Non-adaptive factors determine which equally effective regulatory motif evolves to generate pulses

Kun Xiong[1,2], Mark Gerstein[2*], Joanna Masel[3*]

[1] Department of Molecular and Cellular Biology, University of Arizona, Tucson AZ, United States of America. ORCID: 0000-0003-1431-6586
[2] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven CT, United States of America. ORCID: 0000-0002-9746-3719
[3] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson AZ, United States of America. ORCID: 0000-0002-7398-2127
[*]Corresponding author

Abstract
Transcriptional regulatory networks (TRNs) are enriched for certain subnetworks or "motifs". Motif usage is commonly interpreted as the result of adaptive evolution. But network motifs can also differ in how easy it is to evolve them. Here, we simulated the *de novo* evolution of motifs within TRNs under selection to produce a short, sharp pulse of an effector protein. In agreement with past work in the field, two network motifs, type 1 incoherent feed-forward loops (I1FFLs) and negative feedback loops (NFBLs), evolved as the primary solutions. Different selection conditions changed the relative frequencies of the two solutions, but this was not due to the superior performance of one; under all conditions, either motif can achieve similar top performance. I1FFLs generally evolve more often than NFBLs, unless we selected for a particularly tall pulse. This result suggests that I1FFLs are evolutionary more accessible than NFBLs. When NFBLs do evolve, it is usually from a conjugate containing both I1FFL and NFBL. In contrast, I1FFLs can evolve via a greater variety of trajectories. This difference potentially explains NFBL's lower evolutionary accessibility. To agreement with our simulation results, we found that in the real yeast TRN, output genes of NFBLs had higher expression levels than those of I1FFLs, i.e. selection for taller pulses. These results suggest that evolutionary accessibility, and not relative functionality, shape which networks motifs evolve in TRNs, and do so as a function of the expression levels of particular genes.

Keywords
Pulse generation, adaptationism, mutation-biased adaptation, gene regulatory network, transcriptional regulation

Introduction

The topology of transcriptional regulatory networks (TRNs) is enriched for certain motifs (Lee et al, 2002; Mangan & Alon, 2003; Milo et al, 2002; Shen-Orr et al, 2002). Many argue that these motifs are the result of adaptive evolution where the dynamical behavior of the motif has provided a beneficial function favored by natural selection (Alon, 2007). However, adaptationist claims about TRN organization have also been accused of being just-so stories, with adaptive hypotheses still in need of testing against an appropriate null model of network evolution (Artzy-Randrup et al, 2004; Jenkins & Stekel, 2010; Knabe et al, 2008; Kuo et al, 2006; Lynch, 2007; Mazurie et al, 2005; Payne & Wagner, 2015; Ruths & Nakhleh, 2013; Solé & Valverde, 2006; Tsuda & Kawata, 2010; Wagner, 2003; Widder et

al, 2012). We recently generated such a null model and used it to show that coherent type 1 feed-forward loops can, as hypothesized, evolve adaptively in response to selection to filter out short spurious signals, by combining a fast signaling pathway and a slow signaling pathway with an AND gate (Xiong et al, 2019). Testing the hypothesis in this way was not merely confirmatory, but generated other insights about the existence and nature of alternative adaptive solutions, especially when slow transcriptional regulation is combined with faster response mechanisms such as post-translational regulation (Xiong et al, 2019). Other network motifs and properties have not yet received similar treatment.

At least three different motifs (**Fig. 1A**) are all capable of producing a sharp pulse of expression in response to an increase in input signal (**Fig. 1B**) (Basu et al, 2004; Çağatay et al, 2009; Camas et al, 2006). All depend on an effector first being rapidly activated by a signal, and later, at a slower timescale, being repressed by it. These three motifs are simple auto-repression (AR), negative feed-back loops (NFBLs), and incoherent type 1 feed-forward loops (I1FFLs) (**Fig. 1A**). The three motifs are topologically and functionally similar to each other, differing in whether the slow repression is effected via negative autoregulation by the effector *R* of itself, via negative feedback regulation of *R* using a specialized repressor, or via a separate negative control pathway from the input to the repressor and then the effector.

The high prevalences of I1FFLs and NFBLs in TRNs has been interpreted to occur because these two motifs are adaptations for pulse generation and closely related functions (Ferrell, 2016; Shi et al, 2017; Shoval & Alon, 2010; Shoval et al, 2010). Both I1FFLs and NFBLs allow the steady-state level of the effector, before and after the pulse, to be independent of the signal strength, a property known as chemical adaptation (Ferrell, 2016; Shi et al, 2017). We note that AR is normally hypothesized to perform functions other than pulse generation (Alon, 2007; Wall et al, 2004), but theoretical analysis and experiments show that AR can generate pulses {Amit, 2007 #675;Camas, 2006 #666;Rosenfeld, 2002 #671}. We therefore include AR for the completeness of the study, while focusing on I1FFLs and NFBLs.

Which of the motifs is likely to evolve is often explained by adaptive demands for specific properties of the pulse. For example, although both I1FFLs and NFBLs allow the amplitude of the pulse to be a function of the fold-change of the signal's strength (Shoval et al, 2010), they do so with different functional forms (Adler et al, 2014). I1FFLs and NFBLs can also differ in their ability to filter noise in the signal (Buzi & Khammash, 2016).

Alternatively, non-adaptive causes might be responsible for differences in the occurrence of the three motifs. Whether non-adaptive factors are plausible causes is a question that *in silico* evolution is ideally set up to explore. One key factor that is easily measured and manipulated *in silico* is expression level. An important non-adaptive consideration is that fitness landscapes tend to have many alternative local peaks (Whitlock et al, 1995). Expression levels can change the relative accessibility of different local peaks, in ways that are independent of differences in their heights.

Here we simulate TRN evolution under selection to produce a pulse, and test how subtle differences between scenarios have both adaptive and non-adaptive effects on which motifs evolves. In particular, a highly expressed effector is more able to stimulate its repressor, and we therefore predict that this scenario should be more likely to evolve regulation via an NFBL and correspondingly less likely to evolve an I1FFL. Our simulations confirm that NFBLs are evolutionarily more accessible than I1FFLs under this scenario, but that I1FFLs are more accessible under other scenarios where a highly expressed effector is

not required. Data from real-world yeast TRNs agree with model predictions, showing that the effectors of NFBLs generally have higher expression levels than those of I1FFLs.

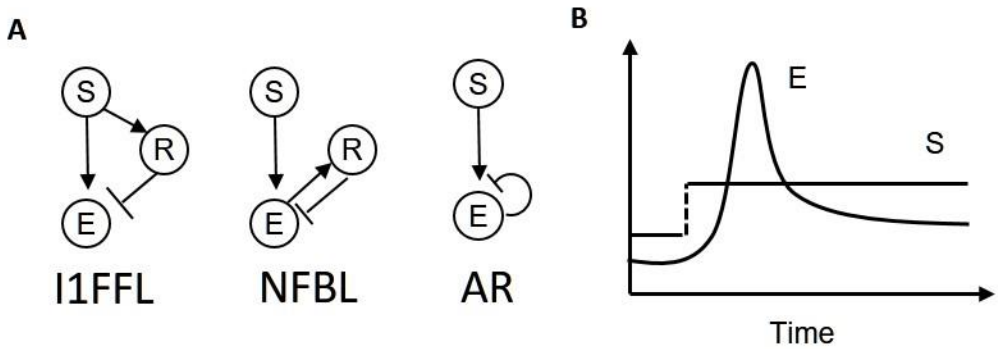

Figure 1. Three motifs (I1FFL, NFBL, and AR) all produce a pulse of effector E expression in response to increased signal S. (A) In all three cases, rapid and direct activation of the effector by the signal is eventually countered by a slower path of repression. The three motifs differ topologically in whether repression is by the effector itself (AR), by a specialized repressor (R) that is activated by the signal (I1FFL), or by a specialized repressor that is activated by the effector (NFBL). Regular arrow tips represent activation and ⊣ represents repression. **(B)** With appropriate parameters, and with a delay between transcriptional activation and protein production in the case of AR, all three motifs can induce a pulse, as the initial increase in expression as S activates E is eventually tamped down by a path of repression.

Results

Model overview

We used a previously described computational model to simulate the expression of genes in a TRN, parameterized by available *Saccharomyces cerevisiae* data (Xiong et al, 2019). The TRN evolves under a realistic mutational spectrum including *de novo* appearance of transcription factor binding sites and frequent gene duplication and deletion. Briefly, each gene encodes either an activating or repressing TF, and is regulated by a 150-bp cis-regulatory sequence accessible to transcription factor binding. Each TF recognizes an 8-bp consensus binding sequence with a characteristic binding affinity. Binding sites with up to 2 mismatches are still recognized, with each mismatch reducing binding affinity according to a thermodynamic model (Xiong et al, 2019, Supplementary Methods/TF Binding). The concentrations of TFs are used to calculate the probabilities that each cis-regulatory region is bound by a given number of activators and repressors (Xiong et al, 2019, Supplementary Methods/TF Occupancy).

We simulate stochastic transitions between an active chromatin state that can initiate transcription, an intermediate primed state capable of becoming either activated or repressed, and a repressed chromatin state. Most transition rates depend on whether activators and/or repressors are bound (Xiong et al, 2019, Methods/Transcriptional regulation, but with no AND-gate capability), with the fastest transition rate to the active state occurring when at least one activator and no repressors are bound. Transcription initiation from the active chromatin state is simulated stochastically, creating transcriptional bursts whose mean duration depends on an evolvable gene-specific parameter (Xiong et al, 2019, Supplementary Methods/$r_{Act_to_Int}$).

