Predicting proteome dynamics using gene expression data

Krzysztof Kuchta1, Joanna Towpik1, Anna Biernacka1, Jan Kutner1, Andrzej Kudlicki2, Krzysztof Ginalski1 & Maga Rowicka2

While protein concentrations are physiologically most relevant, measuring them globally is challenging. mRNA levels are easier to measure genome-wide and hence are typically used to infer the corresponding protein abundances. The steady-state condition (assumption that protein levels remain constant) has typically been used to calculate protein concentrations, as it is mathematically convenient, even though it is often not satisfied. Here, we propose a method to estimate genome-wide protein abundances without this assumption. Instead, we assume that the system returns to its baseline at the end of the experiment, which is true for cyclic phenomena (e.g. cell cycle) and many time-course experiments. Our approach only requires availability of gene expression and protein half-life data. As proof-of-concept, we predicted proteome dynamics associated with the budding yeast cell cycle, the results are available for browsing online at http://dynprot.cent.uw.edu.pl/. The approach was validated experimentally by verifying that the predicted protein concentration changes were consistent with measurements for all proteins tested. Additionally, if proteomic data are available as well, we can also infer changes in protein half-lives in response to posttranslational regulation, as we did for Clb2, a post-translationally regulated protein. The predicted changes in Clb2 abundance are consistent with earlier observations.

Measuring protein abundance provides information that is not apparent from gene expression data but is crucial for the description of the state of a biological system1. Nevertheless, measured mRNA concentrations are often used to linearly approximate the corresponding protein levels, even though such approximation can be very imprecise1. However, mRNA levels (unlike protein abundances) are relatively easy to determine due to RNA and DNA base pair complementarity, which enables precise and high-throughput measurements, such as sequencing and microarrays. Measuring protein levels remains more challenging, due to the different chemical properties of proteins and wide dynamical range of protein abundances. Studies have shown that protein levels cannot be determined from mRNA levels just by correlation1–6. For example, similar mRNA expression levels can be accompanied by a wide range (up to 20-fold difference) of protein abundances and vice versa1.

The relation between mRNA concentration, \([mRNA_i(t)]\), and protein concentration, \([P_i(t)]\), of protein \(i\) can be described in the first approximation by a kinetic equation:

\[
\frac{d[P_i(t)]}{dt} = k_{\text{trans},i} \cdot [mRNA_i(t)] - k_{d,i}[P_i(t)],
\]

where \(k_{d,i} = \frac{\ln(2)}{\tau_i}\), and \(\phi_{d,i}\), \(k_{d,i}\), and \(k_{\text{trans},i}\) are half-life, degradation rate, and translation rate, respectively. Data regarding mRNA levels, protein abundances, degradation rates, and translation rates are required to solve Eq. 1. Among these, only translation rates are not readily available for most model organisms. Eq. 1 is typically solved using the steady-state assumption, which is the easiest mathematical way to solve it, but it is also the least physiologically relevant, since the concentrations of many important proteins and their mRNAs change dynamically. Therefore, instead of using the steady-state assumption, we propose to solve Eq. 1 using alternative boundary conditions: that both mRNA and protein levels will be the same at time 0 and at the certain time T at the end of experiment. Such a condition should be fulfilled in a typical control versus treatment experiment, at the time when treatment wears off as the cells go back to their original (control) state. Here, as proof-of-concept, we
DNA damage repair, genetic recombination during meiosis, and for telomere maintenance

...the simplicity of this model, it has been shown to accurately capture the dynamical changes in protein abundances for a majority of human proteins. These results suggest that the model is suitable for other eukaryotic systems (like *S. cerevisiae*) as well.

As described in detail in Materials and Methods, protein concentration and translation rate can be calculated from a time-course of its gene-expression measurements and its average abundance. As proof-of-concept, we chose five different *S. cerevisiae* cell cycle synchronized gene expression data sets (Table 1): alpha (3395 proteins), brd26 (2840 proteins), brd30 (2699 proteins), brd38 (2751 proteins), cdc15 (3173 proteins) and cdc28 (3424 proteins). First, we used the periodogram to estimate the consensus period for periodically expressed cell cycle genes in each of these data sets (Materials and Methods and Table 1). Second, we mathematically pre-processed raw data on yeast protein half-lives, to remove negative values and improve overall accuracy of half-life estimates (Materials and Methods). Next, we used an existing compendium of the budding yeast mRNA and protein consensus levels to estimate these levels in our conditions (Materials and Methods). Finally, we numerically solved Eq. 1, using the Fixed Point Iteration method, for all periodically expressed proteins in these five data sets. This resulted in predicted time-courses of dynamic protein abundances, with 1-minute resolution during the whole cell cycle, for all budding yeast proteins available in each of five different data sets. All predicted dynamic protein concentrations and translation rates can be browsed, compared, and downloaded via our web server (http://dynprot.cent.uw.edu.pl/).

