Regulating the many to benefit the few: role of weak small RNA targets

Daniel Jost[,†], Andrzej Nowojewski[,†] and Erel Levine[,†*]
[†] Department of Physics and FAS Center for Systems Biology,
Harvard University, Cambridge, MA
[*] Corresponding author: elevine@fas.harvard.edu

Abstract

Small regulatory RNAs are central players in the regulation of many cellular processes across all kingdoms of life. Experiments in mouse and human have shown that a typical small RNA may regulate the expression of many different genes, suggesting that small RNAs act as global regulators. It is noted though that most targets respond only weakly to the presence of the small RNA. At the same time, evidence in bacteria and animals suggest that the phenotypes associated with small RNA mutants are only due to a few of their targets. Here we assume that targets regulated by a small RNA to control function is in fact small, and propose that the role of the many other weak targets is to confer robustness to the regulation of these few "principal" targets. Through mathematical modeling we show that auxiliary targets may significantly buffer both number and kinetic fluctuations of the principal targets, with only minor slowdown in the kinetics of response. Analysis of genomic data suggests that auxiliary targets experience a non-specific evolutionary pressure, playing a role at the system-level. Our work is of importance for studies on small RNA functions, and impacts on the understanding of small RNA evolution.

Key words: post-transcriptional regulation, small regulatory RNAs, auxiliary targets, fluctuations, temporal response, small RNA evolution

Introduction

Small RNA molecules play critical regulatory roles in many different organisms, from small regulatory RNAs regulating stress response in bacteria to microRNAs regulating development and homeostasis in mammals (1, 2). Many small RNA families achieve target-specificity at the post-transcriptional level via base-pairing of a very short (6-8 nucleotides) region with the targeted mRNA. Consequently, one predicts that a particular binding sequence could appear at random in multiple loci across the genome, which may explain why many genes carry binding sites in their 3'UTR. Sequence pairing, however, may not be sufficient for successful interaction, and one might expect that most of these putative sites are non-functional, and would be lost during evolution. Surprisingly, bioinformatic searches for microRNA targets that focus on highly conserved sites still predict a large number of target genes per microRNA (3). In addition, recent experiments in mouse (4) and human (5) demonstrate that the transfection of a single microRNA affects the expression of many genes that carry a seed match, although most of these genes respond to the microRNA very weakly. It is therefore reasonable to hypothesize that these many binding sites mediate real interactions and have some functionality within the cellular context. One tempting interpretation is that microRNAs act as global regulators, affecting directly genes across pathways or modules of the cellular network.

On the other hand, detailed studies focused on particular small regulatory RNAs (srRNAs) draw a very different picture. Studies in bacteria and animals (6–16) suggest that the phenotype associated with mutating srRNAs is due only to few of their targets. For example, the lethal phenotype associated with mutating the recognition sequence in the heterochronic microRNA let-7 of *Caenorhabditis elegans* (6, 7), is rescued by a compensating mutation in only one gene (*lin-41*). Similarly, concurrent re-expression of only 3 targets of the human microRNA miR-31 (10) (out of the hundreds predicted) is enough to abrogate the metastasis suppression phenotype associated with overexpression of the microRNA. These and other examples suggest that only few targets are directly involved in the cellular response to the level of the srRNA.

As the functions of more and more small regulatory RNAs are being elucidated in a variety of organisms and tissues, the mystery remains: what is the role of the many weakly interacting conserved targets (17). Recently, it has been proposed that the number of phenotypically relevant targets regulated by a srRNA is in fact small (18, 19) and that the many other weak targets which we accordingly term "auxiliary" might be competitive inhibitors for the srRNA by preventing the sr-

RNA binding to the few "principal" targets. In this paper we propose that the role of these "auxiliary" targets is to confer robustness and suppress both number and kinetic fluctuations. Through dynamical analysis we show that auxiliary targets can strongly suppress intrinsic fluctuations in their target levels with only minor effect on the kinetics of response. We support these predicted functional role of auxiliary targets by analyzing their conservation across vertebrates, and show that auxiliary targets experience an non-specific evolutionary pressure, suggesting that their role is at the system level. With small RNAs acting in development and response pathways, these features are expected to be of major physiological consequence.

Materials and Methods

Unified model for srRNA regulation

General description of the model

In absence of auxiliary targets, the general picture of srRNA post-transcriptional regulation is well-described by modeling the dynamics of the number of free srRNA s, of the number of principal target mRNAs m and of the number of principal target proteins p, by a set of mass-action equations (20):

$$\frac{ds}{dt} = \alpha_s - \beta_s s - ksm, \quad (1)$$

$$\frac{dm}{dt} = \alpha_m - \beta_m m - ksm, \quad (2)$$

$$\frac{dp}{dt} = \gamma m - \beta_p p. \quad (3)$$

where α_s (α_m) refers to the transcriptional rate of srRNA (mRNA) and β_s (β_m) to its degradation or turnover rate. k represents the interaction rate between the srRNA and its target mRNA. The target proteins are produced at a rate γ per mRNA molecule and self-degrade at a rate β_p. Eqs. 1 and 2 assume that the pairing between the srRNA and the mRNA either leads to degradation of the complex or sequesters it (e.g. while blocking translation) for considerable time. Active co-degradation is believed to occur for many prokaryotic small RNA-mRNA couples (1), while titration may be a more dominant mode of action of eukaryotic miRNAs (2). Within the same framework, it is easy to generalize the model and allow

a fraction of the srRNA to be recycled (see Text S1 Sec.B in the Supporting Material). This does not affect the main results of our work (see Text S1 Sec.C 3c in the Supporting Material).

We generalize this model by accounting for the interaction between the srRNA and the auxiliary targets, leading to the formation of a transient complex (Fig.1). The kinetics of the number of free auxiliary target mRNAs n and of complexed srRNAs c follow the mass-action equations

$$\frac{dn}{dt} = \alpha_n - \beta_n n - k_a s n + k_d c, \quad (4)$$

$$\frac{dc}{dt} = -\beta_c c + k_a s n - k_d c, \quad (5)$$

where α_n is the total production rate over all auxiliary targets and β_n is their average turnover rate. k_a is the association rate between the srRNA and the auxiliary targets and k_d the dissociation rate of the transient complex which is degraded at a rate $\beta_c \ll k_d$. To account for the effect of the auxiliary targets on the srRNA level, one has to augment Eq.1 by $[-k_a s n + k_d c + \beta_c (1 - p_d) c]$, with p_d the probability that the srRNA is also eliminated during the complex degradation.

Accounting for stochasticity

Stochasticity of the underlying biochemical reactions, including the transcriptional burstiness (21, 22) and the effect of transport by diffusion of the interacting molecules (23, 24), are accounted for by augmenting the previous set of mass-action equations with Langevin terms (25, 26) which captured the intrinsic fluctuations of each reactions (see Text S1 Sec. A-D in the Supporting Material for a detailed description of the full stochastic model).

Analysis of steady-state properties

We analyzed the steady-state properties of srRNA regulation in absence or in presence of auxiliary targets within the limit of small noise, using the linear noise approximation (25). The mean steady-state levels were estimated by solving the mass-action equations (1-5). Fluctuations were analyzd by solving the fluctuation-dissipation relation

$$JC + CJ^{\dagger} + N = 0, \quad (6)$$

with C the covariance matrix of the system, J the Jacobian of the set of mass-action equations (1-5) and N is a diagonal matrix whose entries are the amplitudes of the noise (Langevin) terms (see Text S1 Sec. B).

Efficacy of the srRNA regulation and limit between slow and fast transport modes

As a robust measure of the efficacy of the srRNA regulation, we choose an information-theory based measure, which considers the interactions between a regulator and its target as a communication channel, and measure the information capacity of this channel (27, 28). The major advantage of this choice is that it does not require any specific knowledge of the function of the regulator or the precise signal it transduces. Moreover, this measure is insensitive to the particular form of the response function, and can therefore be applied across the entire range of parameters in our model.

In the limit of low noise, the information capacity for srRNA pathways can be approximated by (see Text S1 Sec. C3)

$$I_{max} = \log_2 \left(\frac{1}{\sqrt{2\pi e}} \int d\alpha_s \left| \frac{d\langle p \rangle(\alpha_s)}{d\alpha_s} \right| \frac{1}{\sigma_p(\alpha_s)} \right) \tag{7}$$

where $\langle p \rangle$ and σ_p are, respectively, the steady-state mean and standard deviation of the principal target level. At a fixed set of parameters the impact of auxiliary targets on this information-based efficacy strongly depends on the mobility of the molecules: auxiliary targets increase the capacity of the channel if they diffuse rapidly, but decrease its capacity if diffusion is slow. One can therefore define a critical diffusion rate which marks the boundary between these two behaviors (see Text S1 Sec. C3).

Analysis of kinetic properties

The kinetic response of the pathway was studied using stochastic Gillespie simulations (29), where, for simplicity, we neglect fluctuations in the total number of auxiliary targets. Each simulation starts with a configuration sampled from the steady state distribution of principal targets (Poissonian distribution with mean α_m/β_m). At $t = 0$ the transcription of srRNA is switched on, and the kinetic response is assessed by measuring the distribution of the first passage time when the number of principal mRNAs reaches zero.

Choice of parameter values

Typical values for parameters are only known for bacterial srRNA pathways (20, 30). We choose to use these ranges of values (given in Text S1 Sec.F) to plot

our figures. However, Fig. S2 in the Supporting Material shows that the effects described in the paper are robust over a wide range of parameters values which likely also includes the eukaryotic pathways.

Analysis of genomic data

The conservation of microRNA targets was studied for a set of 87 microRNAs conserved among vertebrates (Dataset S1 in the Supporting Material) and compared with a artificially generated set of 1,000 "mock"-microRNAs (Dataset S2). Seeds of the mock microRNAs were generated as 7-mers that obey the same dinucleotide statistics as the 87 conserved microRNAs in our sample (31). See Text S1 for details.

For each microRNA, a list of predicted targets in 12 vertebrates (*Homo sapiens, Pan troglodytes, Macaca mulatta, Mus musculus, Rattus norvegicus, Canis familiaris, Equus caballus, Bos taurus, Monodelphis domestica, Ornithorhynchus anatinus, Gallus gallus* and *Xenopus tropicalis*) was obtained using the TargetScan algorithm and the aligned 3'UTRs dataset available at the TargetScan website (32). The number of targets $N_{i,s}$ per microRNA i in each species s was normalized by the total number of genes per species $N_{tot,s}$ in the alignment. Conservation of the number of targets was quantified by computing for each microRNA the relative fluctuations of $\{N_{i,s}/N_{tot,s}\}$ across species (see Sec.E in Text S1).

Taking the list of human targets as reference, we represent TargetScan predictions by a binary array $x^t_{i,s}$, where $x^t_{i,s} = 1$ if gene t is a target of microRNA i in human and in species s, and 0 otherwise. The observed frequency f^t_i is then defined as the mean value of $x^t_{i,s}$ across species. The corresponding conservation score C^t_i is obtained by normalizing f^t_i by the number of species that carry a homolog of gene t(see Sec. E in Text S1). For example, $C^t_i = 1$ means that gene t is a target of microRNA i in every species where t is present.

To investigate the conservation of principal and auxiliary targets, we use published experimental data that measured global proteome response to transfection of a microRNA (miR-1, miR-124 and miR-181) in (human) HeLa cells (4). For each transfected microRNA, we extract from the experimental dataset the protein level change of each of its target genes (as defined above), if it had been been measured. Dataset S3 contains the list of these genes for miR-1 (382 genes), miR-124 (249 genes) and miR-181 (345 genes), as well as the corresponding fold-change in protein level (in $\log_2$ unit).

The distributions of fold-repression does not offer a natural separation between

principal and auxiliary targets in HeLa cells. We therefore split the targets based on their fold-change into an subset of "strong targets", representing targets with a fold-repression below a given arbitrary limit ($\log_2$ of the fold-change ≤ -0.6) and the complementary subset of "weak targets". We verified that our conclusions do not depend on the choice of this limit (Sec. E of Text S1).

Results

Simple model for srRNAs with two classes of targets

While it is still unclear which factors determine the efficiency of a given srRNA on a particular target, experiments on mammalian somatic tissues have revealed that most of the putative targets are only weakly affected by the transfection of a single microRNA, and only a small proportion is more significantly repressed (4, 5). Below we identify the phenotypically relevant srRNA targets with this limited subset of strongly affected targets. This assumption may not be always justified, as in some cases small modulation may have strong physiological impact (see Table 2 in (18)), and conversely one could imagine an auxiliary role even for a strongly suppressed target (17, 33) (see Text S1 Sec.C 4 in the Supporting Material).

Despite the fact that small RNA pathways differ in many details (including their biogenesis and mechanism of action), at an abstract level they can all be described by a unified model (20, 26, 30, 34–38), which accounts for synthesis of all RNA species, interaction of the srRNA with its targets, and the consequential suppression of translation and/or promotion of degradation of the mRNA (and, perhaps, the small RNA molecule itself). This model can be translated into a simple mathematical framework (20) that we augment to describe the interactions with two classes of targets: the principal targets and the auxiliary targets (Fig.1). Within a Langevin formalism, our model accounts for the stochastic nature of the underlying biochemical reactions, including the effect of transport by diffusion of the interacting molecules, by taking the limit of weak noise (for details see Materials and Methods as well as Text S1 Sec. A in the Supporting Material).

Auxiliary targets can finely tune the expression of principal targets

In absence of auxiliary targets, the effect of srRNAs on the mean level of principal targets falls into one of three categories (20), depending on the relative strength of the small RNA expression compared with that of the target (Fig.2A, black line). For efficient srRNA production (i.e. in cases where the synthesis rate of mature srRNAs α_s is much larger than that of mRNA, α_m), most of the mRNAs are targeted by the large srRNA pool, leading to complete silencing. Conversely, under conditions where the production of the srRNA is less efficient ($\alpha_s \ll \alpha_m$) the target is almost unaffected by the presence of the small RNA and is normally expressed. Intermediate scenarios, where the production rates of the srRNA and of the mRNA are comparable ($\alpha_s \sim \alpha_m$) allow fine-tuning of the target expression. Under these conditions, quantitative changes in the transcription rate of the srRNA correspond to quantitative changes in the target level. The sharpness of the transition between silenced and expressed regimes is controlled by the strength of the interaction between srRNA and mRNA. For strong interactions, the regulatory logic is a sharp linear-threshold response (Fig. S1 A in the Supporting Material).

Additional targets may affect the level of srRNA available for regulating the principal targets in many ways, and one possibility is that auxiliary targets could significantly change this simple picture. At steady-state, auxiliary targets can be interpreted as stoichiometric weak targets with an effective interaction constant $k_{\text{eff}} = k_a p_d(\beta_c/k_d)/(1+\beta_c/k_d)$ (see Text S1 Sec. C1). The fact that the auxiliary targets are, by definition, weakly affected by the srRNA suggests that $k_{\text{eff}} \ll k$. The level of auxiliary targets would therefore play a measurable role on the principal one only at $k_{\text{eff}}\alpha_n \gtrsim k\alpha_m$ (20). In this situation, the auxiliary targets modify the steady-state level of the srRNA and then indirectly that of principal targets. This could help in finely tuning the position of the transition between the expressed and silent regimes (Fig. 2A) but does not change the regulatory logic of the post-transcriptional regulation. Typical parameters (20, 30) suggests that this effect starts to be significant for thousands of auxiliary targets molecules.

This characteristic regulatory logic has been verified experimentally *in vivo* for bacterial small RNA pathways in *Escherichia coli* (20) and for the mammalian microRNA pathway in HeLa cells (38). These results allow to characterize the effect of a small RNA on a target in terms of three qualitatively different operating regimes: the silencing regime, the tuning (or crossover) regime, and the expression regime. In this language, one may deduce that the effect of many mammalian microRNAs at their normal expression level lies within the tuning category. In-

deed, on one hand, overexpression experiments show at most a 2-fold-repression compared to wild-type (4, 5), suggesting that in normal tissues the targets are not completely silenced. On the other hand, knockdown of the same microRNA leads to at most a 2-fold up-regulation (5), suggesting that targets are already somewhat suppressed in normal tissues. It is therefore possible that many microRNA-target pairs act in the crossover regime, which - as we discuss next - exposes them to amplified fluctuations.

Auxiliary targets reduce the intrinsic noise of the principal targets

The stochastic nature of the biochemical reactions composing gene regulation pathways leads to intrinsic fluctuations around the mean signal levels (39–41). A canonical way to appreciate the strength of protein fluctuations is to consider the Fano factor $\nu = \sigma_p^2/\langle p\rangle$, with σ_p^2 the variance and $\langle p\rangle$ the mean level of the principal protein number p at steady-state. ν is a measure of the noise-to-signal ratio, which confers two advantages: in simple cases it does not depend on the transcription rate, and it allows comparison of the fluctuations in a particular system to a simple memoryless Poisson process, where $\nu = 1$. For example, in absence of srRNA regulation, the noise-to-signal ratio is mainly driven by the burstiness of the protein translation and $\nu \approx 1 + b$ (39) with $b = \gamma/\beta_m$ the protein burst size, that is the average number of protein translated from a single mRNA molecule.