Once transcription is initiated, we simulate a delay before mRNA can be translated at full speed. The delay accounts for the completion of both transcription and the loading of ribosomes to mRNA, and is a function of gene length (Xiong et al, 2019, Supplementary Methods/Transcriptional and Translational delay times). mRNA degradation is simulated stochastically, while protein synthesis and degradation are treated deterministically with ODEs (Xiong et al, 2019, Supplementary Methods/Simulation of gene expression).

To select for pulse generation, we designate an input signal to the TRN, which binds to cis-regulatory regions like any other TF, but whose concentration is set externally rather than being regulated by other TFs in the TRN. The input signal always activates gene expression. Signal concentration is low and constant during a burn-in phase, where genes are initialized with a repressed chromatin state, and begin with zero non-signal mRNA and protein. Then in stage 2, the signal instantly switches to a high level, combined with selection for a pulse in a TF designated to be the "effector". High fitness depends on having low effector expression at the end of stage 1, matching a pre-defined peak effector concentration during stage 2, rapidly increasing effector level after stage 2 begins, and having a low effector level at end of stage 2. Details of the signal and fitness calculation are given in the Methods.

Following Xiong et al. (2019), the evolutionary simulation starts with a randomly generated genotype of 3 activator genes, 3 repressor genes, and an effector gene, which is initialized as an activator, in the process making NFBLs more accessible than ARs (although below we will explore the effects of switching this).

To simulate mutation, each gene has 9 gene-specific characteristics that can mutate as described in Xiong et al. (2019, Supplementary Methods/Mutation): cis-regulatory sequence, rate at which transcriptional bursts end, gene length, mRNA degradation rate, protein synthesis rate, protein degradation rate, preferred binding sequence of the encoded TF, its binding affinity to DNA, whether the TF is an activator vs. a repressor. Point mutations in the cis-regulatory sequences and the preferred binding sequences of TFs can change which TFBSs exist, and whether they have high affinity or low. All quantitative gene-specific parameters except length are subjected to mutational bias, e.g. mutation tends to reduce the affinity of TF binding. In case this is insufficient to ensure the values of the mutatable parameters never go beyond reasonable limits, we also apply hard bounds. All the mutations described above, as well as mutations to cis-regulatory regions and to preferred binding sequences, are treated exactly as described by Xiong et al. (2019, Supplementary Methods/Mutation). We adjust the maximum number of copies that can be created by gene duplication (see Methods).

We calculate the fitness of the resident genotype and a mutant. If the mutant has higher fitness than the resident genotype, the mutant replaces the latter and becomes the new resident (see Methods). Otherwise, we generate a new mutant and repeat the comparison until a replacement happens. We call a replacement an evolutionary step, and stop simulation after 50,000 evolutionary steps.

High peak expression level promotes NFBLs

We evolve TRNs under selection to generate a pulse of effector expression in response to a sudden 10-fold increase in input. While any of the three network motifs can solve this challenge, a highly expressed effector is more capable of stimulating its repressor, and thus this solution should be more likely to evolve regulation via an NFBL and correspondingly less likely to evolve an I1FFL. Note that this prediction is expected on both adaptive grounds of which solution might be superior, and on non-adaptive grounds of which solution is easier for evolution to find.

To test this prediction, we vary the optimal peak level of the effector (see Methods for details of fitness function). *In silico* evolution from a random starting point is not always successful at reaching the target phenotype, so we focus on the most evolutionarily successful simulations. We do this by dividing evolutionary replicates into three categories based on final fitness (**Fig. S1**).

High-fitness solutions rarely involve AR under any of the three selection conditions, while both I1FFLs and NFBLs evolve often (**Fig. 2A**). As predicted, when we select for higher effector expression, we get more NFBLs and fewer I1FFLs (**Fig. 2A**). However, if we prevent either I1FFLs or NFBLs from evolving, similarly high fitness genotypes can be obtained via the other motif (**Figs. 2B and S3**). The reason we get more NFBLs and fewer I1FFLs with selection for higher peak effector expression is therefore not straightforward adaptive superiority of the former.

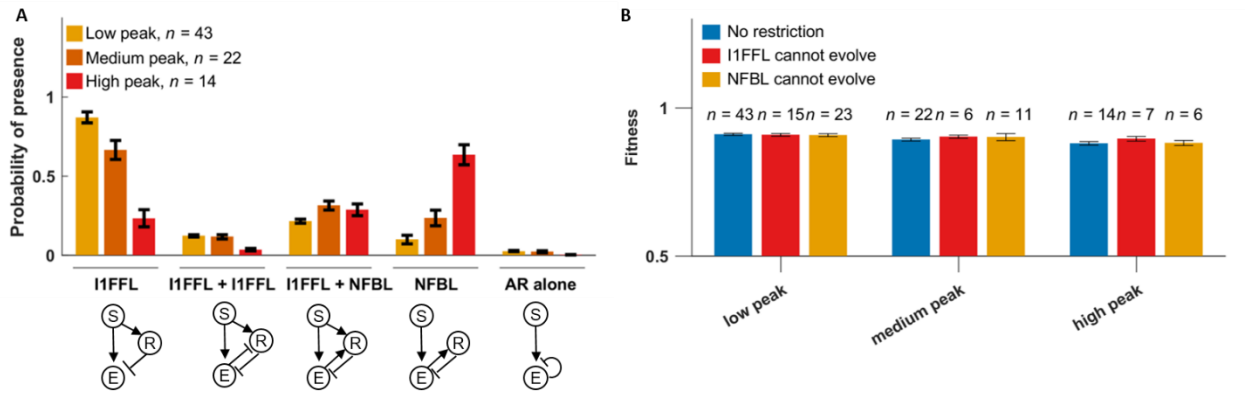

Figure 2. High peak expression levels of effector promote NFBLs. (A) TRNs are evolved under selection for generating pulses of different peak levels. The input signal starts with 100 molecules per cell and increases to 1,000 molecules per cell to trigger a pulse. We simulate evolution under three optimal peak level of the effector: weak (P_{opt} = 5,000 molecules per cell), medium (P_{opt} = 10,000 molecules per cell), and strong (P_{opt} = 20,000 molecules per cell). For each high-fitness genotype (**Fig. S1**), we calculate the proportion of evolutionary steps that contain at least one network motif of the specified type among the last 10,000 evolutionary steps (out of a total of 50,000 evolutionary steps). When scoring for motifs, non-functional spurious TFBSs were excluded (see Methods for details and **Fig. S2** for the occurrence of network motifs when non-functional spurious TFBSs are included). R can be auto-regulating (not shown in circuit diagram). On rare occasions, AR co-occurred with I1FFLs or overlapping I1FFLs (labelled here I1FFL + I1FFL) (**Fig. S3**), and these few cases were included in the scoring of I1FFL and overlapping I1FFL frequencies. **(B)** Preventing either NFBLs or I1FFLs from evolving does not lower the final fitness within high-fitness evolutionary simulations. Instead, genotypes obtained equally high fitness by evolving the other common motif (**Fig. S4**). To prevent NFBLs from evolving, we remove the TF binding activity of effectors; this also prevents the evolution of the AR auto-repression motif. To prevent I1FFLs from evolving, we ignore TFBSs for the signal in the cis-regulatory sequence of any repressors. Because this might have unintended consequences for mutations that convert repressors to/from activators, we set to zero the rate of mutations that effect this conversion. Data are shown as mean ± SE over replicates.

I1FFLs have high evolutionary accessibility in many scenarios

Among replicates that achieve only medium-fitness, we see far more I1FFLs and ARs than among replicates that achieve high-fitness (**Fig. 3**). This suggests a general tendency for I1FFLs and ARs to be evolutionarily more accessible than NFBLs, but more prone to get stuck at subpar local fitness peaks. In

other words, while high fitness I1FFL peaks exist, they cannot always be found from a lower-fitness I1FFL genotype. This problem occurs more often for high effector expression, when we get more subpar I1FFLs (**Fig. 3**).

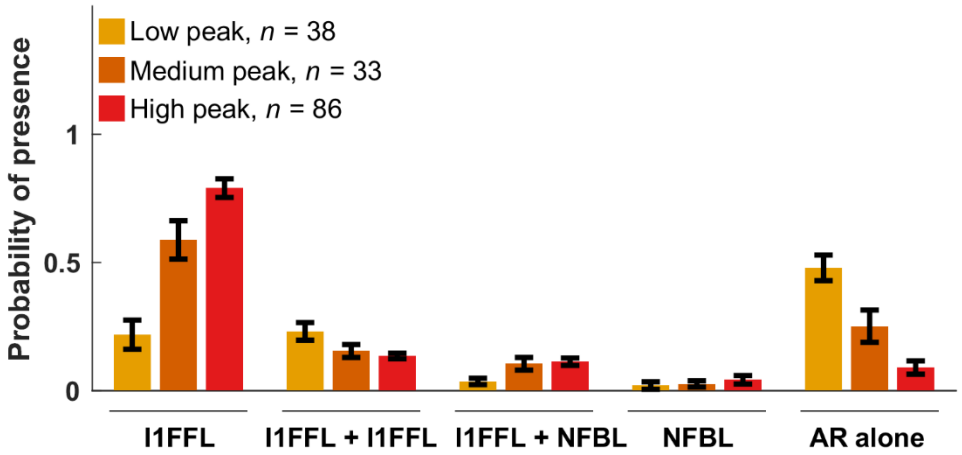

Figure 3. I1FFLs and auto-repression are the primary solution found by medium-fitness genotypes. Similarly to **Fig. 2A**, we calculate the occurrence of different motifs in medium-fitness genotypes evolved under selection to generate a pulse of effector, varying which level of peak effector expression is selected for. Data are shown as mean ± SE over replicates.

If we initialize the simulations with the effector acting as a repressor, instead of as an activator, then the occurrence of NFBLs decreases while that of the I1FFLs increases under the high-peak condition **(Fig. S5)**. This again suggests that whether I1FFLs or NFBLs evolve more often is because of evolutionary accessibility, not because of functional superiority. ARs remain rare even under these initialization conditions that favor them more **(Fig. S5)**.

