rates are not independent, yet small enough to prove that the coupling is not due to estimation variance.

16.7 cDTA reveals different *in vivo* 4tU labeling and incorporation efficiency upon mutation

We can use cDTA to compare 4tU labeling and incorporation efficiencies across different strains. We define three probabilities associated with mRNA labeling process: The *incorporation efficiency* is the probability that a 4sUTP nucleotide is incorporated into a nascent mRNA instead of a UTP. Secondly, the *capture efficiency* is the probability for a 4sUTP nucleotide which is included into a mRNA of being biotinylated, captured and recovered from the streptavidin-coated magnetic beads. Third, the probability that both events occur, assuming independence, is the product of the *incorporation efficiency* and the *capture efficiency*. We refer to this as the *labeling efficiency*. Since the labeled *Sc* and *Sp* RNA from one sample are processed simultaneously, the *capture efficiency* equals each other for that sample. Moreover, the *incorporation efficiency* for *Sp* should also equal each other across different conditions, since we use a common *Sp* standard for all experiments. This circumstance then allows us to conclude to the relative *incorporation efficiency* between a mutant and wild-type strain (Section 11.6). It turns out that the relative *incorporation efficiency* is less than 1 for all three mutants considered by us (Table 2).

rpb1-N488D	*Δccr4*	*Δpop2*
0.51 ± 0.07	0.49 ± 0.05	0.71 ± 0.08

Table 2: Relative incorporation efficiencies were estimated as described in Section 11.6, standard deviations were calculated using replicate measurements.

16.8 Summary & Outlook

A systemic investigation of gene expression requires quantitative monitoring of cellular mRNA metabolism. In particular, a technique is required to quantify absolute mRNA synthesis and decay rates on a genome scale upon genetic perturbation. Here we provide such a technique that we refer to as comparative Dynamic Transcriptome Analysis (cDTA). cDTA is based on non-perturbing metabolic RNA labeling in mutant and wild type budding yeast cells, and the use of fission yeast cells as an internal standard. cDTA is a non-perturbing method for monitoring mRNA turnover and supersedes conventional methods, which require transcription inhibition, resulting in a stress response and perturbation of mRNA metabolism.

cDTA improves our previous DTA protocol (Section 15) [136] in several respects. First, cDTA provides reliable estimates of the absolute synthesis and decay rates, thereby allowing for a direct comparison of rates between different yeast strains. Second, cDTA uses 4tU instead of 4sU for RNA labeling, allowing for standard media and abolishing the need for a nucleoside transporter. Third, cDTA requires only two instead of three microarray measurements per rate estimation. As an immediate result, cDTA revealed that *Sp* and *Sc* cells have similar synthesis rates, but *Sp* RNAs have about five-fold longer mRNA half-lives, leading to similar cellular mRNA concentrations despite a different cell volume.

Application of cDTA to *Sc* cells expressing a Pol II point mutant that elongates mRNA slowly *in vitro* showed that mRNA elongation is a critical determinant for mRNA synthesis in growing cells *in vivo*. It also revealed that cells compensate for low synthesis rates by lowering decay rates, thus stabilizing mRNAs and buffering their levels. Application of cDTA to two mutant strains that lack either one of the two catalytic subunits of the mRNA deadenylase complex Ccr4-Not showed not only the expected defect in mRNA degradation but also a compensatory decrease in mRNA synthesis, also leading to a buffering of mRNA levels. This indicated the existence of a feedback loop that connects mRNA synthesis and degradation, and serves to buffer mRNA levels. These results support published evidence for a global control of mRNA levels in dependence of cell size

[220]. This global control of mRNA levels occurs despite the separation of mRNA synthesis and degradation into nuclear and cytoplasmic compartments.

The mechanisms underlying the synthesis-decay feedback loop and the buffering of mRNA levels are unclear. The feedback loop may be a result of a physical and functional coupling between the various parts of mRNA metabolism. Transcription is coupled to mRNA processing and export [127], and translation is coupled to mRNA degradation [40, 22, 41, 93]. There is also evidence that nuclear and cytoplasmic mRNA metabolism are linked. The Pol II subcomplex Rpb4/7 shuttles between the nucleus and cytoplasm [175], and is involved in transcription [54] and mRNA translation and degradation [122, 123, 82]. The Ccr4-Not complex is involved in mRNA degradation [194], but also in transcription [119, 38, 39, 108].

17 DTA as a tool for miRNA target validation in *D.melanogaster*

In most cases, miRNAs target mRNAs with only partial sequence complementarity (Section 1.3). The limited sequence complementarity can lead to a large number of different mRNA targets regulated by a single miRNA [173, 9]. The seed region (~ 8 nucleotides) can be sufficient for a productive interaction. This complicates the reliable prediction of miRNA - target mRNA interactions. Many mRNAs can in principle be regulated by a given miRNA *in silico* [184, 23, 107, 115, 163, 65, 13]. Consequently, at least 30% of all protein coding genes in animals are predicted to be regulated by miRNAs. miRNAs predominantly act as negative post-transcriptional regulators of gene expression [4, 14, 197], i.e. repress the translation or stimulate the degradation of their target mRNAs. It is however questionable, that such a large number of mRNAs is actually repressed by the respective miRNA [172], which requires the need for experimental validation. To that end, we applied DTA (Section 4) to measure transcriptional and post-transcriptional effects by miR-277, a miRNA that is expressed endogenously in *D.melanogaster* Schneider cells (S2-cells). 2'-*O*-methyl antisense oligonucleotides were used to inhibit miR-277 [96] for three days. Subsequently, the mRNA was pulse-labeled for 60 minutes and analyzed by DTA.
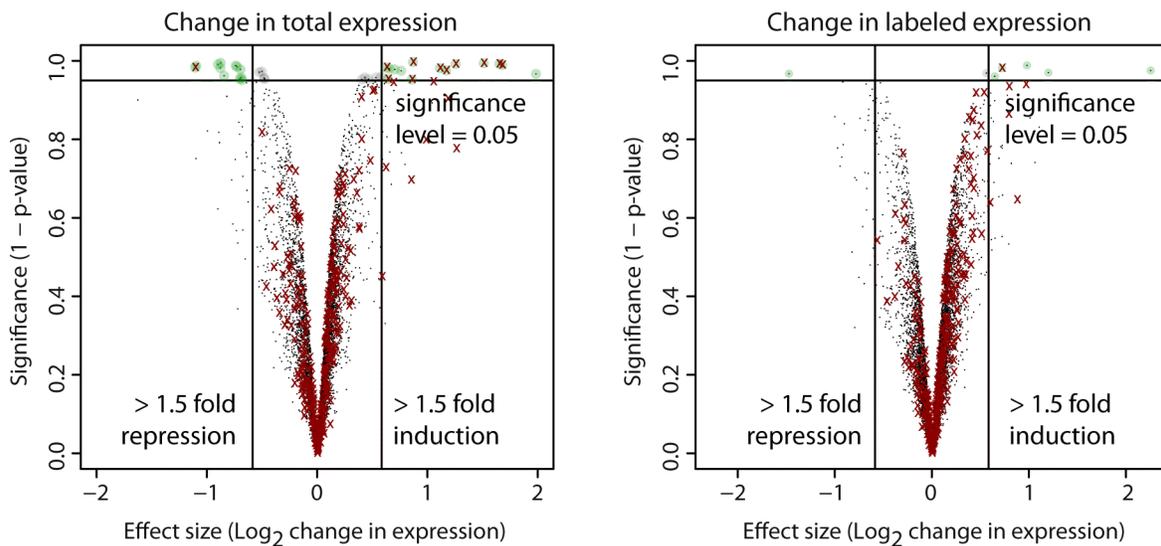

Figure 50: Volcano plots compare total and labeled mRNA levels upon inhibition of miR-277 after 60 min labeling versus reference cells treated with a control oligo. An inhibitor with an unrelated sequence (part of the luciferase gene) served as a control. Each dot corresponds to one gene, the x axis displays the $\log_2$(fold) of that gene, the y axis represents the P-value (without FDR correction, see Section 18.3). 10 out of 14 significantly up-regulated genes in the total mRNA fraction are predicted miR-277 targets (depicted in red). In contrast, the labeled mRNA fraction showed only 1 predicted miR-277 target as significantly up-regulated.

17.1 Inhibition of miR-277 reveals genuine target mRNAs

Inhibition of miR-277 should increase steady-state mRNA levels of its target mRNAs. Changes in mRNA degradation rate can be detected in the unlabeled or total mRNA fraction, whereas synthesis rate changes should be visible in the labeled fraction of the mRNA. No genes were detected as *significantly differentially expressed* with FDR correction (Section 18.3), so it was omitted for the microarray analysis. Without FDR correction, 14 genes were significantly up-regulated in the total mRNA fraction. 10 of these genes are predicted miR-277 targets (Figure 50, Table 3). On the

contrary, the labeled mRNA fraction showed only 1 predicted miR-277 target as significantly up-regulated.

gene	total RNA	unlabeled RNA ($>$1 hr)	labeled RNA ($<$1 hr)	pred. target
CG8199	1.51	1.73	*n.s.*	*yes*
CG5599	1.67	1.69	*n.s.*	*yes*
CG15093	1.17	1.01	*n.s.*	*yes*
CG3267	1.11	1.24	*n.s.*	*yes*
CG2118	1.65	1.93	0.72	*yes*
CG5044	0.63	0.55	*n.s.*	*yes*
CG6984	0.87	1.19	*n.s.*	*yes*
CG6543	1.25	1.25	*n.s.*	*yes*
CG9867	0.65	0.70	*n.s.*	*yes*
CG4594	0.84	n.s.	*n.s.*	*yes*

Table 3: 10 predicted targets were significantly up-regulated in the total mRNA fraction upon inhibition of miR-277. Table shows the $\log_2$-fold of these target mRNAs after inhibition of miR-277 against a reference treatment with a control oligo. Microarray analysis to detect *significantly differentially expressed* mRNAs upon inhibition of miR-277 was done without FDR correction (Section 18.3). Significant changes however appear to be present in the total as well as the unlabeled mRNA fraction.

miR-277 has 691 predicted mRNA targets in the targetscan database (version 6.1) [167]. 358 of these are detected as expressed in our S2-cell microarray analysis but only 10 showed significant up-regulation upon inhibition of miR-277. Most of the predicted miR-277 targets could thus not be validated by us, which might be due to unreliable predictions of miRNA - target mRNA interactions. The small seed region may lead to false positive mRNA target annotations.

