Feed-forward regulation adaptively evolves via dynamics rather than topology when there is intrinsic noise

Short title: Adaptive evolution of feed-forward regulation

Kun Xiong[1], Alex K. Lancaster[2], Mark L. Siegal[3], Joanna Masel[4*]

[1] Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, United States of America

[2] Ronin Institute, Montclair, New Jersey, United States of America

[3] Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York, United States of America

[4] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona, United States of America

* Corresponding author

Email: masel@email.arizona.edu

Abstract

We develop a null model of the evolution of transcriptional regulatory networks, and use it to support an adaptive origin for a canonical “motif”, a 3-node feed-forward loop (FFL) hypothesized to filter out short spurious signals by integrating information from a fast and a slow pathway. Our mutational model captures the intrinsically high prevalence of weak affinity transcription factor binding sites. We also capture stochasticity and delays in gene expression that distort external signals and intrinsically generate noise. Functional FFLs evolve readily under selection for the hypothesized function, but not in negative controls. Interestingly, a 4-node “diamond” motif also emerged as a short spurious signal filter. The diamond uses expression dynamics rather than path length to provide fast and slow pathways. When there is no external spurious signal to filter out, but only internally generated noise, only the diamond and not the FFL evolves.

Introduction

Transcriptional regulatory networks (TRNs) are integral to development and physiology, and underlie all complex traits. An intriguing finding about TRNs is that certain topological "motifs" of interconnected transcription factors (TFs) are over-represented relative to random re-wirings that preserve the frequency distribution of connections. The significance of this finding remains open to debate.

The canonical example is the feed-forward loop (FFL), in which TF A regulates a target C both directly, and indirectly via TF B, and no regulatory connections exist in the opposite direction[1-3]. Each of the three regulatory interactions in a FFL can be either activating or repressing, so there are eight distinct kinds of FFLs (**Fig. S1**)[4]. Given the eight frequencies expected from the ratio of activators to repressors, two of these kinds of FFLs are significantly over-represented[4]. In this paper, we focus on one of these two over-represented types, namely the type 1 coherent FFL (C1-FFL), in which all three links are activating rather than repressing (**Fig. S1**, top left). C1-FFL motifs are an active part of systems biology research today, e.g. they are used to infer the function of specific regulatory pathways[5, 6].

The over-representation of FFLs in observed TRNs is normally explained in terms of selection favoring a function of FFLs. Specifically, the most common adaptive hypothesis is that cells often benefit from ignoring short-lived signals and responding only to durable signals[3, 4, 7]. Evidence that C1-FFLs can perform this function comes from the behavior both of theoretical models[4] and of *in vivo* gene circuits[7]. A C1-FFL can achieve this function when its regulatory logic is that of an "AND" gate, i.e. both the direct path from A to C and the indirect path from A to B to C must be

activated before the response is triggered. In this case, the response will only be triggered if, by the time the signal trickles through the longer path, it is still active on the shorter path as well. This yields a response to long-lived signals but not short-lived signals.

However, just because a behavior is observed, we cannot conclude that the behavior is a historical consequence of past selection favoring that behavior[8, 9]. The explanatory power of this adaptive hypothesis of filtering out short-lived and spurious signals needs to be compared to that of alternative, non-adaptive hypotheses[10]. The over-representation of C1-FFLs might be a byproduct of some other behavior that was the true target of selection[11]. Alternatively, it might be an intrinsic property of TRNs generated by mutational processes – gene duplication patterns have been found to enrich for FFLs in general[12], although not yet C1-FFLs in particular. Adaptationist claims about TRN organization have been accused of being just-so stories, with adaptive hypotheses still in need of testing against an appropriate null model of network evolution[13-23].

Here we develop such a computational null model of TRN evolution, and apply it to the case of C1-FFL over-representation. We include sufficient realism in our model of cis-regulatory evolution to capture the non-adaptive effects of mutation in shaping TRNs. In particular, we consider "weak" TF binding sites (TFBSs) that can easily appear *de novo* by chance alone, and from there be selected to bind a TF more strongly, as well as simulating mutations that duplicate and delete genes.

We also capture the stochasticity of gene expression, which causes the number of mRNAs and hence proteins to fluctuate[24, 25]. This is important, because demand for spurious signal filtering

and hence C1-FFL function may arise not just from external signals, but also from internal fluctuations. Stochasticity in gene expression also shapes how external spurious signals are propagated. Stochasticity is a constraint on what TRNs can achieve, but it can also be adaptively co-opted in evolution[26]; either way, it might underlie the evolution of certain motifs. Most other computational models of TRN evolution that consider gene expression as the major phenotype do not simulate stochasticity in gene expression (but see three notable exceptions[27-29]).

Here we ask whether AND-gated C1-FFLs evolve as a response to selection for filtering out short and spurious external signals. Our new model allows us to compare the frequencies of network motifs arising in the presence of this hypothesized evolutionary cause to motif frequencies arising under non-adaptive control simulations, i.e. evolution under conditions that lack short spurious external signals while controlling both for mutational biases and for less specific forms of selection. We also ask whether other network motifs evolve to filter out short spurious signals, and if so, whether different conditions favor the appearance of different motifs during evolution.

Model overview

We simulate the dynamics of TRNs as the TFs activate and repress one another's transcription over developmental time, to generate gene expression phenotypes on which selection then acts over longer evolutionary timescales. For each moment in developmental time, we simulate the numbers of nuclear and cytoplasmic mRNAs in a cell, the protein concentrations, and the chromatin state of each gene in a haploid genome. Transitions between three possible chromatin states -- Repressed, Intermediate, and Active -- are a stochastic function of TF

104 binding, and transcription initiation from the Active state is also stochastic. An overview of the
105 model is shown in **Fig. 1**. The pattern of TF binding affects chromatin, which affects transcription
106 rates, eventually feeding back to affect the concentration of TFs and hence their binding. The
107 genotype is specified by a set of cis-regulatory sequences that contain TFBSs to which TFs may
108 bind, by which consensus sequence each TF recognizes and with what affinity, and by 5 gene-
109 specific parameters that control gene expression as a function of TF binding: mean duration of
110 transcriptional bursts, mRNA degradation, protein production, and protein degradation rates,
111 and gene length (which affects delays in transcription and translation). An external signal (**Fig.**
112 **1A red**) is treated like another TF, and the concentration of an effector gene (**Fig. 1A blue**) in
113 response is a primary determinant of fitness, combined with a cost associated with gene
114 expression (**Fig. 1B**). Mutants replace resident genotypes as a function of the difference in
115 estimated fitness (**Fig. 1C**). Parameter values, taken as far as possible from *Saccharomyces*
116 *cerevisiae*, are summarized in **Table S1**. Source code in C is available at
117 https://github.com/MaselLab/network-evolution-simulator.

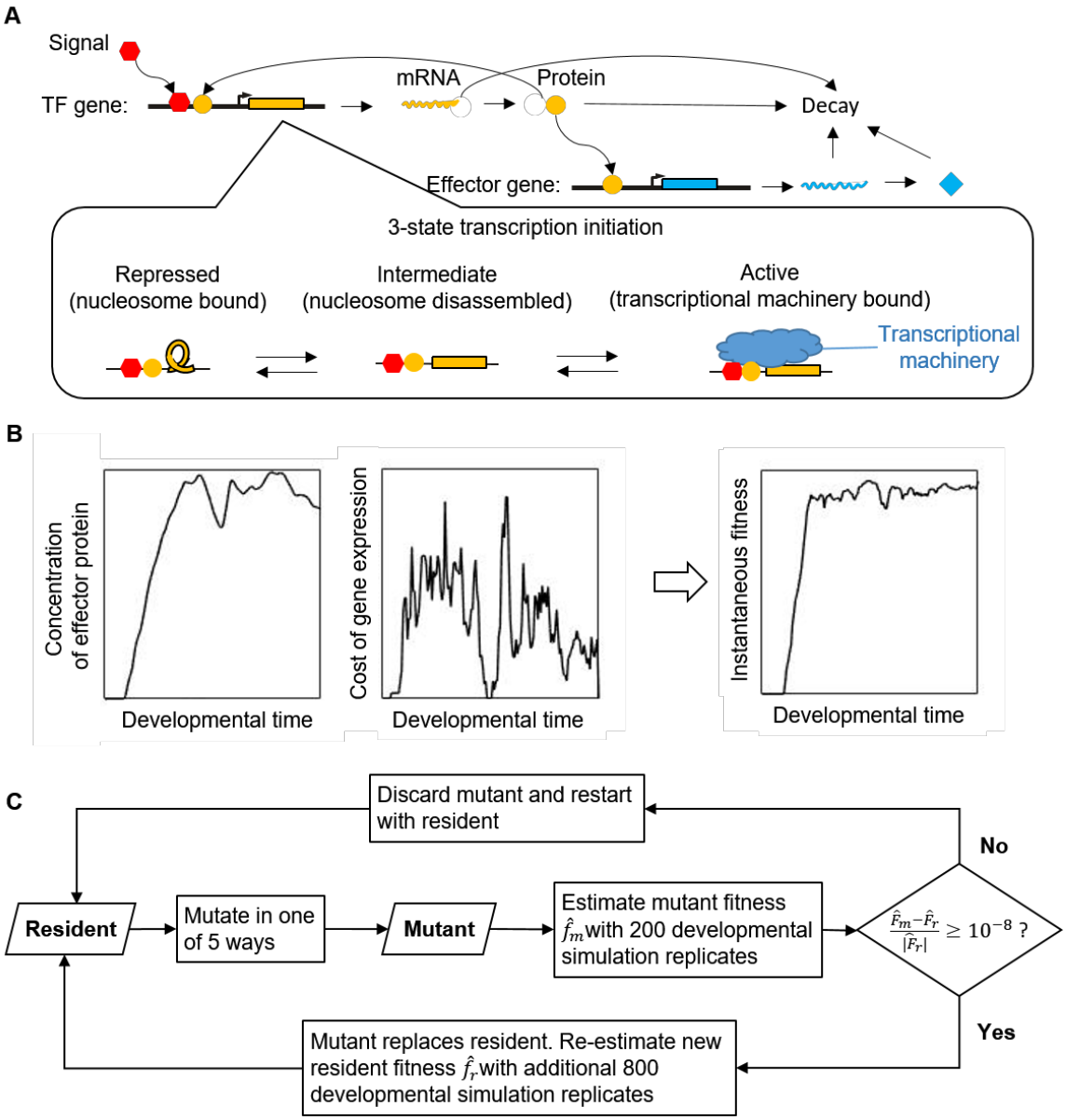


Figure 1. Overview of the model. (A) Simulation of gene expression phenotypes. We show a simple TRN with one TF (yellow) and one effector gene (blue), with arrows for major biological processes simulated in the model. **(B) Phenotype-fitness relationship.** Fitness is primarily determined by the concentration of an effector protein (here shown as beneficial as in Eq. 1, but potentially deleterious in a different environment as in Eq. 2), with a secondary component coming from the cost of gene expression (proportional to the rate of protein production), combined to give an instantaneous fitness at each moment in developmental time. **(C) Evolutionary simulation.** A single resident genotype is replaced when a mutant's estimated fitness is high enough. Stochastic gene expression adds uncertainty to the estimated fitness, allowing less fit mutants to occasionally replace the resident, capturing the flavor of genetic drift.

Transcription factor binding

Transcription of each gene is controlled by TFBSs present within a 150-bp cis-regulatory region. When bound, a TF occupies a stretch of DNA 14 bp long. In the center of this stretch, each TF recognizes an 8-bp consensus sequence, and binds to it with a TF-specific (and mutable) dissociation constant $K_d(0)$. TFs also bind somewhat specifically when there are one or two mismatches, with $K_d(1)$ and $K_d(2)$ values calculated from $K_d(0)$ according to a model of approximately additive binding energy per base pair. With three mismatches, binding occurs at the same background affinity as to any 14 bp stretch of DNA. We model competition between a smaller number of specific higher-affinity binding sites and the much larger number of non-specific binding sites, the latter corresponding to the total amount of nucleosome-free sequence in *S. cerevisiae*. Competition with non-specific binding can be approximated by using an effective dissociation constant $\widehat{K}_d = 10K_d$. See Supplementary Text Section 1 for justification and details of these model choices.

Each TF is either an activator or a repressor. The algorithm for obtaining the probability distribution for *A* activators and *R* repressors being bound to a given cis-regulatory region at a given moment in developmental time is described in Supplementary Text Section 2.

Transcriptional regulation

Activation of the effector gene requires at least two TFBSs to be occupied by activators – not necessarily different activators. The requirement for two activators makes the effector gene capable of evolving an AND-gate via a configuration of TFBSs in which the only way to have two

TFs bound is for them to be different TFs (**Fig. 2**). All other genes are AND-gate-incapable, meaning that their activation requires only one TFBS to be occupied by an activator. P_A denotes the probability of having at least one activator bound for an AND-gate-incapable gene, or two for an AND-gate-capable gene. P_R denotes the probability of having at least one repressor bound.

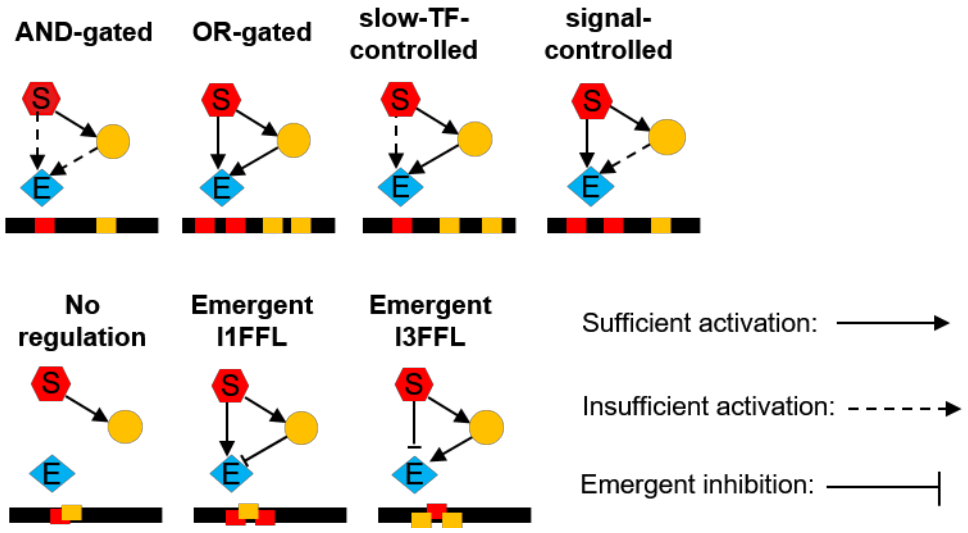


Figure 2. The numbers of TFBSs, and any hindrance between them, determine the regulatory logic of effector expression. We use the pattern of TFBSs (red and yellow bars along black cis-regulatory sequences) to classify the regulatory logic of the effector gene. C1-FFLs are classified first by whether or not they are capable of simultaneously binding the signal and the TF (top vs bottom). Further classification is based on whether either the signal or the TF has multiple non-overlapping TFBSs, allowing it to activate the effector without help from the other (solid arrow). The three subtypes on the bottom (where the signal and TF cannot bind simultaneously) are rarely seen; they are unless otherwise indicated included in "Any logic" and "non-AND-gated" tallies, but are not analyzed separately. Two of them involve emergent repression, creating "incoherent" feed-forward loops (see **Fig. S1** for full FFL naming scheme). Emergent repression occurs when the binding of one activator to its only TFBS prevents the other activator from binding to either of its two TFBSs, hence preventing simultaneous binding of two activators.

Noise in yeast gene expression is well described by a two step process of transcriptional activation[30, 31], e.g. nucleosome disassembly followed by transcription machinery assembly. We denote the three corresponding possible states of the transcription start site as Repressed, Intermediate, and Active (**Fig. 1A**). Transitions between the states depend on the numbers of activator and repressor TFs bound (e.g. via recruitment of histone-modifying enzymes[32, 33]). We make conversion from Repressed to Intermediate a linear function of P_A, ranging from the background rate 0.15 min^{-1} of histone acetylation[34] (presumed to be followed by nucleosome disassembly), to the rate of nucleosome disassembly 0.92 min^{-1} for the constitutively active PHO5 promoter[30]:

$$r_{Rep\_to\_Int} = 0.92P_A + 0.15(1 - P_A).$$

We make conversion from Intermediate to Repressed a linear function of P_R, ranging from a background histone de-acetylation rate of 0.67 min^{-1} [34], up to a maximum of 4.11 min^{-1} (the latter chosen so as to keep a similar maximum:basal rate ratio as that of $r_{Rep\_to\_Int}$):

$$r_{Int\_to\_Rep} = 4.11P_R + 0.67(1 - P_R).$$

We assume that repressors disrupt the assembly of transcription machinery[35] to such a degree that conversion from Intermediate to Active does not occur if even a single repressor is bound. In the absence of repressors, activators facilitate the assembly of transcription machinery[36]. Brown et al.[30] reported that the rate of transcription machinery assembly is 3.3 min^{-1} for a constitutively active PHO5 promoter, and 0.025 min^{-1} when the Pho4 activator of the PHO5 promoter is knocked out. We use this range to set

196

197 $$r_{Int\_to\_Act} = 3.3P_{A\_no\_R} + 0.025P_{notA\_no\_R}$$

198

199 where $P_{A\_no\_R}$ is the probability of having no repressors and either one (for an AND-gate-
200 incapable gene) or two (for an AND-gate-capable gene) activators bound, and $P_{notA\_no\_R}$ is the
201 probability of having no TFs bound (for AND-gate-incapable genes) or having no repressors and
202 not more than one activator bound (for AND-gate-capable genes).

203

204 The promoter sequence not only determines which specific TFBSs are present, but also
205 influences non-specific components of the transcriptional machinery[37, 38]. We capture this via
206 gene-specific but TF-binding-independent rates $r_{Act\_to\_Int}$ with which the machinery disassembles
207 and a burst of transcription ends. In other words, we let TF binding regulate the frequency of
208 "bursts" of transcription, while other properties of the cis-regulatory region regulate their
209 duration. For example, the yeast transcription factor Pho4 regulates the frequency but not
210 duration of bursts of PHO5 expression, by regulating the rates of nucleosome removal and of
211 transition to but not from a transcriptionally active state[30]. Parameterization of $r_{Act\_to\_Int}$ is
212 described in Supplementary Text Section 3.

213

214 **mRNA and protein dynamics**

215

216 All genes in the Active state initiate new transcripts stochastically at rate $r_{max\_transc\_init}$ = 6.75
217 mRNA/min[30], while the time for completing transcription depends on gene length (see
218 Supplementary Text Section 4 for parameterization of gene length and associated delay times).
219 We model a second delay before a newly completed transcript produces the first protein, which

220 we assume is dominated by translation initiation (length-independent) plus elongation (length-
221 dependent) and not splicing or mRNA export (see Supplementary Text Section 5). After the
222 second delay, we model protein production as continuous at a gene-specific rate $r_{protein\_syn}$ (see
223 Supplementary Text Section 5).
224
225 Protein transport into the nucleus is rapid[39] and is approximated as instantaneous and
226 complete, so that the newly produced protein molecules immediately increase the probability of
227 TF binding. Each gene has its own mRNA and protein decay rates, initialized from distributions
228 taken from data (see Supplementary Text Section 6).
229
230 All the rates regarding transcription and translation are listed in **Table S1**, including distributions
231 estimated from data, and hard bounds imposed to prevent unrealistic values arising during
232 evolutionary simulations.
233
234 **Developmental simulation**
235
236 Our algorithm is part stochastic, part deterministic. We use a Gillespie algorithm[40] to simulate
237 stochastic transitions between Repressed, Intermediate, and Active chromatin states, and to
238 simulate transcription initiation and mRNA decay events. Fixed (i.e. deterministic) delay times
239 are simulated between transcription initiation and completion, and between transcript
240 completion and the production of the first protein. Protein production and degradation are
241 described deterministically with ODEs, and updated frequently in order to recalculate TF
242 concentrations and hence chromatin transition rates. Details of our simulation algorithm are

given in the Supplementary Text Section 7. We initialize developmental simulations with no mRNA or protein, and all genes in the Repressed state.

Selection

Filtering out short spurious signals is a special case of signal recognition. In environment 1, expressing the effector is beneficial, and in environment 2 it is deleterious. We select for TRNs that take information from the signal and correctly decide whether to express the effector. Fitness is a weighted average across separate developmental simulations in the two environments, one with a signal and one without. In both cases, we begin each developmental simulation with no signal. To ensure that gene expression changes in response to the signal, and not via an internal timer, we simulate a burn-in phase with duration drawn from an exponential distributed truncated at 30 minutes, with un-truncated mean of 10 minutes. By having no fitness effects of gene expression during the burn-in, we eliminate a significant source of noise in fitness estimation due to variable burn-in duration. In our control condition, at the end of the burn-in, the signal suddenly switches to a constant "on" level in environment 1, and remains off in environment 2. In our test condition (**Fig. 3**), the signal is turned on in the same way in environment 1 but is also briefly turned on (for the first 10 minutes after the burn-in) in environment 2 – selection is to ignore this short spurious signal. The signal is treated as though it were an activating TF whose concentration is controlled externally, with an "off" concentration of zero and an "on" concentration of 1,000 molecules per cell, which is the typical per-cell number of a yeast TF[41].