We also tested how the strength of the input signal affects evolutionary outcomes. As predicted, a strong input signal, which is conducive to the negative regulatory path of an I1FFL, promotes the evolution of I1FFLs and correspondingly reduces that of NFBLs (**Fig. S6A**). Again, if one motif is prevented from evolving, the other can yield similar fitness (**Fig. S6B**).

One aspect of how we select for a pulse generator is to select for rapidity in generating the pulse, assessed as the time taken, after the signal appears, to reach half-peak effector concentration (see Methods). In the results discussed so far, we assume no further advantage for response times shorter than 60 minutes. When we instead select for TRNs that can generate a pulse as fast as possible, I1FFLs become the dominant solution regardless of how high a peak effector expression level we select for (**Fig. S7A**). This result is expected: in I1FFLs, the effector and the repressor can start their expression simultaneously, making it easier to rapidly generate the pulse. What is more surprising is that when I1FFLs were prevented from evolving, the fitness of the best TRNs was just as high (**Fig. S7B**).

Evolution of I1FFLs and NFBLs follows different trajectories

To understand how the topology of TRNs evolves, we looked at the early stages of evolution. The combined frequency of the two motifs rises throughout the long period of evolution, rather than topological solutions being found early and becoming locked in and only incrementally improved on (**Fig. 4**). This suggests that solutions are not easy to find. The relatively early advantage of I1FFLs in the low-

peak and medium peak conditions (**Fig. 4A, B**) compared to the relatively late advantage to NFBLs in the high peak condition (**Fig. 4C**) is again compatible with the idea that I1FFLs are evolutionarily more accessible.

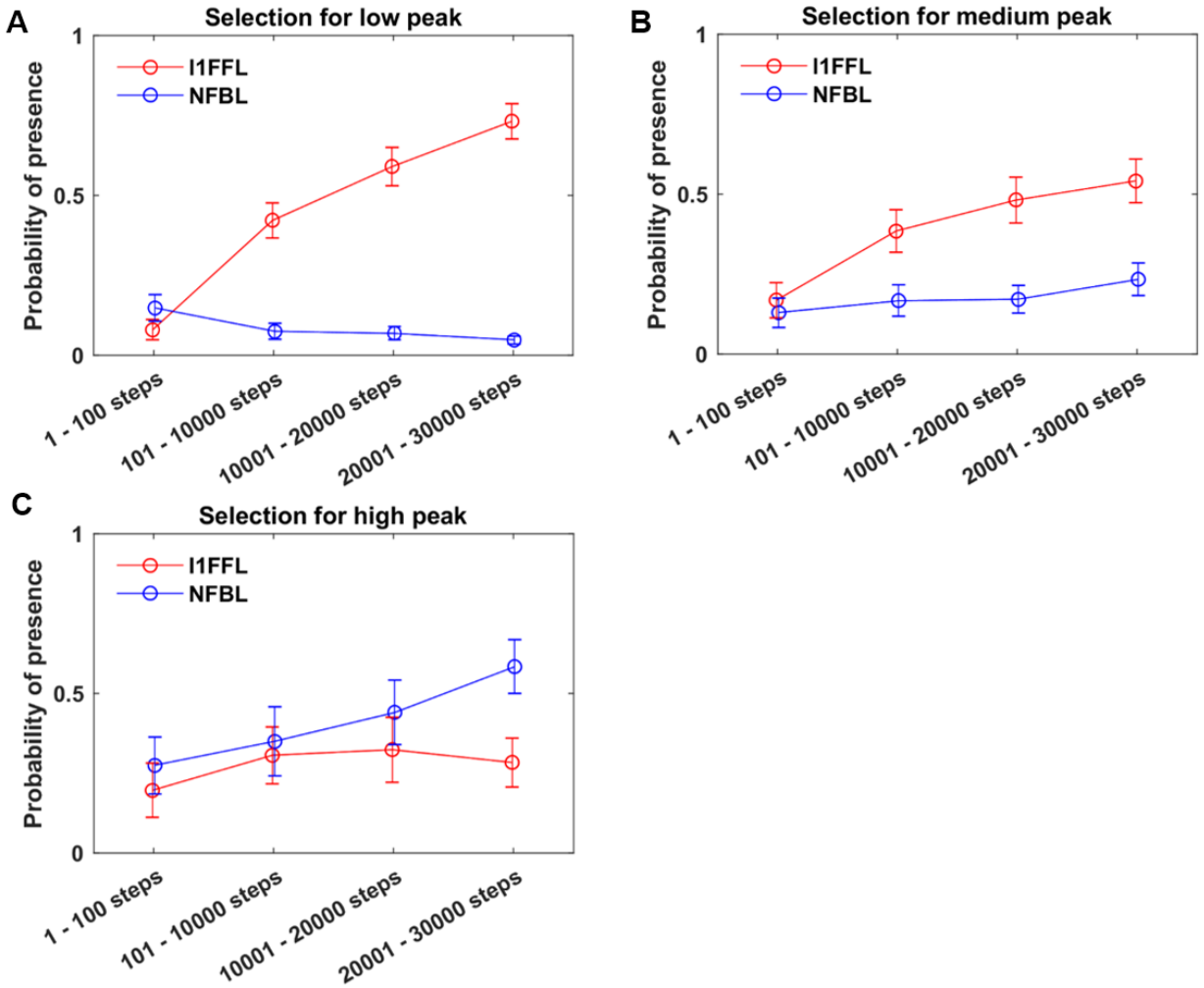

Figure 4. Evolution of I1FFLs and NFBLs follow different trajectories. We score motif occurrence during different time periods along the way to the evolution of the high-fitness replicates shown in **Fig. 2A**. As in **Fig. 2A**, we calculated the proportion of evolutionary steps that contain at least one network motif of the specified type. Data are shown as mean ± SE over replicates.

When we characterized mutations that create functional motifs during the simulations, we found that evolution tried similar numbers of I1FFL-creating mutations and NFBL-creating mutations under low-peak conditions (**Table 1**); it is the higher acceptance rate of I1FFL-creating mutations that drives the evolution of I1FFLs. Similarly, I1FFL-destroying mutations are accepted less often than NFBL-destroying mutations are, in this case regardless of target peak expression (**Table S1**).

peak level		Evolutionary step 101 – 20,000			Evolutionary step 20,001 – 30,000		
		Trialed	Accepted	Acceptance rate	Tried	Accepted	Acceptance rate
Low	I1FFL-creating	2,386	382	0.160	1,041	136	0.131
	NFBL-creating	2,843	196	0.069	1,245	74	0.059
	all mutations	110,852	19,900	0.180	67,181	10,000	0.149
Medium	I1FFL-creating	4,218	417	0.099	2,798	203	0.072
	NFBL-creating	4,528	385	0.085	2,062	163	0.079
	all mutations	111,572	19,900	0.178	65,760	10,000	0.152
High	I1FFL-creating	3,121	259	0.083	2,080	135	0.065
	NFBL-creating	7,620	516	0.068	4,246	305	0.072
	all mutations	120,316	19,900	0.162	70,066	10,000	0.143

Table 1. Summary of mutations that create I1FFLs and/or NFBLs. We identify mutations that increase the number of I1FFLs and/or NFBLs in a TRN to above zero (see Methods for details). For each selection condition and evolutionary stage, we show the average numbers of the qualified mutations that are trialed and/or accepted across all high-fitness replicates shown in Fig. 2A. For comparison, we also show the number of total mutations that are trialed and/or accepted per simulation, many of which will be neutral or nearly neutral. The evolutionary algorithm we use will frequently accept neutral mutations. The acceptance rate shown in the table is the number of accepted mutations across all replicates divided by the number of trialed mutations across all replicates.

Most mutations that create NFBLs, regardless of whether they are accepted, do so by destroying an I1FFL-NFBL conjugate (**Table 2**). Note that "destruction" can mean that a weak affinity binding site that is part of the topology of the I1FFL-NFBL conjugate changes from being critical to function to not critical to function (**Fig. S2**), rather than meaning that the topological connection is removed entirely. The partially destructive nature of such a mutation likely explains its low acceptance rate. The low evolutionary accessibility of NFBLs compared to I1FFLs is reflected in the fact that so many evolutionary paths toward them involve an I1FFL along the way.

It takes considerable evolutionary time to evolve toward high effector expression levels (**Fig. S8**). The advantages of the NFBL only kick in at high expression levels, likely explaining some of the difficulty in evolving NFBLs. In **Table 2**, we see that tolerance of the destruction of I1FFL-NFBL conjugates (assessed in Accepted/Trialed column) is higher for NFBL-creating mutations and for selection for high peak effector expression. Tolerance of conjugate-destroying NFBL-creating mutations increases over evolutionary time in the high peak condition, compatible with a change in evolutionary accessibility once effector expression levels have evolved high enough (**Table 2**). In all three selective conditions, the relative tendency to accept I1FFL-creating rather than NFBL-creating mutations is higher earlier than later (**Table 1**), albeit to a small degree for the low peak selective condition in which there is limited evolution toward higher effector expression (**Fig. S8**).

peak level		Evolutionary step 101-20000			Evolutionary step 20001-30000		
		trialed	accepted	Accepted/trialed	trialed	accepted	Accepted/trialed
Low	I1FFL-creating	0.483	0.317	0.656	0.558	0.345	0.618
	NFBL-creating	0.908	0.717	0.789	0.859	0.657	0.765
Medium	I1FFL-creating	0.670	0.389	0.580	0.610	0.404	0.662
	NFBL-creating	0.931	0.814	0.875	0.914	0.816	0.894
High	I1FFL-creating	0.570	0.167	0.293	0.577	0.254	0.440
	NFBL-creating	0.967	0.869	0.899	0.952	0.905	0.950

Table 2. Most NFBL-creating mutations also destroy I1FFL-NFBL conjugates. Numbers indicate the fraction of I1FFL-NFBL-conjugate-destroying mutations among mutations or among accepted mutations that create the motif in question. The numbers are calculated by pooling qualified mutations from all simulation replicates during a given evolutionary stage. See Methods for details about the identification of mutations that create and/or destroy motifs.