17.2   Summary & Outlook

Predictions of miRNA - target mRNA interactions requires experimental validation as *in silico* prediction tools often bear false positive mRNA target annotations. Transcriptomic validation of targets of a certain miRNA was up to now limited by indistinguishable direct and indirect miRNA-mediated effects. We exemplified DTA as a valuable tool for direct miR-277 target validation in *D.melanogaster* and proved that such an analysis is possible. It may also be a valuable choice for other miRNAs. Additionally, it allows for the separation of indirect from direct effects. Only direct effects are characterized by changes in mRNA stability. Given that DTA is in principle not limited to *D.melanogaster*, we suggest it as a useful strategy for target identification and validation of miRNAs in other organisms.

Part IV
Conclusion

Technological advances in experimental techniques as well as the availability of DNA sequences of whole genomes enable to study biological questions at a global level and so far unreached complexity and detail. The field of computational biology is thus challenged to develop new approaches and efficient algorithms to gain new insights into the underlying biological systems of large scale data sets from high-throughput methods.

In this thesis, we focused on a systemic investigation of gene expression via quantitative monitoring of cellular mRNA metabolism. In particular, we developed a novel method to quantify absolute mRNA synthesis and decay rates on a genome-wide scale upon genetic perturbation (DTA/cDTA). DTA/cDTA is based on non-perturbing metabolic RNA labeling in mutant and wild type budding yeast cells. Fission yeast cells are used as an internal standard. DTA/cDTA is a new experimental approach to simultaneously measure the contributors of mRNA turnover in one single experimental setting. This can be achieved in a noninvasive manner with a superior accuracy compared to previous methods, which require transcription inhibition and thereby perturb mRNA metabolism. In this way, significant changes in mRNA synthesis are observed in newly transcribed mRNA before they have a visible effect on total cellular mRNA. DTA/cDTA is capable of monitoring kinetic parameter changes accurately even during dynamic cellular responses to environmental stimuli, and it does so at a higher sensitivity and temporal resolution than standard transcriptomics. It is a highly valuable tool that can provide quantitative data for modeling complex gene-regulatory systems, which is needed to provide new insights and to uncover misleading drawbacks of traditional methods. It could be used to characterize gene expression programs under defined conditions, and to analyze the influence of changes in environmental conditions on mRNA turnover. Further, DTA/cDTA could be used to decipher the pathways of mRNA decay and the global regulation of mRNA half-lives by RNA-binding proteins.

DTA/cDTA is provided with a statistical methodology and all required bioinformatics steps that allow the accurate absolute quantification and comparison of mRNA turnover. The bioinformatic workflow of DTA/cDTA is implemented in the open source R/Bioconductor package *DTA* [170] as a part of this thesis. The *DTA* package delivers straightforward methods to estimate RNA synthesis and decay rates from pre-processed microarray or RNASeq measurements obtained via the DTA/cDTA protocol. *DTA* fulfills the high standards of the Bioconductor platform, regarding documentation and usability. Therefore, it is easily incorporated in R scripts for pre-processing. Further statistical analysis of the results can readily be carried out by other methods within the R/Bioconductor programming environment.

To deal with rapid advances in high-throughput techniques, bioinformatic approaches will always have to be applied and adapted to progress and improvements of new experimental techniques. It should further be considered whether more complex models can further increase the prediction accuracy.

DTA/cDTA can be applied to reveal rate changes for all kinds of perturbations, e.g. knockout or point mutation strains, responses to stress stimuli or small molecule interfering assays like treatments with miRNA or siRNA inhibitors. In doing so, DTA was shown to be a valuable tool for miRNA target validation. The DTA/cDTA approach is in principle applicable to virtually every organism in which a labeling approach can be established. This covers labeling with 4sU, 4tU as well as 5'-bromo-uridine, one of the new emerging methods (BRIC-seq) [190].

Part V
Appendix

18   Statistical methods

18.1   LOcal Polynomial RegrESSion Fitting

LOESS, or LOWESS (locally weighted scatterplot smoothing) [204, 34] allows to fit a smoothed polynomial surface resp. curve to n-dimensional data. At each data point x a polynomial of low degree is fitted to a subset of the data (neighborhood of x). The size of the neighborhood can be controlled by a parameter α (defaults to $\alpha = 0.75$ of all the data points). The polynomial is fitted by means of a weighted least squares regression [19], weighted to lower the influence of data points with the largest residuals. The value of the dependent variable $f(x)$, i.e. evaluation of the local polynomial using the independent explanatory variable values, is then assigned to the data point x. The degree of the polynomial is usually 1 or 2. The described method is implemented in the *stats* package in R [160] and was used within this environment.

18.2   Methods of moments for gamma Γ distributed sample populations

The method of moments is commonly used to estimate the population parameters of a certain distribution by equating the sample moments with the unobservable population moments. It is based on the assumption that the sample moments converge to the population moments of the underlying distributions if the number of observations $N \to \infty$ [217, 44, 135, 19]. Suppose $X_1, ..., X_N$ are independent identically distributed random variables that follow a gamma distribution with shape k and scale θ:

$$X_1, ..., X_N \underset{iid}{\sim} \Gamma(k, \theta) \tag{142}$$

The first moment - the expected value - is then given as

$$E(X_1) = k\theta \tag{143}$$

and can be approximated as the first sample moment with

$$m_1 = \frac{X_1 + \cdots + X_N}{N} \tag{144}$$

The second moment - the expected value of its square - can be stated as

$$E(X_1^2) = \theta^2 k(k+1) \tag{145}$$

and can be approximated as the second sample moment with

$$m_2 = \frac{X_1^2 + \cdots + X_N^2}{N} \tag{146}$$

Given Equation (144) and (146), the population parameters k and θ can be calculated as

$$k = \frac{m_1^2}{m_2 - m_1^2} \tag{147}$$

and

$$\theta = \frac{m_2 - m_1^2}{m_1} \tag{148}$$

based on the observed samples. Note that $E(X_1)^2 - E(X_1^2) = Var(X_1) = \theta^2 k$.

18.3 Detection of differentially expressed genes

We eventually aim to identify genes that behave differently in the comparison of two groups of genome-wide measurements, be it total mRNA levels, labeled mRNA levels, or synthesis rates. The problem of identifying differentially expressed genes in microarray experiments with arbitrary numbers of groups and mRNA samples was considered by Gordon K. Smyth [181]. His model is formulated as a linear regression problem. The estimators proposed show robust behavior even for small numbers of arrays. The approach is implemented in the R/Bioconductor package *LIMMA* [181], which can be used after appropriate normalization of the used expression profiles (see Section 18.4). Multiple testing correction is done by converting p-values into local false discovery rates [17]. We consider a gene *significantly differentially expressed* if it achieves a local false discovery rate smaller than 5% in the respective two-group comparison. The effect of induction/repression is called *relevant* if the mean expressions of the two groups differ by a factor of at least 2. A gene is called *induced/repressed* if it is significantly and relevantly up-/down-regulated relative to the reference group.

18.4 Baseline normalization

Variations in mRNA extraction efficiencies, amplification steps in the biochemical protocol and scanner calibration of the fluorescence readouts introduce slight differences in the global spot intensity levels on the microarrays. This problem is often solved by centering the medians of the respective expression profiles to a common value [20], often referred to as baseline normalization.

Let G be the set of genes that were measured on the arrays A. The measured gene expression of gene $g \in G$ on a microarray a is denoted by F_{ga}. We further set $m_a = \text{median}\{F_{ga} \mid g \in G\}$, the median of the microarray a. Each microarray $(F_{ga})_{g \in G}$ is rescaled such that $\text{median}\{F_{ga} \mid g \in G\} = m$, where $m = \text{median}\{m_a \mid a \in A\}$ is the median of the median intensities. Namely

$$F_{ga} \text{ is replaced by } \frac{F_{ga} \cdot m}{\text{median}\{F_{ga} \mid g \in G\}} \text{ for all } a \in A \,.$$

This is only reasonable though, if the assumption can be met, that the global level of expression has not changed. In our DTA/cDTA procedure however, this is not necessary for rate extraction. Proportional rescaling of expression profiles is completely compensated by the total least squares regression, which is a feature of our estimation procedure (Section 11.3). For cDTA, it is in fact only required once to calibrate all microarrays to yield the respective absolute median half-life of the wild-type reference (Sections 4.3, 16). Given global changes in total mRNA levels, synthesis or decay rates, as they can be revealed by cDTA, centering the medians of the respective profiles to a common value, is a reasonable approach for further investigation of differential behavior of certain genes referring to the majority of the genes. This can be addressed by considering the internal ranking or the by detecting differential expression (Section 18.3) to find *intra-differential* expressed transcripts, i.e. differential behavior referring to the majority of the genes, and hence regulation that is independent of changes in the global level.

