Supplementary information

October 27, 2013

Contents

List of Figures

1 Supplementary Figures

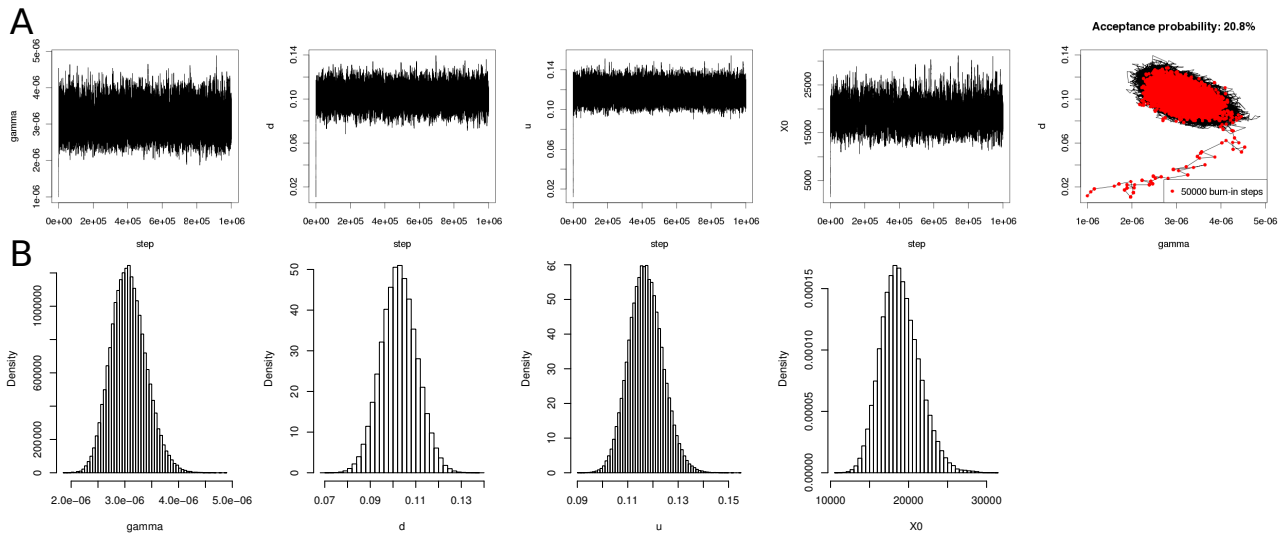

Figure S1: Fitting a 2-exponential function to the time-course of siTK3 small RNA loading into cytoplasmic Ago2, measured by Fluorescence Cross-Correlation Spectroscopy (FCCS) (Ohrt et al., 2008). A: Tracing the progress of Monte Carlo Markov chain (MCMC) sampling for all four model parameters γ, d, u, X_0 over time. B: Posterior probability distributions of the four parameters. See Methods and Materials in the main text.

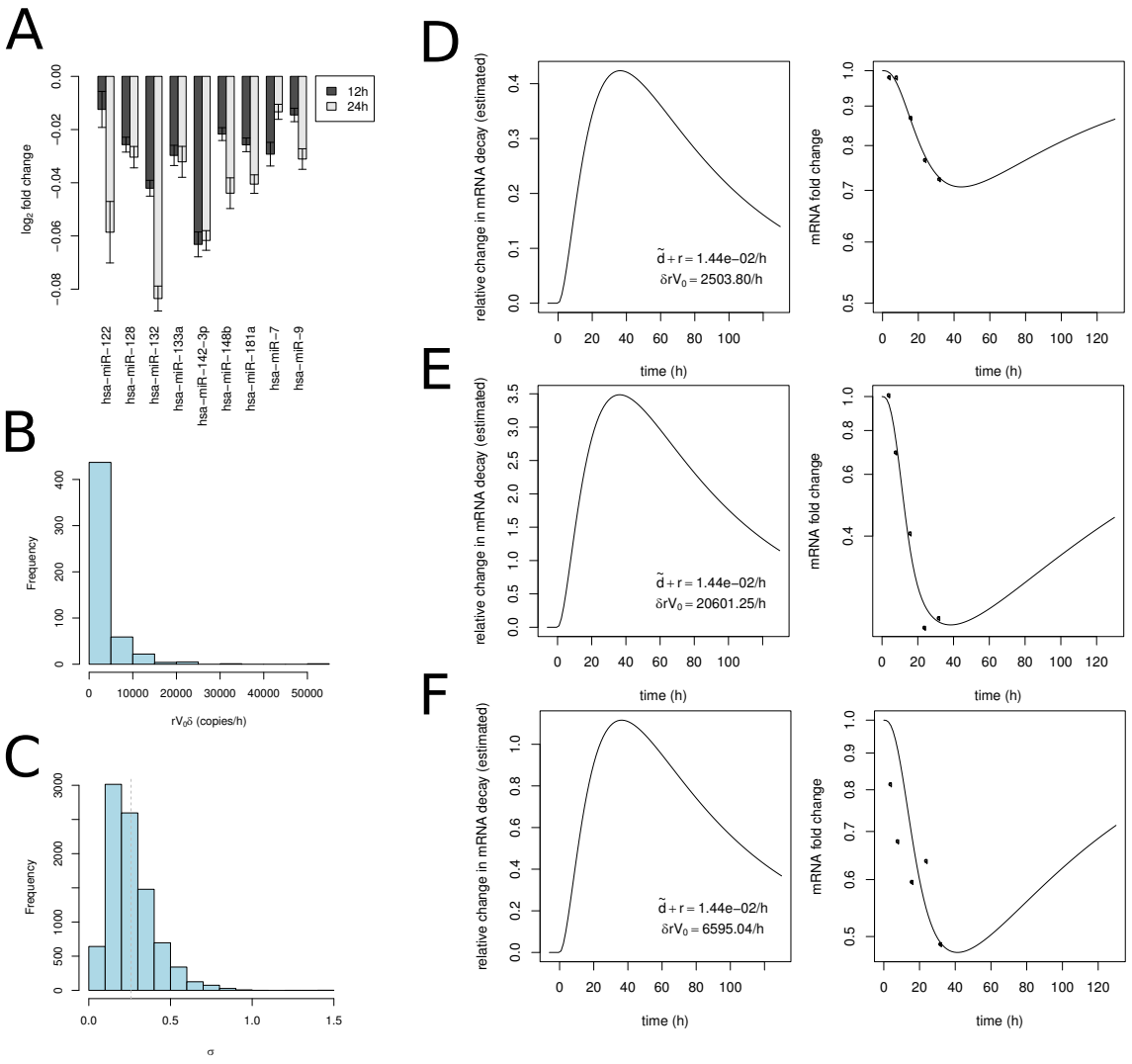

Figure S2: Fitting the targeting efficiency parameter $rV_0\delta$ of the model with the tri-exponential loading function to the mRNA profiling time-course following the transfection of miR-124 in HepG2 cells by Wang and Wang (2006). A: Target mRNAs are significantly more repressed at 24 hours than at 12 hours in 6 out of 9 the miRNA transfection experiments of Grimson et al. (2007). Shown are the average log2 fold changes in mRNAs carrying at least one seed match to the transfected miRNA, with error bars representing 95% confidence intervals on the mean. B: The histogram shows the distribution of maximum likelihood estimates of $rV_0\delta$ for the 1098 down-regulated mRNAs with miR-124 seed match in the 3' UTR. C: Estimating the typical reproducibility of mRNA fold change measurements upon miRNA transfection. Histogram of the standard deviation on the $\log_2$ mRNA fold change of 9009 detected genes upon miR-124 transfection in HEK293 by Karginov et al. (2007). The standard deviation σ was computed across the 6 replicates of the microarray experiment. The dotted vertical lines represents the mean standard deviation on the $\log_2$ mRNA fold change ($\sigma = 0.26$, representing a $2^{0.26} = 1.20$ fold error, or a 20% error on the fold change). D–F: Sample kinetics of miRNA-induced changes in target mRNAs in the miR-124 transfection experiments of Wang and Wang (2006). Shown are changes in mRNA levels of the ANXA4 (D, Entrez ID: 307), RALA (E, Entrez ID: 5898), and KLHL28 (F, Entrez ID: 54813) genes. ANXA4 is among the genes that best fits the model while RALA and KLHL28 respectively fit the model with a 10% and 20% error on the fold change. The left panels show the estimated change in mRNA decay rates following the transfection of miR-124 while the measurements (dots) and model-predicted (lines) changes in mRNA levels following miRNA transfection are shown in the right panels.

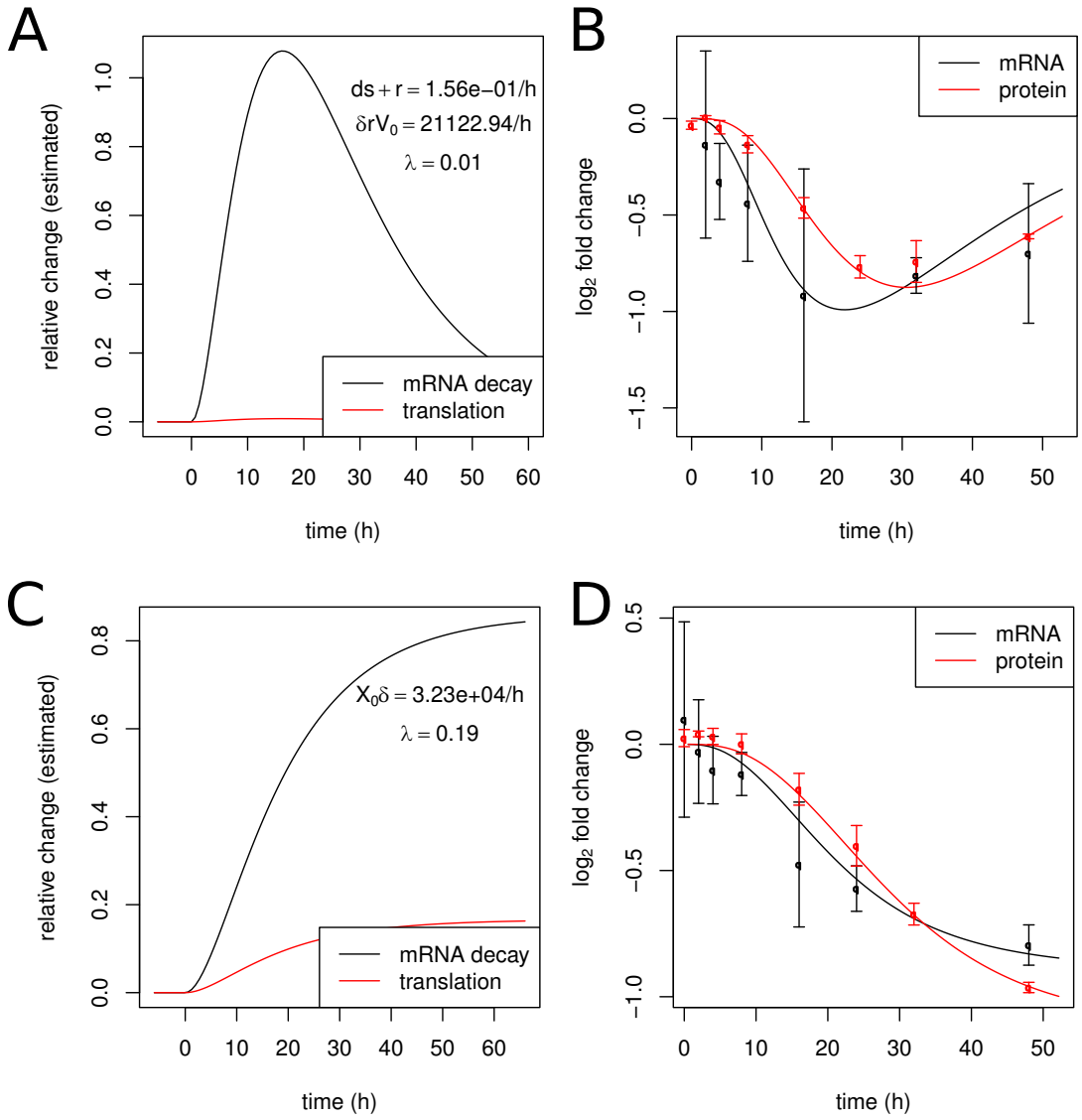

Figure S3: Fit of the kinetic model to changes in luciferase activity and mRNA levels upon miR-199a transfection (A, B) and induction (C, D) in HEK293 cells. A: Estimated change in translation repression and mRNA decay following miR-199a transfection. B: Measured (dots) and model-predicted (lines) changes in mRNA and protein levels. The maximum-likelihood estimate of miRNA clearance parameter $\tilde{d} + r$ was 0.156h^{-1} was one order of magnitude higher than the value obtained based on the transfection experiment of Wang and Wang (2006). This may be due to the difference in the cell lines and transfection reagents used here — Lipofectamine 2000 (Invitrogen) in HEK293 cells — compared to siPORT NeoFX (Ambion) in HepG2 cells by Wang and Wang (2006). On the other hand, the parameter related to the effect of the miRNA on the mRNA level $rV_0\delta = 21'123\text{h}^{-1}$ was within the range of the values inferred for individual genes in the experiment of Wang and Wang (2006) (Fig. S2). Finally, the parameter λ quantifying the extent of translation inhibition was smaller than 1 in both experiments, indicating that the main effect of the miRNA was on mRNA degradation rather than on translation rate.
C: Estimated change in translation repression and mRNA decay following the *induction* of hsa-miR-199a-3p in HEK293 cells, assuming a three compartment Ago loading model and constant miRNA synthesis. D: Measured (dots) and model-predicted (lines) changes in mRNA and protein levels. Error bars represent 95% confidence intervals on the mean.

