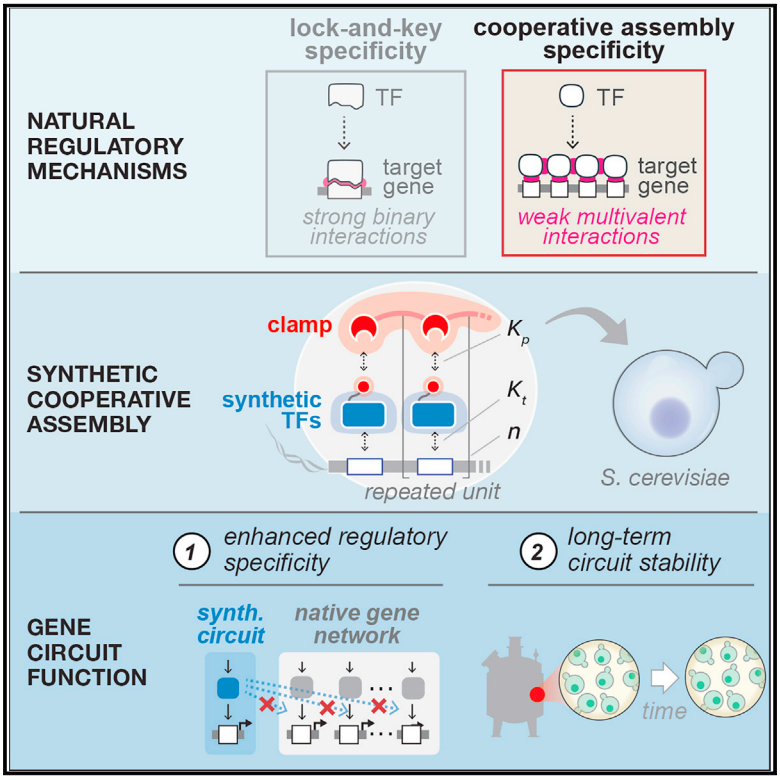
Cell

Article

Cooperative assembly confers regulatory specificity and long-term genetic circuit stability

Graphical abstract



Highlights

- Circuits that utilize artificial zinc-finger TFs impose a fitness burden in yeast
- Wiring circuits using cooperative assembly of weakly binding TFs restores fitness
- TF assembly enhances specificity and reduces off-target misregulation of transcription
- Circuits with cooperative TF assemblies show genetic stability in long-term culture

Authors

Meghan D.J. Bragdon, Nikit Patel, James Chuang, Ethan Levien, Caleb J. Bashor, Ahmad S. Khalil

Correspondence

khalil@bu.edu

In brief

Circuits composed of artificial zinc-finger transcriptional regulators establish cooperative assembly as a mechanism for achieving regulatory specificity and allow for construction of robust gene circuits scalable to a wide range of host organisms and applications.

Bragdon et al., 2023, Cell *186*, 3810–3825
August 31, 2023
https://doi.org/10.1016/j.cell.2023.07.012

CellPress

CellPress
OPEN ACCESS

Cell

Article

Cooperative assembly confers regulatory specificity and long-term genetic circuit stability

Meghan D.J. Bragdon,[1,2,9] Nikit Patel,[1,3,9] James Chuang,[3,4] Ethan Levien,[5] Caleb J. Bashor,[6,7,10] and Ahmad S. Khalil[1,2,3,8,10,11,*]

[1]Biological Design Center, Boston University, Boston, MA 02215, USA
[2]Program in Molecular Biology, Cell Biology and Biochemistry, Boston University, Boston, MA 02215, USA
[3]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA
[4]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[5]Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA
[6]Department of Bioengineering, Rice University, Houston, TX 77030, USA
[7]Department of Biosciences, Rice University, Houston, TX 77030, USA
[8]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA
[9]These authors contributed equally
[10]Senior author
[11]Lead contact
*Correspondence: khalil@bu.edu
https://doi.org/10.1016/j.cell.2023.07.012

SUMMARY

A ubiquitous feature of eukaryotic transcriptional regulation is cooperative self-assembly between transcription factors (TFs) and DNA *cis*-regulatory motifs. It is thought that this strategy enables specific regulatory connections to be formed in gene networks between otherwise weakly interacting, low-specificity molecular components. Here, using synthetic gene circuits constructed in yeast, we find that high regulatory specificity can emerge from cooperative, multivalent interactions among artificial zinc-finger-based TFs. We show that circuits "wired" using the strategy of cooperative TF assembly are effectively insulated from aberrant misregulation of the host cell genome. As we demonstrate in experiments and mathematical models, this mechanism is sufficient to rescue circuit-driven fitness defects, resulting in genetic and functional stability of circuits in long-term continuous culture. Our naturally inspired approach offers a simple, generalizable means for building high-fidelity, evolutionarily robust gene circuits that can be scaled to a wide range of host organisms and applications.

INTRODUCTION

In cells, gene regulatory networks integrate and process external and internal information into appropriate gene expression output responses.[1] Connections in these networks are mediated by the binding of transcription factors (TFs) to DNA *cis*-regulatory motifs (CRMs) located in upstream proximity to sites of transcriptional initiation. Proper cellular function critically depends on the genome-wide fidelity of these interactions: TFs must recognize gene-associated CRMs with high specificity while avoiding off-target interactions that can result in aberrant misregulation (Figure 1). Indeed, there is evidence that native regulatory network fidelity is optimized during evolution, likely through a combination of positive and negative selection processes that, respectively, maximize on-target regulation while minimizing off-target misregulation.[2–10] Disruption of network fidelity due to altered TF specificity or expression levels can lead to a loss in cellular fitness or, in the case of multicellular organisms, to abnormal development or oncogenesis.[11,12]

Extensive and ongoing investigation into the molecular basis of transcriptional regulation has revealed that strategies employed by cells to maintain network fidelity can vary dramatically across phylogeny.[13,14] For example, network connections in prokaryotes are maintained by families of TFs (e.g., helix-turn-helix and winged-helix members) that recognize CRMs via large-footprint, high-affinity interactions capable of specifying unique addresses within small-sized genomes (10^6–10^7 bp).[14–16] By contrast, despite possessing much larger genomes (10^7–10^9 bp), eukaryotic cells primarily regulate transcription using TFs (e.g., zinc-finger [ZF] and homeobox family TFs) that recognize and weakly bind to short, highly degenerate CRMs that occur at locations scattered throughout the genome.[17–23] How then can network fidelity be established with such low specificity TFs that are incapable of cognate CRM recognition

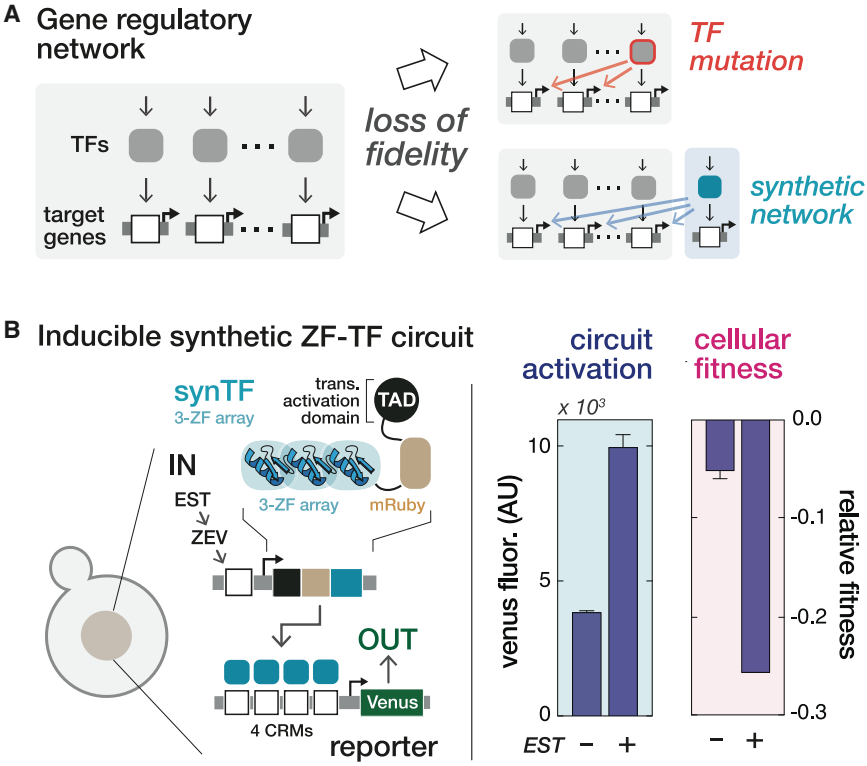


Figure 1. Gene networks rely on specific interactions between transcription factors and target genes for proper cellular function

(A) Cross-reactivity arising from transcription factor (TF) mutations or the introduction of synthetic circuits can drive loss of genome-wide interaction fidelity and disruption of cellular function and fitness.

(B) Activation of synthetic gene circuits, constructed from a common class of artificial zinc-finger (ZF)-based synthetic transcription factors (synTFs), results in observable fitness defects in yeast. The inducible circuit was chromosomally integrated into yeast, induced by addition of β-estradiol (EST), and circuit activation and cellular fitness were quantified by flow cytometry for Venus reporter fluorescence and pairwise growth competition, respectively, 36 h following induction. Bars represent mean values for three biological replicates ± SD.

Related to Figure S1.

within a complex genome? One explanation is that regulatory connections in eukaryotic networks are established via the cooperative assembly of TFs at closely spaced clusters of CRMs located within enhancer regions.[13,24–28] Under this scheme, CRM proximity strengthens TF binding through weak multivalent interactions between TFs and with associated transcriptional cofactors.[13,29–31] Thus, because functional regulatory connections are dependent on the cooperative assembly of multiple TFs, eukaryotes can maintain network fidelity despite the potential for genome-wide CRM occurrence.[32]

Over the last two decades, engineering artificial transcriptional regulatory networks to reprogram cellular behavior has become a major focus for the field of synthetic biology[33,34] and has emerged as a powerful approach for the development of cell-based biotechnologies.[35–38] These engineered networks, often termed gene circuits, are constructed using TF-CRM interactions that specify links between genes or couple gene expression outputs to molecular inputs such as small molecules, proteins, or RNA.[39,40] To date, the predominant focus in the field has been on identifying molecular parts (e.g., engineered TFs and promoters) and validating design strategies that enable the construction of gene circuits with quantitatively precise steady-state and dynamic behavior. Circuits engineered in both prokaryotic and eukaryotic host cells are currently under development for a wide range of applications, including metabolic engineering and cellular therapy. An emerging and critical feature of designing circuits for these applications is their genetic stability.[41] Introducing gene circuits into host cells can impose a fitness cost by creating a metabolic or resource burden or from expression of a toxic protein product.[42–44] Cells harboring mutations that abrogate circuit function can therefore acquire selective growth advantages over those with functionally intact circuits, leading to the progressive loss of circuit-bearing cells from a population. Although design strategies for addressing these issues have been described,[44–48] most rely on challenging *ad hoc* debugging, and generalizable rules for engineering circuit stability remain mostly undefined.

Disruption of transcriptional network fidelity represents another potential source of instability for gene circuits. To date, most circuit engineering efforts have focused on designing TF-CRM interactions to support robust regulatory connections. However, it is seldom investigated whether circuit expression leads to diminished fitness through spurious interactions between circuit TFs and non-cognate CRMs within the host cell genome. Indeed, loss of network fidelity resulting from circuit-associated TF expression may pose an acute challenge for gene circuits engineered in eukaryotic cells, which are often constructed using low-information TFs with potential for off-target misregulation. We recently developed a gene circuit engineering platform in yeast that uses synthetic ZF-derived transcriptional activators to mediate circuit connectivity.[49] As our previous work demonstrates, this framework can be readily utilized to construct diverse synthetic network connectivity, enabling precise control over circuit dose response as well as more complex signal processing behavior.[50] In this study, we investigate the genetic stability of circuits engineered using this framework. We show that an observable fitness cost associated with circuit activity is caused by off-target misregulation of host cell transcription, leading to the gradual loss of circuit function across a cell population. In order to restore network fidelity, we draw upon the organization of natural networks as inspiration and test whether cooperative TF assembly can be used as an engineering strategy to create insulated regulatory connections that limit off-target TF binding. As our results show, circuit

connections that are functionally dependent on multivalent assemblies can be used to effectively mitigate misregulation and restore fitness, resulting in the long-term stabilization of circuit function.

RESULTS

Our recently reported synthetic gene circuit engineering platform recapitulates many of the essential design features of native transcriptional regulation in yeast and other eukaryotes. The platform comprises a set of synthetic TFs (synTFs) constructed from Cys2-His2 ZFs, the most prevalent and conserved DNA-binding domain across eukaryotes.[20,51] The creation of tunable network linkages using synTFs is facilitated by their modular design, and they are composed of 3-finger (3F) ZF domain arrays engineered to bind artificial ~9 bp CRMs that are arranged in clusters upstream of a core promoter. Appending transcriptional activation domains or protein-protein interaction domains to either terminus of the ZF array enables synTFs to, respectively, activate transcription at the core promoter and interact with synTFs bound to adjacent CRMs (Figure S1A). The strength of synTF-mediated circuit linkages can be tuned by adjusting molecular parameters such as the number of CRMs and the strength of ZF binding. Furthermore, we created a collection of 20 distinct ZF species with complementary CRM specificities that facilitate robust construction of circuit designs featuring multiple synTFs. Our work and that of others have demonstrated the utility of programming synTF circuits for a variety of circuit functions in host cells that span eukaryotic phylogeny, including in therapeutically relevant human cells.[49,50,52–60]

Since the DNA-binding sites that our synTFs interact with are of a similarly low-information content as those of native eukaryotic ZF-TFs,[14,61] there is a possibility for off-target interactions between synTFs and genomic CRM sites, potentially leading to perturbation of host cell transcriptional network fidelity. Although the diminished circuit performance or host fitness that accompanies a loss of fidelity may go unobserved during short-timescale experiments, it is possible that such phenotypic defects may manifest during longer-timescale experiments that involve cell growth over many generations. This possibility motivated us to test whether there are measurable fitness costs associated with expression of synTF circuitry in yeast. We constructed a prototype inducible circuit consisting of a single transcriptional network linkage in which expression of a synTF containing a ZF from our collection (42-10) is under the control of an estradiol (EST)-inducible system to activate expression of a Venus reporter gene (Figures 1B, S1A, and S1B; STAR Methods). Following induction with EST, we observed an expected level of reporter activation. However, we also observed a concomitant loss of cellular fitness as measured over 36 h in pairwise growth competition with a reference control strain (Figures 1B and S1C). Control experiments confirmed that synTF expression was the source of both circuit activation and the fitness penalty, and expressing the combination of ZF with transactivation domain (TAD), but not either domain independently, led to a fitness decrease (Figure S1D "high affinity"). To investigate whether this result was specific to ZF 42-10-derived synTFs, we constructed circuits featuring synTFs containing ZFs from our entire collection (Figure S1E). For these 20 synTFs, which contain an average of 97.8 potential genomic binding sites (Figure S1F), we observed a consistent pattern of circuit activation and fitness loss, highlighting the generality of this observation.

Cooperative TF assemblies reduce fitness cost burden while maintaining circuit output

Because of the dependence of the observed growth phenotype on a synTF-TAD fusion, we reasoned that circuit-mediated fitness defects could potentially be the result of altered native gene expression caused by synTFs binding to off-target CRMs throughout the host cell genome. Indeed, analysis of the yeast genome revealed the occurrence of 78 sites that were sequence matches for the core 42-10 CRM and another 1,839 sites containing single base mismatches (Figure S1F). Since this abundance of potential off-target sites would make removal via genome editing time consuming and laborious, we considered less complicated engineering strategies that could mitigate fitness cost while maintaining circuit function. The predominant approach for programming network connections in synthetic circuits is based on TFs that have "one-to-one" specificity, a design strategy that mirrors native prokaryotic gene regulation by relying on high-information content, binary TF-CRM recognition to encode regulatory links with genome-wide specificity. On the other hand, regulatory strategies involving cooperative assembly that are common in eukaryotic cells rely on TFs that are individually weakly binding and low information to establish robust, highly specific connections through multivalent association. Since these TFs have molecular characteristics similar to our synTFs, we hypothesized that circuits incorporating regulation by cooperative assembly could potentially be used to engineer synthetic circuits with enhanced fidelity and diminished fitness defects.