19 Materials and methods for Section 15

19.1 RNA labeling and microarray analysis

We used *S. cerevisiae* strain BY4741 *MATa, his2Δ1, leu2Δ0, met15Δ0, ura3Δ0* (Euroscarf). The strain was transformed with plasmid YEpEBI311 ($2\,\mu$m, *LEU2*) carrying the human equilibrative nucleoside transporter hENT1. Samples for establishing DTA were grown in SD medium overnight, diluted to an OD_{600} of 0.1 the next day and grown up to a mid-log phase (a final OD_{600} of 0.8 corresponding to $1.75 \cdot 10^7$ cells per ml). 4sU (Sigma) was added to the media and made up to a final concentration of $500\,\mu$M, and cells were harvested after different labeling times. Cells were centrifuged at 4000 r.p.m. for 1 min and pellets were flash-frozen in liquid nitrogen. Samples for quantitative RT–PCR and for salt-stress experiments were grown in SILAC medium lacking leucine (6.7 g/l Formedia yeast nitrogen base without amino acids, 2% glucose, 2% agar, 200 mg/l adenine, 200 mg/l tyrosine, 10 mg/l histidine, 10 mg/l methionine, 60 mg/l phenylalanine, 40 mg/l tryptophan, and 20 mg/l uracil) overnight, diluted to an OD_{600} of 0.1 the next day and grown up to a final OD_{600} of 0.8. Cells were harvested 0, 6, 12, 18, 24, 30, and 36 min after addition of NaCl up to a final concentration of 0.8 M. Other steps were as above. Total RNA was extracted with the RiboPure-Yeast Kit (Ambion/Applied Biosystems), following the manufacturer's protocol. Labeled RNA was chemically biotinylated and purified using strepatavidin-coated magnetic beads as described [52]. Labeling of samples for array analysis was performed using the GeneChip 3'IVT labeling assay (Affymetrix) with 100 ng input RNA. Samples were hybridized to GeneChip® Yeast Genome 2.0 Array following the instructions from the supplier (Affymetrix®). Quality control and array processing was done using *GCRMA* [213] for expression quantification and *LIMMA* [182] for elementary array comparisons (Supplementary Sections 12.1–12.5 [136]).

19.2 Pol II *in vitro* transcription

To determine the incorporation efficiency of 4sUTP by Pol II, we performed *in vitro* RNA extension assays with reconstituted elongation complexes containing a synthetic nucleic acid scaffold with an adenine base in templating position $+1$ and $+2$ (Scaffold A, [189]). The assembled ECs were incubated with 0.002, 0.01, 0.05, 0.25, or $1.25\,\mu$M of 4sUTP or UTP. Reactions were stopped at 10 min and product RNAs were separated by gel electrophoresis and quantified with a fluorimager. Lineweaver-Burk plots were used to analyze Michaelis-Menten kinetics K_M values of 13 nM and 3 nM for 4sUTP and UTP, respectively, but the same rate constant k_{cat}. Bead-based assays were as described [49, 189].

19.3 mRNA decay analysis by RT–PCR

mRNA levels were determined for eight genes: *ACT1* (YFL039C), *CTT1* (YGR088W), *GPD1* (YDL022W), *KSS1* (YGR040W), *RDN1* (rRNA locus), *SFG1* (YOR315W), *STL1* (YDR536W), and *TUB2* (YFL037W). The experiment was performed in two steps. Step 1 was performed under normal growth conditions whereas step 2 was performed the same way after addition of 0.8 M NaCl. mRNA levels were analyzed at 0, 2.5, 6, 10, and 16 min after addition of 1,10-phenanthroline ($100\,\mu$g/ml final concentration) and at 0, 12, 30, and 36 min. 4sU labeling was carried out for 6 min at a final concentration of $500\,\mu$M (Supplementary Figure S3 [136]). RNA was extracted as above. cDNA synthesis was performed with 500ng RNA originating from total, unlabeled and labeled mRNA using the iScript cDNA Synthesis Kit (BioRad). Primers were designed with the ProbeFinder software (Roche Applied Science) and individual primer-pair efficiency was tested and ranged between 95 and 100%. Sequence information of primer pairs used in this study is available upon request. PCR reactions contained $1\,\mu$l DNA template, $2\,\mu$l of $10\,\mu$M primer pairs, and $12.5\,\mu$l SsoFast EvaGreen Supermix (BioRad). Quantitative RT–PCR was performed on a Bio-Rad CFX96

Real-Time System (Bio-Rad Laboratories Inc.) using a 3 min denaturing step at 95°C, followed by 49 cycles of 30s at 95°C, 30s at 61°C, and 15s at 72°C. Threshold cycle (Ct) values were determined by application of the corresponding Bio-Rad CFX Manager software version 1.1 using the Ct determination mode 'Regression'. Two biological and three technical replicates were used for each time point and technical variance was minimized using in-plate controls.

19.4 Genomic occupancy profiling

For genomic occupancy profiling by ChIP-chip, we used *S. cerevisiae* strain BY4741 containing a C-terminal tandem affinity purification (TAP) tag on the Pol II subunit Rpb3 (Open Biosystems).We confirmed that the TAP tag was at the correct genomic position, that the tagged Rpb3 subunit was expressed, and that the strain grew normally at 30°C. Yeast cells were grown in YPD medium until exponential phase ($OD_{600} \backsim 0.8$) and then stressed by the addition of 0.8 M NaCl. ChIP-chip was performed for biological replicates 0, 12, and 24 min after salt addition with high-resolution tiling microarrays as described in [129]. The bioinformatics analysis has been done using the R/Bioconductor software package *Starr* [218], see (Supplements, Part II, Section 16–18 [136]).

19.5 Accession codes

Microarray data were deposited in ArrayExpress under accession code E-MTAB-439.

19.6 Supplementary information

Supplementary information is available at the Molecular Systems Biology website http://www.nature.com/msb/journal/v7/n1/suppinfo/msb2010112_S1.html.

20   Materials and methods for Section 16

20.1   Yeast strains and growth curves

Strains *RPB1* and *rpb1-N488D* (GRY 3020 and GRY 3027 respectively) were generously provided by Mikhail Kashlev [125]. Genotypes of GRY 3020 and GRY 3027 are *MATa, his3Δ1, leu2Δ0, lys2Δ0, met15Δ0, trp1Δ::hisG, URA3*::CMV-tTA *RPO21* and *MATa, his3Δ1, leu2Δ0, lys2Δ0, met15Δ0, trp1Δ::hisG, URA3*::CMV-tTA *rpb1-N488D*. For cDTA we used *Sc* wild type strain BY4741 *MATa, his2Δ1, leu2Δ0, met15Δ0, ura3Δ0* (Euroscarf) and the isogenic knockout *Δpop2* and *Δccr4*. *Δpop2* was from the YKO library (OpenBiosystem) and *Δccr4* was generated by substituting the target gene for a KanMX cassette using homologous recombination in the same genetic background [121]. The *rpb1-1* (*rpb1-G1437D*) strain and isogenic wild type strain were generated in our lab. Plasmids pRS316-*RPO21*, pRS315-*RPO21* and pRS315-*rpb1-1* were generated by cloning the respective ORF or mutant ORF plus sequences 500 bp upstream and 250 bp downstream into pRS316 (ATCC) and pRS315 (ATCC) using XhoI/SacI restriction sites. The heterozygous *RPO21/rpo21Δ Sc* yeast strain (BY4743, *rpo21*::KanMX6/*RPO21*) was generated and transformed with pRS316-*RPO21*. Diploids were sporulated and tetrads dissected on YPD plates. After transformation of the shuffle strains with pRS315-*RPO21* or pRS315-*rpb1-1*, the resulting strains were streaked twice on 5-FOA plates and then on SC-Leu. *Sp* strain FY2317 h+, leu1-32::hENT1-leu1+(pJAH29) his7-366::hsv-tk-his7+(pJAH31) ura4-D18 ade6-M210 [88] was kindly provided by Susan Forsburg. YPD medium was inoculated with a single *Sc* colony. *Sp* was grown in YES medium. The culture was grown to stationary phase overnight and diluted to $OD_{600} = 0.1$. Measure points were taken every hour before OD_{600} reached 3. Additional time points were taken until stationary phase was reached. Doubling time was calculated by fitting the log-transformed values of OD_{600} into a linear function.

20.2   Comparative Dynamic Transcriptome Analysis (cDTA)

Sc cells were grown in YPD medium overnight, diluted to an OD_{600} of 0.1, and grown to midlog phase. OD_{600} of 0.8 corresponded to $1.75 \cdot 10^7$ cells per ml. 4-thiouracil (4tU, Sigma, 2M in DMSO) was added to the media at a final concentration of 5mM, and cells were harvested after 6 minutes of labeling by centrifugation at 2465×g 30°C for 1 minute. The supernatant was discarded and the pellet re-suspended in RNAlater solution (Ambion/Applied Biosystems). The cell concentration was determined by Cellometer N10 (Nexus) before flashfreezing in liquid nitrogen. *Sp* cells were grown in YES medium overnight, diluted to $OD_{600} = 0.1$ and grown to $OD_{600} = 0.8$. 4sU was solved in ddH2O (50mM) and added to a final concentration of $500\,\mu$M and cells were labeled for 6 minutes. Cells were harvested by centrifugation at 2465×g for 3 minutes. Other steps were as above. A 4l culture of *Sp* cells was labeled to generate a stock and eliminate errors by variations in the standard. Cells were counted as above. *Sp* cells were mixed with *Sc* cells in a 1:3 ratio, resulting $4 \cdot 10^8$ cells in total. Total RNA extraction, labeled RNA purification as well as sample hybridization and microarray scanning were as previously described [136]. For the cDTA analysis of *rpb1-1* strains, overnight cultures were diluted in fresh medium to OD_{600} of 0.15 (125ml cultures, 160rpm shaking incubator, 30°C). At OD_{600} of 0.9 (time point -18 min) RNA was labeled. 18 minutes later (time point 0 min) cultures were shifted to 37°C by adding the same volume of 42°C-tempered medium. RNA was again labeled 18 minutes and 60 minutes after heat shock (time points $+24$ min and $+66$ min, respectively).