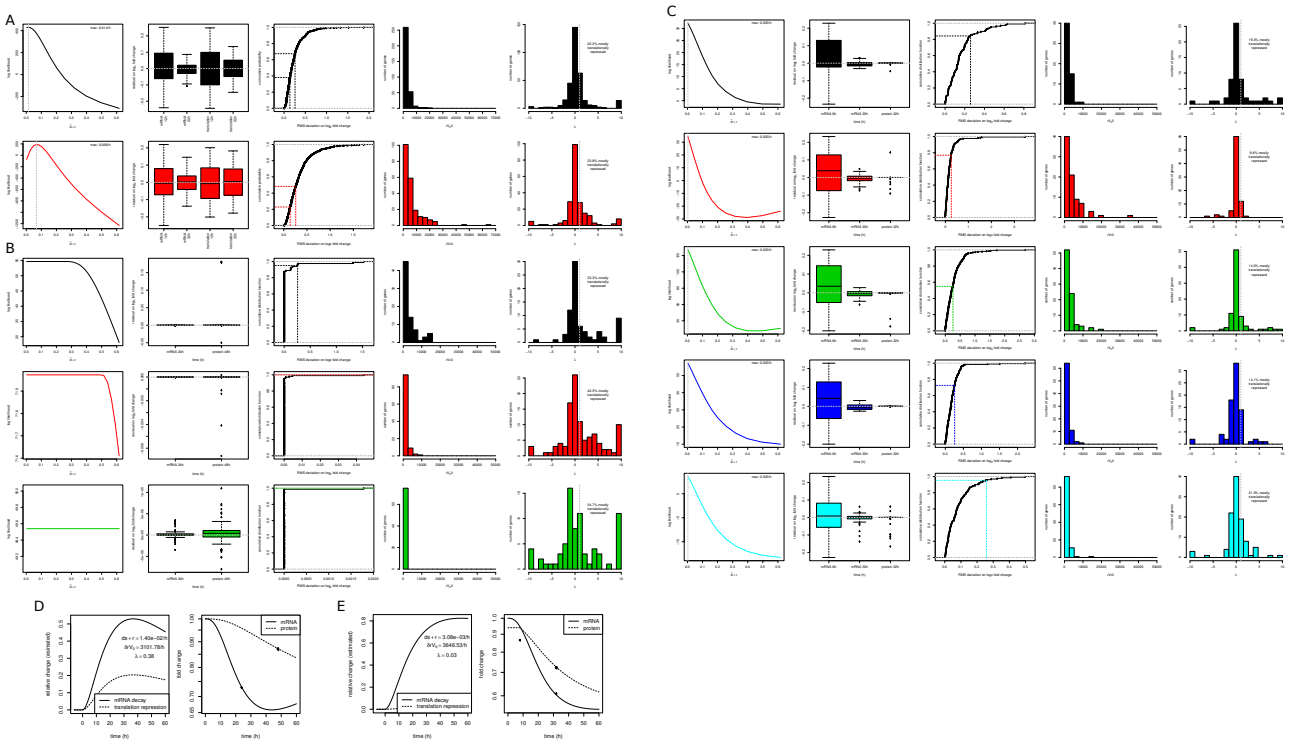

Figure S4: Fit of the model to the changes in mRNA abundance, translation efficiency and protein abundance observed in various experiments.
A: Transfection of miR-155 (black) and miR-1 (red) followed by mRNAseq and Ribosome Protected Fragment sequencing 12h and 32h post transfection by Guo et al. (2010). B: Transfection of miR-124 (black), miR-1 (red) and miR-181a (green) followed by microarrays 24h post-transfection and SILAC proteomics 48h post-transfection by Baek et al. (2008). C: Transfection of let-7b (black), miR-155 (red), miR-16 (green), miR-1 (blue) and miR-30a (cyan) followed by microarrays 8h and 32h post-transfection and pSILAC proteomics 32h post transfection by Selbach et al. (2008).
Among the panels corresponding to each experiment, the left-most figure shows the log-likelihood profile of the $\tilde{d}+r$ parameter given the mRNA profiling data, with the vertical line marking the maximum likelihood estimate of $\tilde{d}+r$. The second figure shows boxplots of the model residuals on $\log_2$ fold changes with boxes spanning the interquartile range and whiskers extending up to 1.5 times the interquartile range. The third figure represents the cumulative distribution of the Root Mean Squared deviations between the model predictions and the data, defined as $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - d_i)^2}$, where m_i are the model predictions, d_i are the measurements and n is the number of data points per gene ($n = 4$ for Guo et al. (2010), $n = 2$ for Baek et al. (2008), $n = 3$ for Selbach et al. (2008)). The dotted line marks the $0.26 = \log_2(1.20)$ cut-off on the RMS deviation which represents a 20% error on the fold change. This error level is typically observed in replicate miRNA transfection experiments (see Fig. S2C). Ingolia et al. (2009) reported a typical measurement error of 0.37 on $\log_2$ translation efficiencies by RPF sequencing, i.e. a 30% error. The fourth figure shows the histogram of the $rV_0\delta$ targeting efficiency parameters for all genes and the fifth figure the histogram of the relative contribution of translation repression λ across genes, together with the fraction of genes for which translation repression dominates over mRNA decay ($\lambda > 1$). Values extending outside the -10 – 10 range were assigned to the -10 or 10 bins.
D: Sample kinetics in the experiments of Baek et al. (2008). The left panel shows best-fitted changes in translation and decay rates of the MYH10 transcript (myosin, heavy chain 10, non-muscle, RefSeq ID: NM_005964) following miR-124 transfection. Measured (dots) and model-predicted (lines) changes in mRNA and protein levels following miRNA transfection appear on the right panel.
E: Sample kinetics in the pSILAC experiments of Selbach et al. (2008). The left panel shows best-fitted changes in translation and decay rates of the HNRNPU gene (heterogeneous nuclear ribonucleoprotein U, RefSeq ID: NM_031844) following miR-1 transfection. Measured (dots) and model-predicted (lines) changes in mRNA and protein levels following miRNA transfection appear on the right panel.

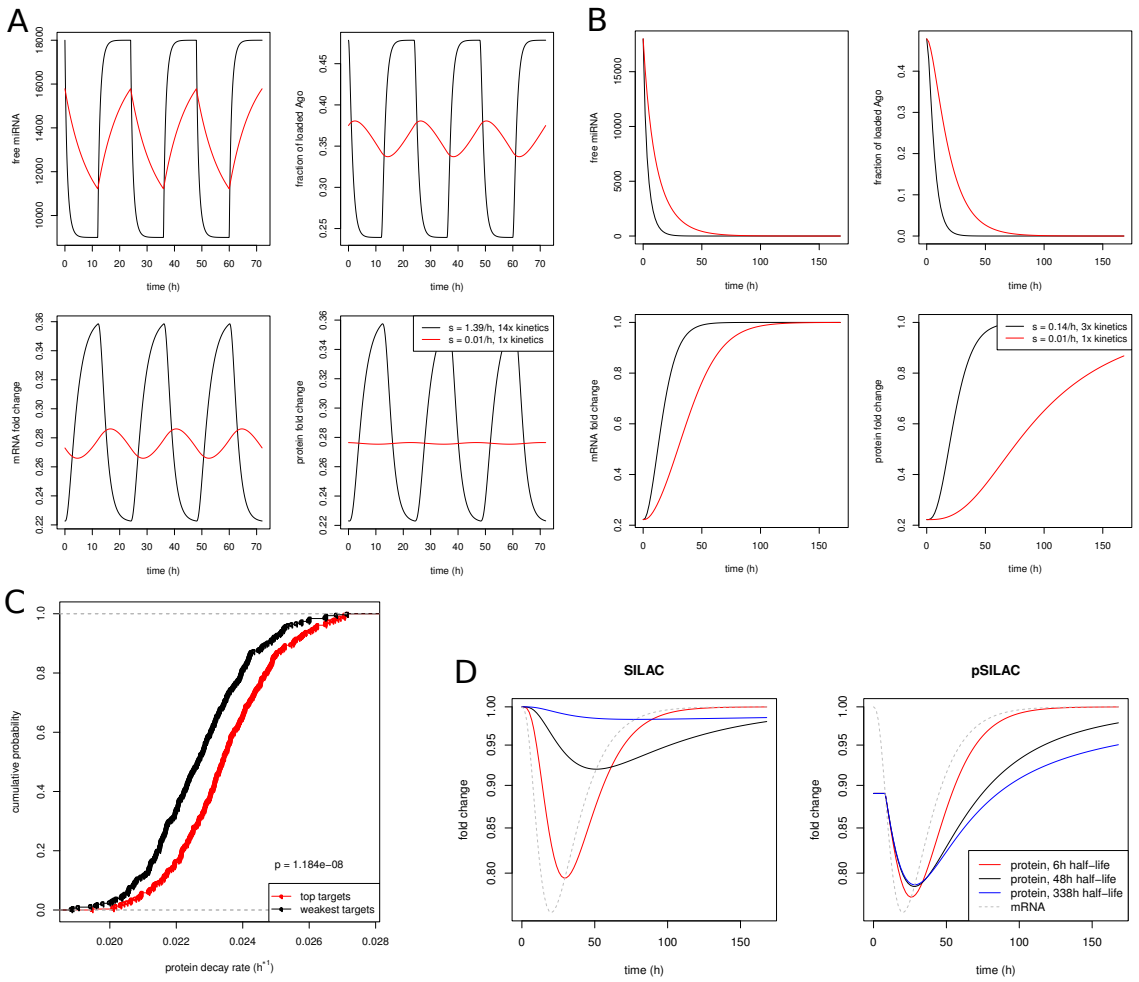

Figure S5: miRNA and protein turn-over create bottlenecks in miRNA-mediated gene regulation. A: Simulated dynamics of the free miRNA, the fraction of Ago loaded with the miRNA, and the target mRNA and protein fold changes in response to a miRNA whose synthesis cycles through half- and full induction every 24h. Simulations were performed assuming parameters derived from experimental data sets (48h protein half-life, miRNA kinetics from biophysics data of Ohrt et al. (2008), red lines) or faster kinetics (30min protein half-life, 14-fold increase in general miRNA kinetics, black lines).
B: Simulated dynamics of the free miRNA, the fraction of Ago loaded with the miRNA, and the target mRNA and protein fold changes upon a sudden drop in miRNA synthesis at time $t = 0$h. Simulations were performed assuming parameters derived from experimental data sets (48h protein half-life, miRNA kinetics from biophysics data of Ohrt et al. (2008), red lines) or faster kinetics (5h protein half-life, three fold increase in general miRNA kinetics, black lines).
C: miRNAs preferentially target protein with fast turn-over. The red line shows the cumulative distribution of the average protein decay rates of the top 50 target genes of 310 miRNA families while the black line represents the distribution of protein decay rates of the 50 weakest target genes of these same 310 miRNA families. From 718 human miRNAs in miRbase (version 13.0, downloaded on January 13th 2010), we built 583 miRNA families, defined as miRNAs sharing the same sequence at positions 2 to 8. Of these families, 310 had at least 100 predicted target genes (Gaidatzis et al., 2007) for which the decay rate of the encoded protein in HeLa cells was measured in the study of Cambridge et al. (2011). We performed the analysis at the Entrez Gene ID level: if several proteins for same gene, we used the average decay rate. Similarly, if ElMMo had predictions for several isoforms of the gene, we used the average ElMMo score. For each of miRNA family, we then computed the average protein decay rate of the top-ranking and weakest 50 targets. We finally compared the protein decay rates of the top and weakest target genes with Wilcoxon's signed rank test which takes pairing into account.
D: In SILAC experiments, changes in protein levels following miRNA transfection strongly depend on the protein decay rates. Left panel: Simulating fold changes in protein levels as measured by SILAC following miRNA transfection. Model parameters were obtained by fitting the fluorescence cross-correlation spectroscopy data of Ohrt et al. (2008) and the microarray time-series of Wang and Wang (2006). We assumed no specific translation repression from the miRNA ($\lambda = 0$). The solid red, black and blue lines represent the simulated trajectories of proteins with half-lives of 6, 48 and 336h, which span 95% of protein half-lives measured by Schwanhäusser et al. (2011). The dotted grey line represents changes in mRNA levels. Right panel: Same for a pulsed SILAC experiment. In this simulation, labeling medium was changed 8h post transfection, like in the experiment of Selbach et al. (2008).

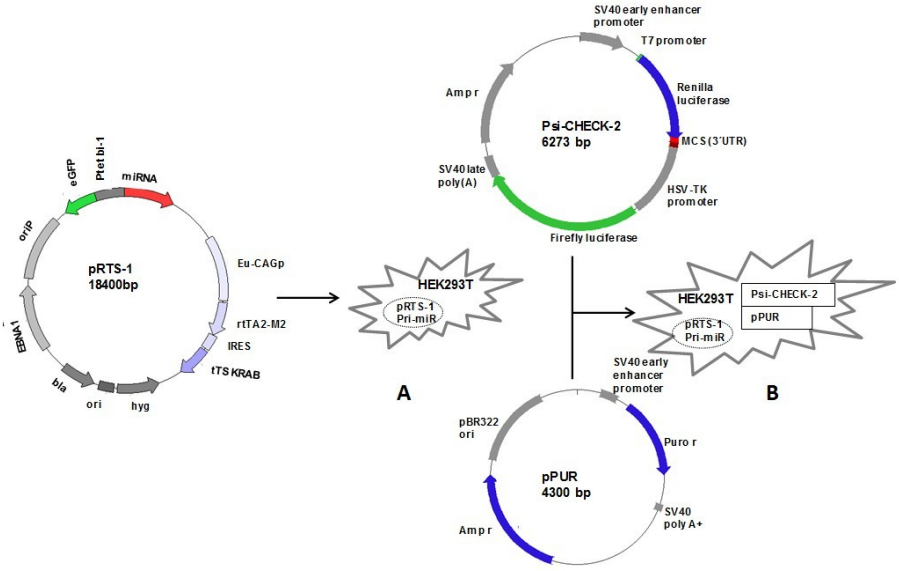

Figure S6: Schematic representation of the work plan followed in the establishment of KTN1 cell line. The 3' UTR of the kinectin 1 (KTN1) gene and the primary miR-199a-3p were PCR amplified and individually cloned into the multiple cloning site of a pGEM-T Easy vector. The 3' UTR was subcloned to a reporter psiCHECK-2 vector, whereas the primary miR-199a-3p was subcloned to an inducible episomal pRTS-1 vector through defined restriction sites. A: First we constructed a stable cell line with inducible miRNA expression from a pRTS-1 vector. B: Into the stable cell line with inducible miRNA expression, the reporter and pPUR vectors were co-transfected for stable integration.

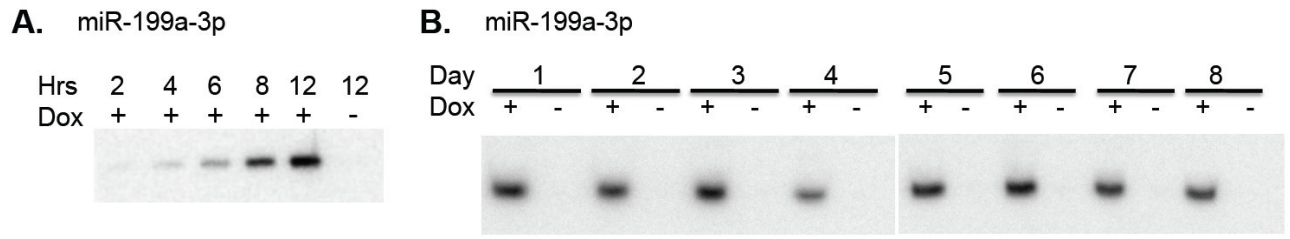

Figure S7: Induction of miR-199a-3p in HEK293 cells containing the pRTS1 construct. A: Northern blot illustrating the progressive increase in miR-199a-3p expression within 12 hours after 1μg/ml Dox induction. B: Northern blot showing stable miR-199a-3p induction over an 8 days time-course, with one time-point per day.

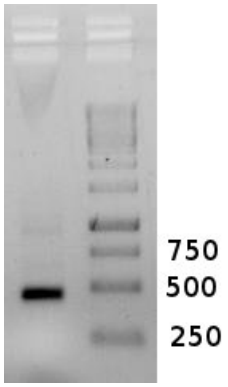

Figure S8: PCR confirming the genomic integration of KTN1 3' UTR-psiCHECK2 vector in the KTN1 cell line.

2 Supplementary Tables

Rank	mature miRNA	reads Mph	reads allPh	% MPh	% allPh	fold change MPh vs allPh	log Psame/Pdiff
1	hsa-miR-21	1.10×10^6	3.37×10^6	4.67	7.79	0.60	-125243
2	hsa-let-7a	828389	803533	3.53	1.86	1.90	-84652.5
3	hsa-let-7f	616610	610403	2.62	1.41	1.86	-59221.9
4	hsa-let-7i	521298	486770	2.22	1.13	1.97	-58191
5	hsa-let-7b	219306	162685	0.93	0.38	2.48	-39196.5
6	hsa-miR-27a	2.23×10^6	4.88×10^6	9.48	11.28	0.84	-26487.8
7	hsa-let-7g	253411	258250	1.08	0.60	1.81	-22143.6
8	has-miR-22	377413	984768	1.61	2.28	0.71	-17800.6
9	hsa-let-7d	170373	164487	0.73	0.38	1.91	-17260.9
10	has-miR-186	441113	1.11×10^6	1.88	2.56	0.73	-16329.2

Table S1: Relative abundance of Ago-bound miRNAs in HeLa cells in M phase (MPh) vs unsynchronized cells (allPh). Data from Kishore et al. (2013), deposited in GEO under the accession GSE43666. Ago was immunoprecipitated with a specific antibody (generously provided by Gunter Meister) and the associated small RNAs were sequenced on an Illumina HiSeq 2000 platform. The sequencing data analyzed on the Clipz server (Khorshid et al., 2011) resulting in 2.35×10^7 reads mapping to miRNAs in the Mph sample and 4.32×10^7 reads in the allPh sample. Shown are number of reads (reads MPh, reads allPh), relative abundance (% MPh, % allPh), fold change of the 10 mature miRNAs which are most significantly differentially expressed in MPh compared to allPh. MiRNAs were ranked by the statistical evidence of differential expression (log Psame/Pdiff, see Berninger et al. (2008)).

3 Supplementary Methods and Results

3.1 Modeling Ago loading following siRNA micro-injection.

A simplified view of the siRNA micro-injection experiment of Ohrt et al. (2008) is the following. At time $t = 0$, X_0 siRNAs are injected into the cell and start binding free Ago molecules. Presumably most of the Ago molecules are already bound to endogenous miRNAs (Khan et al., 2009) and these complexes dissociate at a certain rate. Free small RNAs, whether exogenous (siRNAs) or endogenous (miRNAs), decay. We will first model the turnover of miRNA-Ago complexes that took place before the siRNAs were injected.