Highly expressed effectors tend to be regulated by NFBLs in yeast

Next we test our model predictions about when I1FFLs vs. NFBLs will evolve. From the TRN of *S. cerevisiae*, we compiled relevant motifs for which all nodes were TFs with known expression level. We thereby obtained 42 I1FFLs, 35 NFBLs, and an additional 5 instances of co-occurring I1FFLs and NFBLs. Of these I1FFLs and NFBLs, 5 and 11 also contain a self-repressing effector. We did not find AR in the absence of both of the other motifs. In agreement with model predictions, the effector TF generally has a higher expression level in NFBLs than in I1FFLs, while both the input TF and the auxiliary TF have similar expression levels in the two network motifs (**Fig. 5**). Taken together, whether I1FFLs and NFBLs evolve in *S. cerevisiae* may be affected by the concentration of the effector rather than the input TF.

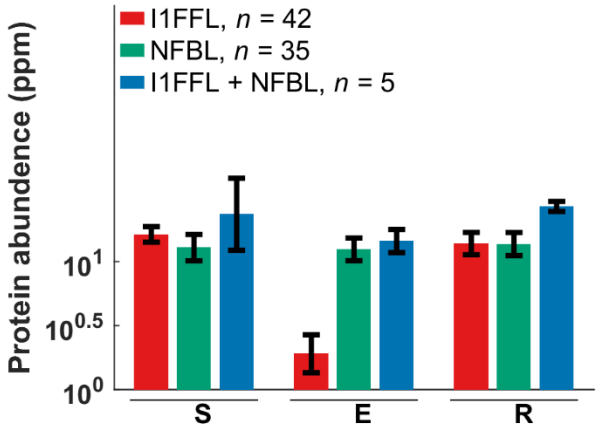

Figure 5. The effector TF has higher expression in NFBLs of *S. cerevisiae* than in I1FFLs. We searched the TRNs of *S. cerevisiae* for I1FFLs, NFBLs, and their conjugates, using data from Yeastract. Protein levels were taken from the "GPM, Aug, 2014" dataset provided by PaxDB (Wang et al, 2015). Similar results were obtained using two more datasets from PaxDB (**Fig. S9**). We take the geometric mean of protein abundance (ppm) for each network position across all instances of the motif found in the TRN of *S. cerevisiae*, and associated SE from log-transformed data.

Discussion

We selected for a pulse generator in an evolutionary simulation model and observed which TRN motifs emerged. As predicted, selecting for high peak expression level of the effector promotes NFBLs over I1FFLs, while a strong input signal promotes I1FFLs. However, if one motif is prevented from evolving, the other motif can evolve to take its place, with no loss of peak fitness, suggesting that the preference between motifs is not adaptive in origin. One predicted pattern is confirmed in the actual TRN of *S. cerevisiae*, where the effector's expression level is higher in NFBLs than in I1FFLs.

We can interpret the evolution of TRNs as ascending a fitness landscape with many local peaks. Which peak is reached depends not only on how high each peak rises, but also on how likely an evolutionary pathway is to begin the climb toward that peak. If we consider two networks motifs that are both capable of generating pulses, evolution is biased towards the one that is easier to find via adaptive, neutral, and nearly neutral mutations searching the fitness landscape. Many previous works have focused on the role of mutation bias at a snapshot in time (Gomez et al, 2020; Stoltzfus & McCandlish, 2017; Yampolsky & Stoltzfus, 2001); evolutionary accessibility is also an emergent property of complete evolutionary paths on landscapes (Bertram & Masel, 2020; Sorrells & Johnson, 2015).

We did not pit the performance of I1FFLs and NFBLs versus the accessibility of the two motifs, because both motifs had indistinguishable performance in our system. We are therefore unable to answer whether the evolutionary accessibility of motifs can alter the evolutionary outcome predicted by performance of motifs. However, finding in our system that high fitness solutions can often be found one way or another is intriguing in its possible generality.

We find that most NFBLs evolve not from connecting previously disconnected genes (e.g. S->E->R), but rather from uncoupling I1FFL-NFBL conjugates in favor of a pure NFBL. We simulate only relatively small TRNs, due to limitations in computational power, and this might restrict the evolutionary trajectories that are capable of generating network motifs. If simulations algorithms that scaled better with TRN size were devised, it would be interesting to explore whether network motifs would evolve via different trajectories in a large TRNs. For example, the use of the same TF for multiple regulatory purposes in real-world TRNs, which of course are larger, can constrain network evolution, requiring complex trajectories to achieve a new regulatory function (Sorrells et al, 2015).

We predicted via simulations that a highly expressed effector promotes the evolution of NFBLs over I1FFLs. Strikingly, this prediction was borne out in empirical data from yeast. A highly expressed TF can more strongly regulate its target, and/or reduce the amount of noise propagated downstream (Jothi et al, 2009; Pedraza & van Oudenaarden, 2005). Once a highly expressed TF gains a TFBS in the target gene, the TFBS may also be easier to retain during evolution. Many studies on TRNs have noted a systematic difference among the expression levels of genes at topologically different positions (Gerstein et al, 2012;

Herrgård et al, 2003; Jothi et al, 2009; Yu et al, 2003), and that highly expressed TFs are often regulators of multiple target genes (Gerstein et al, 2012; Jothi et al, 2009). Our findings also support the idea that the observed network motifs in TRNs are partially shaped by the expression levels of TFs.

Methods

Fitness

In stage 1, low signal levels last for 120 + x minutes, where x is random number drawn from an exponential distribution truncated at 30, and with an un-truncated average of 10. This random duration ensures that TRNs to respond to a change in the signal, rather than evolve a timer mechanism. Stage 2 lasts for 240 minutes. The highest level of effector during stage 2 is denoted p.

The fitness of a TRN has 4 components that are based on the expression of the effector. We note that the model by Xiong et al. (2019, Supplementary Methods/Mutation) allow genes to duplicate and diverge. For the purpose of scoring effector concentration and hence fitness, we use the total protein level of all effector proteins, including those that have diverged to have different regulatory activities. Fitness component 1 ensures the effector matches a pre-defined peak level:

$$f_1 = e^{-\frac{(log(p)-log(p_{opt}))^2}{\sigma^2}}, \quad (1)$$

where p_{opt} is a pre-defined optimal peak level. We set p_{opt} to 5,000 molecules per cell, 10,000 molecules per cell, and 20,000 molecules per cell, corresponding to selection for a low, medium, and high peak level, respectively. The number 10,000 is the average number of a metabolism-related protein per yeast cell (Ghaemmaghami et al, 2003). In other words, we assume that the effector is a metabolism-related protein. The effector also acts as TF; this kind of dual functionality is not uncommon in yeast (Gancedo & Flores, 2008). We set $\sigma^2 = 0.693$ so that when $p = 0.5p_{opt}$, $f_1 = 0.5$.

Fitness component 2 ensures that effector expression level is low at the end of stage 1.

$$f_2 = \begin{cases} \frac{p-s_1}{0.9p}, s_1 > 0.1p \\ 1, s_1 \leq 0.1p \end{cases}, \quad (2)$$

where s_1 is the average effector level over the last 5 minutes of stage 1.

We use fitness component 3 to ensure a rapid turn-on of effector:

$$f_3 = \begin{cases} \frac{240-t_{half_peak}}{240-t_{saturate}}, t_{half_peak} > t_{saturate} \\ 1, t_{half_peak} \leq t_{saturate} \end{cases} \quad (3)$$

where t_{half_peak} is the latest time at which the effect level is at $0.5(s_1 + p)$ after the signal increases and before it hits its peak and $t_{saturate}$ sets a time where further increasing the turn-on speed of the effector no longer increases fitness. We set $t_{saturate}$ to 60 minutes by default, but also explored the evolutionary outcome at $t_{saturation} = 0$ minutes.

To select for the downward slope of a pulse, fitness component 4 ensures expression of the effector falls below 80% of the peak level by the end of stage 2:

$$f_4 = \begin{cases} \frac{s_2}{0.8p}, s_2 < 0.8p \\ 1, s_2 \geq 0.8p \end{cases}, \quad (4)$$

where s_2 is the average effector level in the last 5 minutes of stage 2.

At each point of growth simulation, gene expression incurs a cost that is proportional to the total rate of translation of all genes (Xiong et al, 2019, Supplementary Methods/Cost of gene expression). The estimated fitness of a TRN from one gene expression simulation is the arithmetic mean of the four components minus the cumulative cost of gene expression throughout the last 360 minutes of a growth simulation.

Evolution

As in Xiong et al. (2019), we calculate the relative fitness $f_{resident}$ of the resident genotype with 1,000 growth simulations. The relative fitness f_{mutant} of the mutant is calculated with 200 growth simulations. If f_{mutant} satisfies Eq. 5 below, we replace the resident with the mutant, and re-calculate the fitness of the new resident with another 800 replicate of growth simulation. If f_{mutant} does not satisfy Eq. 5, we discard the mutant and generate a new mutant from the resident genotype. This process is repeated until a replacement happens. We call a replacement an evolutionary step.

$$\frac{f_{mutant} - f_{resident}}{|f_{resident}|} \geq 10^{-8} \quad (5)$$

Each evolutionary replicate continues for 50,000 evolutionary steps, by which time the fitness of the resident genotype has usually plateaued.