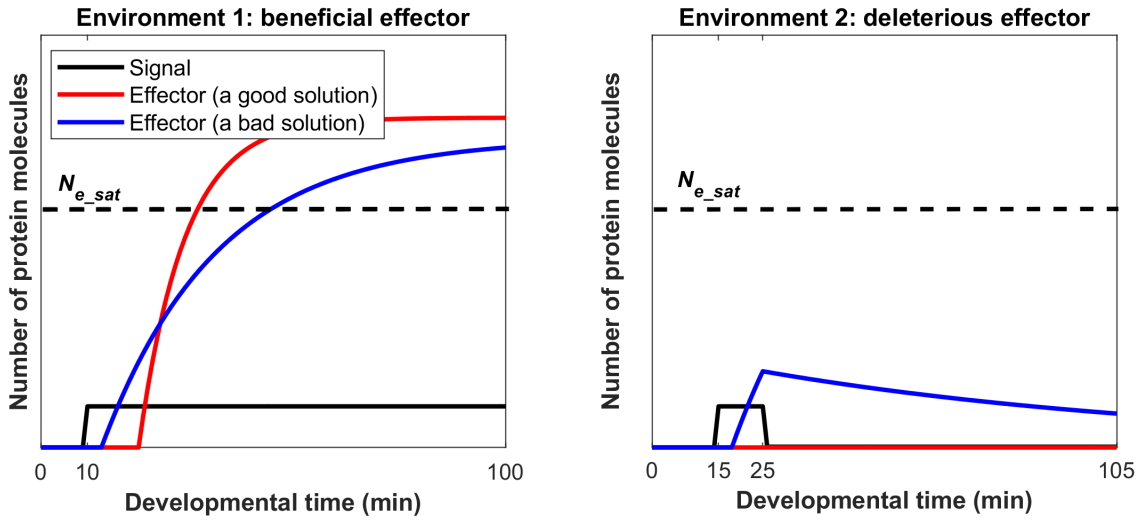


Figure 3. Selection for filtering out short spurious signals. Each selection condition averages fitness across simulations in two environments. The effectors have different fitness effects in the two environments, and the signal also behaves differently in the two environments. Simulations begin with zero mRNA and protein, and all genes at the Repressed state. Each simulation is burned in for a randomly sampled length of time in the absence of signal (shown here as 10 minutes in environment 1, and 15 minutes in environment 2), and continues for another 90 minutes after the burn-in. The signal is shown in black. Red illustrates a good solution in which the effector responds appropriately in each of the environments, while blue shows an inferior solution. See **Fig. S2** for examples of high-fitness and low-fitness evolved phenotypes, where, as shown in this schematic, high-fitness solutions have longer delays followed by more rapid responses thereafter.

We make fitness quantitative in terms of a "benefit" $B(t)$ as a function of the amount of effector protein $N_e(t)$ at developmental time t. Our motivation is a scenario in which the effector protein is responsible for directing resources from a metabolic program favored in environment

2 to a metabolic program favored in environment 1. In environment 1, where the effector produces benefits,

$$B(t) = \begin{cases} b_{max}\frac{N_e(t)}{N_{e\_sat}}, & N_e(t) < N_{e\_sat} \\ b_{max}, & N_e(t) \geq N_{e\_sat} \end{cases}, \quad (1)$$

where b_{max} is the maximum benefit if all resources were redirected, and N_{e_sat} is the minimum amount of effector protein needed to achieve this. Similarly, in environment 2

$$B(t) = \begin{cases} b_{max} - b_{max}\frac{N_e(t)}{N_{e_{sat}}}, & N_e(t) < N_{e_{sat}} \\ 0, & N_e(t) \geq N_{e_{sat}} \end{cases}. \quad (2)$$

We set N_{e_sat} to 10,000 molecules, which is about the average number of molecules of a metabolism-associated protein per cell in yeast[41]. Without loss of generality given that fitness is relative, we set b_{max} to 1.

A second contribution to fitness comes from the cost of gene expression $C(t)$ (**Fig. 1B, middle**). We make this cost proportional to the total protein production rate. We estimate a fitness cost of gene expression of 2×10^{-6} per protein molecule translated per minute, based on the cost of expressing a non-toxic protein in yeast[42] (see Supplementary Text Section 7 for details).

We simulate gene expression for 90 minutes plus the duration of the burn-in (**Fig. 3**). A "cellular fitness" in a given environment is calculated as the average instantaneous fitness $B(t)$-$C(t)$ over the 90 minutes. We consider environment 2 to be twice as common as environment 1 (a

304 "signal" should be for an uncommon event rather than the default), and take the corresponding
305 weighted average.
306
307 **Evolutionary simulation**
308
309 We simulate a novel version of origin-fixation (weak-mutation-strong-selection) evolutionary
310 dynamics, i.e. the population contains only one resident genotype at any time, and mutant
311 genotypes are either rejected or chosen to be the next resident (Fig. 1C). Despite the fact that
312 our mutant acceptance rule (see below) was chosen to maximize computational efficiency, our
313 model usually takes 10 CPUs 1-3 days to complete an evolutionary simulation; modeling a
314 heterogeneous population is clearly out of the question. We note that genetic homogeneity
315 entails ignoring some important population genetic phenomena. First, if there were
316 recombination, heterogeneity would favor mutations that combine well with a range of other
317 genotypes. Second, clonal interference would shift evolution toward beneficial mutations of
318 larger effect[43] (an effect we can mimic by modifying the value 10^{-8} in the equation below). Third,
319 polymorphic populations would evolve mutational robustness[44]. None of these three effects
320 seems *a priori* likely to change our conclusions, although the possibility cannot be ruled out.
321
322 Estimators $\hat{F}$ of genotype fitness are averages of the cellular fitness values of 200
323 developmental replicates per environment in the case of the mutant, plus an additional 800
324 should it be chosen to be the next resident. The mutant replaces the resident if
325
326 $$\frac{\hat{F}_{mutant} - \hat{F}_{resident}}{|\hat{F}_{resident}|} \geq 10^{-8}.$$
327

328 This differs from Kimura's[45] equation for fixation probability, but captures the flavor of genetic
329 drift. Genetic drift allows slightly deleterious mutations to occasionally fix, and beneficial
330 mutations to sometimes fail to do so, even as the probability of fixation is monotonic with
331 fitness. This is also achieved by our procedure, because of stochastic deviations of $\hat{F}$ from true
332 genotype fitness. The number of developmental replicates captures the flavor of effective
333 population size.
334
335 Note that it is possible, especially at the beginning of an evolutionary simulation, for relative
336 fitness to be paradoxically negative. This occurs when a randomly initialized genotype does not
337 express the effector (garnering no fitness benefit), but does express other genes (accruing a cost
338 of expression); this combination makes fitness negative. In this rare case, for simplicity, we use
339 the absolute value of $\hat{F}$ on the denominator.
340
341 If 2,000 successive mutants are all rejected, the simulation is terminated; upon inspection, we
342 found that these resident genotypes had evolved to not express the effector in either
343 environment. We refer to each change in resident genotype as an evolutionary step. We stop
344 the simulation after 50,000 evolutionary steps; at this time, most replicate simulations seem to
345 have reached a fitness plateau (**Fig. S3**); we analyze all replicates except those terminated early.
346 To reduce the frequency of early termination in the case where the signal was not allowed to
347 directly regulate the effector, we used a burn-in phase selecting on a more accessible
348 intermediate phenotype (see Supplementary Text Section 10). In this case, burn-in occurred for
349 1,000 evolutionary steps, followed by the usual 50,000 evolutionary steps with selection for the
350 phenotype of interest (**Fig. S3**, right panels). Most replicates found a stable fitness plateau

within 10,000 evolutionary steps, although some replicates were temporarily trapped at a low fitness plateau (**Fig. S3**).

Genotype Initialization

We initialize genotypes with 3 activator genes, 3 repressor genes, and 1 effector gene. Cis-regulatory sequences and consensus binding sequences contain As, Cs, Gs, and Ts sampled with equal probability. Rate constants associated with the expression of each gene are sampled from the distributions summarized in **Table S1**.

Mutation

A genotype is subjected to 5 broad classes of mutation, at rates summarized in **Table S2** and justified in Supplementary Text Section 9. First are single nucleotide substitutions in the cis-regulatory sequence; the resident nucleotide mutates into one of the other three types of nucleotides with equal probability. Second are single nucleotide changes to the consensus binding sequence of a TF, with the resident nucleotide mutated into recognizing one of the other three types with equal probability. Both of these types of mutation can affect the number and strength of TFBSs.

Third are gene duplications or deletions. Because computational cost scales steeply (and non-linearly) with network size, we do not allow effector genes to duplicate once there are 5 copies, nor TF genes to duplicate once the total number of TF gene copies is 19. We also do not allow the signal, the last effector gene, nor the last TF gene to be deleted.

Fourth are mutations to gene-specific expression parameters. Most of these (L, $r_{Act\_to\_Int}$, $r_{protein\_syn}$, r_{mRNA_deg}, and $r_{protein\_deg}$) apply to both TFs and effector genes, while mutations to the gene-specific values of $K_d(0)$ apply only to TFs. Each mutation to L increases or decreases it by 1 codon, with equal probability unless L is at the upper or lower bound. Effect sizes of mutations to the other five parameters are modeled in such a way that mutation would maintain specified log-normal stationary distributions for these values, in the absence of selection or arbitrary bounds (see Supplementary Text Section 9 for details). Upper and lower bounds (Supplementary Text Section 9) are used to ensure that selection never drives these parameters to unrealistic values.

Fifth is conversion of a TF from being an activator to being a repressor, and vice versa. The signal is always an activator, and does not evolve.

Importantly, this scheme allows for divergence following gene duplication. When duplicates differ due only to mutations of class 4, i.e. protein function is unchanged, we refer to them as "copies" of the same gene, encoding "protein variants". Mutations in classes 2 and 5 can create a new protein.

Table S3 summarizes the tendencies of different mutation types to be accepted, and to contribute to evolution. Acceptance rates are high, indicative of substantial nearly neutral evolution, in which slightly deleterious mutations are fixed and subsequently compensated for.

Results

Functional AND-gated C1-FFLs evolve readily under selection for filtering out a short spurious signal

We begin by simulating the easiest case we can devise to allow the evolution of C1-FFLs for their purported function of filtering out short spurious signals. The signal is allowed to act directly on the effector, after which all that needs to evolve is a single activating TF between the two, as well as AND-logic for the effector (**Fig. 2**, top left; see "Transcriptional regulation" in the Model Overview for how AND-logic evolution is handled). We score network motifs at the end of a set period of evolution (see Supplemental Text Section 11 for details), further classifying evolved C1-FFLs into subtypes based on the presence of non-overlapping TFBSs (**Fig. 2**). The adaptive hypothesis predicts the evolution of the C1-FFL subtype with AND-regulatory logic, which requires the effector to be stimulated both by the signal and by the slow TF. While all evolutionary replicates show large increases in fitness, the extent of improvement varies dramatically, indicating whether or not the replicate was successful at evolving the phenotype of interest rather than becoming stuck at an alternative locally optimal phenotype (**Fig. 4A**). AND-gated C1-FFLs frequently evolve in replicates that reach high fitness outcomes, but not replicates that reach lower fitness (**Fig. 4B**).

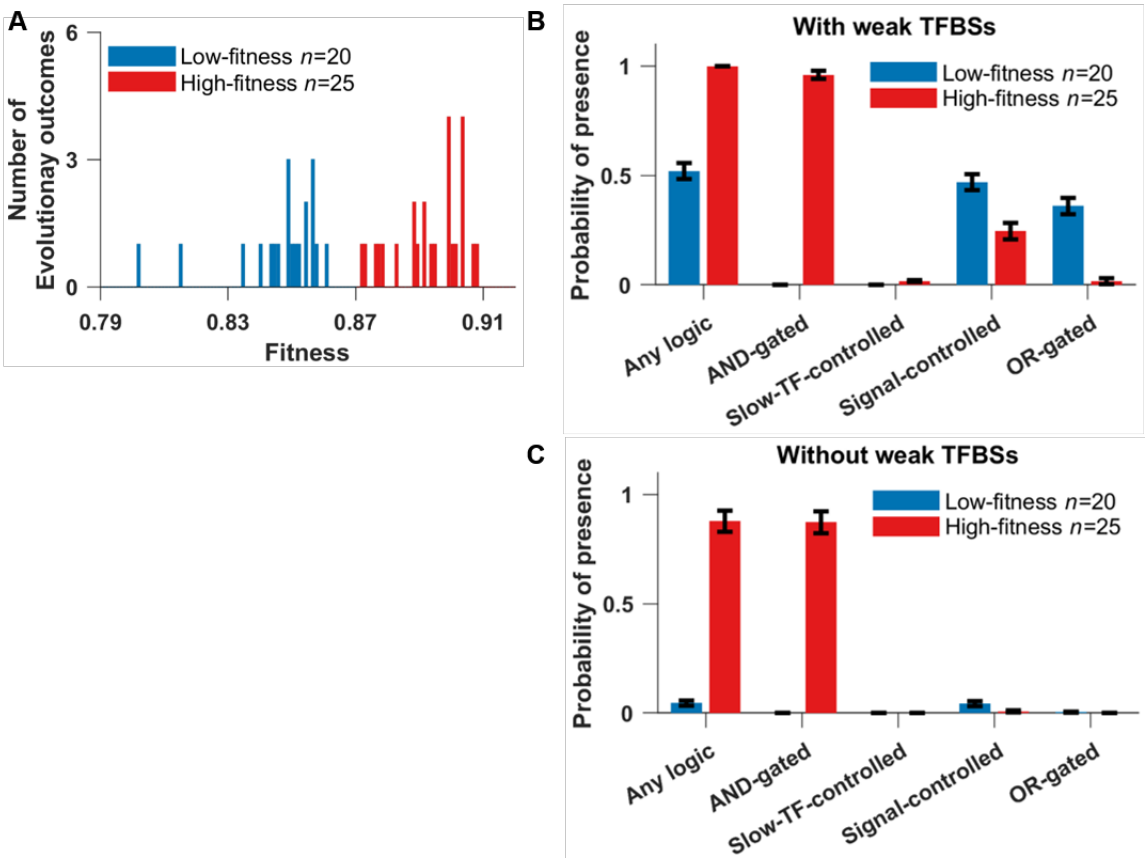


Figure 4. AND-gated C1-FFLs are associated with a successful response to selection for filtering out short spurious signals. (A) Distribution of fitness outcomes across replicate simulations, calculated as the average fitness over the last 10,000 steps of the evolutionary simulation. We divide genotypes into a low-fitness group (blue) and a high-fitness group (red) using as a threshold an observed gap in the distribution. **(B)** High fitness replicates are characterized by the presence of an AND-gated C1-FFL. "Any logic" counts the presence of any of the seven subtypes shown in **Fig. 2B**. Because one TRN can contain multiple C1-FFLs of different subtypes, each of which are scored, the sum of the occurrences of all seven subtypes will generally be more than "Any logic". See Supplementary Text Section 11 for details on the calculation of the y-axis. **(C)** The over-representation of AND-gated C1-FFLs becomes even more pronounced relative to alternative logic-gating when weak (two-mismatch) TFBSs are excluded while scoring motifs. Data are shown as mean±SE of the occurrence over replicate evolution simulations.

We also see C1-FFLs that, contrary to expectations, are not AND-gated. Non-AND-gated motifs are found more often in low fitness than high fitness replicates (**Fig. 4B**), indicating that the preference for AND-gates is associated with adaptation rather than mutation bias. However, some non-AND-gated motifs are still found even in the high fitness replicates. This is because motifs and their logic gates are scored on the basis of all TFBSs, even those with two mismatches and hence low binding affinity. Unless these weak TFBSs are deleterious, they will appear quite often by chance alone. A random 8-bp sequence has probability $\binom{8}{2} \times 0.25^6 \times 0.75^2 = 0.0038$ of being a two-mismatch binding site for a given TF. In our model, a TF has the potential to recognize 137 different sites in a 150-bp cis-regulatory sequence (taking into account steric hindrance at the edges), each with 2 orientations. Thus, by chance alone a given TF will have $0.0038 \times 137 \times 2 \approx 1$ two-mismatch binding sites in a given cis-regulatory sequence (ignoring palindromes for simplicity), compared to only ~0.1 one-mismatch TFBSs. Non-AND-gated C1-FFLs mostly disappear when two-mismatch TFBSs are excluded, but the AND-gated C1-FFLs found in high fitness replicates do not (**Fig. 4C**).

To confirm the functionality of these AND-gated C1-FFLs, we mutated the evolved genotype in two different ways (**Fig. 5A**) to remove the AND regulatory logic. As expected, this lowers fitness in the presence of the short spurious signal but increases fitness in the presence of constant signal, with a net reduction in fitness (**Fig. 5B**). This is consistent with AND-gated C1-FFLs representing a tradeoff, by which a more rapid response to a true signal is sacrificed in favor of the greater reliability of filtering out short spurious signals.

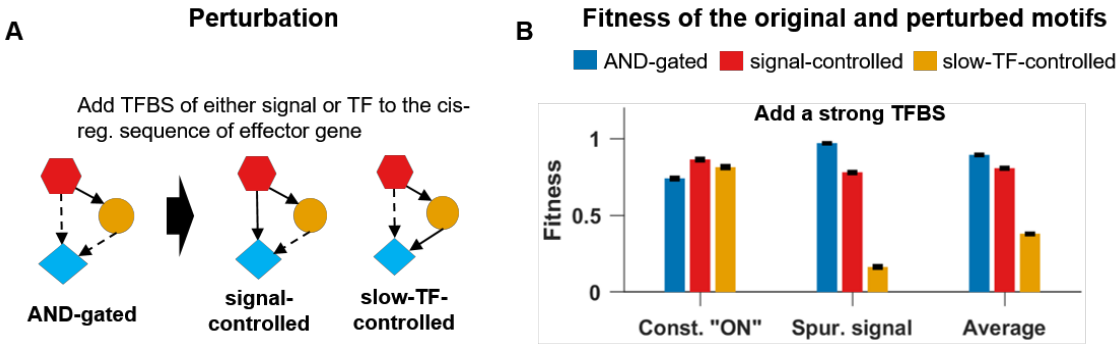


Figure 5. Destroying the AND-logic of a C1-FFL removes its ability to filter out short spurious signals. (**A**) For each of the *n* = 25 replicates in the high fitness group in **Fig. 4**, we perturbed the AND-logic in two ways, by adding one binding site of either the signal or the slow TF to the cis-regulatory sequence of the effector gene. **(B)** For each replicate, the fitness of the original motif (blue) or of the perturbed motif (red or orange) was averaged across the subset of evolutionary steps with an AND-gated C1-FFL and lacking other potentially confounding motifs (see Supplementary Text Section 11 for details). Destroying the AND-logic slightly increases the ability to respond to the signal, but leads to a larger loss of fitness when short spurious signals are responded to. Fitness is shown as mean±SE over replicate evolutionary simulations.

Adaptive motifs are constrained not only in their topology and regulatory logic, but also in the parameter space of their component genes. In particular, there is selection for rapid synthesis of both effector and TF proteins, as well as rapid degradation of effector mRNA and protein (**Table S4**). Fast effector degradation reduces the transient expression induced by the short spurious signal (**Fig. S2**).

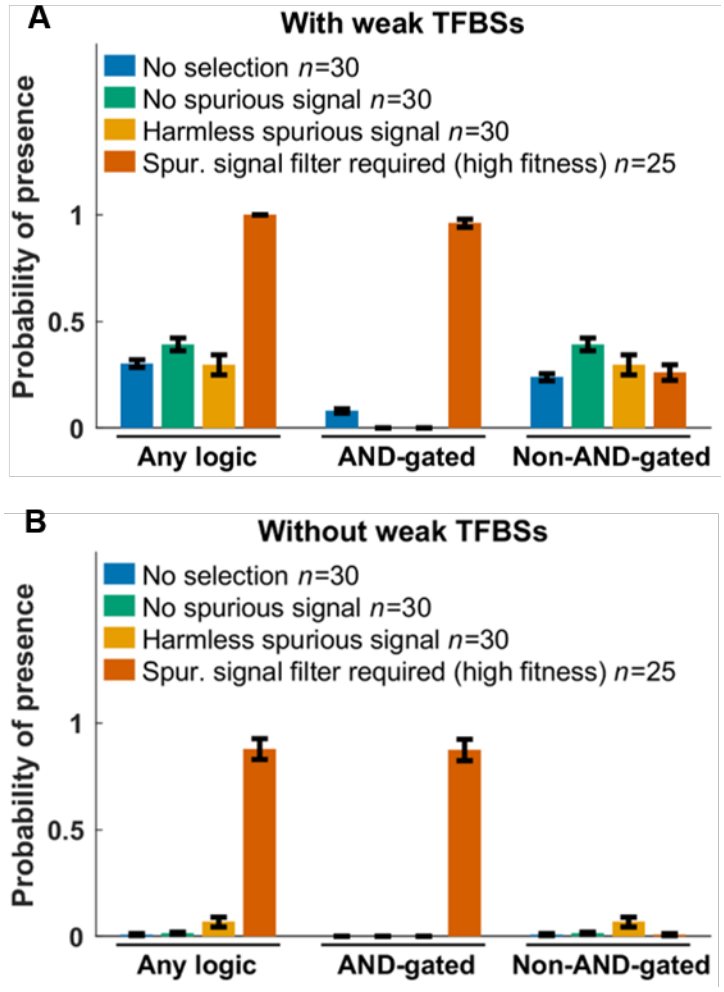


Figure 6. Selection for filtering out short spurious signals is the primary cause of C1-FFLs. TRNs are evolved under different selection conditions, and we score the probability that at least one C1-FFL is present (Supplementary Text Section 11). Weak (two-mismatch) TFBSs are included **(A)** or excluded **(B)** during motif scoring. Data are shown as mean±SE over evolutionary replicates. C1-FFL occurrence is similar for high-fitness and low-fitness outcomes in control selective conditions (**Fig. S4**), and so all evolutionary outcomes were combined. "Spurious signal filter required (high fitness)" uses the same data as in **Fig. 4**.