Let R be the total number of Ago molecules (free and bound to small RNAs), u the dissociation rate of Ago-bound miRNAs, and E the number of free endogenous miRNAs. We then have the following differential equation for the amount of free Ago F:

$$\frac{dF}{dt} = u(R - F) - bEF$$

Or, if we define the fraction of free Ago $f = \frac{F}{R}$, we have

$$\frac{df}{dt} = u(1 - f) - bEf$$

At steady-state, we have

$$f = \frac{u}{bE + u}$$

We can thus assume that, at the beginning of the experiment, the fraction of free Ago is $f_0 = \frac{u}{bE_0+u}$, where E_0 is the amount of free endogenous miRNAs.

We now add the siRNAs that are injected at time $t = 0$. As mentioned above, we assume that only free small RNAs decay at a rate d, and we further assume that the siRNAs bind to and dissociate from Ago at the same rates b and u as endogenous miRNAs. The dynamics of the free siRNAs $X(t)$ is thus described by the equation

$$\frac{dX}{dt} = -dX - bXF + uL, \tag{1}$$

where the dynamics of the amount of Ago loaded with the small RNA, $A(t)$, is determined by

$$\frac{dA}{dt} = bXF - uA \tag{2}$$

and the dynamics of the amount of free Ago $F(t)$ is driven by

$$\frac{dF}{dt} = u(R - F) - b(E + X)F$$

Formation and dissociation of molecular complexes typically happen on a fast time-scale compared to processes such as synthesis and degradation. We therefore make a steady-state approximation for F. We further assume that the amount of free endogenous miRNAs E is not significantly affected by the small RNAs micro-injection, and therefore set $E = E_0$ which leads to

$$\begin{array}{rcl} \frac{dF}{dt} = 0 & \Rightarrow & F = R\frac{u}{b(E_0+X)+u} \\ & \Rightarrow & f = \frac{F}{R} = \frac{u}{b(E_0+X)+u} \end{array} \tag{3}$$

In the limit of f very small, i.e. when there is very little free Ago, the loading of small RNAs into Ago takes place at a rate that is limited by the dissociation of Ago-small RNA complexes u. In this limit, f_0 is close to zero at the start of the experiment. As the small RNAs are added, f decreases somewhat, but unless the amount of small RNAs is much larger than the amount of endogenous miRNA, the decrease in f will be small. In addition, this decrease will be transient, the fraction of free Ago returning to f_0 as the siRNAs decay. It is therefore reasonable to assume that the fraction of free Ago f is small and constant across the time course, i.e. $f(t) \simeq f_0 = \frac{u}{bE_0+u}$. Substituting equation 3 into equation 1 and equation 2 leads to

$$\left\{ \begin{array}{rcl} \frac{dA}{dt} & = & bXRf_0 - uA \\ \frac{dX}{dt} & = & -dX - bXRf_0 + uA \end{array} \right. \tag{4}$$

In a recent study, Ohrt et al. (2008) measured both the fraction $a = \frac{A}{R}$ of Ago in complex with an injected siRNA as well as the fraction $s = \frac{A}{A+X}$ of the siRNAs in complex with Ago over time. a and s can both be obtained from the system 4, which leads to the following system

$$\left\{ \begin{array}{rcl} \frac{da}{dt} & = & bf_0X - ua \\ \frac{dX}{dt} & = & -(d + bf_0R)X + uRa \\ \frac{ds}{dt} & = & bf_0R(1 - s) + s(d(1 - s) - u) \end{array} \right. \tag{5}$$

Because b and f_0 always occur in a product, the two parameters are not identifiable from the time series measurements of the fraction of Ago and siRNAs in complex $a(t)$ and $s(t)$. With therefore substitute $\gamma = bf_0$ which gives

$$\begin{cases} \frac{da}{dt} &= \gamma X - ua \\ \frac{dX}{dt} &= -(d+\gamma R)X + uRa \\ \frac{ds}{dt} &= \gamma R(1-s) + s(d(1-s) - u). \end{cases} \tag{6}$$

By fitting $a(t)$ and $s(t)$ to the data of Ohrt et al. (2008), one can determine the value of the parameters that drive siRNA loading γ, dissociation from Ago u and degradation d.

To be able to model changes in mRNA and protein levels over time however, the critical variable is the fraction of Ago complexed with small RNAs $a(t)$, which depends on $X(t)$ but not on $s(t)$. We therefore sought to find an analytical solution to the ODE system (6). This can be done since the system is linear and homogeneous. Under the initial conditions $a(0) = 0$ and $X(0) = X_0$, one can show (Tenenbaum and Pollard, 1963) that the solution of system 6 is

$$\begin{cases} a(t) &= k\left(e^{-\beta_1 t} - e^{-\beta_2 t}\right) \\ X(t) &= \tilde{k}e^{-\beta_1 t} + \left(X_0 - \tilde{k}\right)e^{-\beta_2 t} \end{cases} \tag{7}$$

with

$$\begin{cases} \beta_1 &= \frac{d+\gamma R+u+\sqrt{(d+\gamma R+u)^2-4ud}}{2} \\ \beta_2 &= \frac{d+\gamma R+u-\sqrt{(d+\gamma R+u)^2-4ud}}{2} \\ k &= \frac{\gamma X_0}{\beta_2-\beta_1} \\ \tilde{k} &= X_0\frac{\beta_2-d-\gamma R}{\beta_2-\beta_1} \end{cases}$$

3.2 Ordinary differential equation model of miRNA-dependent regulation of mRNA and protein levels

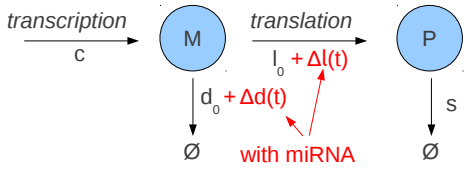

Figure S9: A simple two-compartment ordinary differential equation model of the regulatory action of miRNAs.

Let p_+ and m_+ represent the average protein and mRNA copy number for a gene targeted by the transfected miRNA, and assume that the miRNA transfection alters the translation rate l_0 and the mRNA decay rate d_0 by time-dependent functions $\Delta l(t)$ and $\Delta d(t)$ (see figure above). We obtain the system

$$\begin{cases} \frac{dp_+}{dt} &= (l_0 + \Delta l(t))\, m_+ - sp_+ \\ \frac{dm_+}{dt} &= c - (d_0 + \Delta d(t))\, m_+ \end{cases}$$

In the absence of the miRNA (control/mock transfections), we assume that mRNA translation and decay occur at constant rates, in which case the protein and mRNA levels p_- and m_- of miRNA targets follow the dynamics:

$$\begin{cases} \frac{dp_-}{dt} &= l_0 m_- - sp_- \\ \frac{dm_-}{dt} &= c - d_0 m_- \end{cases} \tag{8}$$

We finally assume that when the miRNA is transfected, the mRNA and protein are at steady-state, i.e.

$$m_+(0) = m_-^* = \frac{c}{d_0}$$

and

$$p_+(0) = p_-^* = \frac{c}{d_0}\frac{l_0}{s}.$$

We can now derive the dynamics of the changes in protein and mRNA levels in miRNA- compared to control/mock-transfected cells, $f_p = \frac{p_+}{p_-}$ and $f_m = \frac{m_+}{m_-}$.

SILAC experiment of Baek et al. (2008). Because we assumed that the system is in steady-state before the miRNA transfection and that the control/mock transfection does not affect the expression of miRNA targets, m_- can be considered constant. This leads to

$$\begin{aligned}
\frac{df_m}{dt} &= \frac{d_0}{c}\frac{dm_+}{dt} && (9)\\
&= \frac{d_0}{c}\left[c - (d_0 + \Delta d(t))\, m_+\right] && (10)\\
&= d_0 - (d_0 + \Delta d(t))\, f_m && (11)\\
&= d_0\left[1 - \left(1 + \frac{\Delta d(t)}{d_0}\right) f_m\right] && (12)
\end{aligned}$$

The initial condition is $f_m(0) = \frac{m_+(0)}{m_-(0)} = \frac{m_-(0)}{m_-(0)} = 1$.

Similarly, for the dynamics of changes in protein levels we have

$$\begin{aligned}
\frac{df_p}{dt} &= s\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - s f_p && (13)\\
&= s\left[\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - f_p\right] && (14)
\end{aligned}$$

with a similar initial condition, $f_p(0) = \frac{p_+(0)}{p_-(0)} = \frac{p_-(0)}{p_-(0)} = 1$. These dynamic equations can be used to analyze data derived from SILAC measurements of protein levels.

Pulsed SILAC experiment of Selbach et al. (2008). The measurements of mRNA dynamics were performed similarly in the studies of Selbach et al. (2008) and of Baek et al. (2008). Consequently, mRNA dynamics upon miRNA transfection is described by equation 12 above.

The measurements of protein expression were performed differently. Namely, cells were either miRNA- or control-transfected and 8 hours after transfection, the media in which the two types of cells were growing were changed. As a consequence, proteins synthesized before and after the 8h time point incorporated different isotope-labeled amino acids. Furthermore, proteins synthesized in the control- and miRNA-transfected cells were labeled with different isotopes. Only proteins that were synthesized after the medium change were quantified. Denoting the levels of these proteins by p_+ (in miRNA-transfected cells) and p_- (in control/mock-transfected cells), we have the initial condition $p_+(8) = p_-(8) = 0$. 32 hours after transfection, cells were collected and protein levels in transfected cells were compared to proteins levels in non-transfected cells. Since mRNA levels do not deviate from the steady-state in the control experiment, $m_-(t) = \frac{c}{d_0}$. Therefore, equation 8 becomes

$$\begin{cases} \frac{dp_-}{dt} &= l_0\frac{c}{d_0} - s p_- \\ \frac{dm_-}{dt} &= 0 \end{cases}$$

which leads to $p_-(t) = \frac{l_0 c}{s d_0}\left(1 - e^{-s(t-8)}\right) = \frac{l_0}{s} m_-\left(1 - e^{-s(t-8)}\right), t \geq 8$. Taking the derivative of $f_p = \frac{p_+}{p_-}$ with respect to time yields

$$\begin{aligned}
\frac{df_p}{dt} &= \frac{\frac{dp_+}{dt}p_- - \frac{dp_-}{dt}p_+}{p_-^2} && (15)\\
&= \frac{1}{p_-}\left[(l_0 + \Delta l(t))\, m_+ - s p_+\right] - \frac{1}{p_-}\frac{dp_-}{dt} f_p && (16)\\
&= \frac{s}{1 - e^{-s(t-8)}}\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - s f_p - \left(\frac{l_0 c}{s p_-} - s\right) f_p && (17)\\
&= \frac{s}{1 - e^{-s(t-8)}}\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - s f_p - \left(\frac{l_0 c}{d_0}\frac{s}{l_0 m_-(1 - e^{-s(t-8)})} - s\right) f_p && (18)\\
&= \frac{s}{1 - e^{-s(t-8)}}\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - s f_p - s\left(\frac{1}{1 - e^{-s(t-8)}} - 1\right) f_p && (19)\\
&= \frac{s}{1 - e^{-s(t-8)}}\left[\left(1 + \frac{\Delta l(t)}{l_0}\right) f_m - f_p\right] && (20)
\end{aligned}$$

The initial condition $f_p(8) = \frac{p_+(8)}{p_-(8)}$ is undefined, so we obtain a limit approximation using L'Hospital's rule

$$\begin{aligned} f_p(8) := \lim_{t\to 8^+} f_p(t) &= \lim_{t\to 8^+} \frac{p_+(t)}{p_-(t)} \\ &= \lim_{t\to 8^+} \frac{\frac{dp_+}{dt}}{\frac{dp_-}{dt}} \\ &= \lim_{t\to 8^+} \frac{(l_0 + \Delta l(t))\, m_+ - s p_+}{l_0 m_- - s p_-} \\ \Rightarrow f_p(8) &= \left(1 + \frac{\Delta l(8)}{l_0}\right) f_m(8) \end{aligned}$$

To simulate changes in mRNA and protein levels over time, one needs to specify the functional form of the time-dependent influence of miRNAs on translation $\frac{\Delta l(t)}{l_0}$ and mRNA decay $\frac{\Delta d(t)}{d_0}$. We make the assumption that these depend on the proportion of Ago proteins that are loaded with the transfected miRNA, which we estimate as described in the next section.

Modeling Ago loading upon miRNA transfection. In miRNA transfection experiments, miRNAs are delivered with liposomes as vectors as opposed to being directly injected into cells. This may introduce an additional delay as miRNAs need to traffic from liposomes to endosomes and cytoplasm. To account for this delay, we introduce an additional compartment V to our model. V_0 miRNAs are delivered to this compartment at time $t = 0$, from where they can either decay with rate $\tilde{d}$ or transfer to the cytoplasm X with rate r. The ODE system describing these dynamics can be written as

$$\left\{ \begin{array}{lcl} \frac{dV}{dt} &=& -\left(\tilde{d} + r\right) V \\ \frac{dX}{dt} &=& rV - (d + \gamma R)X + uRa \\ \frac{da}{dt} &=& \gamma X - ua \end{array} \right. \tag{21}$$

Given the initial condition $V(0) = V_0$, the amount of miRNAs in compartment V decays exponentially as $V(t) = V_0 e^{-\left(\tilde{d}+r\right)t}$. Substituting $V(t)$ into equation 21 leads to

$$\left\{ \begin{array}{lcl} \frac{dX}{dt} &=& rV_0 e^{-\left(\tilde{d}+r\right)t} - (d + \gamma R)X + uRa \\ \frac{da}{dt} &=& \gamma X - ua \end{array} \right. \tag{22}$$

This system is still linear but non-homogeneous. With the method used to solve equation 6 and under the initial conditions $X(0) = 0$ and $a(0) = 0$, one can show that $X(t)$ and $a(t)$ now follow dynamics determined by a linear combination of three exponentially-decaying terms

$$\left\{ \begin{array}{lcl} X(t) &=& k_1 e^{-\beta_1 t} + k_2 e^{-\beta_2 t} - (k_1 + k_2) e^{-\left(\tilde{d}+r\right)t} \\ a(t) &=& \tilde{k}_1 e^{-\beta_1 t} + \tilde{k}_2 e^{-\beta_2 t} - (\tilde{k}_1 + \tilde{k}_2) e^{-\left(\tilde{d}+r\right)t} \end{array} \right. \tag{23}$$

with

$$\left\{ \begin{array}{lcl} \beta_{1,2} &=& \frac{d+\gamma R+u \pm \sqrt{(d+\gamma R+u)^2 - 4ud}}{2} \\ k_1 &=& \frac{rV_0}{\beta_2 - \beta_1} \frac{\beta_2 + d + \gamma R}{\beta_1 + \tilde{d} + r} \\ k_2 &=& -\frac{rV_0}{\beta_2 - \beta_1} \frac{\beta_1 + d + \gamma R}{\beta_2 + \tilde{d} + r} \\ \tilde{k}_1 &=& \frac{\beta_1 + d + \gamma R}{uR} k_1 \\ \tilde{k}_2 &=& \frac{\beta_2 + d + \gamma R}{uR} k_2 \end{array} \right.$$

Combining Ago loading with mRNA and protein dynamics to describe the effects of miRNA transfection. We can now combine equations 12, 14 and 20 which describe changes in mRNA and protein levels with equation 23 that describes the dynamics of Ago loading with miRNAs upon miRNA transfection. Because miRNAs appear to find their targets fast once they are incorporated in Ago (Béthune et al., 2012), one can assume that changes in translation and mRNA decay are proportional to the fraction of loaded Ago. Namely, we describe the fold *increase* in mRNA decay and *decrease* in translation rates by the following equations that depend on the time-dependent fraction of Ago complexed with miRNAs

$$\frac{d_0 + \Delta d(t)}{d_0} = 1 + \delta a(t) \quad \Leftrightarrow 1 + \frac{\Delta d(t)}{d_0} = 1 + \delta a(t) \tag{24}$$

$$\frac{l_0}{l_0 + \Delta l(t)} = 1 + \lambda \delta a(t) \quad \Leftrightarrow 1 + \frac{\Delta l(t)}{l_0} = \frac{1}{1 + \lambda \delta a(t)} \tag{25}$$

Here δ and λ are the proportionality factors relating the fraction of loaded Ago to the relative change in mRNA decay and translation. Note that λ quantifies the relative magnitude of the miRNA effect on translation rate relative to mRNA decay rate. A value of λ between 0 and 1 corresponds to the case where miRNAs impact mostly mRNA decay. $\lambda = 1$ on the other hand indicates that miRNAs have a comparable effect on translation and mRNA decay and $\lambda > 1$ corresponds to the case in which miRNAs function mostly as translational repressors.