20.3   cDTA data analysis

Data was pre-processed array-wise using expresso (R/Bioconductor [160, 71]) with the RMA background correction method [213]. We created our own probe annotation environment (cdf), which excludes probes in probesets that show cross-hybridization between *Sc* and *Sp*. 8708 annotated *Sc*

probes and $13,317$ annotated Sp probes out of a total of $120,855$ probes showed cross-hybridization when a conservative intensity cut-off of 4.5 (log(intensity) values after pre-processing) was used. Cross-hybridizing probes were excluded from further analysis. This included 16 whole probe sets (Fig. 42A, see Supplementary Figures S1 [187]). Note that the standard GC-RMA method is not suitable for our purposes, since its bias model cannot handle bimodal intensity distributions, as caused by the simultaneous hybridization of Sc and Sp transcripts with global differences in RNA abundance (Fig. 43). Labeling bias estimation and correction was done as described (Section 11.1). Between-array normalization of arrays containing mixed Sc and Sp total RNA was done by proportional rescaling, such that the median Sp gene expression level was 1 (Fig. 44B). Accordingly, between-array normalization of arrays containing mixed Sc and Sp labeled RNA was done by proportionally scaling the array to a median labeled Sp gene expression level of c_{Sp} (Fig. 44A). The constant c_{Sp} scales the median half-life of all experiments. We calibrated c_{Sp} in a way that the resulting median Sc wild type mRNA half-life equaled that observed previously (Section 15, Table 1). Now, all Sc RNA levels, no matter if total or labeled, no matter from which experiment, can be compared on an absolute level. Decay rates and synthesis rates were obtained as described (Section II). We assume that the labeled RNA fraction is subject to degradation from the very time it is synthesized. In contrast, Rabani *et al.* 2011 [161] (Supplementary Methods therein) assume that the labeled RNA fraction is mostly nuclear and not degraded at all. We compared the synthesis rate estimates resulting from both alternatives (Supplementary Methods S9 [187]). Given our labeling time, the differences of both approaches are negligible. The whole analysis workflow has been carried out using the open source R/Bioconductor package *DTA* (Section 6) [170].

20.4 RT-qPCR

Sp and Sc cells were grown to $OD_{600} = 0.8$, harvested and flash-frozen in liquid nitrogen. Cells were counted and mixed at 1:1 and 10:1 ratios. Total RNA was extracted, and the mRNA levels of Sc genes *ACT1* (YFL039C), *ADH1* (YOL086C), *HIS4* (YCL030C), and *Rdn1* (rRNA locus) and Sp genes *GDI1* (SPAC22H10.12c) and *GPD1* (SPBC215.05) were determined by RT-qPCR. RT-qPCR was carried out as described (Section 19.3). 500ng RNA was used to reverse transcribe cDNA using the iScript cDNA Synthesis Kit (BioRad). Primers were designed with the ProbeFinder online tool (http://qpcr.probefinder.com/organism.jsp, Roche Applied Science). The primer-pair efficiency was tested individually and ranged between 97 and 100%. PCR reactions contained $1\,\mu$l DNA template, $2\,\mu$l of $10\,\mu$M primer pairs, and $12.5\,\mu$l SsoFast EvaGreen Supermix (BioRad). qPCR was performed on a Bio-Rad CFX96 Real-Time System (Bio-Rad Laboratories Inc.) using a 30 sec denaturing step at 95°C, followed by 40 cycles of 1s at 95°C, 4s at 63°C. Data analysis was performed with the software Bio-Rad CFX Manager 1.6.

20.5 Accession codes

Microarray data were deposited in ArrayExpress under accession codes E-MTAB-760.

20.6 Supplementary information

Supplementary information is available at the Genome Research website
http://genome.cshlp.org/content/early/2012/04/16/gr.130161.111/suppl/DC1.

References

[1] C. C. Adams and D. S. Gross. The yeast heat shock response is induced by conversion of cells to spheroplasts and by potent transcriptional inhibitors. *Journal of bacteriology*, 173(23):7429–7435, Dec. 1991. 3

[2] L. A. Allison, M. Moyle, M. Shales, and C. James Ingles. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*, 42(2):599–610, Sept. 1985. 1.1

[3] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, Florida: Chapman & Hall/CRC Press, 2006. 15.11

[4] V. Ambros. The functions of animal microRNAs. *Nature*, 431(7006):350–355, Sept. 2004. 1.3, 17

[5] S. L. Ameres, J. Martinez, and R. Schroeder. Molecular Basis for Target RNA Recognition and Cleavage by Human RISC. *Cell*, 130(1):101–112, July 2007. 1.3.2

[6] M. J. Amorim, C. Cotobal, C. Duncan, and J. Mata. Global coordination of transcriptional control and mRNA decay during cellular differentiation. *Molecular Systems Biology*, 6, June 2010. 4, 15.5, 16.4

[7] A. F. Andersson, M. Lundgren, S. Eriksson, M. Rosenlund, R. Bernander, and P. Nilsson. Global analysis of mRNA stability in the archaeon Sulfolobus. *Genome biology*, 7(10):R99+, Oct. 2006. 3, 3.2, 5, 15

[8] J.-C. Andrau, L. van de Pasch, P. Lijnzaad, T. Bijma, M. G. Koerkamp, J. van de Peppel, M. Werner, and F. C. P. Holstege. Genome-Wide Location of the Coactivator Mediator: Binding without Activation and Transient Cdk8 Interaction on DNA. *Molecular Cell*, 22(2):179–192, Apr. 2006. 15.11

[9] D. Baek, J. Villan, C. Shin, F. D. Camargo, S. P. Gygi, and D. P. Bartel. The impact of microRNAs on protein output. *Nature*, 455(7209):64–71, Sept. 2008. 1.3.2, 17

[10] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4):1017–1031, Oct. 2007. 1

[11] C. Barreau, L. Paillard, and H. B. Osborne. AU-rich elements and associated factors: are there unifying principles? *Nucleic acids research*, 33(22):7138–7150, 2005. 15.4

[12] D. Bartel. MicroRNAsGenomics, Biogenesis, Mechanism, and Function. *Cell*, 116(2):281–297, Jan. 2004. 1.3.1

[13] D. P. Bartel. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, 136(2):215–233, Jan. 2009. 17

[14] D. P. Bartel and C.-Z. Chen. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet*, 5(5):396–400, May 2004. 1.3, 17

[15] S. Bauer, J. Gagneur, and P. N. Robinson. GOing Bayesian: model-based gene set analysis of genome-scale data. *Nucleic acids research*, 38(11):3523–3532, June 2010. 47

[16] T. H. Beilharz and T. Preiss. Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome. *RNA*, 13(7):982–997, July 2007. 1.2.1

[17] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995. 18.3

[18] J. A. Bernstein, A. B. Khodursky, P.-H. H. Lin, S. Lin-Chao, and S. N. Cohen. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences of the United States of America*, 99(15):9697–9702, July 2002. 3, 3.2, 5, 15

[19] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006. 2, 18.1, 18.2

[20] B. M. Bolstad, R. A. Irizarry, M. Astrand, and T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics (Oxford, England)*, 19(2):185–193, Jan. 2003. 18.4

[21] G. A. Brar, M. Yassour, N. Friedman, A. Regev, N. T. Ingolia, and J. S. Weissman. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science*, 335(6068):552–557, Feb. 2012. 1

[22] M. Brengues, D. Teixeira, and R. Parker. Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science (New York, N.Y.)*, 310(5747):486–489, Oct. 2005. 16.8

[23] J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen. Principles of MicroRNA-Target Recognition. *PLoS biology*, 3(3):e85+, Mar. 2005. 1.3, 1.3.2, 17

[24] C. E. Brown and A. B. Sachs. Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation. *Molecular and cellular biology*, 18(11):6548–6559, Nov. 1998. 1.2.1

[25] F. Brueckner, U. Hennecke, T. Carell, and P. Cramer. CPD Damage Recognition by Transcribing RNA Polymerase II. *Science*, 315(5813):859–862, Feb. 2007. 15.1

[26] S. Buratowski. The CTD code. *Nat Struct Mol Biol*, 10(9):679–680, Sept. 2003. 1.1.2

[27] S. Buratowski. Connections between mRNA 3' end processing and transcription termination. *Current Opinion in Cell Biology*, 17(3):257–261, June 2005. 1.1.3

[28] S. Buratowski. Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell*, 36(4):541–546, Nov. 2009. 1.1, 1.1.2

[29] D. L. Cadena and M. E. Dahmus. Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II. *Journal of Biological Chemistry*, 262(26):12468–12474, Sept. 1987. 1.1

[30] S. B. Cambridge, F. Gnad, C. Nguyen, J. L. Bermejo, M. Krüger, and M. Mann. Systems-wide Proteomic Analysis in Mammalian Cells Reveals Conserved, Functional Protein Turnover. *J. Proteome Res.*, 10(12):5275–5284, Nov. 2011. 1

[31] D. Cao and R. Parker. Computational modeling of eukaryotic mRNA turnover. *RNA*, 7(9):1192–1212, Sept. 2001. 1.2

[32] A. P. Capaldi, T. Kaplan, Y. Liu, N. Habib, A. Regev, N. Friedman, and E. K. O'Shea. Structure and function of a transcriptional network activated by the MAPK Hog1. *Nature Genetics*, 40(11):1300–1306, Oct. 2008. 38, 15.8, 15.9, 15.11

[33] H. C. Causton, B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L. True, E. S. Lander, and R. A. Young. Remodeling of yeast genome expression in response to environmental changes. *Molecular biology of the cell*, 12(2):323–337, Feb. 2001. 3, 15

[34] J. M. Chambers and T. Hastie. *Statistical models in S*. Wadsworth & Brooks/Cole computer science series. Wadsworth & Brooks/Cole Advanced Books & Software, 1992. 18.1

[35] H.-T. Chen and S. Hahn. Binding of TFIIB to RNA Polymerase II. *Molecular Cell*, 12(2):437–447, Aug. 2003. 1.1.1

[36] M. D. Cleary, C. D. Meiering, E. Jan, R. Guymon, and J. C. Boothroyd. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotech*, 23(2):232–237, Feb. 2005. 4.1, 4.2, 5, 15

[37] S. M. Cohen, J. Brennecke, and A. Stark. Denoising feedback loops by thresholding-a new role for microRNAs. *Genes & Development*, 20(20):2769–2772, Oct. 2006. 1.3