Note that our evolutionary algorithm was designed to waste as few beneficial mutations as possible in order to minimize computation. In conventional population genetics, the probability of fixation of a beneficial new mutation is only around twice the selection coefficient (Haldane, 1927); we attempted to accept most of the beneficial mutations we trial, and as a result also many mutations that are neutral or slightly deleterious. In our algorithm, gene expression is stochastic, and hence our assessment of fitness is stochastic, which also makes our fixation probability depend on the strength of selection. The amount of noise in this relationship is set by the number of gene expression simulations, which is smaller than typical values of the effective population size that sets the degree of stochasticity in conventional population genetics.

Mutations

Once a gene has acquired 4 copies, we no longer allow it to duplicate. Once the total number of non-effector TF gene copies reaches 21, we no longer allow any duplications of non-effector TF genes. This helps keep computational demands down, since they scale non-linearly with network size. We never delete or duplicate the signal, nor delete the last copy of the effector.

Counting network motifs

The model allows genes to diverge following duplication. For the purpose of scoring motifs, gene duplications are considered “copies” of the same gene, encoding “protein variants”, until there is a change to preferred binding sequence or to activator vs. repressor identity, at which point they are considered different proteins.

We count I1FFLs and NFBLs formed by the signal, an effector gene, and a repressor gene that is different from the effector gene, and ARs formed by the signal and an effector gene. We score gene A as potentially regulating gene B if there is a TFBS for A in the cis-regulatory sequence of B. A random cis-regulatory sequence carries on average around one two-mismatch TFBS for each TF (see Xiong et al. (2019) for detailed calculation), which because of the weakness of the binding, might have little regulatory effect. **Fig. S11** illustrates five scenarios in which scoring this weak TFBS might give rise to an apparent motif. We therefore consider two-mismatch TFBSs to be potentially "spurious" or non-functional. To test whether a two-mismatch TFBS is in fact spurious, we recalculate the fitness of the TRN without that TFBS in 1,000 growth simulations, and deem it spurious if the recalculated fitness is at least 99% of the original fitness. Spurious connections are ignored while scoring network motifs.

We allow genes in I1FFLs and NFBLs to self-regulate. An overlapping I1FFL in which the effector and the auxiliary TF repress each other is not considered as two I1FFLs. Overlapping I1FFLs evolve rarely.

Mutations that create and destroy motifs

For each evolutionary replicate, we identified the evolutionary steps at which the number of a given motif changes to or from zero. We removing spurious TFBSs before scoring motifs, as described in the last section, with the exception that, to save computation, when examining mutations that were trialed and then rejected by selection, we used only 200 growth simulations to determine fitness, with a threshold of 98% of original fitness. The mutations that are accepted at these evolutionary steps are called motif-creating-mutations and motif-destroying-mutations, respectively. Mutations that change the expression levels of a gene and/or the binding affinity of a TF can potentially change whether a 2-mismatch TFBS is "spurious" in terms of fitness effects, effectively rewiring the TRNs even if they do not create or destroy core TFBSs of the motif in question.

Expression levels of TFs in yeast TRN

We used YeastMine (Balakrishnan et al, 2012) to retrieve 129 *S. cerevisiae* genes that have the GO term "DNA-binding transcription factor activity" or children of this GO term. We then searched Yeastract (Teixeira et al, 2006) for TFs that regulate these 129 TFs, demanding evidence from both DNA binding and gene expression. When the search found new TFs that are not included in list given by YeastMine, the new TFs were added to the list and fed to Yeastract again. We stopped the iterative search when no new TFs were found, and the final list has 203 TFs, 180 of which have protein levels available in an integrated protein abundance data set, namely the "GPM, Aug, 2014" set, from PaxDB (Wang & Purisima, 2005). Yeastract annotates interactions between pairs of TFs as activating, repressing, or both. When annotated as "both" (i.e. likely condition-specific), we interpreted it as whichever interaction mode would be needed in order to complete a motif. We scored I1FFLs, NFBLs, and their conjugates from all combinations of three TFs out of the 180, allowing E and/or R to be self-repressing. We also score ARs from all combinations of two TFs. Note that, by restricting the search to only TF genes, we have forced an effector to also be a TF, but the effector of an I1FFL does not have to also be a TF in natural biological networks.

Acknowledgement

We acknowledge the generous computational resources provided by the High Performance Computation Center of the University of Arizona.

Data Availability

The source code for our computational model is available at https://github.com/MaselLab/network-evolution-simulator/tree/I1_FFLs.

Supplementary materials

peak level		Evolutionary step 101 – 20,000			Evolutionary step 20,001 – 30,000		
		Trialed	Accepted	Acceptance rate	Trialed	Accepted	Acceptance rate
Low	I1FFL-destroying	6,005	381	0.064	3,004	136	0.045
	NFBL-destroying	1,450	196	0.135	617	74	0.120
Medium	I1FFL-destroying	6,268	417	0.067	2,798	203	0.072
	NFBL-destroying	3,389	385	0.114	1,698	163	0.096
High	I1FFL-destroying	4,169	259	0.062	2,148	135	0.063
	NFBL-destroying	4,157	516	0.124	2,533	305	0.120

Table S1. Summary of mutations that remove all I1FFLs and/or NFBLs. For each selection condition, we show the average numbers of the qualified mutations that are trialed and/or accepted across all simulation replicates during the given stage of evolution. The acceptance rate shown in the table is the number of accepted mutations across all replicates divided by the number of trialed mutations across all replicates.

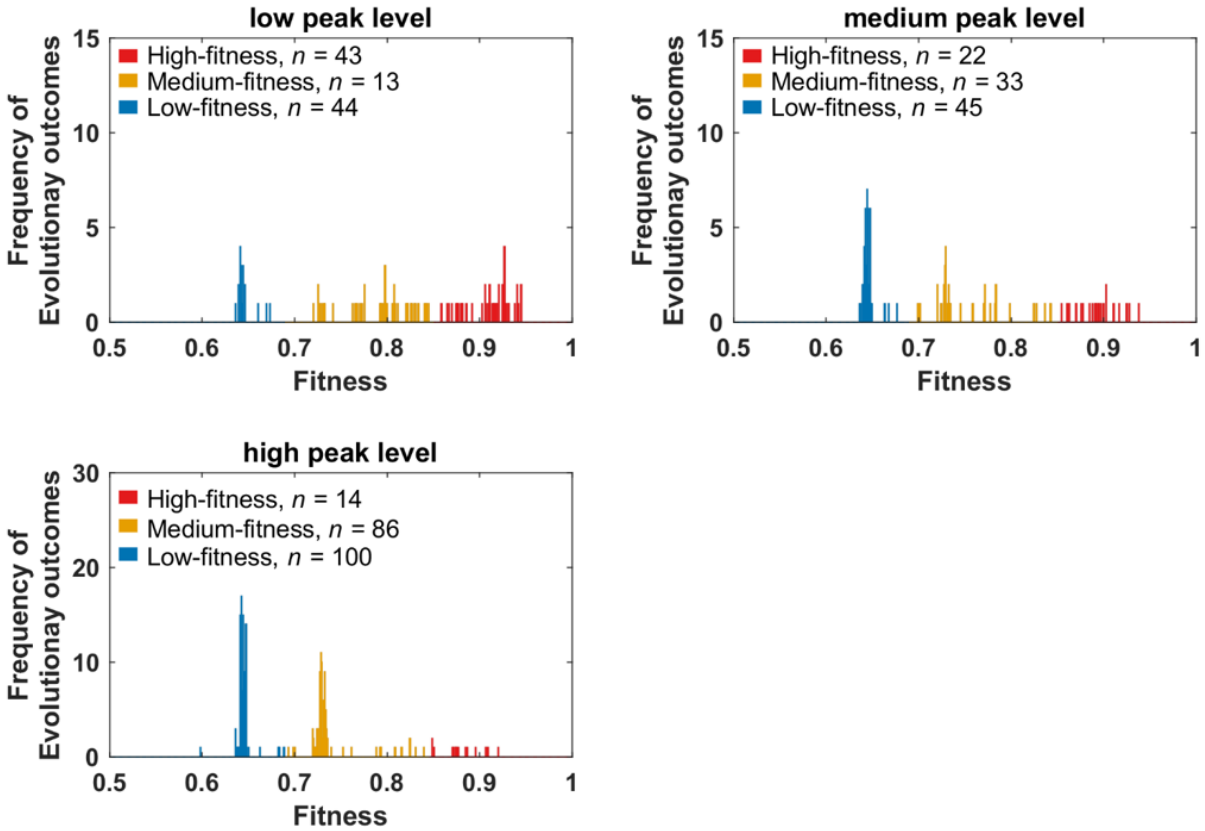

Figure S1. Genotypes evolved with different optimal peak levels of the effector. Fitness distributions of genotypes. We ran 100 evolutionary simulations for the low-peak and the medium-peak conditions, and 200 for the high-peak condition. In all three peak levels, genotypes with fitness above 0.845 are considered as high-fitness genotypes and are further analyzed in **Fig. 2**. We used a fitness of 0.69 to separate medium-fitness genotypes and low-fitness genotypes.