Substituting equation 25 into equation 12 leads to the following differential equation for changes in mRNA levels following miRNA transfection

$$\frac{df_m}{dt} = d_0 \left[1 - (1 + \delta a(t))\, f_m\right] \tag{26}$$

Similarly, substituting equation 25 into equation 14, leads to the following dynamics for the protein changes estimated through SILAC upon miRNA transfection

$$\frac{df_p}{dt} = s\left(\frac{f_m}{1 + \lambda\delta a(t)} - f_p\right). \tag{27}$$

Finally, substituting equation 25 into equation 20 leads to the following dynamics of protein changes estimated through pSILAC following a miRNA transfection

$$\frac{df_p}{dt} = \frac{s}{1 - e^{-s(t-8)}}\left(\frac{f_m}{1 + \lambda\delta a(t)} - f_p\right). \tag{28}$$

Re-parameterization of mRNA and protein dynamics under fixed Ago loading parameters and experimentally measured mRNA and protein decay rates. According to equations 26, 27 and 28, changes in mRNA and protein levels are determined by 10 parameters: $\gamma, u, d, \delta, r, v_0, \tilde{d}, \lambda, s, d_0$.

Estimates of mammalian mRNA half-lives are available from several studies in the literature. Friedel et al. (2009) estimated a median decay rate of $0.15h^{-1}$ in murine fibroblasts and $0.13h^{-1}$ in human B-cells, while Schwanhäusser et al. (2011) reported a median mRNA decay rate of $0.08h^{-1}$ in murine fibroblasts. We therefore set the mRNA decay rate d_0 to the average value of $0.12h^{-1}$, which corresponds to a half-life of 5.8 hours. We obtained the protein decay rates s from measurements of protein turn-over in HeLa of Cambridge et al. (2011) which provide a broader coverage than our targeted proteomics approach (SRM).

If we further fix the Ago loading parameters γ, u, d to the values obtained by fitting the fluorescence cross-correlation spectroscopy-based measurements of Ago and siRNAs in complex as a function of time Ohrt et al. (2008) (see Fig. S1), changes in mRNA levels following miRNA transfection are determined by four parameters: $\delta, r, V_0, \tilde{d}$. According to equation 23, r and V_0 always occur in a product while $\tilde{d}$ and r always occur in a sum. We can therefore reduce the number of parameters to three, namely δ, rV_0 and $\tilde{d}+r$. The number of parameters can be reduced further by introducing a new function $\tilde{a}(t)$ which represents the time-dependent relative change in the mRNA decay rate (equation 25), defined as

$$\tilde{a}(t) = \delta a(t) \tag{29}$$

By plugging equation 29 into equation 23, we can express the relative change in mRNA decay $\tilde{a}(t)$ as a tri-exponential function

$$\tilde{a}(t) = \tilde{k}_1 e^{-\beta_1 t} + \tilde{k}_2 e^{-\beta_2 t} - (\tilde{k}_1 + \tilde{k}_2) e^{-(\tilde{d}+r)t} \tag{30}$$

with

$$\begin{cases} \beta_{1,2} &= \frac{d+\gamma R+u \pm \sqrt{(d+\gamma R+u)^2 - 4ud}}{2} \\ \tilde{k}_1 &= \frac{\beta_1+d+\gamma R}{uR}\frac{\delta r V_0}{\beta_2-\beta_1}\frac{\beta_2+d+\gamma R}{\beta_1+\tilde{d}+r} \\ \tilde{k}_2 &= -\frac{\beta_2+d+\gamma R}{uR}\frac{\delta r V_0}{\beta_2-\beta_1}\frac{\beta_1+d+\gamma R}{\beta_2+\tilde{d}+r} \end{cases}$$

which shows that δ and rV_0 (contained in $\tilde{k}_{1,2}$) always occur in a product. Therefore, under the 3-exponential Ago loading model of equation 23, mRNA dynamics following miRNA transfection are determined by two free parameters, namely $\delta r V_0$ and $\tilde{d} + r$. Substituting $\tilde{a}(t)$ (equation 29) in equations 26 leads to a differential equation describing changes in mRNA levels following miRNA transfection

$$\frac{df_m}{dt} = d_0 \left[1 - (1 + \tilde{a}(t))\, f_m\right] \tag{31}$$

Similarly, substituting $\tilde{a}(t)$ into equation 27, leads to the following dynamics for the protein changes estimated through SILAC upon miRNA transfection

$$\frac{df_p}{dt} = s\left(\frac{f_m}{1 + \lambda\tilde{a}(t)} - f_p\right). \tag{32}$$

Finally, substituting $\tilde{a}(t)$ into equation 20 leads to the following dynamics of protein changes estimated through pSILAC following a miRNA transfection

$$\frac{df_p}{dt} = \frac{s}{1-e^{-s(t-8)}}\left(\frac{f_m}{1+\lambda\tilde{a}(t)} - f_p\right). \tag{33}$$

Modelling mRNA and protein dynamics following miRNA induction When the miRNA is induced (by doxycyclin in the case of our experiment), there is no need to account for the endosomal compartment. We therefore start from equation 6 and slightly alter the model by assuming that miRNA synthesis $c(t)$ is inactive before 0h and follows 0^{th} order kinetics upon doxycyclin induction, that is

$$\left\{\begin{array}{rcl} \frac{dX}{dt} &=& c(t) - (d+\gamma R)X + uRa \\ \frac{da}{dt} &=& \gamma X - ua \end{array}\right. \tag{34}$$

with

$$c(t) = \left\{\begin{array}{ll} 0 & \text{if } t < 0 \\ X_0 d & \text{if } t \geq 0 \end{array}\right.$$

As previously, $a(t)$ represents the fraction of loaded Ago and $X(t)$ the amount of free miRNA in the cell. The parameters d, γ, R, u are defined as in equation 6. Under this model and this parametrization, we have $X = X_0$ and $a = X_0\frac{\gamma}{u}$ at steady-state.

For $t \geq 0$, the solution to equation 34 is

$$\left\{\begin{array}{rcl} X(t) &=& X_0\left(1 - \frac{\beta_2-d}{\beta_2-\beta_1}e^{-\beta_1 t} + \frac{\beta_1-d}{\beta_2-\beta_1}e^{-\beta_2 t}\right) \\ a(t) &=& \frac{X_0\gamma}{u}\left(1 - \frac{\beta_2}{\beta_2-\beta_1}e^{-\beta_1 t} + \frac{\beta_1}{\beta_2-\beta_1}e^{-\beta_2 t}\right) \end{array}\right. \tag{35}$$

with

$$\beta_{1,2} = \frac{d+\gamma R+u \pm \sqrt{(d+\gamma R+u)^2 - 4ud}}{2}.$$

From there, we obtain the relative change in the mRNA decay rate $\tilde{a}(t)$

$$\tilde{a}(t) := \delta a(t) = X_0\delta\frac{\gamma}{u}\left(1 - \frac{\beta_2}{\beta_2-\beta_1}e^{-\beta_1 t} + \frac{\beta_1}{\beta_2-\beta_1}e^{-\beta_2 t}\right). \tag{36}$$

As in the transfection case (equation 30), we use the values infered from the experiments of Ohrt et al. (2008) for the parameters γ, d, u. We model changes in mRNA and protein abundance following miRNA induction by combining equation 36 which describes the change in the mRNA decay rate $\tilde{a}(t)$ and equations 31 and 32 which relate $\tilde{a}(t)$ to mRNA and protein dynamics.

Examining equation 36 shows that mRNA changes upon miRNA induction $f_m(t)$ are driven by one free parameter $X_0\delta$. As in the transfection case, modeling protein dynamics $f_p(t)$ require an additional parameter λ, which accounts for the effect of miRNAs on translation relative to mRNA decay.

3.3 Processing of quantitative proteomics, microarrays and deep sequencing data

mRNA sequences, gene to mRNA mappings, mRNA to protein mappings. The RefSeq mRNA database downloaded from NCBI on Jan 18th 2011 was used for all analyses presented in this manuscript. For each gene, we defined a representative mRNA as the longest mRNA in RefSeq with annotated 5' UTR, coding domain and 3' UTR. This representative mRNA was used for subsequent sequence analyses.

Defining putative miRNA target genes Putative miRNA target genes were defined as genes whose representative mRNA carried at least one match to positions 1-7, 2-8 or 1-8 of the perturbed miRNA in the 3' UTR.

Computational analysis of one-channel Affymetrix microarrays from Selbach et al. (2008) and Wang and Wang (2006). The CEL files of Selbach et al. (2008) were downloaded from http://psilac.mdc-berlin.de/download/ and imported into the R software (www.R-project.org) with the BioConductor affy package (Gentleman et al., 2004). Probe intensities were corrected for optical noise, adjusted for non-specific binding and quantile normalized with the gcRMA algorithm (Wu et al., 2004). Per gene $\log_2$ fold change were obtained through the following procedure. We first fitted a lowess model of the probe $\log_2$ fold change using the probe A/U content. We used this model to correct for the technical bias of A/U content on probe-level $\log_2$ fold change reported by Elkon and Agami (2008). Subsequently, probe set-level $\log_2$ fold changes were defined as the median probe-level $\log_2$ fold change. Probe sets with more than 2 probes mapping ambiguously (more than 1 match) to the genome

were discarded, as were probe sets that mapped to multiple genes. We then collected all remaining probe sets matching a given gene, and averaged their $\log_2$ fold changes to obtain an expression change per gene. Finally, we considered all genes for which at least one probeset was called present in the transfection experiments as expressed, and retained them for further analyses.

Computational analysis of two-channel Agilent microarrays from Baek et al. (2008) and Karginov et al. (2007). The Baek data set was downloaded from the GEO database of NCBI (Barrett et al., 2011) (accession GSE11968) while the text files from the Agilent scanner were kindly provided to us by Ted Karginov. We extracted the rProcessedSignal, gProcessedSignal, LogRatio, rIsWellAboveBG, gIsWellAboveBG fields for each probe, keeping only probes for which both gIsWellAboveBG and rIsWellAboveBG flags were true in all experiments. We then quantile-normalized the green and red channel intensities which we obtained from the rProcessedSignal and gProcessedSignal fields of all experiments together. We computed probe-level $\log_2$ fold changes from the quantile-normalized rProcessedSignals and gProcessedSignals. After discarding probes mapping to multiple genes, we collected all probes matching a given gene, and we estimated the $\log_2$ fold change per gene as the average $\log_2$ fold change of the probe sets associated with it.

Computational analysis of SILAC data from Baek et al. (2008). We downloaded the data provided by the authors in the supplementary material of the paper and used it without additional post-processing.

Computational analysis of pSILAC data from Selbach et al. (2008). We downloaded the 'all peptide evidence' flat file from http://psilac.mdc-berlin.de/download/. We mapped all peptides in the pSILAC data set against the RefSeq Protein database from Aug, 14th 2008 using wu-blastp 2.0 and a seed word length of 5, discarding alignments with gaps or with more than one mismatch. We further discarded peptides that mapped to more than one protein. Per-protein $\log_2$ fold changes were computed for all proteins credited with at least 3 peptides $\log_2$ fold changes across replicates and gel slices.

Computational analysis of the Ribosome Protected Fragment sequencing and mRNAseq data of Guo et al. (2010) We retrieved the processed data from the GEO database of NCBI (Barrett et al., 2011) (accession GSE22004). We only considered genes with at least 1 RPKM in all the 12 HeLa samples (8573 genes) or all the 4 mouse samples (9027 genes). We focused on the 852 and 938 genes that had a seed match to miR-155 and miR-1 and whose mRNA levels were down-regulated following transfection. Fold changes in translation efficiency were defined as the ratio between the fold change in mRNA level and the fold change in RPF.

3.4 The role of miRNA dynamics in the estimation of protein-level effects in miRNAs transfection experiments

The experiments of Baek et al. (2008); Selbach et al. (2008) provided extensive data about the effects of miRNAs on mRNA and protein abundance. As initially shown, perturbing miRNA expression had the expected effect on target mRNA and protein abundance — down-regulation in transfection experiments and up-regulation in knock-out experiments — in 7 of the 9 experiments (Fig. S10A). Yet, in 6 of the 9 experiments, changes in protein abundance were *smaller* than changes in the levels of the cognate mRNAs. This is unexpected because the change in protein should reflect the combined effect of the perturbed miRNA on the level as well as on the translation rate of the mRNA. As a result, the change in protein abundance should be at least as pronounced as the change in mRNA abundance. In addition, the best correlation of mRNA-protein level changes was obtained for the experiment in which miR-223 was knocked out and cells were cultured for 8 days before mRNA and protein profiling. All other experiments that quantified changes in mRNA and protein abundance 24 to 48 hours after transfection lead to smaller correlation coefficients (Table S2). This suggests that the transient nature of the transfection combined with delays in the miRNA-regulatory cascade may obscure the impact of miRNAs on target proteins.

We tested this hypothesis by comparing the ability of two competing models of miRNA-induced changes in mRNA and protein abundance to explain the quantitative proteomics measurements (see detailed methods in Section 3.5). The first model M_0 assumed that miRNAs affect the decay and translation rates of their target mRNA, while the rate of transcription and of protein decay remain constant. With these assumptions, one can show that at steady-state, the fold change in protein level is given by the product of the fold change in translation rate and the fold change in mRNA level (see equation 38). The second model $M_\perp$ assumed that the miRNA affects independently the mRNA and protein abundance of its targets. To apply these models at the level of the entire transcriptome and proteome given that we do not precisely know which transcripts are directly targeted by the miRNA, we further assumed that a fraction ρ of all the transcripts that contain matches to the 'seed' of the perturbed miRNA are *bona fide* miRNA targets. These transcripts contain 'functional' miRNA