[38] M. A. Collart. Global control of gene expression in yeast by the Ccr4-Not complex. *Gene*, 313:1–16, Aug. 2003. 16.8

[39] M. A. Collart and Timmers. *The Eukaryotic Ccr4-Not Complex: A Regulatory Platform Integrating mRNA Metabolism with Cellular Signaling Pathways?*, volume 77, pages 289–322. Elsevier, 2004. 16.8

[40] J. Coller and R. Parker. EUKARYOTIC mRNA DECAPPING. *Annual Review of Biochemistry*, 73(1):861–890, 2004. 16.8

[41] J. Coller and R. Parker. General translational repression by activators of mRNA decapping. *Cell*, 122(6):875–886, Sept. 2005. 16.8

[42] T. A. Cooper, L. Wan, and G. Dreyfuss. RNA and Disease. *Cell*, 136(4):777–793, Feb. 2009. I

[43] J. L. Corden, D. L. Cadena, J. M. Ahearn, and M. E. Dahmus. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 82(23):7934–7938, Dec. 1985. 1.1

[44] H. Cramer. *Mathematical methods of statistics*. Princeton Univ. Press, 1946. 18.2

[45] P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, N.Y.)*, 292(5523):1863–1876, June 2001. 16.5

[46] F. Crick. Central dogma of molecular biology. *Nature*, 227, 1970. I, 1

[47] E. Czeko, M. Seizl, C. Augsberger, T. Mielke, and P. Cramer. Iwr1 Directs RNA Polymerase II Nuclear Import. *Molecular Cell*, 42(2):261–266, Apr. 2011. 1.1

[48] E. de Nadal and F. Posas. Multilayered control of gene expression by stress-activated protein kinases. *The EMBO journal*, 29(1):4–13, Jan. 2010. 15

[49] S. Dengl and P. Cramer. Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro. *Journal of Biological Chemistry*, 284(32):21270–21279, Aug. 2009. 1.1.3, 19.2

[50] A. M. Denli, B. B. J. Tops, R. H. A. Plasterk, R. F. Ketting, and G. J. Hannon. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–235, Nov. 2004. 1.3.1

[51] S. Diederichs and D. A. Haber. Dual Role for Argonautes in MicroRNA Processing and Posttranscriptional Regulation of MicroRNA Expression. *Cell*, 131(6):1097–1108, Dec. 2007. 1.3.2

[52] L. Dölken, Z. Ruzsics, B. Rädle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal, and U. H. Koszinowski. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9):1959–1972, Sept. 2008. 2, 4, 4.1, 4.2, 5, 15, 15.1, 15.5, 19.1

[53] M. Dori-Bachash, E. Shema, and I. Tirosh. Coupled evolution of transcription and mRNA degradation. *PLoS biology*, 9(7):e1001106+, July 2011. 15.4, 16, 16.3, 46

[54] A. M. Edwards, C. M. Kane, R. A. Young, and R. D. Kornberg. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry*, 266(1):71–75, Jan. 1991. 16.8

[55] S. Egloff and S. Murphy. Cracking the RNA polymerase II CTD code. *Trends in Genetics*, 24(6):280–288, June 2008. 1.1.2

[56] X. Escote, M. Zapater, J. Clotet, and F. Posas. Hog1 mediates cell-cycle arrest in G1 phase by the dual targeting of Sic1. *Nat Cell Biol*, 6(10):997–1002, Oct. 2004. 15

[57] S. Esslinger and K. Förstemann. MicroRNAs Repress Mainly through mRNA Decay. *Angewandte Chemie International Edition*, 48(5):853–855, 2009. 4

[58] M. Esteller. Non-coding RNAs in human disease. *Nature reviews. Genetics*, 12(12):861–874, Dec. 2011. I

[59] A. Eulalio, I. Behm-Ansmant, and E. Izaurralde. P bodies: at the crossroads of post-transcriptional pathways. *Nature reviews. Molecular cell biology*, 8(1):9–22, Jan. 2007. 1.2, 1.2.3

[60] A. Eulalio, E. Huntzinger, and E. Izaurralde. Getting to the Root of miRNA-Mediated Gene Silencing. *Cell*, 132(1):9–14, Jan. 2008. 1.3.2

[61] A. Eulalio, E. Huntzinger, T. Nishihara, J. Rehwinkel, M. Fauser, and E. Izaurralde. Deadenylation is a widespread effect of miRNA regulation. *RNA*, 15(1):21–32, Jan. 2009. 1.3.2

[62] Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, Y. Lubling, J. Widom, and E. Segal. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS computational biology*, 4(11):e1000216+, Nov. 2008. 1.1.1

[63] T. M. Franks and J. Lykke-Andersen. The Control of mRNA Decapping and P-Body Formation. *Molecular Cell*, 32(5):605–615, Dec. 2008. 1.2.3

[64] C. C. Friedel, L. Dölken, Z. Ruzsics, U. H. Koszinowski, and R. Zimmer. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Research*, 37(17):e115, Sept. 2009. 4, 5, 15

[65] R. C. Friedman, K. K.-H. K. Farh, C. B. Burge, and D. P. Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, 19(1):92–105, Jan. 2009. 17

[66] N. J. Fuda, M. B. Ardehali, and J. T. Lis. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, 461(7261):186–192, Sept. 2009. 1.1, 1.1.1, 1.1.2

[67] J. Garcıa-Martınez, A. Aranda, and J. E. Perez-Ortın. Genomic Run-On Evaluates Transcription Rates for All Yeast Genes and Identifies Gene Regulatory Mechanisms. *Molecular Cell*, 15(2):303–313, July 2004. 5, 15

[68] N. L. Garneau, J. Wilusz, and C. J. Wilusz. The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol*, 8(2):113–126, Feb. 2007. 1.2.1

[69] A. L. Gartel and E. S. Kandel. RNA interference in cancer. *Biomolecular Engineering*, 23(1):17–34, Mar. 2006. I

[70] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell*, 11(12):4241–4257, Dec. 2000. 3, 15

[71] R. Gentleman, V. Carey, D. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Yang, and J. Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80+, 2004. 20.3

[72] R. C. Gentleman, V. J. Carey, D. M. Bates, and others. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80, 2004. 6.2

[73] A. P. Gerber, D. Herschlag, and P. O. Brown. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS biology*, 2(3), Mar. 2004. 1.2.1

[74] M. Ghildiyal and P. D. Zamore. Small silencing RNAs: an expanding universe. *Nat Rev Genet*, 10(2):94–108, Feb. 2009. 1.3, 1.3.1, 1.3.2

[75] A. C. Goldstrohm, B. A. Hook, D. J. Seay, and M. Wickens. PUF proteins bind Pop2p to regulate messenger RNAs. *Nat Struct Mol Biol*, 13(6):533–539, June 2006. 1.2.1

[76] G. Golub and C. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17:883–893, 1980. 11.3

[77] J. Grigull, S. Mnaimneh, J. Pootoolal, M. D. Robinson, and T. R. Hughes. Genome-Wide Analysis of mRNA Stability Using Transcription Inhibitors and Microarrays Reveals Post-transcriptional Control of Ribosome Biogenesis Factors. *Mol. Cell. Biol.*, 24(12):5534–5547, June 2004. 2, 3, 3.2, 5, 15, 15.4, 16.3, 46

[78] R. A. Gutierrez, R. M. Ewing, J. M. Cherry, and P. J. Green. Identification of unstable transcripts in Arabidopsis by cDNA microarray analysis: rapid decay is associated with a group of touch- and specific clock-controlled genes. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17):11513–11518, Aug. 2002. 3, 3.2, 5, 15

[79] S. Hahn. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*, 11(5):394–403, May 2004. 1.1, 2, 1.1.1

[80] R. E. Halbeisen and A. P. Gerber. Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. *PLoS biology*, 7(5):e1000105+, May 2009. 15.11

[81] B. Haley and P. D. Zamore. Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol*, 11(7):599–606, July 2004. 1.3.2

[82] L. Harel-Sharvit, N. Eldad, G. Haimovich, O. Barkai, L. Duek, and M. Choder. RNA Polymerase II Subunits Link Transcription and mRNA Decay to Translation. *Cell*, 143(4):552–563, Nov. 2010. 1, 16, 16.8

[83] L. He and G. J. Hannon. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*, 5(7):522–531, July 2004. 1.3.1, 1.3.2

[84] D. Hekstra, A. R. Taussig, M. Magnasco, and F. Naef. Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic acids research*, 31(7):1962–1968, Apr. 2003. 3.3, 42

[85] G. A. Held, G. Grinstein, and Y. Tu. Modeling of DNA microarray data by using physical properties of hybridization. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13):7575–7580, June 2003. 3.3, 42

[86] G. A. Held, G. Grinstein, and Y. Tu. Relationship between gene expression and observed intensities in DNA microarrays–a modeling study. *Nucleic acids research*, 34(9):e70, 2006. 3.3, 42

[87] L. M. Hereford and M. Rosbash. Number and distribution of polyadenylated RNA sequences in yeast. *Cell*, 10(3):453–462, Mar. 1977. 15.3, 16.2, 1

[88] J. A. Hodson, J. M. Bailis, and S. L. Forsburg. Efficient labeling of fission yeast Schizosaccharomyces pombe with thymidine and BUdR. *Nucleic acids research*, 31(21), Nov. 2003. 4.1, 15.1, 20.1

[89] S. Hohmann, M. Krantz, and B. Nordlander. Yeast osmoregulation. *Methods in enzymology*, 428:29–45, 2007. 15

[90] F. C. Holstege, E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander, and R. A. Young. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5):717–728, Nov. 1998. 2, 3, 5, 15, 15.4, 16, 16.3, 46

[91] J. Houseley and D. Tollervey. The Many Pathways of RNA Degradation. *Cell*, 136(4):763–776, Feb. 2009. 3, 1.2, 1.2.1, 1.2.2, 1.3, 1.3.2, 4, 16.6