To gain insight into molecular strategies for using cooperative synTF assemblies to construct highly specific circuit connections, we constructed a simple thermodynamic-based model of transcription regulation that extends our previous work[50] (Figure 2A; STAR Methods). This class of model can offer a simplified first-principles framework for predicting gene expression patterns based on key biophysical properties (e.g., protein-DNA interactions and protein-protein interactions) and can be useful as a guide for understanding synthetic systems in which such properties are design-specified.[62–65] Our model considers the simplified case of a TF that can interact with a CRM at both target synthetic (SYN) and "off-target" native (NAT) loci. As an example, we consider a SYN locus with four tandem CRMs and a NAT locus with a single CRM, where all sites are assumed to be identical (Figures S2A and S2B). CRM binding is governed by TF concentration ([TF]) and its affinity for the CRM (K_{TF}), and the energy of the cooperative interactions between TFs bound to adjacent sites (c).[63] We defined a regulatory specificity score as the difference between transcriptional output at the SYN (txn_{SYN}) and NAT (txn_{NAT}) loci (Figure S2C) and then plotted this score as a function of [TF], K_{TF}, and c (Figure 2A right). This analysis revealed that regulatory specificity improves along an axis defined by lowering affinity for DNA and increasing TF cooperativity, a relationship that remained qualitatively similar for cases containing different numbers of binding sites in both the

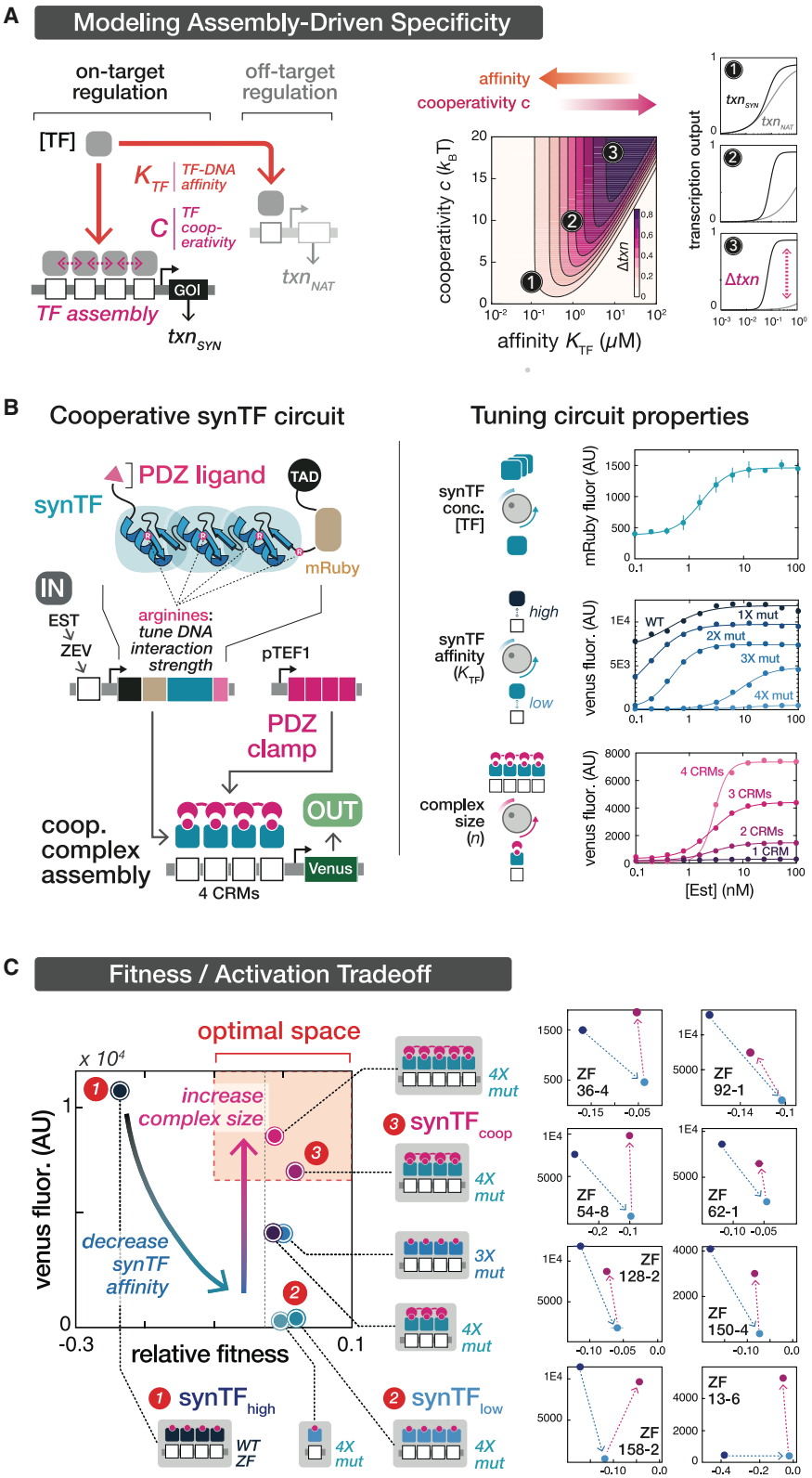


Figure 2. Constructing gene circuits using cooperative assembly minimizes fitness defects while maintaining robust circuit output

(A) A simple model reveals a molecular strategy for using cooperative assemblies to engineer regulatory specificity. Left: thermodynamic-based model of transcription regulation in which a transcription factor (TF) can interact with DNA at both an on-target synthetic circuit (SYN) locus and an off-target native (NAT) locus. Binding is governed by the TF affinity for DNA (K_{TF}) and the level of cooperativity between TFs bound to adjacent sites (*c*). Right: regulatory specificity, defined as transcriptional output from the SYN vs. NAT locus, improves as the TF affinity for DNA is lowered and TF cooperativity is increased.

(B) Experimental platform for constructing circuits composed of cooperative synthetic transcription factor (synTF) assemblies. Left: Inducible circuit architecture. synTF species, under the control of an estradiol (EST)-inducible system, form multivalent assemblies at a reporter locus to drive gene expression. Complex formation is mediated by a clamp protein. The synTF complex architecture and cooperativity are governed by programmable interaction domains (ZF and PDZ) and their respective binding partners (CRM and PDZ ligand). Right: Tunable circuit properties. synTF concentration, DNA affinity, and complex size are tuned by adjusting the EST dose, number of arginine-to-alanine (R → A) mutations at conserved positions in the ZF, and number of CRM sites, respectively. Points represent mean values for three biological replicates ± SD. WT, no R → A mutations; 1X mut, 1 R → A mutation; 2X mut, 2 R → A mutations, etc.

(C) Circuits utilizing cooperative regulatory assemblies of low-affinity synTFs minimize fitness costs while maintaining high circuit output, thus optimizing the fitness-activation tradeoff. Left: cellular fitness vs. circuit activation for various non-clamp and clamp circuit configurations, all constructed from the same ZF (42-10). $synTF_{high}$ uses wild-type ZF 42-10 ($K_d \sim 2$ nM); $synTF_{low}$ and $synTF_{coop}$ use 4X mut ZF 42-10 ($K_d \sim 15$ nM); clamps use syntrophin PDZ domains, which each interact with a VKESLV ligand ($K_d \sim 1.9$ μM). Right: fitness-activation measurements of circuit configurations constructed from different ZF species (with different binding sequences) exhibit similar patterns. Axes labels are identical to the fitness vs. activation plot on the left. Points represent mean values for three biological replicates ± SD.

Related to Figures S1, S2, and S3.

SYN and NAT loci, as well as different formulations of the regulatory specificity score (Figure S2D). This result suggests a design strategy whereby high-specificity circuit connections can be obtained by engineering synTFs that enforce complex formation through strong interaction with each other but interact weakly with DNA.

We sought to test this strategy experimentally using a framework we recently developed for engineering multivalent synTF assemblies.[50] Under this system, interactions between synTFs bound to tandem, core promoter-adjacent CRMs are mediated by a "clamp": a synthetic protein composed of multiple covalently linked PDZ domains that interact with peptide ligands on synTFs to enable multivalent coordination of their binding to DNA (Figures 2B, left and S3A; STAR Methods). Modifying the circuit in Figure 1B with a constitutively expressed clamp enables us to test relationships between molecular parameters underlying complex assembly and regulatory specificity that were suggested by our model (Figure 2B): synTF expression level ([TF]) can be tuned through the addition of EST, whereas K_{TF} for the synTF can be adjusted by introducing alanine mutations (WT-4X mut) to a set of conserved arginine residues in the ZF array that make non-specific interactions with the DNA phosphate backbone[49,66,67] (Figure 2B, right). Additionally, c can be tuned by varying complex valency (n), resulting in altered dose response steepness (Figures 2B, right and S3B).

We constructed various clamp and non-clamp circuit configurations, tested them for activation and competitive growth rate (Figure S1D), and then plotted their mean fluorescence and relative fitness on a two-dimensional "fitness-activation" phenotypic space (Figure 2C). The circuit configuration tested in Figure 1B—a circuit containing a high-affinity synTF (wild-type ZF 42-10, $K_d \sim 2$ nM[50]), termed the synTF$_{high}$ circuit, exhibited high reporter activation but low cellular fitness, placing it in the top-left region of the space (Figure 2C). Reducing K_{TF} by introducing 3 or 4 R → A ZF mutations was sufficient to rescue the fitness, however, this predictably led to significant loss of circuit activation (e.g., see synTF$_{low}$: 4X mut ZF 42-10, $K_d \sim 15$ nM[50]). Consistent with predictions from our model, we found that circuit activation could be restored via expression of clamp (syntrophin PDZ domains: PDZ <> VKESLV ligand $K_d \sim 1.9$ μM) with low-affinity synTFs and n = 4 (synTF$_{coop}$ circuit) or 5 CRM sites, with little apparent loss of fitness (Figure 2C). Furthermore, we found these effects on fitness and circuit output are not due to differential synTF expression, as low-affinity variants are expressed at equivalent, or even modestly increased, levels compared with synTF$_{high}$ (Figure S3C). To determine the generalizability of this result, we tested synTF$_{high}$, synTF$_{low}$, and synTF$_{coop}$ circuit variants for our entire ZF collection and observed the same pattern with all ZFs: a rescue of fitness from synTF$_{high}$ to synTF$_{low}$ and a subsequent improvement of circuit function in synTF$_{coop}$ (Figures 2C, right and S1E). Our data demonstrate that wiring synTF circuits using cooperative assemblies offers a simple and extensible strategy for optimizing both circuit function and host fitness.

Cooperative assembly is sufficient to rescue aberrant gene expression caused by synthetic circuits

To verify that differences in synTF circuit-imposed fitness costs are indeed the result of host cell network misregulation, we performed RNA sequencing (RNA-seq) to assess host cell transcriptomics following induction of synTF$_{high}$, synTF$_{low}$, and synTF$_{coop}$ circuits (all constructed from ZF 42-10). Biological replicates of each strain demonstrated highly correlated gene expression profiles (Figure S4A). We found that synTF$_{high}$ expression led to widespread misregulation of the host transcriptome relative to a reporter control strain (same genetic background with the integrated reporter cassette and neutral spacers integrated into the synTF and clamp loci) (Figure 3A). Consistent with a general model of TAD-dependent off-target gene activation by synTFs, the majority of misregulated genes were upregulated (182/211) (Figure 3B), with such genes more likely to harbor potential synTF binding sites (8/9 bp homology to the CRM) within a 300 bp window upstream of the transcription start site (TSS) (14.8% or 27/182 genes) compared with both downregulated (0/29 genes) or unaffected genes (2.8% or 134/4,827) (Figure S4E).

By contrast, transcriptomes of cells harboring the synTF$_{low}$ and synTF$_{coop}$ circuits demonstrated expression profiles that were similar to one another and to the strain background, indicating minimal effect on native transcription (Figures 3A, 3B, and S4B–S4D). As expected, we found that addition of the clamp shows no effect on endogenous gene misregulation compared with the synTF$_{low}$ case (Figures 3B and S4C), mirroring the observation that synTF$_{low}$ and synTF$_{coop}$ have similar fitness profiles. In fact, the only gene showing differential regulation between synTF$_{coop}$ and synTF$_{low}$ strains was the fluorescent reporter, with the synTF$_{coop}$ circuit showing comparable expression levels to synTF$_{high}$. Altogether, these results implicate transcriptional network misregulation as the basis of the observed growth defect in the synTF$_{high}$ circuit and demonstrate that this defect can be rescued by tuned-down synTF-CRM interaction affinity in the synTF$_{low}$ and synTF$_{coop}$ circuits.

Synthetic cooperative assembly reduces off-target binding in the genome

The data revealed by our RNA-seq experiments are consistent with off-target regulation in the host cell genome underlying the fitness cost associated with expression of a high-affinity synTF. To verify that this misregulation is driven by promiscuous synTF binding events, we performed chromatin immunoprecipitation sequencing (ChIP-seq) analysis of the synTFs across the three circuit strains (synTF$_{high}$, synTF$_{low}$, synTF$_{coop}$), the corresponding strains with synTFs lacking a TAD fusion, and the reporter-only control strain (Figure S5A; STAR Methods). Importantly, we spiked in known quantities of *Schizosaccharomyces pombe*-derived FLAG-tagged DNA, which allowed our data to be normalized to facilitate quantitative comparisons between strains.[68] In addition to the reporter locus, which showed the expected strong enrichment of synTF binding in both the synTF$_{high}$ and synTF$_{coop}$ strains, for both ±TAD conditions (Figures S5B and S5C), we also observed significant enrichment of synTF binding at 23 sites in synTF$_{high}$, 5 in synTF$_{coop}$, and none in synTF$_{low}$ (Figure 4A). To evaluate whether these 28 sites could potentially mediate off-target synTF misregulation, we filtered them on the basis of two criteria: (1) whether the site was robust and not a potential pull-down artifact based on its presence for strains both with and without the TAD fusion,[69] and (2) proximity of the alignment peak (within 700 bp) to a putative synTF CRM, as determined

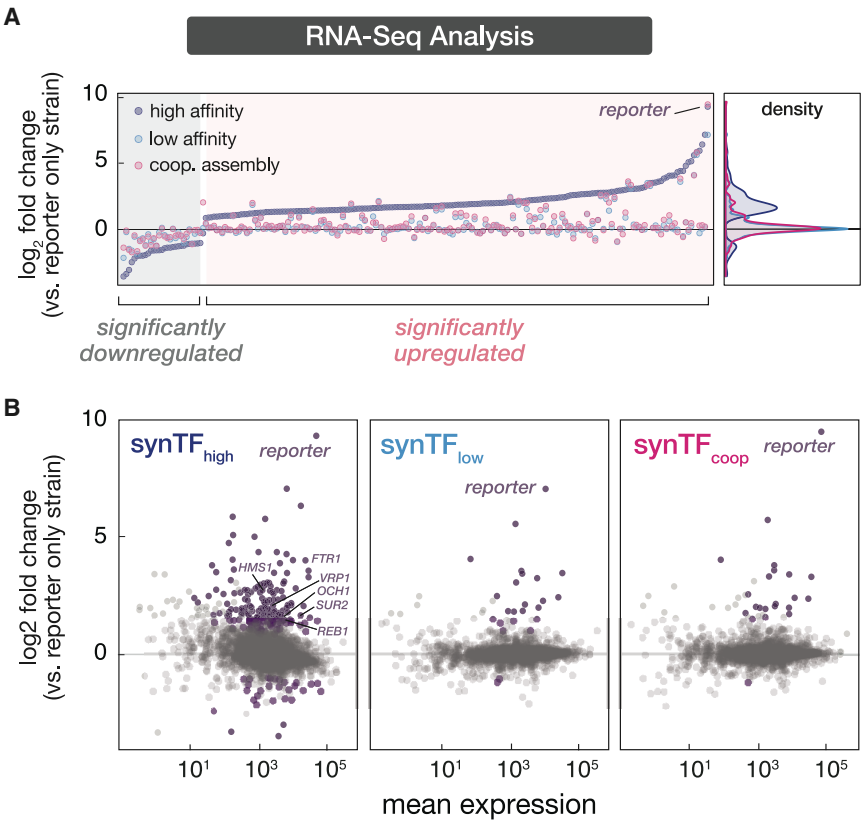


Figure 3. Cooperative assembly rescues aberrant gene expression caused by synthetic circuits

(A) Differential gene expression analysis of RNA-seq measurements following the induction of $synTF_{high}$, $synTF_{low}$, and $synTF_{coop}$ circuits. Plotted are genes that are significantly differentially regulated relative to reporter-only control. The $synTF_{high}$ circuit induces a global misregulation of the host transcriptome, including significantly upregulating 182 genes. Gene expression density distributions of $synTF_{coop}$ and $synTF_{low}$ strains are highly similar to one another and cluster tightly around reporter-only backgrounds.

(B) Differential gene expression profiles for $synTF_{high}$, $synTF_{low}$, and $synTF_{coop}$ strains, plotted for all genes. Purple dots denote genes that are significantly differentially regulated vs. reporter-only strain. The reporter was the only differentially regulated gene in $synTF_{coop}$ vs. $synTF_{low}$.

Related to Figure S4.

by an independent dataset of quantitated 42-10 nuclease DNA cleavage sequence specificities[70] (STAR Methods). Satisfyingly, we found that the same 10 sites were solely and independently isolated by both criteria and were from the $synTF_{high}$ strain (Figure 4A). In most cases, these sites contained motifs that were directly under the ChIP peak maxima, with some sites harboring multiple motifs clustered under the peak (Figure 4B). Additionally, all of the ten ChIP-seq hits contained top-ranked sequence specificities identified from the independent *in vitro* dataset[70] (Figure S5E). By contrast, none of the 3 robust hits for our $synTF_{coop}$ strain had a correlated motif proximate to a ChIP peak. Furthermore, virtually all of the other motifs that we identified (within 2 kb of peaks) that were not correlated with a likely ChIP-seq binding event were low-ranked ZF binding sequences as determined by the independent *in vitro* experiments. These results provide strong evidence that our ChIP-seq analysis likely identified bona fide binding events for synTFs.

To gain further insight into the potential role of the 10 $synTF_{high}$ ChIP-seq-nominated binding events in conferring fitness defects, we plotted alignment peaks from each of the circuit-containing strains atop their corresponding genomic loci (Figures 4B and S5D), classifying binding events as genic or intragenic based on the position of the CRM (and associated peak) relative to the nearest gene.[68] Here, genic denotes a motif located upstream of a gene, where it is more likely to be involved in transcriptional activation, whereas intragenic denotes a motif located within an open reading frame (ORF), where its effect on gene transcription is *a priori* less clear (e.g., positive, negative, no effect). We then plotted the RNA-seq-measured expression changes for the reporter and the ChIP-nominated genes for each of the three circuit strains (relative to reporter-only) and found that transcription of all of the genes associated with genic binding events in the $synTF_{high}$ strain were upregulated relative to the control, whereas those associated with intragenic events showed variable regulation (Figure 4C). Importantly, and as expected, the $synTF_{high}$ misregulation patterns were largely rescued in the low-affinity strains, except at the reporter locus, which showed comparable activation in $synTF_{high}$ and $synTF_{coop}$ strains. Altogether, these results strongly implicate off-target synTF binding as the likely source of host cell transcriptional misregulation in the strains harboring the $synTF_{high}$ circuit, an effect that is minimized by cooperative synTF assembly in our $synTF_{coop}$ circuit.

Cooperative synTF regulatory linkages enhance long-term genetic circuit stability

Motivated by the finding that cooperative synTF assemblies can be used to mitigate loss of transcriptional fidelity and the accompanying fitness cost associated with circuit expression, we investigated whether this strategy could also confer long-term circuit stability in continuously growing cultures. To test this, we utilized a customizable, automated bioreactor platform we recently developed, called eVOLVER,[71,72] to perform 5-day continuous culture of strains expressing the $synTF_{high}$, $synTF_{low}$, and $synTF_{coop}$ circuit designs along with a reporter-only control (Figure 5A). Three biological replicates of each strain (for two different ZFs) were inoculated into separate eVOLVER culture vials, induced with 100 nM β-estradiol, and grown under a turbidostat regime for 130 h to continuously maintain cultures at a constant density (STAR Methods). Growth rates were measured for each culture throughout the experiment, and cultures were periodically sampled to assess circuit output and synTF concentration.