Models of post-transcriptional regulation by srRNAs suggest that this mode of regulation is particularly effective in suppressing intrinsic fluctuations in the silenced regime (26, 35) (Fig.2C, black lines). Indeed, in the limit of efficient transport, the effective life-time of an active mRNA ($[\beta_m + k\langle s\rangle]^{-1}$) is dramatically reduced through the interaction with the srRNA and $\nu \approx 1 + b^*$ with $b^*/b = \langle p\rangle/\langle p\rangle_{\max} = \beta_m/[\beta_m + k\langle s\rangle] \ll 1$ and $\langle p\rangle_{\max} = (\gamma\alpha_m)/(\beta_m\beta_p)$ the expression level in absence of srRNA (see Text S1 Sec. B in the Supporting Material). If transport is inefficient, the slow stochastic diffusion of a very low number of mRNA molecules can lead to relatively high fluctuations in the local mRNA concentration and diffusion noise may dominate (Fig.2B, black lines).

In the crossover regime, where synthesis of srRNA and mRNA are similarly efficient, the ultra-sensitivity of the stoichiometric system leads to a high correlation between the abundance of the srRNA and that of its principal target (42). The cell state becomes broadly distributed and alternates between unrepressed and repressed states, yielding a large distribution for the protein level with a high noise-

to-signal ratio (26, 35, 42) (Fig.2C, black lines). These fluctuations are enhanced by strong RNA-RNA interaction (Fig. S1 B) and amplified by stochasticity in active gene copy number and transcriptional bursting (26, 35). The presence of such large fluctuations deems small RNAs unsuitable for fine-tuning gene expression, and has adverse impact on their suitability for patterning gene expression during development (43). As discussed above, it is likely that many mammalian microRNA-target pairs act within the tuning-regime and are therefore exposed to these amplified fluctuations. In what follows we suggest that the role of auxiliary targets is to suppress these fluctuations, making the tuning regime operational.

Our key observation is that the presence of auxiliary targets serves to diminish the correlation between the free srRNA copy number and that of its principal target, thus reducing fluctuations significantly (Fig.2C). By following the propagation of fluctuations within the model, we find that large intrinsic fluctuations in the srRNA level are quickly absorbed by the dynamic equilibrium between free and complexed srRNAs. To appreciate the effect of auxiliary targets on the fluctuations of principal targets, we first simplify the model by assuming that auxiliary targets have no effect on the mean steady-state level of principal targets ($\beta_c/k_d \to 0$). We will return to this point at the end of this section.

Consider first the case of efficient transport. The presence of auxiliary targets has a weak impact on the steady-state level of *free* srRNA while creating a pool of *complexed* srRNA. This pool plays the role of particle reservoir that absorbs srRNA fluctuations (due e.g. to transcriptional burstiness or ultra-sensitivity of the interaction with the principal targets). In particular, for a high number of auxiliary targets, the noise-to-signal ratio of the principal targets is given by $\nu \approx 1 + b^*$ (Text S1 Sec.C in the Supporting Material). To appreciate this result we call the reader's attention to the fact that in the absence of auxiliary targets this expression is only valid in the repressed regime. The presence of auxiliary targets reduces the fluctuations for *any* level of the srRNA, even in the noise-sensitive crossover regime, thus permitting fine-tuning and better control.

Interaction between srRNAs and their targets occur in a complex environment in the cytoplasm, perhaps in dedicated bodies (44). We reason that the presence of auxiliary targets may alter the delivery of a srRNA to its principal targets and consequently increases fluctuations. We therefore depart from the limit of efficient diffusion, and study the model over a range of diffusion constants governing the srRNA mobility. As suspected, when transport is slower, the presence of auxiliary targets yields a notable increase in fluctuations (Fig. 2B). Indeed, the stochastic diffusion of a small number of free srRNAs (compared with the high number of free auxiliary targets) impacts on the noisy formation of the srRNA-auxiliary tar-

get complexes, leading to an increase of the srRNA fluctuations which propagates into that of principal targets.

We conclude that auxiliary targets have two opposing effects on the fluctuations in the level of a principal targets: they buffer number fluctuations due e.g. to bursty transcription or chromatin fluctuations, while increasing noise due to slow transport. Using the notions of efficacy and information capacity of the regulatory pathway (27) (see Materials and Methods and Text S1 Sec.C), we show that the competition between these two effects is mainly settled by comparing the diffusibility of the RNA molecules and their half-lives (see Fig.3 and Fig. S2 in the Supporting Material). Based on these two parameters (which may be srRNA and condition specific) the principal targets of a srRNA may benefit or suffer from the presence of auxiliary targets. Interestingly, the boundary between these two situations does not depend significantly on the microscopic details of the various interactions in the model. For bursty promoters, the buffering of fluctuations is even stronger and auxiliary targets help maintaining a low noise level even for slow transport.

Which regime characterizes the behavior of naturally occurring small RNAs? It has long been assumed (and in some cases verified experimentally) that srRNAs are stable (45, 46), and that they are actively mobile (46). We therefore expected srRNA to be in the "fast diffusion slow decay" regime, that is favorable for auxiliary targets. Indeed, estimating typical parameters for small RNAs in bacteria, yeast and metazoan taken from the BioNumbers web site (47) (see Table 1) suggests that they all sit safely in the region where auxiliary targets serve to suppress fluctuations. This observation can explain the ability of natural microRNAs present in somatic tissues to regulate their principal targets with low noise, even in the cross-over regime.

To complete our discussion of noise attenuation we comment that the general picture presented above is not changed when allowing the sRNA-auxiliary target complexes to have a short but finite lifetime (see Text S1 Sec. C 3b in the Supporting Material). However, in this case the presence of auxiliary targets may affect the efficacy of the regulation. Indeed, under this conditions auxiliary targets effectively reduce the number of srRNA that are presented to their principal targets. In principle, one could have used this feature to impose a bound on the number of allowed auxiliary targets. However, as discussed in the next section, auxiliary targets act also kinetic traps for the srRNAs and slow-down the regulation process. As it turns out, this imposes even a stronger bound on the number of auxiliary targets.

Auxiliary targets maintain a robust kinetic response

In response to changes in environmental signals or development transitions, cells may need to adapt by stopping the expressions of specific proteins. Post-transcriptional regulation has been suggested as a mechanism to accelerate the response of gene expression to changes in the input signals (26, 30). Indeed, in the absence of auxiliary targets newly produced srRNAs are directly used for depleting the pool of principal mRNA targets. However, it is natural to assume that auxiliary targets could act as kinetic traps and decelerate the adaptation of principal targets to signals that affect the srRNA production. The number of auxiliary targets exceeds significantly that of the principal target (4, 5), and one may wonder if this would not result in an impossible hindrance of cellular response.

To address this question, we compare the kinetic response of two simple scenarios. In Scenario I, the srRNA bares the sole responsibility for suppression of the target. Transcription of the mRNA is unaffected while the production of the srRNA is turned on at a high rate ($\alpha_s \gg \alpha_m$), thus switching from the expressed to the repressed regime of Fig.2A. In Scenario II, which is srRNA-free, transcription of mRNA is stopped (e.g. by binding of transcription factors (TF)) and the remaining mRNAs and proteins are allowed to self-degrade. As we just mentioned, in the absence of auxiliary targets the response time is much faster in the srRNA-based Scenario I than in the TF-based Scenario II (Fig.4, green arrow), and the question is whether this effect is wiped out by the presence of auxiliary targets. To answer this question we extend the analysis by measuring the statistics of the first passage time when the number of principal target mRNAs reaches zero, using exact stochastic simulations (29) (for details see Text S1 Sec.D in the Supporting Material).

Our results suggest that the presence of auxiliary targets does not represent an overwhelming limitation to the reactivity of principal targets. As expected, the mean first-passage time increases as the number of auxiliary targets is increased, imposing an upper bound on that number. To quantify it, we define $n_{\max}$ as the number of auxiliary target molecules that slows down the response time such that it is comparable with the response time in the TF-based Scenario II. Fig. 3 depicts this number for typical sets of parameters. We find that in a significant part of the parameter space this number can be very large (1,000 auxiliary target molecules and more).

What determines the magnitude of $n_{\max}$ are the characteristics of the interaction between auxiliary targets and the srRNA. In particular, the rates of association and dissociation of srRNAs from their auxiliary targets turn out to be crucial for

setting an upper bound on the number of auxiliary targets allowed without dramatically affecting the response time (Fig.5 A). For example, if the srRNA-auxiliary target complex is very unstable (high dissociation constant $K_d \equiv k_d/k_a$), the effect of these targets on the response time is weak, and even a high number of targets can be present with no slow-down in response. On the other hand, if this complex is stable (low K_d), the dynamics is strongly affected even by a small number of auxiliary targets. Typical numbers for the association and dissociation rates ($k_a \approx k$ and $K_d \approx 10 - 100$, (30)) suggest that the number of auxiliary target molecules should be of order $10^3 - 10^4$ to significantly slow down the kinetic response of the principal targets.

The statistics of first-passage times allows us to consider another aspect of fluctuations that is typically ignored. In a fluctuating environment like the cell, the time at which all mRNA molecules initially present are degraded, bares some uncertainty: in one cell this time can be longer than in another. Looking back at the two scenarios described above, we notice that in the absence of auxiliary targets not only the response time is significantly reduced, but also the uncertainty (Fig.4, green arrow). This is yet another aspect of the robustness that small RNAs confer to their regulatory logic. Once again one may be worried that the presence of auxiliary targets might reverse this outcome, and make the response time not only longer but also significantly less predictable.

Our model shows that this is not the case (Fig.4 and Fig.5 B). For example, consider the extreme case where the number of auxiliary targets is $n_{\max}$, such that the mean response time is the same as in the TF-based Scenario II (blue arrow in Fig.4). We find that the variance of this first-passage time is still significantly (at least 2-fold) smaller in Scenario I than in Scenario II. Indeed, over the relevant range of parameters, and for any number of auxiliary targets below $n_{\max}$, the uncertainty in the response time in Scenario I is significantly reduced compared to Scenario II.

Buffering of environmental noise by auxiliary targets

The cellular regulatory network is responsive to environmental changes, such that a change in the environment should lead to a robust response in gene expression. At the same time it is possible that a transient change in the environment should be ignored. We have previously suggested that small RNAs offer an inherent mechanism to buffer transient environmental changes that would result in the activation of a target gene (35). In this mechanism, the accumulation of target mRNAs is delayed with respect to the activation of transcription. Accumulation of the mRNA

only starts after the pool of free srRNAs, that exists when the mRNA transcription is reactivated, is depleted. The size of the srRNA pool in the repressed state dictates the time window during which the changes in environmental signals are buffered.

Here we propose that the existence of auxiliary targets offers a similar mechanism in the opposite direction, i.e. when a change in environmental signals should result in suppression of the target (Scenario I above). We observe that under certain choices of parameters, the auxiliary targets provide the same kind of buffer for the srRNAs (Fig. 5 C,D). When the formation of auxiliary complexes is kinetically favored compared to the formation of principal complexes, srRNAs are first drawn to the pool of free auxiliary targets. Only after a significant fraction of these have been integrated into complexes, the srRNAs can efficiently interact with the principal targets and silence their expression. This introduces a delay in the response time, followed by an abrupt repression of the principal targets. Both parts of this kinetic process are relatively noise free, and together do not take longer than the more continuous (but slower) degradation occurring in Scenario II. This highlights a potential role for the auxiliary targets to delay the onset of the cell response without significantly affecting its efficiency.

Not surprisingly, the buffering effect exists when the rate of complex formation with auxiliary targets k_a is somewhat larger than the corresponding rate for principal targets k. The relevant range of parameters can be identified by observing the plateaus in the mean and the variance of the response time when plotted as a function of k. Such a range can be found for intermediate values of the association rate k_a and of the number of auxiliary targets n, such that $(k_a n)$ is comparable with k, but no fine-tuning is required (cyan lines in Fig. 5 C,D).

Non-specific evolutionary pressure on auxiliary targets

In our model, the precise identity of the auxiliary targets is not significant for their role in suppression of noise. Indeed, the only requirement is that these targets would not be sensitive to a minor down regulation by their ineffective interactions with the srRNA, and that they would be co-expressed (at least partially) with the principal targets. The one important property for the functionality of the auxiliary targets is at the system level: the number of auxiliary targets should be large enough to create a noise-suppressing reservoir, and on the other hand not too large to titrate the srRNA away from its principal targets or to significantly slow down their temporal response. A prediction of this model is that the evolutionary force to maintain the interaction between auxiliary targets and the srRNA is nonspecific,

and permits replacement of one target by another. We therefore ask if one could observe this evolutionary signature in available genomic data.

We focus on the conservation of microRNA targets across vertebrates for a set of 87 conserved microRNAs (Dataset S1 in the Supporting Material). Using the bioinformatic target prediction tool TargetScan (32), we generated a list of predicted targets for each microRNA in a set of 12 genomes. As a control, we employed the same procedure to a set of 1,000 "mock"-microRNAs, that were generated to mimic real microRNAs in many ways but carry a random seed (Dataset S2 in the Supporting Material).

To explore the conservation of target identities, we focused on microRNA-target pairs in human. For each such pair we computed the probability that the target gene interacts with the microRNA in the other 11 species. Fig. 6B shows the cumulative distribution of conservation scores. As expected, predicted targets of real microRNAs are slightly but significantly more conserved than those of mock microRNAs, suggesting a selection pressure acting on the interaction between natural microRNAs and their targets. Importantly, we observe that the total number of targets of each microRNA (but not their identity) remains strongly conserved among the 12 species (Fig.6A).

In order to classify the predicted targets as "principal" and "auxiliary", we used published experimental data following the change in the proteome in human HeLa cells following transfection of three different microRNAs (4). Targets that were strongly repressed (based on some arbitrary threshold) were classified as "principal" targets, while the many predicted targets that showed little or no response were classified as "auxiliary". We note that in this approximation we do not use – or have – any information about the physiological significance of the different targets.

As observed previously (48), our data suggests that targets in the "principal" group are more conserved than those in the "auxiliary" group. Indeed, the data of Fig. 6C, comparing the conservation of these two subsets of targets to those of mock microRNA, supports this claim. This reflects the strong evolutionary pressure that acts to conserve the interactions between microRNAs and their principal targets.

Taken together, we find that the number of targets for each microRNA is highly conserved, and that for "principal" targets this probably comes from the fact that the targets themselves are conserved. For the "auxiliary" targets, however, this target-number conservation is stronger than the conservation of target identity, supporting the notion that it is their collective effect that is important, not their individual regulation.

Discussion

Single microRNAs have been shown to regulate the expression of many different target genes (hundreds if not thousands), although most of them very weakly (4, 5). This seems to be a universal feature, applicable to many small RNA pathways (including those acting in bacteria) that rely on a very short seed for specificity. These findings stimulated the current view in the field, that small RNAs act as global modulators of gene expression at a network scale, rather than gene-by-gene.

On the other hand, in all cases we are aware of, where the functionality of a small RNA has been carefully studied, only a small number of target genes were identified as being phenotypically relevant (6–15). If this is the rule rather than the exception, one is lead to ask what is the function of the other many weak targets of a srRNA. One possibility is that seed matching between most of the targets and a srRNA are completely coincidental and neutral (49). However, it has been reported that most of these interactions are evolutionary conserved, suggesting possible functionality (3, 18).

Our results support an elegant viewpoint of the relationships among srRNA targets, initially formulated by Seitz (18), that is unique to srRNA regulation, but not limited to a particular pathway or organism. Even if the precise molecular mechanisms which control the actual strength of the srRNA regulation on a specific target remain unclear (50), evidence suggest a hierarchy among srRNA targets. ”Principal targets” are the ones behind the physiological role of the srRNA. Examples of such targets include the cancer-related proteins Suz12 for miR-200 (13), ANP32A and SMARCA4 for miR-21 (11), or integrin-α5, radixin and RhoA for miR-31 (10). In contrast, ”auxiliary targets” are not functionally regulated by the srRNA. In addition to the sponge-like effect (51) which allows a fine-tuning of the expression of principal targets (18), our model work suggests functional roles for auxiliary targets in conferring robustness to the regulation and the kinetics of the principal targets. At an affordable price of slowing down the (already accelerated) reactivity of regulation, auxiliary targets significantly reduce fluctuations in the level of the principal targets, while maintaining the sensitivity of the regulatory logic. This observation is robust over the large range of in vivo parameters and does not depend on the details of the model. The buffer effect induced by the presence of auxiliary targets may not be limited to srRNA regulation and has also been suggested as a functional role for the many ”decoy” binding sites of transcription factors along a genome (52).

These features may explain the widespread of srRNAs in development and

stress response pathways (1, 2) where a precise but reactive response is needed. Specificity of each post-transcriptional pathway may have governed the evolution of the number of auxiliary targets for a given srRNA to find a compromise between the loss in reactivity and the gain in fidelity.

