Supplementary materials

"Differences in evolutionary accessibility determine which equally effective regulatory motif evolves to generate pulses"

Xiong et al.

Table S1. Quantitative variables of the model

Variable	Initial values	Bounds[1]	References
Length of gene, L	$10^{N(2.568,\ 0.34)}$ codons	[50, 5000]	(SGD Project) (Balakrishnan et al. 2012)
Transition rate from active to intermediate promoter, $r_{Act\_to\_Int}$	$10^{N(1.27,\ 0.226)}$ min^{-1}[2]	[0.59, 64.7]	(Guillemette et al. 2005; Pelechano et al. 2010; Brown et al. 2013)
Dissociation constant between TF and perfect TFBS, $K_d(0)$	$10^{U(-9,-6)}$ mole per liter[3]	(0, 10^{-5})	(Park et al. 2004; Nalefski et al. 2006)
Rate of mRNA degradation, r_{mRNA_deg}	$10^{N(-1.49,\ 0.267)}$ min^{-1}	[7.5×10^{-4}, 0.54]	(Wang et al. 2002)
Protein synthesis rate, $r_{protein\_syn}$	$10^{N(0.322,\ 0.416)}$ molecule mRNA^{-1} min^{-1}	[4.5×10^{-3}, 61.4]	(Siwiak et al. 2010)
Rate of protein degradation, $r_{protein\_deg}$	$10^{N(-1.88,\ 0.561)}$ min^{-1}	[3.0×10^{-6}, 0.69]	(Belle et al. 2006)
Number of effector genes	1 activator (default) or 1 repressor	[1, 4][4]	
Number of non-effector genes	3 activators and 3 repressors	[2, 20][4]	

[1] The units are the same as in initial values. Parentheses mean the variable cannot take the boundary values; square brackets mean it can. We use these bounds to constrain mutation (see Supplementary Methods).

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

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

[4] In case of gene duplication, each gene can have at most 4 duplicates. See Model Overview for details.

Table S2. Parameters controlling transcriptional regulation and dynamics of gene expression

Parameter	Values	References
Length of cis-regulatory sequence	150 bp	(Yuan et al. 2005)
Length of TF recognition sequence	8 bp	(Wunderlich and Mirny 2009)
Length occupied by a TF on each side of recognition sequence	3 bp	(Zhu and Zhang 1999)
Dissociation constant between TF and non-specific DNA, $K_d(3)$	10^{-5} M	(Maerkl and Quake 2007)
Base rate of transition from Repressed to Intermediate	0.15 min^{-1}	(Katan-Khaykovich and Struhl 2002)
Maximum transition rate from Repressed to Intermediate	0.92 min^{-1}	(Katan-Khaykovich and Struhl 2002; Brown et al. 2013)
Base rate of transition from Intermediate to Repressed	0.67 min^{-1}	(Katan-Khaykovich and 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)
Rate of transcription initiation, $r_{max\_transc\_init}$	6.75 min^{-1}	(Brown et al. 2013)
Speed of transcription elongation	600 codon min^{-1}	(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)
Speed of translation elongation	330 codon min^{-1}	(Siwiak et al. 2010)
Translation initiation time	0.5 min	(Siwiak et al. 2010)

Table S3. Parameters controlling fitness

Parameter	Values	References
Optimal peak concentration of the effector protein, P_{opt}	5000, 10000, 20000 molecules cell^{-1}	(Ghaemmaghami et al. 2003)
Speed r of decrease in fitness component 1 when the peak effector level deviates from P_{opt}	0.693	
Basal expression level s_1 of the effector that maximize fitness	10% of peak level	
Minimum time $t_{saturate}$ to reach half peak that maximizes fitness	0, 60 (default) min	
Post-peak expression level s_2 of the effector that maximizes fitness	80% of peak level	
Fitness cost of protein expression, c_{transl}	2×10^{-6} molecules^{-1} min^{-1}	(Ghaemmaghami et al. 2003; Kafri et al. 2016)

Table S4. Parameters controlling the rates and effects of mutation to variables

	Mutation rate[1]	Mean μ the log10 value of variable in MST[2]	Std σ of variable value in MST[2]	Speed at which variable value regresses to μ[2]
Mutation to L	1.2×10^{-11} per codon	na	na	na
Mutation to $r_{Act\_to\_Int}$	9.5×10^{-12} per codon	1.57	0.773	0.5
Mutation to $K_d(0)$	3.5×10^{-9} per gene	-5	0.776	0.5
Mutation to r_{mRNA_deg}	9.5×10^{-12} per codon	-1.19	0.396	0.5
Mutation to $r_{protein\_syn}$	9.5×10^{-12} per codon	0.021	0.76	0.5
Mutation to $r_{protein\_deg}$	9.5×10^{-12} per codon	-1.88	0.739	0.5
Gene deletion	1.5×10^{-7} per gene	na	na	na
Gene duplication	1.5×10^{-7} per gene	na	na	na
Mutation to consensus sequence of a TF	3.5×10^{-9} per gene	na	na	na
Mutation to TF identity (activator vs. repressor)	3.5×10^{-9} per gene	na	na	na
Single nucleotide substitution in a cis-regulatory sequence	5.25×10^{-8} per gene	na	na	na

[1] Lynch et al. (2008) reports the rates of single nucleotide substitution, indel, gene duplication, and gene deletion in yeast. We use these to derive all rates of mutations in the model. See Supplementary Methods for details.

[2] MST: mutational stationary distribution. Mutation to some 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 Supplementary Methods for details. Variable values are also constrained by the bounds in Table S1.

Optimal peak		**Initial condition 25**		**Initial condition 31**		**Initial condition 79**	
		Evo. step 1 - 10,000	**Evo. step 10,001 – 20,000**	**Evo. step 1 - 5,000**	**Evo. step 5,001 – 10,000**	**Evo. step 1 - 10,000**	**Evo. step 10,001 – 20,000**
Low	**I1FFL-creating**	0.050	0.217	0.680	0.067	0.051	0.586
	NFBL-creating	0.107	0.007	0.006	0.003	0.106	0.013
Medium	**I1FFL-creating**	0.065	0.075	0.811	0.365	0.105	0.047
	NFBL-creating	0.111	0.104	0.016	0.025	0.112	0.171
High	**I1FFL-creating**	0.013	0.034	0.875	0.059	0.048	0.118
	NFBL-creating	0.729	0.413	0.007	0.201	0.259	0.174

Table S5. Mutational bias toward particular motifs can shift over the course of evolution. We focus our analysis on three random TRN initializations (conditions 25, 31, and 79) that evolved to high fitness in all three selection conditions. Under selection for high peak effector expression, all three simulations evolved NFBLs (i.e. the occurrence of NFBL > 0.5 and the occurrence of I1FFL < 0.5). Under selection for low or medium effector expression, all three evolved I1FFLs. As in Table 1, we show the number of mutations normalized by the total number of mutations trialed in resident TRNs that did not contain the motif in question. As an example of a change in mutational bias, initial condition 25 under selection for low peak effector expression initially creates NFBLs more often but later creates I1FFLs more often.

peak level		Evolutionary step 1 – 10,000		Evolutionary step 10,001 – 30,000	
		Trialed	Acceptance rate	Trialed	Acceptance rate
Low	I1FFL-destroying	0.100	0.072	0.065	0.051
	NFBL-destroying	0.141	0.139	0.196	0.128
Medium	I1FFL-destroying	0.129	0.068	0.087	0.069
	NFBL-destroying	0.157	0.108	0.125	0.108
High	I1FFL-destroying	0.117	0.074	0.114	0.057
	NFBL-destroying	0.086	0.120	0.063	0.124

Table S6. Summary of mutations that remove all I1FFLs and/or NFBLs. For each selection condition, we pooled qualified mutations from all high-fitness replicates shown in Fig. 2. A mutation is classed as destroying if it eliminates all instances of the given motif. The total number of qualified mutations were normalized by the total number of mutations trialed in resident TRNs that contained the motif of interest. The acceptance rate is the number of accepted mutations across all replicates divided by the number of trialed mutations across all replicates.

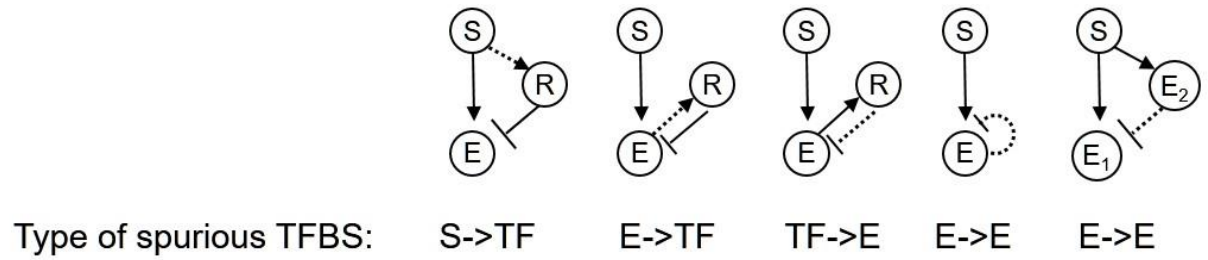

Figure S1. Five scenarios in which apparent but non-functional network motifs can arise from spurious TFBSs. A TFBS containing 2 mismatches can easily appear by chance in a cis-regulatory sequence, but may be deemed spurious if it has negligible functional effect. Spurious E->E TFBSs where both "Es" represent the same effector gene give rise to apparent ARs, whereas if they represent different effector proteins, they give rise to I1FFLs.