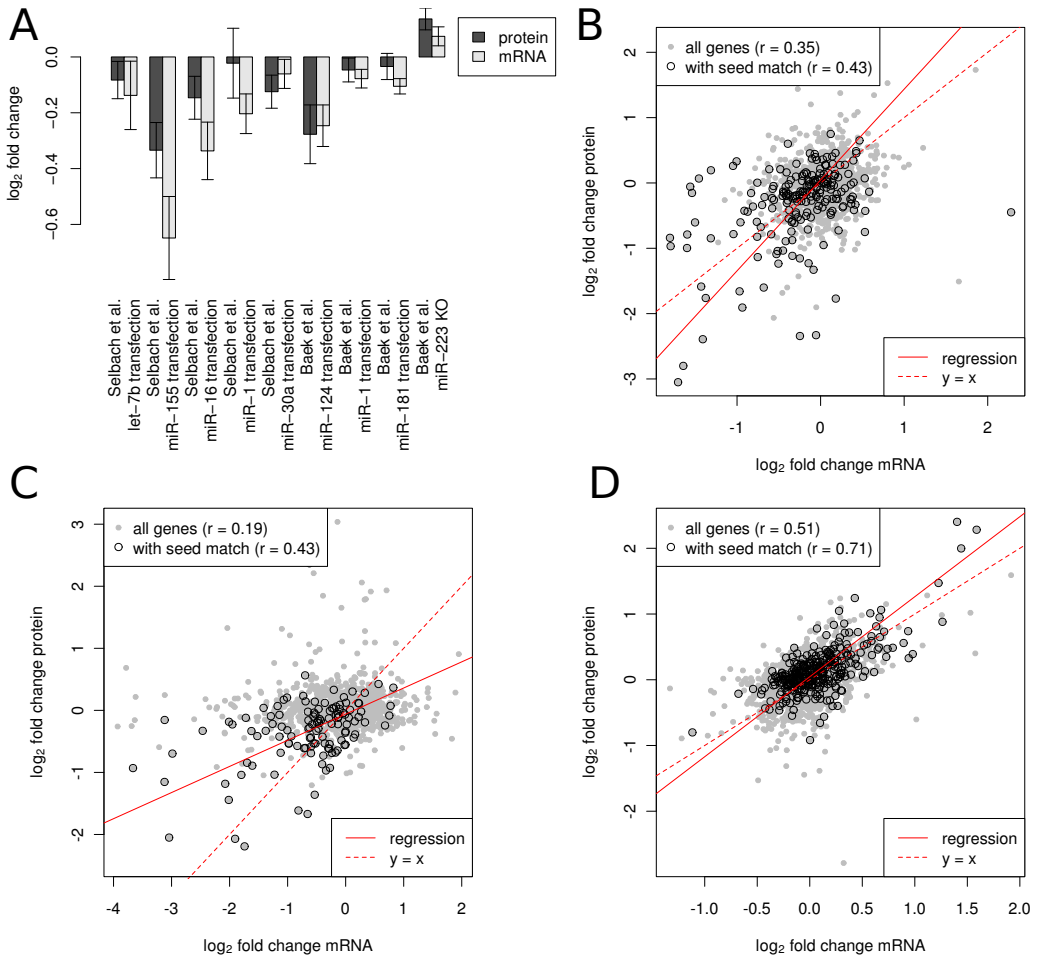

Figure S10: Magnitude of changes mRNA and protein abundance of miRNA targets in miRNA perturbation experiments of various designs. A: Average $\log_2$ fold changes in mRNA and protein levels for genes with seed matches, in the miRNA perturbation experiments of Baek et al. (2008) and Selbach et al. (2008). Error bars represent 95% confidence intervals on the mean. B, C, D: Scatter plots of mRNA-protein level changes for three different experimental designs: B: miR-124 transfection of Baek et al., with mRNA measurements at $t = 24$ hours and protein measurements at 48 hours after transfection. C: miR-155 transfection of Selbach et al., with mRNA and protein measurements at $t = 32$ hours. D: miR-223 knock-out of Baek et al., with mRNA and protein measurements at $t = 8$ days. Each grey dot represents a gene, and circled dots represent genes with seed matches to the perturbed miRNA in the 3' UTR.

binding sites which induce changes in mRNA and protein levels with specific means and variances. On the other hand, transcripts that carry either non-functional 'seed' matches or no 'seed' match at all undergo changes that are described by a different set of means and variances. Simulated data shown in Fig. S11A and Fig. S11B allow us to build an intuition for the mRNA and protein changes that we expect under these two models. Both M_0 and $M_\perp$ produce positive correlations between changes in mRNA and protein levels. Furthermore, with both models, genes that carry functional binding sites to the over-expressed miRNA show the largest differential expression while genes with non-functional binding sites behave like genes without binding site. However, the structures of the two scatters of mRNA-protein changes are quite different, reflecting the different assumptions made by M_0 and $M_\perp$. By comparing the likelihood of the data under the two models (Table S2), we see that the model $M_\perp$ accounts best for the miR-124 and miR-155 transfection experiments of Baek et al. (2008) and Selbach et al. (2008), the corresponding odds being $10^7 : 1$ relative to the M_0 model. Furthermore, the inferred change in translation μ_P in these two experiments is positive under the M_0 model, which is clearly inconsistent with the widely accepted function of miRNAs as translation repressors. Thus, although mRNA and protein levels of miRNA targets decreased in both experiments, changes in protein levels did not reflect changes in mRNA levels. However, the miR-223 knock-out is much better explained by the M_0 model (odds of M_0 relative to $M_\perp$ are $10^{12} : 1$), indicating that on longer time-scales, miRNA-induced changes in protein abundance do reflect changes in mRNA levels. The other six (of the total of 9) miRNA transfection experiments of Baek et al. (2008) and Selbach et al. (2008) are consistent with this conclusion (Fig. S12, S13 and Table S2).

Dataset	Model	ρ	$\mu_{M,+}$	$\sigma_{M,+}$	$\mu_{P,+}$	$\sigma_{P,+}$	$\mathcal{L}(\theta^*\|f_M, f_P)$
Selbach let-7b	M_0	0.07	0.90	0.38	-0.99	0.19	-194.19
Selbach let-7b	$M_\perp$	0.02	-1.66	0.31	-0.76	0.09	-137.45
Selbach miR-155	M_0	0.17	-1.93	0.84	1.45	0.80	-268.97
Selbach miR-155	$M_\perp$	0.40	-1.18	0.94	-0.62	0.72	-237.23
Selbach miR-16	M_0	0.24	-1.22	0.59	0.91	0.63	-371.46
Selbach miR-16	$M_\perp$	0.26	-1.11	0.66	-0.56	0.76	-331.56
Selbach miR-1	M_0	0.30	-0.61	0.46	0.43	0.60	-128.08
Selbach miR-1	$M_\perp$	0.34	-0.58	0.45	-0.30	0.44	-111.46
Selbach miR-30a	M_0	0.15	-0.46	0.44	0.01	0.62	-77.31
Selbach miR-30a	$M_\perp$	0.15	-0.53	0.38	-0.70	0.37	-66.69
Baek miR-124	M_0	0.28	-0.69	0.73	0.02	0.99	-293.55
Baek miR-124	$M_\perp$	0.30	-0.68	0.69	-0.78	0.82	-275.55
Baek miR-1	M_0	0.15	-0.46	0.30	0.54	0.38	-95.02
Baek miR-1	$M_\perp$	0.19	-0.36	0.33	-0.18	0.49	-67.99
Baek miR-181	M_0	0.15	-0.53	0.25	0.55	0.27	-153.14
Baek miR-181	$M_\perp$	0.41	-0.21	0.32	-0.03	0.17	-63.31
Baek miR-223	M_0	0.15	0.45	0.50	0	0.40	-71.03
Baek miR-223	$M_\perp$	0.15	0.53	0.44	0.64	0.55	-99.34

Table S2: Maximum-likelihood parameter estimates under models M_0 and $M_\perp$, upon fitting all transfection experiments of Selbach et al. (2008), as well as the transfection and miRNA knock-out experiments of Baek et al. (2008).

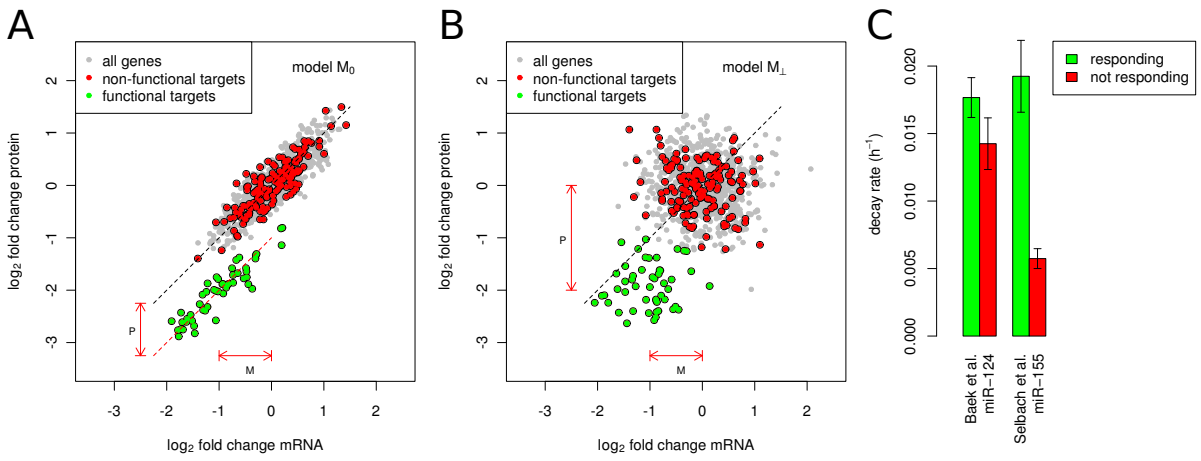

Figure S11: Simulated changes in mRNA and protein levels under model M_0 (A) and $M_\perp$ (B). In both cases, the miRNA was set to induce a 4-fold repression of target proteins and a 2-fold repression of the target mRNAs. μ_M and μ_P represent the average change in target mRNA levels and translation rates. C: Comparison of decay rates of transitions from proteins found to be down-regulated in the miR-124 and miR-155 transfections of Baek et al. (2008) ($n = 100$) and Selbach et al. (2008) ($n = 89$) relative to proteins that did not change in the same transfection experiments ($n = 69$ and $n = 63$). Error bars represent the standard error on the mean. Both sets of proteins have been selected on the basis of the corresponding mRNA having a seed match to the transfected miRNA in the 3'UTR.

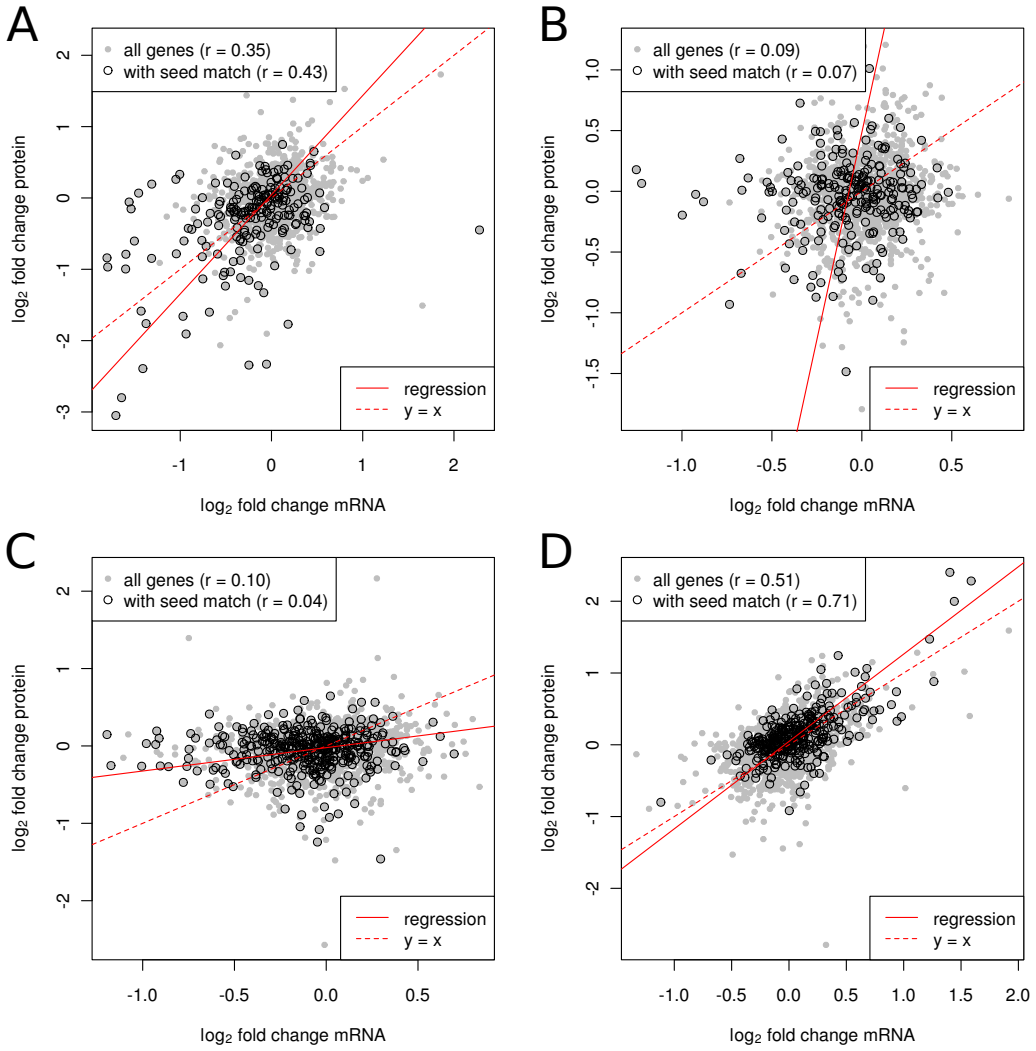

Figure S12: $\log_2$ fold change in protein and mRNA levels following miRNA perturbation in the experiments of Baek et al. (2008). A: miR-124 transfection, B: miR-1 transfection, C: miR-181 transfection and D: miR-223 knock out. Each dot represents a gene with genes carrying seed matches to the perturbed miRNA marked with black circles. The Pearson correlation coefficient for all genes or genes with seed matches is reported in the legend.

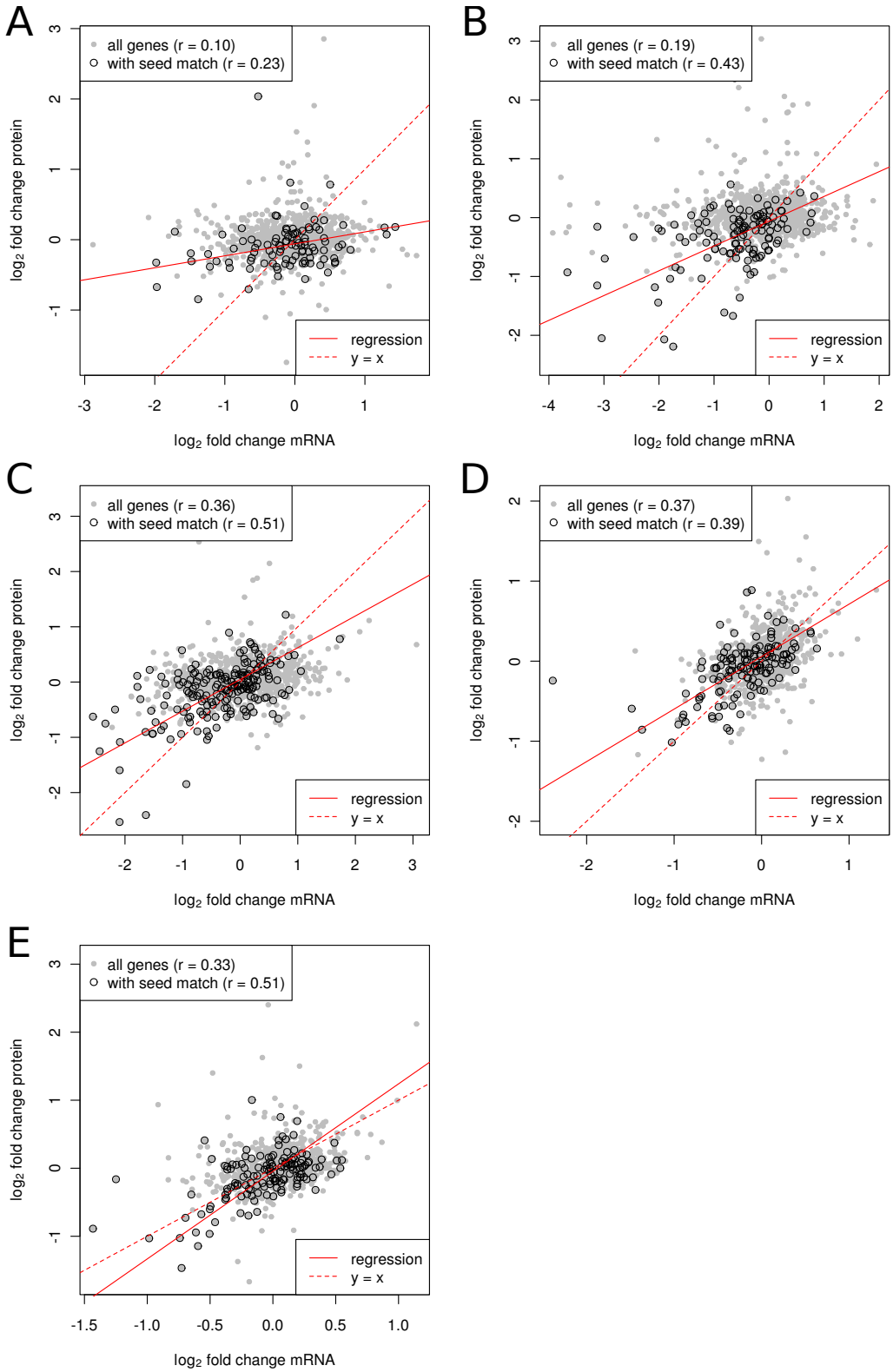

Figure S13: $\log_2$ fold change in protein and mRNA levels following miRNA transfection in the experiments of Selbach et al. (2008). A: let-7b, B: miR-155, C: miR-16, D: miR-1 and E: miR-30a. Each dot represents a gene with genes carrying seed matches to the perturbed miRNA marked with black circles. The Pearson correlation coefficient for all genes or genes with seed matches is reported in the legend.