To test the extent to which AND-gated C1-FFLs are a specific response to selection to filter out short spurious signals, we simulated evolution under three negative control conditions: 1) no

477 selection, i.e. all mutations are accepted to become the new resident genotype; 2) no spurious
478 signal, i.e. selection to express the effector under a constant "ON" signal and not under a
479 constant "OFF" signal; 3) harmless spurious signal, i.e. selection to express the effector under a
480 constant "ON" environment whereas effector expression in the "OFF" environment with short
481 spurious signals is neither punished nor rewarded beyond the cost of unnecessary gene
482 expression. AND-gated C1-FFLs evolve much less often under all three negative control
483 conditions (**Fig. 6**), showing that their prevalence is a consequence of selection for filtering out
484 short spurious signals, rather than a consequence of mutational bias and/or simpler forms of
485 selection. C1-FFLs that do evolve under control conditions tend not to be AND-gated (**Fig. 6A**),
486 and mostly disappear when weak TFBSs are excluded during motif scoring (**Fig. 6B**).
487
488 **Diamond motifs are an alternative adaptation in more complex networks**
489
490 In real biological situations, sometimes the source signal will not be able to directly regulate an
491 effector, and must instead operate via a longer regulatory pathway involving intermediate TFs[46].
492 In this case, even if the signal itself takes the idealized form shown in **Fig. 3**, its shape after
493 propagation may become distorted by the intrinsic processes of transcription. Motifs are under
494 selection to handle this distortion.
495
496 To enforce indirect regulation, we ran simulations in which the signal was only allowed to bind
497 to the cis-regulatory sequences of TFs and not of effector genes. The fitness distribution of the
498 evolutionary replicates has no obvious gaps (**Fig. S5**), so we compared the highest fitness, lowest
499 fitness, and median fitness replicates. In agreement with results when direct regulation is

allowed, genotypes of low and medium fitness contain few AND-gated C1-FFLs, while high fitness genotypes contain many more (**Fig. 7B, left and right**).

While visually examining the network context of these C1-FFLs, we discovered that many were embedded within AND-gated "diamonds". In a diamond, the signal activates the expression of two genes that encode different TFs, and the two TFs activate the expression of an effector gene (**Fig. 7A middle**). When one of the two TF genes activates the other, then a C1-FFL is also present among the same set of genes; we call this topology a "FFL-in-diamond" (**Fig. 7A right**), and the prevalence of this configuration drew our attention toward diamonds. This led us to discover that AND-gated diamonds also occurred frequently without AND-gated C1-FFLs, in the configuration we call "isolated diamonds" (**Fig. 7A middle**). Note that it is in theory possible, but in practice uncommon, for diamonds to be part of more complex conjugates. Systematically scoring the AND-gated isolated diamond motif confirmed its high occurrence (**Fig. 7B and C, middle**).

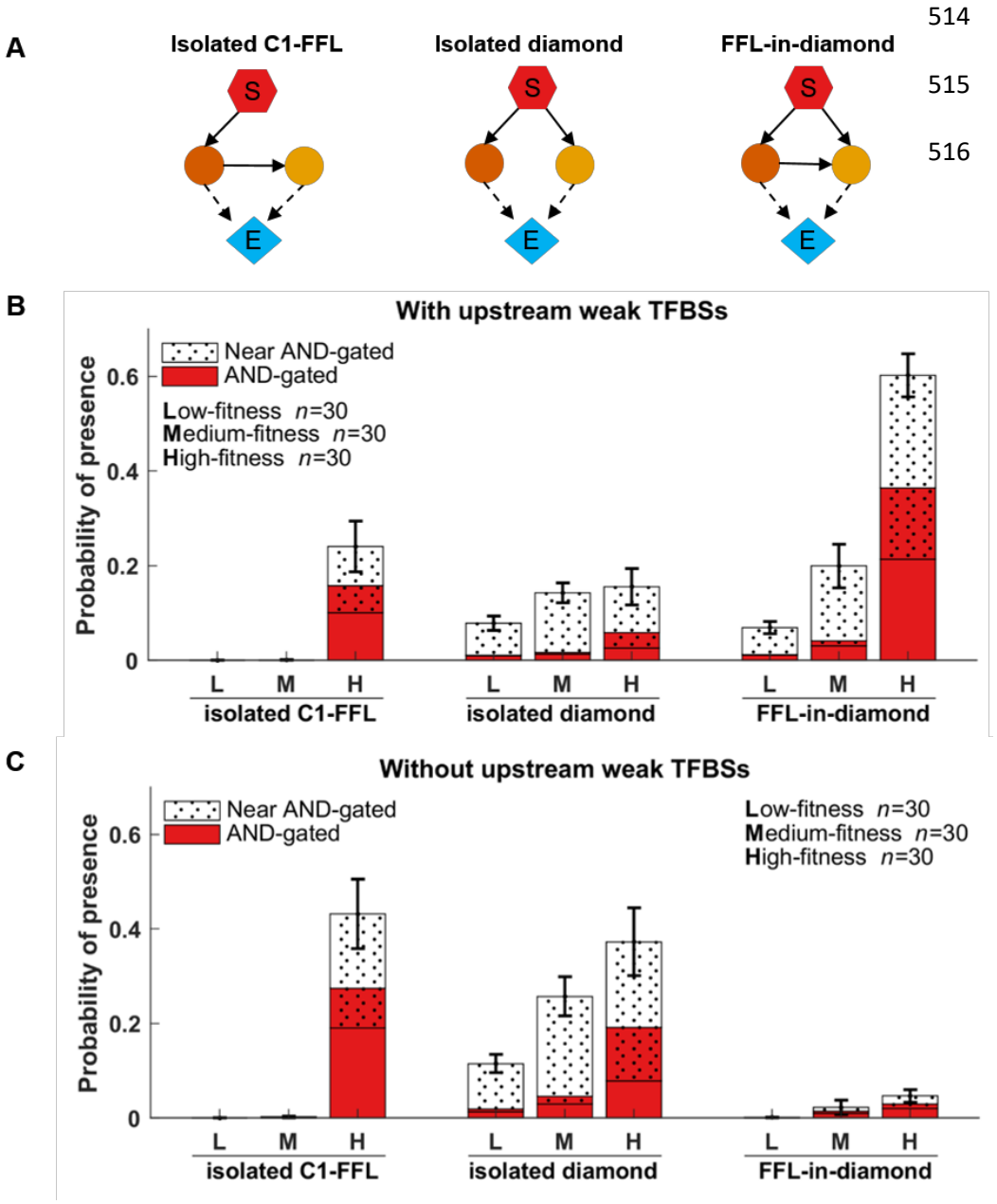
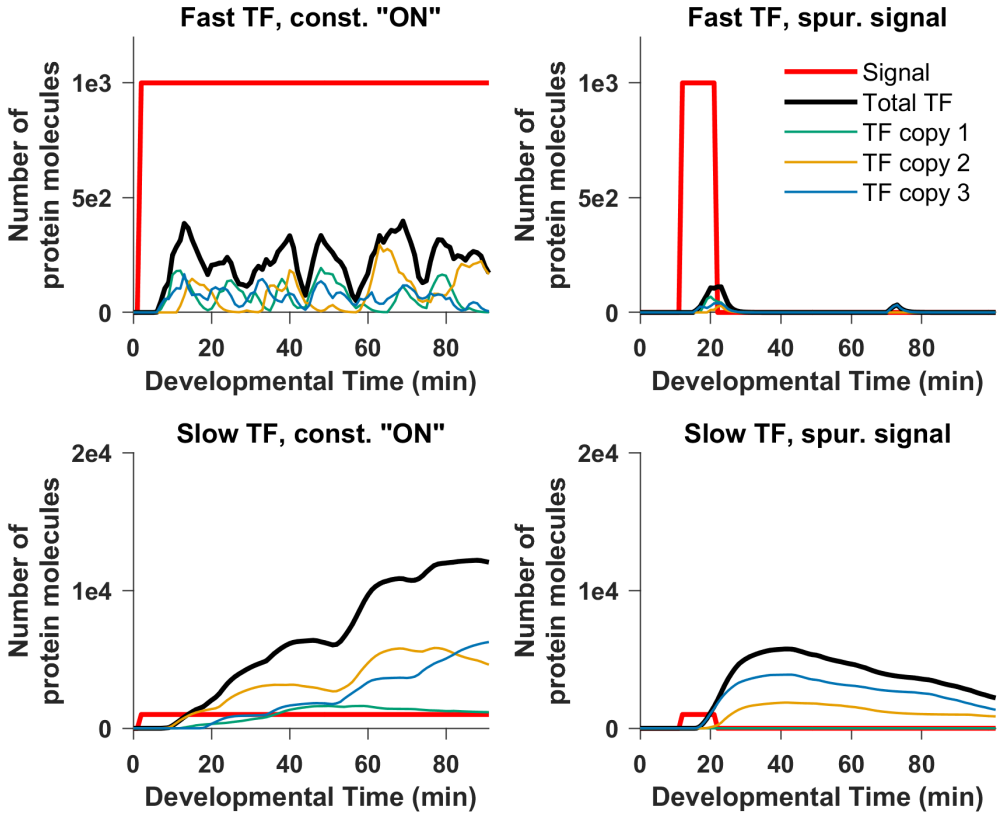


Figure 7. Both AND-gated C1-FFLs and AND-gated diamonds (A) are associated with high fitness in complex networks under selection to filter out short spurious signals. Out of 160 simulations (**Fig. S5**), we took the 30 with the highest fitness (H), the 30 with the lowest fitness (L), and 30 of around median fitness (M). AND-gated motifs are scored while including weak TFBSs in the effectors' cis-regulatory regions, near-AND-gated motifs are those scored only when these weak TFBSs are excluded. It is possible for the same genotype to contain one of each, resulting in overlap between the red AND-gated columns and the dotted near-AND-gated columns. Weak TFBSs upstream in the TRN, i.e. not in the effector, are shown both included (**B**) and excluded (**C**). See Supplementary Text Section 11 for y-axis calculation details. Error bars show mean±SE of the proportion of evolutionary steps containing the motif in question, across replicate evolutionary simulations.

An AND-gated C1-FFL integrates information from a short/fast regulatory pathway with information from a long/slow pathway, in order to filter out short spurious signals. A diamond achieves the same end of integrating fast and slowly transmitted information via differences in the gene expression dynamics of the two regulatory pathways, rather than via topological length (**Fig. 8**). The fast and slow pathways could be distinguished in a number of ways, e.g. by the slope at which the transcription factor concentration increases or the time at which it exceeds a threshold or plateaus. We found it convenient to identify the "fast TF" as the one with the higher protein degradation rate. Specifically, we use the geometric mean of the protein degradation rate over gene copies of a TF in order to differentiate the two TFs. The parameter values of the fast TF are more evolutionarily constrained than those of the slow TF (**Table S5**). In particular, there is selection for rapid degradation of the fast TF protein and mRNA (**Table S5**). Isolated AND-gated C1-FFLs also show pronounced selection for the TF in the fast pathway to have rapid protein degradation (**Table S6**).



542

543 **Figure 8. The two intermediate TFs in an AND-gated "diamond" motif have different**

544 **expression dynamics and propagate the signal at different speeds.** Expression of the two TFs in

545 one representative genotype from the one high-fitness evolutionary replicate in **Fig. 7B** that

546 evolved an AND-gated isolated diamond is shown. Each TF is a different protein, and each is

547 encoded by 3 gene copies, shown separately in color, with the total in thick black. The

548 expression of one TF plateaus faster than that of the other; this is characteristic of the AND-

549 gated diamond motif, and leads to the same functionality as the AND-gated C1-FFL.

550

551 But mutational biases make it difficult to evolve very fast-degrading mRNA and protein. And

552 even when they do evolve, fast degradation keeps the fast TF at low concentrations. To

553 compensate, the fast TF must overcome mutational bias to also evolve high binding affinity and

554 rapid protein synthesis (**Table S5**, **Table S6**).

555

556 Note that a simple transcriptional cascade, signal -> TF -> effector, has also been found
557 experimentally to filter out short spurious signals when the intermediate TF is rapidly degraded,
558 dampening the effect of a brief signal[47]. Two such transcriptional cascades involving different
559 intermediate TFs form a diamond, so the utility of a single cascade is a potential explanation for
560 the high prevalence of double-cascade diamonds. However, in this case we would have no
561 reason to expect marked differences in expression dynamics between the two TFs, as illustrated
562 in **Fig. 8** and **Table S5**. Enrichment for AND-gates (**Fig. 7**) indicates selection to integrate
563 information from the two cascades. On the other hand, we do find some non-AND-gated
564 diamonds, and these might best be considered as cascades. Inspection of their parameter values
565 reveals that in these diamonds, both TFs have fast-degrading mRNAs and proteins so that both
566 TFs shut down rapidly once signal is turned off. This makes such diamonds less vulnerable to
567 spurious signals, reducing the need for the AND gate. The difficulty of evolving not just one but
568 two fast-degrading high-affinity TFs likely explains why non-AND-gated diamonds are rare. As
569 we will see in the next section, these non-AND-gated diamonds are nevertheless scored as AND-
570 gated when weak TFBSs are excluded.

571

572 **Weak TFBSs can change how adaptive motifs are scored even when they do not change**
573 **function**

574

575 Results depend on whether we include weak TFBSs when scoring motifs. Weak TFBSs can either
576 be in the effector's cis-regulatory region, affecting how the regulatory logic is scored, or in TFs
577 upstream in the TRN, affecting only the presence or absence of motifs. When a motif is scored
578 as AND-gated only when two-mismatch TFBSs in the effector are excluded, we call it a "near-

AND-gated" motif. Recall from **Fig. 2** that effector expression requires two TFs to be bound, with only one TFBS of each type creating an AND-gate. When a second, two-mismatch TFBS of the same type is present, we have a near-AND-gate. TFs may bind so rarely to this weak affinity TFBS that its presence changes little, making the regulatory logic still effectively AND-gated. A near-AND-gated motif may therefore evolve for the same adaptive reasons as an AND-gated one. **Fig. 7B** and **C** shows that both AND-gated and near-AND-gated motifs are enriched in the higher fitness genotypes.

When we exclude upstream weak TFBSs while scoring motifs, FFL-in-diamonds are no longer found, while the occurrence of isolated C1-FFLs and diamonds increases (**Fig. 7C**). This makes sense, because adding one weak TFBS, which can easily happen by chance alone, can convert an isolated diamond or C1-FFL into a FFL-in-diamond (added between intermediate TFs, or from signal to slow TF, respectively).

AND-gated isolated C1-FFLs appear mainly in the highest fitness outcomes, while AND-gated isolated diamonds appear in all fitness groups (**Fig. 7C**), suggesting that diamonds are easier to evolve. 25 out of 30 high-fitness evolutionary replicates are scored as having a putatively adaptive AND-gated or near-AND-gated motif in at least 50% of their evolutionary steps when upstream weak TFBSs are ignored (close to addition of bars in **Fig. 7C**, because these two AND-gated motifs rarely coexist in a high-fitness genotype).

Just as for the AND-gated C1-FFLs evolved under direct regulation and analyzed in **Fig. 5**, perturbation analysis supports an adaptive function for AND-gated C1-FFLs and diamonds evolved under indirect regulation (**Fig. 9A.i, 9B.i**). Breaking the AND-gate logic of these motifs by

603 adding a (strong) TFBS to the effector cis-regulatory region reduces the fitness under the
604 spurious signal but increases it under the constant "ON" beneficial signal, resulting in a net
605 decrease in the overall fitness.
606
607 If we add a weak (two-mismatch) TFBS instead, this converts an AND-gated motif to a near-AND-
608 gated motif. This lowers fitness only when the extra link is from the slow TF to the effector, and
609 not when the extra link is from the fast TF to the effector (**Fig. 9A.ii, 9B.ii)**.

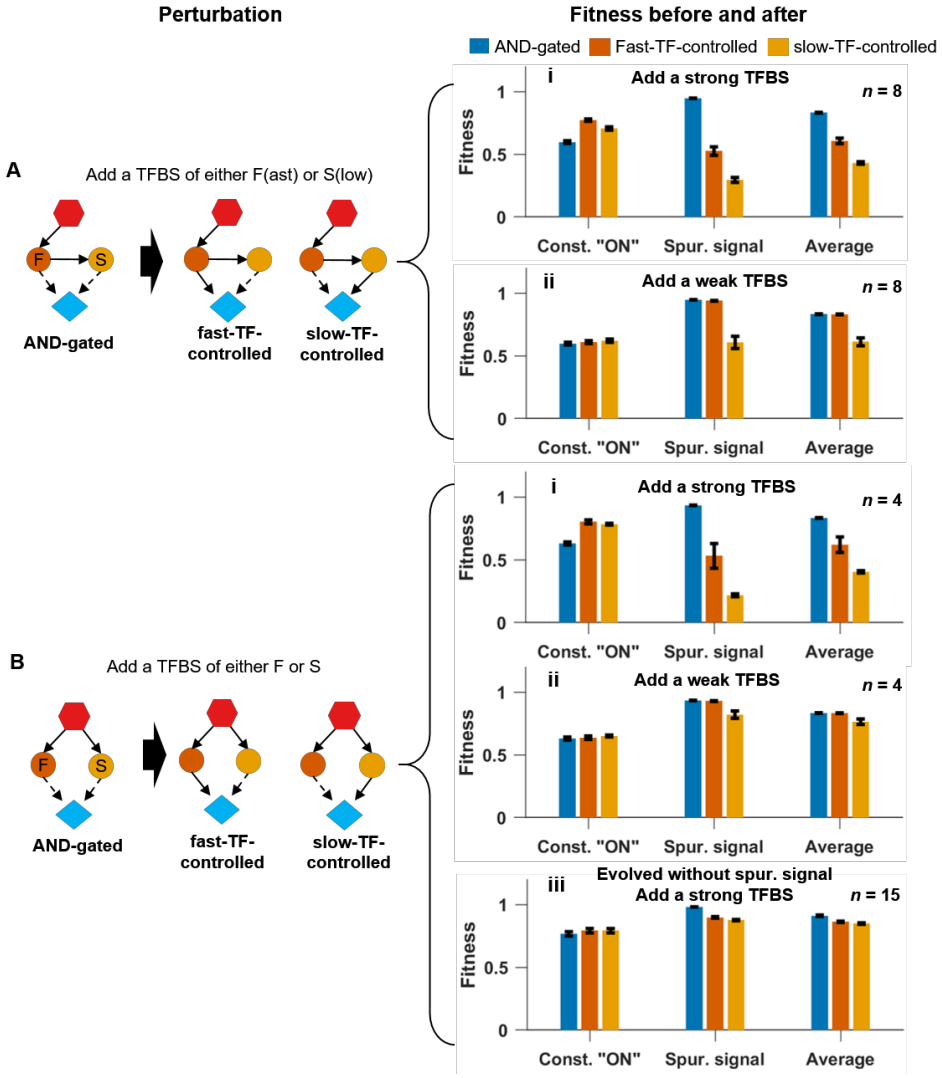
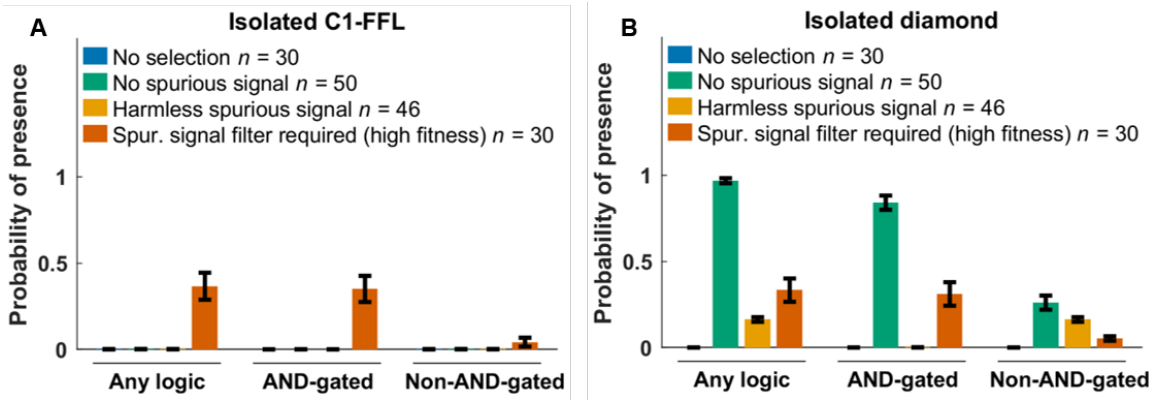


Figure 9. Perturbation analysis shows that AND-gated C1-FFLs (A) and diamonds (B) filter out short spurious signals. We add a strong TFBS **(i)** or a two-mismatch TFBS **(ii)** or **(iii)**; the latter creates near-AND-gated motifs. Allowing the effector to respond to the slow TF alone slightly increases the ability to respond to the signal, but leads to a larger loss of fitness when effector expression is undesirable. Allowing the effector to respond to the fast TF alone does so only when the conversion uses a strong TFBS not a two-mismatch TFBS. **(A)** We perform the perturbation on 8 of the 18 high-fitness replicates from **Fig.7B** that evolved an AND-gated C1-FFL. **(B)** (**i**) and (**ii**) are based on 4 of the 26 high-fitness replicates that evolved an AND-gated diamond in **Fig. 7B**, (**iii**) is based on 15 of the 37 replicates that evolved an AND-gated diamond in response to selection for signal recognition in the absence of an external spurious signal (**Fig. 10B**). Replicate exclusion was based on the co-occurrence of other motifs with the potential to confound results (see Supplementary Text Section 12 for details). Fitness is shown as mean±SE of over replicate evolutionary simulations, calculated as described for **Fig. 5**.