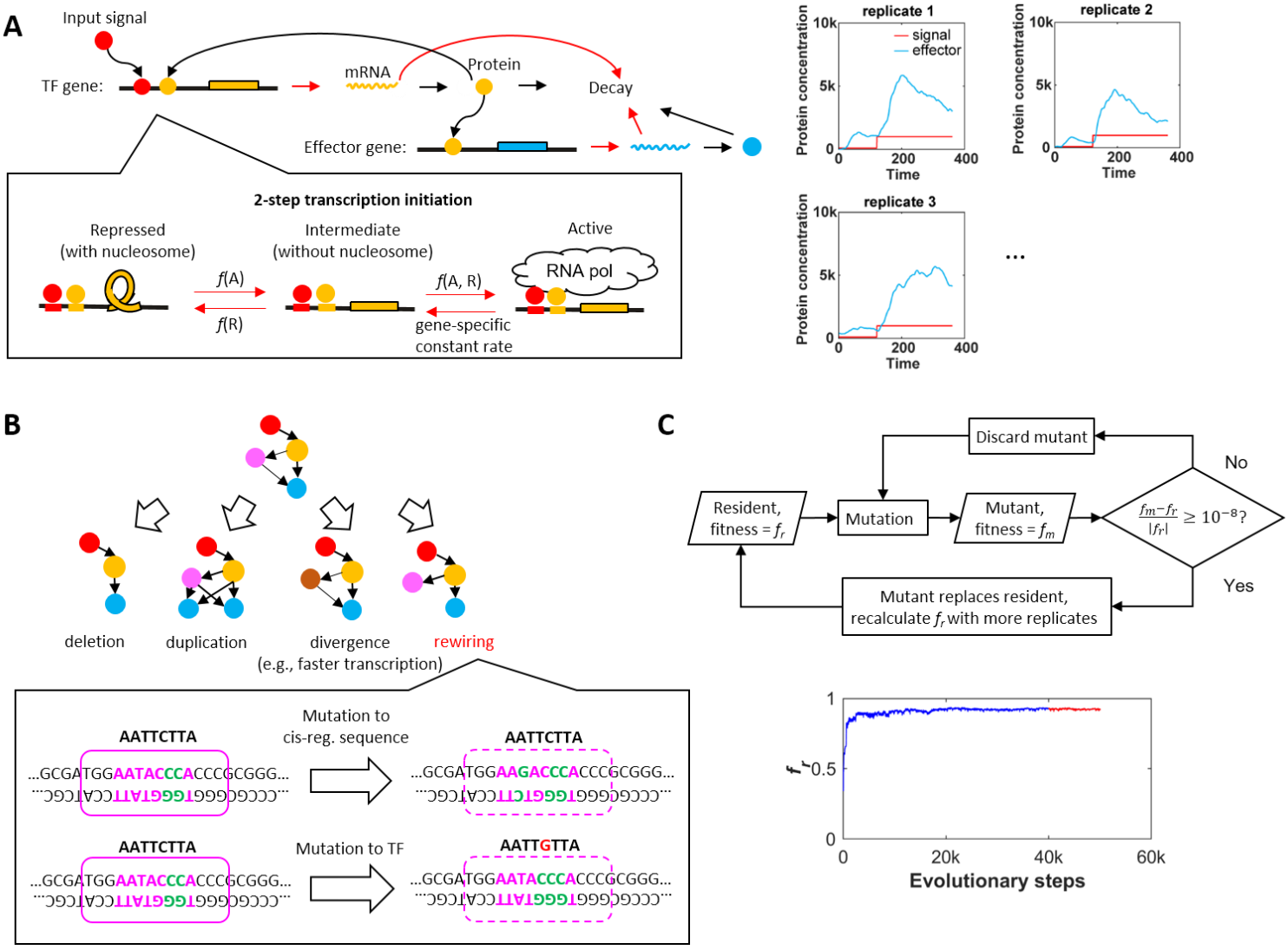

Figure S2. Summary of the model. (A) Simulation of gene expression in a TRN that has two TF genes, one of which is the effector (cyan). Here the input signal, which is simulated as an activator, binds to the cis-regulatory sequence of the non-effector TF gene (TF binding is demonstrated in **(B)**) and induces gene expression. Transcription initiation is a two-step process where most of the transition rates are functions of the concentrations of activators and/or repressors (see Transcriptional regulation in the supplement). Biological processes marked by red arrow are simulated as stochastic processes, and those marked by black arrows are simulated by solving ordinary differential equations (see Simulation of gene expression in the supplement). We use the expression levels of the effector in response to a two-stage input signal to calculate the fitness (see Methods for details). The simulation of gene expression is repeated and the average fitness of the replicates is used as the fitness of the TRN (see Methods for details). The diagram of transcription and translation is revised from Xiong et al. (2019) under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). **(B)** A TRN goes through one of many types of mutation (see Model Overview for details) that change the size of the network, rewire the network, or change one property of a gene in the network. The zoom-in depicts turnover of TF binding sites, which can rewire the network. The purple box represents the TF

and on top of the box is the consensus binding sequence of the TF. At most two mismatches (green letters) to the consensus binding sequences can be tolerated. Point mutations in the cis-regulatory sequence of the target gene and in the consensus binding sequence of the TF can increase mismatch, causing the loss of a TF binding site. Note that the TF occupies additional sequences when it binds to the DNA. **(C)** Evolution of TRNs is simulated as an origin-fixation process. Evolution starts with a random TRN, which is called the resident. if the mutant's fitness is sufficiently high (see Methods for details), it replaces the resident and becomes the new resident (see Methods for details), which is defined as one evolutionary step. Otherwise, new mutants are generated until the replacement happens. The evolution is simulated for 50,000 evolutionary steps, which is generally long enough for the resident's fitness to reach a plateau.

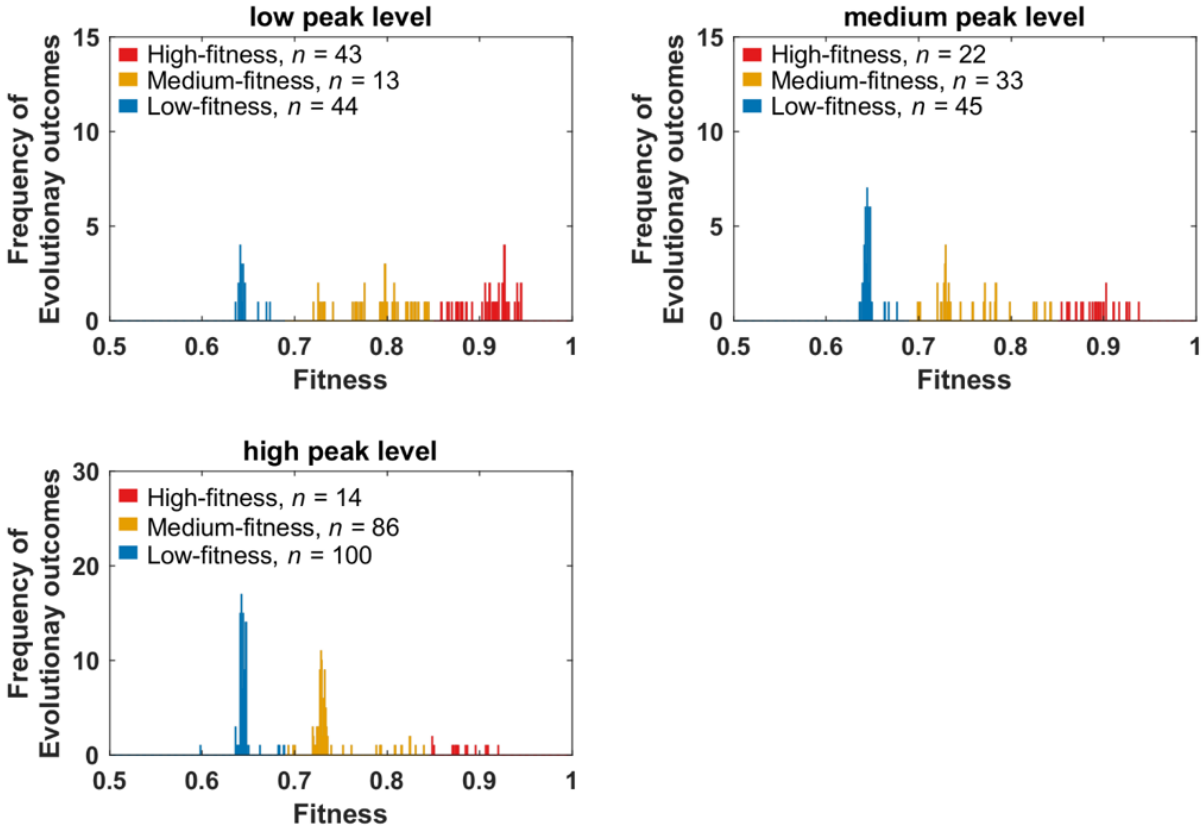

Figure S3. Fitness distributions of genotypes evolved with different optimal peak levels of the effector. We ran 100 evolutionary simulations for the low-peak and the medium-peak conditions, and 200 for the high-peak condition. For each simulation, we calculate the fitness of the evolved genotype as the average fitness of the last 10,000 evolutionary steps. For all three selection conditions, genotypes with fitness above 0.845 are considered as high-fitness genotypes and are further analyzed in **Fig. 2**. We used a fitness cutoff of 0.69 to separate medium-fitness genotypes and low-fitness genotypes.

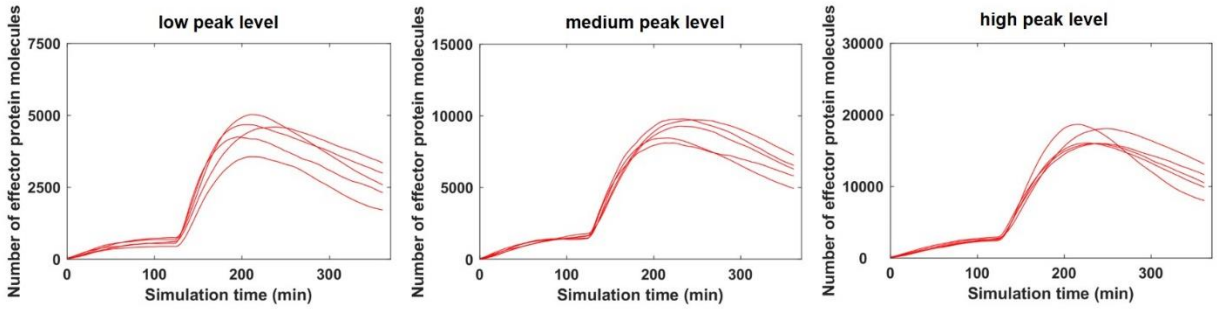

Figure S4. Phenotype of high-fitness replicates. For each selection condition, we randomly picked 5 high-fitness replicates from those defined in Fig. S4. We ran 200 simulations to characterize the expression profile of the effector, as found at evolutionary step 50,000 in each replicate. Each trajectory shows the expression levels of the effector averaged across the 200 simulations, and starts after the burn-in of gene expression (see **Methods** for details).