**Validation of predicted dynamic protein abundances.** In order to verify the temporal protein levels calculated using our model, we utilized western blotting to measure the actual protein concentrations for five representative proteins in cell cycle synchronized yeast culture (Materials and Methods). Representative proteins were chosen from the three groups: (1) proteins with relatively constant mRNA levels and predicted protein levels (Fig. 1A), (2) proteins with highly variable mRNA and relatively constant predicted protein levels (Fig. 1B), (3) proteins with variable mRNA and predicted protein levels during the cell cycle (Fig. 1C). For proteins with variable mRNA levels, we also required that they were transcriptionally regulated during the yeast cell cycle to guarantee that the observed changes in their levels would be meaningful. To confirm mRNA level periodicity in the yeast cell cycle the SCEPTRANS web server was used. The choice of individual proteins within a group was based on availability of commercial antibodies. The first group is represented by Rad50p, a protein required for DNA damage repair, genetic recombination during meiosis, and for telomere maintenance (Materials and Methods).

| Data set      | S. cerevisiae strain | Cycle period | Data granul | Reference |
|---------------|----------------------|--------------|-------------|-----------|
| alpha         | DBY8724 (GAL2 urs3 bar1::URA3) | 56 min     | 7 min     | Spellman et al. |
| brd26         | BY2125 (W303-MATa ade2-1 trp1-1 can1-1000 leu2-3, 115 his3-11 ura3 ho ssd1 d) | 60 min     | 5 min     | Pramila et al. |
| brd30         | BY2125 (W303-MATa ade2-1 trp1-1 can1-1000 leu2-3, 115 his3-11 ura3 ho ssd1 d) | 60 min     | 5 min     | Pramila et al. |
| brd38         | W303 (cdc15-2)       | 60 min      | 10 min    | Spellman et al. |
| cdc15         | K3445 (YNN553) contains cdc28-13 allele | 116 min    | 10 min    | Cho et al. |
| cdc28         | K3445 (YNN553) contains cdc28-13 allele | 79 min     | 10 min    | Cho et al. |

Table 1. Pearson and Spearman correlations between average mRNA and average protein concentrations.
and colleagues' data, Cdc5 and Clb2 half-lives are 10 and 22 min, respectively. Our model predicted that Cdc5 and Clb2 concentrations would exhibit strong variability during the yeast cell cycle (Fig. 1C). Indeed, the levels of Cdc5p and Clb2p, as determined by western blotting, varied strongly, reaching peaks at 65 and 115 min (M phase), and 55 and 110 min (G2/M transition), respectively (Fig. 2C). However, assuming that Clb2 has a constant half-life of 22 min (as calculated based on the data of Belle et al.), gives less than ideal agreement of predicted protein concentrations with western blot measurements (Fig. 2C).
Extending the model to accommodate post-translational regulation. Discrepancies between predicted and experimental protein levels during the cell cycle may be caused by known inaccuracies of the western blot (up to 2-fold) or by post-translational regulation. To address this question, we also constructed a more complex model, allowing variable half-life throughout the cell cycle, to verify if considering dynamical half-lives would result in much better agreement between predictions and experimental data. We tested the expanded model on the case of Clb2, since it was the only protein tested showing discrepancy with the predicted model beyond that expected from western blot measurement errors. First, we calculated predicted Clb2 temporal abundance based on static experimental half-lives from two different studies \(^9,14\) (Fig. 3A, B). Next, we utilized our expanded model, which allows Clb2 to switch between longer and shorter half-lives depending on the stage of the cell cycle. We generated such models for Clb2 with half-lives ranging from 1 to 40 minutes, with 1-minute step, and changing throughout the cell cycle. We chose the model which best fit the western blot data, which turned out to be the model assuming a very short Clb2 half-life up to minute 30 and after the minute 55 after the alpha-factor release and longer during the rest of the cell cycle (Fig. 3C). Indeed, it was reported earlier that the Clb2 half-life was less than 1 min for cells arrested in G1 by \(\alpha\) factor \(^9,14\) and in our best-fitting models the Clb2 half-life was 1 minute (shorter values were not considered) during the G1 phase (Fig. 3C). Clb2 had a longer half-life, closer to the value measured in \(^9,14\) during the Clb2 activity window, which is at the G2/M transition. These results show another important application of our method: if half-life (and/or translation rates) are unavailable, they can be estimated with good accuracy from corresponding gene expression and proteomic time-courses, even in very challenging cases in which the half-life is variable and the protein time-course is inferred from relatively inaccurate western blots.