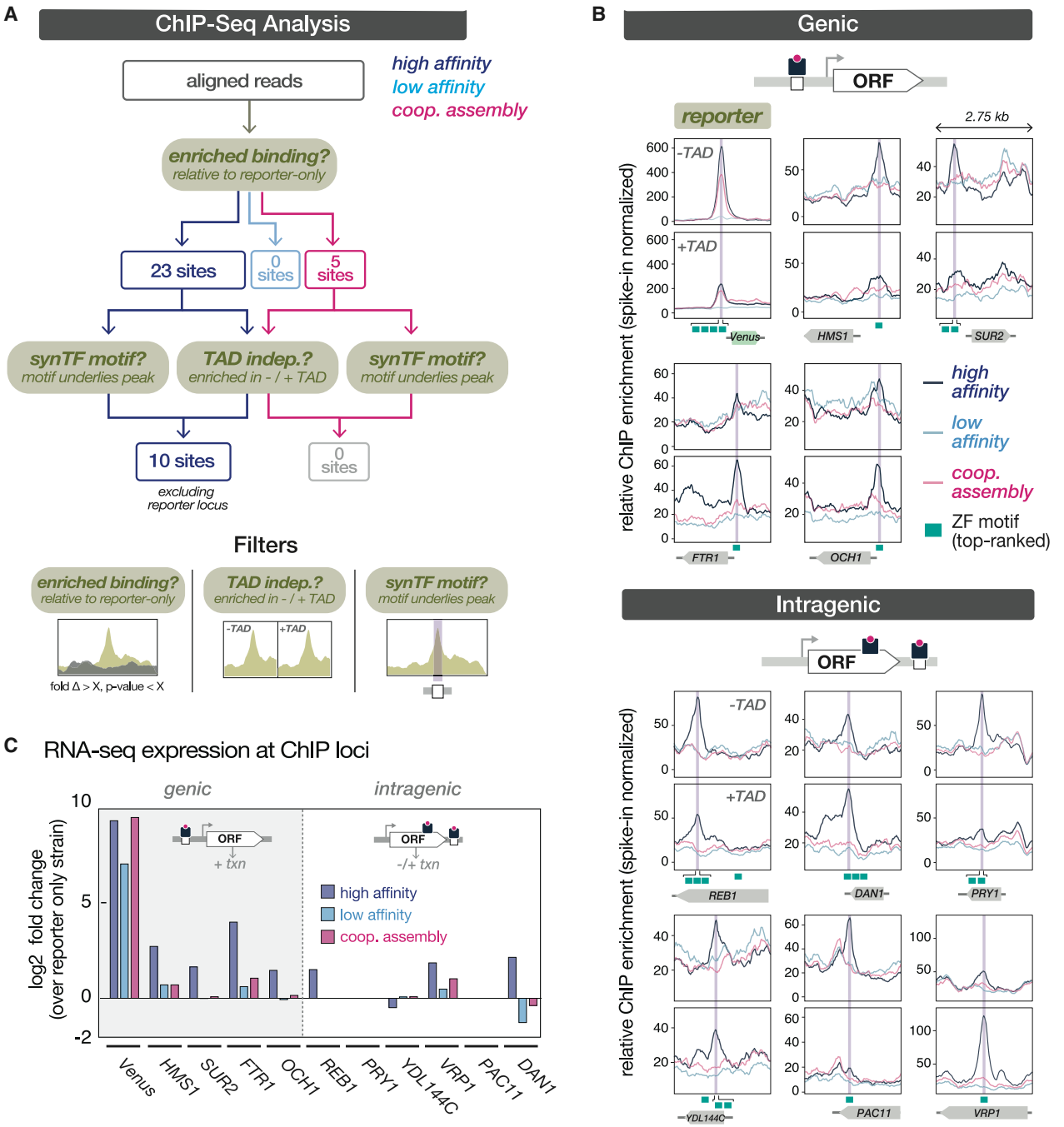


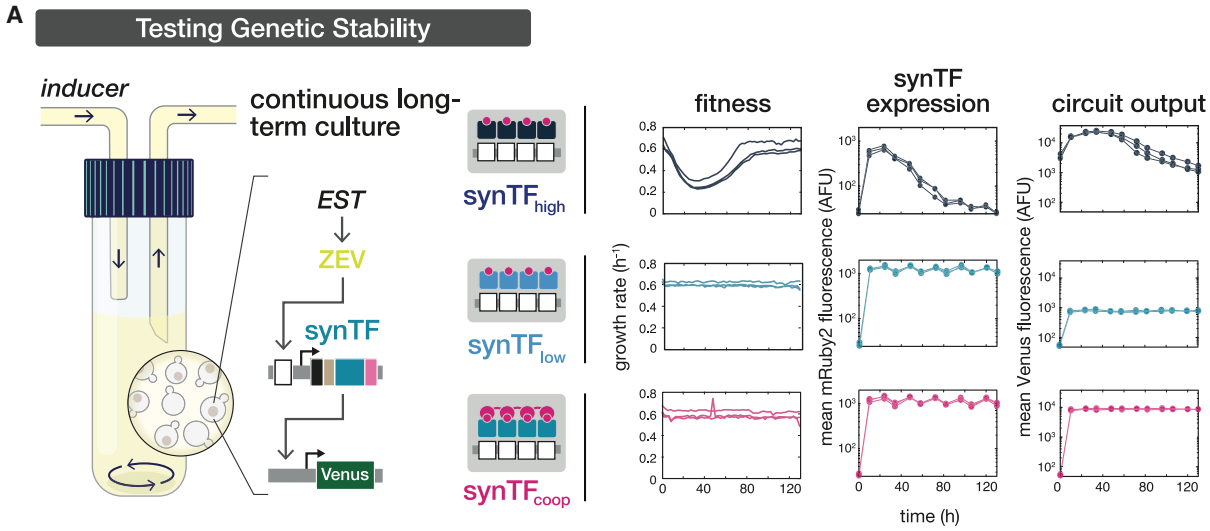
Figure 4. Synthetic cooperative assembly reduces off-target binding in the host genome

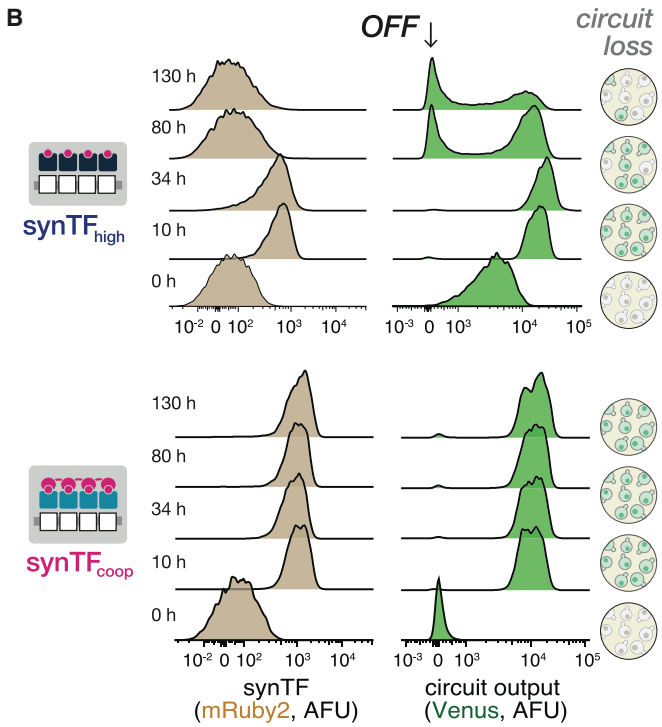
(A) ChIP-seq analysis pipeline for identifying genome-wide binding events by synTFs. Significant binding enrichment relative to the reporter-only strain was observed at 28 endogenous sites. These hits were subsequently filtered based on two criteria: (1) presence in strains with and without TAD fusion and (2) proximity to synTF motif. This analysis yielded 10 sites enriched in the $synTF_{high}$ strain that were identified as binding events. The synthetic reporter locus was the only enriched site that met these criteria in the $synTF_{coop}$ strain (along with the $synTF_{high}$ strain).

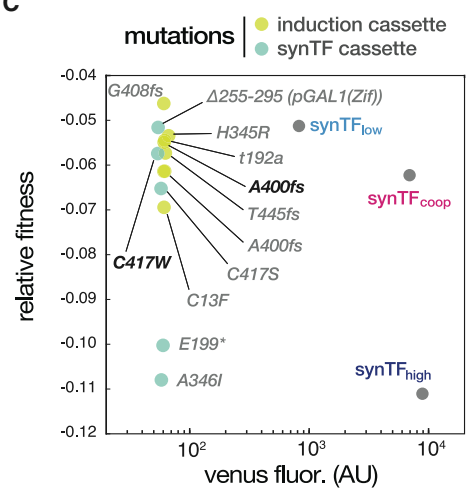
(B) synTF ChIP enrichment patterns at the 10 nominated binding sites, classified as genic if located upstream of a gene TSS or intragenic if located within a gene body. Location of top-ranked binding sequences (as determined by *in vitro* studies) are denoted by green boxes and were highly correlated with bound regions. Relative ChIP enrichment was normalized to FLAG-tagged *S. pombe* spike-in DNA that was produced in parallel with the *Saccharomyces cerevisiae* samples.

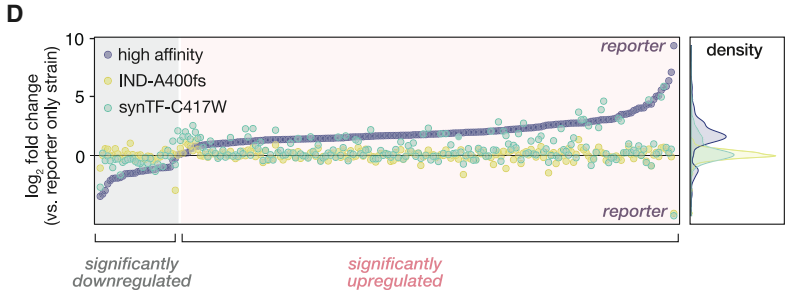
(C) RNA-seq differential expression of genes associated with a synTF ChIP binding event. In general, $synTF_{high}$ genic binding events were associated with higher gene expression, except at the reporter locus where $synTF_{high}$ and $synTF_{coop}$ exhibited similar expression levels. Bars represent the log_2 transformed fold change in transcription for each strain ($synTF_{high}$, $synTF_{low}$, $synTF_{coop}$) over the reporter-only control at each labeled gene.

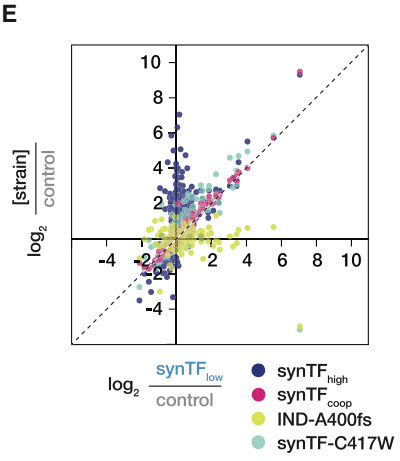
Related to Figure S5.











(legend on next page)

Following circuit induction, we observed a rapid collapse in the growth rate of the synTF$_{high}$ strain, followed by a slow recovery phase over ~40 h (Figure 5A). This was accompanied by a concomitant decay in the synTF concentration (mRuby) to pre-induction levels, followed by loss of reporter expression (Venus). Single-cell fluorescence-activated cell sorting (FACS) distributions over the time course revealed the gradual emergence of a growing "circuit-off" subpopulation, which appeared concurrently with a population-wide loss of synTF-mRuby expression (Figure 5B, top). By contrast, cultures inoculated with synTF$_{low}$ and synTF$_{coop}$ circuits maintained a growth rate that was similar to the control strain following induction and throughout the duration of the experiment (Figure 5A), with synTF concentration and reporter activation remaining unchanged after reaching a post-induction steady state and retaining a sharply defined "circuit-on" population (Figure 5B, bottom). The same patterns of growth and circuit expression were observed with circuits featuring a different member of our library (13-6) (Figures S6C and S6D).

Adaptive circuit-breaking mutations target synTF expression and function

A plausible explanation for the growth patterns we observed in our eVOLVER experiment is the emergence of adaptive mutations that rescue fitness costs by disabling circuit function, and then ultimately fix within the population by outcompeting cells with intact circuits. To gain insight into whether such mutations could account for our observations, we created a simple computational model designed to simulate populations of cells harboring both functional and broken circuits (STAR Methods). The model accounted for the average synTF concentration in each subpopulation of cells while assuming both a fitness cost proportional to the probability of off-target synTF binding, as well as a constant mutation rate capable of disrupting synTF activity and relieving the fitness cost (Figure S6F). Consistent with our observed experimental results, our model predicted a decrease in the average culture fitness within 20 h after induction, followed by a recovery. Furthermore, as the TF-DNA affinity decreases, the model predicts that fitness is improved and the time to recovery increases, whereas decreasing cooperativity decreases recovery time and reduces fitness (Figure S6F). These predictions are consistent with the occurrence of mutations that select against functional synTF expression underlying the observed culture dynamics.

We next experimentally assessed mutational paths to loss of circuit function by analyzing endpoint genotypes from 50 individual colonies from synTF$_{high}$ and synTF$_{coop}$ cultures. No mutations were observed in the circuit genotype from synTF$_{coop}$-derived colonies, while we found a total of 27 mutated synTF$_{high}$-derived colonies with mutations occurring within the induction cassette (10 distinct) and in the synTF cassette (5 distinct) (Figure 5C). Mutations were found in each component of the induction cassette, with one TAD residue (A400fs, frameshift) targeted in more than a third of all of the colonies (Figure S6B). Distinct mutations in the synTF cassette were found in the promoter as well as the coding sequence, with two mutations (C417S and C417W) found in the same cysteine of the synTF Cys2His2 ZF backbone, suggesting that disrupting the ability of the synTF to bind DNA is sufficient for fitness rescue and corresponds with the loss of reporter expression in the mutated strains. To confirm that these circuit mutations drive fitness recovery, we tested the effects of each individually in a clean synTF$_{high}$ background (Figure 5C). All of the mutations were shown to disable circuit output, while all but two restored fitness to control levels.

We verified that the mutations were selected for their ability to restore loss of fitness through rescue of host gene network misregulation by performing RNA-seq analysis on two circuit mutants, one from each class: A400fs in the induction cassette (IND-A400fs) and C417W in the synTF cassette (synTF-C417W). We found that either mutation was sufficient to mostly rescue the pattern of gene misregulation induced by synTF$_{high}$ (Figure 5D). Interestingly, the transcriptomic profile of the IND-A400fs mutant showed no significant gene misregulation over the reporter-only control, while the profile of the synTF-C417W mutant circuit showed similarity to synTF$_{low}$ and synTF$_{coop}$ strains (Figure 5E). These results reinforce functional synTF expression as the basis for synTF$_{high}$ circuit instability and, furthermore, indicate that our engineering strategy for rescuing this fitness defect by lowering synTF affinity recapitulates the growth phenotype of adaptive circuit-breaking mutations.

Cooperative assembly mediated positive feedback circuits for stable long-term activation memory in continuous culture

We next sought to use our cooperative assembly scheme to engineer more complex circuit architectures to validate that our strategy can scale across network architectures featuring >1

Figure 5. Cooperative regulatory linkages enhance the long-term genetic stability of synthetic circuits

(A) Testing long-term stability of synTF circuits in eVOLVER, an automated continuous culture system with real-time measurements of cellular fitness. Individual bioreactor vials were inoculated with three biological replicates of each strain (synTF$_{high}$, synTF$_{low}$, synTF$_{coop}$, reporter-only). Cultures were induced and continuously grown in inducer media (100 nM β-estradiol) using a turbidostat routine. Samples were periodically taken to measure synTF concentration and circuit output by flow cytometry. Points represent a sample from each of three eVOLVER vials per strain type.

(B) Single-cell flow cytometry distributions of synTF and circuit reporter expression over the time course of the continuous culture experiment.

(C) Characterizing circuit genotype mutations selected from the eVOLVER continuous culture experiment. Two classes of mutations were identified from synTF$_{high}$-derived colonies: mutations in the induction cassette (yellow) and synTF cassette (green) (see Figure S6B). Each mutation was introduced into a clean synTF$_{high}$ circuit background and quantified for fitness (by growth competition) and reporter expression (by Venus fluorescence). All of the mutations disabled circuit output, while all except two were sufficient to restore fitness to control levels. Points represent mean values for three colonies ± SD.

(D) Adaptive circuit-breaking mutations rescue the pattern of gene misregulation induced by the synTF$_{high}$ circuit. RNA-seq differential gene expression analysis for the synTF$_{high}$ circuit and two mutant genotypes: IND-A400fs (induction cassette) and synTF-C417W (synTF cassette). Plotted are genes that are significantly differentially regulated relative to the reporter-only control.

(E) Correlation of transcriptomes for various circuit genotypes versus the synTF$_{low}$ circuit genotype. Control, reporter-only genotype.

Related to Figure S6.

assembly regulated nodes and to potentially address a biotechnological need. We investigated the design of positive feedback loops, which are central to the function of numerous natural and engineered networks.[57,73–77] A well-established property of these motifs is their ability to extend the lifetime of network activation after an input signal is removed, thus providing a basis for cellular memory.[78] This capability could be valuable in biomanufacturing applications, where the production of chemicals and proteins can be improved by inducing temporally discreet expression of biosynthetic pathway genes under particular conditions in a fed-batch or continuous culture process.[79] However, because it is costly to continuously supply chemical inducers in large-volume bioreactors to maintain gene activation, one attractive possibility is to deploy positive feedback circuits that enable sustained expression following a transient pulse of chemical inducer.

To engineer positive feedback circuits, we expanded our existing 2-node inducible circuit design to three nodes (Figures 6A and S7A). Here, a first synTF species (containing ZF: 43-8) acts as a ''trigger'' to drive expression of a second synTF (containing ZF: 42-10), which autoregulates its own transcription and also activates the downstream Venus reporter. Both synTF species in this circuit are low affinity, with complexes at the autoregulatory and reporter nodes both mediated by clamp binding, which we hypothesized would enable genetic stability. We simulated transient induction in a bioreactor by testing circuit activation in a 5-day eVOLVER experiment. We compared the 3-node positive feedback circuit to a no-feedback control (no autoregulatory 42-10 CRMs at the second node) and our existing two-node circuit, adding inducer (100 nM β-estradiol) for 12 h to continuously growing strains, and then switching back to uninduced growth media for the rest of the time course (Figure 6A). As expected, the positive feedback circuit exhibited robust activation memory relative to the non-feedback circuits, which decayed rapidly after inducer removal (Figures 6A and 6B). Consistent with these circuit output dynamics, expression of the autoregulated synTF was maintained in the positive feedback circuit but diminished over time in controls. The role of positive feedback in maintaining circuit activity was further demonstrated by versions of the 3-node circuit where we weakened the feedback loop, either by reducing the number of CRMs in the feedback complex or lowering clamp binding affinity (Figure S7), and both circuits demonstrated more rapid signal decays compared with our original feedback design. Importantly, we observed that all 3-node circuit designs (both feedback and non-feedback controls) maintained a consistent and high growth rate throughout the time course, suggesting maintenance of genetic stability (Figures 6A and S7B). Following the 5 h time course, we re-administered inducer to each of the circuits, demonstrating their full reactivation and further supporting the conclusion that signal decay seen in weak feedback and control strains is a consequence of dynamic circuit properties and not mutation-driven loss of circuit function. Altogether, these results demonstrate that cooperative assembly mediated feedback circuits enable robust and tunable activation for a circuit function that could address bioproduction and other biotechnology needs. In addition, they validate the scalability of our framework, which has the potential to generate more complex circuits with high regulatory specificity, fitness, and long-term genetic stability.