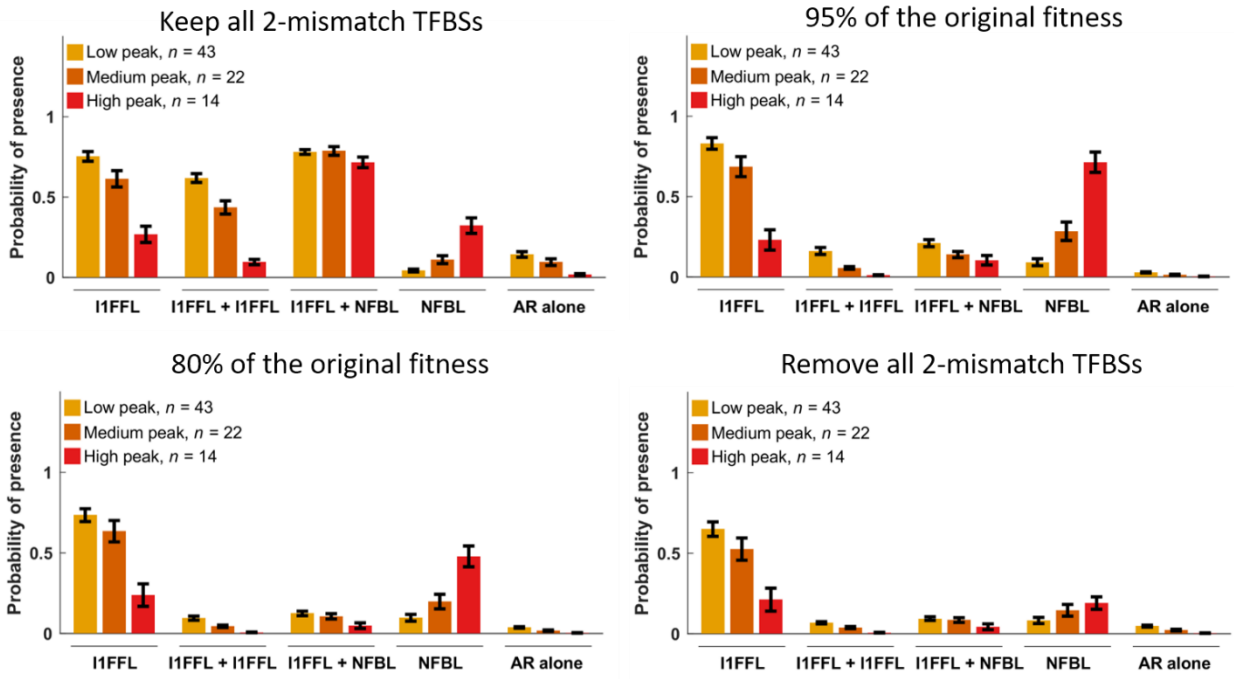

Figure S5. The relative occurrences of motifs do not depend strongly on the criteria for removing spurious 2-mismatch TFBSs. Results are from the same high-fitness evolutionary replicates shown in Fig. 2A, where sets of TFBSs were excluded when their removal yielded fitness of at least 99% of the fitness observed in their presence. Data are shown as mean ± SE over replicates.

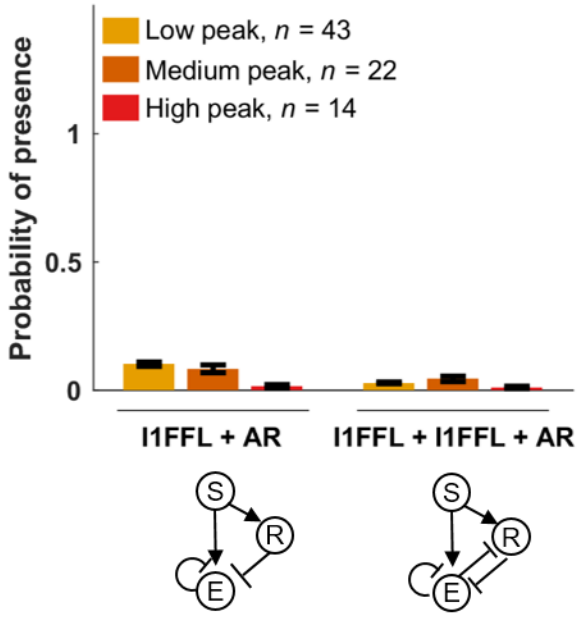

Figure S6. Auto-repression (AR) rarely evolves with I1FFLs or overlapping I1FFLs. In high-fitness genotypes evolved under selection for pulse generation, there were few auto-repressing effectors co-occurring with other motifs, and for simplicity, they were therefore grouped in **Fig. 2** with the motif 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.

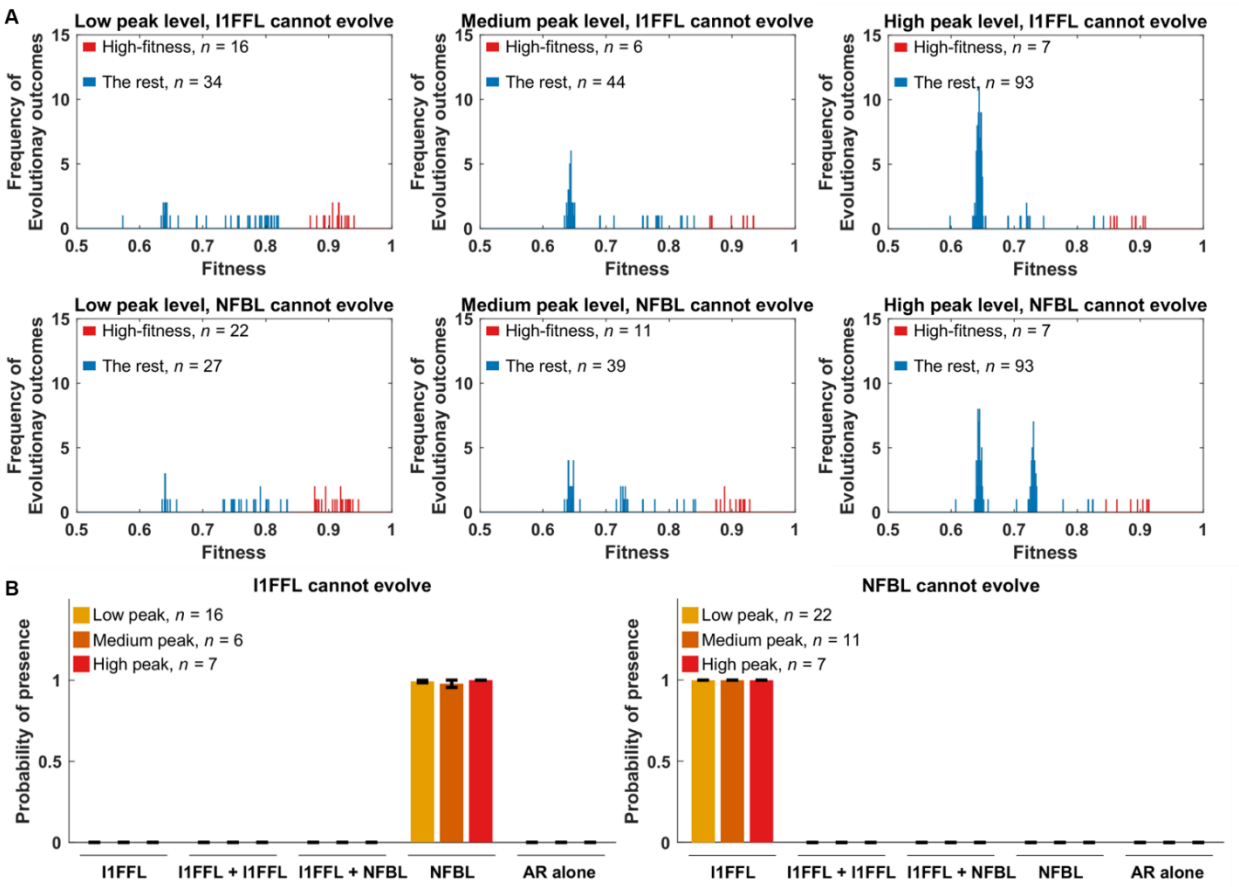

Figure S7. Fitness distributions and network occurrences of genotypes with restricted solutions under different selection conditions. (A) Each panel under low peak and medium peak selection summarized 50 evolutionary simulations, and the two panels under high peak each summarize 100 evolution simulations. Under the condition where we select for a low peak and prevent NFBLs from evolving, we removed one simulation that was terminated prematurely before evolving 50,000 evolutionary steps. This particular simulation failed to find a mutant that has higher fitness than the resident phenotype even after 2,000 trials. To classify a genotype as high-fitness (red), we apply the same fitness cutoff as in **Fig. S3**. 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.