Correlation between mRNAs and protein abundances in time-course data. It is typically assumed that with an increase in the quality of both gene expression and proteomic data, the correlation between mRNA and protein abundance would grow. However, a significant correlation between mRNA and protein concentration can be expected only for some groups of proteins. Greenbaum et al. \(^15\) showed a significant increase in correlation between mRNA and protein levels for proteins localized in the same cell compartment or with the same MIPS functional category. O’Shea and colleagues \(^4\) later showed that proteins of similar function tend to have similar half-lives. So far, the highest correlations between mRNA and protein concentrations have been achieved by Futcher et al. \(^16\), who found relatively high correlations \((r = 0.76)\) after copula-transforming the data to normal distributions. The \(r = 0.7–0.8\) range likely represents the highest possible correlation to achieve. On the other hand, protein half-lives are known to have a dynamic range of several orders of magnitude \(^9\), and therefore even similar mRNA expression levels can be accompanied by a wide range of protein abundance levels, and vice versa \(^1\). In general, it is increasingly recognized that mRNA abundances are only a weak surrogate for the corresponding protein concentrations, mainly because of post-transcriptional control of gene expression. Our studies allow us to look deeper at this problem. We found that even though the Spearman and Pearson correlation between average protein and mRNA concentrations is highly significant (Table 2), the temporal profiles of protein and mRNA concentrations are only weakly correlated (Fig. 4), with typical correlation not higher than 0.2. As expected, the
highest correlations between temporary protein and mRNA abundances were observed for proteins with short half-lives, when protein levels follow closely behind mRNA concentrations (Fig. 5). These data show that even in the simplified case of not considering post-translational modification, mRNA levels are good estimates of temporal protein abundances during the whole cell cycle only for a handful of proteins, highlighting the usefulness of the modeling described above.

**Estimating translation rates.** Translation rate ($TR$, denoted by $k_{trans,i}$ in Eq. 1) is the output of protein production relative to the amount of mRNA. Translation rates are not easy to measure directly, and are traditionally estimated utilizing a steady-state condition ($TR_{ss}$, Material and Methods, Eq. 7). However, the steady-state assumption is usually not fulfilled in physiological conditions. Moreover, there is growing evidence that unlike the degradation rate, the translation rate is very plastic and is a mechanism to control protein abundances, in response to changing mRNA levels (e.g.17). Our approach provides a method for estimating condition-specific translation rates requiring neither the steady-state condition nor knowing protein abundance, but using time-series gene expression data instead ($TR_{tc}$, Material and Methods, Eqs 2 and 3). To compare translation rates calculated using these two different approaches we computed the relative difference between steady-state and timecourse-derived rates $TR_{diff}$ (Materials and Methods, Eq. 6), which varies from 0 to 1 depending on how different $TR_{ss}$ and $TR_{tc}$ are. We found that there are relatively few proteins for which $TR_{diff}$ is greater than 0.1 (65 out of 3395 in the alpha data set, 59 out of 2840 in the brd26 data set, 25 out of 2699 in the brd30, 2751 proteins, brd38 (2751 proteins), cdc15 (3173 proteins) and cdc28 (3424 proteins)).

| Data set | Pearson p-value | Spearman p-value | Pearson p-value | Spearman p-value |
|----------|----------------|-----------------|----------------|-----------------|
| alpha    | 0.51           | 4.0e-224        | 0.58           | 2.3e-301        |
| brd26    | 0.43           | 2.1e-127        | 0.56           | 8.2e-236        |
| brd30    | 0.47           | 5.6e-148        | 0.57           | 3.0e-231        |
| brd38    | 0.53           | 4.0e-196        | 0.56           | 6.2e-227        |
| cdc15    | 0.52           | 1.1e-218        | 0.58           | 1.4e-279        |
| cdc28    | 0.52           | 2.0e-232        | 0.58           | 3.9e-302        |

**Table 2.** Data sources.
these two proteins we obtained \(TR_{\text{diff}} > 0.1\). (b) The second likely source of differences is measurement errors of data used: here mRNA and protein concentrations and degradation rates. (c) Third, as we will discuss below, some half-lives are time-dependent and neither the steady-state nor the time-course based method used so far can fully accommodate such time dependency. Due to the very different approaches to estimating TR using either the steady-state or time-course method, it is not surprising that time-dependence of actual protein half-lives would affect these calculations in a different manner, causing the observed discrepancies. In summary, the main source of differences in translation rates we computed is related to our experimental conditions, with additional effects resulting from using time-course, not average expression values, and from measurement errors.