623 Indeed, these extra links are tolerated during evolution too. If we take the 16 high-fitness
624 replicates that contain a near-AND-gated C1-FFL in at least 1% of the evolutionary steps, then
625 for 15 replicates of the 16, at least 88% of the near-AND-gated C1-FFLs in each of the 15
626 replicates are only near-AND-gated because of extra weak TFBSs for the fast TF. In the remaining
627 1 replicate, 93% of the near-AND-gated C1-FFLs have extra weak TFBSs specific for each of the
628 TFs (and are therefore scored as OR-gated). In this last replicate, the two TFs in these OR-gated
629 C1-FFLs have high and similar protein degradation rates, reducing the need for an AND gate for
630 reasons discussed earlier. We similarly examine high-fitness replicates that, when upstream
631 weak TFBSs are excluded, contain a near-AND-gated diamond in at least 1% of the evolutionary
632 steps. In 15 of these 24 evolutionary replicates, the near-AND regulatory logic is in most
633 evolutionary steps due to an extra weak TFBS of the fast TF, in 8 replicates (all of them OR-
634 gated, like the OR-gated C1-FFL already discussed) it is due to weak TFBSs for each of the TFs,
635 and in only 1 replicate is it due to an extra TFBS for the slow TF. For the latter two categories,
636 both TFs in near-AND-gated diamonds have high and similar protein degradation rates. By
637 chance alone, fast and slow TF should be equally likely to contribute the weak TFBS that makes a
638 motif near-AND-gated rather than AND-gated. This expected 50:50 ratio can be rejected from
639 our observed 15:0 and 15:1 ratios with $p = 3 \times 10^{-5}$ and $p = 3 \times 10^{-4}$, respectively
640 (cumulative binomial distribution, one-sided test). This non-random occurrence of weak TFBSs
641 creating near-AND-gates illustrates how even weak TFBSs can be shaped by selection against
642 some (but not all) motif-breaking links.
643
644 **AND-gated isolated diamonds also evolve in the absence of external spurious signals**
645

646 We simulated evolution under the same three control conditions as before, this time without
647 allowing the signal to directly regulate the effector. In the "no spurious signal" and "harmless
648 spurious signal" control conditions, motif frequencies are similar between low and high fitness
649 genotypes (**Fig. S6, Fig. S7)**, and so our analysis includes all evolutionary replicates. When weak
650 (two-mismatch) TFBSs are excluded, AND-gated isolated C1-FFLs are seen only after selection
651 for filtering out a spurious signal, and not under other selection conditions (**Fig. 10A**). However,
652 AND-gated isolated diamonds also evolve in the absence of spurious signals, indeed at even
653 higher frequency (**Fig. 10B**). Results including weak TFBSs are similar (**Fig. S8**).
654



655 **Figure 10. Selection for filtering out a short spurious signal is the primary way to evolve AND-**
656 **gated isolated C1-FFLs (A), but AND-gated isolated diamonds also evolve in the absence of**
657 **spurious signals (B).** The selection conditions are the same as in **Fig. 6**, but we do not allow the
658 signal to directly regulate the effector. When scoring motifs, we exclude all two-mismatch
659 TFBSs; more comprehensive results are shown in **Fig. S8**. Many non-AND-gated diamonds have
660 the "no regulation" logic in **Fig. 2**, perhaps as an artifact created by the duplication and
661 divergence of intermediate TFs; we excluded them from the "Any logic" and "Non-AND-gated"
662 tallies in (B). See Supplementary Text Section 11 for the calculation of y-axis. Data are shown as
663 mean±SE over evolutionary replicates. We reused data from **Fig. 7** for "Spurious signal filter
664 required (high fitness)".

Perturbing the AND-gate logic in these isolated diamonds reduces fitness via effects in the environment where expressing the effector is deleterious (**Fig. 9B.iii**). Even in the absence of external short spurious signals, the stochastic expression of intermediate TFs might effectively create short spurious signals when the external signal is set to "OFF". It seems that AND-gated diamonds evolve to mitigate this risk, but that AND-gated C1-FFLs do not. The duration of internally generated spurious signals has an exponential distribution, which means that the optimal filter would be one that does not delay gene expression[48]. The two TFs in an AND-gated diamond can be activated simultaneously, but they must be activated sequentially in an AND-gated C1-FFL; the shorter delays possible with AND-gated diamonds might explain why only diamonds and not FFLs evolve to filter out intrinsic noise in gene expression.

Discussion

Adaptive nature of AND-gated C1-FFLs

There has never been sufficient evidence to satisfy evolutionary biologists that motifs in TRNs represent adaptations for particular functions. Critiques by evolutionary biologists to this effect[13-23] have been neglected, rather than answered, until now. While C1-FFLs can be conserved across different species[49-52], this does not imply that specific "just-so" stories about their function are correct. In this work, we study the evolution of AND-gated C1-FFLs, which are hypothesized to be adaptations for filtering out short spurious signals[3]. Using a novel and more mechanistic computational model to simulate TRN evolution, we found that AND-gated C1-FFLs evolve readily under selection for filtering out a short spurious signal, and not under control conditions. Our results support the adaptive hypothesis about C1-FFLs.

AND-gated C1-FFLs express an effector after a noise-filtering delay when the signal is turned on, but shut down expression immediately when the signal is turned off, giving rise to a "sign-sensitive delay"[3, 7]. Rapidly switching off has been hypothesized to be part of their selective advantage, above and beyond the function of filtering out short spurious signals[48]. We intended to select only for filtering out a short spurious signal, and not for fast turn-off; specifically, we expected effector expression to evolve a delay equal to the duration of the spurious signal. However, evolved solutions still expressed the effector in the presence of short spurious signals (**Fig. S2**), and thus benefitted from rapidly turning off this spurious expression. In other words, we effectively selected for both delayed turn-on and rapid turn-off, despite our intent to only select for the former.

It is difficult to distinguish adaptations from "spandrels"[8]. Standard procedure is to look for motifs that are more frequent than expected from some randomized version of a TRN[2, 53]. For this method to work, this randomization must control for all confounding factors that are non-adaptive with respect to the function in question, from patterns of mutation to a general tendency to hierarchy – a near-impossible task. Our approach to a null model is not to randomize, but to evolve with and without selection for the specific function of interest. This meets the standards of evolutionary biology for inferring the adaptive nature of a motif[13-23].

Technical lessons learned

Previous studies have also attempted to evolve adaptive motifs in a computational TRN, successfully under selection for circadian rhythm and for multiple steady states[54], and

unsuccessfully under selection to produce a sine wave in response to a periodic pulse[23]. Other studies have evolved adaptive motifs in a mixed network of transcriptional regulation and protein-protein interaction[55-57]. Our successful simulation might offer some methodological lessons, especially a focus on high-fitness evolutionary replicates, which was done by us and by Burda et al.[54] but not by Knabe et al.[23].

Knabe et al.[23] suggested that including a cost for gene expression may suppress unnecessary links and thus make it easier to score motifs. However, when we removed the cost of gene expression term ($C(t) = 0$ in Supplementary Section 8), AND-gated C1-FFLs still evolved in the high-fitness genotypes under selection for filtering out a spurious signal (**Fig. S9**). In our model, removing the cost of gene expression did not, via permitting unnecessary links, conceal motifs.

While simplified relative to reality, our model is undeniably complicated. An important question is which complications are important for what. One complication is our nucleotide-sequence-level model of cis-regulatory sequences. This has the advantage of capturing weak TFBSs, realistic turnover, and other mutational biases. The disadvantage is that calculating the probabilities of TF binding is computationally expensive and scales badly with network size. Future work might design a more schematic model of cis-regulatory sequences to improve computation while still capturing realistic mutation biases. A second complication of our approach is the stochastic simulation of gene expression. This is essential for our question, because intrinsic noise in gene expression can mimic the effects of a spurious signal, but may be less important in other scenarios, e.g. where the focus is on steady state behavior.

The ubiquity of weak TFBSs complicates motif scoring

Our model, while complex for a model and hence capable of capturing intrinsic noise, is inevitably less complex than the biological reality. However, we hope to have captured key phenomena, albeit in simplified form. One key phenomenon is that TFBSs are not simply present vs. absent but can be strong or weak, i.e. the TRN is not just a directed graph, but its connections vary in strength. Our model, like that of Burda et al.[54] in the context of circadian rhythms, captures this fact by basing TF binding affinity on the number of mismatch deviations from a consensus TFBS sequence. While in reality, the strength of TF binding is determined by additional factors, such as broader nucleic context and cooperative behavior between TFs (reviewed in Inukai et al.[58]), these complications are unlikely to change the basic dynamics of frequent appearance of weak TFBSs and greater mutational accessibility of strong TFBSs from weak TFBSs than de novo. Similarly, AND-gating can be quantitative rather than qualitative[59], a phenomenon that weak TFBSs in our model provide a simplified version of.

Core links in adaptive motifs almost always involve strong not weak TFBSs. However, weak (two-mismatch) TFBSs can create additional links that prevent an adaptive motif from being scored as such. Some potential additional links are neutral while others are deleterious; the observed links are thus shaped by this selective filter, without being adaptive. Note that there have been experimental reports that even weak TFBSs can be functionally important[60, 61]; these might, however, better correspond to 1-mismatch TFBSs in our model than two-mismatch TFBSs. Ramos et al.[61] and Crocker et al.[60] identified their “weak” TFBSs in comparison to the strongest possible TFBS, not in comparison to the weakest still showing affinity above baseline.

Different solutions for filtering out short spurious signals

761

762 A striking and unexpected finding of our study was that AND-gated diamonds evolved as an
763 alternative motif for filtering out short spurious external signals, and that these, unlike FFLs,
764 were also effective at filtering out intrinsic noise. Multiple motifs have previously been found
765 capable of generating the same steady state expression pattern[21]; here we find multiple motifs
766 for a much more complex function.

767

768 Diamonds are not overrepresented in the TRNs of bacteria[2] or yeast[62], but are overrepresented
769 in signaling networks (in which post-translational modification plays a larger role)[63], and in
770 neuronal networks[1]. In our model, we treated the external signal as though it were a
771 transcription factor, simply as a matter of modeling convenience. In reality, signals external to a
772 TRN are by definition not TFs (although they might be modifiers of TFs). This means that our
773 indirect regulation case, in which the signal is not allowed to directly turn on the effector, is the
774 most appropriate one to analyze if our interest is in TRN motifs that mediate contact between
775 the two. Note that if under indirect regulation we were to score the signal as not itself a TF, we
776 would observe adaptive C1-FFLs but not diamonds, in agreement with the TRN data. However,
777 this TRN data might miss functional diamond motifs that spanned levels of regulatory
778 organization, i.e. that included both transcriptional and other forms of regulation. The greatest
779 chance of finding diamonds within TRNs alone come from complex and multi-layered
780 developmental cascades, rather than bacterial or yeast[64]. Multiple interwoven diamonds are
781 hypothesized to be embedded with multi-layer perceptrons that are adaptations for complex
782 computation in signaling networks[65].

783

Previous work has also identified alternatives to AND-gated C1-FFLs. Specifically, in mixed networks of transcriptional regulation and protein-protein interactions, FFLs did not evolve under selection for delayed turn-on (as well as rapid turn-off)[57]. Indeed, even when a FFL topology was enforced, with only the parameters allowed to evolve, two alternative motifs remained superior[57]. However, one alternative motif, which the authors called "positive feedback" is essentially still an AND-gated C1-FFL, specifically one in which the intermediate TF expression is also AND-gated, requiring both itself and the signal for upregulation. The other is a cascade in which the signal inhibits the expression of an intermediate TF protein that represses the expression of the effector. The cost of constitutive expression of the intermediate TF in the absence of the signal was not modeled[57], giving this cascade an unrealistic advantage.

The importance of dynamics and intrinsic noise

Most previous research on C1-FFLs has used an idealized implementation (e.g. a square wave) of what a short spurious signal entails[4, 48, 66]. In real networks, noise arises intrinsically in a greater diversity of forms, which our model does more to capture. Even when a "clean" form of noise enters a TRN, it subsequently gets distorted with the addition of intrinsic noise[67]. Intrinsic noise is ubiquitous and dealing with it is an omnipresent challenge for selection. Indeed, we see adaptive diamonds evolve to suppress intrinsic noise, even when we select in the absence of extrinsic spurious signals.

The function of a motif relies ultimately on its dynamic behavior, with topology merely a means to that end. To create two pathways that regulate the effector in different speeds, the C1-FFL motif uses a pair of short and long pathways, but these also correspond to fast-degrading and

808 slow-degrading TFs. This same function was achieved entirely non-topologically in our
809 adaptively evolved diamond motifs. This agrees with other studies showing that topology alone
810 is not enough to infer activities such as spurious signal filtering from network motifs [68-70].
811

812 Acknowledgements

813 Work was supported by the University of Arizona and by a Pew Scholarship to JM, John
814 Templeton Foundation grant 39667 to JM, and by National Institutes of Health grants
815 R35GM118170 to MLS and R01GM076041 to JM. We thank Hinrich Boeger for helpful
816 discussions and careful reading of the manuscript, Jasmin Uribe for early work on this project,
817 and the high-performance computing center at the University of Arizona for generous
818 allocations.
819

820 Author Contributions

821 K.X. and J.M. designed the simulations, analyzed the results, and wrote the manuscript. K.X.
822 performed the simulations and statistical analyses. K.X., A.L., and J.M. wrote the simulation
823 code. M.L.S. and J.M. conceptualized the initial design of the simulations.
824

825 **Competing Interests**: The authors declare no conflicts of interest.

References

1. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: Simple building blocks of complex networks. *Science* **298**, 824-827 (2002).

2. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet* **31**, 64-68 (2002).

3. Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-461 (2007).

4. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100**, 11980-11985 (2003).

5. Jaeger KE, Pullen N, Lamzin S, Morris RJ, Wigge PA. Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis. *The Plant Cell* **25**, 820-833 (2013).

6. Peter IS, Davidson EH. Assessing regulatory information in developmental gene regulatory networks. *Proc Natl Acad Sci USA* **114**, 5862-5869 (2017).

7. Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* **334**, 197-204 (2003).

8. Gould SJ, Lewontin RC. The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proc R Soc Lond, B, Biol Sci* **205**, 581-598 (1979).

9. Graur D, Zheng Y, Price N, Azevedo RBR, Zufall RA, Elhaik E. On the Immortality of Television Sets: “Function” in the Human Genome According to the Evolution-Free Gospel of ENCODE. *Genome Biol Evol* **5**, 578-590 (2013).

10. Masel J, Promislow DEL. Answering evolutionary questions: A guide for mechanistic biologists. *BioEssays* **38**, 704-711 (2016).

11. Widder S, Solé R, Macía J. Evolvability of feed-forward loop architecture biases its abundance in transcription networks. *BMC Syst Biol* **6**, 7 (2012).

12. Cordero OX, Hogeweg P. Feed-forward loop circuits as a side effect of genome evolution. *Mol Biol Evol* **23**, 1931-1936 (2006).

13. Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L. Comment on "Network Motifs: Simple Building Blocks of Complex Networks" and "Superfamilies of Evolved and Designed Networks". *Science* **305**, 1107 (2004).

14. Jenkins D, Stekel D. De Novo Evolution of Complex, Global and Hierarchical Gene Regulatory Mechanisms. *J Mol Evol* **71**, 128-140 (2010).

15. Lynch M. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* **8**, 803-813 (2007).

16. Mazurie A, Bottani S, Vergassola M. An evolutionary and functional assessment of regulatory network motifs. *Genome Biol* **6**, R35 (2005).

17. Solé RV, Valverde S. Are network motifs the spandrels of cellular complexity? *Trends Ecol Evol* **21**, 419-422 (2006).

18. Tsuda ME, Kawata M. Evolution of Gene Regulatory Networks by Fluctuating Selection and Intrinsic Constraints. *PLoS Comput Biol* **6**, e1000873 (2010).

19. Wagner A. Does Selection Mold Molecular Networks? *Sci STKE* **2003**, pe41 (2003).

20. Kuo PD, Banzhaf W, Leier A. Network topology and the evolution of dynamics in an artificial genetic regulatory network model created by whole genome duplication and divergence. *BioSystems* **85**, 177-200 (2006).

21. Payne JL, Wagner A. Function does not follow form in gene regulatory circuits. *Scientific Reports* **5**, 13015 (2015).

22. Ruths T, Nakhleh L. Neutral forces acting on intragenomic variability shape the *Escherichia coli* regulatory network topology. *Proc Natl Acad Sci USA* **110**, 7754-7759 (2013).

23. Knabe JF, Nehaniv CL, Schilstra MJ. Do motifs reflect evolved function?—No convergent evolution of genetic regulatory network subgraph topologies. *Biosystems* **94**, 68-74 (2008).

24. Raser JM, O'Shea EK. Noise in Gene Expression: Origins, Consequences, and Control. *Science* **309**, 2010-2013 (2005).

25. Kærn M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451-464 (2005).

26. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).

27. Draghi J, Whitlock M. Robustness to noise in gene expression evolves despite epistatic constraints in a model of gene networks. *Evolution* **69**, 2345-2358 (2015).

28. Jenkins DJ, Stekel DJ. A New Model for Investigating the Evolution of Transcription Control Networks. *Artificial Life* **15**, 259-291 (2009).

29. Henry A, Hemery M, François P. ϕ-evo: A program to evolve phenotypic models of biological networks. *PLoS Comput Biol* **14**, e1006244 (2018).

30. Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. Linking Stochastic Fluctuations in Chromatin Structure and Gene Expression. *PLoS Biol* **11**, e1001621 (2013).

31. Mao C*, et al.* Quantitative analysis of the transcription control mechanism. *Mol Syst Biol* **6**, 431 (2010).

32. Shahbazian MD, Grunstein M. Functions of Site-Specific Histone Acetylation and Deacetylation. *Annu Rev Biochem* **76**, 75-100 (2007).

33. Voss TC, Hager GL. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69-81 (2013).

34. Katan-Khaykovich Y, Struhl K. Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev* **16**, 743-752 (2002).

35. Courey AJ, Jia S. Transcriptional repression: the long and the short of it. *Genes Dev* **15**, 2786-2796 (2001).

36. Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* **48**, 575-608 (2013).

37. Decker KB, Hinton DM. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annu Rev Microbiol* **67**, 113-139 (2013).

38. Roy AL, Singer DS. Core promoters in transcription: old problem, new insights. *Trends Biochem Sci* **40**, 165-171 (2015).

39. van Drogen F, Stucke VM, Jorritsma G, Peter M. MAP kinase dynamics in response to pheromones in budding yeast. *Nat Cell Biol* **3**, 1051 (2001).

40. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 2340-2361 (1977).

41. Ghaemmaghami S*, et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737-741 (2003).

42. Kafri M, Metzl-Raz E, Jona G, Barkai N. The Cost of Protein Production. *Cell Reports* **14**, 22-31 (2016).

43. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. *Genetica* **102**, 127 (1998).

44. van Nimwegen E, Crutchfield JP, Huynen M. Neutral evolution of mutational robustness. *Proc Natl Acad Sci USA* **96**, 9716-9720 (1999).

45. Kimura M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713-719 (1962).

46. Balázsi G, Barabási AL, Oltvai ZN. Topological units of environmental signal processing in the transcriptional regulatory network of *Escherichia coli*. *Proc Natl Acad Sci U S A* **102**, 7841-7846 (2005).

47. Hooshangi S, Thiberge S, Weiss R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA* **102**, 3581-3586 (2005).

48. Dekel E, Mangan S, Alon U. Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Physical Biology* **2**, 81 (2005).

49. Boyle AP*, et al.* Comparative analysis of regulatory information and circuits across distant species. *Nature* **512**, 453-456 (2014).

50. Kemmeren P*, et al.* Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell* **157**, 740-752 (2014).

51. Stergachis AB*, et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365-370 (2014).

52. Madan Babu M, Teichmann SA, Aravind L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358**, 614-633 (2006).

53. Kashtan N, Itzkovitz S, Milo R, Alon U. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* **20**, 1746-1758 (2004).

54. Burda Z, Krzywicki A, Martin OC, Zagorski M. Motifs emerge from function in model gene regulatory networks. *Proc Natl Acad Sci USA* **108**, 17263-17268 (2011).

55. François P, Hakim V. Design of genetic networks with specified functions by evolution in silico. *Proc Natl Acad Sci USA* **101**, 580-585 (2004).

56. François P, Siggia ED. Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* **137**, 2385-2395 (2010).

57. Warmflash A, François P, Siggia ED. Pareto evolution of gene networks: An algorithm to optimize multiple fitness objectives. *Physical Biology* **9**, 56001-56007 (2012).

58. Inukai S, Kock KH, Bulyk ML. Transcription factor–DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* **43**, 110-119 (2017).

59. Wang D*, et al.* Loregic: A Method to Characterize the Cooperative Logic of Regulatory Factors. *PLoS Comput Biol* **11**, e1004132 (2015).

60. Crocker J*, et al.* Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**, 191-203 (2015).

61. Ramos AI, Barolo S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos Trans R Soc Lond, B, Biol Sci* **368**, 20130018 (2013).

62. Lee TI*, et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799-804 (2002).

63. Ma'ayan A*, et al.* Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* **309**, 1078-1083 (2005).

64. Rosenfeld N, Alon U. Response Delays and the Structure of Transcription Networks. *J Mol Biol* **329**, 645-654 (2003).

65. Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC (2007).

66. Hayot F, Jayaprakash C. A feedforward loop motif in transcriptional regulation: induction and repression. *J Theor Biol* **234**, 133-143 (2005).

67. Pedraza JM, van Oudenaarden A. Noise Propagation in Gene Networks. *Science* **307**, 1965-1969 (2005).

68. Ingram PJ, Stumpf MP, Stark J. Network motifs: structure does not determine function. *BMC genomics* **7**, 108-108 (2006).