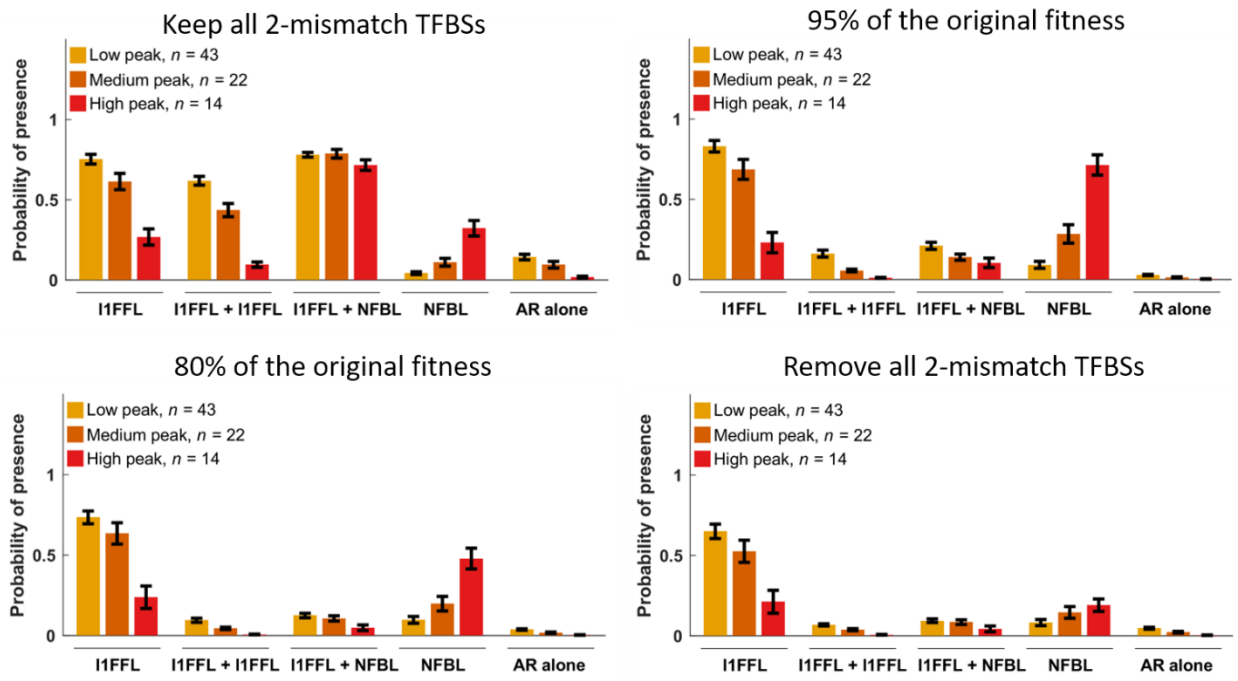

Figure S2. The relative occurrences of motifs are similar when we removed 2-mismatch TFBSs by using different thresholds. The replicates are the same as the high-fitness replicates shown in Fig. 2A. For comparison, we also show the occurrences of motifs when all 2-mismatch TFBSs are kept or removed. Data are shown as mean ± SE over replicates.

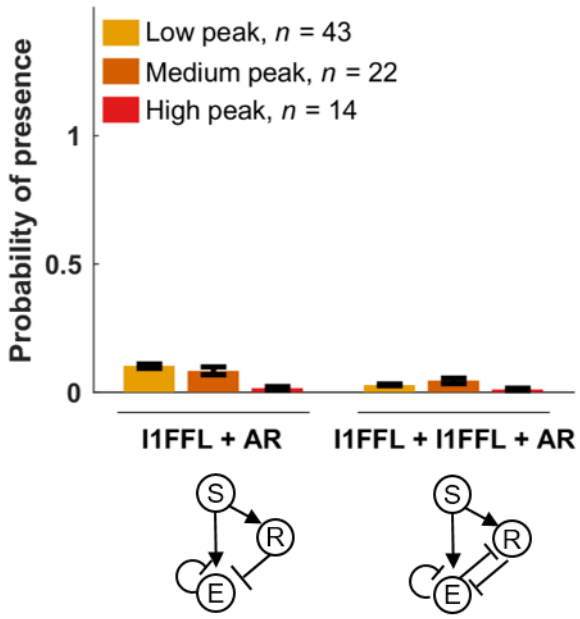

Figure S3. Auto-repression (AR) rarely evolves with I1FFLs or overlapping I1FFLs. In high-fitness genotypes evolved under selection for pulse generation, there were few auto-repressing effectors co-occurring with other motifs, and for simplicity, they were therefore grouped in **Fig. 2** with the motif with which they co-occurred. We note that when the repressor of an I1FFL-NFBL conjugate is an effector, this effector can form auto-repression. We classified such case as a stand-alone AR, because from the

perspective of this effector, it is not regulated by an I1FFL, NFBL, overlapping I1FFL, or I1FFL-NFBL conjugate. Data are shown as mean ± SE over replicates.

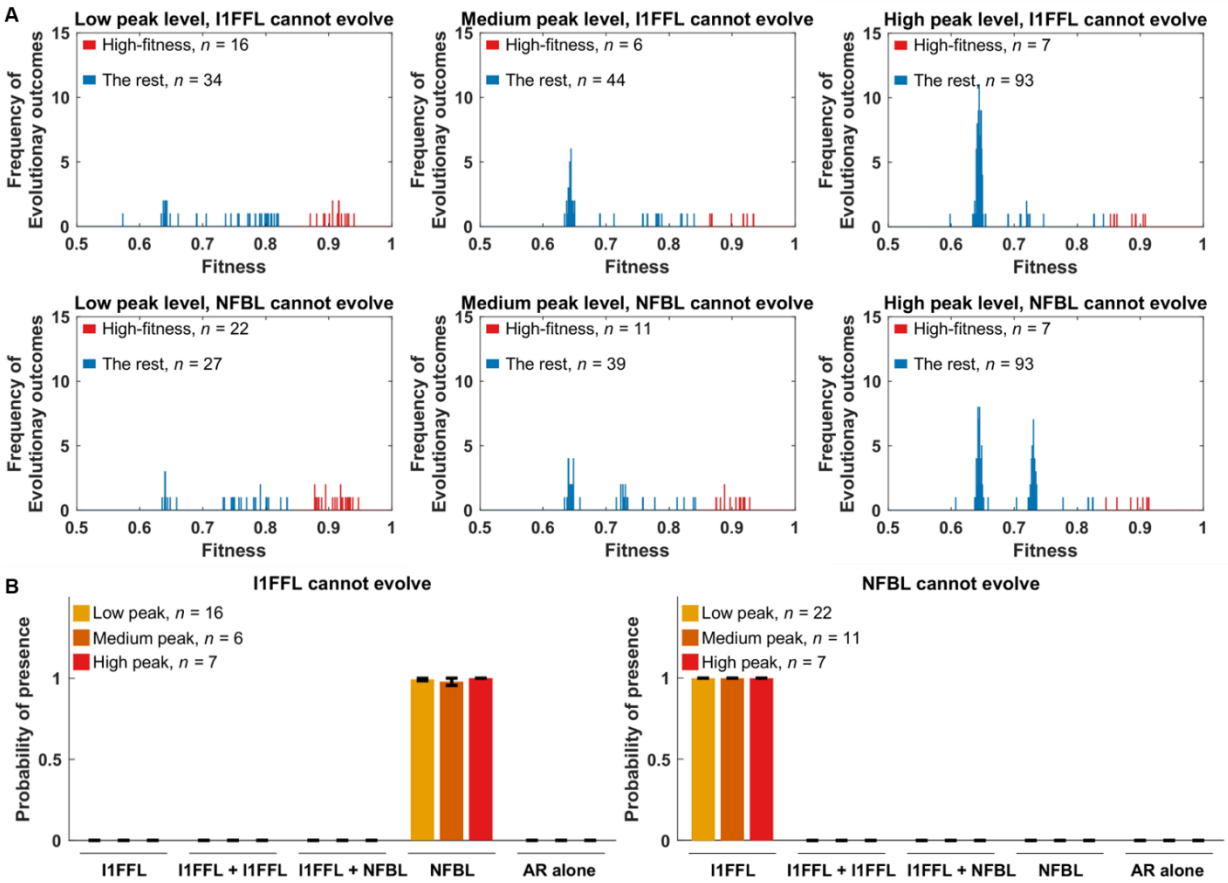

Figure S4. Fitness distributions and network occurrences of genotypes with restricted solutions under different optimal peak levels. (A) Each panel under low peak and medium peak selection had 50 evolutionary simulations, and the two panels under high peak each had 100 evolution simulations. Under the condition where we select for a low peak and forbid NFBLs to evolve, we removed one simulation that was terminated pre-maturely before evolving 50,000 evolutionary steps. This particular simulation failed to find a mutant that has higher fitness than the standing phenotype even after 2,000 trials, and was terminated according to the design of our program. To classify a genotype as high-fitness (red), we apply the same fitness cutoff as in **Fig. S1**. The average fitness of the high-fitness genotypes is shown in **Fig. 3B**. See legend of **Fig. 2B** for description of modifications to prevent the evolution of NFBLs or I1FFLs. **(B)** In the high-fitness genotypes, when either I1FFL or NFBL is not allowed to evolve, the other motif almost always evolves. Data are shown as mean ± SE over replicates.

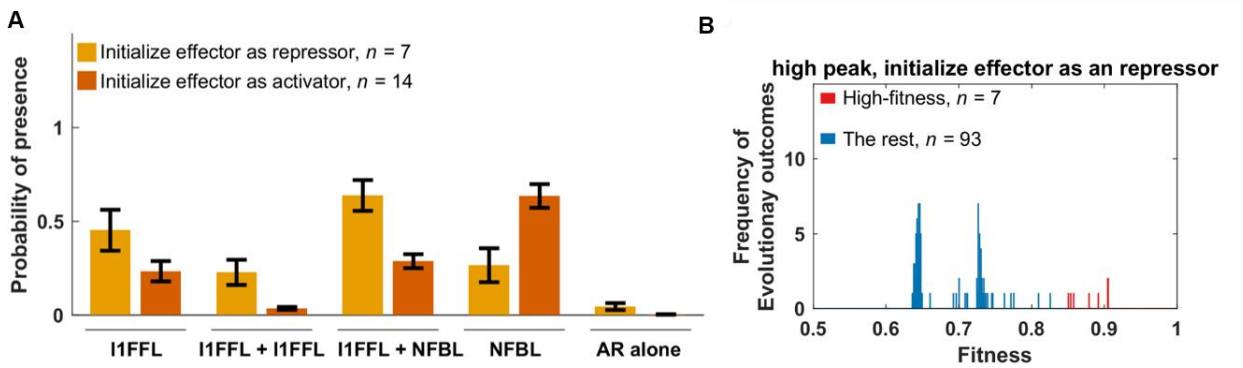

Figure S5. Initializing the effector as a repressor facilitates the evolution of I1FFLs. We repeated evolution under selection for a high peak level of the effector, but initialized the effector as a repressor. **(A)** Motif occurrence when the effector is initialized as a repressor vs. as an activator (latter data as in **Fig. 2**). Data are shown as mean ± SE over replicates. **(B)** Fitness of the evolved TRNs. Similar to Fig. S1, TRNs with a minimal fitness of 0.845 are considered high-fitness.