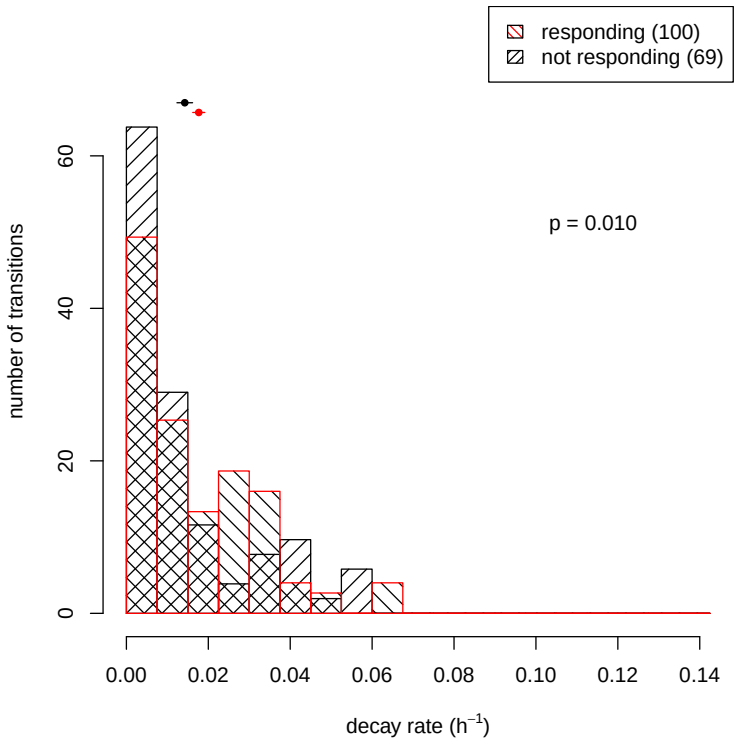

Figure S14: Comparison of decay rates of transitions corresponding to responding (red) and non-responding (black) proteins in the miR-124 transfection experiment of Baek et al. (2008). The red and black dots show the average decay rate of transitions from responding and non-responding proteins with error bars representing the standard error. The decay rates of transitions corresponding to responding and non-responding proteins were compared using Wilcoxon's rank sum test ($p = 0.01$).

A key assumption of the M_0 model is that mRNA and protein levels are measured at steady-state, when changes in protein abundance do not depend on protein decay rates. That this model does not explain well the data from transfection experiments suggests that mRNA and protein levels are not at steady-state, with fast-decaying proteins showing larger regulation than stable proteins. We tested for this possibility by selecting miRNA targets genes based on the presence of a 'seed' match to miR-124 or miR-155 in the 3' UTR and on the protein-level response to the miRNA transfection. Namely, we selected 'responding' proteins, that underwent the strongest repression upon miRNA transfection and 'non-responding' proteins that changed least upon transfection. By selected reaction monitoring (SRM) of isotopically labeled proteins (Lange et al., 2008), we then measured the decay rates of 'responding' and 'non-responding' proteins in the miR-124 and miR-155 transfection experiments of Baek et al. (2008) and Selbach et al. (2008) (as described in Section 3.6). Proteins that did not 'respond' in the miR-124 transfection SILAC experiment of Baek et al. (2008) decayed 24% slower ($p = 0.01$, Wilcoxon's rank sum test, Fig. S14) and those that did not 'respond' in the miR-155 pSILAC experiment of (Selbach et al., 2008) decayed 3 fold slower ($p < 10^{-8}$, Fig. S15) than proteins that were considered 'responders' in these experiments whereas proteins that emerged as 'responders' in these two experiments had comparable decay rates.

Thus, miRNA targets typically changed less at the protein compared to the mRNA level in most transfection experiments. Changes in protein abundance appeared uncoupled from changes in mRNA abundance, and fast-decaying proteins were more likely to be detected as miRNA targets than stable proteins. These observations suggested that a deeper understanding of miRNA-mediated gene regulation would benefit from a quantitative framework in which to interpret the experiments.

3.5 Relationship between changes in target mRNA and protein abundance of targets in response to miRNA perturbation

We now describe in detail the M_0 and $M_\perp$ models that we used in the previous section to infer that changes in mRNA and protein levels are temporarily decoupled in transfection experiments.

The M_0 model is based on the steady-state solution of an ordinary differential equation (ODE) model of miRNA-dependent regulation originally introduced by Khanin and Higham (2009) and illustrated in the figure

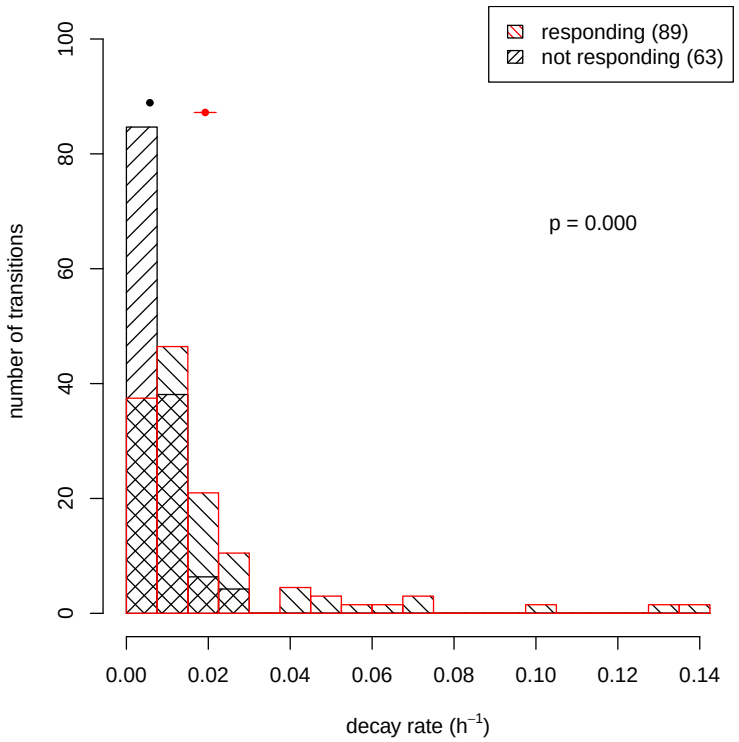

Figure S15: Comparison of decay rates of transitions corresponding to responding (red) and non-responding (black) proteins in the miR-155 transfection experiment of Selbach et al. (2008). The red and black dots show the average transition decay rate of responding and non-responding with error bars representing the standard error. The decay rates of transitions corresponding to responding and non-responding proteins were compared using Wilcoxon's rank sum test ($p < 10^{-8}$).

above, namely

$$\begin{cases} \frac{dM}{dt} &= c - d_M M \\ \frac{dP}{dt} &= lM - d_P P \end{cases} \tag{37}$$

M is the mRNA concentration, P is the protein concentration, c is the transcription rate, d_M the mRNA decay rate, l the translation rate and d_P the protein decay rate. We assumed that miRNAs regulate the decay and translation rates of their target mRNA, leaving transcription and protein decay rates unchanged. Unless specified otherwise, we also assumed that miRNAs have instant action on target mRNA decay and translation rates following the perturbation. Therefore, when the expression of the miRNA is perturbed at time $t = 0$, the mRNA decay and translation rates d_M and l are changed into d'_M and l'.

The steady state of the system is then:

$$\begin{cases} M^* &= \frac{c}{d'_M} \\ P^* &= \frac{l'c}{d_M d_P} \end{cases}$$

If the system was at steady-state prior to the miRNA perturbation, we have the following initial condition:

$$\begin{cases} M(0) &= \frac{c}{d_M} \\ P(0) &= \frac{lc}{d_M d_P} \end{cases}$$

And therefore, the fold changes in mRNA and protein levels following miRNA perturbation are

$$\begin{cases} \frac{M^*}{M(0)} &= \frac{d_M}{d'_M} \\ \frac{P^*}{P(0)} &= \frac{l' d_M}{l d'_M}. \end{cases}$$

Thus, at steady-state, the log fold change in protein level can be expressed in terms of the sum of the log fold change in translation rate and mRNA level:

$$\log \frac{P^*}{P(0)} = \log \frac{l'}{l} + \log \frac{M^*}{M(0)} \tag{38}$$

The $M_\perp$ model on the other hand assumed that miRNAs-induced changes in protein levels are independent from changes in their cognate mRNAs levels. Thus, rather than depending on the change in the mRNA level, the change in protein is given by some (unknown) constant k:

$$\log \frac{P^*}{P(0)} = k$$

This model is of course unrealistic from a mechanistic point of view. It aims at capturing unknown and potentially complex lags and feedbacks in the miRNA regulation pathway in the simplest possible way.

Fitting the M_0 and $M_\perp$ models To determine which of these two models best explains the combined transcriptomics and proteomics data we used the following probabilistic framework. Let f_M and f_P be two vectors that represent the $\log_2$ changes in mRNA and protein levels in one miRNA perturbation experiment. Let us assume that genes that contain seed matches to the transfected miRNA can be split into two categories: functional targets of the perturbed miRNA (labeled as $+$) and non-functional targets of the miRNA (labeled as $-$), and that each category has its associated distributions of mRNA and protein level changes. Under the M_0 model, we write the likelihood of the data for gene i as:

$$\left\{\begin{array}{lcl} P(f_{M,i}|+) & = & \mathcal{N}\left(\mu_{M,+}, \sigma^2_{M,+}\right) \\ P(f_{P,i}|+) & = & \mathcal{N}\left(f_{M,i} + \mu_{P,+}, \sigma^2_{P,+}\right) \\ P(f_{M,i}|-) & = & \mathcal{N}\left(\mu_{M,-}, \sigma^2_{M,-}\right) \\ P(f_{P,i}|-) & = & \mathcal{N}\left(f_{M,i} + \mu_{P,-}, \sigma^2_{P,-}\right) \end{array}\right.$$

where $\mathcal{N}$ stands for the density of a Gaussian probability distribution. Parameters $\mu_{M,-}$, $\mu_{P,-}$, $\sigma_{M,-}$, $\sigma_{P,-}$ capture the secondary effects on mRNAs and proteins that are not functional targets of the transfected miRNA and are fitted from the genes carrying no seed match to the transfected miRNA. $\mu_{M,+}$, $\mu_{P,+}$, $\sigma_{M,+}$ and $\sigma_{P,+}$ capture the regulatory effect of direct targeting by the miRNA. We see that the model implies that $f_P = f_M + \mu_P$ on average, that is `protein change = mRNA change + translation change`, consistent with the ODE model underlying M_0. $\mu_{M,+}$, $\mu_{P,+}$, $\sigma_{M,+}$ and $\sigma_{P,+}$ cannot be estimated directly since we do not know *a priori* which genes are functional miRNA targets and which others are not. However, if we knew $\mu_{M,+}$, $\mu_{P,+}$, $\sigma_{M,+}$ and $\sigma_{P,+}$, we could compute which genes are likely to be functional miRNA targets. The reverse is true as well: if we knew which genes are functional miRNA targets, we could estimate $\mu_{M,+}$, $\mu_{P,+}$, $\sigma_{M,+}$ and $\sigma_{P,+}$. Therefore, we estimate $\mu_{M,+}$, $\mu_{P,+}$, $\sigma_{M,+}$ and $\sigma_{P,+}$ by Expectation Maximization (EM) (Dempster et al., 1977) over a single hidden variable $X_{i,+}$, namely the probability that gene i is a functional miRNA target.

We first write the posterior probability of observing a functional miRNA target site given the data (f_P, f_M):

$$\begin{array}{rcl} P(+|f_M, f_P) & = & \frac{P(f_M, f_P|+)P(+)}{P(f_P, f_M)} \\ & = & \frac{P(f_P|f_M,+)P(f_M|+)P(+)}{P(f_P|f_M)P(f_M)} \\ & = & \frac{P(f_P|f_M,+)P(f_M|+)\rho}{P(f_P|f_M,+)P(f_M|+)\rho + P(f_P|f_M,-)P(f_M|-)(1-\rho)} \end{array}$$

where ρ is the fraction of functional targets among the n genes whose representative mRNA carries a seed match to the perturbed miRNA in the 3'UTR. We can now compute the log-likelihood of the parameters $\theta = (\sigma_{M,+}, \sigma_{P,+}, \mu_{M,+}, \mu_{P,+}, \rho)$:

$$\left\{\begin{array}{lcl} \mathcal{L}(\theta|f_{P,i}, f_{M,i}, +) & = & \log\rho - \log\sigma_{P,+} - \log\sigma_{M,+} - \log 2\pi \\ & & -\frac{1}{2}\left(\frac{f_{P,i} - (f_{M,i} + \mu_{P,+})}{\sigma_{P,+}}\right)^2 - \frac{1}{2}\left(\frac{f_{M,i} - \mu_{M,+}}{\sigma_{M,+}}\right)^2 \\ \mathcal{L}(\theta|f_{P,i}, f_{M,i}, -) & = & \log\rho - \log\sigma_{P,-} - \log\sigma_{M,-} - \log 2\pi \\ & & -\frac{1}{2}\left(\frac{f_{P,i} - (f_{M,i} + \mu_{P,-})}{\sigma_{P,-}}\right)^2 - \frac{1}{2}\left(\frac{f_{M,i} - \mu_{M,-}}{\sigma_{M,-}}\right)^2 \end{array}\right.$$

and we can write the E-step of the expectation maximization algorithm as:

$$X^{(t)}_{i,+} = P\left(+|f_P, f_M, \theta^{(t)}\right)$$

and the M-step as:

$$\rho^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} X^{(t)}_{i,+}$$

$$
\begin{aligned}
\mu_{M,+}^{(t+1)} &= \frac{\sum_{i=1}^{n} X_{i,+}^{(t)} f_{M,i}}{\sum_{i=1}^{n} X_{i,+}^{(t)}} \\
\sigma_{M,+}^{(t+1)} &= \sqrt{\frac{\sum_{i=1}^{n} X_{i,+}^{(t)} \left(f_{M,i} - \mu_{M,+}^{(t+1)}\right)^2}{\sum_{i=1}^{n} X_{i,+}^{(t)}}} \\
\mu_{P,+}^{(t+1)} &= \frac{\sum_{i=1}^{n} X_{i,+}^{(t)} (f_{P,i} - f_{M,i})}{\sum_{i=1}^{n} X_{i,+}^{(t)}} \\
\sigma_{P,+}^{(t+1)} &= \sqrt{\frac{\sum_{i=1}^{n} X_{i,+}^{(t)} \left(f_{P,i} - f_{M,i} - \mu_{P,+}^{(t+1)}\right)^2}{\sum_{i=1}^{n} X_{i,+}^{(t)}}}
\end{aligned}
$$

We then iterate between the E- and M- steps until the log-likelihood $\mathcal{L}(\theta | f_{P,i}, f_{M,i})$ converges.