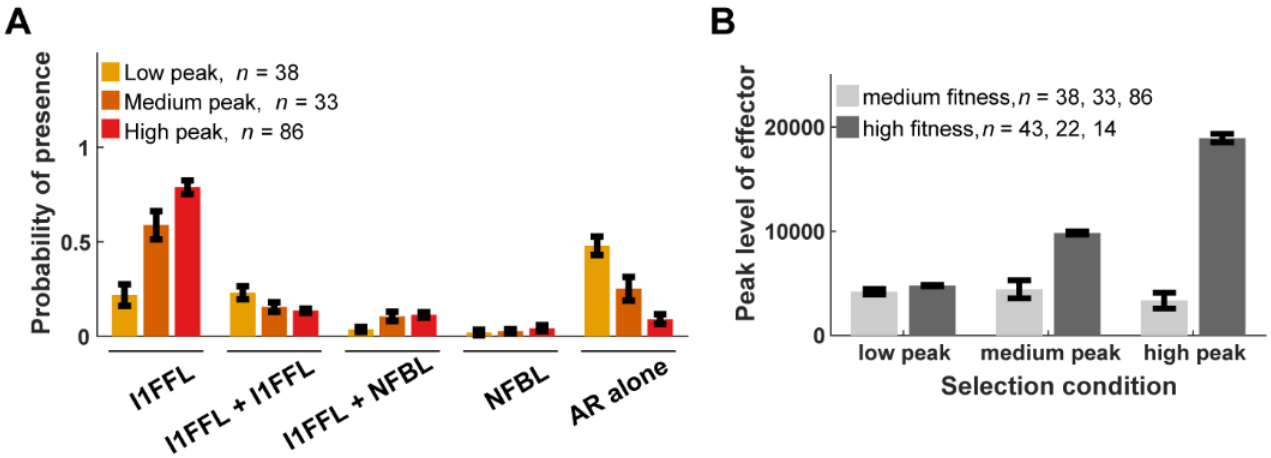

Figure S8. Medium-fitness genotypes fail to achieve high peak effector expression, and primarily evolve I1FFLs and auto-repression. (A) Methods are the same as for **Fig. 2A**, applied here to medium-fitness evolutionary replicates. **(B)** For each high-fitness and medium-fitness replicate shown in Fig. S4, we average the peak protein levels of the effector over 200 replicate simulations of gene expression. Data are shown as mean ± SE over replicates.

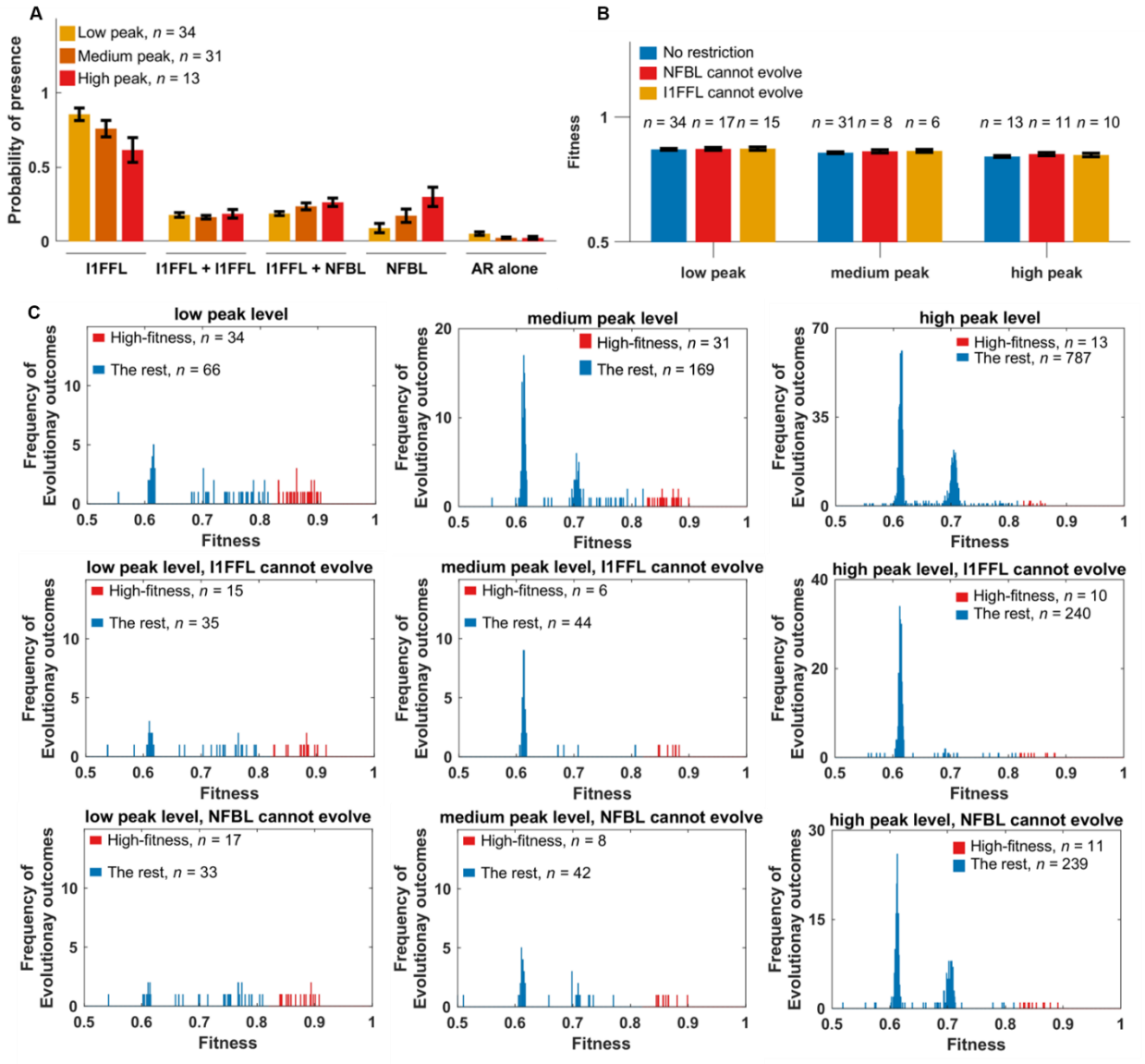

Figure S9. Selection for a fast response promotes I1FFLs, with optimal peak expression playing a secondary role. Similarly to **Fig. 2**, we tested evolution under selection for three peak levels, but we set $t_{saturate}$ to 0 to select for TRNs that generate a peak as soon as possible. **(A)** Occurrence of different motifs in high-fitness genotypes. We performed the Jonckeere-Terpstra test (R package DescTools, version 0.99.42, setting "alternative" to "increasing" (i.e. a single tailed test) and "nperm" to 10,000) to show that the occurrence of NFBLs increases with the peak level (p = 5e-4). **(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.

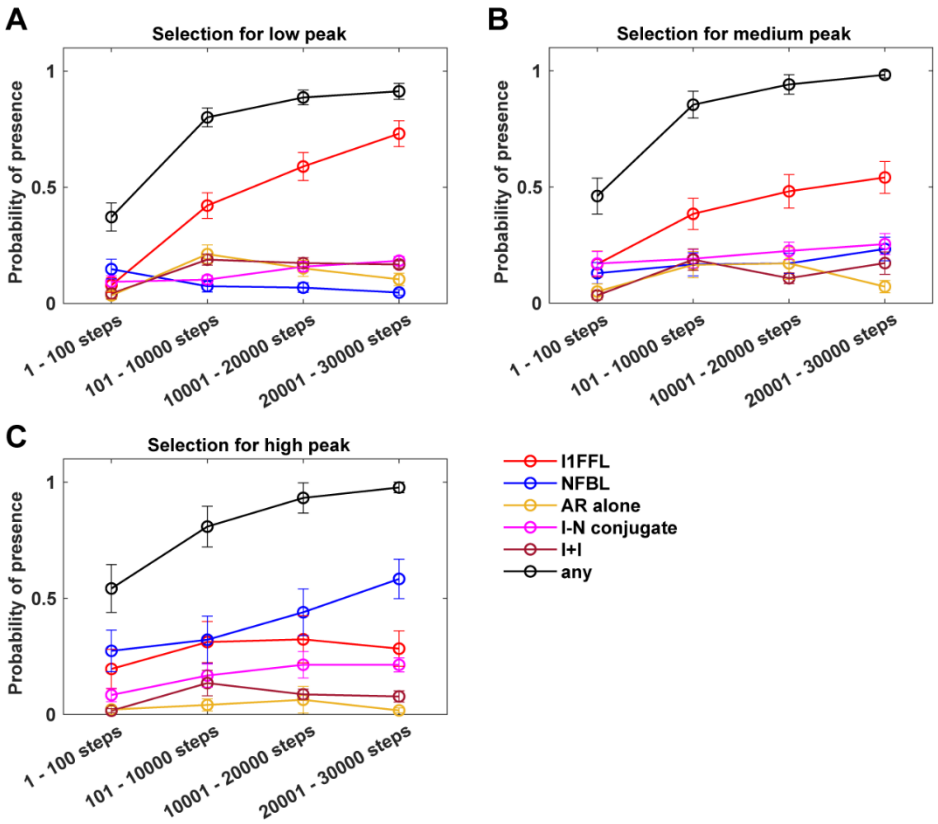

Figure S10. The occurrence of all motifs during evolution. We calculated the proportion of evolutionary steps that contain at least one motif of that type. Details are the same as for **Fig. 3**, except here we show a broader range of motifs as shown in **Fig. 2**. Data are shown as mean ± SE over the high-fitness replicates.

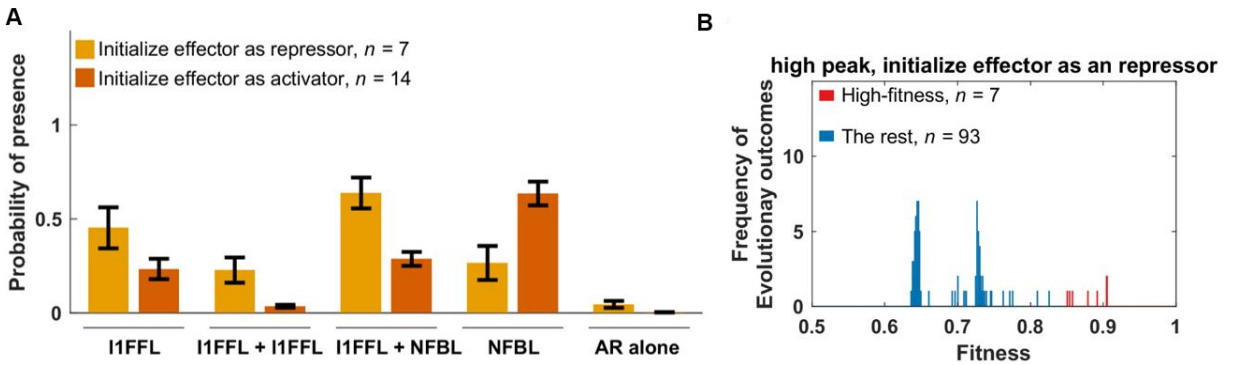

Figure S11. Initializing the effector as a repressor facilitates the evolution of I1FFLs. We repeated evolution under selection for high peak effector expression, but initialized the effector as a repressor rather than as an activator. **(A)** Motif occurrence compared to the activator-initialized evolutionary conditions given in **Fig. 2**. Data are shown as mean ± SE over replicates. **(B)** Fitness of the evolved TRNs. Similar to Fig. S1, TRNs with fitness of 0.845 or higher are considered high-fitness.