What are the auxiliary targets? The emerging picture suggests that auxiliary targets should be co-expressed with the ”principal” targets of a srRNA. For a srRNA with multiple ”principal” targets, that are not co-expressed, we expect a set of auxiliary targets associated with each principal target. Any gene that share a pattern of expression with a principal target can be recruited as an auxiliary target, since the expression of these genes is hardly affected by the srRNA. Our results indicate that the exact number of auxiliary targets and their exact level of expression is of little consequence.

The role of most srRNA targets as auxiliary targets predicts that the evolutionary pressure in place to maintain the interaction between a srRNA and its auxiliary targets is non-specific (18), and cannot be explained in evolutionary models that accounts for targets ”one at a time” (53). The pressure to maintain auxiliary targets should also be contingent on the conservation of the principal targets. This type of evolutionary pressure should have unique fingerprints, and we expect it to be detectable as more whole-genome sequences become available.

Acknowledgments

We thank Craig Hunter, Philip Zamore, David Bartel and Benedikt Obermayer for discussions. This work was supported by the National Science Foundation under Grant No. MCB1121057.

Supporting citations

References (54–57) appear in the Supporting Material.

References

1. Gottesman, S., 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet* 21:399–404. http://dx.doi.org/10.1016/j.tig.2005.05.008.

2. Ghildiyal, M., and P. D. Zamore, 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10:94–108. http://dx.doi.org/10.1038/nrg2504.

3. Thomas, M., J. Lieberman, and A. Lal, 2010. Desperately seeking microRNA targets. *Nat Struct Mol Biol* 17:1169–1174. http://dx.doi.org/10.1038/nsmb.1921.

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

5. Selbach, M., B. Schwanhäusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 455:58–63. http://dx.doi.org/10.1038/nature07228.

6. Reinhart, B. J., F. J. Slack, M. Basson, A. E. Pasquinelli, J. C. Bettinger, A. E. Rougvie, H. R. Horvitz, and G. Ruvkun, 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 403:901–906. http://dx.doi.org/10.1038/35002607.

7. Slack, F. J., M. Basson, Z. Liu, V. Ambros, H. R. Horvitz, and G. Ruvkun, 2000. The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* 5:659–669.

8. Massé, E., C. K. Vanderpool, and S. Gottesman, 2005. Effect of RyhB small RNA on global iron use in Escherichia coli. *J Bacteriol* 187:6962–6971. http://dx.doi.org/10.1128/JB.187.20.6962-6971.2005.

9. Liu, K., Y. Liu, W. Mo, R. Qiu, X. Wang, J. Y. Wu, and R. He, 2011. MiR-124 regulates early neurogenesis in the optic vesicle and forebrain, targeting NeuroD1. *Nucleic Acids Res* 39:2869–2879. http://dx.doi.org/10.1093/nar/gkq904.

10. Valastyan, S., N. Benaich, A. Chang, F. Reinhardt, and R. A. Weinberg, 2009. Concomitant suppression of three target genes can explain the impact of a microRNA on metastasis. *Genes Dev* 23:2592–2597. http://dx.doi.org/10.1101/gad.1832709.

11. Schramedei, K., N. Mörbt, G. Pfeifer, J. Läuter, M. Rosolowski, J. M. Tomm, M. von Bergen, F. Horn, and K. Brocke-Heidrich, 2011. MicroRNA-21 targets tumor suppressor genes ANP32A and SMARCA4. *Oncogene* 30:2975–2985. http://dx.doi.org/10.1038/onc.2011.15.

12. Hyun, S., J. H. Lee, H. Jin, J. Nam, B. Namkoong, G. Lee, J. Chung, and V. N. Kim, 2009. Conserved MicroRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* 139:1096–1108. http://dx.doi.org/10.1016/j.cell.2009.11.020.

13. Iliopoulos, D., M. Lindahl-Allen, C. Polytarchou, H. A. Hirsch, P. N. Tsichlis, and K. Struhl, 2010. Loss of miR-200 inhibition of Suz12 leads to polycomb-mediated repression required for the formation and maintenance of cancer stem cells. *Mol Cell* 39:761–772. http://dx.doi.org/10.1016/j.molcel.2010.08.013.

14. Melton, C., R. L. Judson, and R. Blelloch, 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature* 463:621–626. http://dx.doi.org/10.1038/nature08725.

15. Zhao, C., G. Sun, S. Li, M.-F. Lang, S. Yang, W. Li, and Y. Shi, 2010. MicroRNA let-7b regulates neural stem cell proliferation and differentiation by targeting nuclear receptor TLX signaling. *Proc Natl Acad Sci U S A* 107:1876–1881. http://dx.doi.org/10.1073/pnas.0908750107.

16. Mavrakis, K. J., A. L. Wolfe, E. Oricchio, T. Palomero, K. de Keersmaecker, K. McJunkin, J. Zuber, T. James, A. A. Khan, C. S. Leslie, J. S. Parker, P. J. Paddison, W. Tam, A. Ferrando, and H.-G. Wendel, 2010. Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. *Nat Cell Biol* 12:372–379. http://dx.doi.org/10.1038/ncb2037.

17. Flynt, A. S., and E. C. Lai, 2008. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 9:831–842. http://dx.doi.org/10.1038/nrg2455.

18. Seitz, H., 2009. Redefining microRNA targets. *Curr Biol* 19:870–873. http://dx.doi.org/10.1016/j.cub.2009.03.059.

19. Swami, M., 2010. Small RNAs: Pseudogenes act as microRNA decoys. *Nat Rev Genet* 11:530–531. http://dx.doi.org/10.1038/nrg2835.

20. Levine, E., Z. Zhang, T. Kuhlman, and T. Hwa, 2007. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol* 5:e229. http://dx.doi.org/10.1371/journal.pbio.0050229.

21. Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer, 2006. Transcriptional pulsing of a developmental gene. *Curr Biol* 16:1018–1025. http://dx.doi.org/10.1016/j.cub.2006.03.092.

22. So, L.-H., A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, 2011. General properties of transcriptional time series in Escherichia coli. *Nat Genet* 43:554–560. http://dx.doi.org/10.1038/ng.821.

23. Berg, H. C., and E. M. Purcell, 1977. Physics of chemoreception. *Biophys J* 20:193–219. http://dx.doi.org/10.1016/S0006-3495(77)85544-6.

24. Bialek, W., and S. Setayeshgar, 2005. Physical limits to biochemical signaling. *Proc Natl Acad Sci U S A* 102:10040–10045. http://dx.doi.org/10.1073/pnas.0504321102.

25. van Kampen, N., 2001. Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam.

26. Mehta, P., S. Goyal, and N. S. Wingreen, 2008. A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol Syst Biol* 4:221. http://dx.doi.org/10.1038/msb.2008.58.

27. Tkacik, G., C. G. Callan, and W. Bialek, 2008. Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci U S A* 105:12265–12270. http://dx.doi.org/10.1073/pnas.0806077105.

28. Cover, T., and J. Thomas, 1991. Elements of Information Theory. John Wiley and Sons.

29. Gillespie, D., 1977. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.

30. Shimoni, Y., G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit, 2007. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* 3:138. http://dx.doi.org/10.1038/msb4100181.

31. Gentles, A. J., and S. Karlin, 2001. Genome-scale compositional comparisons in eukaryotes. *Genome Res* 11:540–546. http://dx.doi.org/10.1101/gr.163101.

32. Lewis, B. P., C. B. Burge, and D. P. Bartel, 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120:15–20. http://dx.doi.org/10.1016/j.cell.2004.12.035.

33. Bartel, D. P., and C.-Z. Chen, 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* 5:396–400. http://dx.doi.org/10.1038/nrg1328.

34. Mitarai, N., A. M. C. Andersson, S. Krishna, S. Semsey, and K. Sneppen, 2007. Efficient degradation and expression prioritization with small RNAs. *Phys Biol* 4:164–171. http://dx.doi.org/10.1088/1478-3975/4/3/003.

35. Levine, E., M. Huang, Y. Huang, T. Kuhlman, Z. Zhang, and T. Hwa, 2008. On noise and silence in gene regulation by small RNA.

36. Mitarai, N., J.-A. M. Benjamin, S. Krishna, S. Semsey, Z. Csiszovszki, E. Massé, and K. Sneppen, 2009. Dynamic features of gene expression control by small regulatory RNAs. *Proc Natl Acad Sci U S A* 106:10655–10659. http://dx.doi.org/10.1073/pnas.0901466106.

37. Jost, D., A. Nowojewski, and E. Levine, 2011. Small RNA biology is systems biology. *BMB Rep* 44:11–21. http://dx.doi.org/10.5483/BMBRep.2011.44.1.11.

38. Mukherji, S., M. S. Ebert, G. X. Y. Zheng, J. S. Tsang, P. A. Sharp, and A. van Oudenaarden, 2011. MicroRNAs can generate thresholds in target gene expression. *Nat Genet* 43:854–859. http://dx.doi.org/10.1038/ng.905.

39. Thattai, M., and A. van Oudenaarden, 2001. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* 98:8614–8619. http://dx.doi.org/10.1073/pnas.151588598.

40. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain, 2002. Stochastic gene expression in a single cell. *Science* 297:1183–1186. http://dx.doi.org/10.1126/science.1070919.

41. Paulsson, J., 2004. Summing up the noise in gene networks. *Nature* 427:415–418. http://dx.doi.org/10.1038/nature02257.

42. Elf, J., J. Paulsson, O. G. Berg, and M. Ehrenberg, 2003. Near-critical phenomena in intracellular metabolite pools. *Biophys J* 84:154–170. http://dx.doi.org/10.1016/S0006-3495(03)74839-5.

43. Levine, E., P. McHale, and H. Levine, 2007. Small regulatory RNAs may sharpen spatial expression patterns. *PLoS Comput Biol* 3:e233. http://dx.doi.org/10.1371/journal.pcbi.0030233.

44. Goldie, B. J., and M. J. Cairns, 2012. Post-transcriptional trafficking and regulation of neuronal gene expression. *Mol Neurobiol* 45:99–108. http://dx.doi.org/10.1007/s12035-011-8222-0.

45. Le Derout, J., I. Boni, P. Regnier, and E. Hajnsdorf, 2010. Hfq affects mRNA levels independently of degradation. *BMC Molecular Biology* 11:17. http://www.biomedcentral.com/1471-2199/11/17.

46. Krol, J., I. Loedige, and W. Filipowicz, 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* 11:597–610. http://dx.doi.org/10.1038/nrg2843.

47. Milo, R., P. Jorgensen, U. Moran, G. Weber, and M. Springer, 2010. BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res* 38:D750–D753. http://dx.doi.org/10.1093/nar/gkp889.

48. Friedman, R. C., K. K.-H. Farh, C. B. Burge, and D. P. Bartel, 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19:92–105. http://dx.doi.org/10.1101/gr.082701.108.

49. Chen, K., and N. Rajewsky, 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8:93–103. http://dx.doi.org/10.1038/nrg1990.

50. Grimson, A., K. K.-H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim, and D. P. Bartel, 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27:91–105. http://dx.doi.org/10.1016/j.molcel.2007.06.017.

51. Ebert, M. S., and P. A. Sharp, 2010. Emerging roles for natural microRNA sponges. *Curr Biol* 20:R858–R861. http://dx.doi.org/10.1016/j.cub.2010.08.052.

52. Burger, A., A. M. Walczak, and P. G. Wolynes, 2010. Abduction and asylum in the lives of transcription factors. *Proc Natl Acad Sci U S A* 107:4016–4021. http://dx.doi.org/10.1073/pnas.0915138107.

53. Shomron, N., D. Golan, and E. Hornstein, 2009. An evolutionary perspective of animal microRNAs and their targets. *J Biomed Biotechnol* 2009:594738. http://dx.doi.org/10.1155/2009/594738.

54. Berg, O. G., 1978. On diffusion-controlled dissociation. *Chemical Physics* 31:47 – 57. http://www.sciencedirect.com/science/article/pii/0301010478870256.

55. Hao, Y., L. Xu, and H. Shi, 2011. Theoretical analysis of catalytic-sRNA-mediated gene silencing. *J. Mol. Biol.* 406:195–204.

56. Walczak, A. M., G. Tkacik, and W. Bialek, 2010. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys Rev E Stat Nonlin Soft Matter Phys* 81:041905.

57. Mugler, A., B. Grinshpun, R. Franks, and C. H. Wiggins, 2011. Statistical method for revealing form-function relations in biological networks. *Proc Natl Acad Sci U S A* 108:446–451. http://dx.doi.org/10.1073/pnas.1008898108.

Tables

Table 1: **Diffusion mode in different organisms**

	E. coli	Budding yeast	HeLa cells
D (μm^{2}s^{-1})	5	60	100
Ω (μm^3)	0.7	33	2000
β (min^{-1})	1/10	1/20	1/600
β/d	0.05	0.09	0.11

Estimation of typical parameters for different organisms. The transition between fast and slow diffusion modes is mainly controlled by the ratio between the RNA self-degradation rate β and the diffusion rate $d \equiv D\ell/\Omega$ with D the effective diffusion constant, Ω the volume of the cell (or of the cytoplasm) and $\ell \sim 5$ nm the length-scale of the reaction volume (see Text S1 Sec.A). A small (high) ratio means fast (slow) transport mode. Numbers are taken from the BioNumbers database (47).

Figure Legends

Figure 1. Interactions of a srRNA with its targets. In the pathway, the main difference between auxiliary and principal targets resides in the fate of the mRNA-srRNA complexes: for principal targets, the complexes are rapidly degraded leading to a strong effect on the overall principal protein level, while for auxiliary targets, the complexes preferentially dissociate, leading to only a weak effect on the auxiliary protein level.

Figure 2. Steady-state properties of the principal targets in absence or presence of auxiliary targets. (**A**) Mean protein level of the principal targets as a function of the ratio α_s/α_m between the production rates of srRNA and principal mRNAs, and for different numbers n of auxiliary mRNA molecules (black: 0, red: 100, blue: 1,000, green: 10,000). (**B, C**) Effect of the number of auxiliary targets on the noise-to-signal ratio (Fano factor) of the principal target proteins in the slow (**B**) and fast (**C**) transport limit.

Figure 3. Impact of the srRNA degradation rate on the transition between fast and slow diffusion limit in presence of auxiliary targets. In the fast transport region (colored zone), auxiliary targets help maintaining a low noise level for the principal targets. In the slow transport region (black zone), diffusion noise dominates and the presence of auxiliary targets increases the internal fluctuations of the principal targets. Colors represent the upper bound of the number of auxiliary targets $n_{\max}$ imposed by limiting the slow-down in temporal response (see text).

Figure 4. Impact of auxiliary targets on the temporal response of the principal targets. The temporal response is quantified by the first-passage time, when the number of principal target mRNAs reaches zero. Solid lines are the mean first-passage time for the srRNA-based Scenario I and the TF-based Scenario II, as a function of the number of auxiliary targets. Dashed lines represent the standard deviation from the mean.

Figure 5. Impact of kinetic parameters on the efficiency of the temporal response of the post-transcriptional repression. (**A**,**B**) Number $n_{\max}$ of auxiliary targets (**A**) at which the mean time required for full suppression of the principal target by the srRNAs is equal to the corresponding time required by TF-based regulation, and the uncertainty in the response time (**B**) of the post-transcriptional

regulation for $n = n_{\max}$ (normalized by the corresponding value for the transcriptional regulation). (**C**,**D**) Mean and variance of the first-passage time (FPT) as a function of the ratio k/k_a between the interaction rate of the principal targets and the one of the auxiliary targets, for different numbers of auxiliary target molecules n and different values of the dissociation constant for the formation of the srRNA-auxiliary target complexes $K_d = 10$ (full line), 1 (dashed line), 0.1 (dotted and dashed line) and 0.01 (dotted line). Results are normalized by the corresponding values for the TF-repression.

Figure 6. Conservation of microRNA targets across vertebrates. (**A**) Cumulative distributions for the relative fluctuations of the number of targets per microRNA across species for natural (red) and fake (black) microRNAs. Differences between the two distributions are significant (p-value $< 10^{-4}$ for the Kolmogorov-Smirnov test). Number of targets is more conserved for natural microRNAs. (**B**) Cumulative distributions for the conservation scores of predicted targets of natural (red) and fake (black) microRNAs. Differences between the two distributions are small but significant (p-value $< 10^{-10}$, K.-S. test). Proportion of conserved targets is slightly more important for natural microRNAs. (**C**) Cumulative distributions for the conservation scores of a subset of genes which are weak (red dashed line) or strong (red dotted line) targets of 3 natural microRNAs (miR-1, miR-124 and miR-181) in human (see Materials and Methods). The red full line is the distribution for the merged ensemble (weak and strong). The black lines represent the corresponding cumulative distributions for the same genes but when associated with fake microRNAs. Differences between natural and random microRNAs are always significant (p-values $< 10^{-10}$, K.-S. test). Proportion of conserved targets is larger for strong targets of natural microRNAs, then for weak targets of natural microRNAs, then for targets of fake microRNAs.