\(TR\) was expected to correlate with many factors known to contribute to protein production, such as protein abundance, ribosome density, ribosome occupancy, mRNA concentration, the codon adaptation index (CAI), or the tRNA adaptation index (TAI)\(^{20,21}\). However, the \(TR\) we computed (Fig. 6A) does not show high correlation with features that had been expected to be correlated with \(TR\). For example, it seems intuitive and it has been proposed in Arava et al.\(^{20}\) that \(TR\) would be proportional to the product of ribosomal density (i.e. number of ribosomes bounded to mRNA) and ribosomal occupancy (number of mRNA associated with ribosomes), denoted in Fig. 6A as TAI. However, we did not observe such correlation using the Spearman or Pearson coefficient (Fig. 6A). Although this could suggest that neither ribosomal density nor occupancy contribute meaningfully to translation rates, the lack of high positive correlation between \(TR\) and the proposed \(TR\) contributing factors is in fact the result of high standard deviations of \(k_mRNA\), the proportionality factor between translation rate and average protein concentration for the protein \(i\). Indeed, the factors mentioned earlier, that are reported as likely to correlate with \(TR\) in some publications, are highly correlated with average protein concentration (Fig. 6B). \(TR\) is associated with average protein concentration, however, this correlation is not very high (0.18 for cdc15, 0.20 for brd30 and 0.19 for others) due to the important impact of protein half-life, which can vary by at least two orders of magnitude, on protein concentration (Eq. 8). Another interesting observation is that very complex attempts at modeling translation rate, such as the Ribosomal Flow Model, do not fare better than simpler models: in our comparison the complex RFM approach of\(^{21}\) is outperformed by simpler methods.

To visualize which cell compartments and protein functions are associated with high or low half-life and translation rate, we analyzed different MIPS functional categories and localizations using SCEPTRANS webserver (Fig. 7A–D). Global analysis shows that half-lives and translation rates have almost the same levels in all functional categories. However, there are some interesting exceptions to this principle: in the cell wall and extracellular categories there are proteins with relatively short half-lives (that is high degradation rates) and high translation rates (Fig. 7B and D). Additionally, proteins involved in protein synthesis have much shorter half-lives than average (Fig. 7A).

In summary, the proposed model (Eq. 1), combined with a periodic data set (other time-course data sets can be used as well) allowed us to estimate not only genome wide changes in protein abundances, but also both translation and degradation rates of proteins. The model performs especially well in the most interesting case of substantially dynamic changes in protein abundances over time. It is also capable of detecting post-translational regulation of proteins for which corresponding time-course abundance data are available. Finally, the calculated protein concentration time-courses were validated experimentally for several proteins.
Discussion

Taking advantage of the high availability of genome-wide data of mRNA levels, we propose a model which predicts dynamic levels of protein abundances based on the time-course of gene expression levels and measured or predicted half-lives. We experimentally verified the proposed computational approach in the model organism...
of transcriptional regulation in the process (measured e.g. as concentration) is required at every step of the numerical integration. In the present case, the cell cycle data sets, smaller than a typical resolution of time-course gene expression experiments. Therefore, estimation of mRNA expression13,26,27.

tome dynamics, such as our MaxEnt model-based approach to infer high-resolution changes in gene expression and protein half-life data. Currently, all our predictions for proteome dynamics in the budding yeast study presented in this paper, to provide predicted protein time-course profiles based on user-provided gene analysis in a variety of organisms and conditions, we are developing a webserver, based on the proof-of-concept potential temporal changes to half-life can be calculated, following the approach we used for Clb2. To allow such predictions to undergo a post-translational modification, such as methylation23 or phosphorylation24, it can be flagged for potential lower accuracy of our predictions. If the corresponding proteomic timecourse is available, the case of very dynamic expression data (e.g. yeast Metabolic Cycle)25, where characteristic scale of the process is highly unstable, but also do not undergo substantial post-translational regulation in the conditions studied. As was shown in our verification, and as should be expected, depending on half-life, protein abundance profiles may show anywhere between no resemblance, to very high resemblance to the underlying mRNA expression profiles. Therefore, our predicted protein profiles can provide a valuable resource for scientists interested in dynamic changes of protein abundances during their process of interest, but who only have gene expression profiles available, which are much easier and less expensive to measure than protein levels. Moreover, if a protein is known or predicted to undergo a post-translational modification, such as methylation23 or phosphorylation24, it can be flagged for potential lower accuracy of our predictions. If the corresponding proteomic time course is available, potential temporal changes to half-life can be calculated, following the approach we used for Clb2. To allow such analysis in a variety of organisms and conditions, we are developing a webserver, based on the proof of concept study presented in this paper, to provide predicted protein time-course profiles based on user-provided gene expression and protein half-life data. Currently, all our predictions for proteome dynamics in the budding yeast in different conditions can be conveniently browsed and visualized at http://dynprot.cent.uw.edu.pl/.

To ensure the accuracy of numerical integration, the integration step ∆t in Eq. (2) should be very small, smaller than a typical resolution of time-course gene expression experiments. Therefore, estimation of mRNA concentration is required at every step of the numerical integration. In the present case of the cell cycle data sets, we obtained it from linear interpolation. Here, such approximation is justified, since the characteristic time-scales of transcriptional regulation in the process (measured e.g. as 100 nt).

Ribosome occupancy is a fraction of transcripts associated with ribosomes, i.e. engaged in translation, with values in the [0,1] interval.