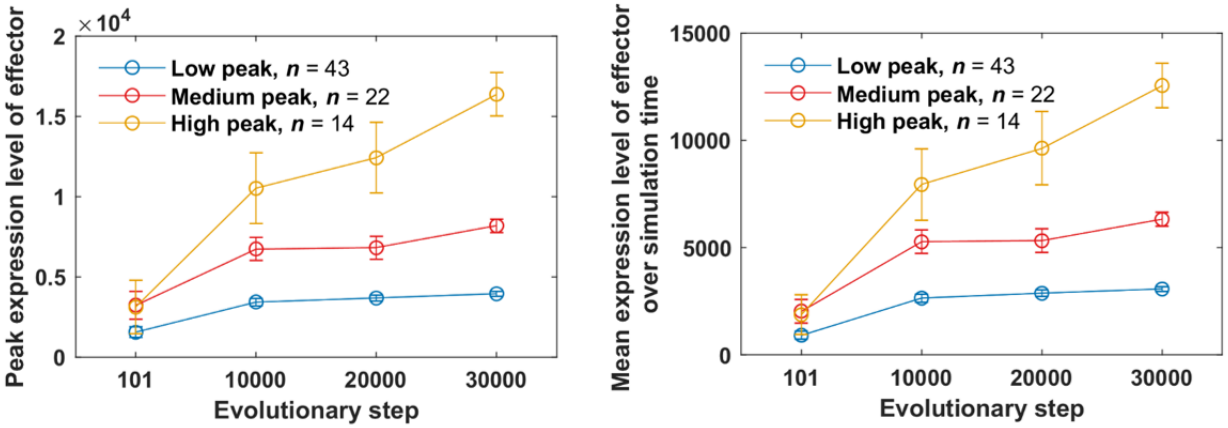

Figure S12. High peak effector expression evolves slowly. For each high-fitness replicate shown in Fig. 2A, we average the peak protein levels of the effector over 200 replicate simulations of gene expression. Data are shown as mean ± SE over replicates.

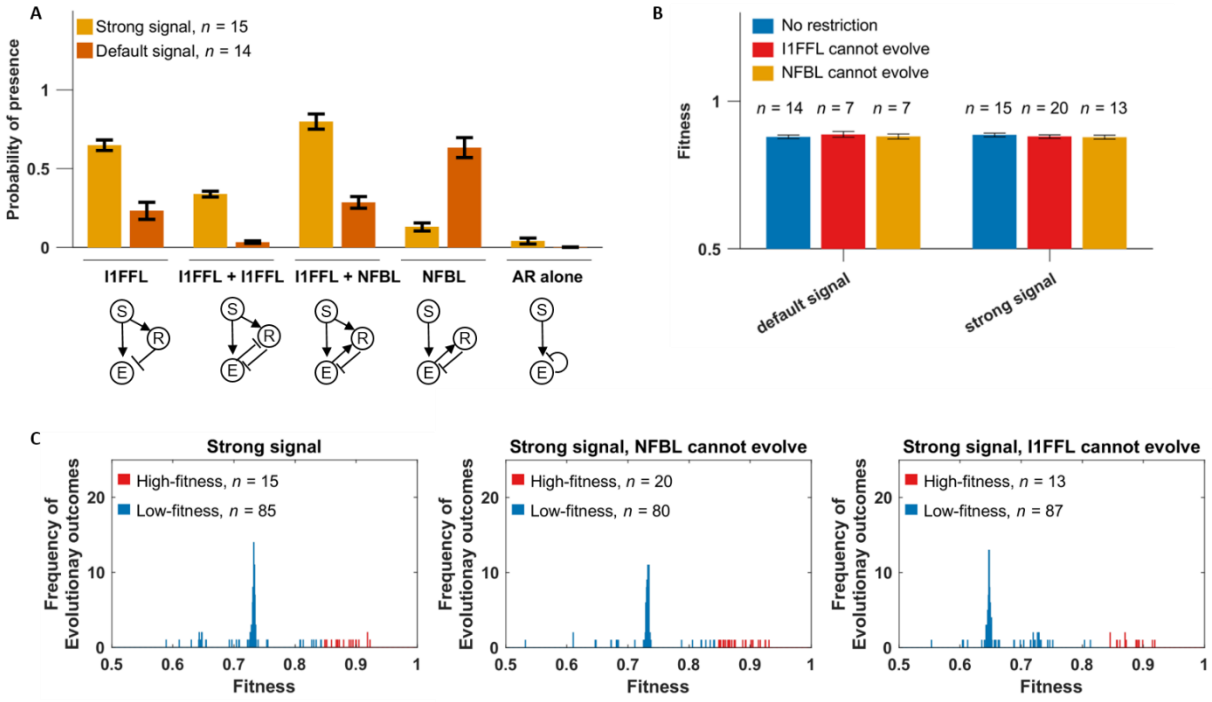

Figure S13. A stronger signal increases I1FFL prevalence. We compared evolution under the default signal, where signal strength increases from 100 molecules per cell to 1,000 molecules per cell, to a stronger signal, where the signal strength increases from 1,000 molecules per cell to 10,000 molecules per cell. **(A)** Occurrence of different motifs in high-fitness genotypes. **(B)** I1FFLs or NFBLs can yield similar fitness under a given signal regime. Data are shown as mean ± SE over replicates. **(C)** Fitness distribution of genotypes evolved with a strong signal without (leftmost) or with (middle and rightmost) restrictions on evolution. Similarly to Figs. S1 and S3, we define high-fitness genotypes to be those with fitness greater or equal to 0.845.

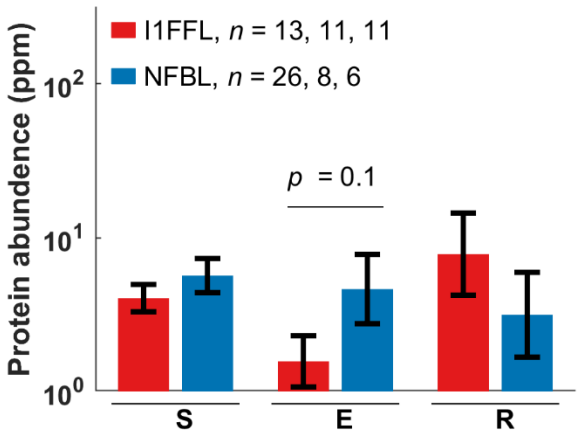

Figure S14. Effectors have higher protein expression in NFBLs than in I1FFLs in *S. cerevisiae* across a more comprehensive dataset. Protein expression levels are from the "GPM, Aug, 2014" dataset provided by PaxDB (Wang et al. 2015). Data are shown as mean ± SE (of log-transformed data in the case of protein expression) over each network position across all instances of the motif, excluding positions where the data are not available. For each motif type, we list the numbers of genes with available expression level data at signal nodes, effector nodes, and repressor nodes. Statistical significance is assessed using two-tailed t-tests.

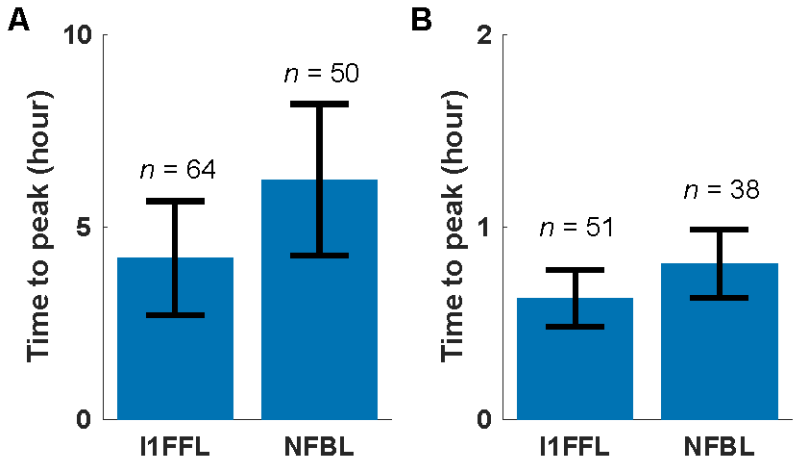

Figure S15. Pulse generation is not significantly faster in I1FFLs than in NFBLs. Similar to Fig. 4A, for each I1FFL and NFBL that shows pulse-like RNA expression in response to one of the 9 stimuli in Gasch et al. (2000), we calculated the time the motif spent to reach the peak level after experiencing the stimulus. We excluded data on diauxic shift, because dataset does not provide the exact time for each data point but only general order. Responses to nitrogen depletion and to stationary growth phase growth were measured for up to 5 days, much longer than for the other 7 stimuli. We therefore show the speed of pulse generation both when nitrogen depletion and stationary phase of growth are included **(A)** and excluded **(B)**. Data is shown as the mean ± SE over the motif-stimulus combinations. *n* is the number of motif-stimulus combinations.