69. Wall ME. Structure–function relations are subtle in genetic regulatory networks. *Mathematical Biosciences* **231**, 61-68 (2011).

70. Wall ME, Dunlop MJ, Hlavacek WS. Multiple Functions of a Feed-Forward-Loop Gene Circuit. *J Mol Biol* **349**, 501-514 (2005).

Supporting information for "Feed-forward regulation adaptively evolves via dynamics rather than topology when there is intrinsic noise" by Xiong et al.

Table of contents

22

Table S1. Major model parameters

Parameter	Values[1]	Bounds[2]	References
Length of cis-regulatory sequence	150 bp		(Yuan et al. 2005)
Length of TF recognition sequence	8 bp		(Wunderlich & Mirny 2009)
Length occupied by a TF on each side of recognition sequence	3 bp		(Zhu & Zhang 1999)
Dissociation constant between TF and perfect TFBS, $K_d(0)$	**$10^{U(-9,-6)}$ mole/liter**[3]	$(0, 10^{-5})$	(Park et al. 2004; Nalefski et al. 2006)
Dissociation constant between TF and non-specific DNA, $K_d(3)$	10^{-5} M		(Maerkl & Quake 2007)
Base rate of transition from Repressed to Intermediate	0.15 min^{-1}		(Katan-Khaykovich & Struhl 2002)
Maximum transition rate from Repressed to Intermediate	0.92 min^{-1}		(Katan-Khaykovich & Struhl 2002; Brown et al. 2013)
Base rate of transition from Intermediate to Repressed	0.67 min^{-1}		(Katan-Khaykovich & Struhl 2002)
Maximum transition rate from Intermediate to Repressed	4.11 min^{-1}		Chosen to give same dynamic range and Repressed to Intermediate
Base rate of transition from Intermediate to Active	0.025 min^{-1}		(Brown et al. 2013)
Maximum transition rate from Intermediate to Active	3.3 min^{-1}		(Brown et al. 2013)
Transition rate from Active to Intermediate, $r_{Act\_to\_Int}$	**$10^{N(1.27, 0.226)}$ min^{-1}**[4]	[0.59, 64.7]	(Guillemette et al. 2005; Pelechano et al. 2010; Brown et al. 2013)
Length of gene, L	**$10^{N(2.568, 0.34)}$ codons**	[50, 5000]	(SGD Project)
Rate of transcription initiation, $r_{max\_transc\_init}$	6.75 min^{-1}		(Brown et al. 2013)
Speed of transcription elongation	600 codon/min		(Dujon 1996; Larson et al. 2011; Hocine et al. 2013)
Time for transcribing UTRs and for terminating transcription	1 min		(Dujon 1996; Larson et al. 2011; Hocine et al. 2013)
Rate of mRNA degradation, r_{mRNA_deg}	**$10^{N(-1.49, 0.267)}$ min^{-1}**	$[7.5\times10^{-4}, 0.54]$	(Wang et al. 2002)
Speed of translation elongation	330 codon/min		(Siwiak et al. 2010)
Translation initiation time	0.5 min		(Siwiak et al. 2010)
Protein synthesis rate, $r_{protein\_syn}$	**$10^{N(0.322, 0.416)}$ molecule $mRNA^{-1}$ min^{-1}**	$[4.5\times10^{-3}, 61.4]$	(Siwiak et al. 2010)
Rate of protein degradation, $r_{protein\_deg}$	**$10^{N(-1.88, 0.561)}$ min^{-1}**	$[3.0\times10^{-6}, 0.69]$	(Belle et al. 2006)
Saturation concentration of effector protein, N_{e_sat}	10,000 molecules/cell		(Ghaemmaghami et al. 2003)
Fitness cost of protein expression for a gene with $L = 10^{2.568}$, c_{transl}	2×10^{-6} $(molecules/min)^{-1}$		(Ghaemmaghami et al. 2003; Kafri et al. 2016)
Maximum number of effector gene copies	5		
Maximum number of TF gene copies, excluding the signal	19		

[1] Parameters in bold can be altered by mutation, and the table shows the distributions from which their initial values are sampled. Estimation of N_{e_sat} is described in the Methods; estimation of the other parameters is described in the Supplementary Text (Sections 1, 2 – 7, and 8).

[2] Same units as the parameter values. Parentheses mean the parameter cannot take the boundary values; square brackets mean it can. We also use these bounds to constrain mutation (see Section 9).

[3] The uniform distribution is denoted U(min, max).

[4] The normal distribution is denoted N(mean, SD).

Table S2. Mutation rates and effect sizes

Mutation	Relative rate	Effect of mutation[1]
Single nucleotide substitution	5.25×10^{-8} per gene	
Gene deletion	1.5×10^{-7} per gene[2]	
Gene duplication	1.5×10^{-7} per gene[2]	
Mutation to consensus sequence of a TF	3.5×10^{-9} per gene	
Mutation to TF identity (activator vs. repressor)	3.5×10^{-9} per gene	
Mutation to $K_d(0)$	3.5×10^{-9} per gene	k = 0.5, μ = -5[2], σ = 0.776
Mutation to L	1.2×10^{-11} per codon	
Mutation to $r_{protein\_syn}$	9.5×10^{-12} per codon	k = 0.5, μ = 0.021[2], σ = 0.760
Mutation to $r_{protein\_deg}$	9.5×10^{-12} per codon	k = 0.5, μ = -1.88, σ = 0.739
Mutation to $r_{Act\_to\_Int}$	9.5×10^{-12} per codon	k = 0.5, μ = 1.57[2], σ = 0.773
Mutation to r_{mRNA_deg}	9.5×10^{-12} per codon	k = 0.5, μ = -1.19, σ = 0.396

[1] Mutation to these quantitative rates takes the form $\log_{10}x' = \log_{10}x + \mathrm{Normal}(k(\mu - \log_{10}x), \sigma)$, where x is the original value of the rate and x' is the value after mutation. See Section 9 for details.

[2] The value of this parameter is different during burn-in. See Section 9 for details.

	Probability that mutation of this type is accepted, given it occurs		Probability that an accepted mutation is of this type, given that it is accepted	
	First 1000 evol. steps	Last 1000 evol. steps	First 1000 evol. steps	Last 1000 evol. steps
Substitution	0.34 ± 0.01	0.35 ± 0.00	0.180 ± 0.005	0.213 ± 0.008
Deletion	0.27 ± 0.01	0.21 ± 0.01	0.360 ± 0.003	0.345 ± 0.005
Duplication	0.34 ± 0.01	0.32 ± 0.01	0.368 ± 0.003	0.343 ± 0.005
TF recognition seq.	0.30 ± 0.02	0.19 ± 0.02	0.009 ± 0.001	0.005 ± 0.000
$r_{Act\_to\_Int}$	0.33 ± 0.02	0.25 ± 0.01	0.012 ± 0.001	0.010 ± 0.001
r_{mRNA_deg}	0.34 ± 0.02	0.27 ± 0.01	0.014 ± 0.001	0.016 ± 0.002
$r_{protein\_syn}$	0.32 ± 0.02	0.23 ± 0.01	0.013 ± 0.001	0.013 ± 0.001
$r_{protein\_deg}$	0.35 ± 0.01	0.26 ± 0.01	0.014 ± 0.001	0.015 ± 0.002
$K_d(\mathbf{0})$	0.28 ± 0.02	0.21 ± 0.02	0.006 ± 0.000	0.005 ± 0.001
TF identity	0.29 ± 0.01	0.29 ± 0.02	0.008 ± 0.000	0.008 ± 0.001
Locus length	0.33 ± 0.01	0.36 ± 0.01	0.017 ± 0.001	0.026 ± 0.002

Table S3. Summary of mutations that replaced the resident genotype. Data is shown as mean ± SE over the 45 evolutionary replicates under selection for filtering out a spurious signal, with the signal allowed to regulate the effector directly. Without selection, each mutation would have probability 50% of replacing the resident; selection reduces this to around one in three at the beginning of the simulation, down to around one in four at the end. This high rate of accepting mutations after fitness has plateaued suggests significant nearly neutral evolution, i.e. that slightly deleterious mutations fix and are then compensated for. The estimated selection coefficient need only be 10^{-8} for a mutant to replace the resident, which can be easily occur for a slightly deleterious mutation through the error in fitness estimation (see Evolution Simulation in the main text). Single nucleotide substitutions are particularly prone to nearly neutral evolution, whereas changes to the consensus sequence recognized by a TF are under stronger stabilizing selection. Deletion and duplication mutations are the most common forms of substitution not because they are more likely to be accepted, but because they occur at higher mutation rates.

	Signal		TFs		Effector	
	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n
$r_{Act\_to\_Int}$	NA	NA	0.89	0.18	8.26	0.13
r_{mRNA_deg}	NA	NA	2.09	0.98	13.4	2.55
$r_{protein\_syn}$	NA	NA	1.51	8.03	43.1	62.4
$r_{protein\_deg}$	NA	NA	1.28	0.56	7.23	12.5
$K_d(0)$	0.68	0.002	0.67	0.009	NA	NA
Locus length	NA	NA	1.01	0.72	2.07	0.79

Table S4. Evolutionary constraint on parameters in AND-gated C1-FFLs. Adaptive AND-gated C1-FFLs are taken from the 25 high-fitness replicates evolved for filtering out a spurious signal, where the signal directly regulates the effector. For each replicate, we sample one of the last 10,000 evolutionary time steps, and then sample one AND-gated C1-FFL in that genotype, should there be more than one (or resample a time step for that replicate, if there are none). We then take the variance V_s of each C1-FFL parameter value across the 25 replicates. We repeat this sampling process 100 times (using the same 25 replicates) and take the mean in order to obtain a better estimator of the variance in each parameter value. We compare this by a comparable variance V_n given no selection. We obtain these from 30 evolutionary replicates under no selection (from **Fig. 6**), sampling parameter values from the signal, from one TF gene copy, and from one effector gene, without the requirement for C1-FFL presence. Variances are calculated for log-transformed parameter values, except for locus length. For locus length, we use the coefficient of variation rather than variance, i.e. we divide both variances by the square of the average locus length. The table also shows the how the parameter values M_s in adaptive AND-gated C1-FFLs differ from the expected value M_n given no selection. M_s and M_n are calculated as arithmetic means for locus length and as geometric means for all other parameters. The variance ratio is greater than 1 (indicating constraint), for all parameters except $K_d(0)$, where the ratio of mean parameter values indicates that $K_d(0)$ is nevertheless subject to strong directional selection. Effectors are more constrained than TFs, likely because the former are less redundant, having evolved fewer gene copies (4.7 on average for effectors vs. 8.6 for TFs). High degradation rates of effector mRNA and protein suggest selection to shorten the impact of transient expression in response to a short spurious signal (**Fig. S2**). High degradation rates of effector mRNA and protein are also seen in **Tables S5** and **S6**.

	Signal		Fast TFs		Slow TFs		Effector	
	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n
$r_{Act\_to\_Int}$	NA	NA	1.49	0.44	1.15	0.18	6.64	0.1
r_{mRNA_deg}	NA	NA	5.27	8.21	1.07	0.81	7.99	2.34
$r_{protein\_syn}$	NA	NA	2.10	16.2	1.09	4.96	139	57.8
$r_{protein\_deg}$	NA	NA	12.5	45.3	1.53	0.99	25.7	11.3
$K_d(0)$	0.65	0.005	0.30	0.004	0.18	0.007	NA	NA
Locus length	NA	NA	3.43	0.47	3.40	0.47	5.97	0.74

Table S5. Evolutionary constraint on parameters in isolated AND-gated diamonds. V_n, V_s, M_n, and M_s are defined in the same way as in **Table S4**, and are calculated from 18 high-fitness evolutionary replicates (**Fig. 7B**) in which isolated AND-gated diamonds occur in at least 100 of the last 10,000 evolutionary steps. Because they occur at low rates, we sample 50 times per evolutionary replicate, instead of 100 times as in **Tables S4** and **S6**. There is more constraint on fast TFs than on slow TFs. The fast TFs usually have more gene copies than the slow TFs, therefore redundancy is not the reason for this difference in constraint. As seen for the C1-FFLs in **Table S4**, effectors are more constrained than either TF, $K_d(0)$ shows strong selection for high affinity combined with high variance, and effectors evolve rapid degradation. Fast TFs exhibit not just fast protein degradation (which was used for their identification), but also fast mRNA degradation.

	Signal		Signal-regulated TFs		TF-regulated TFs		Effector	
	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n	V_n / V_s	M_s / M_n
$r_{Act\_to\_Int}$	NA	NA	2.16	0.33	1.03	0.26	6.81	0.13
r_{mRNA_deg}	NA	NA	10.8	8.5	1.40	0.74	12.4	2.36
$r_{protein\_syn}$	NA	NA	4.34	24.9	2.35	9.83	119	58.6
$r_{protein\_deg}$	NA	NA	73.6	49.4	1.50	0.34	34.1	9.92
$K_d(\mathbf{0})$	0.51	0.005	0.29	0.009	0.24	0.002	NA	NA
Locus length	NA	NA	2.52	0.71	2.45	0.71	3.35	0.73

Table S6. Evolutionary constraint on parameters in isolated AND-gated C1-FFLs. V_n, V_s, M_n, and M_s are defined in the same way as in **Table S4**, and are calculated from 12 high-fitness evolutionary replicates (**Fig. 7B**) evolved when the signal cannot directly regulate the effector, and in which isolated AND-gated C1-FFLs occur in at least 1,000 out of the last 10,000 evolutionary steps. Note that the signal-regulated TFs, which are identified via network topology, also have high protein degradation rates, as is used to identify their fast TF counterparts in diamonds – they can thus be seen as a kind of fast TF. Consistent with results on C1-FFLs when direct regulation is allowed (**Table S4**) and results on isolated AND-gated diamonds (**Table S5**), effectors are more constrained than signal-regulated (fast) TFs, which are more constrained than TF-regulated (slow) TFs, despite an opposite trend in gene copy number. Note that selection promotes fast mRNA and protein degradation in fast TFs, but promotes slow degradation of slow TFs; this result is also found more weakly in **Table S5**.

Supplementary Figures

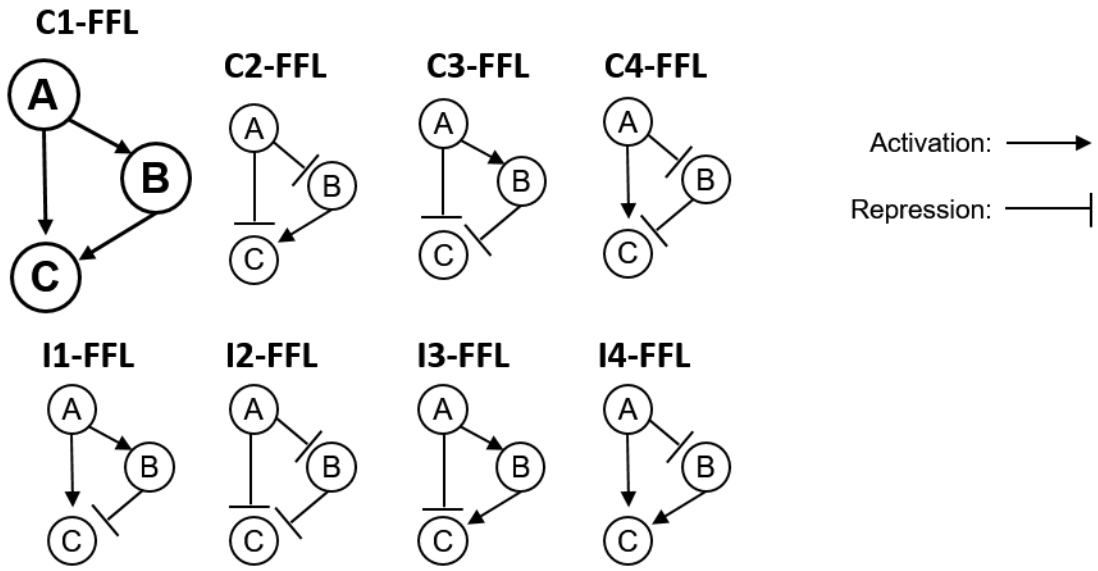


Fig. S1. Feed-forward loops come in eight subtypes. TF A and TF B can activate (indicated by arrows) or repress (indicated by bars) expression of the effector C as well as other TFs. Auto-regulation is allowed, but not shown. Following Milo et al. (2002), we exclude the case in which A and B regulate one another, rather than treating this case as two overlapping FFLs. C stands for coherent and I for incoherent.

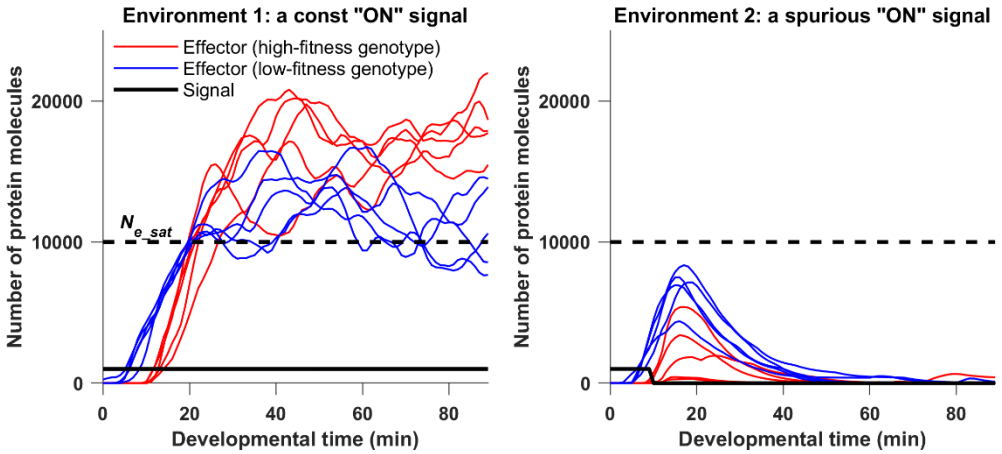


Fig. S2 Examples of evolved phenotypes under selection for filtering out a short spurious signal. The figure shows trajectories of the effector protein in one randomly chosen high-fitness replicate (red) and one randomly chosen low-fitness replicate (blue), as defined in **Fig. 4A**. The genotype of the final evolutionary step is used, and other genotypes were confirmed to behave similarly. Each genotype is illustrated by 5 replicate developmental simulations in each of the two environments. The high-fitness genotype has a longer delay followed by more rapid response given a consistent signal, with this longer delay reducing but not eliminating effector expression given a short spurious signal. The signal is allowed to directly regulate the effector in these simulations. The burn-in period is not shown, with developmental time zero corresponding to the moment the signal is turned on. Among developmental replicates of the same genotype, the concentration at a given time usually has an approximately log-normal distribution, but in environment 2 the distribution has two modes after the spurious signal turns off. One mode corresponds to expression at the basal rate, the other to a burst of expression that has yet to turn off. Because of this bimodality, we plot sample trajectories rather than mean concentration over many replicates.

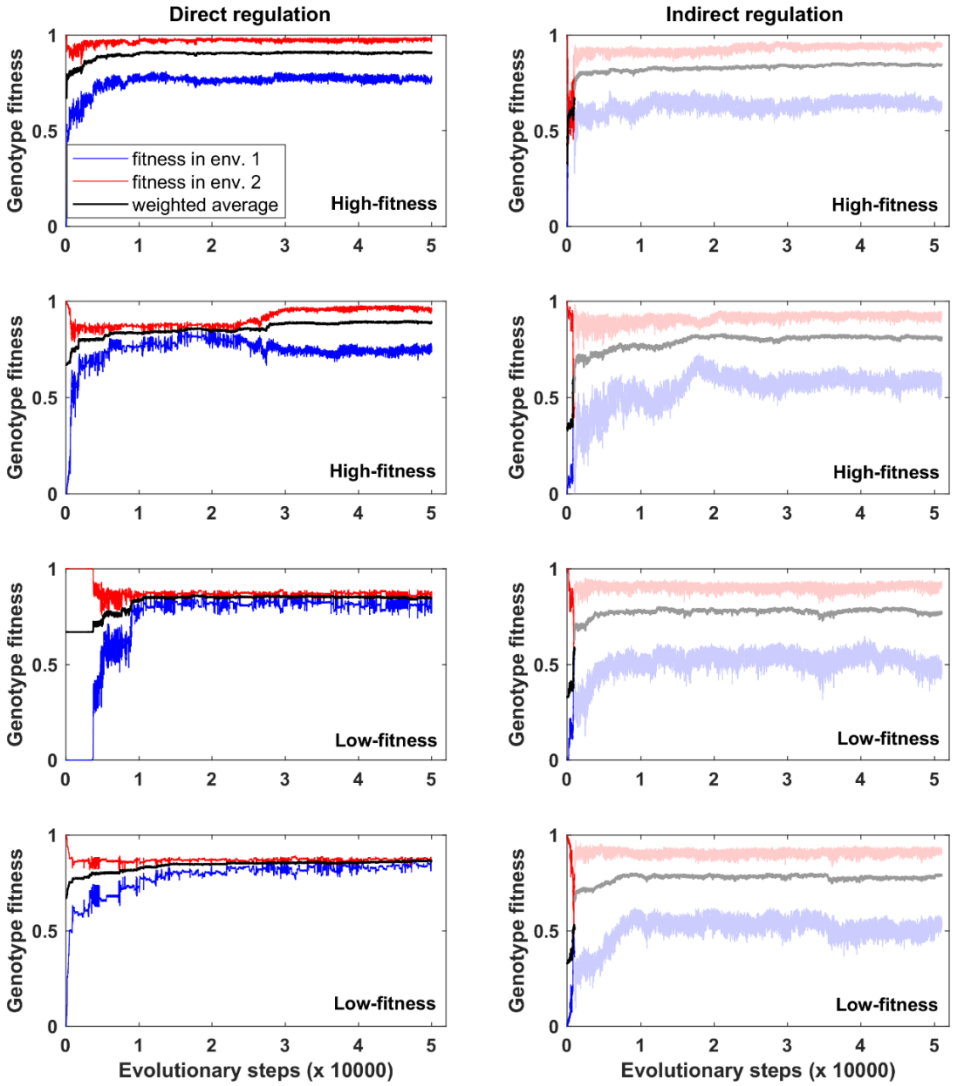


Fig. S3 Representative fitness trajectories under selection to filter out short spurious signals. Left panels: The signal is allowed to directly regulate the effector genes. Panels 1 and 3 correspond to the two genotypes shown in **Fig. S2**. Right panels: the signal cannot directly regulate the effector genes. Average fitness (black) is a weighted average of the blue and red trajectories, with environment 2 (where the signal is spurious) being considered twice as common as environment 1 (where the signal is sustained and real). When the signal cannot directly regulate the effector genes, evolutionary simulations begins with a burn-in phase that lasts 1000 evolutionary steps (see Evolutionary Simulation in the Main Text). We show the burn-in phase in undiluted color, and dilute color after burn-in. Most replicates quickly reach a stable fitness plateau (first and third rows). Certain replicates can be temporarily trapped at a low fitness plateau (second and third rows on the left).