Figures

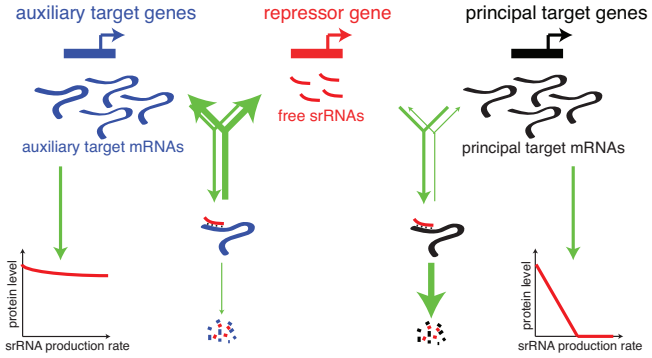

Figure 1:

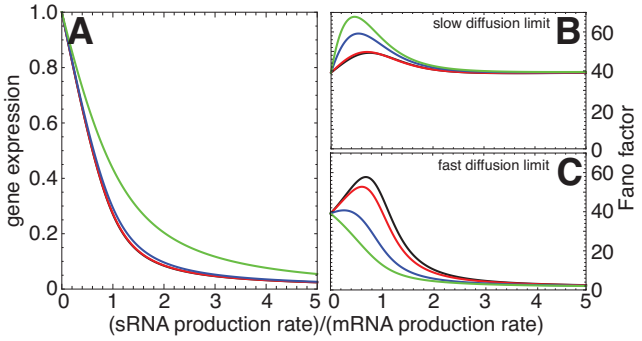
A
B
C
slow diffusion limit
fast diffusion limit
gene expression
Fano factor
(sRNA production rate)/(mRNA production rate)
0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
20
40
60

Figure 2:

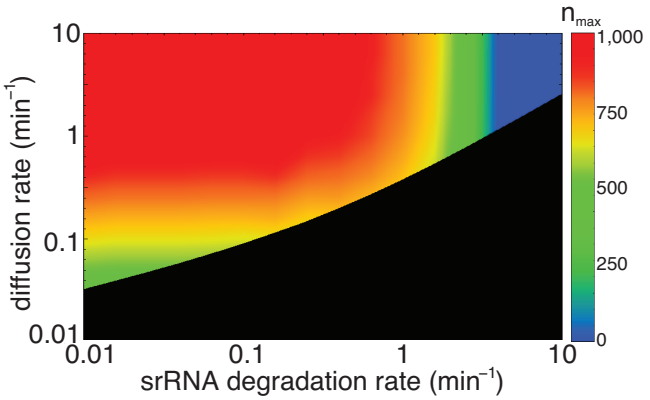
n_max
1,000
750
500
250
0
10
1
0.1
0.01
diffusion rate (min⁻¹)
0.01
0.1
1
10
srRNA degradation rate (min⁻¹)

Figure 3:

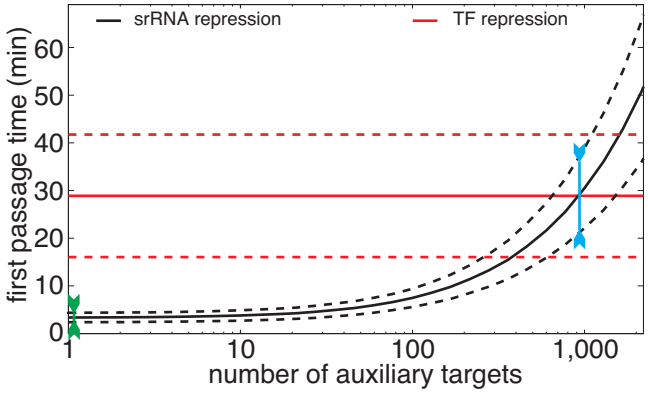
srRNA repression
TF repression
first passage time (min)
60
50
40
30
20
10
0
1
10
100
1,000
number of auxiliary targets

Figure 4:

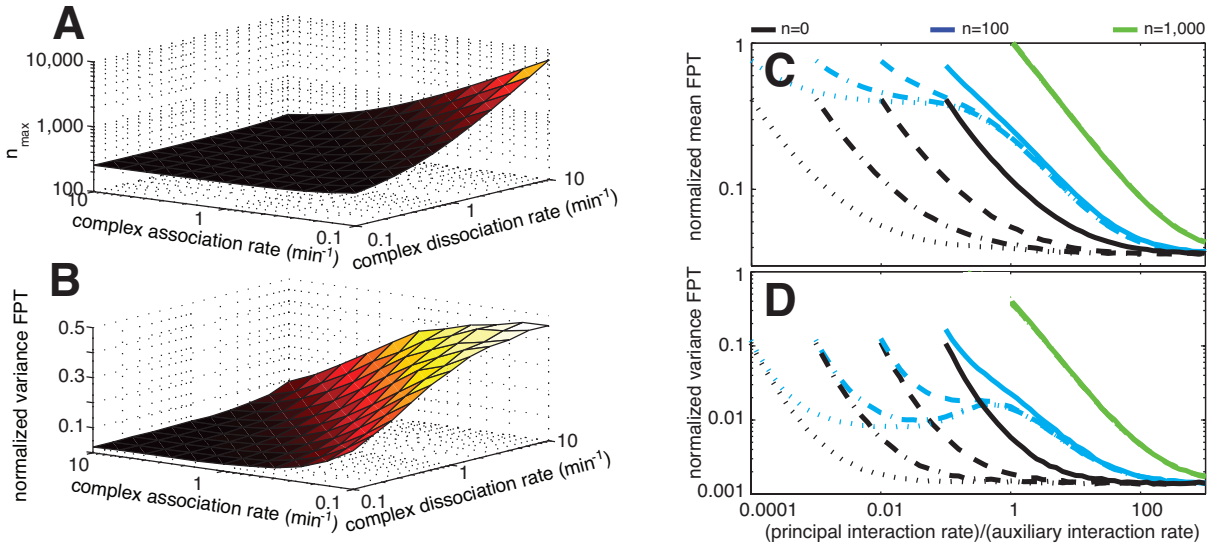
A
10,000
1,000
100
n_{max}
10
1
0.1
complex association rate (min^{-1})
0.1
1
10
complex dissociation rate (min^{-1})
B
normalized variance FPT
0.5
0.3
0.1
10
1
0.1
complex association rate (min^{-1})
0.1
1
10
complex dissociation rate (min^{-1})
n=0
n=100
n=1,000
C
normalized mean FPT
1
0.1
D
normalized variance FPT
1
0.1
0.01
0.001
0.0001
0.01
1
100
(principal interaction rate)/(auxiliary interaction rate)

Figure 5:

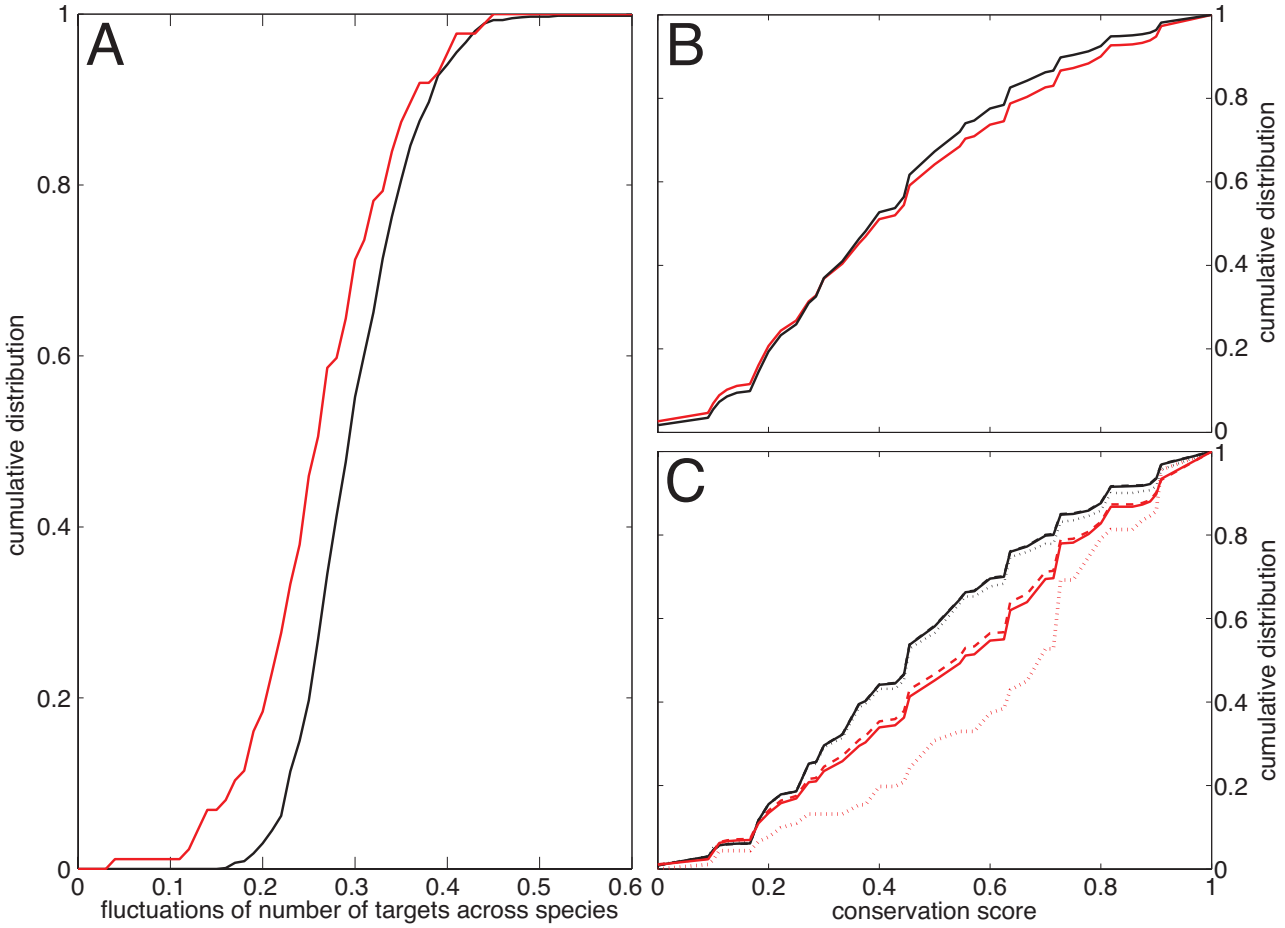
A
B
C
cumulative distribution
fluctuations of number of targets across species
conservation score
cumulative distribution
cumulative distribution
0
0.1
0.2
0.3
0.4
0.5
0.6
0.2
0.4
0.6
0.8
1

Figure 6:

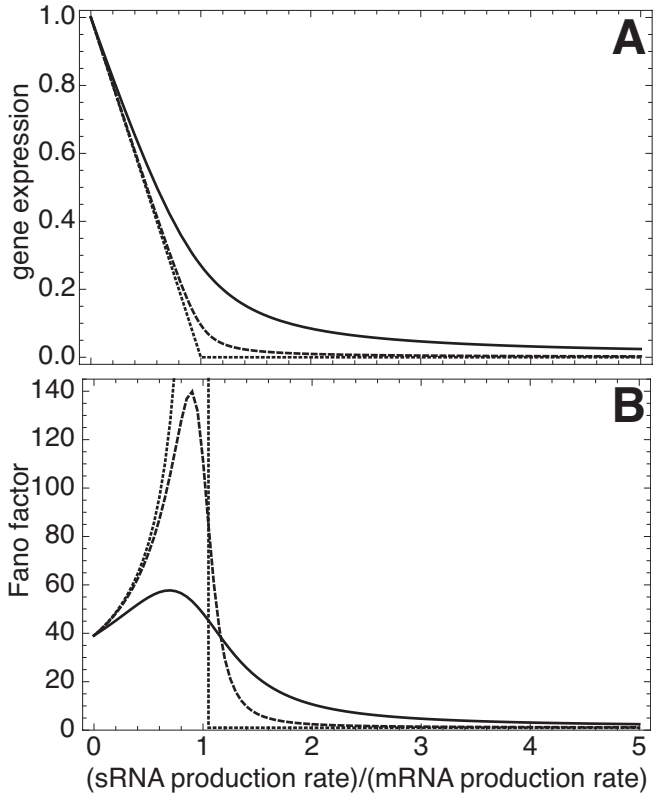

Fig. S1: Review of the steady-state properties of a protein regulated by a srRNA in the absence of auxiliary targets. Mean protein level (A) and the corresponding Fano factor (B) are plotted as a function of the ratio between the srRNA and the mRNA transcription rates for different values of the interaction rate k (full lines: 0.1 min^{-1}, dashed: 1, dotted: ∞). Strong interactions lead to sharp ("ultra-sensitive") linear-threshold response. Fixed parameters are (in min^{-1}) $\alpha_m = 1$, $\beta_m = 0.1$, $\beta_s = 0.1$, $\gamma = 4$, $\beta_p = 1/200$ (see SI Text, Sec.A for a detailed definition of each parameter).

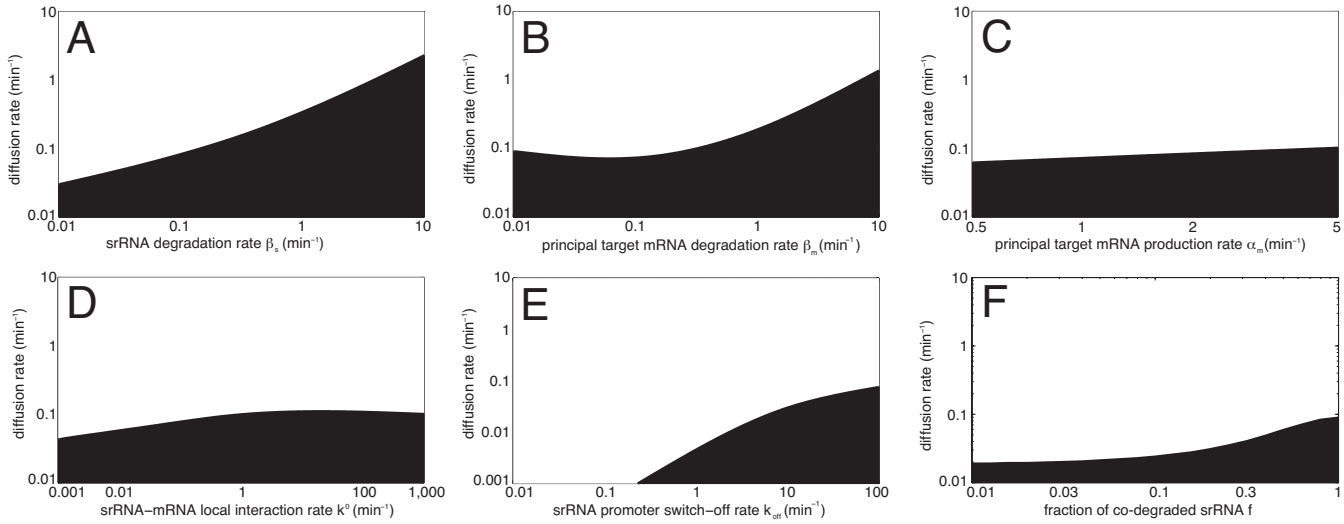

Fig. S2: Impact of parameters on the transition between fast and slow diffusion limit in presence of auxiliary targets. In the fast transport regions (white zones), auxiliary targets help maintaining a low level of intrinsic noise for the principal targets. In the slow transport regions (black zones), diffusion noise dominates and the presence of auxiliary targets increases the internal fluctuations of the principal targets. The frontier between fast and slow modes was computed using the notion of information capacity of the regulatory pathway (see SI Text Sec. C for details). In (E), we augment the burstiness of the srRNA promoter by decreasing the switch-off rate. Fixed parameters as in Fig. S1, completed by (in min^{-1}) $\beta_n = 0.1$, $\beta_c = 0$, $k_0 = 0.1$, $k_{a0} = 0.1$, $k_{d0} = 1$, (see SI Text Sec. A for a detailed definition of each parameter).

Text S1
Regulating the many to benefit the few: role of weak small RNA targets

Daniel Jost, Andrzej Nowojewski, and Erel Levine
Department of Physics, FAS Center for Systems Biology,
Harvard University, Cambridge, MA 02138, USA

Contents

Appendix A: Unified model for small RNA regulation

In this section, we present in details the model we used to study the srRNA-mediated regulation of principal targets in presence of auxiliary targets.

1. Stochastic model without auxiliary targets

The general picture of srRNA post-transcriptional regulation is well-described by modeling the dynamics of the number of srRNA s, of the number of principal mRNA targets m and of the number of target proteins p, by a set of mass-action equations [1–5] where intrinsic fluctuations are captured by adding Langevin-like noises ξ_i accounting for the stochasticity of the underlying reactions [6]:

$$\frac{ds}{dt} = \alpha_s - \beta_s s - ksm + \xi_s + \xi_k, \tag{A1}$$

$$\frac{dm}{dt} = \alpha_m - \beta_m m - ksm + \xi_m + \xi_k, \tag{A2}$$

$$\frac{dp}{dt} = \gamma m - \beta_p p + \xi_p, \tag{A3}$$

where α_i refers to the transcriptional rate of species i, β_i to its degradation or turnover rate, and γ to the translational rate of the target protein. k represents the second-order kinetic constant between the srRNA and its target mRNA. Eq.A1 and A2 assume that the pairing between the srRNA and the mRNA leads to a rapid and full degradation of the complex. This type of active co-degradation is believed to occur for many prokaryotic small RNA-mRNA couples [7]. In eukaryotic pathways, evidence suggests that the degradation of the mRNA in the complex does not always yield the degradation of the srRNA [8]. However, in many cases the mRNA-srRNA pair is sequestered for considerable time, thus titrating the srRNA. Moreover, within the same framework it is easy to extend the model such that only a fraction of the srRNA is co-degraded with the mRNA (see section B 3). This does not affect the results of our work (see section C 3 c). In the following, for simplicity, we consider the minimal model given above, except in the sections B 3 and C 3 c.