In the alternative model $M_\perp$, changes in protein level depend only on the action of miRNA on the translation rate, not on changes in the level of the cognate mRNAs:

$$
\left\{
\begin{array}{lcl}
P(f_{M,i}|+) &=& \mathcal{N}\left(\mu_{M,+}, \sigma_{M,+}^2\right) \\
P(f_{P,i}|+) &=& \mathcal{N}\left(\mu_{P,+}, \sigma_{P,+}^2\right) \\
P(f_{M,i}|-) &=& \mathcal{N}\left(\mu_{M,-}, \sigma_{M,-}^2\right) \\
P(f_{P,i}|-) &=& \mathcal{N}\left(\mu_{P,-}, \sigma_{P,-}^2\right)
\end{array}
\right.
$$

Like with model M_0, we use EM to fit the parameters. This requires two small changes changes in the likelihood function used by the E-step:

$$
\left\{
\begin{array}{lcl}
\mathcal{L}(\theta | f_{P,i}, f_{M,i}, +) &=& \log \rho - \log \sigma_{P,+} - \log \sigma_{M,+} - \log 2\pi \\
& & -\frac{1}{2}\left(\frac{f_{P,i} - \mu_{P,+}}{\sigma_{P,+}}\right)^2 - \frac{1}{2}\left(\frac{f_{M,i} - \mu_{M,+}}{\sigma_{M,+}}\right)^2 \\
\mathcal{L}(\theta | f_{P,i}, f_{M,i}, -) &=& \log \rho - \log \sigma_{P,-} - \log \sigma_{M,-} - \log 2\pi \\
& & -\frac{1}{2}\left(\frac{f_{P,i} - \mu_{P,-}}{\sigma_{P,-}}\right)^2 - \frac{1}{2}\left(\frac{f_{M,i} - \mu_{M,-}}{\sigma_{M,-}}\right)^2
\end{array}
\right.
$$

and the M-step becomes:

$$
\rho^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} X_{i,+}^{(t)}
$$

$$
\begin{aligned}
\mu_{M,+}^{(t+1)} &= \frac{\sum_{i=1}^{n} X_{i,+}^{(t)} f_{M,i}}{\sum_{i=1}^{n} X_{i,+}^{(t)}} \\
\sigma_{M,+}^{(t+1)} &= \sqrt{\frac{\sum_{i=1}^{n} X_{i,+}^{(t)} \left(f_{M,i} - \mu_{M,+}^{(t+1)}\right)^2}{\sum_{i=1}^{n} X_{i,+}^{(t)}}} \\
\mu_{P,+}^{(t+1)} &= \frac{\sum_{i=1}^{n} X_{i,+}^{(t)} f_{P,i}}{\sum_{i=1}^{n} X_{i,+}^{(t)}} \\
\sigma_{P,+}^{(t+1)} &= \sqrt{\frac{\sum_{i=1}^{n} X_{i,+}^{(t)} \left(f_{P,i} - \mu_{P,+}^{(t+1)}\right)^2}{\sum_{i=1}^{n} X_{i,+}^{(t)}}}
\end{aligned}
$$

Log odds ratio to compare the goodness of fit of M_0 and $M_\perp$ To determine which of the two models, M_0 or $M_\perp$, agrees most with the data (f_M, f_P), we computed the odds ratio between the two models given the data. Using Bayes' theorem and equal *a priori* probability of the two models $P(M_0) = P(M_\perp)$, we obtained

$$
\begin{aligned}
\log_{10}\left(\frac{P(M_0|f_M,f_P)}{P(M_\perp|f_M,f_P)}\right) &= \log_{10}\left(\frac{P(f_M,f_P|M_0)P(M_0)}{P(f_M,f_P)} \frac{P(f_M,f_P)}{P(f_M,f_P|M_\perp)P(M_\perp)}\right) \\
&= \log_{10}\left(\frac{P(f_M,f_P|M_0)}{P(f_M,f_P|M_\perp)}\right)
\end{aligned}
$$

To calculate odds ratios, we need to integrate the probability of the data under each model, weighted by the a priori probability of the parameters:

$$
P(f_M, f_P|M_0) = \int P(f_M, f_P|\theta, M_0) P(\theta) d\theta.
$$

This can be done in principle by multi-dimensional numerical integration. Alternatively, we can approximate $P(\theta)$ by a Dirac delta function:

$$
P(\theta) \simeq \left\{
\begin{array}{ll}
1 & \text{if } \theta = \theta^* \\
0 & \text{if } \theta \neq \theta^*
\end{array}
\right.
$$

where θ^* are the parameters inferred by the EM algorithm. Filling this approximation into the expression of the integral gives

$$
P(f_M, f_P|M_0) \simeq P(f_M, f_P|\theta^*, M_0),
$$

where $P(f_M, f_P|\theta^*, M_0)$ is the probability of the data under the maximum likelihood parameter values, whose logarithm is reported in the column $\mathcal{L}(\theta^*|f_M, f_P)$ of Table S2. To summarize,

$$\log_{10}\left(\frac{P(M_0|f_M, f_P)}{P(M_\perp|f_M, f_P)}\right) \simeq \frac{1}{\log 10}\left(\mathcal{L}(\theta^*|f_M, f_P, M_0) - \mathcal{L}(\theta^*|f_M, f_P, M_\perp)\right)$$

3.6 Experimental estimation of protein decay rates by Specific Reaction Monitoring (SRM)

For each protein, we monitored the decay of 3 peptides, each represented by 3 SRM transitions over 9 time-points spanning a 96 hours time-course. After filtering out low quality transitions, we collected a total of 100 and 89 'responding' transitions, and 69 and 63 'non-responding' transitions. These corresponded to 18 and 25 'responding' proteins and 14 and 12 'non-responding' proteins from the miR-124 and miR-155 transfection experiments, respectively.

Sample preparation. HeLa cells were grown on light medium (normal DMEM with unlabeled amino acids (Dundee Cell Products, cat. no LM014, supplemented with 10% dialysed fetal calf serum, Dundee Cell Products, cat. no DS1003) in triplicates for two weeks. Cells were then collected 0, 1.5, 3, 6, 12, 24, 48, 72 and 96h after changing the medium to heavy SILAC medium containing 13C and 15N labeled arginine, and 13C and 15N labeled lysine (R10K8, Dundee Cell Products, cat. no LM015 with 10% dialysed calf serum). Protein isolation was achieved by resuspending the cell pellets in 100 μl lysis buffer containing 8 M Urea (Sigma-Aldrich, USA), 1 mM DTT (Sigma-Aldrich), 50 mM Tris HCl (Sigma-Aldrich), pH 7.5. The lysis buffer was supplemented with complete, EDTA-free protease inhibitor cocktail from Roche. After four freeze-thaw cycles, the sample was centrifuged at 16000xg for 15 min at 4°C and protein concentration was assessed with the BCA Protein Assay Kit (Thermo, Rockford, US). Proteins were digested and prepared for the SRM analysis as described previously (Selevsek et al., 2011).

Targeted proteomics approach. For each protein of interest, a set of proteotypic peptides (i.e uniquely associated with the protein of interest) together with their SRM coordinates were extracted from the spectral library of the Human SRM Atlas (http://www.srmatlas.org/). The SRM coordinates include the fragment ion masses, their relative intensities, the charge state distribution of the precursor and the elution time. The samples were analyzed in SRM mode as previously described in Selevsek et al. (2011) and SRM data were processed using the SRM skyline software (MacLean et al., 2010). Peptides with the following criteria where used for the quantification: i) good correlation between ion ratios obtained for the heavy and the light form, ii) good correlation between the ion ratios obtained for both forms and the ion ratios obtained in the MS/MS spectra present in the SRM spectral library, iii) transitions intensities of the heavy and the light form larger than 10. The three transitions for each heavy-light pair were used to quantify the peptide unless signals of co-eluting interferences were detected. In addition, peptides where the heavy transitions were not present in all the measurements were also discarded from the study.

An exponential decay model of protein decay. Following the method of Schwanhäusser et al. (2011), we assumed that the amount of light-labeled protein in the sample $x(t)$ decays exponentially

$$x(t) = x_0 e^{-\lambda t}$$

with x_0 being the amount of protein in the sample at time $t = 0$ and λ the protein decay rate. Because protein levels are at steady-state, the total amount of proteins, light – $x(t)$ – or heavy – $y(t)$ – grows exponentially with cell growth

$$x(t) + y(t) = x_0 2^{t/T}$$

with T the period of cell division. The fraction of light-labeled protein is therefore

$$\begin{aligned}\frac{x(t)}{x(t)+y(t)} &= \frac{x_0 e^{-\lambda t}}{x_0 2^{t/T}} \\ &= e^{-\left(\lambda + \frac{\log 2}{T}\right)t} \\ &= e^{-(\lambda+\gamma)t}\end{aligned}$$

with $\gamma = \frac{\log 2}{T}$ the cell growth rate.

Let l_{ij} and h_{ij} be the light and heavy transition intensity measurements the jth biological replicate of the samples collected at time t_i. We obtained these intensities by subtracting the peak area from the background signal estimated from the spectra baseline. Let x_{ij} and y_{ij} be the corresponding amount of light and heavy

protein in the sample, which assumed to be proportional to the measured transition intensities. Because the intensity measured by the mass spectrometer depends on the peptide and the fragment of the peptide observed, we introduce the transition-dependent proportionality constant c. This factor is the same for light and heavy transitions as the intensity measured by the spectrometer is not influenced by the isotope used. The measured transition intensity further depends on a sample-dependent proportionality factor that can be further decomposed in a factor k_{ij} which accounts for the error in attempting to spike the same number of cells at different samples. This leads to

$$\begin{aligned} x_{ij} &= ck_{ij}l_{ij} \\ y_{ij} &= ck_{ij}h_{ij} \end{aligned}$$

By computing the fraction of the light transition in the total intensity, one can eliminate the proportionality factors, obtaining

$$\frac{x_{ij}}{x_{ij}+y_{ij}} = \frac{l_{ij}}{l_{ij}+h_{ij}}$$

and therefore,

$$f_{ij} = \frac{l_{ij}}{l_{ij}+h_{ij}} = e^{-(\lambda+\gamma)t_i}, \tag{39}$$

where the relative intensity of the light transition f_{ij} decays exponentially at the rate that depends on the protein decay λ and the growth γ. Here, we set γ to 0.031 which corresponds to a cell cycle period of 22.5h, typically observed in HeLa. That the growth rate γ is not known with precision is not an issue in our case because we are not interested in the absolute decay rate of the proteins. Instead, we only want to test for differences in the decay rates of different proteins. To do so, it is sufficient to determine the apparent decay rate $\lambda+\gamma$, as the cell growth rate γ is the same for all proteins.

While the majority of the transitions give a good fit to the exponential decay model, some transitions featured one or two data points that were far off the general trend. We reasoned that more precise decay rates could be estimated if these outliers were identified and their weight moderated during fitting. In addition, a minority of transitions did not fit the exponential decay pattern at all. These transitions would obscure the comparison of the decay rates of different protein sets (e.g. responders vs non-responders, see main text) and we thus developed a probabilistic method to identify them.

Fitting protein decay rates in the presence of outliers. To take into account the possibility that a minority of data points are uninformative when it comes to estimating decay rates, we assumed that every measurement f_{ij} can be explained by one of two competing models: M_{exp} and M_{noise}. Under the M_{exp} model, the data point is explained by the exponential decay of rate λ plus a Gaussian measurement error with variance σ^2,

$$P(f_{ij}|\lambda,\sigma,M_{exp}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{f_{ij}-e^{-(\lambda+\gamma)t}}{\sigma}\right)^2} \tag{40}$$

Under the M_{noise} model, the data point is assumed to be drawn from a uniform distribution across the range R of all observed f_{ij} of the data set

$$P(f_{ij}|M_{noise}) = \frac{1}{R}$$

Given the parameters λ and σ, the probability X_{ij} that f_{ij} is an outlier can be computed using the Bayes theorem

$$\begin{aligned} X_{ij} &= P(M_{noise}|f_{ij},\lambda,\sigma) \\ &= \frac{P(f_{ij}|M_{noise},\lambda,\sigma)P(M_{noise}|\lambda,\sigma)}{P(f_{ij}|\lambda,\sigma)} \\ &= \frac{\frac{1}{R}\rho}{P(f_{ij}|M_{noise},\lambda,\sigma)P(M_{noise}|\lambda,\sigma)+P(f_{ij}|M_{exp},\lambda,\sigma)P(M_{exp}|\lambda,\sigma)} \\ &= \frac{\frac{1}{R}\rho}{\frac{1}{R}\rho+P(f_{ij}|\lambda,\sigma,M_{exp})(1-P(M_{noise}|\lambda,\sigma))} \\ &= \frac{\frac{1}{R}\rho}{\frac{1}{R}\rho+(1-\rho)P(f_{ij}|\lambda,\sigma,M_{exp})}, \end{aligned}$$

where $\rho = P(M_{noise})$ is the *a priori* fraction of outlying data points in the time-course and $P(f_{ij}|\lambda,\sigma,M_{exp})$ is defined by equation 40. The parameters λ,σ,ρ are unknown and we fit them by Expectation Maximization (Dempster et al., 1977), where we just derived the E-step in the last equation.

The M-step is obtained by maximizing the likelihood of the data $P(f_{ij}|\lambda,\sigma,\rho,X_{ij})$ assuming X_{ij} is known. The likelihood of the parameters given the time course data f can be written as

$$\begin{aligned} L(\lambda,\sigma,\rho|f,X) &= \prod_{i,j} P(f_{ij}|\lambda,\sigma,\rho,X) \\ &= \prod_{i,j} \left[X_{ij}P(f_{ij}|\lambda,\sigma,\rho,M_{noise}) + (1-X_{ij})P(f_{ij}|\lambda,\sigma,\rho,M_{exp})\right] \\ &= \prod_{i,j} \left[X_{ij}\frac{\rho}{R} + (1-X_{ij})\frac{1-\rho}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{f_{ij}-e^{-(\lambda+\gamma)t}}{\sigma}\right)^2}\right] \end{aligned}$$

Consequently, the log-likelihood is

$$\begin{aligned} \mathcal{L}(\lambda,\sigma,\rho|f,X) &= \log L(\lambda,\sigma,\rho|f,X) \\ &= \sum_{i,j} X_{ij}\left[\log\rho - \log R\right] + (1-X_{ij})\left[\log(1-\rho) - \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\left(\frac{f_{ij}-e^{-(\lambda+\gamma)t}}{\sigma}\right)^2\right] \end{aligned}$$

We can now compute the M-step for ρ by maximizing the log-likelihood with respect to ρ, which occurs when $\frac{\partial\mathcal{L}}{\partial\rho} = 0$. Taking the derivative of the log-likelihood with respect to ρ yields

$$\frac{\partial\mathcal{L}}{\partial\rho} = \sum_{ij} \frac{X_{ij}}{\rho} - \frac{1-X_{ij}}{1-\rho}.$$

Therefore,

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\rho} &= 0 \\ \Leftrightarrow (1-\rho)\sum_{ij} X_{ij} &= \rho\sum_{ij}(1-X_{ij}) \\ \Leftrightarrow \frac{1}{\rho} - 1 &= \frac{\sum_{ij} 1-X_{ij}}{\sum_{ij} X_{ij}} \\ \Leftrightarrow \rho &= \frac{1}{n}\sum_{ij} X_{ij}, \end{aligned}$$

with n the number of points in the time-course f. The M-step for λ and σ consists in maximizing $\mathcal{L}(\lambda,\sigma,\rho|f,X)$:

$$\begin{aligned} &\text{argmax}_{\lambda,\sigma}\mathcal{L}(\lambda,\sigma,\rho|f,X) \\ =\ &\text{argmax}_{\lambda,\sigma}\sum_{i,j} X_{ij}\left[\log\rho - \log R\right] + (1-X_{ij})\left[\log(1-\rho) - \frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\left(\frac{f_{ij}-e^{-(\lambda+\gamma)t}}{\sigma}\right)^2\right] \\ =\ &\text{argmax}_{\lambda,\sigma}\sum_{i,j}(1-X_{ij})\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\left(\frac{f_{ij}-e^{-(\lambda+\gamma)t}}{\sigma}\right)^2\right]. \end{aligned}$$

This cannot be done analytically because of the combination of the weighted sum and the exponential. We therefore maximize the log-likelihood numerically with respect to λ and σ under the constraint that the resulting decay rate is positive and larger than 15 min ($\lambda < 2.77$, which helps keeping estimates numerically stable when fitting very noisy transitions) with the method of Nelder and Mead (1965). From initial values of $\rho = 0.1$, $\lambda = 0.06$ (10h half-life) and $\sigma = \frac{1}{2}\sqrt{\frac{1}{n}\sum_{ij}\left(f_{ij}-\bar{f}\right)^2}$ (i.e. half the sample standard deviation), we iterate between the E-step and the M-step until convergence. Convergence is considered to be reached when the parameters ρ,λ,σ change by less than 10^{-5} between two successive M-steps, that is

$$|\rho_i - \rho_{i-1}| + |\lambda_i - \lambda_{i-1}| + |\sigma_i - \sigma_{i-1}| < 10^{-5}$$

Fig. S16 shows the result of this procedure on a representative transition.