[92] C. L. Hsu and A. Stevens. Yeast cells lacking 5' to 3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Molecular and Cellular Biology*, 13(8):4826–4835, Aug. 1993. 1.2.2

[93] W. Hu, T. J. Sweet, S. Chamnongpol, K. E. Baker, and J. Coller. Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature*, 461(7261):225–229, Sept. 2009. 1.2.2, 16, 16.8

[94] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, and M. Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18(suppl 1):S96–S104, July 2002. 13.1

[95] G. Hutvágner, J. McLachlan, A. E. Pasquinelli, E. Bálint, T. Tuschl, and P. D. Zamore. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science (New York, N.Y.)*, 293(5531):834–838, Aug. 2001. 1.3.1

[96] G. Hutvágner, M. J. Simard, C. C. Mello, and P. D. Zamore. Sequence-specific inhibition of small RNA function. *PLoS biology*, 2(4):e98+, Apr. 2004. 17

[97] J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai. Revealing modular organization in the yeast transcriptional network. *Nature genetics*, 31(4):370–377, Aug. 2002. 15.9

[98] O. Isken and L. E. Maquat. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet*, 9(9):699–712, Sept. 2008. 1.2

[99] S. Jimeno-González, L. L. Haaning, F. Malagon, and T. H. Jensen. The Yeast 5' to 3' Exonuclease Rat1p Functions during Transcription Elongation by RNA Polymerase II. *Molecular Cell*, 37(4):580–587, Feb. 2010. 16.5

[100] P. Jorgensen, J. L. Nishikawa, B.-J. J. Breitkreutz, and M. Tyers. Systematic identification of pathways that couple cell growth and division in yeast. *Science (New York, N.Y.)*, 297(5580):395–400, July 2002. 16.4

[101] L. D. Kapp and J. R. Lorsch. The molecular mechanics of eukaryotic translation. *Annual Review of Biochemistry*, 73(1):657–704, 2004. 1.2.1

[102] M. Kenzelmann, S. Maertens, M. Hergenhahn, S. Kueffer, A. Hotz-Wagenblatt, L. Li, S. Wang, C. Ittrich, T. Lemberger, R. Arribas, S. Jonnakuty, M. C. Hollstein, W. Schmid, N. Gretz, H. J. Gröne, and G. Schütz. Microarray analysis of newly synthesized RNA in cells and animals. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15):6164–6169, Apr. 2007. 4.1, 5, 15, 15.1

[103] R. F. Ketting, S. E. J. Fischer, E. Bernstein, T. Sijen, G. J. Hannon, and R. H. A. Plasterk. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Development*, 15(20):2654–2659, Oct. 2001. 1.3.1

[104] M. Kim, N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522, Nov. 2004. 1.1.3

[105] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. *Systems Biology: A Textbook*. Wiley, 2009. 3.3

[106] D. Kostrewa, M. E. Zeller, K.-J. Armache, M. Seizl, K. Leike, M. Thomm, and P. Cramer. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271):323–330, Nov. 2009. 1.1.1

[107] A. Krek, D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, and N. Rajewsky. Combinatorial microRNA target predictions. *Nat Genet*, 37(5):495–500, May 2005. 1.3, 17

[108] J. A. Kruk, A. Dutta, J. Fu, D. S. Gilmour, and J. C. Reese. The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes & development*, 25(6):581–593, Mar. 2011. 16.8

[109] J. N. Kuehner, E. L. Pearson, and C. Moore. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol*, 12(5):283–294, May 2011. 1.1.3

[110] E. C. Lai. Predicting and validating microRNA targets. *Genome biology*, 5(9), 2004. 1.3.2

[111] L. Lam, O. Pickeral, A. Peng, A. Rosenwald, E. Hurt, J. Giltnane, L. Averett, H. Zhao, R. E. Davis, M. Sathyamoorthy, L. Wahl, E. Harris, J. Mikovits, A. Monks, M. Hollingshead, E. Sausville, and L. Staudt. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, 2(10), 2001. 2, 5

[112] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V. N. Kim. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956):415–419, Sept. 2003. 1.3.1

[113] Y. Lee, K. Jeon, J.-T. T. Lee, S. Kim, and V. N. Kim. MicroRNA maturation: stepwise processing and subcellular localization. *The EMBO journal*, 21(17):4663–4670, Sept. 2002. 1.3.1

[114] Y. Lee, M. Kim, J. Han, K.-H. Yeom, S. Lee, S. H. Baek, and V. N. Kim. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO Journal*, 23(20):4051–4060, Sept. 2004. 1.3.1

[115] B. P. Lewis, C. B. Burge, and D. P. Bartel. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell*, 120(1):15–20, Jan. 2005. 1.3, 1.3.2, 17

[116] B. P. Lewis, I.-h. Shih, M. W. Jones-Rhoades, D. P. Bartel, and C. B. Burge. Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7):787–798, Dec. 2003. 1.3.2

[117] X. Li, J. J. Cassidy, C. A. Reinke, S. Fischboeck, and R. W. Carthew. A MicroRNA Imparts Robustness against Environmental Fluctuation during Development. *Cell*, 137(2):273–282, Apr. 2009. 1.3

[118] H. Liu, N. D. Rodgers, X. Jiao, and M. Kiledjian. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. *The EMBO journal*, 21(17):4699–4708, Sept. 2002. 1.2.2

[119] H. Y. Liu, V. Badarinarayana, D. C. Audino, J. Rappsilber, M. Mann, and C. L. Denis. The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *The EMBO journal*, 17(4):1096–1106, Feb. 1998. 16, 16.8

[120] D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Norton, and E. L. Brown. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotech*, 14(13):1675–1680, Dec. 1996. 3.3

[121] M. S. Longtine, A. Mckenzie, D. J. Demarini, N. G. Shah, A. Wach, A. Brachat, P. Philippsen, and J. R. Pringle. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, 14(10):953–961, July 1998. 20.1

[122] R. Lotan, V. G. Bar-On, L. Harel-Sharvit, L. Duek, D. Melamed, and M. Choder. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes & Development*, 19(24):3004–3016, Dec. 2005. 1, 16, 16.8

[123] R. Lotan, V. Goler-Baron, L. Duek, G. Haimovich, and M. Choder. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *The Journal of Cell Biology*, 178(7):1133–1143, Sept. 2007. 1, 16, 16.8

[124] J. Macia, S. Regot, T. Peeters, N. Conde, R. Sole, and F. Posas. Dynamic Signaling in the Hog1 MAPK Pathway Relies on High Basal Signal Transduction. *Sci. Signal.*, 2(63):ra13+, Mar. 2009. 15, 15.6

[125] F. Malagon, M. L. Kireeva, B. K. Shafer, L. Lubkowska, M. Kashlev, and J. N. Strathern. Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics*, 172(4):2201–2209, Apr. 2006. 16.5, 20.1

[126] D. A. Mangus, N. Amrani, and A. Jacobson. Pbp1p, a factor interacting with Saccharomyces cerevisiae poly(A)-binding protein, regulates polyadenylation. *Molecular and cellular biology*, 18(12):7383–7396, Dec. 1998. 1.2.1

[127] T. Maniatis and R. Reed. An extensive network of coupling among gene expression machines. *Nature*, 416(6880):499–506, Apr. 2002. 16.8

[128] T. Margaritis and F. C. P. Holstege. Poised RNA Polymerase II Gives Pause for Thought. *Cell*, 133(4):581–584, May 2008. 1.1.2

[129] A. Mayer, M. Lidschreiber, M. Siebert, K. Leike, J. Soding, and P. Cramer. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol*, 17(10):1272–1278, Oct. 2010. 1.1.2, 15.10, 15.11, 19.4

[130] C. Mazzoni and C. Falcone. mRNA stability and control of cell proliferation. *Biochemical Society transactions*, 39(5):1461–1465, Oct. 2011. 1.2

[131] A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, and P. Cramer. A structural perspective of CTD function. *Genes & Development*, 19(12):1401–1415, June 2005. 1.1.2

[132] D. Melamed, L. Pnueli, and Y. Arava. Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *RNA (New York, N.Y.)*, 14(7):1337–1351, July 2008. 15

[133] W. T. Melvin, H. B. Milne, A. A. Slater, H. J. Allen, and H. M. Keir. Incorporation of 6-Thioguanosine and 4-Thiouridine into RNA. *European Journal of Biochemistry*, 92(2):373–379, 1978. 4.1, 15

[134] J. T. Mettetal, D. Muzzey, C. Gómez-Uribe, and A. van Oudenaarden. The Frequency Dependence of Osmo-Adaptation in Saccharomyces cerevisiae. *Science*, 319(5862):482–484, Jan. 2008. 15.11

[135] A. S. M.G. Kendall. *The advanced theory of statistics*. Griffin, 1987. 18.2

[136] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. E. Martin, A. Tresch, and P. Cramer. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology*, 7, Jan. 2011. 4, 6, 4.1, 7, 11.5, 15.1, 36, 15.2, 15.3, 15.4, 15.5, 15.7, 15.8, 15.9, 41, 15.11, 16.1, 16.2, 16.8, 19.1, 19.3, 19.4, 20.2

[137] M. R. Miller, K. J. Robinson, M. D. Cleary, and C. Q. Doe. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Meth*, 6(6):439–441, June 2009. 15

[138] P. Mitchell and D. Tollervey. Musing on the structural organization of the exosome complex. *Nat Struct Mol Biol*, 7(10):843–846, Oct. 2000. 1.2.2

[139] C. Molin, A. Jauhiainen, J. Warringer, O. Nerman, and P. Sunnerhagen. mRNA stability changes precede changes in steady-state mRNA amounts during hyperosmotic stress. *RNA (New York, N.Y.)*, 15(4):600–614, Apr. 2009. 15, 15.11

[140] D. Muhlrad and R. Parker. Premature translational termination triggers mRNA decapping. *Nature*, 370(6490):578–581, Aug. 1994. 1.2.2