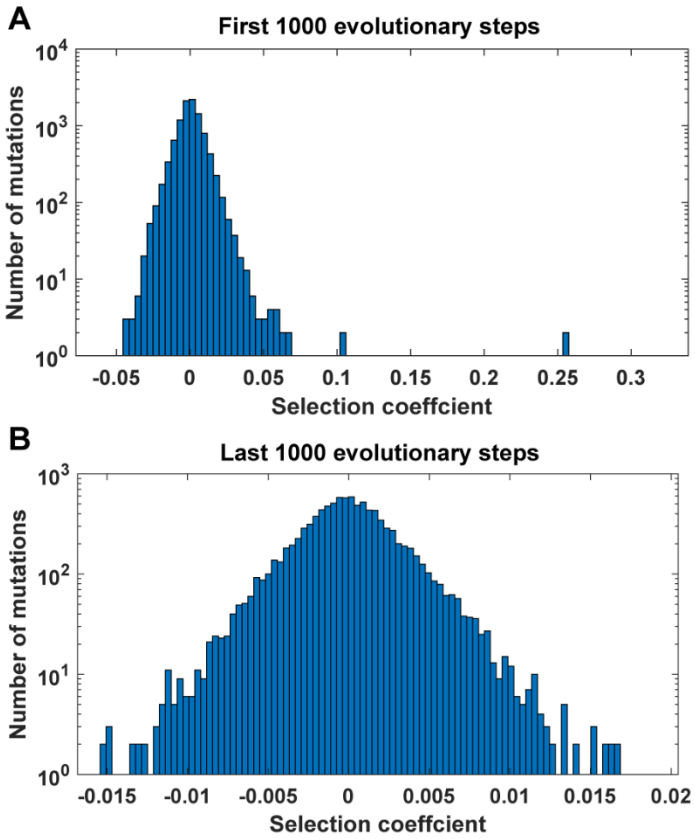

Figure S16. Evolutionary paths include slightly deleterious mutations. We pooled all accepted mutations from 10 evolutionary simulations under selection for high peak effector expression. Selection coefficients were calculated from the average fitness across 1,000 simulations of gene expression. Note while fitness is therefore biased by the 200 replicates used to decide to accept that mutation, this bias applies to both resident and mutant. We measure noise on top of the true distribution of fitness effects, suggesting that the underlying distribution is narrower than shown here. **(A)** Data restricted to the first 1,000 evolutionary steps, during which fitness generally increases rapidly. (**B**) Data restricted to the last 1,000 evolutionary steps, during which almost all simulations have reached a fitness plateau.

Searching for network motifs in Yeast TRNs

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. 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 203, allowing E and/or R to be self-repressing.

Details of the model

TF binding

In our model, each gene is controlled by a 150-bp cis-regulatory region, corresponding to a typical yeast nucleosome-free region within a promoter (Yuan et al. 2005). TFBSs can evolve in the cis-regulatory region, and we set the length of a consensus binding sequence to be 8 bp. Assuming that only one of the four nucleotides is a good match at each of the 8 base pairs, then the 8-bp consensus binding sequence has an information of 16 bits, which is slightly larger than that of a typical yeast TF (13.8 bits) (Wunderlich and Mirny 2009). We assume a higher information content than seen empirically in order to reduces the number of TFBSs within the cis-regulatory regions to a point that our computational power can handle. We allow up to 2 mismatches in the consensus binding sites, based on the finding that, with up to 2 mismatches in the 6-bp binding sequence, some yeast TFs can still bind DNA at above background level (Maerkl and Quake 2007). To capture competitive binding between TFs, we assume that two TFs cannot simultaneously occupy overlapping stretches, which we assume extend beyond the recognition sequence to occupy a total of 14 bp (Zhu and Zhang 1999).

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 per liter (Maerkl and 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). We ignore cooperativity in binding. Dissociation constants of eukaryotic TFs for perfect TFBSs can range from 10^{-5} mole per liter (Park et al. 2004) to 10^{-11} mole per 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 per liter. Substituting $m = 0$ and $m = 3$ into

$$\Delta G_m = -RT\ln K_{\mathrm{d}}(m) = \Delta G_0 - \min(m,3)\Delta G_{\mathrm{bp}}, \qquad (1)$$

where R is the gas constant and T is temperature, 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).

We rescale K_d values to effective K_d values to account for the "dilution" of TFs by non-specific TF binding sites (NSBSs) in the genome. A haploid *S. cerevisiae* genome is 12 Mb, 80% of which is wrapped in nucleosomes (Lee et al. 2007), yielding approximately 10^6 potential NSBSs. In a yeast nucleus of volume 3×10^{-15} liters, the NSBS concentration is of order 10^{-4} mole per liter. To find the concentration of free TF [TF] in the nucleus given a total nucleic TF concentration of C_{TF}, we consider

$$K_{\mathrm{d}} = \frac{[\mathrm{binding\_site}][\mathrm{TF}]}{[\mathrm{binding\_site}\cdot\mathrm{TF}]}, \quad (2)$$

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

$$[\mathrm{TF}] = \frac{K_{\mathrm{d}}(3)}{K_{\mathrm{d}}(3)+[\mathrm{NSBS}]} C_{\mathrm{TF}} = \frac{10^{-5}}{10^{-5}+10^{-4}} C_{\mathrm{TF}} \approx 0.1 C_{\mathrm{TF}}. \quad (3)$$

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 $\widehat{K}_{\mathrm{d}}(m) = 10K_{\mathrm{d}}(m)$ to account for the reduction in the amount of available TF due to non-specific binding. We also convert $\widehat{K}_{\mathrm{d}}$ from the units of mole per liter in which K_d is estimated empirically to the more convenient molecules per nucleus. The rescaling factor r for which $\widehat{K}_{\mathrm{d}}$ (in molecule per nucleus) = $r\widehat{K}_{\mathrm{d}}$ (in mole per liter) is 3×10^{-15} liter per nucleus $\times\ 6.02 \times 10^{23}$ molecule mole^{-1} = 1.8×10^9 molecule cell^{-1} liter mole^{-1}. Taken together, $\widehat{K}_{\mathrm{d}}$ (molecule per nucleus) = $10rK_d$ (mole per liter), where the factor 10 accounts for non-specific TF binding.

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 gene expression time. First we note that if we consider TF i binding to TFBS j in isolation from all other TFs and TFBSs, Supplementary Equation 4 gives us the probability of being bound:

$$P_{\mathrm{b}}(j) = 1 - P_{\mathrm{u}}(j) = \frac{C_i}{\widehat{K}_{\mathrm{d}}+C_i} \quad (4)$$

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 $(n+1)^{\mathrm{th}}$ binding site belongs to an activator. The case where this activator is not bound contributes $P_{A,R}^{(n)} P_{\mathrm{u}}(n+1)$ 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 $P_{A-1,R}^{(n-H)} P_{\mathrm{b}}(n+1) \prod_{j=n-H+1}^{n} P_{\mathrm{u}}(j)$. Taken together,

$$P_{A,R}^{(n+1)} = P_{A,R}^{(n)} P_{\mathrm{u}}(n+1) + P_{A-1,R}^{(n-H)} P_{\mathrm{b}}(n+1) \prod_{j=n-H+1}^{n} P_{\mathrm{u}}(j). \quad (5)$$

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

$$P_{A,R}^{(n+1)} = P_{A,R}^{(n)} P_{\mathrm{u}}(n+1) + P_{A,R-1}^{(n-H)} P_{\mathrm{b}}(n+1) \prod_{j=n-H+1}^{n} P_{\mathrm{u}}(j). \quad (6)$$

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, $P_{0,0}^{(1)} \propto P_{\mathrm{u}}(1)$ and if the binding site belongs to an activator $P_{1,0}^{(1)} \propto P_{\mathrm{b}}(1)$; otherwise, $P_{1,0}^{(1)} \propto P_{\mathrm{b}}(1)$. For a gene where the total number N of TFBSs is 1, $P_{0,0}^{(1)}$, $P_{1,0}^{(1)}$, and $P_{0,1}^{(1)}$ sum to 1 and normalization is unnecessary. For higher values of $N = N_{\mathrm{Act}} + N_{\mathrm{Rep}}$ TFBSs, where N_{Act} and N_{Rep} are the total numbers of activator binding sites and repressor binding sites, respectively, we normalize $P_{A,R}^{(N)}$ at the end of the recursion by dividing by $\sum_{A=0}^{N_{\mathrm{Act}}} \sum_{R=0}^{N_{\mathrm{Rep}}} P_{A,R}^{(N)}$ to get the probability of binding configurations that include exactly A activators and R repressors.

Transcriptional regulation

We model transcription initiation as a two-step transition between the repressed, intermediate, and active state. Only the active state leads to transcription. The transition rates are linear functions of four probabilities: 1) the probability P_{A} of having at least one activator bound to a gene, 2) the probability P_{R} of having at least one repressor bound, 3) the probability $P_{\mathrm{A\_no\_R}}$ of having no repressors and one activators bound, and 4) the probability $P_{\mathrm{notA\_no\_R}}$ of having no TFs bound.

We set the transition rate from the repressed state to the active state to

$$r_{\mathrm{Rep\_to\_Int}} = 0.92 P_{\mathrm{A}} + 0.15(1 - P_{\mathrm{A}}),$$

using as bounds for our linear function 0.15 min^{-1} as the background rate of histone acetylation (Katan-Khaykovich and Struhl 2002) (which leads to nucleosome disassembly) and 0.92 min^{-1} as the rate of nucleosome disassembly for the constitutively active PHO5 promoter (Brown et al. 2013).

We set the transition rate from the intermediate state to the active state to

$$r_{\mathrm{Int\_to\_Rep}} = 4.11 P_{\mathrm{R}} + 0.67(1 - P_{\mathrm{R}}),$$

where 0.67 min^{-1} is a background histone de-acetylation rate (Katan-Khaykovich and Struhl 2002) and 4.11 min^{-1} is chosen so as to keep a similar maximum:basal rate ratio as that of $r_{\mathrm{Rep\_to\_Int}}$.