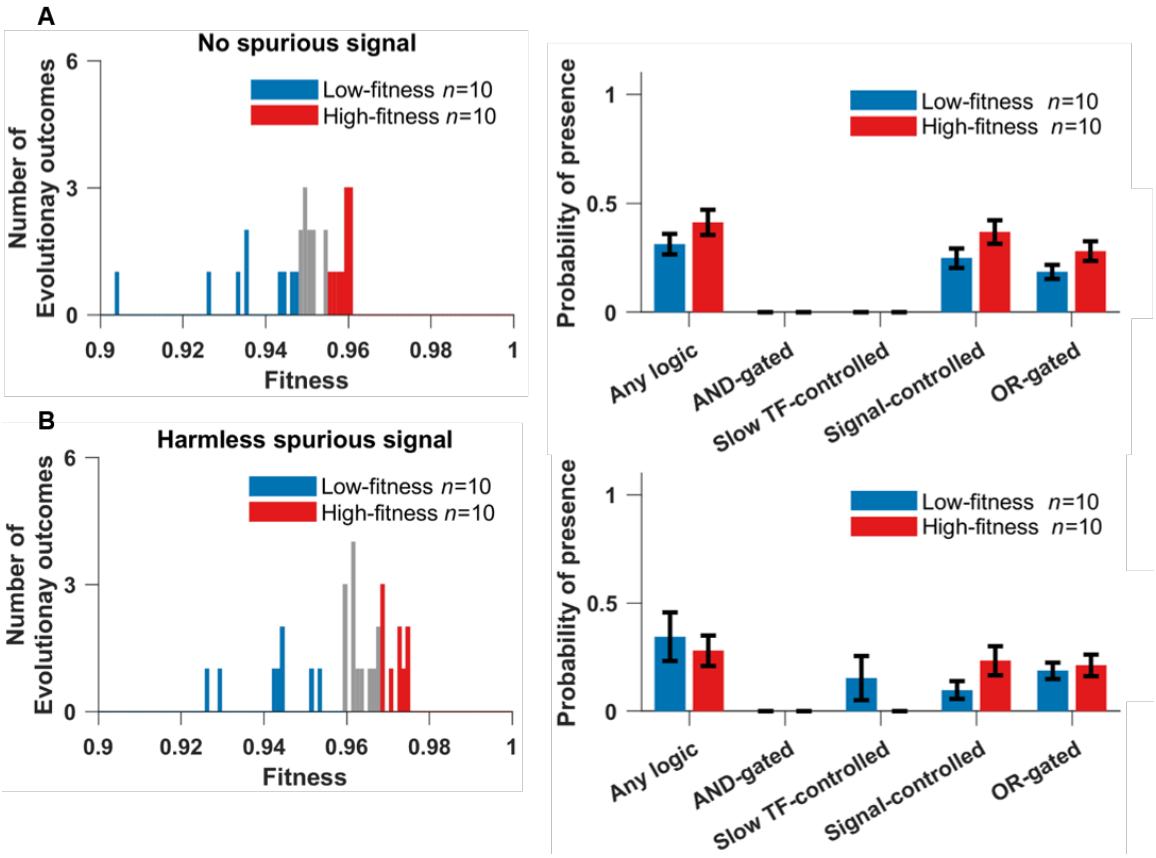


Fig. S4 Genotypes evolved under control selective conditions: (A) "no spurious signal", and (B) "harmless spurious signal". There is no clear evidence of a multimodal distribution of fitness outcomes among replicates (left), and C1-FFLs occur equally in the 10 genotypes of the highest fitness vs. the 10 genotypes of the lowest fitness (right), and so the entire distribution (left) was used to produce **Fig. 6**. Data are shown as mean±SE over evolutionary replicates.

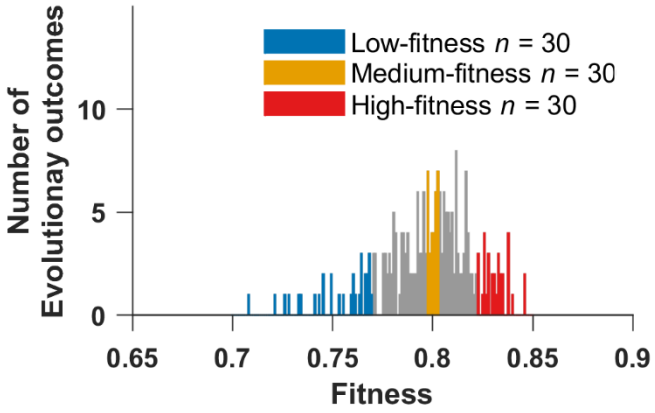
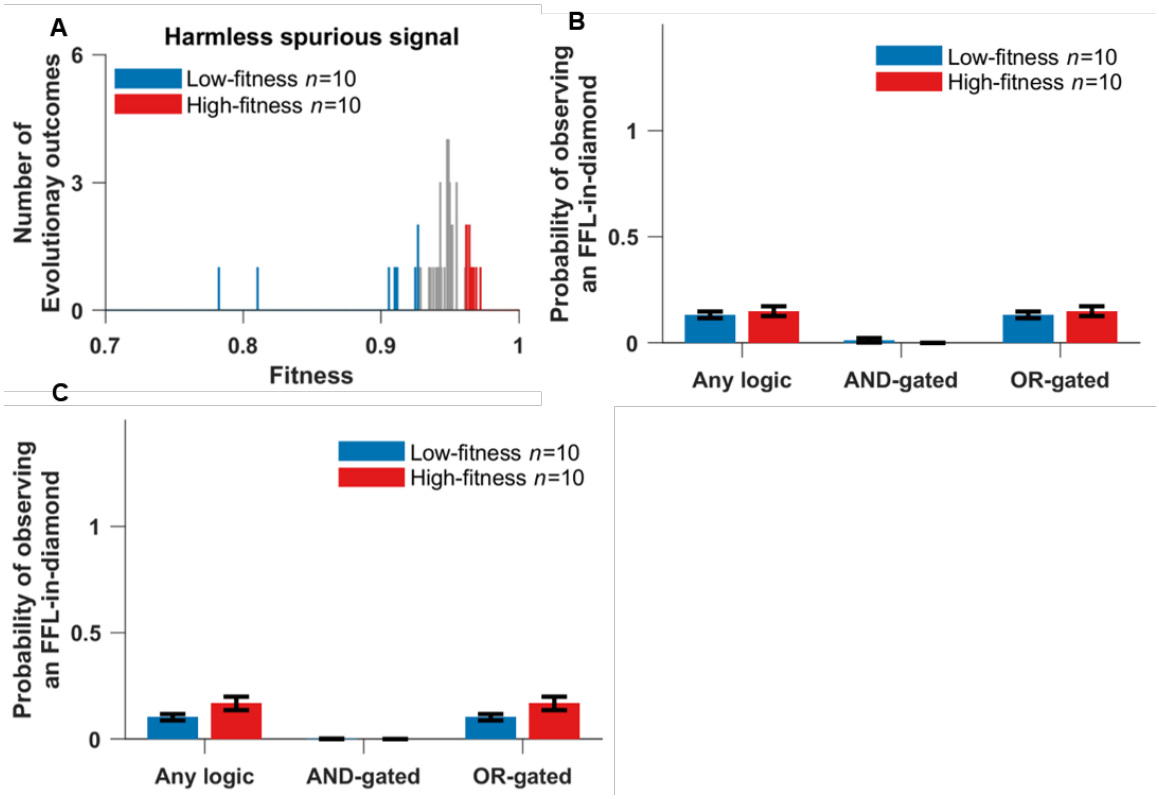


Fig. S5 Fitness distribution of 258 evolutionary replicates under selection for filtering out short spurious signals, when the signal cannot directly regulate the effector. The fitness of a replicate is the average genotype fitness over the last 10,000 evolutionary steps. Colors indicate replicates analyzed elsewhere.



139

140 **Fig. S6 Evolution when responding to a spurious signal is harmless, when the signal is not**

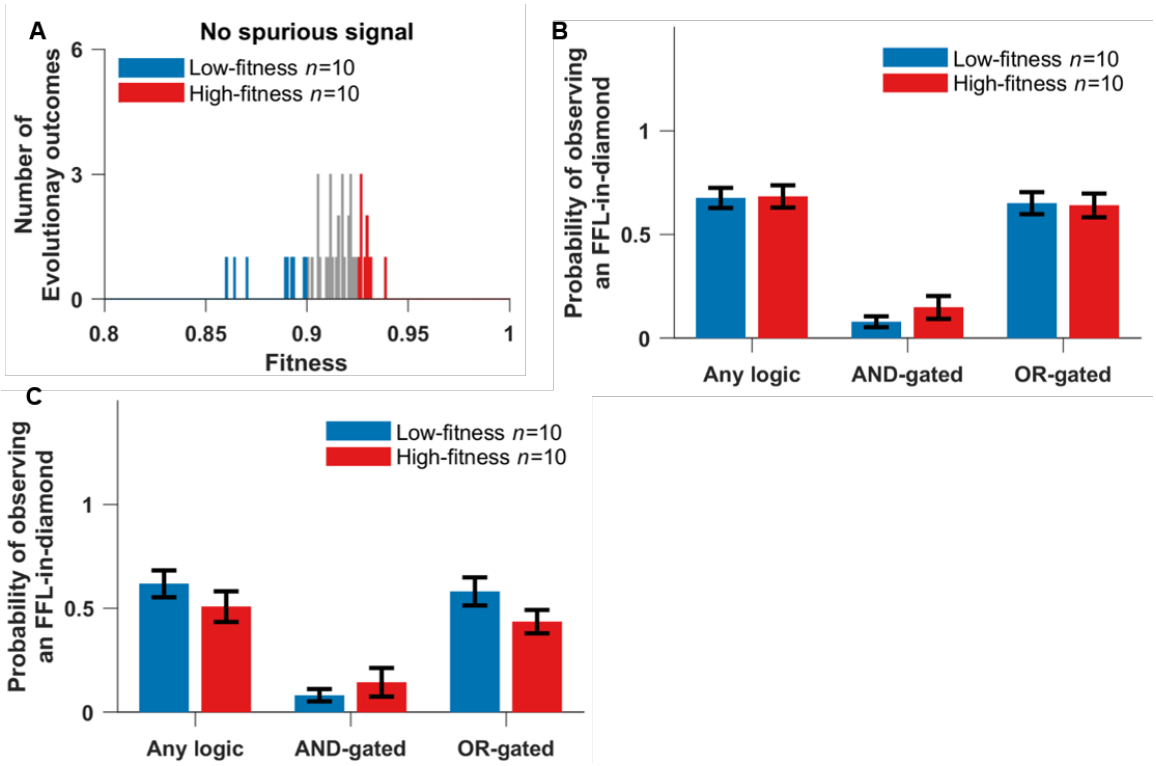
141 **allowed to directly regulate the effector. (A)** Fitness distribution of 50 replicate simulations.

142 The occurrence of both **(B)** FFL-in-diamonds and **(C)** isolated diamonds were similar in the 10

143 genotypes with the highest fitness vs. in 10 genotypes with the lowest fitness. Weak (two-

144 mismatch) TFBSs are included when scoring motifs. Data are shown as mean±SE over replicates.

145 Isolated C1-FFLs rarely evolve under this condition, therefore their occurrence is not plotted.



146

147 **Fig. S7 Evolution when there is no spurious signal, when the signal is not allowed to directly**

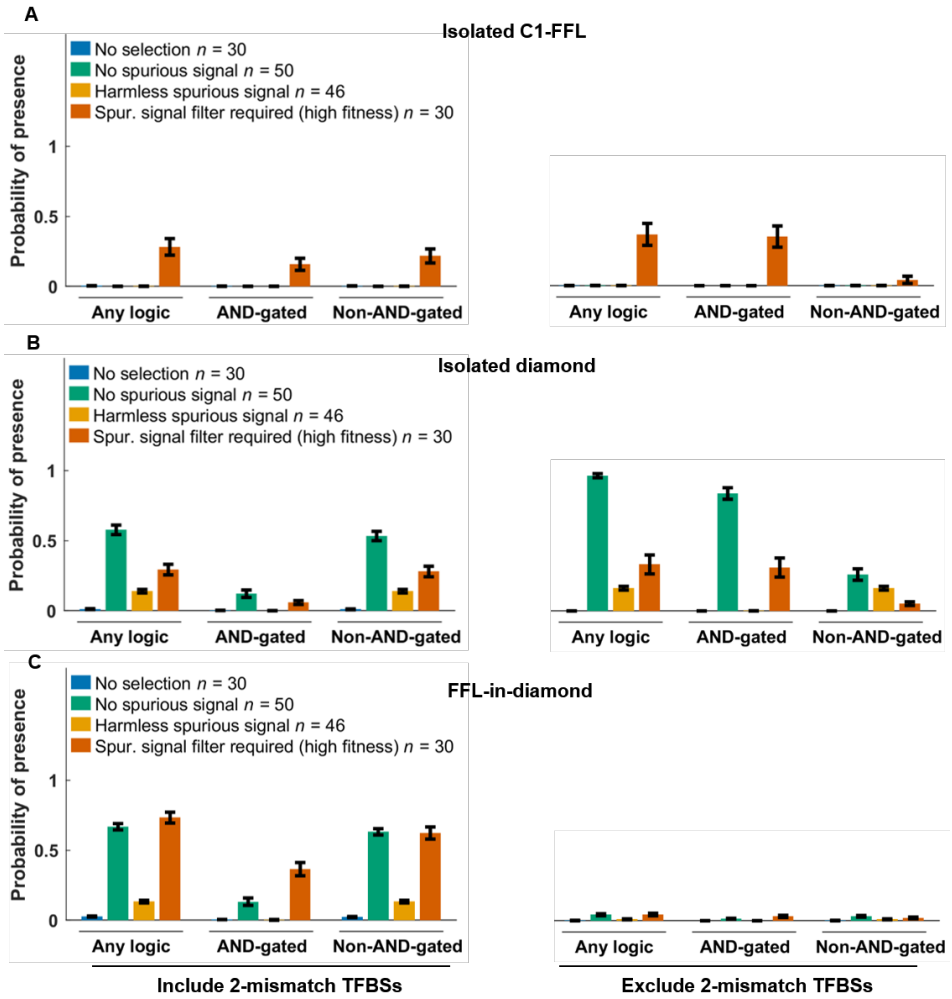
148 **regulate the effector. (A)** Fitness distribution of 46 replicate simulations. The occurrence of both

149 **(B)** FFL-in-diamonds and **(C)** isolated diamonds were similar in the 10 genotypes with the highest

150 fitness vs. in the 10 genotypes with the lowest fitness. Weak (two-mismatch) TFBSs are included

151 when scoring motifs. Data are shown as mean±SE over replicates. Isolated C1-FFLs rarely evolve

152 under this condition, therefore their occurrence is not plotted.



153 **Fig. S8 Selection for filtering out a short spurious signal is the primary way to evolve AND-**
154 **gated C1-FFLs (A), but AND-gated isolated diamonds also evolve in the absence of spurious**
155 **signals (B).** The signal is not allowed to directly regulate the effector, and the right panels of **(A)**
156 and **(B)** are identical to **Fig. 10**. When scoring motifs, we either include (left) or exclude (right) all
157 two-mismatch TFBSs in the cis-regulatory sequences of intermediate TF genes and effector
158 genes. We excluded "no regulation" (**Fig. 2**) diamonds from the "Any logic" and "Non-AND-
159 gated" tallies in **(B)**; this was necessary because of their high occurrence due to duplication and
160 divergence of intermediate TFs. See Section 11 for the calculation of y-axis. Data are shown as
161 mean±SE over evolutionary replicates.

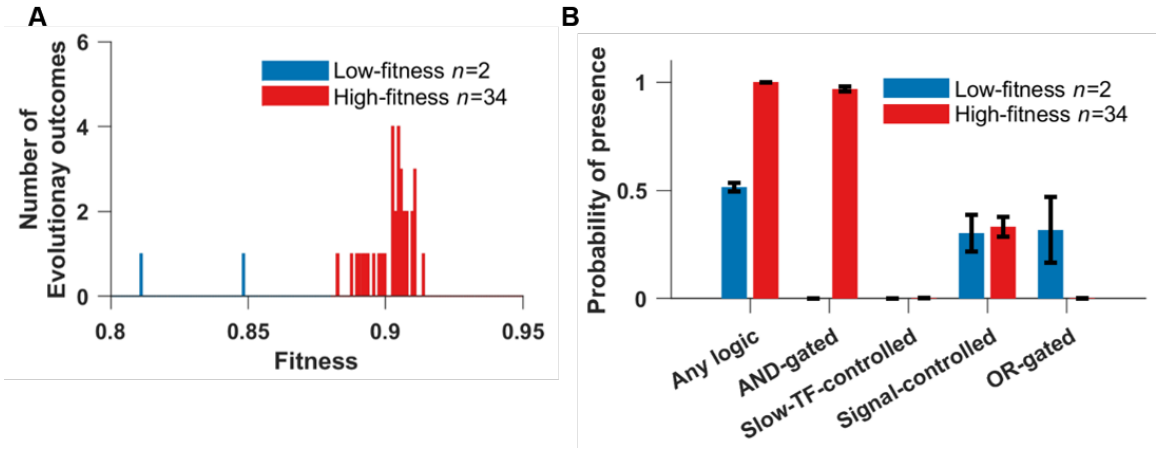


Fig. S9 After removing cost of gene expression, AND-gated C1-FFLs are still associated with a successful response to selection for filtering out a short spurious signal. The signal can directly regulate the effector genes. **(A)** We arbitrarily divide the 36 replicate simulations into high-fitness (red) and low-fitness (blue) groups. **(B)** The high-fitness replicates still evolve AND-gated C1-FFLs. Bars are mean±SE of the occurrence over replicate evolutionary simulations.

Supporting Text

1. TF binding

Transcription of each gene is controlled by TFBSs present within a 150-bp cis-regulatory region, corresponding to a typical yeast nucleosome-free region within a promoter (Yuan et al. 2005). The perfect TFBS for a typical yeast TF has information content equivalent to 13.8 bits (Wunderlich & Mirny 2009); this means that in a simplified model of binding where only one of the four nucleotides is a good match at each site, ~7 bp are recognized as an optimal consensus binding site. Maerkl & Quake (2007) reported that the TFBSs of two yeast TFs, Pho4p and Cbf1p, can have up to 2 mismatched sites within their 6 bp consensus binding sequence, while still binding the TF above background levels (Maerkl & Quake 2007). Our model therefore tracks TFBSs with up to 2 mismatches. This low information content implies a higher density of TFBSs within our cis-regulatory regions than our algorithm was able to handle, so we instead assigned each TF an 8-bp consensus sequence. Two TFs cannot simultaneously occupy overlapping stretches (Fig. S10), which we assume extend beyond the recognition sequence to occupy a total of 14 bp (Zhu & Zhang 1999); this captures competitive binding. The consequences of hindrance between TFBSs for the regulation of effector gene expression are shown in **Fig. 2**.

Fig. S10 TFs (white boxes) recognize 8 bp (red) sites while occupying and thus excluding other TFs from a 14 bp long space. TFs are assumed to bind in either orientations (Sharon et al. 2012). The sequence on the left allows simultaneous binding but that on the right does not.

We denote the dissociation constant of a TFBS with m mismatches as $K_d(m)$. Sites with m>3 mismatches are assumed to still bind at a background rate equal to m=3 mismatches, with dissociation constant $K_d(3) = 10^{-5}$ mole/liter (Maerkl & Quake 2007) for all TFs. We assume that each of the last three base pairs makes an equal and independent additive contribution $\Delta G_{bp} < 0$ to the binding energy (Benos et al. 2002): although not always true, this approximates average behavior well (Maerkl & Quake 2007). We ignore cooperativity in binding. Dissociation constants of eukaryotic TFs for perfect TFBSs can range from 10^{-5} mole/liter (Park et al. 2004) to 10^{-11} mole/liter (Nalefski et al. 2006). We initialize each TF with its own value of $\log_{10}(K_d(0))$ sampled from a uniform distribution between -6 and -9, with mutation capable of further expanding this range, subject to $K_d(0) < 10^{-5}$ mole/liter. Substituting m=0 and m=3 into

$$\Delta G_m = -RTlnK_d(m) = \Delta G_0 - \min(m, 3)\Delta G_{bp},$$

we can solve for ΔG_{bp} and ΔG_0, and thus obtain $K_d(1)$ and $K_d(2)$ (the dissociation constants for TFBS with one and two mismatches, respectively).