Within the Langevin framework, each reaction is treated as an independent Poissonian process with white delta-correlated noise

$$\langle \xi_i(t) \rangle = 0, \tag{A4}$$

$$\langle \xi_i(t)\xi_j(t') \rangle = \delta_{ij} N_i \delta(t-t'), \tag{A5}$$

with δ_{ij} the Kronecker symbol, $\delta(t)$ the Dirac function, $N_s = \alpha_s + \beta_s\langle s\rangle$, $N_m = \alpha_m + \beta_m\langle m\rangle$, $N_k = k\langle s\rangle\langle m\rangle$ and $N_p = \gamma\langle m\rangle + \beta_p\langle p\rangle$, where $\langle\cdot\rangle$ represents the expectation value. The Langevin approach can be viewed as an approximation of the corresponding Master-equation for large numbers of molecules [9]. One of its advantage is that it allows to add extra contribution to the noise (such as diffusion or external noise) in a simple way (see below).

Within the same framework, transcriptional burstiness [10–13] can be efficiently accounted for by allowing the promoter g to switch between two (on/off) states [6, 14]:

$$\begin{aligned} \frac{dg}{dt} &= k_{\text{on}}(1-g) - k_{\text{off}} g + \xi_g, \\ \text{with} \quad & \langle \xi_g(t) \rangle = 0, \\ & \langle \xi_g(t)\xi_g(t') \rangle = 2(k_{\text{on}} + k_{\text{off}})\langle g\rangle(1-\langle g\rangle)\delta(t-t') \\ \text{and} \quad & \alpha_i = \alpha_i^0 g\,. \end{aligned} \tag{A6}$$

Here k_{on} and k_{off} are respectively the on and off rate of the promoter, and α_i^0 the maximal transcription rate when the promoter is fully induced ($g = 1$). Arguably, the Langevin formalism is not natural to describe a two-state system. Indeed, instead of considering a binary switch dynamics between the two allowed states (on/off), g is a continuous variable, and its equation of evolution (Eq.A6) is equivalent to a diffusion process in a quadratic potential centered around $\langle g\rangle$. However, steady-state values for the mean and the variance of the promoter occupancy are perfectly described by the Langevin framework (see Fig. S4). As seen below, this requirement is enough to provide excellent agreement with stochastic simulations in the regime where the linear noise approximation is adequate.

2. Effect of diffusion

This Langevin formalism implicitly assumes that the different species are well-mixed in the cell and that spatial correlations could be neglected. This is true so long as the typical diffusion time across the cell (of typical dimension L) $\tau_c = L^2/D$ is much shorter that the different relevant time-scales of the system. However, even in a well-mixed solution, diffusion plays an effective role by renormalizing the interaction constant between two reactants [15]

$$k = \frac{4\pi D\ell k^0/\Omega}{4\pi D\ell + k^0}, \tag{A7}$$

with $\Omega \propto L^3$ the volume of the cell, D the relative diffusion constant between the two reactants, ℓ the typical scale of the reaction volume and k^0 the microscopic reaction rate which accounts for the local effective interaction between the two molecules. If $k^0 \gg D\ell$ then $k \approx 4\pi D\ell/\Omega$ and the macroscopic reaction is diffusion-limited. If $k^0 \ll D\ell$ then $k \approx k^0/\Omega$ and the interaction is reaction-limited.

Moreover, interaction between two molecules occurs in a very small volume ($\ell \ll L$). Therefore, diffusive arrival of molecules inside the reaction volume is also a stochastic process and leads to fluctuations in the local concentrations of the different species [14, 16, 17]. The corresponding variation in the rate ksm would be given in the linear noise approximation [9] by

$$\delta(ksm) \approx k\left(\langle s\rangle \delta m + \langle m\rangle \delta s\right), \tag{A8}$$

with δs (δm) the random variable characterizing the fluctuations of s (m) due to diffusion. As long as the typical diffusion time in the reaction volume ($\sim l^2/D$) is much shorter than the other characteristic times of the system, terms in Eq.A8 could be considered as a white Langevin noise defined as

$$\langle \delta(ksm)\rangle = 0, \tag{A9}$$

$$\langle \delta(ksm)(t)\delta(ksm)(t')\rangle = k^2\tau(\langle s\rangle^2|\delta m|^2 + \langle m\rangle^2|\delta s|^2)\delta(t-t'), \tag{A10}$$

where τ is a typical time-scale, and $|\delta s|^2$ and $|\delta m|^2$ are the strengths of the diffusion fluctuations [16]

$$|\delta s|^2 = \frac{\Omega/(D\ell)}{\tau}\langle s\rangle \quad , \quad |\delta m|^2 = \frac{\Omega/(D\ell)}{\tau}\langle m\rangle \quad , \tag{A11}$$

and $D = D_s + D_m$ is the relative diffusion constant between the two species.

Grouping together all the terms related to the intrinsic fluctuations of the interaction between the srRNA and the mRNA, the diffusion noise can be captured by correcting the amplitude of the Langevin noise ξ_k by

$$N_k = k\langle s\rangle\langle m\rangle\left(1 + \Sigma(\langle s\rangle + \langle m\rangle)\right), \tag{A12}$$

with $\Sigma = k\Omega/D\ell$. For diffusion-limited interactions, the prefactor Σ is constant and independent of D. For reaction-limited interactions, $\Sigma \sim k^0/(D\ell) \ll 1$, and the diffusion noise is negligible.

3. Accounting for interactions with auxiliary targets

Next, we generalize the framework discussed above by considering the interaction between the srRNA and the auxiliary targets. These targets represent mRNAs molecules which could interact with the srRNA (i.e., containing a binding site) but whose mean levels are just weakly affected by the srRNA. Interaction between a srRNA and an auxiliary target leads to the formation of transient complexes. The kinetics of the numbers of auxiliary targets n and of complexes c follow the Langevin-like equations

$$\frac{dn}{dt} = \alpha_n - \beta_n n - k_a sn + k_d c + \xi_n + \xi_a + \xi_d, \tag{A13}$$

$$\frac{dc}{dt} = -\beta_c c + k_a sn - k_d c + \xi_c - \xi_a - \xi_d, \tag{A14}$$

where α_n is the total production rate over all auxiliary targets. $k_a = (4\pi D\ell k_a^0/\Omega)/(4\pi D\ell + k_a^0)$ is the association rate between the srRNA and the auxiliary targets and $k_d = (4\pi D\ell k_d^0)/(4\pi D\ell + k_a^0)$ the dissociation rate of the transient complex. The dissociation constant $K_d = k_d/k_a = k_d^0\Omega/k_a^0$ is independent of the diffusion constant. Degradation of

the complex is accounted for via β_c. The Langevin noises are characterized by their amplitudes: $N_n = \alpha_n + \beta_n\langle n\rangle$, $N_c = \beta_c\langle c\rangle$, $N_a = k_a\langle s\rangle\langle n\rangle(1 + (k_a\Omega/D\ell)[\langle s\rangle + \langle n\rangle])$ and $N_d = k_d\langle c\rangle$.

To account for the effect of the auxiliary targets on the srRNA level, one has to augment Eq.A1 by $(-k_a sn + k_d c + \beta_c(1-p_d)c + \xi_a + \xi_d + \xi_c')$, with p_d the probability that the srRNA is also eliminated during the complex degradation, and $N_c' = \beta_c(1-p_d)\langle c\rangle$.

Note that for each type (principal and auxiliary) of targets, we focus on the dynamics of the whole ensemble. We do not consider possible heterogeneities (e.g. in the values of k or α_m) within each ensemble that, for example, may lead to hierarchical crosstalk between principal targets [1] (see Section C 4).

Appendix B: Steady-state properties without auxiliary targets

In this section, we review the steady-state properties of srRNA regulation for a single target.

1. Mean levels

The mean steady-state levels could be estimated by setting to zero all the time derivatives and the Langevin noises present in our set of equations (Eq.A1-A3) [1]:

$$\langle s\rangle = \frac{\alpha_s - \alpha_m - \lambda + \sqrt{(\alpha_m - \alpha_s - \lambda)^2 + 4\alpha_m\lambda}}{2\beta_s}, \tag{B1}$$

$$\langle m\rangle = \frac{\alpha_m - \alpha_s - \lambda + \sqrt{(\alpha_m - \alpha_s - \lambda)^2 + 4\alpha_m\lambda}}{2\beta_m}, \tag{B2}$$

$$\langle p\rangle = \frac{\gamma\langle m\rangle}{\beta_p}, \tag{B3}$$

with $\lambda = \beta_s\beta_m/k$ the leakage rate which control the efficiency of the regulation [1]. These equations describe a linear-threshold response for the protein level (see Fig. S1 A) where three different regimes could be identified: (i) an unrepressed regime ($\alpha_s \ll \alpha_m$) where the srRNA-induced degradation of the mRNA target is small and the protein is expressed ($\langle p\rangle \approx p_{\max}$); (ii) a crossover regime around the threshold ($\alpha_s = \alpha_m$) where both the srRNA and the mRNA levels are low; and (iii) a repressed regime ($\alpha_s \gg \alpha_m$) where most of the mRNAs are targeted by the large srRNA pool and the expression of the protein is very low ($\langle p\rangle/p_{\max} \approx \lambda/\alpha_s \ll 1$) .

2. Fluctuations

The stochastic nature of the biochemical reactions composing gene regulation pathways leads to intrinsic fluctuations around the mean signal levels [18–20]. Total intrinsic fluctuations in the output protein level are the results of the propagation of the different sources of intrinsic noise along the regulatory pathway [20]. A canonical way to appreciate the strength of protein fluctuations is to consider the Fano factor $\nu = C_{p,p}/\langle p\rangle$, with $C_{p,p}$ the variance of p at steady-state. ν is a measure of the noise-to-signal ratio, that allows estimating the deviation of the corresponding distribution of protein number from the Poisson limit ($\nu = 1$).

a. Linear noise approximation (LNA)

To compute $C_{p,p}$, we apply the linear noise approximation [9] to the set of Langevin equations. If we consider small perturbations around the steady-state ($s = \langle s\rangle + \delta s$, $m = \langle m\rangle + \delta m$ and $p = \langle p\rangle + \delta p$), the system of equations driving the evolution of the vector $\vec{\delta} = \{\delta s, \delta m, \delta p\}$ could be written in the general form

$$\frac{d}{dt}\vec{\delta} = J\vec{\delta} + \vec{\xi}, \tag{B4}$$

where $(J_{i,j})$ is the Jacobian of the full system and ξ_i the different noise contributions ($\langle\xi_i\rangle = 0$, $\langle\xi_i(t)\xi_j(t')\rangle = N_{i,j}\delta(t-t')$).

There are two standard (and equivalent) approaches to estimate correlations between species in the weak noise limit. The first consists in taking the Fourier transform of Eq.B4, leading to

$$\tilde{\vec{\delta}} = A^{-1}\,\tilde{\vec{\xi}}, \tag{B5}$$

with $A_{i,i} = \imath\omega - J_{i,i}$, $A_{i,j} = -J_{i,j}$ $(i \neq j)$, and $\langle\tilde{\xi}_i(\omega)\tilde{\xi}_j^*(\omega')\rangle = 2\pi N_{i,j}\delta(\omega-\omega')$. The power spectrum $S(\omega)$ is then obtained from

$$\begin{aligned}\langle\tilde{\vec{\delta}}(\omega)\tilde{\vec{\delta}}^\dagger(\omega')\rangle &= A^{-1}\langle\tilde{\vec{\xi}}(\omega)\tilde{\vec{\xi}}^\dagger(\omega')\rangle(A^{-1})^\dagger \\ &= 2\pi\delta(\omega-\omega')[A^{-1}N(A^{-1})^\dagger] \equiv 2\pi\delta(\omega-\omega')S(\omega).\end{aligned} \tag{B6}$$

Using the Wiener-Khintchine theorem, the covariance matrix is then given by

$$C \equiv \langle\vec{\delta}\vec{\delta}^\dagger\rangle = \int\frac{d\omega}{2\pi}S(\omega) = \int\frac{d\omega}{2\pi}[A^{-1}N(A^{-1})^\dagger]. \tag{B7}$$

The second method starts by explicitly writing the time-derivative of the covariance matrix

$$\frac{d}{dt}C \equiv \frac{d}{dt}\langle\vec{\delta}\vec{\delta}^\dagger\rangle = JC + CJ^\dagger + \langle\vec{\delta}\vec{\xi}^\dagger + \vec{\xi}\vec{\delta}^\dagger\rangle. \tag{B8}$$

Formally, the solution of Eq.B4 is given by

$$\vec{\delta} = \exp[Jt]\vec{\delta}_0 + \int_{t_0}^{t} dt'\exp[J(t-t')]\vec{\xi}(t'). \tag{B9}$$

Therefore,

$$\begin{aligned}\langle\vec{\delta}\vec{\xi}^\dagger + \vec{\xi}\vec{\delta}^\dagger\rangle &= \exp[Jt]\langle\vec{\delta}_0\vec{\xi}^\dagger\rangle + \int_{t_0}^{t}dt'\exp[J(t-t')]\langle\vec{\xi}(t')\vec{\xi}^\dagger(t)\rangle + \langle\vec{\xi}\vec{\delta}_0^\dagger\rangle\exp[J^\dagger t] + \int_{t_0}^{t}dt'\langle\vec{\xi}(t)\vec{\xi}^\dagger(t')\rangle\exp[J^\dagger(t-t')] \\ &= \exp[Jt]\langle\vec{\delta}_0\rangle\langle\vec{\xi}^\dagger\rangle + \int_{t_0}^{t}dt'\exp[J(t-t')]\delta(t-t')N + \langle\vec{\xi}\rangle\langle\vec{\delta}_0^\dagger\rangle\exp[J^\dagger t] + \int_{t_0}^{t}dt'N\delta(t-t')\exp[J^\dagger(t-t')] \\ &= N.\end{aligned} \tag{B10}$$

Using this result in Eq.B8, one arrives at

$$\frac{d}{dt}C = JC + CJ^\dagger + N, \tag{B11}$$

(which can also be obtained using the Ω-expansion of the Master-equation [9]). The steady-state covariance matrix is given by solving the linear system $JC + CJ^\dagger + N = 0$ (often referred to as the fluctuation-dissipation theorem).

In the following, analytical results were derived using the first method whereas numerical results were computed using the second.

b. Protein fluctuations

The Fourier transform method applied to Eq.A3 gives

$$\langle|\delta\tilde{p}|^2\rangle = \frac{2\beta_p p}{\omega^2+\beta_p^2} + \frac{\gamma^2}{\omega^2+\beta_p^2}\langle|\delta\tilde{m}|^2\rangle, \tag{B12}$$

$$C_{pp} = \langle\delta p^2\rangle = \int\frac{d\omega}{2\pi}\langle|\delta\tilde{p}|^2\rangle = p + \gamma^2\int\frac{d\omega}{2\pi}\frac{\langle|\delta\tilde{m}|^2\rangle}{\omega^2+\beta_p^2}. \tag{B13}$$

The last equation illustrates that the noise in the protein level is the sum of a characteristic Poisson noise and of the propagation of the mRNA noise through a lowpass filter of frequency β_p. Generally speaking, one expects the protein lifetime to be longer than interaction time-scales and mRNA lifetimes. Therefore, the lowpass frequency can be considered much lower than other typical frequencies in the system, and Eq.B13 reduces to

$$C_{pp} = p + \frac{\gamma^2}{2\beta_p}\langle|\delta\tilde{m}|^2(0)\rangle. \tag{B14}$$

The Fourier transform method applied to the couple (s, m) leads to

$$\delta\tilde{m} = \frac{(\imath\omega + \beta_s + km)\tilde{\xi}_m + (\imath\omega + \beta_s)\tilde{\xi}_k - km\tilde{\xi}_s}{(\imath\omega + \beta_s)(\imath\omega + \beta_m + ks) + km(\imath\omega + \beta_m)}. \tag{B15}$$

Accounting for transcriptional burstiness leads to the same expression if we renormalize $\tilde{\xi}_{s,m} \to \tilde{\xi}_{s,m} + \alpha^0_{s,m}\tilde{\xi}_{g_{s,m}}/(\imath\omega + k_{on_{s,m}} + k_{off_{s,m}})$.