We assume that the binding of a single repressor can prevent the transition from the intermediate state to the active state (Courey and Jia 2001). In the absence of repressors, activators facilitate the assembly of transcription machinery (Poss et al. 2013). Under these assumptions, we set the transition rate from the intermediate state to the active state to

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

where 3.3 min^{-1} is the rate of transcription machinery assembly for a constitutively active PHO5 promoter (Brown et al. 2013), and 0.025 min^{-1} is same rate when the PHO4 activator of the PHO5 promoter is knocked out.

We set the transition rate $r_{\text{Act\_to\_Int}}$ from the active state to the intermediate state to be gene-specific and independent of TF binding. This is because the promoter sequence not only determines which specific TFBSs are present, but also influences non-specific components of the transcriptional machinery (Decker and Hinton 2013). See next section for the parameterization of $r_{\text{Act\_to\_Int}}$. Note that we allow $r_{\text{Act\_to\_Int}}$ and K_d to evolve, making them variables of the model. We summarize them with other variables in Supplementary Table S1. The basal and maximum values of $r_{\text{Rep\_to\_Int}}$, $r_{\text{Int\_to\_Rep}}$, and $r_{\text{Int\_to\_Act}}$ are constant, and are summarized with other parameters that control transcriptional regulation and dynamics of gene expression in Supplementary Table S2.

$r_{\text{Act\_to\_Int}}$

Transcription initiation over an interval of time $r_{\text{transc\_init}}$ 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:

$$\frac{r_{\text{transc\_init}}}{r_{\text{max\_transc\_init}}} = \frac{r_{\text{Int\_to\_Act}}}{r_{\text{Int\_to\_Act}} + r_{\text{Act\_to\_Int}}} P_{\text{Int\_or\_Act}}, \quad (7)$$

where $P_{\text{Int\_or\_Act}}$ is the probability a gene is at Intermediate or Active. We set $r_{\text{max\_transc\_init}}$ (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. $P_{\text{Int\_or\_Act}} = 0.8$. We take these two values as universal for constitutively expressed genes, and assume that variation in $r_{\text{Act\_to\_Int}}$ is responsible for variation in $r_{\text{transc\_init}}$. 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 $r_{\text{transc\_init}}$ to the production rate of mRNA of these 225 genes, and solve for gene-specific $r_{\text{Act\_to\_Int}}$ from Eq. S7. We fit the solutions to a log-normal distribution and arrive at $10^{N(1.27,\ 0.226)}$ min^{-1}.

To initialize values of $r_{\text{Act\_to\_Int}}$ 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 $r_{\text{Act\_to\_Int}}$ beyond these bounds, we set the value of $r_{\text{Act\_to\_Int}}$ to equal to boundary value. We set the lower bound for $r_{\text{Act\_to\_Int}}$ 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 $r_{\text{transc\_init}} = 0.5$, which gives a solution of 32.34 min^{-1}; we double this to set the upper bound as 64.7 min^{-1}.

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; YeastMine (Balakrishnan et al. 2012) was used to query the database and to download data). 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 5,000

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 per 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.

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 per 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.

mRNA and protein decay rates

We fit a 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 (3×10^{-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.

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 as described later below.

The total rate of all Gillespie events is

$$r_{\mathrm{total}} = \sum_{i=1}^{\mathrm{Rep}} r_{\mathrm{Rep\_to\_Int\_i}} + \sum_{i=1}^{\mathrm{Int}} (r_{\mathrm{Int\_to\_Rep\_i}} + r_{\mathrm{Int\_to\_Act\_i}}) + \sum_{i=1}^{\mathrm{Act}} (r_{\mathrm{Act\_to\_Int\_i}} + r_{\mathrm{transc}}) + \sum_{i=1}^{N_{\mathrm{copies}}} r_{\mathrm{mRNA\_deg\_i}} N_{\mathrm{mRNA\_i}}, \quad (8)$$

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, N_{mRNA_i} is the number of completely transcribed mRNA molecules from gene *i*, and N_{copies} is the total number of gene copies. 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_G before the next Gillespie event is

$$\Delta t_{\mathrm{G}} = \frac{x}{r_{\mathrm{total}}}, \quad (9)$$

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 $\Delta t_F < \Delta t_G$, we advance time by Δt_F, 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_F, the remaining "labor" required to trigger the Gillespie event planned earlier is reduced. The new waiting time, $\Delta t_G'$, to trigger the planned Gillespie event is

$$\Delta t_{\mathrm{G}}' = \frac{x - r_{\mathrm{total}} \Delta t_{\mathrm{F}}}{r_{\mathrm{total}}'}. \quad (10)$$

Gene duplication creates $n \geq 1$ genes copies 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 $N_{\text{mRNA\_aft\_delay\_i}}(t)$. The total concentration of a protein obeys:

$$N'_{\text{protein}}(t) = \sum_i^n \left(r_{\text{protein\_syn\_i}} N_{\text{mRNA\_aft\_delay\_i}}(t) - r_{\text{protein\_deg\_i}} N_{\text{protein\_i}}(t) \right) \quad (11)$$

Protein concentrations are updated using a closed-form integral of Supplementary Equation 11

$$N_{\text{protein}}(t_1) = \sum_i^n \left(\frac{r_{\text{protein\_syn\_i}} N_{\text{mRNA\_aft\_delay\_i}}}{r_{\text{protein\_deg\_i}}} + \left(N_{\text{protein\_i}}(t_0) - \frac{r_{\text{protein\_syn\_i}} N_{\text{mRNA\_aft\_delay\_i}}}{r_{\text{protein\_deg\_i}}} \right) \mathrm{e}^{-r_{\text{protein\_deg\_i}}(t_1 - t_0)} \right) \quad (12)$$

with this expression updated every time a Gillespie or fixed event at time t_1 changes the value of $N_{\text{mRNA\_aft\_delay\_i}}$.

In between updates, values of P_A, P_R, $P_{A\_no\_R}$, and $P_{notA\_no\_R}$, and hence chromatin transition rates, are calculated under the approximation of constant N_{protein}. Additional updates, above and beyond fixed and Gillespie events, are performed in order to ensure that chromatin transition rates do not change too dramatically from one update to the next. We use a target of $D = 0.01$ 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 schedule updates after time Δt_U, which are triggered when neither a Gillespie event nor a fixed event occurs before this time has elapsed, i.e. when $\Delta t_U < \Delta t_F$ and $\Delta t_U < \Delta t_G$.

There is the greatest potential for large changes after an update that changes the value of $N_{\text{mRNA\_aft\_delay\_i}}$. In this case, we solve for the time interval for which the probability that TF *i* would be bound to a single perfect and non-overlapping TFBS would change by D, by choosing $\Delta t_U > 0$ that satisfies

$$\left| \frac{N_i(t)}{N_i(t) + \bar{K}_{\text{d\_i}}(0)} - \frac{N_i(t + \Delta t_U)}{N_i(t + \Delta t_U) + \bar{K}_{\text{d\_i}}(0)} \right| = D. \quad (13)$$

where the two left-hand terms are derived from Supplementary Equation 4. A solution for Δt_U may not exist, e.g. if the concentration of TF *i* is decreasing but $P_{b\_i}(t) < D$. In such cases, we set Δt_U to infinity.

When the previous update does not change any $N_{\text{mRNA\_aft\_delay\_i}}$ values, then we modify Δt_U 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, and Δt be the advance in time between the last two updates. We then schedule an update at

$$\Delta t_U{}' = \frac{D}{d} \Delta t. \quad (14)$$

After an update that changes the value of $N_{\text{mRNA\_aft\_delay\_i}}$, we use the smaller value from Supplementary Equations 13 and 14. These additional update times are discarded and recalculated when a Gillespie or fixed event occurs first. Supplementary Figure 12 of Xiong et al. (2019) shows that simulations rarely exceed the target of $D = 0.01$, and do so only modestly.

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. Setting the growth rate of the yeast when mCherry is not expressed, i.e. the fitness, to one, 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 $c_{\text{transl}}L$ / 370, where 370 is the geometric mean length of a yeast protein as described above in "Transcription delay times". Results using the length of mCherry instead, i.e. a slightly higher cost of expression of $c_{\text{transl}}L$ / 236, are unlikely to be significantly different.

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

$$C(t) = c_{\text{transl}}\left(\sum_{i=1}^{N_{\text{copies}}} \frac{L_i}{10^{2.568}} r_{\text{transl\_init\_i}} N_{\text{mRNA\_aft\_delay\_i}}(t) + \sum_{i=1}^{N_{\text{copies}}} \frac{L_i}{10^{2.568}} \frac{r_{\text{transl\_init\_i}}}{2} N_{\text{mRNA\_during\_delay\_i}}(t)\right). \quad (15)$$

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.

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, so we chose values 10-fold lower, making both gene duplication and gene deletion occur at rate 1.5×10^{-7} per gene. This preserves their numerical excess but reduces its magnitude.

Our model contains 8 gene-specific variables, namely L, $r_{\text{Act\_to\_Int}}$, $r_{\text{protein\_deg}}$, $r_{\text{protein\_syn}}$, $r_{\text{mRNA\_deg}}$, the $K_d(0)$ of a TF, whether a TF is an activator vs. repressor, and the consensus binding sequence of a TF. We assume mutations to L are caused by relatively neutral small indels, which we assume to be 20% of all small indels; mutation to L therefore occurs at rate 1.2×10^{-11} per codon, i.e. $1.2 \times 10^{-11}L$ for a gene of length L. For $r_{\text{Act\_to\_Int}}$, we assume that it is altered by 10% of all the point mutations (single nucleotide substitution and small indels) to the core promoter of a gene. The length of a core promoter is about 100 bp and is relatively constant among genes (Roy and Singer 2015), yielding a mutation rate of $r_{\text{Act\_to\_Int}}$ of 3.5×10^{-9} per gene.