Because TFs bind non-specifically to DNA at a high background rate, each nucleosome-free stretch of 14 bp can be considered to be a non-specific binding site (NSBS). A haploid *S. cerevisiae* genome is 12 Mb, 80% of which is wrapped in nucleosomes (Lee et al. 2007), yielding approximately 10^6 potential non-specific binding sites (NSBSs). In a yeast nucleus of volume 3×10^{-15} liters, the NSBS concentration is of order 10^{-4} mole/liter. To find the concentration of free TF [TF] in the nucleus given a total nucleic TF concentration of C_{TF}, we consider

$$K_d = \frac{[\text{binding\_site}][\text{TF}]}{[\text{binding\_site} \cdot \text{TF}]},$$

in the context of NSBSs, substitute [TF·NSBS] with C_{TF} - [TF], and solve for

$$[\text{TF}] = \frac{K_d(3)}{K_d(3)+[\text{NSBS}]} C_{TF} = \frac{10^{-5}}{10^{-5}+10^{-4}} C_{TF} \approx 0.1 C_{TF}.$$

Thus, about 90% of total TFs are bound non-specifically, leaving about 10% free. The relatively small number of specific TFBSs is not enough to significantly perturb the proportion of free TFs, and so for the specific TFBSs with m<3 that are of interest in our model, we simply use $\widehat{K_d}(m)$ = $10K_d(m)$ to account for the reduction in the amount of available TF due to non-specific binding. We also convert $\widehat{K_d}$ from the units of mole/liter in which K_d is estimated empirically to the more convenient molecules/nucleus. The rescaling factor r for which $\widehat{K_d}$(in molecule/nucleus) = $r\widehat{K_d}$ (in mole/liter) is 3×10^{-15} liter/nucleus $\times$ 6.02×10^{23} molecule/mole = 1.8×10^{9} molecule cell^{-1} liter mole^{-1}. Taken together, $\widehat{K_d}$(molecule/nucleus) = $10rK_d$ (mole/liter), where the factor 10 accounts for non-specific TF binding.

2. TF occupancy

Here we calculate the probability that there are A activators and R repressors bound to a given cis-regulatory region at a given moment in developmental time. First we note that if we consider TF i binding to TFBS j in isolation from all other TFs and TFBSs, Eq. S1 gives us probability of being bound:

$$P_b(j) = 1 - P_u(j) = \frac{c_i}{\widehat{K_d}+c_i}. \quad \text{(S1)}$$

Let $P_{A,R}^{(n)}$ be a term proportional (for a given value of n) to the combined probability of all binding configurations in which exactly *A* activators and *R* repressors are bound to the first *n* binding sites along the cis-regulatory sequence. We calculate $P_{A,R}^{(n)}$ recursively, considering one additional TFBS at each step. Note that if two different TFs bind to exactly the same location on a cis-regulatory region, we treat this as two TFBSs, not as one, and treat first one and then the other in our recursive algorithm.

Consider the case where the (*n*+1)th binding site belongs to an activator. The case where this activator is not bound contributes $P_{A,R}^{(n)} P_u(n+1)$ to $P_{A,R}^{(n+1)}$. If it is bound, then we must also take into account that the (*n*+1)th binding site overlaps (partially or completely) with the last $H \geq 0$ sites, and so contributes $P_{A-1,R}^{(n-H)} P_b(n+1) \prod_{j=n-H+1}^{n} P_u(j)$. Taken together,

$$P_{A,R}^{(n+1)} = P_{A,R}^{(n)} P_u(n+1) + P_{A-1,R}^{(n-H)} P_b(n+1) \prod_{j=n-H+1}^{n} P_u(j).$$

Similarly, if the (*n*+1)th site belongs to a repressor, we have

$$P_{A,R}^{(n+1)} = P_{A,R}^{(n)} P_u(n+1) + P_{A,R-1}^{(n-H)} P_b(n+1) \prod_{j=n-H+1}^{n} P_u(j).$$

By definition, $P_{A,R}^{(n)} = 0$ for binding configurations that are impossible, e.g. those with negative *A* or negative *R*. We initialize the recursion at *n* = 0, where the only valid binding configuration is for *A* = *R* = 0, i.e. $P_{0,0}^{(0)} = 1$. At *n* = 1, $P_{0,0}^{(1)} \propto P_u(1)$, and if the binding site belongs to an activator, $P_{1,0}^{(1)} \propto P_b(1)$; otherwise, $P_{0,1}^{(1)} \propto P_b(1)$. For *N* = 1, the two probabilities sum to 1 and

normalization is unnecessary. For higher values of $N = N_A + N_R$ TFBSs, we normalize $P_{A,R}^{(N)}$ at the end of the recursion by dividing by $\sum_{A=0}^{N_A}\sum_{R=0}^{N_R} P_{A,R}^{(N)}$ to get the probability of binding configurations that include exactly *A* activators and *R* repressors.

3. $r_{Act\_to\_Int}$

Transcription initiation over an interval of time r_{transc_init} is proportional to the proportion of time spent in the Active state. Assuming a steady state between Repressed, Intermediate, and Active states, as a function of current TF concentrations, we have:

$$\frac{r_{transc\_init}}{r_{max\_transc\_init}} = \frac{r_{Int\_to\_Act}}{r_{Int\_to\_Act}+r_{Act\_to\_Int}} P_{Int\_or\_Act}, \quad \text{(S2)}$$

where $P_{Int\_or\_Act}$ is the probability a gene is at Intermediate or Active. We set $r_{max\_transc\_init}$ (the rate of transcription given 100% Active state) to 6.75 min^{-1}, based on the corresponding rate when a model of the *PHO5* promoter is fit to data (Brown et al. 2013). In this model fit, the constitutively expressed *PHO5* promoter is free of nucleosomes 80% of the time, i.e. $P_{Int\_or\_Act} = 0.8$. We take these two values as universal for constitutively expressed genes, and assume that variation in $r_{Act\_to\_Int}$ is responsible for variation in r_{transc_init}. To identify a set of constitutively expressed genes, we identified 225 genes that have mRNA production rate of at least 0.5 molecule min^{-1} from genome-wide measurements (Pelechano et al. 2010); this threshold corresponds to low H2A.Z occupancy (Guillemette et al. 2005). We set r_{transc_init} to the production rate of mRNA of these 225 genes, and solve for gene-specific $r_{Act\_to\_Int}$ from Eq. S2. We fit the solutions to a log-normal distribution and arrive at $10^{N(1.27,\ 0.226)}$ min^{-1}.

To initialize values of $r_{Act\_to\_Int}$ for each gene, we sample from this distribution. We also set lower and upper bounds for allowable values; if either the initial sample or subsequent mutation put $r_{Act\_to\_Int}$ beyond these bounds, we set the value of $r_{Act\_to\_Int}$ to equal to boundary value. We set the lower bound for $r_{Act\_to\_Int}$ at 0.59 min^{-1}, half the minimum of the values inferred from the set of 225 genes. To set an upper bound, we use the low H2A.Z occupancy bound of $r_{transc\_init} = 0.5$, which gives a solution of 32.34 min^{-1}; we double this to set the upper bound as 64.7 min^{-1}.

4. Transcription delay times

Yeast protein lengths fit a log-normal distribution of $10^{N(2.568,\ 0.34)}$ amino acids (from the Saccharomyces Genome Database (SGD Project), excluding mitochondrial proteins). We sample ORF length *L* from this distribution. To constrain the values of *L*, we set a lower bound of 50 amino acids and an upper bound of 5000 amino acids; the longest protein in SGD is 4910 amino acids. If either initialization or mutation put *L* beyond these bounds, we set the value of *L* to the boundary value.

With an mRNA elongation rate of 600 codon/min (Larson et al. 2011; Hocine et al. 2013), it takes *L* / 600 minutes to transcribe the ORF of an mRNA. Also including time for transcribing UTRs and for transcription termination, and ignoring introns for simplicity, it takes 290 seconds to complete transcription of the yeast *GLT1* gene (Larson et al. 2011), whose ORF is 6.4kb. Putting the two together, we infer that transcribing the UTRs and terminating transcription takes around 1 minute for *GLT1*. Generalizing to assume that transcribing UTRs and terminating transcription takes exactly 1 minute for all genes, producing an mRNA from a gene of length *L* takes 1 + *L* / 600 minutes.

5. Translation delay times and $r_{protein\_syn}$

We model a second delay between the completion of a transcript and the production of the first protein from it. The delay comes from a combination of translation initiation and elongation; it ends when the mRNA is fully loaded with ribosomes all the way through to the stop codon and the first protein is produced. We ignore the time required for mRNA splicing; introns are rare in yeast (Dujon 1996). mRNA transportation from nucleus to cytosol, which is likely diffusion-limited (Niño et al. 2013; Smith et al. 2015), is fast even in mammalian cells (Mor et al. 2010) let alone much smaller yeast cells, and the time it takes is also ignored. The median time in yeast for initiating translation is 0.5 minute (Table 1 in Siwiak et al. 2010), and the genomic average peptide elongation rate is 330 codon/min (Siwiak et al. 2010). After an mRNA is produced, we therefore wait for $0.5 + L / 330$ minutes, and then model protein production as continuous at a gene-specific rate $r_{protein\_syn}$.

To calculate $r_{protein\_syn}$, we combine the gene-specific ribosome densities D along the mRNAs and the gene-specific peptide elongation rates E, both measured in yeast (Siwiak et al. 2010). The values of DE across yeast genes fit the log-normal distribution $10^{N(0.322,\ 0.416)}$ molecule mRNA^{-1} min^{-1}; we initialize $r_{protein\_syn}$ for each gene by sampling from this distribution. We set the lower bound for $r_{protein\_syn}$ at half the minimum observed value of DE (4.5×10^{-3} molecule mRNA^{-1} min^{-1}). The upper bound corresponds to an mRNA fully occupied by rapidly moving ribosomes. Each ribosome occupies about 10 codons (Siwiak et al. 2010), and the peptide elongation rate can be as high as 614 codon/min (Waldron et al. 1977). If ribosomes are packed closely together at 10 codons apart, a protein comes off the end of production in the time taken to elongate 10 codons, i.e. proteins are produced at 61.4 molecules per minute. If either initialization or mutation put $r_{protein\_syn}$ beyond these bounds, we set the value of $r_{protein\_syn}$ to the boundary value.

6. mRNA and protein decay rates

We fit the log-normal distribution $10^{N(-1.49,\ 0.267)}$ min^{-1} to yeast mRNA degradation rates (Wang et al. 2002), and initialize the mRNA degradation rate r_{mRNA_deg} for each gene by sampling from this distribution. We set lower and upper bounds for r_{mRNA_deg} at half the minimum and twice the maximum observed values (7.5×10^{-4} min^{-1} and 0.54 min^{-1}), respectively. If either initialization or mutation put r_{mRNA_deg} beyond these bounds, we set the value of r_{mRNA_deg} to the boundary value.

Expressing the estimated half-lives of yeast proteins (Belle et al. 2006) in terms of protein degradation rates, they fit the log-normal distribution $10^{N(-1.88,\ 0.56)}$ min^{-1}; we initialize gene-specific protein degradation rates $r_{protein\_deg}$ by sampling from this distribution. We ignore the additional reduction in protein concentration due to dilution as the cell grows and thus increases in volume. We set lower and upper bounds for $r_{protein\_deg}$ at half the minimum and twice the maximum observed degradation rate (3.0×10^{-6} min^{-1} and 0.69 min^{-1}), respectively. If either initialization or mutation put $r_{protein\_deg}$ beyond these bounds, we set the value of $r_{protein\_deg}$ to the boundary value.

7. Simulation of gene expression

Our algorithm is part-stochastic, part-deterministic. We use a Gillespie algorithm (Gillespie 1977) to simulate stochastic transitions between Repressed, Intermediate, and Active chromatin states, and to simulate transcription initiation and mRNA decay events. We refer to these as "Gillespie events". The completion of transcription to produce a complete mRNA, and subsequent ribosomal loading onto the mRNA, are referred to as "fixed events" (they require fixed times of $1 + L / 600$ minutes and $0.5 + L / 330$ minutes, respectively). Scheduled changes in

the strength of the external signal are also fixed events. Protein production and degradation are described deterministically with ODEs, and updated frequently in order to recalculate TF concentrations and hence chromatic transition rates. Updates occur at the time of Gillespie and fixed events, and also in between.

The total rate of all Gillespie events is

$$r_{total} = \sum_i^{Rep} r_{Rep\_to\_Int\_i} + \sum_i^{Int}(r_{Int\_to\_Rep\_i} + r_{Int\_to\_Act\_i}) + \sum_i^{Act}(r_{Act\_to\_Int\_i} + r_{transc}) + \sum_i^{genes} r_{mRNA\_deg\_i} N_{mRNA\_i},$$

where *Rep*, *Int*, and *Act* are the numbers of gene copies in our haploid model that are in the Repressed, Intermediate, and Active chromatin states, respectively, and N_{mRNA_i} is the number of completely transcribed mRNA molecules from gene *i*. We only simulate degradation of full transcribed mRNA, and not that of mRNA that are still being transcribed, because the latter are already captured implicitly by $r_{max\_transc\_init}$, which is based on mRNAs that complete transcription (Brown et al. 2013). Once an mRNA finishes transcription, it is subjected to degradation regardless of whether ribosome loading is complete.

The waiting time Δt before the next Gillespie event is

$$\Delta t = \frac{x}{r_{total}},$$

where x is random number drawn from an exponential distribution with mean 1. Which Gillespie event takes place next is sampled only if a different update does not happen first. If a fixed event is scheduled to happen first at $\Delta t_1 < \Delta t$, we advance time by Δt_1, update the state of the cell, and calculate a new r_{total}'. Since the cellular activity has been going on with the old rate r_{total} for Δt_1, the remaining "labor" required to trigger the Gillespie event planned earlier is reduced. The new waiting time, $\Delta t'$, to trigger the planned Gillespie event is

$$\Delta t' = \frac{x - r_{total}\Delta t_1}{r_{total}'}.$$

Gene duplication creates $n \geq 1$ genes producing the same protein, where each copy i might have diverged in their production rate $r_{protein\_syn\_i}$ and degradation rate $r_{protein\_deg\_i}$. Complete proteins are produced continuously once an mRNA molecule is fully loaded with ribosomes, which occurs 0.5 + L / 330 minutes after transcription is complete – the concentration of such molecules is denoted $N_{mRNA\_aft\_delay\_i}(t)$. Total protein concentration obeys:

$$N'_{protein}(t) = \sum_i^n \left(r_{protein\_syn\_i} N_{mRNA\_aft\_delay\_i}(t) - r_{protein\_deg\_i} N_{protein\_i}(t) \right). \quad \text{(S3)}$$

Protein concentrations are updated using a closed-form integral of Eq. S3

$$N_{protein}(t_1) = \sum_i^n \left(\frac{r_{protein\_syn\_i} N_{mRNA\_aft\_delay\_i}}{r_{protein\_deg\_i}} + \left(N_{protein\_i}(t_0) - \frac{r_{protein\_syn\_i} N_{mRNA\_aft\_delay\_i}}{r_{protein\_deg\_i}} \right) e^{-r_{protein\_deg\_i}(t_1 - t_0)} \right) \quad \text{(S4)}$$

399 with this expression updated every time a Gillespie or fixed event at time t_1 changes the value of
400 $N_{mRNA\_aft\_delay\_i}$.
401
402 In between updates, values of P_A, P_R, $P_{A\_no\_R}$, and $P_{notA\_no\_R}$, and hence chromatin transition rates,
403 are calculated under the approximation of constant $N_{protein}$. Additional updates, above and
404 beyond fixed and Gillespie events, are performed in order to ensure that chromatin transition
405 rates do not change too dramatically from one update to the next. We use a target of D = 0.01
406 for the amount of change tolerated in the values of P_A, P_R, $P_{A\_no\_R}$, and $P_{notA\_no\_R}$, in order to
407 schedule updates after time Δt^*, which are triggered when neither a Gillespie event nor a fixed
408 event occurs before this time has elapsed, i.e. when $\Delta t^* < \Delta t_1$ and $\Delta t^* < \Delta t$.
409
410 There is the greatest potential for large changes after an update that changes the value of
411 $N_{mRNA\_aft\_delay\_i}$. In this case, we use Eq. S1 to solve for the time interval for which the probability
412 that TF i would be bound to a single perfect and non-overlapping TFBS would change by D, by
413 choosing $\Delta t^* > 0$ that satisfies
414

415 $$\left| \frac{N_i(t)}{N_i(t)+K^*_{d,i}(0)} - \frac{N_i(t+\Delta t^*)}{N_i(t+\Delta t^*)+K^*_{d,i}(0)} \right| = D. \quad \text{(S5)}$$

416
417 A solution for Δt may not exist, e.g. if the concentration of TF i is decreasing but $P_b(t_2) < D$. In
418 such cases, we set Δt^* to infinity.
419
420 When the previous update does not change any $N_{mRNA\_aft\_delay\_i}$ values, then we modify Δt^*
421 adaptively. Let d be the maximum of ΔP_A, ΔP_R, $\Delta P_{A\_no\_R}$, and $\Delta P_{notA\_no\_R}$ during the last update. We
422 then schedule an update at

$$\Delta t^{*\prime} = \frac{D}{d}\Delta t^{*}. \quad \text{(S6)}$$

After an update that changes the value of $N_{mRNA\_aft\_delay\_i}$, we use the smaller value from Eqs. S5 and S6. These additional update times are discarded and recalculated when a Gillespie or fixed event occurs first.

In **Fig. S11**, we see that simulations rarely exceed our target of D=0.01, and do so only modestly.

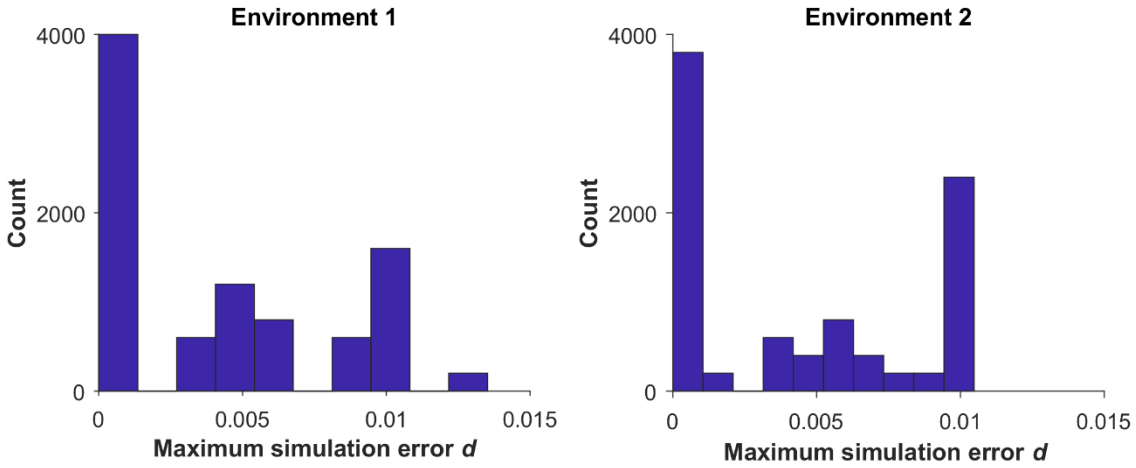


Fig. S11 Our updating algorithm is able to limit simulation errors. The distribution across 9,000 simulations of the maximum value of d over the course of development. For each of the 45 evolutionary replicates in **Fig. 4**, we run 200 simulations of development of the final evolved genotype. These genotypes were the outcome of evolution under selection for filtering out short spurious signals, in which direct regulation of the effector by the signal is not allowed. In environment 1 a genotype responds to a constant "ON" signal and in environment 2 it responds to a short spurious signal (**Fig. 3**).

8. Cost of gene expression

The cost of gene expression comes from some combination of the act of expression and from the presence of the resulting gene product. Yeast cells with plasmids carrying fast-degrading GFP had as much growth impairment as those carrying wild-type GFP (Fig. 3 of Kafri et al. 2016), suggesting that the former cost dominates. Universal costs stemming from the act of gene expression include the consumption of energy (Wagner 2005; Wagner 2007) and the opportunity cost of not using ribosomes to make other gene products (Scott et al. 2014). While some costs arise from transcription (Kafri et al. 2016), we simplify our model by attributing all of the cost of expression to the act of translation.

Kafri et al. (2016) reported that, in rich media, the growth rate of haploid yeast is reduced by about 1% when mCherry is expressed to about 2% of proteome. With b_{max} = 1 giving the growth rate of the yeast when mCherry is not expressed, we have the cost of gene expression equal to 0.01. Next, we estimate the production rate of mCherry in Kafri et al. (2016) by assuming that mCherry is in steady state between production and dilution due to cell division; fluorescent proteins tend to be stable such that degradation can be ignored (Snapp 2009). Ghaemmaghami et al. (2003) estimated that a haploid yeast cell contains about 5×10^7 protein molecules, 2% of which are now mCherry. Over a 90 minute cell cycle in Kafri et al. (2016), about 5×10^5 mCherry molecule per cell need to be expressed in order to double in numbers. This yields a production rate of about 5×10^3 mCherry molecules per minute per cell. Because the total cost of gene expression is 0.01, the cost at a protein production rate of one mCherry molecule per minute per cell, c_{transl}, is 2×10^{-6}. Long genes should be more expensive to express than short ones; for a gene of length L, we assume its cost of expression is $c_{transl}L$ / 370, where 370 is the geometric mean length of a yeast protein as described above in Section 4. Results using the length of

mCherry instead, i.e. a slightly higher cost of expression of $c_{transl}L$ / 236, are unlikely to be significantly different.