DISCUSSION

In this study, we investigated the fundamental question of how exquisite regulatory specificity is achieved in gene regulatory networks, despite the widespread prevalence of natural TF CRMs with surprisingly low-information content. Using synthetic gene circuits, genome-wide measurements of transcription (RNA-seq) and TF binding (ChIP-seq), and mathematical models, we found that high specificity emerges simply from cooperative interactions among TF regulatory proteins that individually interact weakly and non-specifically. Further, our results show that cooperative TF assemblies can be used to engineer highly specific regulatory connections in gene circuits, offering a means for enhancing circuit performance and minimizing circuit-imposed fitness costs in eukaryotic cells. We initially observed that expression of synthetic gene circuits constructed from ZF-based synTFs results in observable growth defects due to misregulation of the native transcriptional network in yeast. Using long-term continuous culture experiments, we demonstrated that these fitness costs drive the gradual loss of circuits from the population as adaptive mutants with abrogated circuit function acquire a selective growth advantage over circuit-bearing cells. In agreement with simple models of gene regulation and evolutionary dynamics, we found that network fidelity and host cell fitness could be restored, and circuits stabilized, by engineering cooperative complexes that render circuit connections functionally dependent on multivalent assembly of weakly interacting synTFs. Collectively, this work demonstrates that our naturally inspired strategy can be harnessed to effectively insulate synthetic circuits from cross-talk with host regulatory networks, thus enabling the rapid development of circuits with enhanced stability against evolutionary pressures (Figure 7).

In recent years, numerous studies have revealed that synthetic circuits are susceptible to unintended interactions with endogenous cellular processes.[42–44] These interactions generally impede circuit function, though in some cases they have been shown to serendipitously support it.[80,81] Thus, examining the interface between synthetic circuits and the host, and developing strategies to functionally insulate circuits from the host cell have become central objectives in synthetic biology.[82,83] Recent studies characterizing synthetic circuits in *Escherichia coli* have established that unintended circuit-host coupling can arise when competition for cellular resources leads to circuit-imposed burden.[42,44] These observations have motivated the development of numerous circuit insulation strategies.[43–45,84,85]

In this study, we offer evidence that transcriptional misregulation resulting from off-target genomic binding constitutes another class of fitness-reducing circuit-host interaction—one that is potentially a primary source of disruption to circuit function in eukaryotic host cells due to their genomic complexity (Figure 7B). Results from our RNA-seq and ChIP-seq experiments provide evidence that the misregulation of host transcription caused by synTF circuits is likely the result of interactions with a select subset of genomic CRMs located primarily, but not exclusively, adjacent to sites of native gene transcriptional

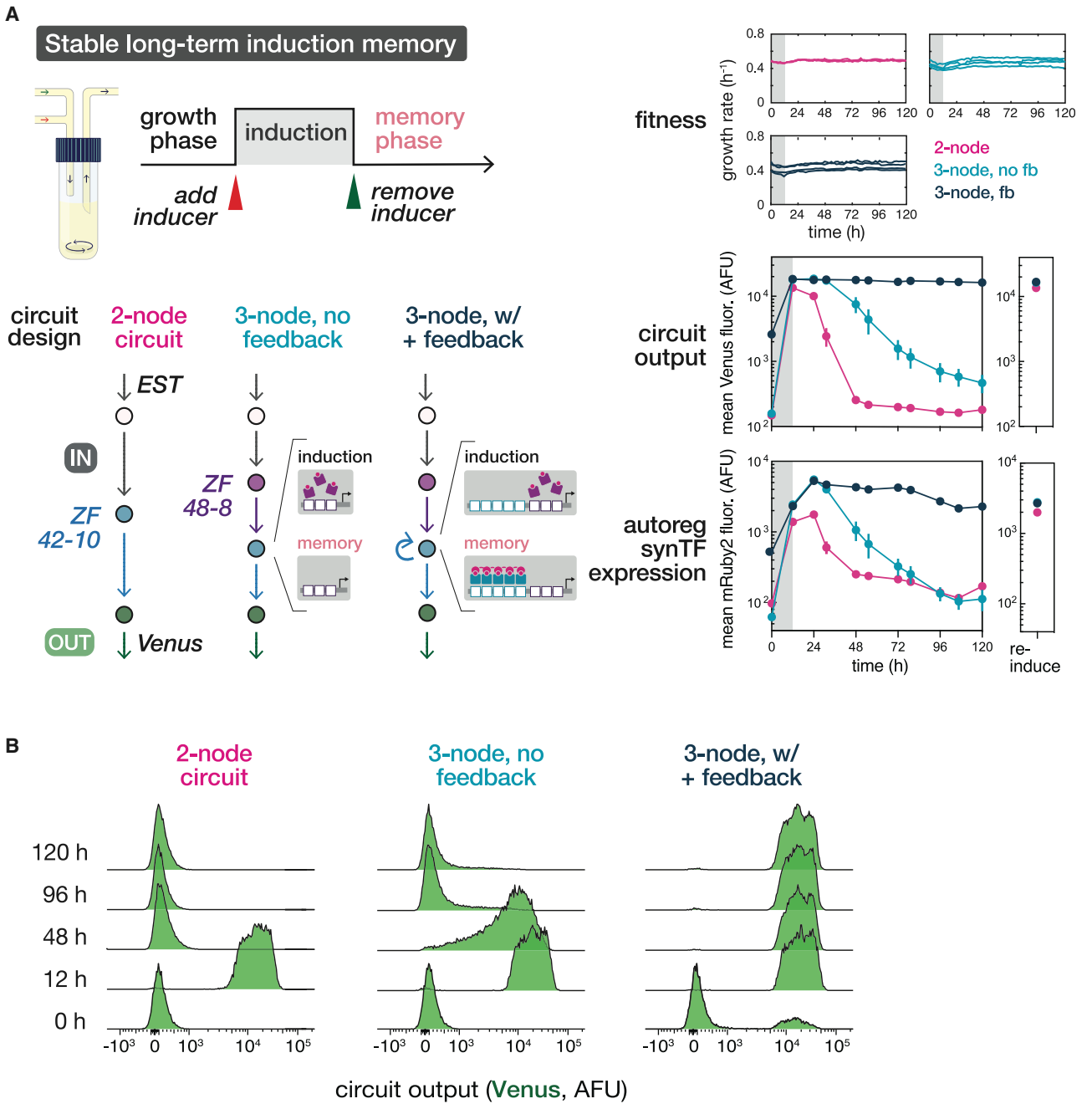


Figure 6. Using cooperative assembly to engineer positive feedback circuits for stable activation memory in long-term continuous culture

(A) Testing activation memory of synTF circuits in eVOLVER following transient induction. Individual bioreactor vials were inoculated with four biological replicates of each strain (2-node circuit; 3-node, no fb; 3-node, fb) and grown continuously using a turbidostat routine. Following inoculation and growth stabilization, cultures were first transiently induced by growing in inducer media (100 nM β-estradiol) for 12 h, followed by growth in uninduced media for the remainder of the time course. Samples were periodically taken to measure circuit output and autoregulated synTF concentration by flow cytometry. Points represent mean values for the replicates ± SD.

(B) Single-cell flow cytometry distributions of circuit reporter expression over the time course of the continuous culture experiment.

Related to Figure S7.

initiation. *A priori* prediction of off-target regulatory interactions is challenging and complicated by many factors (e.g., regulatory context, chromatin architecture, cell type), even when quantitative measurements of DNA sequence recognition are available from *in vitro* experiments[50,70]. This motivates the broader question: what are general design strategies that could give

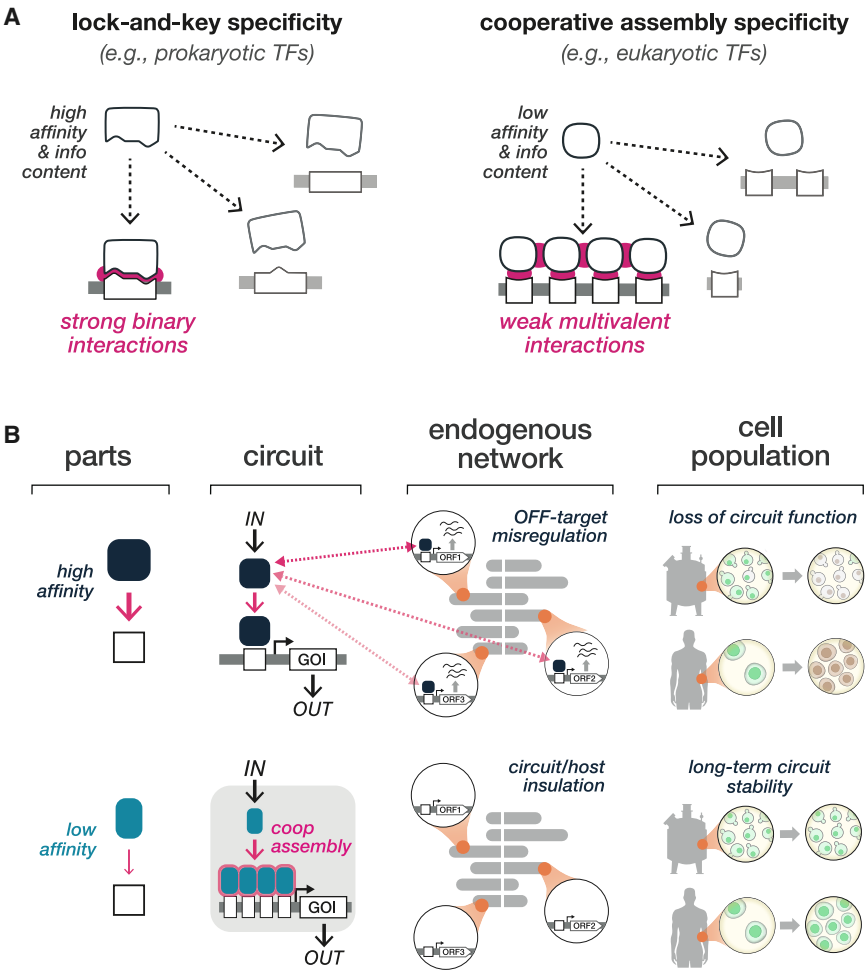


Figure 7. A naturally inspired strategy to engineer regulatory specificity and long-term genetic circuit stability

(A) Different strategies employed by cells to maintain regulatory network specificity. One strategy, common in prokaryotic gene regulation, uses high affinity, large-footprint interactions between TFs and CRMs (lock-and-key specificity). A second strategy, common to gene regulatory networks of higher-order organisms, uses multivalent associations between weakly interacting, low-specificity components (cooperative assembly specificity).

(B) Wiring circuit connections using cooperative assembly is sufficient to enhance regulatory specificity, insulate synthetic circuits from aberrant misregulation of the host genome, and promote long-term stabilization of circuit function.

researchers the ability to rapidly construct new synthetic circuits that confer transcriptional fidelity, especially given the relatively few synthetic biology tools available for optimizing circuits in eukaryotes?

Our solution of using cooperative synTF assemblies to address this challenge draws inspiration from natural eukaryotic transcriptional networks, where combinatorial and cooperative regulation by TFs has been proposed as a key mechanism for the robust and specific rewiring of transcriptional circuitry over evolutionary time.[86–88] This solution is thought to be facilitated by the distribution of binding energy among multiple protein-DNA and protein-protein interactions, which accommodate mutational strengthening or weakening of individual interactions with minimal loss of regulatory robustness.[89] Further, this drift enables the formation and stabilization of new TF-TF and TF-CRM interactions that can facilitate assembly of novel complexes and establish new regulatory connections.[86] Our present work highlights an additional role for cooperative TF assembly as a mechanism to maintain transcriptional network fidelity. Extensive studies have revealed that eukaryotic TFs tend to bind CRMs that are overwhelmingly too short and degenerate to specify unique addresses in the large genome[14,22] (Figure 7A). One potential solution to this "specificity paradox" is employing clusters of low-affinity binding sites, thus making specificity and regulatory robustness dependent on the collective action of multiple TFs.[17] Indeed, CRMs are often shorter and further from consensus in promoter regions regulated by multiple TFs,[90] while frequently interacting TF pairs have been shown to generate composite motifs with unique binding specificity.[19,91] Collectively, these observations suggest that by relaxing the importance of any single interaction within a complex, individual TF-DNA interaction are less likely to be functional and deleterious upon the likely appearance of spurious binding sites in a large genome—a strategy that amounts to optimizing the "hub" rather than individually addressing the "spokes."

From a synthetic biology perspective, our work demonstrates that programming cooperative assembly is a robust, generalizable design strategy for engineering insulated synthetic gene circuitry that minimizes cycles of *ad hoc* design. Unlike prevailing strategies that rely on sophisticated biomolecular engineering to develop highly specialized regulatory components for wiring connections, circuits that employ cooperative assemblies can be constructed from existing parts by weakening their interaction affinity and engineering cooperative interactions between combinations of components. This approach requires no *a priori* knowledge of binding and misregulation profiles and, furthermore, minimizes the need to fine-tune the regulation of expression levels to manage component toxicity. In addition to simplifying circuit design, engineering cooperative assemblies may provide useful and complementary approaches to examine design principles governing how specificity is encoded in natural regulatory systems.[82,92]

Finally, because our approach offers a potential means for engineering gene circuit stability, it could prove impactful in biotechnology applications that demand maintenance of circuit function over many generations (Figure 7B). For example, in

metabolic engineering, strains harboring circuit-controlled biosynthesis pathways must maintain function when they reach bioreactor capacity during growth phases.[93,94] Similarly, cell-based therapy applications typically require the expansion of genetically engineered cells to achieve products that are sufficiently large for patient dosing. In both cases, any burden imposed through circuit-host interactions would not only slow production but could potentially give rise to circuit-deficient subpopulations. In the case of metabolic engineering, this might result in uncontrolled or early activation of metabolic pathways that lower yield, whereas in cell therapy applications, potential effects on product potency and purity could diminish both the safety and efficacy of a treatment. Although post-expansion induction of circuits using exogenously activated transcriptional switches offers one potential solution, the opportunity for misregulation still exists, and the requirement to add an inducer molecule imposes an additional cost on the process. By relieving circuit burden through regulatory insulation, our approach offers a solution to both of these issues that can be applied to existing circuit design strategies by engineering interactions to accommodate regulatory assemblies. Finally, it is possible that the design strategies we developed here could be translated more broadly to other molecular settings, including the engineering of post-translational networks mediated by protein-protein interactions, where specificity of cooperative assemblies encode specific subcellular localization or maintain orthogonality from native interaction network.[95]

Limitations of the study

Despite the apparent generalizability of our cooperative assembly scheme, we anticipate a number of challenges that may be encountered when scaling or deploying this strategy in other subcellular or organismal settings. Although our demonstration of multi-node orthogonal circuit function offers strong initial evidence of the scalability of our approach, limitations may be encountered when constructing higher-order networks, including effects on stoichiometric distribution of the clamp across multiple regulatory complexes. Clamp concentration may be a critical factor for tuning the system, especially when porting our system to other organismal settings. Although clamp was expressed at a concentration permissible for complex assembly in this study, it is possible that lower expression of clamp may not be sufficient to drive assembly, while an excessively high clamp concentration could limit activation due to synTF squelching. Overcoming this challenge may require tuning or even developing additional clamp species to mediate complex assembly. As we discussed above, it may be possible to use our cooperative assembly strategy to engineer specificity in other types of molecular networks (e.g., intracellular signaling or cell-cell communication). However, our synTF toolkit is relatively mature compared with other synthetic part sets, with well-understood and quantitatively predictable biophysical properties. Molecular assembly strategies for other types of networks would likely require extensive component set validation to reach the same level of programmability. For example, developing synthetic pathways that use cooperatively assembling complexes to wire connections may require a suite of new engineered protein components created through a combination of computational design and careful *in vivo* experimental validation.[47,96,97] Finally, we may encounter challenges in porting our strategy to other organismal hosts, particularly to human and other mammalian cells where extensive use of PDZ domains in signaling pathways may preclude our current clamp design. Additionally, programming cooperative assembly using the simple strategy of mutually reinforced binding of synTFs to adjacent genomically integrated CRMs may not be sufficient to achieve strong cooperativity or activation since it does not account for chromatin regulatory mechanisms that underlie much of mammalian transcriptional regulation. This may be especially challenging in primary cells, where transgene silencing is particularly acute.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Strains
- METHOD DETAILS
  - Cloning and plasmid construction
  - Flow cytometry
  - Fitness assay
  - Chromatin immunoprecipitation sequencing (ChIP-seq)
  - Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR)
  - RNA sequencing (RNA-seq)
  - eVOLVER continuous culture
  - Thermodynamic model
  - Population genetics model
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.07.012.

ACKNOWLEDGMENTS

We thank members of the Khalil laboratory, Jané Kondev, and Zeba Wunderlich for their helpful discussions. We are grateful to Fred Winston and his laboratory for providing FLAG-tagged *S. pombe* and for their technical advice on ChIP-seq experiments. This work was supported by the US National Institutes of Health (NIH) (grant nos. R01EB029483 and R01EB027793), the National Science Foundation (NSF) (grant nos. CCF-2027045 and EF-1921677), the Office of Naval Research (ONR) (grant no. N00014-21-1-4006), the Department of Defense Vannevar Bush Faculty Fellowship (no. N00014-20-1-2825), and Schmidt Science Polymath Award (no. G-22-63292).

AUTHOR CONTRIBUTIONS

N.P., C.J.B., and A.S.K. conceived the study. M.D.J.B., N.P., C.J.B., and A.S.K. designed the study. M.D.J.B. and N.P. designed and generated all

genetic constructs and strains. M.D.J.B. performed all experiments and analyzed data with N.P. N.P. developed the thermodynamic model of transcription regulation. E.L. developed the evolutionary dynamics model. J.C. assisted with optimization of the RNA-seq and ChIP-seq protocols and performed all the genomics analyses. C.J.B. and A.S.K. supervised the study. M.D.J.B., C.J.B., and A.S.K. drafted the manuscript with input from all authors.

DECLARATION OF INTERESTS

N.P., C.J.B., and A.S.K. are co-inventors on a patent related to engineered cooperativity and control of gene expression. A.S.K. is a scientific advisor for and holds equity in Senti Biosciences and Chroma Medicine and is a co-founder of Fynch Biosciences and K2 Biotechnologies.