Using Eq.B14 and B15 and after some algebra, we can find simple expressions for the Fano factor in the unrepressed and repressed regimes. In the unrepressed regime, the Fano factor is given by

$$\nu^{unrep} \approx (1 + b) + b\left(1 - \frac{\langle p\rangle}{p_{max}}\right)\left[\left(\frac{\alpha_m - \lambda}{\alpha_m + \lambda}\right) + \left(\frac{\alpha_m\beta_s\lambda}{(\alpha_m + \lambda)^2}\right)\left(\frac{\Omega}{2D\ell}\right)\right], \tag{B16}$$

where $b = \gamma/\beta_m$ is the protein burst size (average number of proteins produced per mRNA) and $p_{max} = \gamma\alpha_m/(\beta_p\beta_m)$ is the maximal mean protein number obtained when $\alpha_s = 0$. The first term of Eq.B16 reflects the burstiness of the translation of the mRNA [18]. Diffusion noise intervenes in the second (first-order) term and has only a very small effect on the global noise in this regime.

In the repressed regime, we find

$$\nu^{rep} \approx (1 + b^*) + b\beta_m\left(\frac{\Omega}{2D\ell}\right)\left(1 - \frac{b^*}{b}\right), \tag{B17}$$

where $b^* = b/(1 + k\langle s\rangle/\beta_m) = b\langle p\rangle/p_{max} \ll b$ is the effective protein burst size. b^* is much smaller than the natural value b since the effective lifetime of the mRNA ($[\beta_m + k\langle s\rangle]^{-1}$) is greatly reduced by srRNA-induced degradation. The last term in Eq.B17 is the signature of the diffusion noise (Eq.A12) which tends to increase the expression basal level. For slow diffusion (compared with β_m), the slow stochastic diffusion of a very low number of mRNA molecules can lead to relatively high fluctuations in the local mRNA concentration and diffusion noise may dominate. In the limit of high diffusion, intrinsic fluctuations are greatly suppressed by a strong srRNA regulation ($\nu^{rep} \ll \nu^{unrep}$).

In the crossover regime, the mRNA and srRNA levels are both very low. The ultra-sensitivity of the system leads to large (near critical) fluctuations [21] (see Fig. S1 B). Indeed, near the threshold, the cell state becomes broadly distributed and alternates between unrepressed and repressed states, yielding to a large distribution for the protein level with a high noise-to-signal ratio [22]. This effect is significantly enhanced by strong interaction ($k \gg 1$). In this regime, varying the diffusion modifies the strength of the fluctuations mainly by changing the sensitivity of the system via k (Eq.A7).

c. Comparison with stochastic simulations

We now compare the results of the linear noise approximation to those of stochastic simulation. Simulations were implemented using the Gillespie algorithm [23]. Fig. S3A shows the good agreement between the solutions of the rate equations (Eqs.B1-B3) and of the full-stochastic system. Small deviations are observed around the threshold where correlations between the levels of srRNA and mRNAs are important. As already pointed out by Mehta et al [6], Fig. S3B shows the good quality of the LNA to describe the mean and the variance of the different species, even in the cross-over regime where numbers of molecules are low and the fluctuations are high. The position and amplitude of the peak of the Fano factor, predicted by the LNA, deviate only slightly from the result from Gillespie simulations. However, while the LNA well describes the behavior of the covariance matrix of the system, it is known [21] that the classic LNA fails to reproduce the bistable distribution due to the ultra-sensitivity of the system close to the threshold.

As discussed above, the LNA perfectly describes the steady-state values for the mean and the variance of the promoter occupancy even when we consider transcriptional burstiness (Fig. S4), despite the intrinsic binary nature of the promoter states.

3. Effect of recycling the srRNA

Here we assume that only a fraction f of the srRNA is co-degraded with the principal target mRNA [24]. The system of mass-action equations describing the dynamics of s, m, p and c_0 (number of srRNA-mRNA complexes) is

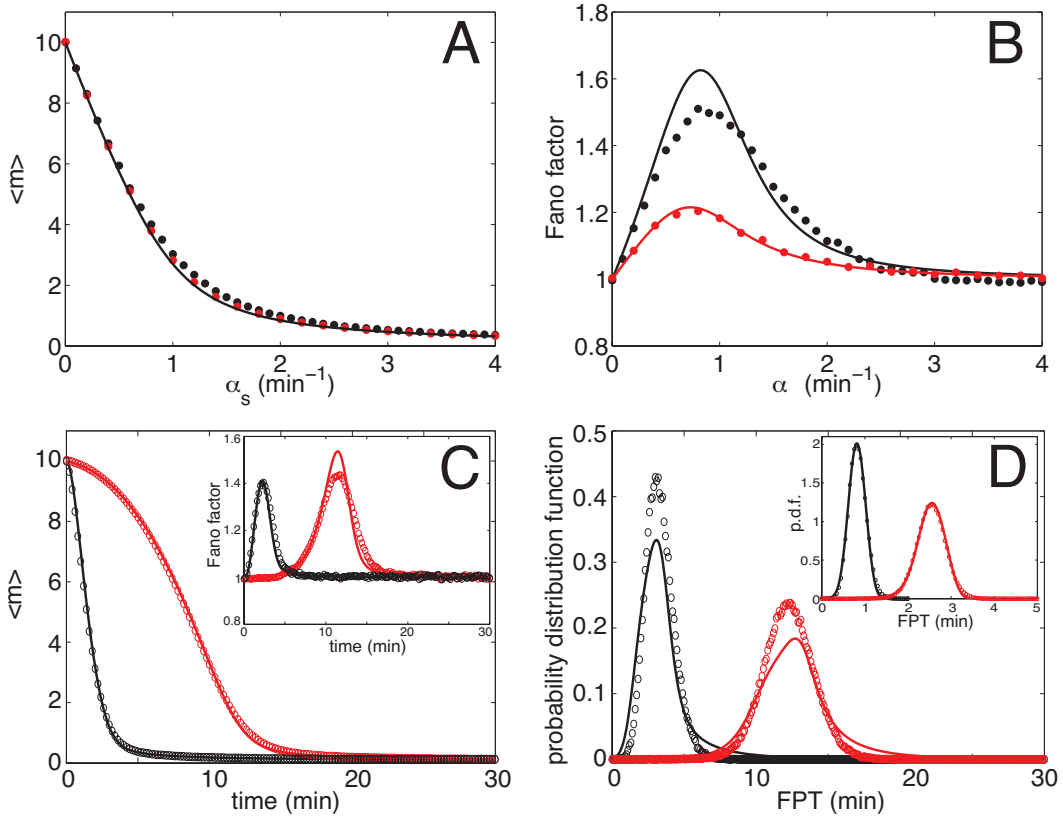

FIG. S3: Comparison of the linear-noise approximation (full lines) and stochastic simulations (dots/circles). Steady-state (A,B) and temporal response (C) for the mean number of principal targets $\langle m \rangle$ (A,C) and the corresponding Fano factor ν (B,inset in C) as a function of the srRNA production rate α_s (A,B) or time (C), in absence (black) or presence (red) of auxiliary targets ($\langle n \rangle = 10$ in A,B; and $n_{\rm tot} = 100$ in C,D). (D) Probability distribution function (p.d.f) of the first-passage time (FPT) for m ($\alpha_m = 1$ min^{-1}) to reaches 0, after the activation of the srRNA transcription ($\alpha_s = 10$ min^{-1}) . The inset in D shows the p.d.f of the FPT for m ($\alpha_m = 5$ min^{-1}) to reach $[\langle m \rangle(0)]/2$ ($\alpha_s = 50$ min^{-1}). Same parameters as in Fig. S1, completed by $k_a = k_d = 1$ min^{-1} and $\beta_n = 0.1$ min^{-1}.

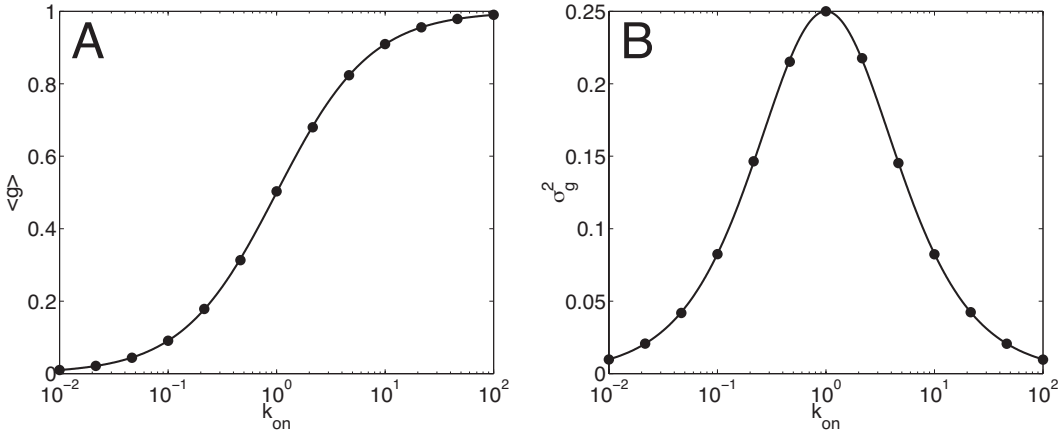

FIG. S4: Comparison of the linear-noise approximation (full lines) and stochastic simulations (dots) for the steady-state mean (A) and variance (B) of the promoter state g (see Eq.A6) if one considers transcriptional burstiness in the model ($k_{\rm off} = 1$).

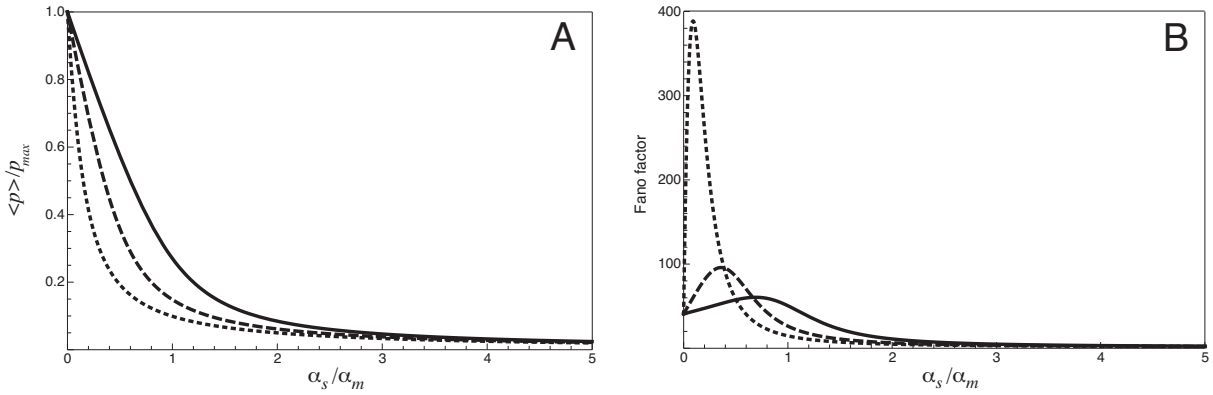

FIG. S5: Steady-state properties of the principal target proteins in absence of auxiliary targets when we consider a partial recycling of the srRNA. Mean protein level $\langle p\rangle/p_{max}$ (A) and Fano factor (B) as a function of the ratio α_s/α_m for different values of the fraction f of srRNA co-degraded with the principal target mRNA: $f = 1$ (full lines), 0.5 (dashed lines) and 0.1 (dotted lines). Fixed parameters as in Fig. S1.

given by

$$\frac{ds}{dt} = \alpha_s - \beta_s s - k_+ sm + k_- c_0 + (1-f)\beta_{c_0} c_0 + \xi_s + \xi_{k_+} + \xi_{k_-} + \xi'_{c_0}, \tag{B18}$$

$$\frac{dm}{dt} = \alpha_m - \beta_m m - k_+ sm + k_- c_0 + \xi_m + \xi_{k_+} + \xi_{k_-}, \tag{B19}$$

$$\frac{dc_0}{dt} = k_+ sm - k_- c_0 - \beta_{c_0} c_0 - \xi_{k_+} - \xi_{k_-} + \xi_{c_0}, \tag{B20}$$

$$\frac{dp}{dt} = \gamma m - \beta_p p + \xi_p, \tag{B21}$$

with k_+ the association rate between s and m, k_- the dissociation rate of c_0 and $\beta_{c_0} \gg k_-$ the active degradation rate of the complexes. The amplitude of the Langevin noise terms are: $N_{k_+} = k_+\langle s\rangle\langle m\rangle$, $N_{k_-} = k_-\langle c_0\rangle$, $N_{c_0} = \beta_{c_0}\langle c_0\rangle$ and $N'_{c_0} = \beta_{c_0}(1-f)\langle c_0\rangle$.

Substituting the steady-state mean value of c_0 ($\langle c_0\rangle = k_+\langle s\rangle\langle m\rangle/(k_- + \beta_{c_0})$) in Eq.B18 and B19 leads to the system

$$0 = \alpha_m - \beta_m\langle m\rangle - k\langle s\rangle\langle m\rangle \tag{B22}$$

$$0 = \alpha_s - \beta_s\langle s\rangle - kf\langle s\rangle\langle m\rangle \tag{B23}$$

with $k = k_+\beta_{c_0}/(k_- + \beta_{c_0})$. The mean gene expression of the principal targets is then given by

$$\langle p\rangle = \frac{1}{2\alpha_m}\left(\alpha_m - \alpha_s/f - \lambda' + \sqrt{[\alpha_m - \alpha_s/f - \lambda')^2 + 4\alpha_m\lambda'}\right) p_{max} \tag{B24}$$

with $\lambda' = \beta_s\beta_m/(fk)$. Thus, accounting for f renormalizes the leakage rate λ and the threshold position α_s/f, but does not change the general shape of the linear-threshold response. Since a fraction of srRNA are recycled, for a fixed srRNA transcription rate, the regulation will be more efficient for smaller f (Fig. S5A).

Fig. S5B shows a significant increase in the noise-to-signal ratio of the principal target proteins when the srRNA is highly recycled. Indeed, as the fluctuations of s and m are highly correlated in the ultra-sensitive cross-over regime [6, 21, 22], the intrinsic noise of the output proteins will suffer from the increase of the srRNA fluctuations due to the stochastic degradation of the complexes c_0.

Appendix C: Steady-state properties with auxiliary targets

In this section, we discuss the impact of auxiliary target on steady-state properties of the principal targets.

1. Mean levels

Substituting the steady-state mean value of c ($\langle c\rangle = k_a\langle s\rangle\langle n\rangle/(k_d+\beta_c)$) in Eq.A1 and A13, allows interpreting the auxiliary targets as stoichiometric weak targets of the srRNA with an effective interaction constant $k_{\text{eff}} = k_a p_d(\beta_c/k_d)/(1+\beta_c/k_d) \ll k$:

$$\frac{d\langle m\rangle}{dt} = 0 \;=\; \alpha_m - \beta_m\langle m\rangle - k\langle s\rangle\langle m\rangle \tag{C1}$$

$$\frac{d\langle s\rangle}{dt} = 0 \;=\; \alpha_s - \beta_s\langle s\rangle - k\langle s\rangle\langle m\rangle - k_{\text{eff}}\langle s\rangle\langle n\rangle \tag{C2}$$

$$\frac{d\langle n\rangle}{dt} = 0 \;=\; \alpha_n - \beta_n\langle n\rangle - \frac{k_{\text{eff}}}{p_d}\langle s\rangle\langle n\rangle \tag{C3}$$

The explicit solution of this system is quite cumbersome so we do not give its exact form here.

The level of auxiliary targets would start to play a significant role only when $k_{\text{eff}}\langle n\rangle \sim k\langle m\rangle$ [1], i.e when $\alpha_n \sim (k/k_{\text{eff}})\alpha_m \gg \alpha_m$. In this situation, the auxiliary targets modify the steady-state level of the srRNA and then indirectly the one of the principal target. This could help to finely tune the position of the transition between repressed and unrepressed regime (Fig.2A of the main text) but this effect does not change the regulatory logic of the post-transcriptional regulation.

In the following we focus on the regime where the auxiliary targets have a negligible effect on the steady-state mean levels of the free srRNAs and of the principal targets.

2. Fluctuations

To appreciate the effect of auxiliary targets on the fluctuations of principal targets, we first simplify the model by assuming that auxiliary targets have no effect on the mean steady-state level of principal targets ($\beta_c/k_d \to 0$). Effects of a finite β_c on the fluctuations are briefly discussed in the end of Sec.C 3. Results are obtained, as previously, within the linear noise approximation.

Applying the Fourier transform method to n and c (Eqs. A13 and A14), we find expressions for $\delta\tilde{n}$ and $\delta\tilde{c}$ as a function of $\delta\tilde{s}$. Substituting these expressions in the linearized Fourier equation for s gives

$$\delta\tilde{s} = \frac{-k\langle s\rangle\delta\tilde{m} + \xi_s + \xi_k + \imath\omega\frac{(\imath\omega+\beta_n)}{E}(\xi_a+\xi_d) - \imath\omega\frac{k_a\langle s\rangle}{E}\xi_n}{\beta_s + \imath\omega(k_a\langle n\rangle)\frac{(\imath\omega+\beta_n)}{E} + k\langle m\rangle}, \tag{C4}$$

with $E = (\imath\omega+k_d)(\imath\omega+\beta_n) + \imath\omega k_a\langle s\rangle$.