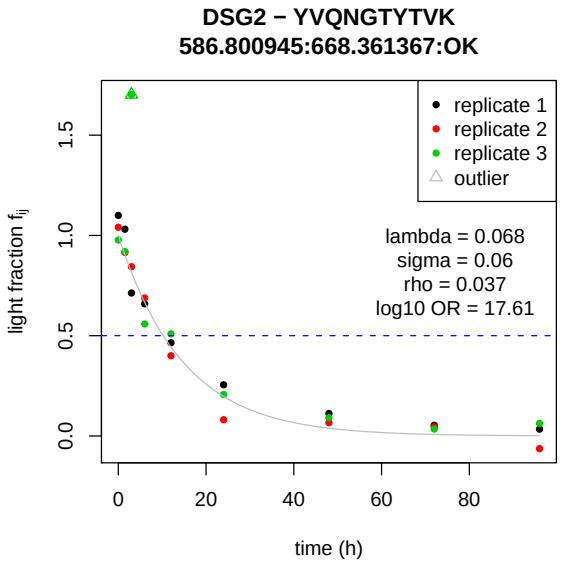

Figure S16: Fitting the decay rate of the 586.8:668.4 transition of peptide YVQNGTYTVK from protein DSG2. Each point represents the fraction of light peptide (y-axis) at a given time-point (x-axis) and in one of the three replicates (labeled in black, red and green). The inferred decay rate here is $\lambda = 0.068$ which corresponds to a half-life of 10.2h, with the best-fit exponential decay shown as a grey line. The inferred noise level is $\sigma = 0.06$, with an estimated $\rho = 3.7\%$ of the data points being outliers (marked as triangles). The $\log_{10}$ odds ratio ($\log_{10}$ OR) compares the likelihood of the data under the exponential decay model *vs* the 'no-trend' model (dashed blue line). The 'no-trend' model has the same number of parameters as the exponential decay model but assumes that the fraction of light peptide is constant over time. With a $\log_{10}$ odds ratio of 17.61 in the present case, the exponential decay model is 10^{18} times likelier than the 'no-trend' model, which indicates that this transition is compatible with exponential decay. This transition can thus be used to infer the decay rate of the YVQNGTYTVK peptide of protein DSG2.

The 'no-trend model': detecting transitions that cannot be explained by exponential decay. As mentioned above, a minority of transitions cannot be described by an exponential decay pattern and including them in the estimation of decay rates would only blur the comparisons of protein decay rates. Therefore, we developed a method to determine whether a transition is likely to reflect an exponential decay pattern as opposed to no clear decay pattern. We assume that in the 'no-trend' case, the measurements were scattered around a constant average value μ plus Gaussian noise of variance σ^2:

$$P(f_{ij}|\lambda,\sigma,M_{no-trend}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{f_{ij}-\mu}{\sigma}\right)^2} \tag{41}$$

Comparing the agreement of the data with the exponential decay model and 'no-trend' model requires the two models to be as similar as possible. Therefore, we again consider the possibility that a fraction of outlying data points do not fit the model. As in the M_{exp} model introduced in the previous paragraph, these outlying data points are distributed uniformly across the R range of all observed f_{ij} of the data set

$$P(f_{ij}|M_{noise}) = \frac{1}{R}$$

As previously, the probability X_{ij} that f_{ij} is an outlier given the parameters μ and σ can be computed using

Bayes' theorem

$$\begin{aligned}
X_{ij} &= P(M_{noise}|f_{ij},\mu,\sigma) \\
&= \frac{P(f_{ij}|M_{noise},\mu,\sigma)P(M_{noise}|\mu,\sigma)}{P(f_{ij}|\mu,\sigma)} \\
&= \frac{\frac{1}{R}\rho}{P(f_{ij}|M_{noise},\mu,\sigma)P(M_{noise}|\mu,\sigma)+P(f_{ij}|M_{no-trend},\mu,\sigma)P(M_{no-trend}|\mu,\sigma)} \\
&= \frac{\frac{1}{R}\rho}{\frac{1}{R}\rho+P(f_{ij}|\mu,\sigma,M_{no-trend})(1-P(M_{noise}|\mu,\sigma))} \\
&= \frac{\frac{1}{R}\rho}{\frac{1}{R}\rho+(1-\rho)P(f_{ij}|\mu,\sigma,M_{no-trend})},
\end{aligned}$$

where $\rho = P(M_{noise})$ is the *a priori* fraction of outlier data points in the time-course and $P(f_{ij}|\mu,\sigma,M_{no-trend})$ is defined by equation 41. The parameters μ,σ,ρ are unknown and we fit them by Expectation Maximization (Dempster et al., 1977), where we just derived the E-step in the last equation.

The M-step is obtained by maximizing the likelihood assuming X_{ij} is known. The likelihood of the parameters given the time course measurements f can be written as

$$\begin{aligned}
L(\mu,\sigma,\rho|f,X) &= \prod_{i,j} P(f_{ij}|\mu,\sigma,\rho,X) \\
&= \prod_{i,j} \left[X_{ij}P(f_{ij}|\mu,\sigma,\rho,M_{noise})+(1-X_{ij})P(f_{ij}|\mu,\sigma,\rho,M_{no-trend})\right] \\
&= \prod_{i,j} \left[X_{ij}\frac{\rho}{R}+(1-X_{ij})\frac{1-\rho}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{f_{ij}-\mu}{\sigma}\right)^2}\right]
\end{aligned}$$

Consequently, the log-likelihood is

$$\begin{aligned}
\mathcal{L}(\mu,\sigma,\rho|f,X) &= \log L(\mu,\sigma,\rho|f,X) \\
&= \sum_{i,j} X_{ij}\left[\log\rho-\log R\right]+(1-X_{ij})\left[\log(1-\rho)-\frac{1}{2}\log(2\pi\sigma^2)-\frac{1}{2}\left(\frac{f_{ij}-\mu}{\sigma}\right)^2\right]
\end{aligned}$$

We can now compute the M-step for ρ by maximizing the log-likelihood, which is maximal with respect to ρ when $\frac{\partial\mathcal{L}}{\partial\rho}=0$. Taking the derivative of the log-likelihood with respect to ρ yields

$$\frac{\partial\mathcal{L}}{\partial\rho} = \sum_{ij}\frac{X_{ij}}{\rho}-\frac{1-X_{ij}}{1-\rho}$$

Therefore,

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\rho} &= 0 \\
\Leftrightarrow (1-\rho)\sum_{ij}X_{ij} &= \rho\sum_{ij}(1-X_{ij}) \\
\Leftrightarrow \frac{1}{\rho}-1 &= \frac{\sum_{ij}1-X_{ij}}{\sum_{ij}X_{ij}} \\
\Leftrightarrow \rho &= \frac{1}{n}\sum_{ij}X_{ij},
\end{aligned}$$

with n the number of points in the time-course f. The M-step for μ and σ consists in maximizing $\mathcal{L}(\mu,\sigma,\rho|f,X)$:

$$\begin{aligned}
&\text{argmax}_{\mu,\sigma}\mathcal{L}(\mu,\sigma,\rho|f,X) \\
=\ &\text{argmax}_{\mu,\sigma}\sum_{i,j}X_{ij}\left[\log\rho-\log R\right]+(1-X_{ij})\left[\log(1-\rho)-\frac{1}{2}\log(2\pi\sigma^2)-\frac{1}{2}\left(\frac{f_{ij}-\mu}{\sigma}\right)^2\right] \\
=\ &\text{argmax}_{\mu,\sigma}\sum_{i,j}(1-X_{ij})\left[-\frac{1}{2}\log(2\pi\sigma^2)-\frac{1}{2}\left(\frac{f_{ij}-\mu}{\sigma}\right)^2\right]
\end{aligned}$$

As in the previous paragraph, the maximum-likelihood estimates of μ and σ are the X_{ij}-weighted average of f_{ij} and the X_{ij}-weighted average of the squared deviation of f_{ij} from the sample mean $\bar{f}$. However, to treat the 'no-trend' and exponential model similarly, we again maximize the log-likelihood numerically with respect to μ and σ with the method of Nelder and Mead (1965). From the initial values $\mu = \frac{1}{n}\sum_{ij} f_{ij}$ and $\sigma = \sqrt{\frac{1}{n}\sum_{ij}\left(f_{ij} - \bar{f}\right)^2}$, we iterate between the E-step and the M-step until convergence, defined as

$$|\rho_i - \rho_{i-1}| + |\mu_i - \mu_{i-1}| + |\sigma_i - \sigma_{i-1}| < 10^{-5}$$

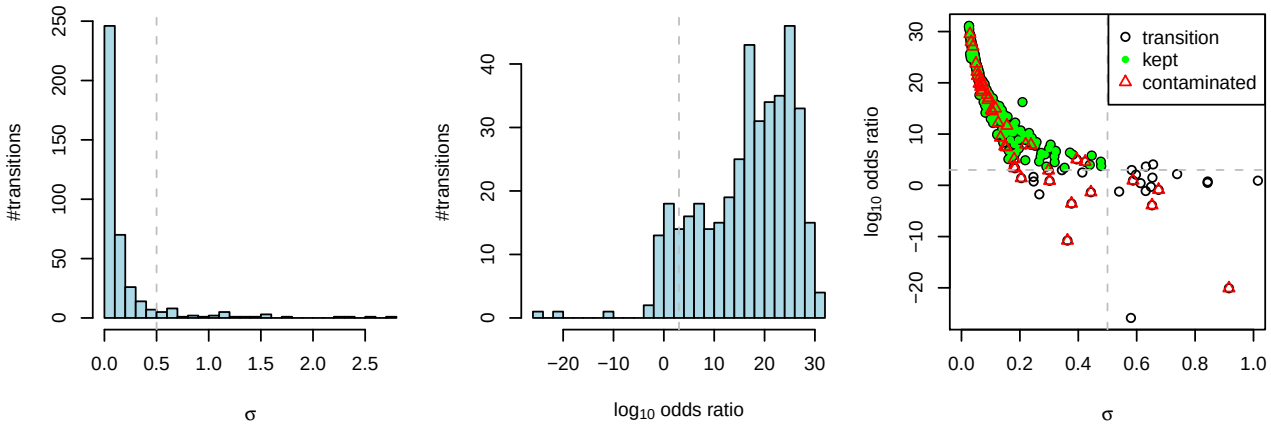

Figure S17: Filtering out uninformative transitions. Left: histogram of the inferred amount of noise σ under the exponential decay model. Center: Histogram of the $\log_{10}$ odds ratio of the exponential decay model and the "no-trend" model. Higher $\log_{10}$ odds ratios indicate that the exponential decay model explains the time-course best. Right: Scattering the noise σ against the $\log_{10}$ odds ratio shows that transitions with least noise are more compatible with the exponential decay model. Each dot represents a transition time-course, with green dots representing transitions that were used for statistical comparison of decay rates. Red triangles mark transitions that were contaminated with co-eluting peptides of different half-lives according to manual spectra examination. In all panels, dashed grey lines show the cut-offs set to discard uninformative transitions ($\sigma < 0.5$, $\log_{10}$ odds > 3).

Summary Under the exponential decay model, 316 of the 398 transitions we measured show low noise levels $\sigma < 0.2$ (Fig. S17, left panel). Relative to the maximum signal 1, such a noise level corresponds to a signal-to-noise ratio of at least 5 for the vast majority of transitions. Also, 380 transitions had positive $\log_{10}$ odds ratio (Fig. S17, central panel), which suggests that most transitions were better explained by the exponential decay model than by the 'no-trend' model. By visual inspection of the exponential decay fits, we determined that the cut-offs $\sigma < 0.5$ and $\log_{10}$ odds ratio > 3 define transitions that agree well with an exponential decay pattern and with noise levels low enough to lead to accurate estimates of the peptide decay rate λ. We further discarded transitions that appeared to be contaminated by co-eluting peptides. 77 transitions were hence filtered out, leaving a total of 321 transitions for further statistical analysis (Fig. S17, right panel).

References

Baek, D., Villén, J., Shin, C., Camargo, F., Gygi, S., and Bartel, D. P. (2008). The impact of microRNAs on protein output. Nature *455*, 64–71. doi:10.1038/nature07242.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., et al. (2011). NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic acids research *39*, D1005–10. doi:10.1093/nar/gkq1184.

Berninger, P., Gaidatzis, D., Nimwegen, E. V., and Zavolan, M. (2008). Computational analysis of small RNA cloning data. Methods *44*, 13–21.

Béthune, J., Artus-Revel, C. G., and Filipowicz, W. (2012). Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. EMBO reports pages 1–8. doi:10.1038/embor.2012.82.

Cambridge, S. B., Gnad, F., Nguyen, C., Bermejo, J. L., Krüger, M., and Mann, M. (2011). Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. Journal of proteome research *10*, 5275–84. doi:10.1021/pr101183k.

Dempster, A., Laird, N., Rubin, D., and Others (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) *39*, 1–38.

Elkon, R., and Agami, R. (2008). Removal of AU bias from microarray mRNA expression data enhances computational identification of active microRNAs. PLoS computational biology *4*, e1000189. doi: 10.1371/journal.pcbi.1000189.

Friedel, C. C., Dölken, L., Ruzsics, Z., Koszinowski, U. H., and Zimmer, R. (2009). Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic acids research *37*, e115. doi: 10.1093/nar/gkp542.

Gaidatzis, D., van Nimwegen, E., Hausser, J., and Zavolan, M. (2007). Inference of miRNA targets using evolutionary conservation and pathway analysis. BMC bioinformatics *8*, 69. doi:10.1186/1471-2105-8-69.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80. doi:10.1186/gb-2004-5-10-r80.

Grimson, A., Farh, K., Johnston, W., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell *27*, 91–105.

Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840. doi:10.1038/nature09267.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–23. doi: 10.1126/science.1168978.

Karginov, F. V., Conaco, C., Xuan, Z., Schmidt, B. H., Parker, J. S., Mandel, G., and Hannon, G. J. (2007). A biochemical approach to identifying microRNA targets. Proceedings of the National Academy of Sciences of the United States of America *104*, 19291–6. doi:10.1073/pnas.0709971104.

Khan, A. A., Betel, D., Miller, M. L., Sander, C., Leslie, C. S., and Marks, D. S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nature biotechnology *27*, 549–55. doi:10.1038/nbt.1543.

Khanin, R., and Higham, D. J. (2009). Mathematical and computational modelling of post-transcriptional gene regulation by microRNAs. In Handbook of Statistical Systems Biology, chapter 10.

Khorshid, M., Rodak, C., and Zavolan, M. (2011). CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic acids research *39*, D245–52. doi: 10.1093/nar/gkq940.

Kishore, S., Gruber, A. R., Jedlinski, D. J., Syed, A. P., Jorjani, H., and Zavolan, M. (2013). PAR-CLIP of snoRNA core proteins and small RNA-seq identify novel human snoRNA loci and give insights into snoRNA processing. Genome Biology *in press*.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular systems biology *4*, 222. doi:10.1038/msb.2008.61.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics (Oxford, England) *26*, 966–8. doi: 10.1093/bioinformatics/btq054.

Nelder, J., and Mead, R. (1965). A simplex method for function minimization. The Computer Journal *7*, 308–313.

Ohrt, T., Mütze, J., Staroske, W., Weinmann, L., Höck, J., Crell, K., Meister, G., and Schwille, P. (2008). Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. Nucleic acids research *36*, 6439–49. doi: 10.1093/nar/gkn693.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342. doi:10.1038/nature10098.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58–63.

Selevsek, N., Matondo, M., Sanchez Carbayo, M., Aebersold, R., and Domon, B. (2011). Systematic quantification of peptides/proteins in urine using selected reaction monitoring. Proteomics *11*, 1135–47. doi: 10.1002/pmic.201000599.

Tenenbaum, M., and Pollard, H. (1963). Ordinary differential equations (New York: Harper & Row, Publishers, Inc.).

Wang, X., and Wang, X. (2006). Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic acids research *34*, 1646–52. doi:10.1093/nar/gkl068.

Wu, Z., Irizarry, R., Gentleman, R., Martinez-Murillo, F., and Spencer, F. (2004). A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association *99*, 909–917.