Received: May 16, 2022
Revised: May 17, 2023
Accepted: July 10, 2023
Published: August 7, 2023

REFERENCES

1. Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. Proc. Natl. Acad. Sci. USA *102*, 4936–4942.
2. Berg, J., Willmann, S., and Lässig, M. (2004). Adaptive evolution of transcription factor binding sites. BMC Evol. Biol. *4*, 42.
3. Froula, J.L., and Francino, M.P. (2007). Selection against spurious promoter motifs correlates with translational efficiency across bacteria. PLoS One *2*, e745.
4. Gera, T., Jonas, F., More, R., and Barkai, N. (2022). Evolution of binding preferences among whole-genome duplicated transcription factors. eLife *11*, e73225.
5. Hahn, M.W., Stajich, J.E., and Wray, G.A. (2003). The effects of selection against spurious transcription factor binding sites. Mol. Biol. Evol. *20*, 901–906.
6. McKeown, A.N., Bridgham, J.T., Anderson, D.W., Murphy, M.N., Ortlund, E.A., and Thornton, J.W. (2014). Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell *159*, 58–68.
7. Moses, A.M., Chiang, D.Y., Kellis, M., Lander, E.S., and Eisen, M.B. (2003). Position specific variation in the rate of evolution in transcription factor binding sites. BMC Evol. Biol. *3*, 19.
8. SenGupta, A.M., Djordjevic, M., and Shraiman, B.I. (2002). Specificity and robustness in transcription control networks. Proc. Natl. Acad. Sci. USA *99*, 2072–2077.
9. Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. Nature *443*, 415–420.
10. Zarrinpar, A., Park, S.H., and Lim, W.A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. Nature *426*, 676–680.
11. Gillinder, K.R., Ilsley, M.D., Nébor, D., Sachidanandam, R., Lajoie, M., Magor, G.W., Tallack, M.R., Bailey, T., Landsberg, M.J., Mackay, J.P., et al. (2017). Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability. Nucleic Acids Res. *45*, 1130–1143.
12. Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.
13. Morgunova, E., and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. Curr. Opin. Struct. Biol. *47*, 1–8.
14. Wunderlich, Z., and Mirny, L.A. (2009). Different gene regulation strategies revealed by analysis of binding motifs. Trends Genet. *25*, 434–440.
15. Browning, D.F., and Busby, S.J. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. *2*, 57–65.
16. Madan Babu, M., and Teichmann, S.A. (2003). Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res. *31*, 1234–1244.
17. Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell *160*, 191–203.
18. Jaeger, S.A., Chan, E.T., Berger, M.F., Stottmann, R., Hughes, T.R., and Bulyk, M.L. (2010). Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. Genomics *95*, 185–195.
19. Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. Cell *152*, 327–339.
20. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *175*, 598–599.
21. Li, X.Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. *6*, e27.
22. Tanay, A. (2006). Extensive low-affinity transcriptional interactions in the yeast genome. Genome Res. *16*, 962–972.
23. Zhu, C., Byers, K.J., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. *19*, 556–566.
24. Bell, S.P., Learned, R.M., Jantzen, H.M., and Tjian, R. (1988). Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. Science *241*, 1192–1197.
25. Friedlander, T., Prizak, R., Guet, C.C., Barton, N.H., and Tkačik, G. (2016). Intrinsic limits to gene regulation by global crosstalk. Nat. Commun. *7*, 12307.
26. Panne, D. (2008). The enhanceosome. Curr. Opin. Struct. Biol. *18*, 236–242.
27. Ptashne, M. (1988). How eukaryotic transcriptional activators work. Nature *335*, 683–689.
28. Todeschini, A.L., Georges, A., and Veitia, R.A. (2014). Transcription factors: specific DNA binding and specific gene regulation. Trends Genet. *30*, 211–219.
29. Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328.
30. Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.
31. Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.
32. Gao, A., Shrinivas, K., Lepeudry, P., Suzuki, H.I., Sharp, P.A., and Chakraborty, A.K. (2018). Evolution of weak cooperative interactions for biological specificity. Proc. Natl. Acad. Sci. USA *115*, E11053–E11060.
33. Cameron, D.E., Bashor, C.J., and Collins, J.J. (2014). A brief history of synthetic biology. Nat. Rev. Microbiol. *12*, 381–390.
34. Meng, F., and Ellis, T. (2020). The second decade of synthetic biology: 2010–2020. Nat. Commun. *11*, 5174.
35. Fischbach, M.A., Bluestone, J.A., and Lim, W.A. (2013). Cell-based therapeutics: the next pillar of medicine. Sci. Transl. Med. *5*, 179ps177.
36. Kitada, T., DiAndreth, B., Teague, B., and Weiss, R. (2018). Programming gene and engineered-cell therapies with synthetic biology. Science *359*, eaad1067.
37. Voigt, C.A. (2020). 2020–2030: six commercially-available products that are changing our world. Nat. Commun. *11*, 6379.

38. Xie, M., and Fussenegger, M. (2018). Designing cell function: assembly of synthetic gene circuits for cell biology applications. Nat. Rev. Mol. Cell Biol. *19*, 507–525.

39. Bashor, C.J., and Collins, J.J. (2018). Understanding biological regulation through synthetic biology. Annu. Rev. Biophys. *47*, 399–423.

40. Brophy, J.A., and Voigt, C.A. (2014). Principles of genetic circuit design. Nat. Methods *11*, 508–520.

41. Son, H.I., Weiss, A., and You, L. (2021). Design patterns for engineering genetic stability. Curr. Opin. Biomed Eng. *19*, 100297.

42. Borkowski, O., Ceroni, F., Stan, G.B., and Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. Curr. Opin. Microbiol. *33*, 123–130.

43. Ceroni, F., Algar, R., Stan, G.B., and Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat. Methods *12*, 415–418.

44. Gorochowski, T.E., Espah Borujeni, A., Park, Y., Nielsen, A.A., Zhang, J., Der, B.S., Gordon, D.B., and Voigt, C.A. (2017). Genetic circuit characterization and debugging using RNA-seq. Mol. Syst. Biol. *13*, 952.

45. Ceroni, F., Boo, A., Furini, S., Gorochowski, T.E., Borkowski, O., Ladak, Y.N., Awan, A.R., Gilbert, C., Stan, G.B., and Ellis, T. (2018). Burden-driven feedback control of gene expression. Nat. Methods *15*, 387–393.

46. Müller, I.E., Rubens, J.R., Jun, T., Graham, D., Xavier, R., and Lu, T.K. (2019). Gene networks that compensate for crosstalk with crosstalk. Nat. Commun. *10*, 4028.

47. Ng, A.H., Nguyen, T.H., Gómez-Schiavon, M., Dods, G., Langan, R.A., Boyken, S.E., Samson, J.A., Waldburger, L.M., Dueber, J.E., Baker, D., et al. (2019). Modular and tunable biological feedback control using a de novo protein switch. Nature *572*, 265–269.

48. Riglar, D.T., Giessen, T.W., Baym, M., Kerns, S.J., Niederhuber, M.J., Bronson, R.T., Kotula, J.W., Gerber, G.K., Way, J.C., and Silver, P.A. (2017). Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. Nat. Biotechnol. *35*, 653–658.

49. Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., and Collins, J.J. (2012). A synthetic biology framework for programming eukaryotic transcription functions. Cell *150*, 647–658.

50. Bashor, C.J., Patel, N., Choubey, S., Beyzavi, A., Kondev, J., Collins, J.J., and Khalil, A.S. (2019). Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. Science *364*, 593–597.

51. Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. *29*, 183–212.

52. Donahue, P.S., Draut, J.W., Muldoon, J.J., Edelstein, H.I., Bagheri, N., and Leonard, J.N. (2020). The COMET toolkit for composing customizable genetic programs in mammalian cells. Nat. Commun. *11*, 779.

53. Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., and Khalil, A.S. (2014). Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell *158*, 110–120.

54. Nguyen, T.H., Dods, G., Gómez-Schiavon, M., Muziyue, W., Chen, Z., Kibler, R., Baker, D., El-Samad, H., and Ng, A.H. (2022). Competitive displacement of de novo designed heterodimers can reversibly control protein–protein interactions and implement feedback in synthetic circuits. Gen Biotechnol. *1*, 91–100.

55. Park, M., Patel, N., Keung, A.J., and Khalil, A.S. (2019). Engineering epigenetic regulation using synthetic read-write modules. Cell *176*, 227–238.e20.

56. Zhu, I., Liu, R., Garcia, J.M., Hyrenius-Wittsten, A., Piraner, D.I., Alavi, J., Israni, D.V., Liu, B., Khalil, A.S., and Roybal, K.T. (2022). Modular design of synthetic receptors for programmed gene regulation in cell therapies. Cell *185*, 1431–1443.e16.

57. Zhu, R., Del Rio-Salgado, J.M., Garcia-Ojalvo, J., and Elowitz, M.B. (2022). Synthetic multistability in mammalian cells. Science *375*, eabg9765.

58. Newby, G.A., Kiriakov, S., Hallacli, E., Kayatekin, C., Tsvetkov, P., Mancuso, C.P., Bonner, J.M., Hesse, W.R., Chakrabortee, S., Manogaran, A.L., et al. (2017). A genetic tool to track protein aggregates and control prion inheritance. Cell *171*, 966–979.e18.

59. Lohmueller, J.J., Armel, T.Z., and Silver, P.A. (2012). A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. Nucleic Acids Res. *40*, 5180–5187.

60. Li, H.S., Israni, D.V., Gagnon, K.A., Gan, K.A., Raymond, M.H., Sander, J.D., Roybal, K.T., Joung, J.K., Wong, W.W., and Khalil, A.S. (2022). Multidimensional control of therapeutic human cell function with synthetic gene circuits. Science *378*, 1227–1234.

61. Stewart, A.J., Hannenhalli, S., and Plotkin, J.B. (2012). Why transcription factor binding sites are ten nucleotides long. Genetics *192*, 973–985.

62. Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

63. Buchler, N.E., Gerland, U., and Hwa, T. (2003). On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA *100*, 5136–5141.

64. Garcia, H.G., Kondev, J., Orme, N., Theriot, J.A., and Phillips, R. (2011). Thermodynamics of biological processes. Methods Enzymol. *492*, 27–59.

65. Gertz, J., Siggia, E.D., and Cohen, B.A. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature *457*, 215–218.

66. Elrod-Erickson, M., Rould, M.A., Nekludova, L., and Pabo, C.O. (1996). Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure *4*, 1171–1180.

67. Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809–817.

68. Doris, S.M., Chuang, J., Viktorovskaya, O., Murawska, M., Spatt, D., Churchman, L.S., and Winston, F. (2018). Spt6 is required for the fidelity of promoter selection. Mol. Cell *72*, 687–699.e6.

69. Kidder, B.L., Hu, G., and Zhao, K. (2011). ChIP-Seq: technical considerations for obtaining high-quality data. Nat. Immunol. *12*, 918–922.

70. Pattanayak, V., Ramirez, C.L., Joung, J.K., and Liu, D.R. (2011). Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. Nat. Methods *8*, 765–770.

71. Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., and Khalil, A.S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat. Biotechnol. *36*, 614–623.

72. Heins, Z.J., Mancuso, C.P., Kiriakov, S., Wong, B.G., Bashor, C.J., and Khalil, A.S. (2019). Designing automated, high-throughput, continuous cell growth experiments using eVOLVER. J. Vis. Exp. *147*, e59652.

73. Becskei, A., Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J. *20*, 2528–2535.

74. Ferrell, J.E., Jr. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. Curr. Opin. Cell Biol. *14*, 140–148.

75. Ingolia, N.T., and Murray, A.W. (2007). Positive-feedback loops as a flexible biological module. Curr. Biol. *17*, 668–677.

76. Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A.J., Elowitz, M.B., and Alon, U. (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat. Genet. *36*, 147–150.

77. Xiong, W., and Ferrell, J.E., Jr. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. Nature *426*, 460–465.

78. Jiang, Y., and Hao, N. (2021). Memorizing environmental signals through feedback and feedforward loops. Curr. Opin. Cell Biol. *69*, 96–102.

79. Holtz, W.J., and Keasling, J.D. (2010). Engineering static and dynamic control of synthetic pathways. Cell *140*, 19–23.

80. Cookson, N.A., Tsimring, L.S., and Hasty, J. (2009). The pedestrian watchmaker: genetic clocks from engineered oscillators. FEBS Lett. *583*, 3931–3937.

81. Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., and Hasty, J. (2008). A fast, robust and tunable synthetic gene oscillator. Nature *456*, 516–519.

82. Nandagopal, N., and Elowitz, M.B. (2011). Synthetic biology: integrated gene circuits. Science *333*, 1244–1248.

83. Purnick, P.E., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.

84. Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., and Sommer, M.O.A. (2018). Diverse genetic error modes constrain large-scale bio-based production. Nat. Commun. *9*, 787.

85. Sleight, S.C., Bartley, B.A., Lieviant, J.A., and Sauro, H.M. (2010). Designing and engineering evolutionary robust genetic circuits. J. Biol. Eng. *4*, 12.

86. Sorrells, T.R., Johnson, A.N., Howard, C.J., Britton, C.S., Fowler, K.R., Feigerle, J.T., Weil, P.A., and Johnson, A.D. (2018). Intrinsic cooperativity potentiates parallel cis-regulatory evolution. eLife *7*, e37563.

87. Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

88. Tuch, B.B., Li, H., and Johnson, A.D. (2008). Evolution of eukaryotic transcription circuits. Science *319*, 1797–1799.

89. Sorrells, T.R., Booth, L.N., Tuch, B.B., and Johnson, A.D. (2015). Intersecting transcription networks constrain gene regulatory evolution. Nature *523*, 361–365.

90. Bilu, Y., and Barkai, N. (2005). The design of transcription-factor binding sites is affected by combinatorial regulation. Genome Biol. *6*, R103.

91. Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388.

92. Crocker, J., and Ilsley, G.R. (2017). Using synthetic biology to study gene regulatory evolution. Curr. Opin. Genet. Dev. *47*, 91–101.

93. Lee, S.Y., and Kim, H.U. (2015). Systems strategies for developing industrial microbial strains. Nat. Biotechnol. *33*, 1061–1072.

94. Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T.R., and Mukhopadhyay, A. (2019). Engineering robust production microbes for large-scale cultivation. Trends Microbiol. *27*, 524–537.

95. Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

96. Chen, Z., Boyken, S.E., Jia, M., Busch, F., Flores-Solis, D., Bick, M.J., Lu, P., VanAernum, Z.L., Sahasrabuddhe, A., Langan, R.A., et al. (2019). Programmable design of orthogonal protein heterodimers. Nature *565*, 106–111.

97. Langan, R.A., Boyken, S.E., Ng, A.H., Samson, J.A., Dods, G., Westbrook, A.M., Nguyen, T.H., Lajoie, M.J., Chen, Z., Berger, S., et al. (2019). De novo design of bioactive protein switches. Nature *572*, 205–210.

98. Gopalakrishnan, R., Marr, S.K., Kingston, R.E., and Winston, F. (2019). A conserved genetic interaction between Spt6 and Set2 regulates H3K36 methylation. Nucleic Acids Res. *47*, 3888–3903.

99. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

100. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

101. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

102. Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

103. Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth. Biol. *4*, 975–986.

104. Kryazhimskiy, S., Rice, D.P., Jerison, E.R., and Desai, M.M. (2014). Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science *344*, 1519–1522.

105. Köster, J., and Rahmann, S. (2012). Snakemake–a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522.

106. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

107. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

108. Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

109. Shea, M.A., and Ackers, G.K. (1985). The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. J. Mol. Biol. *181*, 211–230.

110. Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information integration and energy expenditure in gene regulation. Cell *166*, 234–244.

111. Mirny, L.A. (2010). Nucleosome-mediated cooperativity between transcription factors. Proc. Natl. Acad. Sci. USA *107*, 22534–22539.

112. Desai, M.M., Fisher, D.S., and Murray, A.W. (2007). The speed of evolution and maintenance of variation in asexual populations. Curr. Biol. *17*, 385–394.

113. Lang, G.I., and Murray, A.W. (2008). Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics *178*, 67–82.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-FLAG tag monoclonal	Sigma Aldrich	Cat# F1804; RRID:AB_262044
Chemicals, peptides, and recombinant proteins		
Dynabeads Protein G	Thermo Fisher Scientific	Cat# 10004D
Adenine hemisulfate	Sigma Aldrich	Cat# A9126
B-Estradiol	Sigma Aldrich	Cat# E8875-5G
ChIP DNA Clean and Concentrator Column	Zymo Research	Cat# D5205
RnaseA	Thermo Fisher Scientific	Cat# FEREN0531
LightCycler® 480 SYBR Green I Master	Roche	Cat# 4887352001
Yeast extract	VWR	Cat# 90000-726
Bacto peptone	VWR	Cat# 90000-368
D-glucose	Sigma Aldrich	Cat# G7528-1KG
Yeast nitrogen base	VWR	Cat# 90004-146
Complete supplement mixture (CSM) media	Sunrise Science Products	Cat# 1001-100
CSM without uracil	Sunrise Science Products	Cat# 1004-100
CSM without leucine	Sunrise Science Products	Cat# 1005-100
CSM without uracil and leucine	Sunrise Science Products	Cat# 1038-100
Critical commercial assays		
Rneasy Plus Mini Kit	QIAGEN	Cat# 74134
YeaSTAR RNA kit	Zymo Research	Cat# E1004
Deposited data		
Data files for RNA-seq	This study	GEO: GSE203146
Data files for ChIP-seq	This study	GEO: GSE203146
All other raw data files	This study	https://doi.org/10.5061/dryad.zpc866tdg
Experimental models: Organisms/strains		
S. cerevisiae: Strain background: YPH500	ATCC	76626
S. pombe: Spike-in control for ChIP-qPCR: FWP5607	Gopalakrishnan et al.[98]	FWP5607
Oligonucleotides		
ChIP-qPCR primer fwd: 5'-gcgatcacagacattaacccacag-3'	This study	N/A
ChIP-qPCR primer rev: 5'-tggcggatctgggatccga-3'	This study	N/A
Software and algorithms		
FlowJo V8	FlowJo, LLC	N/A
GraphPad Prism	GraphPad Software	N/A
Bowtie2	Langmead and Salzburg[99]	N/A
SAMtools	Li et al.[100]	N/A
BEDTools	Quinlan and Hall[101]	N/A
Custom ChIP-seq analysis code	This study	https://doi.org/10.5281/zenodo.8083144
Custom RNA-seq analysis code	This study	https://doi.org/10.5281/zenodo.8083146
Modeling code: Thermodynamic model and Population Genetics model	This study	https://doi.org/10.5281/zenodo.8083150

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Epoch 2 Microplate Spectrophotometer	BioTek	Epoch 2
Attune NxT Flow Cytometer	Thermo Fisher Scientific	Attune NxT
LightCycler 480 Instrument II	Roche	Cat# 00015243001

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact: Ahmad S. Khalil (khalil@bu.edu).