For high numbers of auxiliary targets ($\langle n\rangle/K_d \gg 1$), Eq.C4 leads to

$$\langle|\delta\tilde{s}|^2\rangle \approx \langle s\rangle\frac{\Omega}{D\ell}. \tag{C5}$$

The fluctuations due to the interaction with the principal targets ($-k\langle s\rangle\delta\tilde{m}$ and ξ_k) as well as those due to the srRNA production (including transcriptional burstiness) and degradation (ξ_s) have been absorbed by the large pool of complexed srRNAs, leaving only the contribution due to diffusion noise. Since fluctuations of s propagate in those of m and p, the Fano factor is given by

$$\nu = 1 + b\left[\frac{1+(k^2\langle m\rangle\langle s\rangle/\alpha_m)(2\langle m\rangle+\langle s\rangle)(\Omega/2D\ell)}{1+k\langle s\rangle/\beta_m}\right]. \tag{C6}$$

For high diffusion constant, this expression reduces to $\nu = 1+b^*$. Note that unlike the case with no auxiliary target, where this expression was only valid in the repressed regime (see Eq.B16), here the reduction of fluctuations holds for any level of the srRNA.

For an arbitrary number of auxiliary targets, no simple expression can be found for ν. Figures 2 B,C of the main text show numerical results of the Fano factor for different values of the parameters.

It should be noted that the relevant parameter here is not $\langle n\rangle$ but $\langle n\rangle/K_d$, since the efficacy of the fluctuation absorption depends on the size of the complexed pool ($\langle c\rangle = (\langle n\rangle/K_d)\langle s\rangle$). Changing α_n, β_n, k_a and k_d while keeping $\langle n\rangle/K_d$ constant does not impact significantly on the behavior of the process.

As before, Fig. S3B shows the remarkable ability of the LNA to describe the general behavior of post-transcriptional regulation in presence of auxiliary targets.

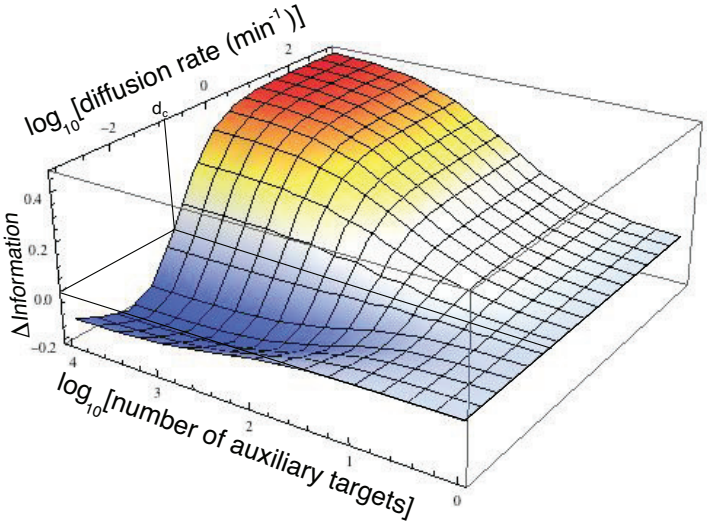

FIG. S6: Change in the information capacity due to the presence of auxiliary targets, for different values of the diffusion rate $d = D\ell/\Omega$. Same parameters as in Fig. S2

3. Information capacity

In the last section we saw that interactions with auxiliary targets reduce or increase the noise-to-signal ratio, depending on the importance of diffusion noise in the system. Next we aimed to quantify the impact of these results on the function of the srRNA regulatory pathway. Motivated by its recent successful applications to gene regulation [25–28], we decide to focus on the capacity of this pathway to convey information.

In information theory, transfer of information between an input X and an output Y via a given channel is well-characterized by the concept of mutual information [29]:

$$I(X;Y) = \int dX \int dY\, P(X,Y) \log_2 \left[\frac{P(X,Y)}{P(X)P(Y)} \right], \tag{C7}$$

where $P(X,Y)$ is the joint probability distribution function of X and Y, $P(X)$ and $P(Y)$ are the marginal probability distribution functions. Intuitively, the mutual information represents the number of different output levels achievable by varying the input, and depends on the intrinsic noise properties of the channel. The maximal mutual information with respect to the distribution of inputs is called the information capacity of the channel.

Since biological circuits could be seen as information channel between environmental or internal signals and internal responses, information theory concepts have been recently applied to genetic networks. It can be shown [25, 26] that the information capacity of such circuit could be approximated, in the low-noise limit, by

$$I_{\max} = \log_2 \left(\frac{1}{\sqrt{2\pi e}} \int dX \left| \frac{d\langle Y \rangle(X)}{dX} \right| \frac{1}{\sigma_Y(X)} \right), \tag{C8}$$

with σ_Y^2 the variance of the output for a given input level. The information capacity therefore depends directly on the variation in the mean output level in response to a change in the input level and on the intrinsic noise properties of the channel (via σ_Y). For example, a system with a flat or a very noisy response cannot carry information, and would have a low information capacity.

a. *Effect of auxiliary targets*

We use Eq.C8 to estimate the global effect of auxiliary targets and diffusion on the information capacity of the srRNA post-transcriptional regulation. We define the input as the transcription rate α_s of the srRNA and the output as the protein level p. Fig. S6 shows the effect of auxiliary targets on the information capacity of the srRNA regulation. For slow transport, diffusion noise dominates and the information capacity is reduced by an increase of the number of targets. In contrast, for fast transport, the pool of complexed srRNA allows an efficient absorption of the srRNA

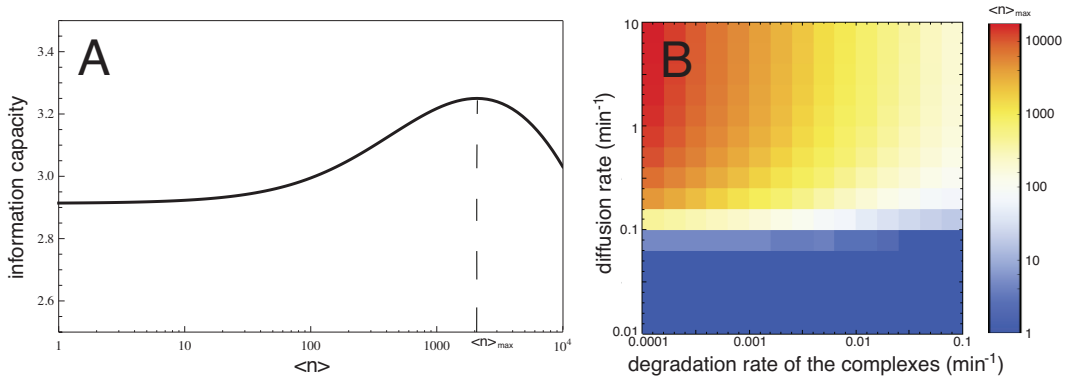

FIG. S7: Limitation of information gain due to the degradation of the complex when $d > d_c$. (A) Information capacity as a function of the number of auxiliary targets ($d \to \infty$, $\beta_c = 1/200$). (B) Number of auxiliary targets optimizing the information capacity as a function of the diffusion rate d and of the degradation rate β_c of the complexes. Same fixed parameters as in Fig. S2.

fluctuations in the crossover region, and promotes a good information transfer. More precisely, there exists a value for the diffusion parameter $d \equiv D\ell/\Omega$ such that, for $d > d_c$, adding auxiliary targets in the system decreases globally the intrinsic noise of the output protein and improve the information capacity of the post-transcriptional regulation; for $d < d_c$, auxiliary targets increase the output noise and the information transmission is less efficient.

The value of d_c depends on the different parameters of the system in a non trivial way, as shown in Fig. S2. As expected the value of the production rate of the principal targets does not influence the position of the transition (Fig. S2 C), since α_m mainly defined the position of the different regulation regimes but not the general behavior of the process. More surprisingly, d_c does not depend significantly on the microscopic reaction rate k^0 (Fig. S2 D).

The main contribution comes from the srRNA and mRNA turnover rates (Fig. S2 A,B). Since β_m controls the strength of the diffusion noise of the srRNA-mRNA interaction (see Eq.B17), high mRNA degradation rate will have an impact on d_c. Regarding β_s we note that at the transition the fluctuations of the srRNA with (see Eq.C5) or without ($\langle|\delta\tilde{s}|^2\rangle \sim \langle s\rangle/\beta_s$) auxiliary targets have the same amplitude. This implies that $\beta_s \sim d$. Therefore, increasing β_s reduces the influence of burstiness absorption, and diffusion noise starts to dominate at higher values of the diffusion constant.

When the srRNA transcription is highly bursty, one gain in information capacity even at slow diffusion (Fig. S2 E). This is due to buffering of transcriptional noise, which is higher for a bursty promoter, by the auxiliary targets. Diffusion noise, however, does not depend on the nature of the transcription process. Thus, for bursty promoters the advantage of auxiliary targets surpasses the disadvantage dues to diffusion noise even at slow diffusion.

b. Effect of the degradation of the complex

The previous part suggests that if $d > d_c$, the more auxiliary targets the better. However, as explained in Sec.C 1, at high number of auxiliary targets, the level of the free srRNAs and therefore that of the principal targets become influenced by the presence of the auxiliary targets due to the (slow but finite) degradation of the complexes. If we set β_c to a non-zero value, in addition to affect the fluctuations ($\sim \sigma_Y$) of principal targets, the number of auxiliary targets will also influence the sensitivity of the regulatory logic of the principal targets ($\sim d\langle Y\rangle/dX$). Therefore, from Eq.C8 we expect a change in the behavior of the information capacity as a function of n. While the boundary between slow and fast transport mode does not significantly depend on the degradation rate of the complex (Fig. S7B), information capacity in the fast diffusion regime exhibits a maximum at a finite number of auxiliary targets (Fig. S7A). This then defines an upper limit for the number of auxiliary targets that improves the information transfer. Auxiliary targets act also as kinetic traps for the srRNAs and slow-down the regulation process, as discussed below, imposing a stronger bound on the number of auxiliary targets.

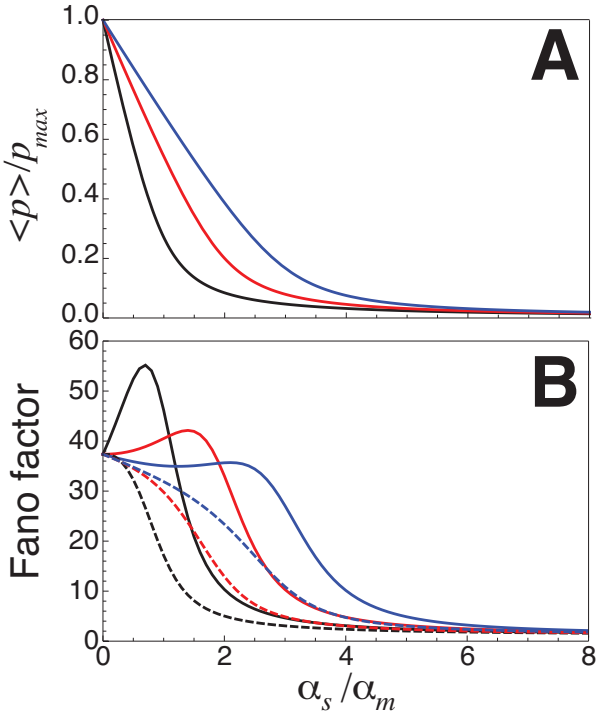

FIG. S8: Impact of strong auxiliary targets on the steady-state mean (A) and Fano factor (B) of the principal targets for $\alpha_a/\alpha_m = 0$ (black), 1 (red) and 2 (blue), in absence (full lines) or in presence (dashed lines) of weak auxiliary targets. Fixed parameters as in Fig. S1

c. *Effect of recycling the srRNA*

We saw in section B 3 that the partial recycling of the srRNA during the degradation of the complex srRNA-principal mRNA increased the intrinsic fluctuations of the principal targets. Therefore, as in the situation of a bursty srRNA transcription, one gain in information capacity even at slow diffusion (Fig. S2 F). Indeed, auxiliary targets buffer the intrinsic noise of the free srRNAs which is more important at low f, and diffusion noise does not impact the degradation of the complex c_0. Thus, the advantage of auxiliary targets surpasses the disadvantage due to diffusion noise even at slow diffusion.

4. Effect of strong auxiliary targets

In our work, we have identified the principal targets with the few strongly affected targets. However, some of these targets may not be phenotypically relevant and may be defined as "strong auxiliary" targets [30]. In this section, we briefly discuss the effect of such targets on the steady-state properties of the principal targets by augmenting our set of Langevin equations by

$$\frac{dm_a}{dt} = \alpha_a - \beta_m m_a - ks\, m_a + \xi_a + \xi_{ka} \tag{C9}$$

with m_a the number of strong auxiliary mRNAs.

Fig. S8A shows that strong auxiliary targets have a significant impact on the regulatory logic of principal targets, shifting the threshold position ($\alpha_s \sim \alpha_m + \alpha_a$) to higher values. They play a role a sponge by sequestering the srRNA molecules away from the principal targets and therefore delaying the entry in the silenced regime.

However, Fig. S8B illustrates that the buffer effect played by weak auxiliary targets is not perturbed by the presence of strong auxiliary targets.

Appendix D: Kinetic properties

In this section, we describe the numerical method used to compute the first passage time (FPT) statistics. We also present results on FPT if we account for diffusion noise.

1. Method

We use the Gillespie algorithm [23] to sample exact stochastic trajectories of the process where, for simplicity, we neglect fluctuations in the total number of auxiliary targets $n_{\rm tot}$. Each simulation starts with a configuration sampled from the steady state distribution of principal targets (Poissonian distribution with mean α_m/β_m). At $t = 0$ the transcription of srRNA is switched on, and the kinetic response is assessed by measuring the statistics of the first passage time when the number of principal targets reaches zero.

2. Effect of diffusion

To estimate the effect of diffusion on the first-passage time distribution, we use the linear-noise approximation (see Sec.B 2 a). In the LNA, fluctuations around the mean are assumed to be small and normally distributed [9]. The total probability $P(m, s, c; t)$ to observe the microstate (m, s, c) at time t is then given by a multivariate normal distribution

$$P(m, s, c; t) = \frac{1}{(2\pi)^{3/2}|det[C(t)]|^{1/2}} \exp[-X(t)^\dagger C^{-1}(t)X(t)/2], \tag{D1}$$

with the vector $X = (m - \langle m\rangle(t); s - \langle s\rangle(t); c - \langle c\rangle(t))$. The evolution of the covariance matrix $C(t)$ follows Eq.B11. Denoting $S_{m_0}(t) = \int_{m_0}^{\infty} dm \int dsdc\, P(m, s, c; t)$ the probability for m to be larger than m_0 at time t, the first-passage time probability $h_{m_0}(t)$ at m_0 is given by

$$h_{m_0}(t) = \frac{-\frac{d}{dt}\left[S_{m_0}(t)\right]}{S_{m_0}(0) - S_{m_0}(\infty)}. \tag{D2}$$

Fig. S3C shows the very good agreement between results from the LNA and those from the Gillespie algorithm for the time evolution of the mean and the variance of m. However, the distribution of the FPT to reach $m_0 = 0$ predicted by the LNA has a long tail which dramatically increases (by almost 8 times) the predicted variance for the FPT (see Fig. S3D). This underlines the limits of the LNA to describe the exact probability distribution functions in systems with small number of particles.

In order to study the effect of diffusion on the FPT distribution, we place ourselves in a regime where the LNA should well characterize also the first-passage time. We choose higher production rates for m and s ($\alpha_m = 5$ min^{-1}, $\langle m\rangle(0) = 50$ and $\alpha_s = 50$ min^{-1}) and we study the first-passage time to reach $m_0 = \langle m\rangle(0)/2 = 25$. The inset in Fig. S3D confirms the success of the LNA in this regime.

Fig. S9A and B show the mean and variance of the FPT as a function of the diffusion rate and of the number of auxiliary targets. Low diffusion rates exhibit higher mean and variance. Even for a small number of auxiliary targets where the mean response of the srRNA-repression outperforms the one of the TF-like repression, the uncertainty of the response in the slow transport mode could be very large.

The entire probability distribution function for the FPT as a function of the diffusion rate is shown in Fig. S9C, D and E. For fast diffusion, as expected, the distributions are well peaked around a mean value, which increases with the number of auxiliary targets. When the diffusion is slow, however, all interactions become diffusion limited, and diffusion dominates the noise. In this limit the distribution is peaked at short first-passage times, due to possible avoidance of the auxiliary targets, but exhibits a long tail, due to possible long searches for the principal targets.

Appendix E: Conservation of microRNA targets

In this section we briefly describe the methods used to analyze the evolutionary data on microRNA targets.