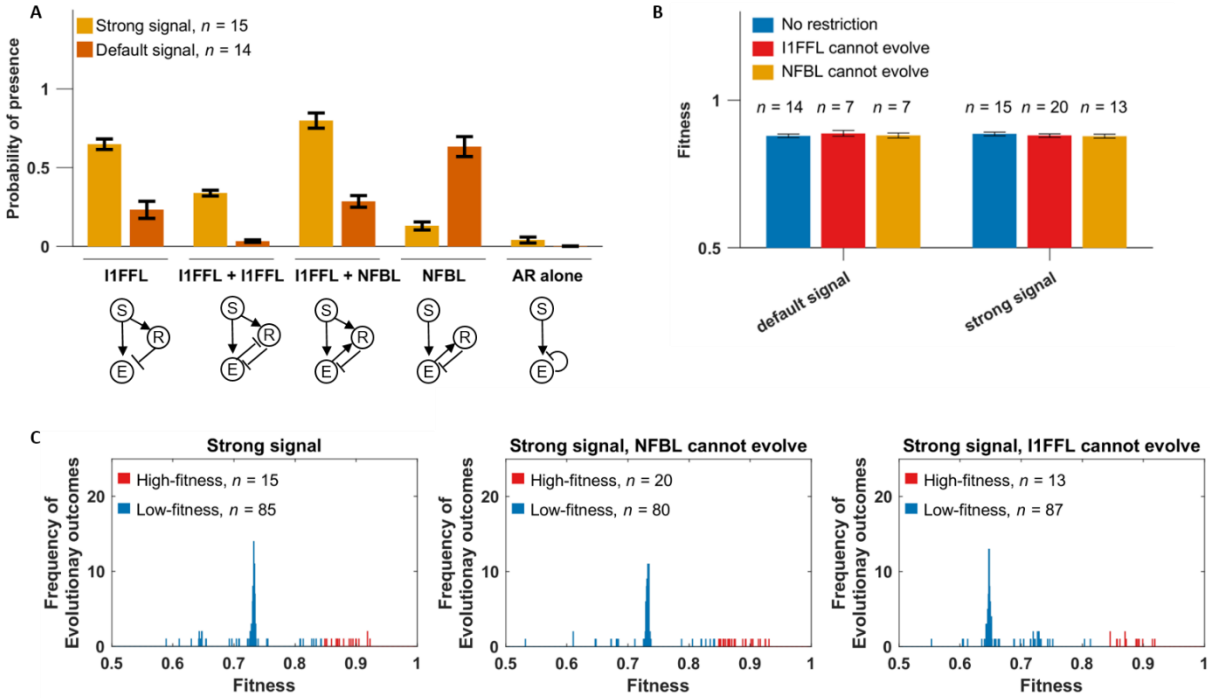

Figure S6. A strong signal promotes I1FFLs. We tested evolution under the default signal, where signal strength increases from 100 molecules per cell to 1,000 molecules per cell, and a strong signal, where the signal strength increases from 1,000 molecules per cell to 10,000 molecules per cell. **(A)** Occurrence of different motifs in high-fitness genotypes. **(B)** Either I1FFLs or NFBLs alone can yield similar fitness under a given signal regime. Data are shown as mean ± SE over replicates. **(C)** Fitness distribution of genotypes evolved under the strong signal without (leftmost) or with (middle and rightmost) restrictions on evolution. Similar to Fig. S1 and S3, high-fitness genotypes must have fitness greater or equal to 0.845.

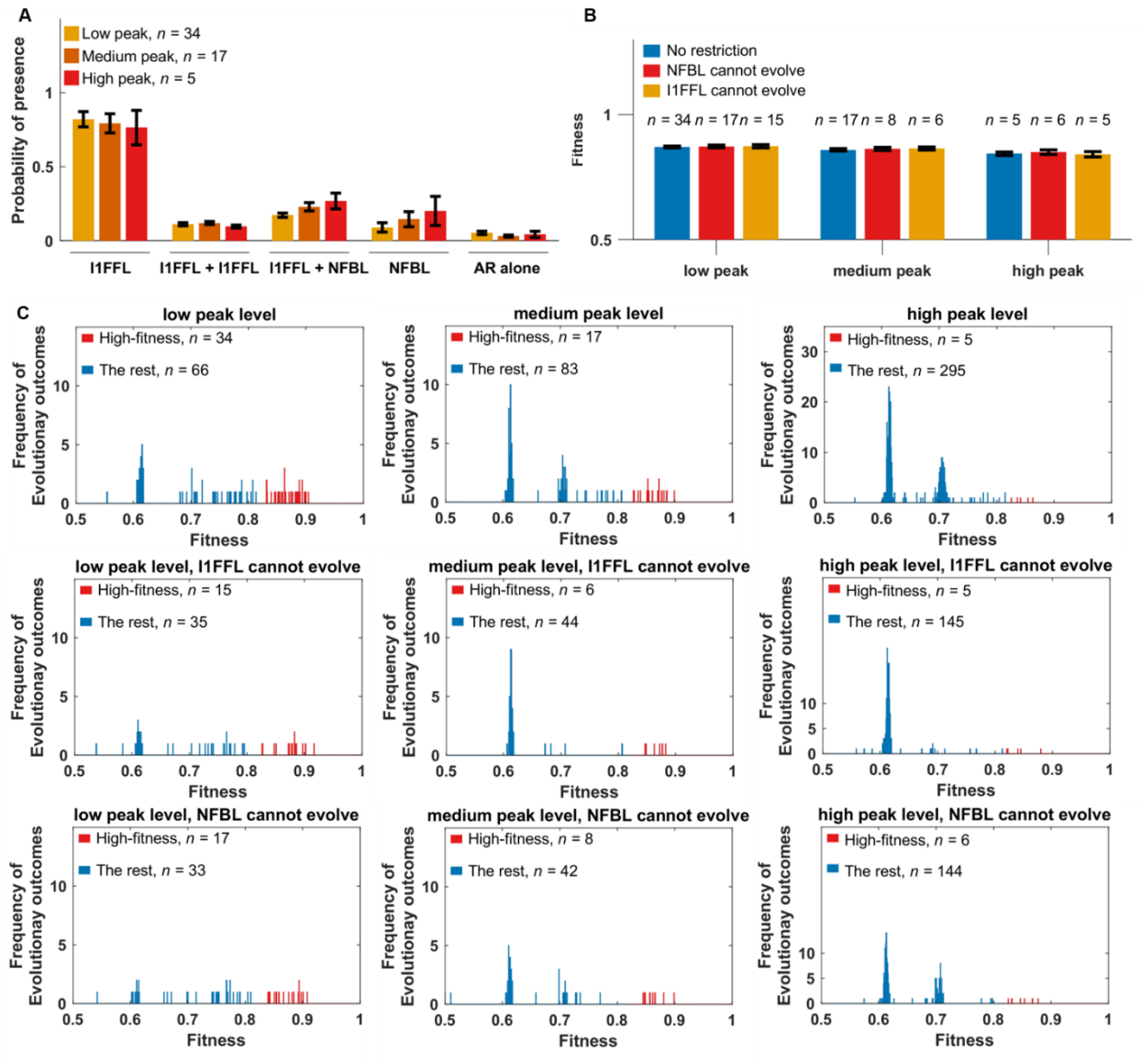

Figure S7. Selection for a fast response promotes I1FFLs. Similarly to Fig. 2, we tested evolution under selection for three peak levels, but we set $t_{saturate}$ to 0 to select for TRNs that generate a peak as soon as possible. **(A)** Occurrence of different motifs in high-fitness genotypes. **(B)** Either I1FFLs or NFBLs alone can yield similar fitness under a given signal regime. Data are shown as mean ± SE over replicates. **(C)** Fitness distribution of genotypes evolved under selection for the three peak levels, and when either I1FFLs or NFBLs were not allowed to evolve under each peak level. Because changing the value of $t_{saturate}$ also alters the fitness landscape, we use a fitness cutoff of 0.82 to classify high-fitness genotypes throughout this figure.