Quantitative model of gene expression. Using periodic gene expression data enables us to eliminate the value of translation rate, \( k_{\text{trans,}i} \), from equation [Eq. 1]. In order to do that, we introduced the function \( R(t) \) defined as follows:

**Methods**

**Definitions.** Ribosome density is an average number of ribosomes bound to mRNA per unit of mRNA length (100 nt).

Ribosome occupancy is a fraction of transcripts associated with ribosomes, i.e. engaged in translation, with values in the [0,1] interval.

Quantitative model of gene expression. Using periodic gene expression data enables us to eliminate the value of translation rate, \( k_{\text{trans,}i} \), from equation [Eq. 1]. In order to do that, we introduced the function \( R(t) \) defined as follows:
For a small time interval $\Delta t$:

$$\int f(t)dt = \frac{1}{2}(f(t + \Delta t) + f(t)) \cdot \Delta t,$$

and the first order differential equation, [Eq. 1],

$$\frac{d[P(t)]}{dt} = k_{trans,i} \cdot [mRNA_i(t)] - k_{d,i} \cdot [P(t)],$$

can be rewritten in the form:

$$[R_i(t + \Delta t)] = \frac{2 - k_{d,i} \cdot \Delta t}{2 + k_{d,i} \cdot \Delta t} \cdot [R_i(t)] + \frac{\Delta t}{2 + k_{d,i} \cdot \Delta t} \cdot ([mRNA_i(t + \Delta t)] + [mRNA_i(t)]).$$

Detailed derivation of [Eq. 2] is provided in the Appendix. The boundary condition for the [Eq. 1]:

$$[P(t)] = [P(t + T)]$$

is equivalent to:

$$[R_i(t)] = [R_i(t + T)],$$

where $T$ is the period of the cyclic phenomenon, e.g. the length of the cell cycle. The proportionality factor $k_{trans,i}$ can be obtained from the following formula:

$$k_{trans,i} = \frac{\langle[P_i]\rangle}{\langle[R_i]\rangle},$$

where $\langle[P_i]\rangle$ and $\langle[R_i]\rangle$ are the mean values of $[P_i]$ and $[R_i]$ over time $T$, respectively.

**Data sets used and data pre-processing.** The average protein and mRNA concentrations have been taken from previous studies of Beyer et al.22. Test data sets alpha, brd26, brd30, brd38, cdc15 and cdc28 are cell-cycle synchronized gene expression data sets described in detail in Table 1. The data sets alpha and cdc15 have been published by Spellman et al.32; cdc28 by Cho et al.33 and brd26, brd30 and brd38 by Pramila et al.34. The gene expression log2 ratios, $L_i(t)$, were transformed to mRNA concentrations [molecules/cell] according to the following relation:

$$[mRNA_i(t)] = 2^{L_i(t)} \cdot \frac{\langle[M_i(t)]\rangle}{\langle[2^{L_i(t)}]\rangle},$$

where $2^{L_i(t)}$ is the arithmetic average of $2^{L_i(t)}$ in one cell cycle period and $[M_i(t)]$ is the cell-cycle average mRNA concentration in molecules per cell, based on literature22. Linear interpolation was used to approximate the value of mRNA concentration in every minute during cell cycle, based on computed values at points of measurements (equation above).

**Estimating the consensus period for periodically expressed genes.** The set of genes transcriptionally regulated during the cell cycle will be defined as the genes with a transcriptional modulation consistent with the periodicity $T$ of the mitotic cell division. We utilized the measure of periodicity defined as the periodogram, $P_{35-37}$, of transcript concentration:

$$\int_{-\pi}^{\pi} \sigma_x E_x(T) \cdot \exp[-P_i(T)],$$

where $a$ and $b$ are the beginning and end of the time-course, respectively, $E$ is the transcript concentration and $\sigma$ is the standard deviation of gene expression $E$. To accommodate uneven distribution of time points, we estimate $P(T)$ using the unbiased formula of $P_{38}$. The statistical significance of a single frequency (corresponding to periodicity with period $T$) in the periodogram, assuming a Gaussian null hypothesis, is expressed by:

$$z = \exp[-P_i(T)],$$

Since no reliable value of the period $T$ measured independently from the transcriptome profiles was available, therefore, similar as in $6,25$, before applying Eq. 4, we estimated the most likely period of transcriptional oscillation in the system from the expression data. We have followed the Maximum Likelihood approach, using Eqs 4 and 5 for each gene independently over a range of possible periods, computing the logarithms of likelihood of periodicity for every gene and every period. These logarithms summed over all genes yield the total likelihood of every
period, and the period with the maximum total likelihood has been adopted as the consensus period of regulation in the system. Estimated cell cycle periods for different data sets are described in Table 1.