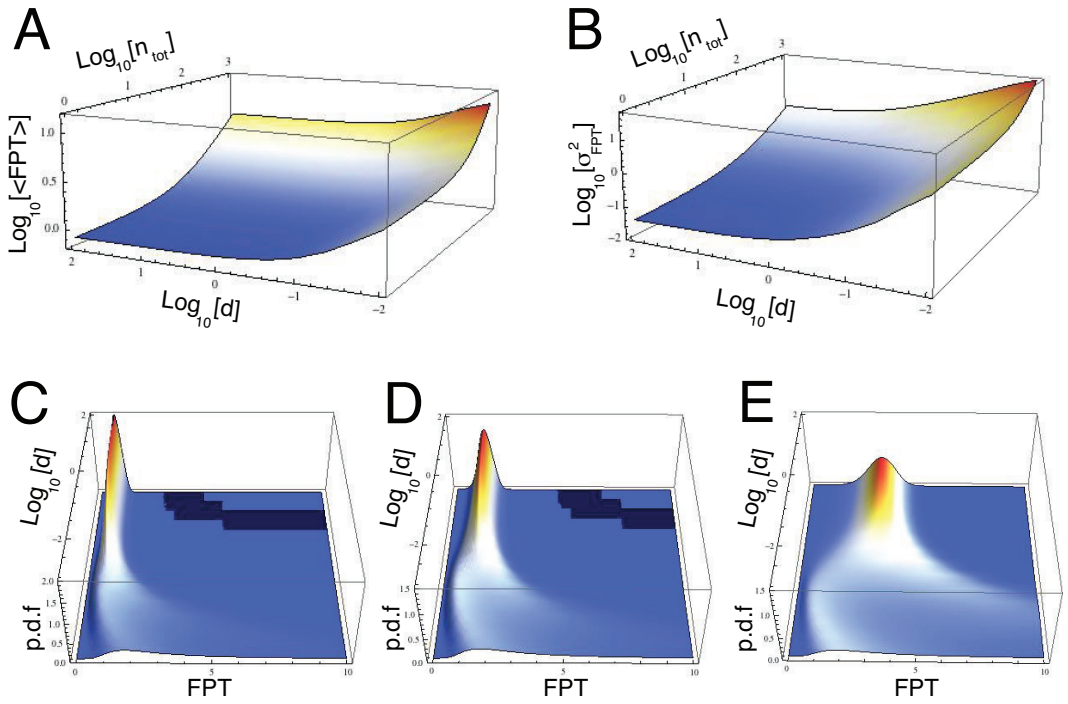

FIG. S9: Effect of diffusion on the kinetic response of srRNA regulation. Mean (A) and variance (B) of the first-passage time (FPT) as a function of the diffusion rate $d = D\ell/\Omega$ and of the number of auxiliary targets $n_{\rm tot}$. And probability distribution function (p.d.f) of the FPT as a function of the diffusion rate for $n_{\rm tot} = 0$ (C), 100 (D) and 500 (E). Results were obtained using the linear-noise approximation. Same fixed parameters as in Fig. S3.

1. Data set

We study the conservation of microRNA targets on a set of 87 microRNA families conserved among vertebrates (Dataset S1). The dataset, including sequences of seed regions and mature species for each microRNA in each of the 12 studied vertebrate genomes, was downloaded from the TargetScan web site (www.targetscan.org). The list of aligned 3'UTRs (corresponding to 30,887 genes) needed to search for target sites was downloaded from the same website.

As a control, we performed parallel analysis on a set of 1,000 "mock" microRNAs (Dataset S2). The mature sequences of the mock microRNAs were constructed by randomly picking one of the 87 natural mature sequences, and replacing the seed regions by randomly generated sequence having the same dinucleotide composition as the seed regions of the 87 natural microRNAs (Fig. S10).

2. Analysis

a. Defining target sites and conservation score

We used the TargetScan algorithm (implemented as perl script, available for download from the TargetScan website) to obtain the list of putative binding sites for each microRNA (real or mock) in the corresponding genome. The list of gene targets of a particular microRNA is defined as the list of all genes that have at least one putative binding sites of that microRNA in their 3'UTR. Fig. S11 shows that for all species the distributions of target numbers is similar for real and random seeds.

To investigate the conservation of the number of targets across species, we normalize the predicted number of targets $N_{i,s}$ per microRNA i in each species s by the total number of genes per species $N_{tot,s}$ in the alignment. This allows to correct for the loss of gene homologs as the evolutionary distance with human increases. The relative fluctuation of the number of targets for the microRNA i across species is then defined as the ratio between the standard deviation and the mean of the ensemble of values for $N_{i,s}/N_{tot,s}$ (s=human, chimpanzee,..., frog).

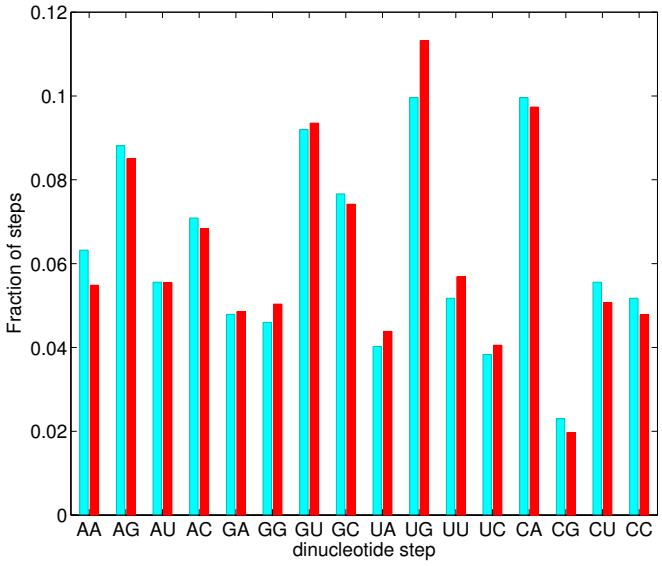

FIG. S10: Dinucleotide composition of the seed regions of the 87 natural microRNAs (blue bars) and of the 1,000 fake microRNAs (red bars).

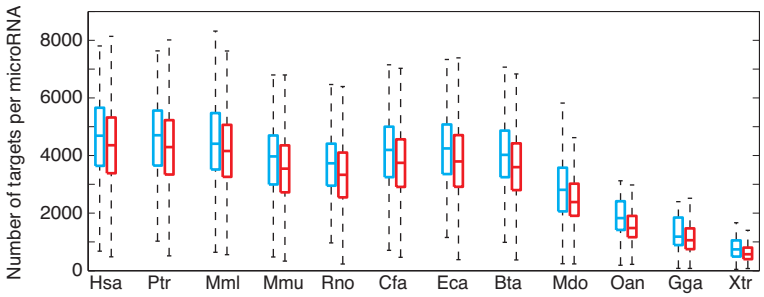

FIG. S11: Boxplots of the number of targets per microRNA in human (Hsa), chimpanzee (Ptr), rhesus (Mml), mouse (Mmu), rat (Rno), dog (Cfa), horse (Eca), cow (Bta), opossum (Mdo), platypus (Oan), chicken (Gga) and frog (Xtr), for natural (blue) and fake (red) microRNAs.

Taking the list of human targets as reference, we represent the predictions of Targetscan by a binary array $x^t_{i,s}$, where $x^t_{i,s} = 1$ if gene t is a target of microRNA i in human and in species s, and 0 otherwise. The observed frequency f^t_i of t for microRNA i is then defined as the mean value of $x^t_{i,s}$ across the species

$$f^t_i = \frac{1}{11} \sum_{s \neq \text{human}} x^t_{i,s}. \tag{E1}$$

We exclude human from the sum since $x^t_{i,s}$ is always equal to 1 in human by definition. To estimate the conservation of a human target in vertebrates, we define a conservation score C^t_i as the ratio between f^t_i and the background frequency f^t_0 ($C^t_i = f^t_i / f^t_0$) where $f^t_0 = (1/11) \sum_{s \neq \text{human}} \delta_{t,s}$ with $\delta_{t,s} = 1$ if gene t is present in species s, 0 otherwise. f^t_0 represents an upper bound for f^t_i. For example, $C^t_i = 1$ means that gene t is a target of microRNA i in every species where t is present. This target-dependent normalization of f^t_i permits unbiased comparisons between all the genes even if some are not present in every species.

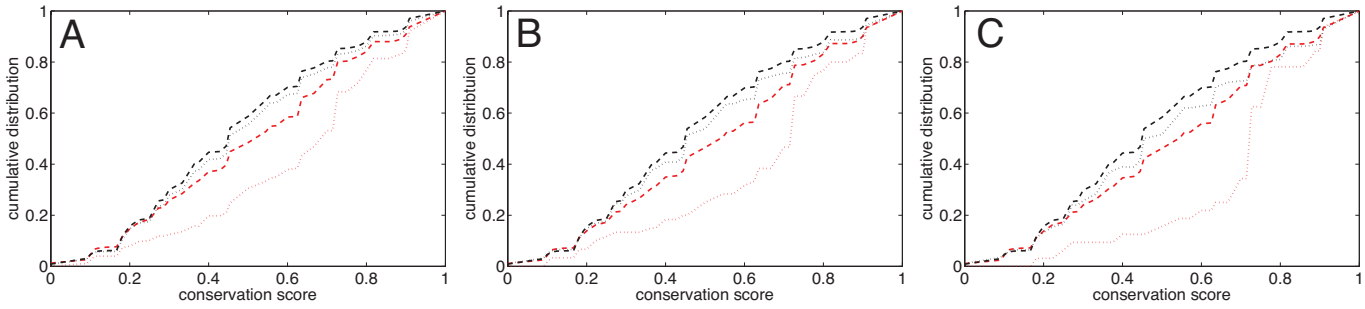

FIG. S12: Same as in Fig. 6B of the main text but for different definitions of the strong and weak ensembles. The limit (in $\log_2$ of the fold-change) is fixed to -0.4 (A), -0.8 (B) or -1 (C).

b. Conservation of weak and strong targets

To investigate the conservation of principal and auxiliary targets, we use published experimental data that measured global proteome response to transfection of a microRNA (miR-1, miR-124 and miR-181) in (human) HeLa cells [31]. For each transfected microRNA, we extract from the experimental dataset the protein level change of each of its target genes (as defined above), if it had been measured. Dataset S3 contains the list of these genes for miR-1 (382 genes), miR-124 (249 genes) and miR-181 (345 genes), as well as the corresponding fold-change in protein level (in $\log_2$ unit).

The distribution of fold-repression does not offer a natural separation between principal and auxiliary targets in HeLa cells. We therefore split the targets based on their fold-change into an subset of "strong targets", representing targets with a fold-repression below a given arbitrary limit ($\log_2$ of the fold-change ≤ -0.6) and the complementary subset of "weak targets". We choose this arbitrary limit in order to have enough statistics for the strong target ensemble. We verified that our conclusions do not depend on the choice of this limit (see Fig. S12).

For each gene t of a given subset (weak or strong), we estimate its conservation scores relatively to the transfected natural microRNAs, or to the random microRNAs (($\{C_i^t\}$ for $i \in$ {miR-1, miR-124, miR-181} or $i \in$ {random seeds}). Comparisons between the corresponding distributions of conservation scores allows to check the specificity of conservation properties of weak and strong targets of natural microRNAs.

Appendix F: Parameters used in the main text

Typical values for parameters are only known for bacterial srRNA pathways [1, 4]. Therefore, we choose to use these ranges of values to plot our figures. However, Fig. S2 shows that the effects described in the paper are robust over a wide range of parameters values which likely also includes the eukaryotic pathways.

1. Figure 2 A

Fixed parameters are (in min-1) $\alpha_m = 1$, $\beta_m = \beta_s = 0.1$, $\gamma = 4$, $\beta_p = 1/200$, $k = 0.1$, $k_a = k$, $k_d = 1$, $\beta_n = 0.1$, $\beta_c = 1/200$, $p_d = 1/20$.

2. Figures 2 B,C and 3

Fixed parameters are (in min-1) $\alpha_m = 1$, $\beta_m = \beta_s = 0.1$, $\gamma = 4$, $\beta_p = 1/200$, $\beta_n = 0.1$, $\beta_c = 0$, $k^0 = 0.1$, $k_a^0 = 0.1$, $k_d^0 = 1$, $d = \infty$ (A) or 0.05 (B).

3. Figures 4 and 5

Fixed parameters are (in min-1) $\alpha_m = 1$, $\alpha_s = 10$, $\beta_m = \beta_s = 0.1$, $k = k_a = 0.1$, $k_d = 1$.

---

[1] Levine, E., Zhang, Z., Kuhlman, T., and Hwa, T. Quantitative characteristics of gene regulation by small rna. *PLoS Biol* **5**(9), e229, Sep (2007).

[2] Mitarai, N., Andersson, A. M. C., Krishna, S., Semsey, S., and Sneppen, K. Efficient degradation and expression prioritization with small rnas. *Phys Biol* **4**(3), 164–171, Sep (2007).

[3] Mitarai, N., Benjamin, J.-A. M., Krishna, S., Semsey, S., Csiszovszki, Z., Massé, E., and Sneppen, K. Dynamic features of gene expression control by small regulatory rnas. *Proc Natl Acad Sci U S A* **106**(26), 10655–10659, Jun (2009).

[4] Shimoni, Y., Friedlander, G., Hetzroni, G., Niv, G., Altuvia, S., Biham, O., and Margalit, H. Regulation of gene expression by small non-coding rnas: a quantitative view. *Mol Syst Biol* **3**, 138 (2007).

[5] Jost, D., Nowojewski, A., and Levine, E. Small rna biology is systems biology. *BMB Rep* **44**(1), 11–21, Jan (2011).

[6] Mehta, P., Goyal, S., and Wingreen, N. S. A quantitative comparison of srna-based and protein-based gene regulation. *Mol Syst Biol* **4**, 221 (2008).

[7] Gottesman, S. Micros for microbes: non-coding regulatory rnas in bacteria. *Trends Genet* **21**(7), 399–404, Jul (2005).

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

[9] van Kampen, N. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, (2001).

[10] Chubb, J. R., Trcek, T., Shenoy, S. M., and Singer, R. H. Transcriptional pulsing of a developmental gene. *Curr Biol* **16**(10), 1018–1025, May (2006).

[11] So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., and Golding, I. General properties of transcriptional time series in escherichia coli. *Nat Genet* **43**(6), 554–560, Jun (2011).

[12] Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**(6), 1025–1036, Dec (2005).

[13] Paré, A., Lemons, D., Kosman, D., Beaver, W., Freund, Y., and McGinnis, W. Visualization of individual scr mrnas during drosophila embryogenesis yields evidence for transcriptional bursting. *Curr Biol* **19**(23), 2037–2042, Dec (2009).

[14] Bialek, W. and Setayeshgar, S. Physical limits to biochemical signaling. *Proc Natl Acad Sci U S A* **102**(29), 10040–10045, Jul (2005).

[15] Berg, O. G. On diffusion-controlled dissociation. *Chemical Physics* **31**(1), 47 – 57 (1978).

[16] Berg, H. C. and Purcell, E. M. Physics of chemoreception. *Biophys J* **20**(2), 193–219, Nov (1977).

[17] Tkacik, G. and Bialek, W. Diffusion, dimensionality, and noise in transcriptional regulation. *Phys Rev E Stat Nonlin Soft Matter Phys* **79**(5 Pt 1), 051901, May (2009).

[18] Thattai, M. and van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* **98**(15), 8614–8619, Jul (2001).

[19] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**(5584), 1183–1186, Aug (2002).

[20] Paulsson, J. Summing up the noise in gene networks. *Nature* **427**(6973), 415–418, Jan (2004).

[21] Elf, J., Paulsson, J., Berg, O. G., and Ehrenberg, M. Near-critical phenomena in intracellular metabolite pools. *Biophys J* **84**(1), 154–170, Jan (2003).

[22] Levine, E., Huang, M., Huang, Y., Kuhlman, T., Zhang, Z., and Hwa, T. On noise and silence in gene regulation by small rna. (2008).

[23] Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340–2361 (1977).

[24] Hao, Y., Xu, L., and Shi, H. Theoretical analysis of catalytic-srna-mediated gene silencing. *J. Mol. Biol.* **406**, 195–204 (2011).

[25] Tkacik, G., Callan, C. G., and Bialek, W. Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci U S A* **105**(34), 12265–12270, Aug (2008).

[26] Tkacik, G., Walczak, A. M., and Bialek, W. Optimizing information flow in small genetic networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **80**(3 Pt 1), 031920, Sep (2009).

[27] Walczak, A. M., Tkacik, G., and Bialek, W. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Phys Rev E Stat Nonlin Soft Matter Phys* **81**(4 Pt 1), 041905, Apr (2010).

[28] Mugler, A., Grinshpun, B., Franks, R., and Wiggins, C. H. Statistical method for revealing form-function relations in biological networks. *Proc Natl Acad Sci U S A* **108**(2), 446–451, Jan (2011).

[29] Cover, T. and Thomas, J. *Elements of Information Theory*. John Wiley and Sons, (1991).

[30] Flynt, A. S. and Lai, E. C. Biological principles of microrna-mediated regulation: shared themes amid diversity. *Nat Rev Genet* **9**(11), 831–842, Nov (2008).

[31] Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. The impact of microrna on protein output. *Nature* **455**(7209), 64–71, Sep (2008).