The overall cost of gene expression at time *t*, *C*(*t*) is:

$$C(t) = c_{transl}(\sum_1^n \frac{L_i}{10^{2.568}} r_{transl\_init\_i} N_{mRNA\_aft\_delay\_i}(t) + \sum_1^n \frac{L_i}{10^{2.568}} \frac{r_{transl\_init\_i}}{2} N_{mRNA\_during\_delay\_i}(t)).$$

The second term represents transcripts that are on average half-loaded with ribosomes, and hence experiencing on average half the cost of translation. We integrate *C*(*t*) within segments of constant *C*(*t*) to obtain the overall cost of gene expression during a simulation.

9. Mutation

Because we use an origin-fixation approach, only the relative and not the absolute values of our mutation rates matter. In *S. cerevisiae,* the rates of small indels and of single nucleotide substitutions have been estimated as 0.2×10^{-10} per base pair and 3.3×10^{-10} per base pair, respectively (Lynch et al. 2008). Thus, cis-regulatory sequences are primarily shaped by single nucleotide substitutions. We do not model small indels in the cis-regulatory sequence, but increase the single nucleotide substitution up to 3.5×10^{-10} per base pair to compensate. This corresponds to a rate of 5.25×10^{-8} per 150 bp cis-regulatory sequence.

Lynch et al. (2008) also report a rate of gene duplication of 1.5×10^{-6} per gene and of deletion of 1.3×10^{-6} per gene (not including non-deletion-based loss of function mutations). These values turned out to swamp the evolution of TFBSs and hence significantly slow down our simulations,

489 so we chose values 10-fold lower, making both gene duplication and gene deletion occur at rate
490 1.5×10^{-7} per gene. This preserves their numerical excess but reduces its magnitude.
491
492 Our model contains 8 gene-specific parameters, namely L, $r_{Act\_to\_Int}$, $r_{protein\_deg}$, $r_{protein\_syn}$, r_{mRNA_deg},
493 the $K_d(0)$ of a TF, whether a TF is an activator vs. repressor, and the consensus binding sequence
494 of a TF. We assume mutations to L are caused by relatively neutral small indels, which we
495 assume to be 20% of all small indels; mutation to L therefore occurs at rate 1.2×10^{-11} per codon,
496 i.e. $1.2\times10^{-11}L$ for a gene of length L. For $r_{Act\_to\_Int}$, we assume that it is altered by 10% of all the
497 point mutations (single nucleotide substitution and small indels) to the core promoter of a gene.
498 The length of a core promoter is about 100 bp and is relatively constant among genes (Roy &
499 Singer 2015), yielding a mutation rate of $r_{Act\_to\_Int}$ of 3.5×10^{-9} per gene.
500
501 The remaining 6 gene-specific parameter mutation rates are parameterized with lower accuracy
502 due to lack to data; the principal decision is which to make dependent vs. independent of gene
503 length. TF binding to DNA depends on particular peptide motifs whose length is likely
504 independent of TF length, therefore we make mutation rates independent of gene length for
505 mutations to $K_d(0)$, to the consensus binding sequence of a TF, and to the activating vs
506 repressing identity of a TF. We set the rate of each of the three mutation types to 3.5×10^{-9} per
507 gene. In contrast, because the stability of an mRNA mainly depends on its codon usage (Cheng
508 et al. 2017) and thus more codons means more opportunities for change, we assume the rate of
509 mutation to r_{mRNA_deg} does depend on gene length, as do mutations to protein stability $r_{protein\_deg}$.
510 $r_{protein\_syn}$ is determined by the density of ribosomes on mRNA and the elongation rate of
511 ribosomes, and therefore is affected both by ribosome loading speed and by slow spots forming
512 queues in the mRNA. Ribosome loading often relies on the 5'UTR of mRNA (Hinnebusch 2011),

and 5'UTR length is positively correlated with ORF length (Tuller et al. 2009). Slow-spots in mRNA can be due to secondary structure or to suboptimal codons, therefore are also more likely to appear by mutation to long mRNAs, so we assume the rate of mutation to $r_{protein\_syn}$ depends on gene length. We set the mutation rates of $r_{protein\_deg}$, $r_{protein\_syn}$, and r_{mRNA_deg} each to 9.5×10^{-12} per codon; in other words, each mutation rate is 3.5×10^{-9} for a yeast gene of average length (on a log-scale) $10^{2.568}$ = 370 codons.

$r_{Act\_to\_Int}$, $r_{protein\_syn}$, $K_d(0)$, $r_{protein\_deg}$, and r_{mRNA_deg} evolve as quantitative traits. They are assumed to have, in the absence of selection, a log-normal stationary distribution with mean μ and standard deviation σ, with values estimated below and listed in **Table S2**. Denote the values of a parameter as x before mutation and x' after mutation; mutation takes the form:

$$\log_{10}x' = \log_{10}x + \mathrm{Normal}(k(\mu - \log_{10}x), \sigma), \qquad \text{(S7)}$$

where k controls the speed of regressing back to the stationary distribution; we set k = 0.5 for all 5 parameters. To set values of μ, central tendency estimates of these five values (from **Table S1)** are adjusted according to our expectations about mutation bias. We assume a mutation bias toward faster mRNA degradation r_{mRNA_deg}, faster $r_{Act\_to\_Int}$ (Decker & Hinton 2013; Roy & Singer 2015), slower translation initiation $r_{protein\_syn}$ (Hinnebusch 2011), and larger $K_d(0)$. We assume that the observed log-normal means of r_{mRNA_deg}, $r_{protein\_syn}$, and $r_{Act\_to\_Int}$ differ by 2-fold from the mean expected from mutational bias; for example, the mean of $\log_{10}(r_{mRNA\_deg})$ is -1.49, so the value of μ for r_{mRNA_deg} is -1.49 + $\log_{10}(2)$ = -1.19. We assume a larger bias for $K_d(0)$, namely that mutation is likely to reduce the affinity of a TF for a TFBS down to non-specific levels. Thus, we set $\mu = \log_{10}(K_d(3))$ = -5 for $K_d(0)$; note that in this case μ is equal to one of the boundary values,

which will be hit far more often than during the evolution of other parameters. We assume that the observed central tendency estimate of protein stability does not depart from mutational equilibrium, therefore the value of μ for $r_{protein\_deg}$ is the mean of $\log_{10}(r_{protein\_deg})$ =-1.88.

The value of σ controls mutational effect size. We set the value of σ such that 1% of mutational changes from x=10^{μ} go beyond the boundary values, for simplicity approximating by considering only the closer of the two boundary values on a log scale, i.e. we solve Eq. S8 for σ:

$$\begin{cases} P(\mu + \text{Normal}(0,\sigma) \geq \log_{10} U) = 0.01, & \text{if the upper bound U is closer} \\ P(\mu + \text{Normal}(0,\sigma) \leq \log_{10} L) = 0.01, & \text{if the lower bound L is closer} \end{cases} \quad \text{(S8)}$$

For example, the upper and the lower bounds of r_{mRNA_deg} are 0.54 min^{-1} and 7.5×10^{-4} min^{-1}; on a log-scale, the upper bound is closer to 10^{μ} = $10^{-1.19}$ min^{-1}. Plugging these values in Eq. S8 and solving for σ, we have σ = 0.396. We set the values of σ for $r_{protein\_syn}$, and $r_{protein\_deg}$ in the same way. However for $r_{Act\_to\_Int}$, σ is set according to the lower bound, even though it is the more distant from 10^{μ}, because otherwise a stable preinitiation complex will evolve too rarely. Under this high mutational variance, evolutionary outcomes at the two bounds are still only observed 5% of the time. For $K_d(0)$, because its upper bound is equal to 10^{μ}, we set σ to 0.776, such that 1% of mutations can change the values of $K_d(0)$ by 100-fold or more.

Mutant values of L, $r_{Act\_to\_Int}$, $r_{protein\_syn}$, $r_{protein\_deg}$, and r_{mRNA_deg} are constrained by the same bounds that constrain the initial values of these parameters (Sections 3-6). If a mutation increases the value of any of these 5 parameters to beyond the corresponding upper bound, we set the mutant value to the upper bound; similarly for a mutant value that is smaller than the lower bound of the corresponding parameter. For mutation to $K_d(0)$, we resample if $x' \geq K_d(3)$,

because otherwise the mutation effectively "deletes" the TF by reducing its affinity to non-specific levels.

10. Burn-in evolutionary simulation conditions

When the signal is not allowed to regulate the effector genes directly, most simulations under selection either to filter out short spurious signals or for simple signal recognition in the absence of spurious signals rapidly found a local optimal solution in which effector genes are never expressed. This local optimum exists in part because we assume that the environment in which the effector is deleterious is twice as common as the environment in which it is beneficial (**Fig. 3**). When the signal is not allowed to directly turn on the effector, then to escape this local optimum, at least one activator must be induced by the signal and then induce the effector. Such activators are rare when genotypes are randomly initialized. Making matters worse, mutation tends to reduce expression after initialization (see Section 9).

To reduce the frequency of this problem, we added a burn-in stage to simulations in which the signal is not allowed to regulate the effector directly. During burn-in, we switch the frequencies of the two environments, so that selection to express the effector is stronger. We also change the mutational bias in $r_{Act\_to\_Int}$, $r_{protein\_syn}$, and $K_d(0)$ to favor higher expression and stronger binding. For $r_{Act\_to\_Int}$ and $r_{protein\_syn}$, we use 0.1 instead of 0.01 as the tolerated fraction of extreme mutations in Eq. S8. For $K_d(0)$, we decrease μ from -5 to -7.5, biasing mutation toward the mean value at which we initialize (**Table S1**). Evolving an activator that can reliably turn on the effector when the signal is "ON" primarily relies on forming strong binding sites and appropriate kinetic constants in expression, assisted by the change in mutational bias above. To better focus the simulations on sampling appropriate mutations during the burn-in phase, we

585 reduce the rate of gene duplication and the rate of deletion to 5.25×10^{-9} per gene, and limit
586 the maximum number of TF genes to 9 and that of effector genes to 2. Each simulation is run
587 under burn-in conditions for 1000 steps, after which normal model settings and selection
588 conditions are restored. The same burn-in mutational settings are used for the control selection
589 conditions (no selection, no spurious signal, and harmless spurious signal).
590

591 **11. Quantifying occurrence of network motifs**

592 Scoring the presence of a C1-FFL motif (e.g. **Fig. 4B**) or diamond motif (e.g. **Fig. 7**) is based on
593 scoring whether TF *x* regulates gene *y*. Gene duplication and divergence complicate this scoring,
594 because different gene copies might encode functionally identical proteins, but one copy of
595 gene *y* might have a TFBS for TF *x* and the other might not. For the purpose of scoring motifs,
596 our algorithm begins by simply treating each gene copy as though it were a unique gene.
597

598 Following Milo et al. (2002), a C1-FFL is scored if activating TF A can bind to the cis-regulatory
599 sequence of activating TF B and to the effector, if B can also bind to that of the effector, and if B
600 does not bind to that of A. Auto-regulation is allowed. We exclude C1-FFLs in which A and B
601 encode the same TF or variants of the same TF. In the case of direct regulation, A can be the
602 signal rather than a TF. C1-FFLs can then be subdivided into categories based on overlap
603 between the TFBSs in the cis-regulatory region of the effector (**Fig. 2**).
604

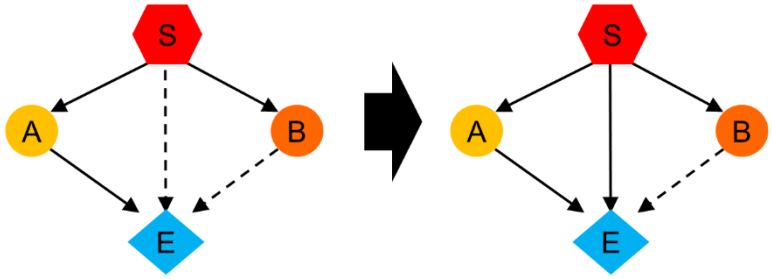
605 A diamond is scored if two signal-regulated activating TFs, A and B, do not bind to each other's
606 cis-regulatory region, but both bind to that of the effector. We allow auto-regulation and
607 require A and B to not encode the same TF or variants of the same TF.
608

609 A FFL-in-diamond is scored if one signal-regulated activating TF A binds to the cis-regulatory
610 region of another signal-regulated activating TF B, but B does not bind to that of A, and both A
611 and B bind to that of the effector. Again, auto-regulation is allowed, and A and B must not
612 encode the same TF or variants of the same TF.
613
614 Occurrence within one evolutionary replicate is calculated as the fraction of the last 10,000
615 evolutionary steps in which at least one motif of the type of interest is present. The mean and
616 standard error of this occurrence metric is then calculated across replicates.
617

618 **12. Perturbing network motifs**

619 In **Fig. 5** and **Fig. 9**, we add a TFBS to the cis-regulatory sequence of the effector gene, in order
620 to destroy the AND-gate logic of an isolated C1-FFL or diamond. The new TFBS is chosen such
621 that it does not overlap with any existing TFBSs, and has the same affinity as the strongest TFBS
622 that is already present in the cis-regulatory sequence of the effector gene for the signal/fast TF
623 (to convert from an AND-gate to signal-controlled/fast TF-controlled), or for the slow TF (to
624 convert from an AND-gate to slow TF-controlled).
625
626 When a TRN has multiple AND-gated motifs of interest, we convert all of them. A perturbation
627 can also affect the logic of other, potentially non-AND-gated motifs in the same TRN (e.g. **Fig.**
628 **S12**), making it hard to attribute the fitness effect to the AND-gate logic of the targeted motif.
629 For this reason, we perform the perturbation analysis not on a single potentially problematic
630 genotype, but on the last 10,000 evolutionary steps of an evolutionary simulation. Within those
631 10,000 related genotypes, we exclude those that also contain other motifs that might influence
632 our results. For simulations where the signal is allowed to directly regulate the effector, this

633 means excluding those with non-AND-gated C1-FFLs. For simulations where the signal is not
634 allowed to directly regulate the effector, we exclude genotypes with either AND-gated or non-
635 AND-gated motifs other those of interest (e.g. if we intend to perturb AND-gated isolated C1-
636 FFLs, we exclude genotypes that also contain either an AND-gated isolated diamond or a non-
637 AND-gated C1-FFL). Both pre-perturbation fitness and post-perturbation fitness are averaged
638 over the remaining genotypes. If no evolutionary step meets our requirement, we exclude the
639 entire evolutionary simulation; this occurs only when the signal cannot directly regulate the
640 effector genes.
641



642

643 **Fig. S12 Examples of confounding motifs in perturbation analysis.** The TRN on the left contains
644 a slow TF-controlled C1-FFL (S-B-E) and an AND-gated C1-FFL (S-C-E). To convert S-C-E into a
645 signal-controlled C1-FFL, we need to add one TFBS for the signal to the cis-regulatory sequence
646 of E. However, this change also makes S-B-E OR-gated, making it difficult to conclude whether it
647 is the AND gate logic of S-B-E that matters for fitness.

References

Belle, A, Tanay, A, Bitincka, L, Shamir, R, and O'Shea, EK. 2006. Quantification of protein half-lives in the budding yeast proteome. *PNAS* 103:13004-13009.

Benos, PV, Bulyk, ML, and Stormo, GD. 2002. Additivity in protein-DNA interactions: how good an approximation is it? *Nucleic Acids Res* 30:4442-4451.

Brown, CR, Mao, C, Falkovskaia, E, Jurica, MS, and Boeger, H. 2013. Linking Stochastic Fluctuations in Chromatin Structure and Gene Expression. *PLoS Biol* 11:e1001621.

Cheng, J, Maier, KC, Avsec, Ž, Rus, P, and Gagneur, J. 2017. Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. *RNA* 23:1648-1659.

Decker, KB, and Hinton, DM. 2013. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annu Rev Microbiol* 67:113-139.

Dujon, B. 1996. The yeast genome project: what did we learn? *Trends Genet* 12:263-270.

Ghaemmaghami, S, Huh, W-K, Bower, K, Howson, RW, Belle, A, Dephoure, N, O'Shea, EK, and Weissman, JS. 2003. Global analysis of protein expression in yeast. *Nature* 425:737-741.

Gillespie, DT. 1977. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340-2361.

Guillemette, B, Bataille, AR, Gevry, N, Adam, M, Blanchette, M, Robert, F, and Gaudreau, L. 2005. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 3:e384.

Hinnebusch, AG. 2011. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiology and Molecular Biology Reviews* 75:434-467.

Hocine, S, Raymond, P, Zenklusen, D, Chao, JA, and Singer, RH. 2013. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature Methods* 10:119-121.

Kafri, M, Metzl-Raz, E, Jona, G, and Barkai, N. 2016. The Cost of Protein Production. *Cell Reports* 14:22-31.

Katan-Khaykovich, Y, and Struhl, K. 2002. Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev* 16:743-52.

Larson, DR, Zenklusen, D, Wu, B, Chao, JA, and Singer, RH. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* 332:475-478.

Lee, W, Tillo, D, Bray, N, Morse, RH, Davis, RW, Hughes, TR, and Nislow, C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet* 39:1235-44.

Lynch, M, Sung, W, Morris, K, Coffey, N, Landry, CR, Dopman, EB, Dickinson, WJ, Okamoto, K, Kulkarni, S, Hartl, DL, and Thomas, WK. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272-9277.

Maerkl, SJ, and Quake, SR. 2007. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science* 315:233-237.

Milo, R, Shen-Orr, S, Itzkovitz, S, Kashtan, N, Chklovskii, D, and Alon, U. 2002. Network motifs: Simple building blocks of complex networks. *Science* 298:824-827.

Mor, A, Suliman, S, Ben-Yishay, R, Yunger, S, Brody, Y, and Shav-Tal, Y. 2010. Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature Cell Biol* 12:543-552.

Nalefski, EA, Nebelitsky, E, Lloyd, JA, and Gullans, SR. 2006. Single-molecule detection of transcription factor binding to DNA in real time: Specificity, equilibrium, and kinetic parameters. *Biochemistry* 45:13794-13806.

693 Niño, CA, Hérissant, L, Babour, A, and Dargemont, C. 2013. mRNA nuclear export in yeast.
694 *Chemical Reviews* 113:8523-8545.
695 Park, S, Chung, S, Kim, KM, Jung, KC, Park, C, Hahm, ER, and Yang, CH. 2004. Determination of
696 binding constant of transcription factor myc-max/max-max and E-box DNA: The effect of
697 inhibitors on the binding. *Biochim Biophys Acta, Gen Subj* 1670:217-228.
698 Pelechano, V, Chávez, S, and Pérez-Ortín, JE. 2010. A Complete Set of Nascent Transcription
699 Rates for Yeast Genes. *PLoS ONE* 5:e115560.
700 Roy, AL, and Singer, DS. 2015. Core promoters in transcription: old problem, new insights.
701 *Trends Biochem Sci* 40:165-171.
702 Scott, M, Klumpp, S, Mateescu, EM, and Hwa, T. 2014. Emergence of robust growth laws from
703 optimal regulation of ribosome synthesis. *Mol Syst Biol* 10:747.
704 SGD Project. https://yeastmine.yeastgenome.org accessed April 2, 2018.
705 Sharon, E, Kalma, Y, Sharp, A, Raveh-sadka, T, Levo, M, Zeevi, D, Keren, L, Yakhini, Z,
706 Weinberger, A, and Segal, E. 2012. Articles Inferring gene regulatory logic from high-
707 throughput measurements of thousands of systematically designed promoters. *Nature*
708 *Biotechnol* 30:521-530.
709 Siwiak, M, Zielenkiewicz, P, Jacobson, A, Grebogi, C, and Kito, K. 2010. A Comprehensive,
710 Quantitative, and Genome-Wide Model of Translation. *PLoS Comput Biol* 6:e1000865.
711 Smith, C, Lari, A, Derrer, CP, Ouwehand, A, Rossouw, A, Huisman, M, Dange, T, Hopman, M,
712 Joseph, A, Zenklusen, D, Weis, K, Grunwald, D, and Montpetit, B. 2015. In vivo single-
713 particle imaging of nuclear mRNA export in budding yeast demonstrates an essential
714 role for Mex67p. *J Cell Biol* 211:1121-1130.
715 Snapp, EL. 2009. Fluorescent proteins: a cell biologist's user guide. *Trends in Cell Biology* 19:649-
716 655.

717 Tuller, T, Ruppin, E, and Kupiec, M. 2009. Properties of untranslated regions of the *S. cerevisiae*
718 genome. *BMC Genomics* 10:391.
719 Wagner, A. 2005. Energy constraints on the evolution of gene expression. *Mol Biol Evol* 22:1365-
720 1374.
721 Wagner, A. 2007. Energy costs constrain the evolution of gene expression. *J Exp Zool Part B*
722 308B:322-324.
723 Waldron, C, Jund, R, and Lacroute, F. 1977. Evidence for a high proportion of inactive ribosomes
724 in slow-growing yeast cells. *Biochemical Journal* 168:409-415.
725 Wang, Y, Liu, CL, Storey, JD, Tibshirani, RJ, Herschlag, D, and Brown, PO. 2002. Precision and
726 functional specificity in mRNA decay. *Proc Natl Acad Sci USA* 99:5860-5865.
727 Wunderlich, Z, and Mirny, LA. 2009. Different gene regulation strategies revealed by analysis of
728 binding motifs. *Trends Genet* 25:434-440.
729 Yuan, GC, Liu, YJ, Dion, MF, Slack, MD, Wu, LF, Altschuler, SJ, and Rando, OJ. 2005. Genome-
730 scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309:626-630.
731 Zhu, J, and Zhang, MQ. 1999. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*.
732 *Bioinformatics* 15:607-611.