Correcting the estimated protein degradation half-lives. Belle et al.\textsuperscript{9} reported protein half-lives, as estimated from the observed degradation rate, that sometimes have very high values, and, at times, negative ones. Since such values are not realistic, we adopted the following algorithm to estimate the most likely true half-lives for these proteins. We assumed that the measured quantity (degradation rate $k_{d,i}$, which is related to the half-life $\tau_{d,i}$, by $k_{d,i} = \ln(2) / \tau_{d,i}$) may include an error that has a Gaussian distribution, with a variance corresponding to the inverse of 300 minutes (the maximum reliably measureable value according to\textsuperscript{9}) divided by the scaling factor $\ln(2)$. The negative reported half-lives result from experimental error, therefore, to correct the data we used the described above error model and prior assumption that a half-life must be positive. The true degradation rate was computed by integrating the normal distribution, limited and normalized to the positive part of its domain, and the inverse of this value multiplied by $\ln(2)$ was adopted as the corrected half-life. The correction was small for half-lives significantly shorter than 300 minutes, but significant for values longer than 300 minutes or negative reported values.

Calculating protein concentrations. We used the Fixed Point Iteration numerical method to solve Eq. 2 for each protein and mRNA data set. As a starting point for iterations we used $[P](0) = 0$ and $\Delta t = 1$ minute. We continued iterative calculations until convergence, specifically until the condition $|[P](T) - [P](0)| \leq 5 \cdot 10^{-10}$ had been met.

Comparison between steady-state and time-course based translation rates. To determine the differences between steady-state derived translation rate, $TR_{ss}$, and time-course derived translation rate, $TR_{tc}$, we defined the coefficient $TR_{diff}$:

$$TR_{diff} = \frac{|TR_{ss} - TR_{tc}|}{\min(TR_{ss}, TR_{tc})},$$

(6)

where the time-course derived translation rate, $TR_{tc}$, is defined by Eq. 3 and the steady-state derived translation rate, $TR_{ss}$, is defined by:

$$TR_{ss,i} = k_{d,i} \cdot \frac{[P]}{[M_j]}.$$

(7)

Incorporating post-translational regulation. To accommodate post-translational regulation, we expanded our approach by allowing time-dependent variation of degradation rates. We will use Clb2 as an example to illustrate detecting post-translational modifications. For Clb2, fitting constant degradation rate results in poor fit, both for half-lives based on the report of O’Shea and colleagues\textsuperscript{9} (Fig. 3B) and for the much shorter half-life reported by Amon et al.\textsuperscript{34} (Fig. 3A). Therefore, instead we propose a time-dependent half-life function that will also be periodic in the consecutive cell cycles. To describe a half-life that is modified by post-translational regulation within K minute window starting at the time $t_0$ within the cell cycle with the period T, we propose the following step function $\vartheta(t)_i$:

$$\vartheta(t)_i = \begin{cases} \vartheta_0^1, & t_0 \leq t \leq t_0 + K \\ \vartheta_0^2, & 0 < t < t_0 \text{ and } t_0 + K < t \leq T \end{cases}.$$

(8)

To find values of $\vartheta_0^1$, $\vartheta_0^2$, $t_0$ and K optimally describing time dependence of Clb2 half-life we numerically optimized these parameters, considering for half-lives $\vartheta_0^1$ and $\vartheta_0^2$ all values in the range from 1 minute to 40 minutes, with 1 minute steps, and for $t_0$ and K all possible times from the first to the last minute of the cell cycle, again with 1 minute steps. For each set of parameters for the function $\vartheta(t)_i$, we solved Eq. 2, as described previously (Calculating protein concentrations). The set of parameters offering the best fit with experimental data was chosen as the best estimate of true Clb2 half-life. Thus, we were also able to calculate the time-dependent degradation rate for Clb2 as $k_{d,i} = \frac{\ln(2)}{\vartheta_0^2}$. The best fit was achieved for variable half-life, with the Clb2 protein becoming extremely unstable outside of the window of its activity during the cell cycle (Fig. 3C). This result shows that our approach allows one to re-discover, \textit{ab initio}, the timing of post-translational regulation of a protein, if only gene expression and proteomic time-courses are available.

\(\alpha\)-Factor based synchronization. Yeast strain DBY8724 (Mat a GAL2 ura3 bar1::URA3) was kindly provided by P. T. Spellman. Obtained S. cerevisiae cells were synchronized by \(\alpha\)-factor arrest as described by Spellman et al.\textsuperscript{32} and Pramila et al.\textsuperscript{34}. Cells were grown to an OD\textsubscript{600} of 0.2 in YEP glucose pH 5.5, an asynchronous sample was taken and \(\alpha\)-factor (Sigma Aldrich) was added to a concentration of 25 ng/ml. After 2 hours cells were released from \(\alpha\)-factor arrest by pelleting and re-suspended in fresh medium to an OD\textsubscript{600} of 0.2 (Fig. 8C, time 0). Every 5 min, for the next 120 min, 25 samples were taken (25ml for western blot analysis, 1ml for FACS analysis and 1ml to count budding index). Cell cycle progression was monitored by bud counting and DNA content analysis (FACS) (Fig. 8A–C).
Budding index calculation and FACS analysis. For budding index calculation, two hundred cells were examined at every time point. The budding percentage was calculated as the number of budded cells divided by the number of all cells. To monitor DNA synthesis, samples were prepared as described previously and DNA content was measured using a BD FACSCalibur Flow Cytometer.