Materials Availability

Key plasmids have been deposited at Addgene for distribution. DNA constructs and strains are available from the lead contact.

Data and Code Availability

- Raw RNA-seq and ChIP-seq data for transcriptome and binding analyses, respectively, have been deposited in the NCBI GEO database. Accession number is listed in the key resources table. All other raw datasets have been deposited on Dryad. DOI is listed in the key resources table.
- All original code is available on Github and has been deposited on Zenodo. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Strains

The background strain used for all experiments in this study was *S. cerevisiae* YPH500 (α, *ura3-52, lys2-801, ade2-101, trp1, his3, leu21*) (Stratagene). Strains were constructed by sequential plasmid transformations, stably and singly integrating synthetic cassettes into the yeast genome using standard lithium acetate-based transformation techniques and growth on selective minimal media (Sunrise Science Products), using the *URA3*, *LEU2*, and hygromycin B phosphotransferase (*HPH*, integrated into the *HO* locus) genes as selectable markers. Induction and synTF cassettes were integrated into the *HO* locus, reporters were integrated into the *URA3* locus, and clamp (or random spacer in cases without clamp) was integrated into the *LEU2* locus. For the feedback circuits in Figure 6, induction and synTF 1 (ZF 43-8) cassettes were integrated into the *LEU2* locus, reporters were integrated into the *URA3* locus, synTF 2 (ZF 42-10) cassettes were integrated into the *HO* locus, and the clamp was either integrated alongside the reporter in the *URA3* locus or integrated separately in a custom locus with a *TRP1* marker. Genotypes for experimentally tested strains are listed in Table S2. Experimental replicates comprised distinct colonies picked from a transformation plate following construct integration and selection.

METHOD DETAILS

Cloning and plasmid construction

Plasmid constructs used in this study are listed in Table S1 and their designs described in Figures S1 and S3. All plasmids in this study were constructed using Golden Gate Assembly[102] and formatted with the Yeast MoClo Toolkit[103] (Addgene Kit #1000000061). ORFs encoding previously described zinc finger and clamp proteins[49,50] were codon optimized for yeast, adapted for Golden Gate assembly, and synthesized (IDT). BsmBI, T7 DNA Ligase, and T4 DNA Ligase Buffer (NEB) were used to construct Level 0 and Level 2 plasmids. The Golden Gate Assembly Master Mix BsaI-HF v1 and v2 (NEB) was used to construct Level 1 plasmids.

Flow cytometry

Yeast colonies were picked from plates and cultured overnight in 500 μL synthetic defined (SD) media prepared without the appropriate amino acids required for auxotrophic selection. SD media was prepared with Yeast Nitrogen Base without Amino Acids (VWR), 2% w/v D-glucose (Sigma Aldrich), and appropriate CSM amino acid dropout mixture (Sunrise Science Products). Cultures were then diluted 1:50 into 500 μL of non-selective SD media (SDC) and grown for 7 h at 30°C in the presence or absence of inducer (ß-estradiol).

Prior to flow cytometry reading, cells were diluted 1:20 into 200 μL of PBS treated with 20 μg/mL cyclohexamide to inhibit protein synthesis, and stored at 25°C, in the dark, for 1 h to allow for complete fluorophore maturation. Plates were then stored at 4°C

overnight. Typically, 10,000 events were acquired using an Attune Nxt Flow Cytometer equipped with a high throughput autosampler (Thermo Fisher Scientific), and data was processed using FlowJo (Treestar Software). Events were gated by forward and side scatter, and geometric means of the fluorescence distributions were calculated. Flow cytometer laser/filter configurations used in this study were mVenus (488 nm, 574/26), mRuby (561 nm, 620/15), and mTurquoise (405 nm, 512/25).

Fitness assay

We adapted a previously described fitness assay based on competitive growth between a "reference" strain and "query" strain.[104] A single colony for a reference strain constitutively expressing (*pTDH3*) an mTurquoise reporter was grown in 2 mL SDC overnight with shaking at 30°C, then diluted 1:100 in 20 mL SDC and grown overnight with shaking at 30°C. Single colonies for three biological replicates for each query strain expressing synTF circuits or the reporter only were each cultured overnight in 500 μL SD -ura/-leu media in 96 well plates. The reference strain culture was diluted 1:50 into 500 μL of SDC in 96 well plates in the presence or absence of inducer (1 μM ß-estradiol) across four 96 well culture plates, and each query strain was added to the reference strain-containing wells at 1:50 and mixed. A 10 μL sample was immediately sampled from each well and fixed in PBS + 20 μg/mL cycloheximide to obtain a t_0 measurement of the cocultures prior to induction. The cocultures were then diluted 1:50 into SDC with or without inducer every 12 h and samples were isolated at 16.5 h (t_1) and 36 h (t_2) corresponding to ~7 and 15 generations, respectively, and fixed in PBS + 20 μg/mL cyclohexamide for flow cytometry analysis. We determined the relative abundances of reference and query strains at t_0, t_1, and t_2 for each coculture. Abundance was derived from the fraction of cells in each well expressing the mTurquoise reporter (reference). Fitness was computed for each query strain by calculating changes in abundance from t_0 to the experimental endpoint, t_2:

$$F = \frac{1}{t_2 - t_0} log\left(\frac{n(t_2)}{n_r(t_2)} \Big/ \frac{n(t_0)}{n_r(t_0)}\right)$$

where n (t) and n_r (t) are the cell counts for the query and reference strains, respectively, at time t after coculturing.[104]

Chromatin immunoprecipitation sequencing (ChIP-seq)

Preparation, immunoprecipitation, and sequencing

250 mL flasks of SDC were inoculated with overnight cultures and grown for 1 h before induction with 100 nM ß-estradiol, then grown for an additional 8 h to an OD600 of 0.525–0.625. All cultures were diluted to OD600 0.525, then cells were crosslinked with 1% formaldehyde for 9 min at 30°C with shaking. Fixation was quenched with a final concentration of 125 mM glycine (EMD 4840 OmniPur) for 10 min at 30°C with shaking. Cells were pelleted for 10 min at 4°C at 3000 RPM (Haraeus Multifuge X3R), washed twice with ice-cold TE (Tris-HCl, EDTA), transferred to 4 bead-beater tubes/strain and frozen at -80°C. Cell pellets were resuspended in 400 μL ice cold lysis buffer (50 mM HEPES, 140 mM NaCl, 1mM ethylenediaminetetraacetic acid, 1% Triton X-100, 0.1% Na-Deoxycholate, 1 mM phenylmethylsulfonyl fluoride, 200 μL Roche cOmplete protease inhibitors). 0.5 mm diameter glass beads were added to 1 mm below the meniscus. Cells were lysed by bead beating on a MagNA Lyser (Roche) three times for 45 s each at 4500 RPM with 2 min rests at 4°C. Lysate was collected by puncturing the tube with a 21G needle and centrifugation at 2000 g for 2 min into a 2 mL microtube. The pellet was resuspended in lysis buffer, then sonicated for 6 pulses using a probe sonicator (Fisher Scientific FB120) for 20 s at 25% amplitude with 120 s intervening rests on ice, achieving a range of 150-1500 bp DNA fragments. Cell debris was pelleted by centrifugation at max speed for 15 min at 4°C.

FLAG-tagged *S. pombe* (generously provided by the Winston Lab, FWP567) was used as a spike-in control and was prepared similarly to the *S. cerevisiae* cultures with a few modifications: grown in 250 mL YES media to OD600 0.65, split into 5 tubes, underwent 4 lysis steps on the bead beater and 5 sonication steps. The supernatant from the 4 preps of each strain (5 preps of *S. pombe*) were mixed together in a new low retention tube (Thermo Fisher Scientific 02-681-320). To determine DNA concentration, 50 μL samples from each strain were isolated. Samples were brought up to 200 μL with elution buffer, then incubated with 50 μg of RNAse A (Thermo Fisher Scientific) at 37°C for 30 min to remove RNA. Then 100 μg of Proteinase K (Thermo Fisher Scientific) was added and samples were incubated overnight (~16 h) at 60°C to degrade proteins and reverse crosslinks. Samples were then purified with the ChIP DNA Clean and Concentrator kit (Zymo Research), eluted with 100 μL water and concentrations were determined by Qubit 4 Fluorometer (Thermo Fisher Scientific). 50 μL were brought to 13.5 ng/μL concentration and split into 4 separate tubes, then diluted to 1 mL in lysis buffer. Input samples were concurrently isolated at 10% of the DNA concentration for the IP samples and brought to 100 uL lysis buffer. 1 ug anti-FLAG (Sigma F1804) was added to each IP sample. The prepared (lysed and sonicated) FLAG-tagged *S. pombe* chromatin was added as a spike-in control to 10% of the sample DNA concentration for all IP and input samples. Input samples were stored at 4°C and IP samples rotated overnight at 4°C.

30 μL Dynabeads Protein G (10004D, Thermo Fisher Scientific) per culture was added to a low retention tube and washed 3 times with 1 mL ice cold lysis buffer. Dynabeads were resuspended in 100 μL lysis buffer per culture. 100 μL of Dynabead solution was added to each antibody-pull-down sample and incubated at 4°C for 4 h while rotating. Dynabeads were washed at room temperature on a magnet (twice with 1 mL lysis buffer, twice with 1 mL lysis buffer/500 mM NaCl, twice with 10 mM TrisHCl-pH8/250 mM LiCl/0.5% NP-40/0.5% sodium deoxycholate/1 mM EDTA, and once with 1 mL TE). Bound material was eluted by adding 200 μL of 50 mM Tris-HCl ph8/10 mM EDTA/1% SDS and incubating at 65°C for 30 min. A second elution with the same buffer was combined with the

first and tubes were incubated at 65°C overnight to reverse crosslinks. Input samples were brought to 400 μL elution buffer and stored at 65°C with antibody-pull-down samples. 50 μg of RNase A was added to each pull-down and input sample and incubated at 37°C for 30 min. 100 μg Proteinase K was added to each sample, then incubated at 55°C for 4 h. DNA was purified with the ChIP DNA Clean and Concentrator kit (Zymo Research): 4 preps/strain were concentrated into two preps in two columns and eluted with 25 μL of water for IP samples/column (50 μL total) or 200 μL water for input samples and stored at -80°C.

Sample concentrations were measured with the Qubit and 38 μL of each sample was submitted to the Tufts Genomics Core for TruSeq ChIP library preparation (Illumina). Tufts Genomics Core subsequently sequenced all samples, paired end, on a NextSeq 550 (Illumina) to 75 bp.

ChIP-seq analysis

ChIP-seq data analyses were performed using the Snakemake workflow management system.[105] Code and raw data suitable for reproducing all ChIP- and RNA-seq analyses are publicly available (see Resource Availability).

ChIP-seq library processing

Adapter removal and 3' quality trimming of paired-end reads was performed using cutadapt (http://journal.embnet.org/index.php/embnetjournal/article/view/200). Reads were aligned using Bowtie2[99] to a combined genome consisting of *S. pombe* genome ASM294v2 concatenated with *S. cerevisiae* genome build R64-2-1 modified to include the mVenus reporter at the *URA3* locus. Correctly paired uniquely mapping reads mapping to *S. cerevisiae* were selected using SAMtools.[100] Coverage of fragments and fragment midpoints were generated using SAMtools[100] and bedtools,[101] and normalized to the number of fragments in the library. Quality statistics of raw, cleaned, non-aligning, and correctly paired mapping reads were assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Transcription factor ChIP-seq peak calling

Transcription factor peak calling was performed for each strain by calling peaks in each replicate using MACS2,[106] followed by filtering for reproducibility among replicates by the Irreproducible Discovery Rate (IDR) method (https://doi.org/10.1214/11-AOAS466). The size of the small and large local regions used by MACS2 to model expected counts were set to 500 and 2000 bp, respectively, and the IDR threshold was set to 0.01.

Transcription factor ChIP-seq differential binding analysis

For transcription factor ChIP-seq differential binding analysis, transcription factor peaks were called as described above. A non-redundant list of peaks called in the strains being compared was generated using bedtools,[101] and the counts of fragment midpoints from both input and IP samples over these peaks were used as the input to a differential binding analysis with DESeq2,[107] in which the linear model coefficient extracted represents the change in IP/input enrichment in the query strain versus the control strain. We investigated a set of 132 peaks as candidates for specific binding in any of the $synTF_{high}$, $synTF_{low}$, or $synTF_{coop}$ strains over the reporter-only control strain, at a false discovery rate of 0.05.

Chromatin immunoprecipitation quantitative PCR (ChIP-qPCR)

Samples were prepared as described for ChIP-seq. qPCR was performed on a LightCycler 480 Instrument II (Roche) with LightCycler 480 SYBR Green I Master Kit (Roche) according to manufacturer's instructions. A total reaction volume of 10 μL (2 μL of 1:50 dilution of input DNA or 1:20 dilution of IP DNA, 0.5 μM of forward primer, 0.5 μM of reverse primer, 5 μL of 2X SYBR Green Master Mix), using the following cycle conditions: (i) pre-incubation: 95°C for 10 min; (ii) amplification (45 cycles): 95°C for 10 s, 57°C for 20 s, 72°C for 8 s; (iii) melting curve: 95°C for 5 s, 65°C for 1 min, 97°C at ramp rate 0.11C/s; (iv) cooling: 40°C for 10 s. PCR primer sequences were designed to flank the cis-regulatory motifs (CRMs) at the synthetic promoter: gcgatcacagacattaacccacag; tggcggatctgggatccga. Fold enrichment over the reporter-only control strain was then computed from the resulting qPCR Ct values using the ΔΔCt method.

RNA sequencing (RNA-seq)

Preparation and sequencing

RNA-seq measurements were performed on two biological replicates per strain type. Our results were reproduced with a technical replicate for each biological replicate in two separate experiments, aside from $synTF_{high}$, for which we reported on two biological replicates and a single technical replicate. Total RNA was purified from ~$5x10^7$ cells following the "Purification of Total RNA" from the "Yeast Mechanical Disruption" protocol in the RNAeasy Plus Mini Kit handbook: 50 mL of cells from an overnight culture were induced with β-estradiol and cultured for 7 h in a 30°C shaking incubator. Cells were brought to the same concentration, spun down for 5 min at 1000 RCF at 4°C, liquid was removed and the pellets were resuspended in 600 μL RLT buffer + β-mercaptoethanol. ~600 μL of 0.5 mm diameter glass beads were added and cells were lysed by bead beating on a MagNA Lyser (Roche) three times for 45 s each at 4500 RPM with 2 min rests at 4°C. ~300 μL of supernatant was moved into a clean tube, 300 μL of 70% ethanol was added, and samples were processed using the RNeasy Plus Mini Kit (QIAGEN) according to the manufacturer's instructions. Sequencing libraries were prepared at the Tufts University Core Facility (TUCF Genomics) using the TruSeq Stranded mRNA Library Prep Kit (Illumina). 50-bp single-end reads were sequenced on an Illumina HiSeq 2500.

RNA-seq analysis

RNA-seq data analyses were performed using the Snakemake workflow management system.[105] Code and raw data suitable for reproducing all ChIP- and RNA-seq analyses are publicly available (see resource availability).

RNA-seq library processing

Adapter removal and 3′ quality trimming were performed using cutadapt (http://journal.embnet.org/index.php/embnetjournal/article/view/200). Reads were aligned using TopHat2 without a reference transcriptome, against S. cerevisiae genome build R64-2-1 modified to include the Venus reporter at the URA3 locus. Uniquely mapping reads were selected using SAMtools.[100] Coverage of the 5′-most base of the read (3′-most base of the RNA fragment) was extracted using bedtools genomecov,[101] and normalized to the total number of uniquely mapped reads. Quality statistics of raw, cleaned, non-aligning, and uniquely aligning reads were assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

RNA-seq differential expression analysis

RNA-seq differential expression analysis was performed for transcripts of verified coding genes, using an annotation of transcript boundaries based on TIF- [108] and TSS-seq[68] data that was modified to accommodate the Venus reporter at the *URA3* locus. Read counts over these transcripts were input to differential expression analysis with DESeq2.[107]

eVOLVER continuous culture

eVOLVER experiments were run as previously described,[71,72] with the following modifications. Following eVOLVER sterilization, each vial was inoculated from saturated overnight cultures and run at 25 mL total volume at 30°C with stirring. For all experiments, yeast strains were grown in synthetic complete (SC) media supplemented with adenine hemisulfate (50 μg/mL), with or without inducer as detailed below. eVOLVER was operated in turbidostat mode, which uses feedback on OD to trigger dilutions and maintain cultures in a defined exponential-phase density range; OD is continuously measured, from which the growth rate is calculated (github.com/FYNCH-BIO/). Cultures were periodically sampled to assess reporter and synTF expression via flow cytometry.

For the experiments of Figure 5, cultures were continuously maintained between OD 0.2–0.5. Following inoculation, strains were first grown in inducer-free media for 18.5 h to stabilize cultures, and then induced with media containing 100 nM ß-estradiol (time = 0) for the remainder of the time-course.

For the activation memory experiments of Figure 6, cultures were continuously maintained between OD 0.25–0.5. Following inoculation, strains were first grown in inducer-free media for 24 h to stabilize cultures, prior to induction with 100nM ß-estradiol (time = 0). After 12 h of induction, media bottles were exchanged for inducer-free media for the remainder of the time-course.

Thermodynamic model

Model description

We constructed a simple thermodynamic model to gain insight into how cooperative assembly could be used to engineer specific regulatory connections in gene circuits, drawing on the rich history of describing transcriptional regulation by a thermodynamic treatment.[62,63,65,109] The model presented in this paper is a simplified version of our previously described model framework for cooperatively interacting synthetic transcription factors (synTFs) in yeast.[50] Note that the previously described model was intended to be a molecularly-detailed, quantitative, and predictive design tool that could be parameterized by our experiments and then used to select molecular configurations that would yield a desired logic or dynamic output. Instead, the goal of the current model is to capture the minimal features of interest that we wanted to tune/control – e.g., DNA-binding affinity, valency, and TF cooperativity – thus providing a general framework to develop quantitative intuition about the relationship between those properties in driving regulatory specificity. Below we provide a detailed description of the features and assumptions that underlie the current model, highlighting its key differences and simplifications relative to the previously described framework.

The model is composed of four key parameters: transcription factor concentration ([TF]), TF-DNA affinity (K_{TF}), TF-TF cooperativity (c), and the number of binding sites at a given locus (n). We begin by enumerating all possible TF-bound promoter configurations for n binding sites. Each promoter state is assigned a transcriptional rate (r) and a thermodynamic weight (w). The transcriptional rate for a particular state is proportional to the number of TFs bound to a promoter. For simplicity, maximum transcriptional rate for a promoter is set to 1. Transcriptional weights describe the relative free energies of each state and are computed based on the number and affinity of interactions within each state, as previously described.[50] Transcriptional output is computed by averaging the relative contributions from each TF-bound promoter state:

$$txn = \sum_i r_i \cdot w_i \Big/ \sum_j w_j$$

where i are transcriptionally active states and j are all promoter states.