Mutation rates for the remaining 6 gene-specific variables are parameterized with lower accuracy due to lack of data; the principal decision is which to make dependent vs. independent of gene length. TF binding to DNA depends on particular peptide motifs whose length is likely independent of TF length, therefore we make mutation rates independent of gene length for mutations to $K_d(0)$, to the consensus binding sequence of a TF, and to the activating vs repressing identity of a TF. We set the rate of each of the three mutation types to 3.5×10^{-9} per gene. In contrast, because the stability of an mRNA mainly depends on its codon usage (Cheng et al. 2017) and thus more codons means more opportunities for change, we assume the rate of mutation to $r_{\text{mRNA\_deg}}$ does depend on gene length, as do mutations to protein stability $r_{\text{protein\_deg}}$. $r_{\text{protein\_syn}}$ is determined by the density of ribosomes on mRNA and the elongation rate of ribosomes, and therefore is affected both by ribosome loading speed and by slow spots forming 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 $r_{\text{protein\_syn}}$ depends on gene length. We set the mutation rates of $r_{\text{protein\_deg}}$, $r_{\text{protein\_syn}}$, and $r_{\text{mRNA\_deg}}$ 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.

$r_{\text{Act\_to\_Int}}$, $r_{\text{protein\_syn}}$, $K_d(0)$, $r_{\text{protein\_deg}}$, and $r_{\text{mRNA\_deg}}$ 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 Supplementary Table 2. Denote the values of a variable as x before mutation and x' after mutation; mutation takes the form:

$$\log_{10}x' = \log_{10}x + \text{Normal}(k(\mu - \log_{10}x), \sigma), \quad (16)$$

where k controls the speed of regressing back to the stationary distribution; we set k = 0.5 for all 5 variables. To set values of μ, central tendency estimates of these five values (from Supplementary Table 1**)** are adjusted according to our expectations about mutation bias. We assume a mutation bias toward faster mRNA degradation $r_{\text{mRNA\_deg}}$, faster $r_{\text{Act\_to\_Int}}$ (Decker and Hinton 2013; Roy and Singer 2015), slower translation initiation $r_{\text{protein\_syn}}$ (Hinnebusch 2011), and larger $K_d(0)$. We assume that the observed log-normal means of $r_{\text{mRNA\_deg}}$, $r_{\text{protein\_syn}}$, and $r_{\text{Act\_to\_Int}}$ differ by 2-fold from the mean expected from mutational bias; for example, the mean of $\log_{10}(r_{\text{mRNA\_deg}})$ is -1.49, so the value of μ for $r_{\text{mRNA\_deg}}$ 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 variables. We assume that the observed central tendency estimate of protein stability does not depart from mutational equilibrium, therefore the value of μ for $r_{\text{protein\_deg}}$ is the mean of $\log_{10}(r_{\text{protein\_deg}})$ = -1.88.

The value of σ controls mutational effect size. We set the value of σ such that 1% of mutational changes from $x = 10^{\mu}$ 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 Supplementary Equation 17 for σ:

$$\begin{cases} P(\mu + \text{Normal}(0,\sigma) \geq \log_{10} B_{\text{U}}) = 0.01, \text{if the upper bound } B_{\text{U}} \text{ is closer} \\ P(\mu + \text{Normal}(0,\sigma) \leq \log_{10} B_{\text{L}}) = 0.01, \text{if the lower bound } B_{\text{L}} \text{ is closer} \end{cases} \quad (17)$$

For example, the upper and the lower bounds of $r_{\text{mRNA\_deg}}$ are 0.54 min^{-1} and 7.5×10^{-4} min^{-1}; on a log-scale, the upper bound is closer to $10^{\mu} = 10^{-1.19}$ min^{-1}. Plugging these values in Eq. S8 and solving for σ, we have $\sigma = 0.396$. We set the values of σ for $r_{\text{protein\_syn}}$, and $r_{\text{protein\_deg}}$ in the same way. However for $r_{\text{Act\_to\_Int}}$, 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 $K_{\text{d}}(0)$, 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 $K_{\text{d}}(0)$ by 100-fold or more.

Mutant values of L, $r_{\text{Act\_to\_Int}}$, $r_{\text{protein\_syn}}$, $r_{\text{protein\_deg}}$, and $r_{\text{mRNA\_deg}}$ are constrained by the same bounds that constrain the initial values of these variables (see previous sections). If a mutation increases the value of any of these 5 variables 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 $K_{\text{d}}(0)$, we resample if $x' \geq K_{\text{d}}(3)$, because otherwise the mutation effectively "deletes" the TF by reducing its affinity to non-specific levels.

Evolutionary simulation

Standard origin-fixation evolutionary simulations explicitly calculate a probability of fixation for each mutation and compare it to a pseudo-random number to decide whether fixation occurs. Our model achieves a similar exploration of nearly neutral evolutionary paths by using the intrinsic uncertainty in the stochastic estimation of fitness. Our approach wastes as few beneficial mutations as possible, minimizing computation, rather than discard most beneficial mutations through the use of a fixation probability that is only around twice the selection coefficient (Haldane 1927). For example, in our simulations, we accepted 0.5 million out of 1.9 million trialed mutations across 10 evolution replicates in the high-peak condition, of which only a minority can be presumed to have achieved true fitness increases (**Fig. S16**). Importantly, fixation probability in our algorithm still depends on the size of the true underlying fitness difference, which controls the probability that the estimated selection coefficient in Eq. 5 will be positive.

References

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G et al. . 2012. YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. Database:bar062.

Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK. 2006. Quantification of protein half-lives in the budding yeast proteome. PNAS 103(35):13004-13009.

Benos PV, Bulyk ML, Stormo GD. 2002. Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Research 30(20):4442-4451.

Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H. 2013. Linking Stochastic Fluctuations in Chromatin Structure and Gene Expression. PLoS Biology 11(8):e1001621.

Cheng J, Maier KC, Avsec Ž, Rus P, Gagneur J. 2017. Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. RNA 23(11):1648-1659.

Courey AJ, Jia S. 2001. Transcriptional repression: the long and the short of it. Genes & Development 15:2786-2796.
Decker KB, 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 in Genetics 12(7):263-270.
Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A et al. . 2003. Global analysis of protein expression in yeast. Nature 425(6959):737-741.
Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 81(25):2340-2361.
Guillemette B, Bataille AR, Gevry N, Adam M, Blanchette M et al. . 2005. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol 3(12):e384.
Haldane JBS. 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23(7):838-844.
Hinnebusch AG. 2011. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. Microbiology and Molecular Biology Reviews 75(3):434-467.
Hocine S, Raymond P, Zenklusen D, Chao JA, Singer RH. 2013. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nature Methods 10(2):119-121.
Kafri M, Metzl-Raz E, Jona G, Barkai N. 2016. The Cost of Protein Production. Cell Reports 14(1):22-31.
Katan-Khaykovich Y, 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 & development 16(6):743-52.
Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH. 2011. Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332(6028):475-478.
Lee W, Tillo D, Bray N, Morse RH, Davis RW et al. . 2007. A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics 39(10):1235-44.
Lynch M, Sung W, Morris K, Coffey N, Landry CR et al. . 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proceedings of the National Academy of Sciences of the United States of America 105:9272-9277.
Maerkl SJ, Quake SR. 2007. A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. Science 315(5809):233-237.
Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y et al. . 2010. Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nature Cell Biology 12(6):543-552.
Nalefski EA, Nebelitsky E, Lloyd JA, Gullans SR. 2006. Single-molecule detection of transcription factor binding to DNA in real time: Specificity, equilibrium, and kinetic parameters. Biochemistry 45(46):13794-13806.
Niño CA, Hérissant L, Babour A, Dargemont C. 2013. mRNA nuclear export in yeast. Chemical Reviews 113(11):8523-8545.
Park S, Chung S, Kim KM, Jung KC, Park C et al. . 2004. Determination of binding constant of transcription factor myc-max/max-max and E-box DNA: The effect of inhibitors on the binding. Biochimica et Biophysica Acta - General Subjects 1670(3):217-228.
Pelechano V, Chávez S, Pérez-Ortín JE. 2010. A Complete Set of Nascent Transcription Rates for Yeast Genes. PLoS ONE 5(11):e115560.
Poss ZC, Ebmeier CC, Taatjes DJ. 2013. The Mediator complex and transcription regulation. Critical Reviews in Biochemistry and Molecular Biology 48(6):575-608.
Roy AL, Singer DS. 2015. Core promoters in transcription: old problem, new insights. Trends in Biochemical Sciences 40(3):165-171.

Scott M, Klumpp S, Mateescu EM, Hwa T. 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Molecular Systems Biology 10(8):747.
SGD Project. Available from: https://yeastmine.yeastgenome.org
Siwiak M, Zielenkiewicz P, Jacobson A, Grebogi C, Kito K. 2010. A Comprehensive, Quantitative, and Genome-Wide Model of Translation. PLoS Computational Biology 6(7):e1000865.
Smith C, Lari A, Derrer CP, Ouwehand A, Rossouw A et al. . 2015. In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. Journal of Cell Biology 211(6):1121-1130.
Snapp EL. 2009. Fluorescent proteins: a cell biologist's user guide. Trends in Cell Biology 19(11):649-655.
Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR et al. . 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. Nucl. Acids Res. 34(suppl_1):D446-451.
Tuller T, Ruppin E, Kupiec M. 2009. Properties of untranslated regions of the *S. cerevisiae* genome. BMC Genomics 10:391.
Wagner A. 2005. Energy constraints on the evolution of gene expression. Molecular Biology and Evolution 22(6):1365-1374.
Wagner A. 2007. Energy costs constrain the evolution of gene expression. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B(3):322-324.
Waldron C, Jund R, Lacroute F. 1977. Evidence for a high proportion of inactive ribosomes in slow-growing yeast cells. Biochemical Journal 168(3):409-415.
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(18):3163-3168.
Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D et al. . 2002. Precision and functional specificity in mRNA decay. Proceedings of the National Academy of Sciences of the United States of America 99(9):5860-5865.
Wunderlich Z, Mirny LA. 2009. Different gene regulation strategies revealed by analysis of binding motifs. Trends in Genetics 25(10):434-440.
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(1):2418.
Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF et al. . 2005. Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309(5734):626-30.
Zhu J, Zhang MQ. 1999. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. Bioinformatics 15(7):607-611.