Western blot analysis. Cell extracts were prepared by TCA precipitation and then subjected to western blot analysis. Protein samples were separated on Mini-PROTEAN TGX 4–20% (Bio-Rad) gels and transferred to PureNitrocellulose Paper 0.45 μm (Bio-Rad). Blots were blocked using 0.2% I-Block buffer (Applied Biosystems), cut horizontally and probed with primary antibodies followed by incubation with appropriate horseradish peroxidase-conjugated secondary antibodies. The primary antisera used to detect selected proteins were from Santa Cruz Biotechnology (Rad50, Cdc5, and Clb2), Abcam (H3), Agrisera (Rnr1) and Millipore (Act1) and the secondary antisera were from Dako. Protein bands were visualized with the Immoblilon Western (Millipore) and scanned in a G-Box imaging system (Syngene). Band intensities were quantified using Gene-Snap software (Syngene).

References
1. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19, 1720–1730 (1999).
2. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci USA 94, 814–819 (1997).
3. von der Haar, T. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. BMC Syst Biol 2, 87, https://doi.org/10.1186/1752-0509-2-87 (2008).
4. Cohen, A. A. et al. Protein dynamics in individual human cells: experiment and theory. PLoS One 4, e4901, https://doi.org/10.1371/journal.pone.0004901 (2009).
5. Kudlicki, A., Rowicka, M. & Otwonowski, Z. SCEPTRANS: an online tool for analyzing periodic transcription in yeast. Bioinformatics 23, 1559–1561, https://doi.org/10.1093/bioinformatics/btm126 (2007).
6. Symington, L. S. Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev 66, 630–670, table of contents (2002).
7. Krogh, B. O. & Symington, L. S. Recombination proteins in yeast. Annu Rev Genet 38, 233–271, https://doi.org/10.1146/annurev.genet.38.072902.091500 (2004).
8. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. Proc Natl Acad Sci USA 103, 13004–13009, https://doi.org/10.1073/pnas.0605420103 (2006).
9. Elledge, S. J. & Davis, R. W. Two genes differentially regulated in the cell cycle and by DNA-damaging agents encode alternative regulatory subunits of ribonucleotide reductase. Genes Dev 4, 740–751 (1990).
10. Amon, A., Miller, M. P., Ee, L. S., Elkin, S. K. & Amon, A. Polo kinase Cdc5 is a central regulator of meiosis I. Proc Natl Acad Sci USA 110, 14278–14283, https://doi.org/10.1073/pnas.1311845110 (2013).
11. Veis, J., Klug, H., Koranda, M. & Ammerer, G. Activation of the G2/M-specific gene CLB2 requires multiple cell cycle signals. Mol Cell Biol 27, 8364–8373, https://doi.org/10.1128/MCB.01253-07 (2007).
12. Rowicka, M., Kudlicki, A., Tu, B. P. & Otwonowski, Z. High-resolution timing of cell cycle-regulated gene expression. Proc Natl Acad Sci USA 104, 16892–16897, https://doi.org/10.1073/pnas.0706022104 (2007).
13. Ammer, A., Irriger, S. & Nasmyth, K. Closing the cell cycle circle in yeast: G2 cyclin proteolysis initiated at mitosis persists until the activation of G1 cyclins in the next cycle. Cell 77, 1037–1050 (1994).
15. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**, 117, https://doi.org/10.1186/gb-2003-4-9-117 (2003).

16. Fuchter, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. J. A sampling of the yeast proteome. *Mol Cell Biol.* **19**, 7357–7368 (1999).

17. Kristensen, A. R., Goponer, I. & Foster, L. J. Protein synthesis rate is the predominant regulator of protein expression during differentiation. *Mol Syst Biol* **9**, 689, https://doi.org/10.1038/msb.2013.47 (2013).

18. Chan, R. K. & Otte, C. A. Isolation and genetic analysis of Saccharomyces cerevisiae mutants supersensitive to G1 arrest by a factor and alpha factor pheromones. *Mol Cell Biol.* **2**, 11–20 (1982).

19. Cappellaro, C., Mrsa, V. & Tanner, W. New potential cell wall glucanases of Saccharomyces cerevisiae and their involvement in mating. *J Bacteriol* **180**, 5030–5037 (1998).

20. Arava, V. et al. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci USA* **100**, 3889–3894, https://doi.org/10.1073/pnas.0035171100 (2003).