[141] S. E. Munchel, R. K. Shultzaberger, N. Takizawa, and K. Weis. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Molecular Biology of the Cell*, 22(15):2787–2795, Aug. 2011. 4.1, 16.1, 16.2, 16.3, 46

[142] D. Muzzey, C. A. Gómez-Uribe, J. T. Mettetal, and A. van Oudenaarden. A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation. *Cell*, 138(1):160–171, July 2009. 15.11

[143] R. Narsai, K. A. Howell, A. H. Millar, N. O'Toole, I. Small, and J. Whelan. Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *The Plant cell*, 19(11):3418–3436, Nov. 2007. 3, 3.2, 5, 15

[144] S. Nechaev and K. Adelman. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1809(1):34–45, Jan. 2011. 1.1.2

[145] P. Nelson, M. Radosavljević, and S. Bromberg. *Biological Physics: Energy, Information, Life*. W.H. Freeman and Company, 2007. 2

[146] F. R. Neumann and P. Nurse. Nuclear size control in fission yeast. *The Journal of cell biology*, 179(4):593–600, Nov. 2007. 16.4

[147] D. B. Nikolov, H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, 377(6545):119–128, Sept. 1995. 1.1.1

[148] M. Nonet, C. Scafe, J. Sexton, and R. Young. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Molecular and cellular biology*, 7(5):1602–1611, May 1987. 3.1, 16.3

[149] I. Novoa, J. Gallego, P. G. Ferreira, and R. Mendez. Mitotic cell-cycle progression is regulated by CPEB1 and CPEB4-dependent translational control. *Nat Cell Biol*, 12(5):447–456, May 2010. 1.2.1

[150] S.-E. Ong and M. Mann. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, 1(6):2650–2660, Jan. 2007. 1

[151] H. Y. Y. Park, A. R. Buxbaum, and R. H. Singer. Single mRNA tracking in live cells. *Methods in enzymology*, 472:387–406, 2010. 15.3

[152] R. Parker and U. Sheth. P Bodies and the Control of mRNA Translation and Degradation. *Molecular Cell*, 25(5):635–646, Mar. 2007. 1.2, 1.2.3

[153] R. Parker and H. Song. The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol*, 11(2):121–127, Feb. 2004. 1.2, 1.2.1, 1.2.2

[154] V. Pelechano, S. Jimeno-González, A. Rodríguez-Gil, J. García-Martínez, J. E. Pérez-Ortín, and S. Chávez. Regulon-specific control of transcription elongation across the yeast genome. *PLoS genetics*, 5(8), Aug. 2009. 15.11

[155] V. Pelechano and J. E. Pérez-Ortín. There is a steady-state transcriptome in exponentially growing yeast cells. *Yeast*, 27(7):413–422, 2010. 5, 15.3

[156] H. P. Phatnani and A. L. Greenleaf. Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development*, 20(21):2922–2936, Nov. 2006. 1.1, 1.1.2

[157] M. Proft and K. Struhl. MAP Kinase-Mediated Stress Relief that Precedes and Regulates the Timing of Transcriptional Induction. *Cell*, 118(3):351–361, Aug. 2004. 15, 15.9, 15.11

[158] N. J. Proudfoot. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences*, 14(3):105–110, Mar. 1989. 1.1.3

[159] N. J. Proudfoot, A. Furger, and M. J. Dye. Integrating mRNA Processing with Transcription. *Cell*, 108(4):501–512, Feb. 2002. 1.1.2

[160] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0. 6.1, 6.2, 18.1, 20.3

[161] M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, and A. Regev. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotech*, 29(5):436–442, May 2011. 20.3

[162] A. Raghavan, R. L. Ogilvie, C. Reilly, M. L. Abelson, S. Raghavan, J. Vasdewani, M. Krathwohl, and P. R. Bohjanen. Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic acids research*, 30(24):5529–5538, Dec. 2002. 3, 3.2, 5, 15

[163] N. Rajewsky. microRNA target predictions in animals. *Nat Genet*, 38 Suppl 1(6s):S8–S13, May 2006. 17

[164] P. Richard and J. L. Manley. Transcription termination by nuclear RNA polymerases. *Genes & Development*, 23(11):1247–1269, June 2009. 1.1.3

[165] A. Rodríguez-Gil, J. García-Martínez, V. Pelechano, M. d. l. C. d. e. . L. Muñoz Centeno, V. Geli, J. E. Pérez-Ortín, and S. Chávez. The distribution of active RNA polymerase II along the transcribed region is gene-specific and controlled by elongation factors. *Nucleic acids research*, 38(14):4651–4664, Aug. 2010. 15.11

[166] L. Romero-Santacreu, J. Moreno, J. E. Pérez-Ortín, and P. Alepuz. Specific and global regulation of mRNA stability during osmotic stress in Saccharomyces cerevisiae. *RNA (New York, N.Y.)*, 15(6):1110–1120, June 2009. 15, 15.11

[167] J. G. Ruby, A. Stark, W. K. Johnston, M. Kellis, D. P. Bartel, and E. C. Lai. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Research*, 17(12):1850–1864, Dec. 2007. 17.1

[168] B. Schwalb, D. Schulz, M. Sun, B. Zacher, S. Dümcke, D. E. Martin, P. Cramer, and A. Tresch. Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA). *Bioinformatics*, 28(6):884–885, Mar. 2012. 4, 6.3

[169] B. Schwalb, A. Tresch, and R. Francois. *LSD: Lots of Superior Depictions*, 2011. R package version 2.5. 6.2, 6.3, 9

[170] B. Schwalb, B. Zacher, S. Duemcke, and A. Tresch. *DTA: Dynamic Transcriptome Analysis*, 2011. R package version 2.0.2. 5, 6.2, 6.3, 34, IV, 20.3

[171] B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, May 2011. 1, 2

[172] H. Seitz. Redefining MicroRNA Targets. *Current Biology*, 19(10):870–873, May 2009. 17

[173] M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209):58–63, Sept. 2008. 1.3.2, 17

[174] D. W. Selinger, R. M. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow. Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome research*, 13(2):216–223, Feb. 2003. 3, 3.2, 5, 15

[175] M. Selitrennik, L. Duek, R. Lotan, and M. Choder. Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. *Eukaryotic cell*, 5(12):2092–2103, Dec. 2006. 16.8

[176] O. Shalem, O. Dahan, M. Levo, M. R. Martinez, I. Furman, E. Segal, and Y. Pilpel. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular Systems Biology*, 4(1), Oct. 2008. 2, 3, 5, 15, 15.4, 16.3, 46

[177] U. Sheth and R. Parker. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science*, 300(5620):805–808, May 2003. 1.2.3

[178] A. B. Shyu, M. E. Greenberg, and J. G. Belasco. The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. *Genes & Development*, 3(1):60–72, Jan. 1989. 3

[179] T. W. Sikorski and S. Buratowski. The basal initiation machinery: beyond the general transcription factors. *Current Opinion in Cell Biology*, 21(3):344–351, June 2009. 1.1.1

[180] D. Skvortsov, D. Abdueva, C. Curtis, B. Schaub, and S. Tavaré. Explaining differences in saturation levels for Affymetrix GeneChip arrays. *Nucleic acids research*, 35(12):4154–4163, June 2007. 3.3, 42

[181] G. K. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3(1), Feb. 2004. 18.3

[182] G. K. Smyth and T. Speed. Normalization of cDNA microarray data. *Methods*, 31(4):265–273, Dec. 2003. 19.1

[183] B. Soufi, C. D. Kelstrup, G. Stoehr, F. Frohlich, T. C. Walther, and J. V. Olsen. Global analysis of the yeast osmotic stress response by quantitative proteomics. *Mol. BioSyst.*, 5(11):1337–1346, Nov. 2009. 15.11

[184] A. Stark, J. Brennecke, R. B. Russell, and S. M. Cohen. Identification of Drosophila MicroRNA targets. *PLoS biology*, 1(3):e60+, Dec. 2003. 17

[185] E. J. Steinmetz, C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari, and D. A. Brow. Genome-Wide Distribution of Yeast RNA Polymerase II and Its Control by Sen1 Helicase. *Molecular Cell*, 24(5):735–746, Dec. 2006. 15.11

[186] K. Struhl. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell*, 98(1):1–4, July 1999. 1.1.1

[187] M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivière, K. C. Maier, M. Seizl, A. Tresch, and P. Cramer. Comparative Dynamic Transcriptome Analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Research*, Mar. 2012. 4, 8, 16.1, 42, 16.2, 16.2, 16.4, 16.4, 16.5, 16.6, 20.3

[188] J. Q. Svejstrup. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1677(1-3):64–73, Mar. 2004. 1.1, 2

[189] J. F. Sydow, F. Brueckner, A. C. M. Cheung, G. E. Damsma, S. Dengl, E. Lehmann, D. Vassylyev, and P. Cramer. Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Molecular Cell*, 34(6):710–721, June 2009. 15.1, 19.2

[190] H. Tani, R. Mizutani, K. A. Salam, K. Tano, K. Ijiri, A. Wakamatsu, T. Isogai, Y. Suzuki, and N. Akimitsu. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Research*, 22(5):947–956, May 2012. IV

[191] D. Teixeira and R. Parker. Analysis of P-body assembly in Saccharomyces cerevisiae. *Molecular biology of the cell*, 18(6):2274–2287, June 2007. 15.11

[192] M. C. Thomas and C.-M. M. Chiang. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, 41(3):105–178, Jan. 2006. 1.1.1

[193] I. Tirosh and N. Barkai. Two strategies for gene regulation by promoter nucleosomes. *Genome Research*, 18(7):1084–1091, July 2008. 1.1.1

[194] M. Tucker, R. R. Staples, M. A. Valencia-Sanchez, D. Muhlrad, and R. Parker. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in Saccharomyces cerevisiae. *The EMBO journal*, 21(6):1427–1436, Mar. 2002. 1.2.1, 16.6, 16.8