To model cooperative synTF assembly, we include a promoter state weighted by an additional cooperativity (c) term. In our study, this represents the additional free energy contribution by the clamp molecule on fully bound promoters. However, unlike in our previous model framework, we do not account for the concentration of the clamp in the cell, nor do we explicitly enumerate all possible clamp-bound assembly states. Instead, we chose to capture and vary TF cooperativity through the single term, c, exclusively accounting for it when all TFs are bound on the promoter. This choice is justified by our previous work, in which we obtained cooperativity (c) values by fitting a thermodynamic model to experimental data collected from yeast cells in which clamp-mediated TF assemblies of different sizes (n=2, 3, or 4) drove transcription of a fluorescent reporter.[50] The resulting fit predicted that TF cooperativity increases by ~75-fold for every additional TF bound to a promoter. Thus, it is reasonable to only consider TF cooperativity on a

fully occupied promoter because the transcriptional contribution from that state will be orders of magnitude larger than that of partially bound promoters. The result is a model framework with a simplified description of cooperativity captured through the single term, c, which importantly allows us to derive general principles for the relationship between cooperativity and regulatory specificity. Because our accounting for cooperativity is general, future work could integrate more molecularly-detailed aspects of clamp-based cooperativity or even model other reported mechanisms of TF cooperativity[13,110,111] to inform optimal circuit design that minimizes regulatory crosstalk. Finally, we assume that cooperativity does not affect the maximum rate of transcription at a promoter. Total transcription is calculated as described above. An example calculation of transcription at n=1 binding site is shown in Figure S2A and transcription at n=4 binding sites with cooperativity is shown in Figure S2B.

Regulatory Specificity

We extended our model to investigate how biophysical parameters governing synTF assembly could be used to design specific regulatory connections. The model considers the simplified case of a TF that can interact with binding sites at a target synthetic (SYN) locus and an 'off-target' native (NAT) locus (Figures 2A and S2). To model different synTF assembly sizes, we considered a SYN locus with binding site clusters of n = 3 – 5; to model the spurious appearance of a CRM in the genome, we considered a NAT locus with n = 1 binding site.

We defined a regulatory specificity score as the difference between transcriptional output at the SYN (txn_{SYN}) and NAT (txn_{NAT}) loci (Figure S2C). Using the thermodynamic model, we computed regulatory specificity scores across a range of DNA affinities (K_{TF} = 10^{-2} – 10^{2} μM), cooperativities (c = 0 – 20 k_BT), and SYN binding site numbers (n = 3 – 5). For simplicity, TF concentration was set to 1 μM for all simulations. We repeated this analysis for different formulations of the regulatory specificity score (Figure S2D).

All MATLAB code associated with this model is publicly available (see Resource Availability).

Population genetics model

We developed a population genetics model to explain the observed fitness dynamics in the eVOLVER continuous culture experiments. Generally, the dynamics of mutant progenies in adapting populations are shaped by both deterministic (e.g., natural selection) and stochastic forces (e.g., demographic fluctuations). It can be shown that the dynamics of a mutant progeny will be dominated by fluctuations when the population size is less than the inverse selective advantage (defined as the normalized fitness difference between the mutant and functional population).[112] In our experiments, a new mutant cell will obtain a fitness advantage on a time-scale comparable to the doubling time. As a result, the mutant population will grow deterministically after about one doubling and we can safely neglect demographic fluctuations (that is, fluctuations caused by finite cell numbers). To this end, we use an Ordinary Differential Equation (ODE) model to describe the population genetics.

In our model, we assume that cells grow at a rate $F(z)$ where z is the concentration of a synTF. Before a synTF is induced, cells double approximately every 1.5 hrs (λ). After induction, cells pay a fitness penalty proportional to the fraction of 'off-target' NAT sites that are occupied by synTFs. We model off-target binding using our thermodynamic framework for n = 1 binding site, as before. This leads to the fitness function:

$$F(z) = \lambda\left(1 - \beta\frac{z}{K_z+z}\right)$$

where β is the maximum fitness cost imposed by a synTF.

In our thermodynamic framework (Figure S2), K_{TF} can range from 10^{-2} to 10^{2} μM, but we work in units of TFs per cell. If a typical yeast cell is about 10^{-15} L, this translates to a range of 10^{0} to 10^{4} for K_z. When induced, functional cells produce synTF at a rate α (per cell). Cells can mutate the transcriptional circuit at rate μ per unit time. We assume that the mutation rate does not depend on the doubling time. We know that the mutation rate per generation is roughly 3.5×10^{-10}.[113] If we assume that there are ~100s of potential mutants that can break the circuit, then the per hour rate to get a circuit-breaking mutation in a single lineage is 2.53×10^{-8}.

Since there are roughly 10^{8} cells in the population, the average time to see a mutation is on the order of 1 h. Since cells are grown in our continuous culture experiments for 18 h prior to induction, it is reasonable to expect mutants in the population at the time of induction. However, prior to induction the mutations are nearly neutral (they incur no fitness benefit) and therefore the size of the mutant lineages will be determined solely by stochastic fluctuations. Standard theory dictates that if a mutant colony survives until the time of induction, its size will be on the order of the number of generations between the mutation and induction. This will be on the order of 100 cells, which we take as the initial mutant clone size. Using this order-of-magnitude estimate will be sufficient for our purposes, since we are ultimately interested in predicting qualitative features of the dynamics.

We model the number of functional (x_f) and mutant (x_m) cells in a growing population as:

$$\frac{d}{dt}x_f = F(z_f)x_f - \mu x_f$$

$$\frac{d}{dt}x_m = F(z_m)x_m + \mu x_f$$

Letting TF_f denote the number of synTFs in functional cells, we have

$$\frac{d}{dt}TF_f = \alpha x_f - \mu TF_f$$

The second term comes from the fact that μx_f functional cells mutate per unit time, each taking z_f=TF_f / x_f synTFs with it. Therefore, the number of synTFs in functional cells is described by the following equation:

$$\frac{d}{dt}z_f = \frac{d}{dt}\frac{TF_f}{x_f} = \alpha - \mu z_f - z_f(F(z_f) - \mu)$$

$$\alpha - F(z_f)z_f$$

The number of synTFs in mutant (broken-circuit) cells is defined by:

$$\frac{d}{dt}TF_m = \mu z_f x_f$$

This implies that the change in the number of mutant (broken-circuit) cells can be described by:

$$\frac{d}{dt}z_b = \mu z_f\frac{x_f}{x_m} - z_m\left(F(z_m) + \mu\frac{x_f}{x_m}\right)$$

$$-z_m F(z_m) + \mu\frac{x_f}{x_m}(z_f - z_m)$$

Note that we have assumed the population is growing exponentially, rather than in a finite culture. However, from the perspective of TF concentration and fitness only the species fractions, which are identical for exponentially growing and finite populations, are relevant.

We tested the qualitative features of this model by varying the maximum fitness cost (β) and synTF binding affinity (K_z) while holding all other parameters constant. Changing the maximum fitness cost (Figure S6F) is analogous to choosing a different member of our zinc finger library, with varying DNA-binding specificities that could lead to differing levels of 'OFF'-target interactions. Strains with a maximum high fitness cost show a severe growth defect after induction and are quickly out competed. This simulation result is similar to our eVOLVER continuous culture experiment using a second, high affinity zinc finger (13–16) (Figure S6C). Strains with comparably lower fitness costs are also lost over time but at a slower rate. Changing synTF binding affinity in the model is analogous to testing the high and low affinity ZF variants. As with the evolution experiments, functional cells harboring low affinity synTFs are retained for longer periods of time compared to those with high affinity synTFs.

QUANTIFICATION AND STATISTICAL ANALYSIS

FlowJo was used to extract geometric mean fluorescence values or the percentage of mTurquoise, mVenus or mRuby activated cells from flow cytometry measurements. Microsoft Excel and GraphPad Prism software were used to process data. Statistical details such as number of replicates and error calculations are provided in figure legends.

CellPress
OPEN ACCESS

Supplemental figures

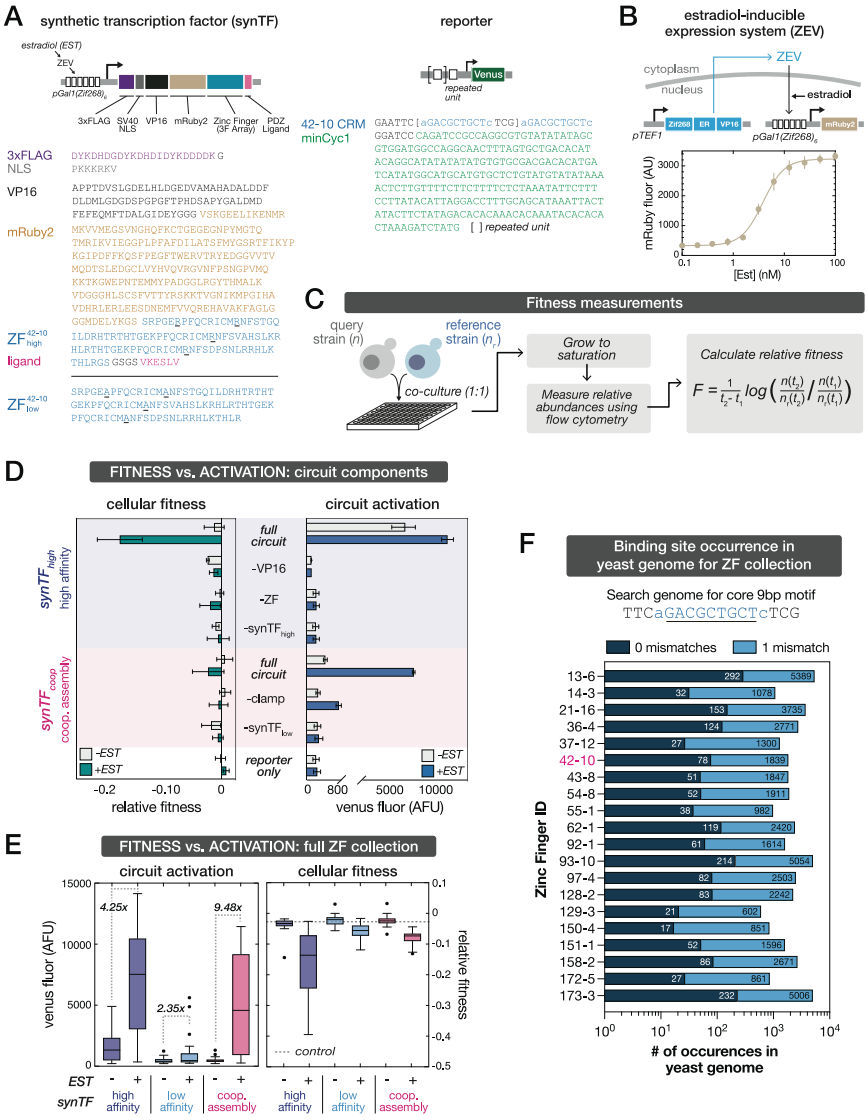


Figure S1. Design and characterization of synthetic circuits, related to Figures 1 and 2

(A) Design details of inducible synTF circuit components including amino acid sequences common to all synTFs used in this study. SynTF activators are composed of a 3-finger (3F) ZF array fused to a herpes simplex VP16 transactivation domain (TAD), 3× repeat FLAG epitope tag (Sigma), mRuby fluorescent protein, syntrophin PDZ ligand, and SV40-derived nuclear localization sequence (NLS). For zinc-finger affinity alleles, mutated arginine residues are underlined. DNA sequences for engineered synTF promoters, composed of a minimal CYC1 (minCyc1) promoter with upstream synTF binding sequences (CRMs). All sequences were yeast-optimized and chromosomally integrated into *S. cerevisiae*.

(B) Characterization of inducible expression system. mRuby2 fluorescent protein expression was used as a proxy to quantify the β-estradiol-inducible promoter. Flow cytometry measurements were made at mid-log phase, and error bars indicate standard deviation from three biological replicates. [Est], β-estradiol concentration.

(C) Workflow for the competition co-culture experiment used to quantify fitness across this study. Query strains (and associated control strains) were each co-cultured 1:1 with a reference strain (parental strain constitutively expressing mTurquoise reporter). Three biological replicates (separate colonies) of each query or control strain were measured in separate wells, and duplicate experiments were performed in media with and without 1 μM β-estradiol inducer. Cocultures were sampled when initially mixed (T0) and every 12 h to determine relative abundancies of reference vs. query or control strain, derived from the fraction of cells in each well expressing mTurquoise. Fitness measurements were equated for each query strain by calculating changes in mTurquoise expression from T0 to the experimental endpoint.

(legend continued on next page)

(D) Deletion of either the VP16 transactivation domain or the ZF DNA-binding domain is sufficient to rescue the fitness cost induced by $synTF_{high}$ circuits. Strains lacking designated circuit components were constructed, and cellular fitness and circuit activation measured as previously described. Bars represent mean values for three biological replicates ± SD (measured in two separate experiments).
(E) Cellular fitness and circuit activation for synTF circuits constructed from a collection of distinct ZF species. Fitness and activation profiles were quantified at 36 h following induction, in conditions with and without β-estradiol inducer. In this case, "control" denotes a strain with no integrated circuit components, since reporter-only strains are different for each synTF. Tukey boxplots represent the range of means for 20 ZF synTF library members.
(F) Prevalence of synthetic ZF binding sites in the yeast genome. Occurrences of the full and single mismatches of the predicted core (9 bp) binding motifs for the full ZF collection.

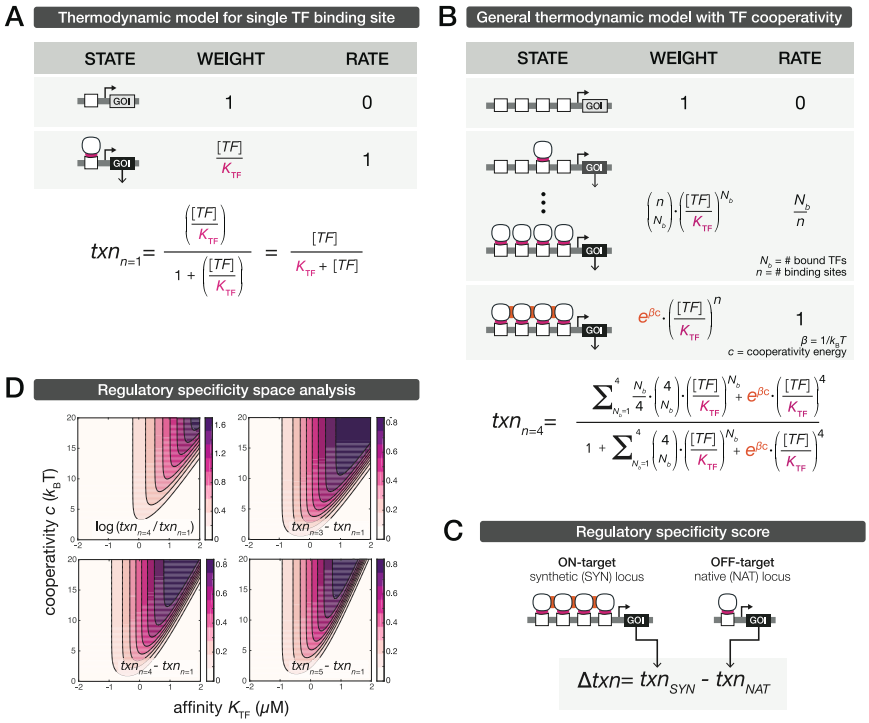


Figure S2. Description and analysis of thermodynamic model for cooperative-assembly-driven specificity, related to Figure 2

(A) Description of thermodynamic model for a single TF binding site. Gray boxes: enumeration of promoter states with corresponding thermodynamic weights and transcriptional rates (proportional to the number of bound TFs). Equation: transcriptional output ($txn_{n=1}$) is obtained by averaging the relative transcriptional contributions of all promoter states. K_{TF}, TF-DNA-binding affinity.

(B) Description of a generalized thermodynamic model incorporating TF cooperative assembly. Gray boxes: enumeration of promoter states with corresponding thermodynamic weights and transcriptional rates (proportional to the number of bound TFs). Equation: transcriptional output for n = 4 binding sites. K_{TF}, TF-DNA-binding affinity; c, cooperativity term that defines the additional stability provided by the multivalent TF interactions.

(C) Regulatory specificity score is defined as the difference between SYN on- vs. NAT off-target transcription.

(D) Regulatory specificity space is qualitatively preserved for different model formulations. Alternative regulatory specificity score with n = 4 binding sites (top left). Specificity score as described in (C) for different number of binding sites at the SYN locus: n = 3 (top right), n = 4 (bottom left), and n = 5 (bottom right).

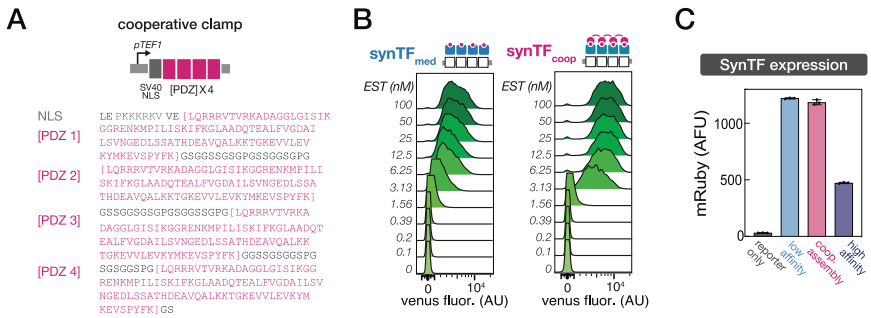


Figure S3. Design of the cooperative clamp and characterization of clamp-mediated cooperative synthetic circuits, related to Figure 2

(A) Design and sequence details of the cooperative clamp. Cooperative clamps are composed of the same SV40-derived NLS followed by repeat units of the syntrophin PDZ domain. An n = 4 clamp sequence is depicted with the repeat units highlighted. Sequences were yeast-optimized and chromosomally integrated into *S. cerevisiae*. The binding affinity of the syntrophin PDZ <> VKESLV ligand pair was experimentally determined to have a $K_d \sim 1.97$ μM.

(B) Single-cell dose response behaviors for the independent ($synTF_{med}$) and cooperative ($synTF_{coop}$) synTF circuits. Flow cytometry analysis of the dose responses show a linear shift from OFF to ON with the non-clamp synTF and a characteristically non-linear shift from OFF to ON with the clamp-mediated cooperative synTF architecture in response to increasing inducer concentrations.