21. Revenu, S., Meliljison, I., Kupiec, M., Ruppin, E. & Tuller, T. Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput Biol* **7**, e1002127, https://doi.org/10.1371/journal.pcbi.1002127 (2011).

22. Beyer, A., Hollunder, J., Nasheuer, H. P. & Wilhelm, T. Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. *Mol Cell Proteomics* **3**, 1083–1092, https://doi.org/10.1074/mcp.M400959-MCP200 (2004).

23. Szczepinska, T. et al. Probabilistic approach to predicting substrate specificity of methyltransferases. *PLoS Comput Biol* **10**, e1003514, https://doi.org/10.1371/journal.pcbi.1003514 (2014).

24. Plewczynski, D., Tkacz, A., Godzik, A. & Rychlewski, L. A support vector machine approach to the identification of phosphorylation sites. *Cell Mol Biol Lett* **10**, 73–89 (2005).

25. Yu, B. P., Kudlicki, A., Rowicka, M. & McKnight, S. L. Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science* **310**, 1152–1158, https://doi.org/10.1126/science.1120499 (2005).

26. Fongang, B. & Kudlicki, A. Comparison between Timelines of Transcriptional Regulation in Mammals, Birds, and Teleost Fish Somitogenesis. *PLoS One* **11**, https://doi.org/10.1371/journal.pone.0155802 (2016).

27. Fongang, B. & Kudlicki, A. The precise timeline of transcriptional regulation reveals causation in mouse somitogenesis network. *Bmc Dev Biol* **13**, 42, https://doi.org/10.1186/1471-213X-13-42 (2013).

28. Yen, H. C. S., Xu, Q. K., Chou, D. M., Zhao, Z. M. & Elledge, S. J. Global Protein Stability Profiling in Mammalian Cells. *Science* **322**, 918–923, https://doi.org/10.1126/science.1160489 (2008).

29. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342, https://doi.org/10.1038/nature10098 (2011).

30. van der Lee, R. et al. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep* **8**, 1832–1844, https://doi.org/10.1016/j.celrep.2014.07.035 (2014).

31. Fishbain, S. et al. Sequence composition of disordered regions fine-tunes protein half-life. *Nat Struct Mol Biol* **22**, 214–221, https://doi.org/10.1038/nsmb.2958 (2015).

32. Spellman, P. T. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* **9**, 3273–3297 (1998).

33. Cho, R. J. et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* **2**, 65–73 (1998).

34. Pramila, T., Wu, W., Miles, S., Noble, W. S. & Breeden, L. L. The Forkhead transcription factor Hcm1 regulates chromosome segregation and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev* **20**, 2266–2278, https://doi.org/10.1101/gad.1450606 (2006).

35. Fisher, R. A. Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A.* **125**, 54–59 (1929).

36. Lomb, N. R. Least-Squares Frequency-Analysis of Unequally Spaced Data. *Astrophysics and Space Science* **39**, 447–462 (1976).

37. Schuster, A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism and Atmospheric Electricity* **3**, 13–41 (1898).

38. Kudlicki, A., Rowicka, M. & Otwinowski, Z. Significance-testing of periodogram for short time series. *Proceedings of the 2008 International Conference on Bioinformatics and Computational Biology*, 424–430 (2008).

39. Foss, E. J. Tedp regulates DNA damage responses during S phase in Saccharomyces cerevisiae. *Genetics* **157**, 567–577 (2001).

40. Oficjalska-Pham, D. et al. General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Mafl. *Mol Cell* **22**, 623–632, https://doi.org/10.1016/j.molcel.2006.04.008 (2006).

41. Li, L. et al. Budding yeast SSD1-V regulates transcript levels of many longevity genes and extends chronological life span in purified quiescent cells. *Mol Cell Biol* **20**, 3851–3864, https://doi.org/10.1091/mbc.E09-04-0347 (2009).

Acknowledgements

This work was supported by National Institutes of Health grant 5R01GM112131, Foundation for Polish Science (TEAM), National Science Centre (2011/02/A/ N22/00014, 2014/15/B/NZ1/03357), European Regional Development Fund under Innovative Economy Programme (POIG.02.02.00-14/02/08-00, POIG.02.03.00-14–128/13). We are grateful to Andrzej Dziembowski and Maciej Sykulski for insightful suggestions, Paul T. Spellman for DBY8724 α strain and help with α-factor based synchronization, and Tamir Tuller for RFM translation rates data.

Author Contributions

M.R. conceived the computational method, K.G. conceived the validation experiments, M.R. and K.G. coordinated the project. K.K. implemented the computational method and developed software, J.T., A.B. and J.K. performed validation experiments. A.K. processed half-life data and contributed to statistical data analysis. M.R., K.K., K.G., A.B. and A.K. wrote the manuscript; all authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-31752-4.

Competing Interests: The authors declare no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