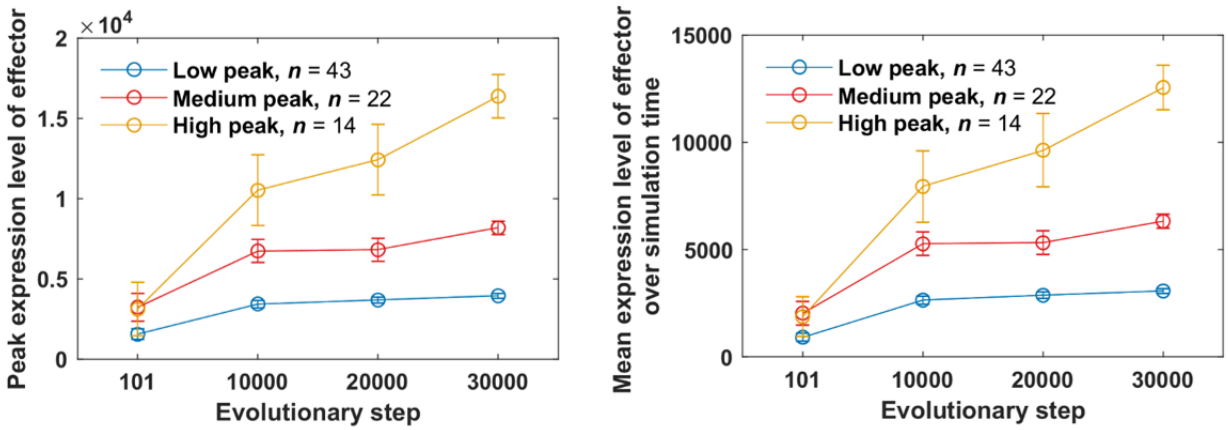

Figure S8. Peak expression levels of effector. For each high-fitness replicate shown in Fig. 2A, we average the peak protein levels of the effector over 200 replicate simulations of gene expression. Data are shown as mean ± SE over replicates.

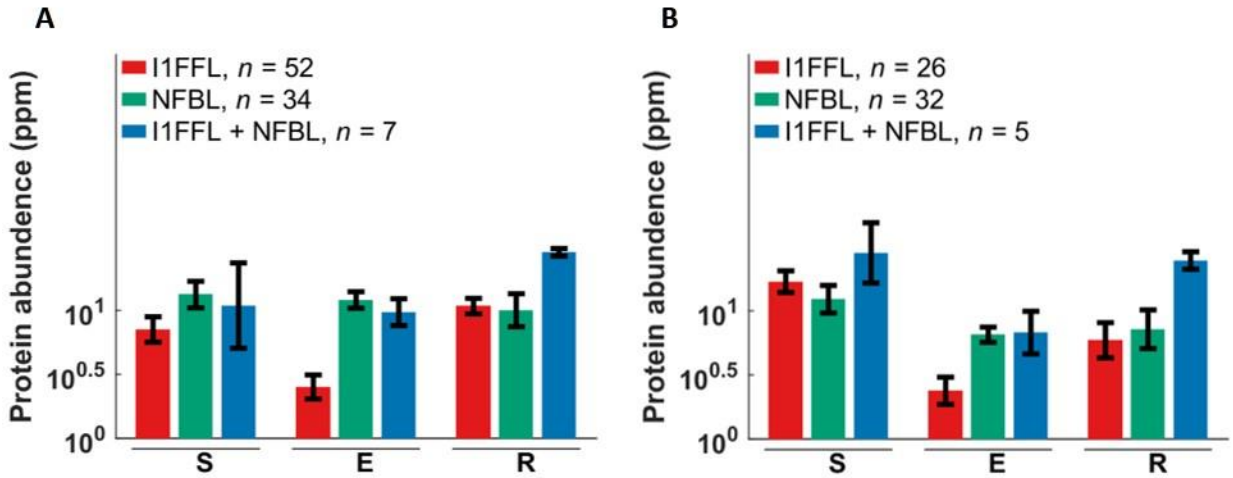

Figure S9. Alternative data for protein abundance also shows that the effector have higher expression in NFBLs of *S. cerevisiae* than in I1FFLs. We took protein expression levels from two alternative datasets provided by PaxDB (Wang et al, 2015), the "PeptideAtlas, March, 2013" **(A)** and the "PeptideAtlas, May, 2009" **(B)**. The "2013" datasets provides the expression levels of 163 out of the 203 TFs, and the "2009" datasets provides 138. Similarly to **Fig. 5**, we searched for I1FFLs, NFBLs, and their conjugates in the two TRNs formed by the 163 and 138 TFs, respectively. We show the geometric mean of protein abundance (ppm) for each network position across all instances of the motif. Data are shown as mean ± SE over instances of network motifs. SEs were calculated from log-transformed data.

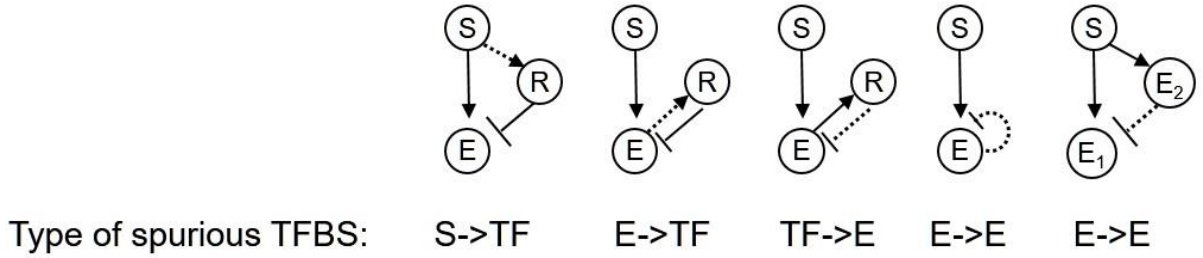

Figure S10. Five scenarios in which apparent but non-functional network motifs can arise from spurious TFBSs. A TFBS containing 2 mismatches can easily appear by chance in a cis-regulatory sequence, but may be deemed spurious if it has negligible functional effect. Spurious E->E TFBSs where both "Es" represent the same effector gene give rise to apparent ARs, whereas if they represent different effector genes, they give rise to I1FFLs.

References

Adler M, Mayo A, Alon U (2014) Logarithmic and Power Law Input-Output Relations in Sensory Systems with Fold-Change Detection. *PLoS Comput Biol* **10:** e1003781

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

Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, Amariglio N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y (2007) A module of negative feedback regulators defines growth factor signaling. *Nature Genet* **39:** 503-512

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

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Michael Cherry J (2012) YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database***:** bar062

Basu S, Mehreja R, Thiberge S, Chen M-T, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA* **101:** 6355-6360

Bertram J, Masel J (2020) Evolution rapidly optimizes stability and aggregation in lattice proteins despite pervasive landscape valleys and mazes. *Genetics* **214:** 1047-1057

Buzi G, Khammash M (2016) Implementation Considerations, Not Topological Differences, Are the Main Determinants of Noise Suppression Properties in Feedback and Incoherent Feedforward Circuits. *PLoS Comput Biol* **12:** e1004958

Çağatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM (2009) Architecture-Dependent Noise Discriminates Functionally Analogous Differentiation Circuits. *Cell* **139:** 512-522

Camas FM, Blázquez J, Poyatos JF (2006) Autogenous and nonautogenous control of response in a genetic network. *Proc Natl Acad Sci USA* **103:** 12718-12723

Ferrell JE (2016) Perfect and Near-Perfect Adaptation in Cell Signaling. *Cell Systems* **2:** 62-67

Gancedo C, Flores C-L (2008) Moonlighting proteins in yeasts. *Microbiology and Molecular Biology Reviews* **72:** 197-210

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu J, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ,

Cheng Y et al (2012) Architecture of the human regulatory network derived from ENCODE data ENCODE Encyclopedia of DNA Elements. *Nature* **488**

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

Gomez K, Bertram J, Masel J (2020) Mutation bias can shape adaptation in large asexual populations experiencing clonal interference. *Proc R Soc Lond, B, Biol Sci* **287:** 20201503

Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* **23:** 838-844

Herrgård MJ, Covert MW, Palsson BØ (2003) Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res* **13:** 2423-2434

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

Jothi R, Balaji S, Wuster A, Grochow JA, Gsponer J, Przytycka TM, Aravind L, Madan Babu M (2009) Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol Syst Biol* **5:** 294

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

Kuo PD, Banzhaf W, Leier A (2006) 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

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK et al (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799-804

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

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

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

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

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

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

Rosenfeld N, Elowitz MB, Alon U (2002) Negative Autoregulation Speeds the Response Times of Transcription Networks. *J Mol Biol* **323:** 785-793

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

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

Shi W, Ma W, Xiong L, Zhang M, Tang C (2017) Adaptation with transcriptional regulation. *Scientific Reports* **7:** 42648

Shoval O, Alon U (2010) SnapShot: Network Motifs. *Cell* **143:** 326-326.e321

Shoval O, Goentoro L, Hart Y, Mayo A, Sontag E, Alon U (2010) Fold-change detection and scalar symmetry of sensory input fields. *Proc Natl Acad Sci USA* **107:** 15995-16000

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

Sorrells TR, Booth LN, Tuch BB, Johnson AD (2015) Intersecting transcription networks constrain gene regulatory evolution. *Nature* **523:** 361-365

Sorrells TR, Johnson AD (2015) Making Sense of Transcription Networks. *Cell* **161:** 714-723

Stoltzfus A, McCandlish DM (2017) Mutational Biases Influence Parallel Adaptation. *Mol Biol Evol* **34:** 2163-2172

Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sa-Correia I (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucl Acids Res* **34:** D446-451

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

Wagner A (2003) Does Selection Mold Molecular Networks? *Sci STKE* **2003:** pe41

Wall ME, Hlavacek WS, Savageau MA (2004) Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5:** 34-42

Wang E, Purisima E (2005) Network motifs are enriched with transcription factors whose transcripts have short half-lives. *Trends Genet* **21:** 492-495

Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C (2015) Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15:** 3163-3168

Whitlock MC, Phillips PC, Moore FB-G, Tonsor SJ (1995) MULTIPLE FITNESS PEAKS AND EPISTASIS. *Annual Review of Ecology and Systematics* **26:** 601-629

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

Xiong K, Lancaster AK, Siegal ML, Masel J (2019) Feed-forward regulation adaptively evolves via dynamics rather than topology when there is intrinsic noise. *Nature Communications* **10:** 2418

Yampolsky LY, Stoltzfus A (2001) Bias in the introduction of variation as an orienting factor in evolution. *Evolution and Development* **3:** 73-83

Yu H, Luscombe NM, Qian J, Gerstein M (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet* **19:** 422-427