[195] M. Tucker, M. A. Valencia-Sanchez, R. R. Staples, J. Chen, C. L. Denis, and R. Parker. The Transcription Factor Associated Ccr4 and Caf1 Proteins Are Components of the Major Cytoplasmic mRNA Deadenylase in Saccharomyces cerevisiae. *Cell*, 104(3):377–386, Feb. 2001. 1.2.1, 16

[196] Y. Uesono and A. Toh-e. Transient Inhibition of Translation Initiation by Osmotic Stress. *Journal of Biological Chemistry*, 277(16):13848–13855, Apr. 2002. 15

[197] M. A. Valencia-Sanchez, J. Liu, G. J. Hannon, and R. Parker. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development*, 20(5):515–524, Mar. 2006. 1.3, 17

[198] J. van de Peppel, P. Kemmeren, H. van Bakel, M. Radonjic, D. van Leenen, and F. C. Holstege. Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO reports*, 4(4):387–393, Apr. 2003. 16

[199] A. van Hoof and R. Parker. The Exosome. *Cell*, 99(4):347–350, Nov. 1999. 1.2.2

[200] S. Vasudevan, Y. Tong, and J. A. Steitz. Switching from repression to activation: microRNAs can up-regulate translation. *Science (New York, N.Y.)*, 318(5858):1931–1934, Dec. 2007. 1.3

[201] B. J. Venters and B. F. Pugh. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research*, 19(3):360–371, Mar. 2009. 15.11

[202] B. J. Venters and B. F. Pugh. How eukaryotic genes are transcribed. *Critical reviews in biochemistry and molecular biology*, 44(2-3):117–141, June 2009. 1.1.1

[203] B. J. Venters, S. Wachi, T. N. Mavrich, B. E. Andersen, P. Jena, A. J. Sinnamon, P. Jain, N. S. Rolleri, C. Jiang, C. Hemeryck-Walsh, and B. F. Pugh. A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces. *Molecular Cell*, 41(4):480–492, Feb. 2011. 1.1.1

[204] E. G. W. S. Cleveland and W. M. Shyu. *Local regression models.* 1992. 18.1

[205] Y. Wang, C. L. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, 99(9):5860–5865, Apr. 2002. 2, 3, 5, 15, 15.4, 16, 16.3, 46

[206] J. Warringer, M. Hult, S. Regot, F. Posas, and P. Sunnerhagen. The HOG pathway dictates the short-term translational response after hyperosmotic shock. *Molecular biology of the cell*, 21(17):3080–3092, Sept. 2010. 15.11

[207] Y. Wen and A. J. Shatkin. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & Development*, 13(14):1774–1779, July 1999. 1.1.2

[208] M. Wickens, D. S. Bernstein, J. Kimble, and R. Parker. A PUF family portrait: 3'UTR regulation as a way of life. *Trends in Genetics*, 18(3):150–157, Mar. 2002. 1.2.1

[209] K. Wiederhold and L. A. Passmore. Cytoplasmic deadenylation: regulation of mRNA fate. *Biochemical Society transactions*, 38(6):1531–1536, Dec. 2010. 3, 1.2, 1.2.1, 1.2.2

[210] T. Wild and P. Cramer. Biogenesis of multisubunit RNA polymerases. *Trends in Biochemical Sciences*, 37(3):99–105, Mar. 2012. 1.1

[211] L. Wodicka, H. Dong, M. Mittmann, M.-H. Ho, and D. J. Lockhart. Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nat Biotech*, 15(13):1359–1367, Dec. 1997. 3.3

[212] L. Wu and J. G. Belasco. Let Me Count the Ways: Mechanisms of Gene Regulation by miRNAs and siRNAs. *Molecular Cell*, 29(1):1–7, Jan. 2008. 1.3, 1.3.2

[213] Z. Wu, R. A. Irizarry, R. Gentleman, F. M. Murillo, and F. Spencer. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association*, 99(468), 2004. 19.1, 20.3

[214] X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, and M. Kellis. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434(7031):338–345, Mar. 2005. 1.3

[215] E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J. E. Darnell. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome research*, 13(8):1863–1872, Aug. 2003. 3, 3.2, 5, 15

[216] N. Yudkovsky, J. A. Ranish, and S. Hahn. A transcription reinitiation intermediate that is stabilized by activator. *Nature*, 408(6809):225–229, Nov. 2000. 1.1.3

[217] Y. R. Yu.V. Prohorov. *Probability theory, basic concepts. Limit theorems, random processes.* Springer, 1969. 18.2

[218] B. Zacher, P. Kuan, and A. Tresch. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics*, 11(1):194+, 2010. 19.4

[219] D. Zenklusen, D. R. Larson, and R. H. Singer. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol*, 15(12):1263–1271, Dec. 2008. 16.2, 1

[220] J. Zhurinsky, K. Leonhard, S. Watt, S. Marguerat, J. Bähler, and P. Nurse. A Coordinated Global Control over Cellular Transcription. *Current Biology*, 20(22):2010–2015, Nov. 2010. 16.8

List of Figures

Results & Discussion   63

List of Tables

Results & Discussion   63

Curriculum vitae

Björn Schwalb

Personal Information

Date of Birth   29. October 1979

Place of Birth   Heilbronn

Citizenship   German

Doctorate

since 11/2008   **Graduate studies in Bioinformatics**
Gene Center, Ludwig-Maximilians-Universität Munich Munich Center of
Excellence
Supervisors: Prof. Dr. Patrick Cramer and Prof. Dr. Achim Tresch.
PhD thesis: „Dynamic transcriptome analysis (DTA): Kinetic modeling of
synthesis and decay of mRNA transcripts upon perturbation in *S.cerevisiae*,
S.pombe and *D.melanogaster*"
Development of a dynamic kinetic modeling approach to yield precise rates
for mRNA synthesis and decay based upon DTA measurements facilitated by
bias correction, validation and simulation studies.

07/2012   **Graduation**
dissertation (grade 0.7)
disputation (grade 0.7)
doctorate (overall grade: summa cum laude).

Studies

10/2002 – 07/2008	**Undergraduate studies in mathematics** at the Johann Wolfgang Goethe Universität Frankfurt am Main.
07/2008	**Diploma** pure mathematics (grade 1.0) applied mathematics (grade 1.0) special subject (grade 1.0) subsidiary subject: economics (grade 2.0) Diploma (overall grade: 1.0).
09/2007 – 03/2008	**Diploma thesis** advised by Prof. Dr. Jürgen Wolfart Title: „On the existence of intermediate groups in multi-level inclusions of triangle groups" (grade: 1.0).
04/2006 – 07/2008	**Student assistant** Analysis III, Prof. Dr. M. Reichert-Hahn Johann Wolfgang Goethe Universität Frankfurt am Main.
10/2005 – 02/2006	**Student assistant** Analysis II, Prof. Dr. M. Reichert-Hahn Johann Wolfgang Goethe Universität Frankfurt am Main.
04/2005 – 07/2005	**Student assistant** Linear algebra I, Prof. Dr. G. Burde Johann Wolfgang Goethe Universität Frankfurt am Main.
10/2004 – 02/2005	**Student assistant** Mathematics for physicists I, Prof. Dr. J. Baumeister Johann Wolfgang Goethe Universität Frankfurt am Main.
10/2004	**Intermediate diploma**

Education / Civilian service

09/2000 – 07/2001	**Civilian service** Dialysis centre Heilbronn
08/1997 – 06/2000l	**Secondary school** Gustav von Schmoller Schule in Heilbronn General qualification for university entrance

Publications

2011 **Drosophila miR-277 controls TOR activity and lifespan via tuning of branched-chain amino acid metabolism**
S. Esslinger, **B. Schwalb**, S. Helfer, H. Witte, K. Maier, D. Martin, B. Michalke, A. Tresch, P. Cramer, K. Förstemann
Manuscript in preparation.

2011 **Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA)**
B. Schwalb, D. Schulz, M. Sun, B. Zacher, S. Dümcke, D. Martin, P. Cramer, A. Tresch
Bioinformatics, Vol. 28, No. 6, pp. 884-885.
http://bioinformatics.oxfordjournals.org/content/28/6/884

2011 **Comparative Dynamic Transcriptome Analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation**
M. Sun*, **B. Schwalb***, D. Schulz*, N. Pirkl, S. Etzold, L. Larivière, K. Maier, M. Seizl, A. Tresch, P. Cramer (* joint first authorship)
Genome Research.
http://genome.cshlp.org/content/early/2012/03/30/gr.130161.111

2011 **Mediator phosphorylation is involved in stress response**
C. Miller, I. Matic, K. Maier, S. Roether, **B. Schwalb**, K. Straesser, A. Tresch, M. Mann, P. Cramer
Manuscript in preparation.

2010 **Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast**
C. Miller*, **B. Schwalb***, K. Maier*, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. Martin, A. Tresch, P. Cramer (* joint first authorship)
Molecular Systems Biology 7, 458.
http://www.nature.com/msb/journal/v7/n1/full/msb2010112.html

2008 **On the existence of intermediate groups in multi-level inclusions of triangle groups**
B. Schwalb, Publications system Johann Wolfgang Goethe Universität Frankfurt am Main
http://publikationen.ub.uni-frankfurt.de/volltexte/2008/5933/.

Software

2011   **R/Bioconductor package 'DTA Dynamic Transcriptome Analysis'**
B. Schwalb, B. Zacher, S. Duemcke, A. Tresch
The Comprehensive R Archive Network and Bioconductor
http://www.bioconductor.org/packages/release/bioc/html/DTA.html.
Implements the DTA kinetic modeling approach capable of the precise
determination of synthesis and decay rates from individual microarray or
RNAseq measurements.

2010   **R package 'LSD Lots of Superior Depictions'**
B. Schwalb, R. Francois, A. Tresch
The Comprehensive R Archive Network
http://cran.r-project.org/web/packages/LSD/index.html.
Collection of sophisticated plotting routines to create high level plots in a
plethora of variations.