(C) synTF expression levels measured by quantifying synTF-mRuby2 fluorescence following circuit induction. Bars represent mean values for three biological replicates ± SD.

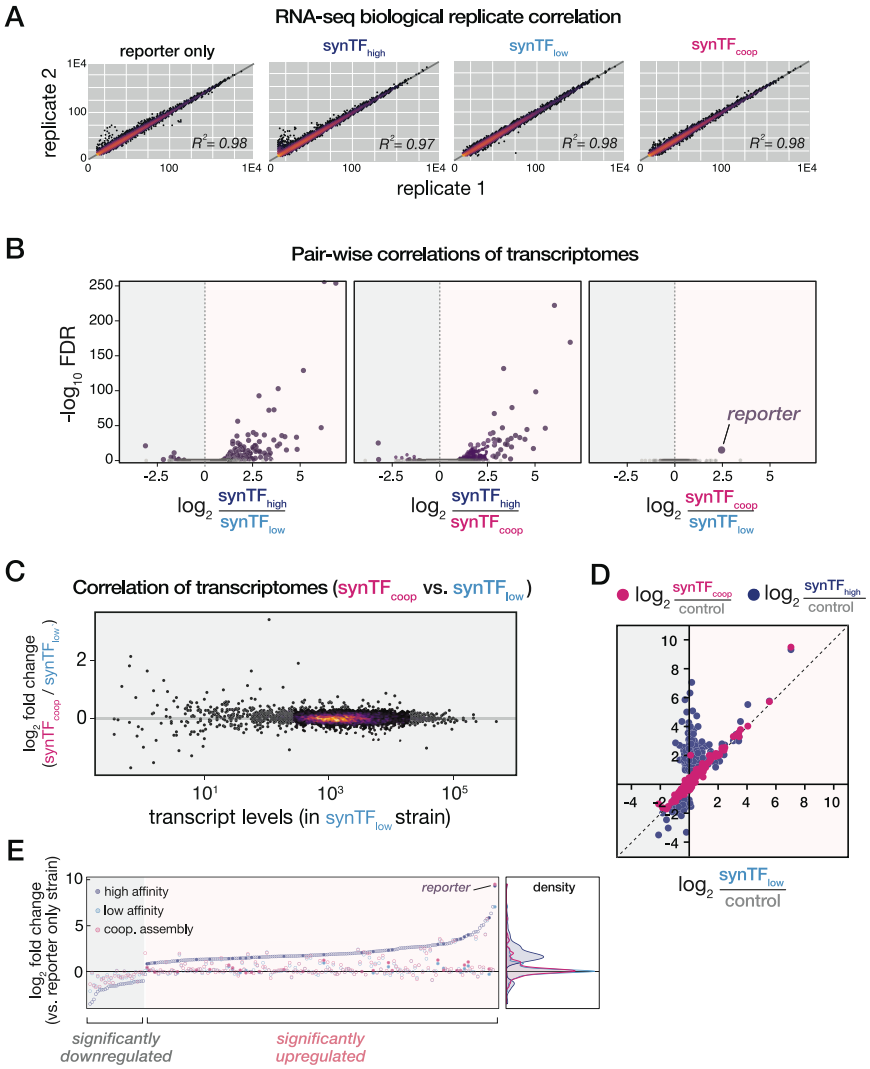


Figure S4. Correlation of transcriptomes of strains expressing synTF circuit variants, related to Figure 3

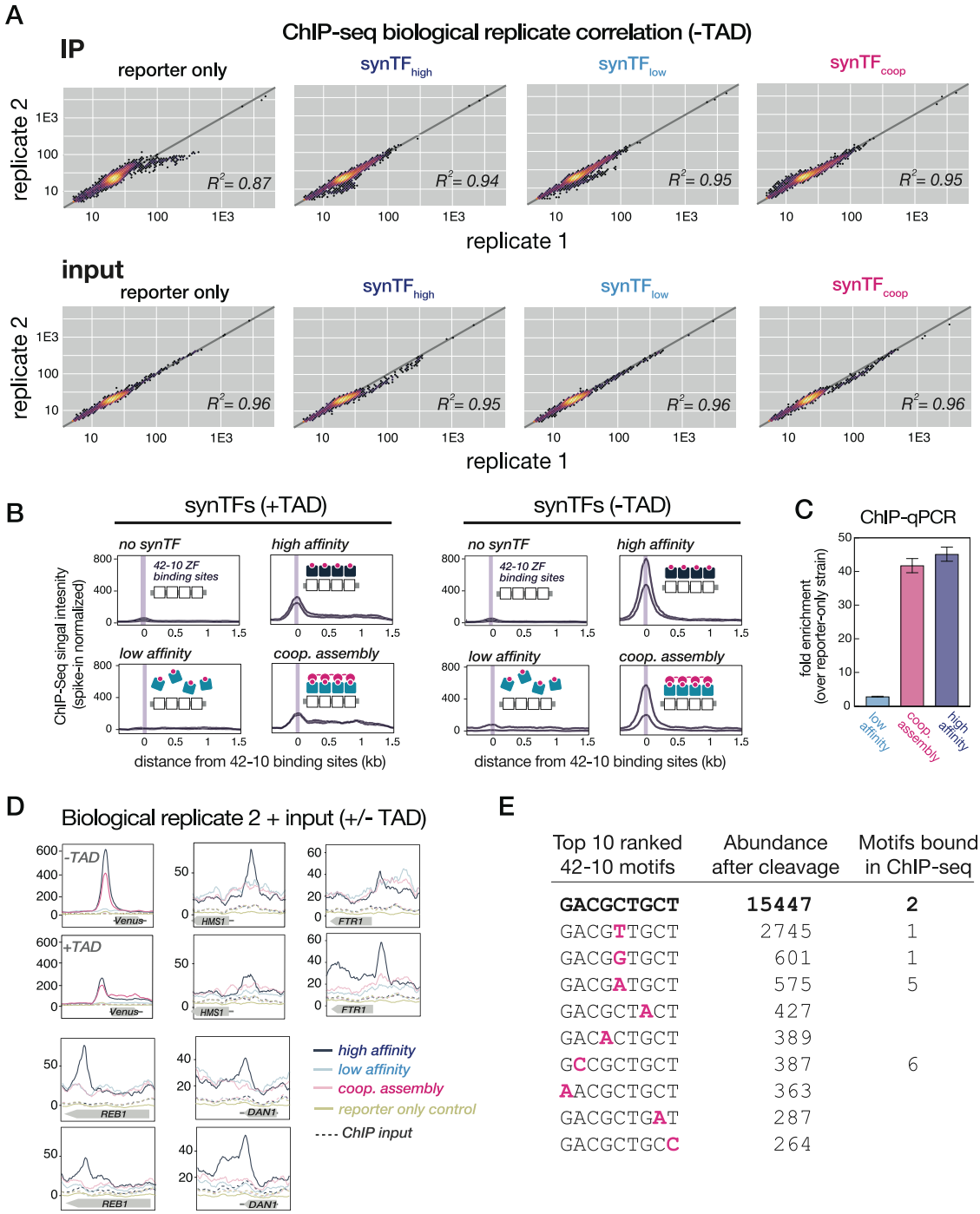
(A) Correlation of RNA-seq measurements for biological replicates.

(B) Pairwise correlations of transcriptomes for the three synTF circuit configurations. The synthetic reporter was the only gene differentially regulated by $synTF_{coop}$ vs. $synTF_{low}$. Two biological replicates for each strain are reported. FDR, false discovery rate.

(C) Transcript levels are highly correlated between strains expressing $synTF_{coop}$ and $synTF_{low}$ circuits.

(D) Differential gene expression values for $synTF_{high}$ and $synTF_{coop}$ against $synTF_{low}$. Differential gene expression for the $synTF_{coop}$ strain correlates highly with $synTF_{low}$, with one notable exception: the synthetic reporter gene, which is differentially expressed to equivalent levels as the $synTF_{high}$ strain.

(E) Differential gene expression analysis of RNA-seq measurements following induction of $synTF_{high}$, $synTF_{low}$, and $synTF_{coop}$ circuits. Plotted are genes that are significantly differentially regulated relative to the reporter-only control. The $synTF_{high}$ circuit induces a global misregulation of the host transcriptome, including significantly upregulating 182 genes. Gene expression density distributions of $synTF_{coop}$ and $synTF_{low}$ strains are highly similar to one another and cluster tightly around the reporter-only background. Filled circles represent genes with a motif located within 300-bp upstream of the TSS that has at least 8/9 bp homology to the cognate motif.



Top 10 ranked 42-10 motifs	Abundance after cleavage	Motifs bound in ChIP-seq
GACGCTGCT	**15447**	**2**
GACGTTGCT	2745	1
GACGGTGCT	601	1
GACGATGCT	575	5
GACGCTACT	427	
GACACTGCT	389	
GCCGCTGCT	387	6
AACGCTGCT	363	
GACGCTGAT	287	
GACGCTGCC	264	

Figure S5. Validation of ChIP-seq methods and synTF binding profiles, related to Figure 4

(A) Correlation of ChIP-seq measurements for biological replicates. Correlations are plotted for both immunoprecipitated (IP) and input control samples for synTFs lacking a TAD fusion. Correlation coefficients were calculated using normalized counts in non-overlapping windows over the genome.

(B) ChIP enrichment profiles at the synthetic reporter locus for no synTF, $synTF_{low}$, $synTF_{high}$, and $synTF_{coop}$ (±TAD). Binding enrichment for two biological replicates of each strain are shown. The maximum peak heights (purple line) are highly correlated with the genomic location of the ZF 42-10 binding sites. Relative ChIP enrichment was normalized to FLAG-tagged *S. pombe* spike-in DNA that was produced in parallel with the *S. cerevisiae* samples.

(C) synTF enrichment at the synthetic reporter locus measured using ChIP-quantitative PCR (ChIP-qPCR). Fold enrichment is determined for each condition compared with a reporter-only control. Enrichment patterns recapitulate those observed with ChIP-seq. Bars represent mean values for three technical replicates ± SD.

(legend continued on next page)

(D) ChIP enrichment profiles for each synTF (low, high, coop) strain, reporter-only control strain, and input control sample at five representative loci. Relative ChIP enrichment was normalized to FLAG-tagged *S. pombe* spike-in DNA that was produced in parallel with the *S. cerevisiae* samples.
(E) Top-ranked interaction motifs for ZF 42-10, as determined by an independent dataset based on an *in vitro* DNA cleavage profiling assay.[70] Abundance after cleavage quantifies the frequency that a motif has been targeted (and cleaved) by a nuclease version of our candidate ZF. The 15 synTF$_{high}$ binding events nominated by our ChIP-seq analysis occurred at the seven most preferred motifs.

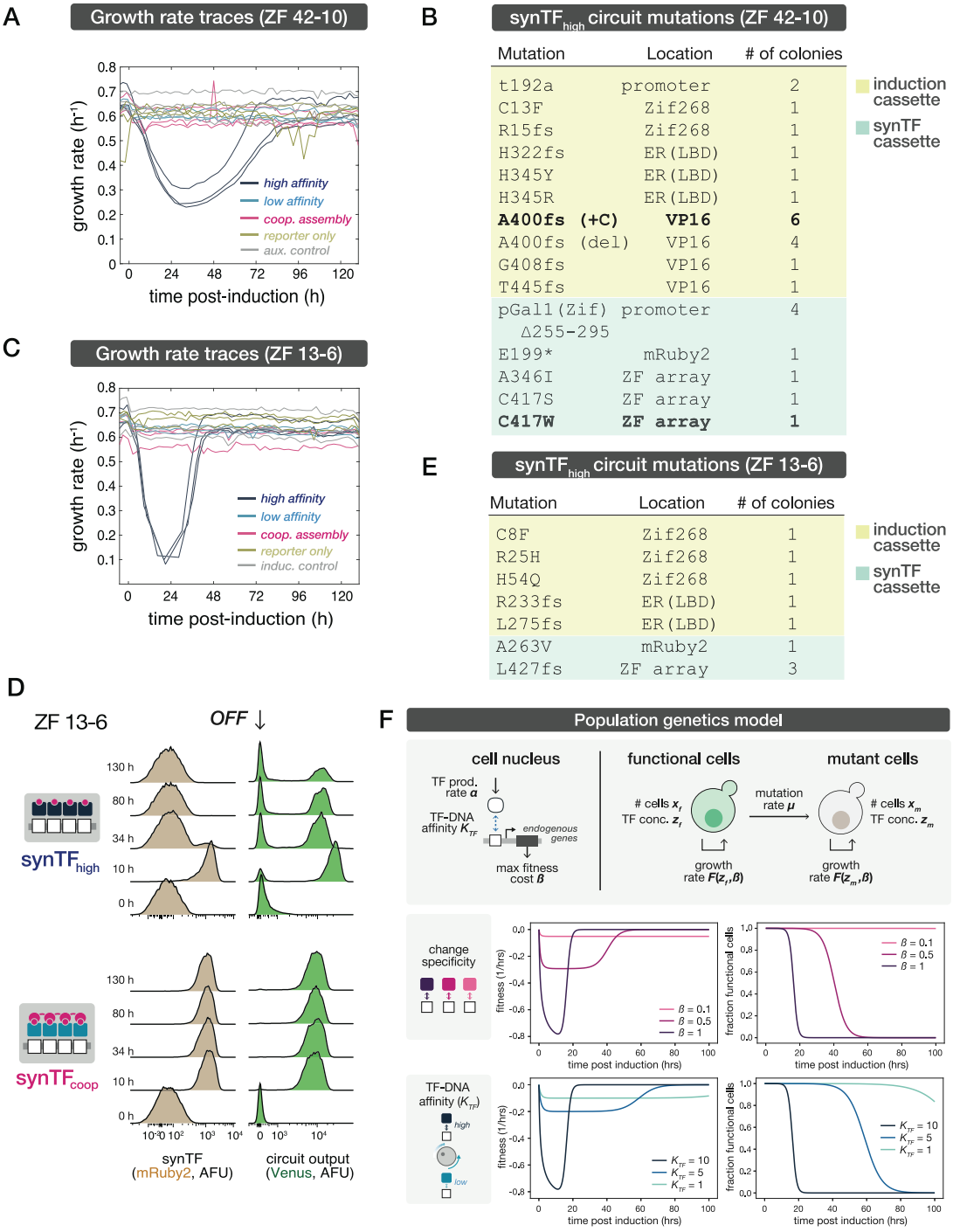


Figure S6. Growth, circuit expression, and mutational patterns are conserved in a replicate long-term culture experiment and captured by a population genetics model, related to Figure 5

(A) Raw growth rate traces for three biological replicates of synTF and control strains in long-term eVOLVER continuous culture. Each replicate was cultured in a separate eVOLVER vial. The auxiliary (aux.) control has a scrambled placeholder sequence integrated into each of the three loci into which circuit components are integrated.

(B) Mutational analysis of the circuit genotype from $synTF_{high}$-derived colonies following eVOLVER long-term culture. Mutations were identified within the induction cassette (yellow) and synTF cassette (green). A single residue in the induction cassette was highly targeted, with mutations identified in 10 of the sequenced colonies.

(legend continued on next page)

(C) Raw growth rate traces for three biological replicates of a second synTF species (ZF 13-6) and control strains in long-term eVOLVER continuous culture. Each replicate was cultured in a separate eVOLVER vial. The inducer (induc.) control has the induction cassette driving an mRuby2 fluorescent reporter in place of the same induction cassette driving the synTFs.

(D) Single-cell flow cytometry distributions of synTF and circuit reporter expression over the time course of the continuous culture experiment for a second synTF species (ZF 13-6).

(E) Mutational analysis of the circuit genotype from $synTF_{high}$-derived colonies following eVOLVER long-term culture with a second synTF species (ZF 13-6).

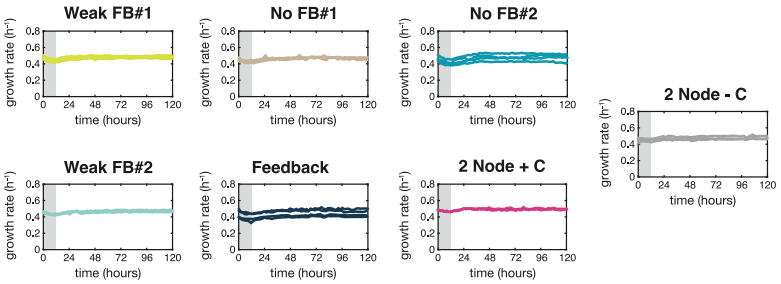
(F) A population genetics model captures the population fitness and circuit retention dynamics observed in long-term culture experiments. Description of the model (top). A TF is produced at a constant rate (α) and has an affinity for DNA (K_{TF}), which is proportional to the maximum fitness cost (β) it imposes on a host cell. The number of functional and mutant cells in a population is defined by x_f and x_m, respectively. The concentration of TF in each cell type is defined by z_f and z_m, respectively. Functional cells are converted to mutant cells at a constant rate μ. The growth rate of each population (F) is a function of the concentration and maximum fitness cost of each TF. Population fitness (middle left) and circuit retention (middle right) dynamics for a range of TF fitness cost values (β), where $K_{TF} = 1$. Population fitness (bottom left) and circuit retention (bottom right) dynamics for a range of TF-DNA-binding affinities (K_{TF}), where $\beta = 1$.

A Feedback circuit strain descriptions

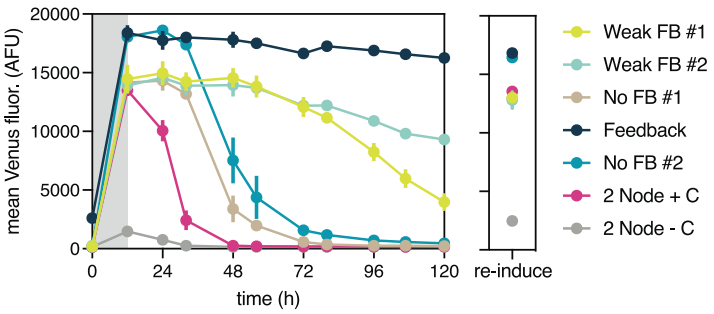
STRAIN	Node 2 promoter	42-10 SynTF PDZ	Clamp Valency
Weak FB #1	(42-10 CRM1)x4, (43-8 CRM1)x3	VKESLV	4
Weak FB #2	(42-10 CRM1)x5, (43-8 CRM1)x3	VKESLV	4
No FB #1	(43-8 CRM1)x3	VKESLV	4
Feedback	(42-10 CRM2)x5, (43-8 CRM1)x3	IRETIL	5
No FB #2*	(42-10 CRM2)x5, (43-8 CRM1)x3	IRETIL	5
2 Node + C*	-	VKESLV	4
2 Node - C	-	VKESLV	4

* highlighted in Figure 6

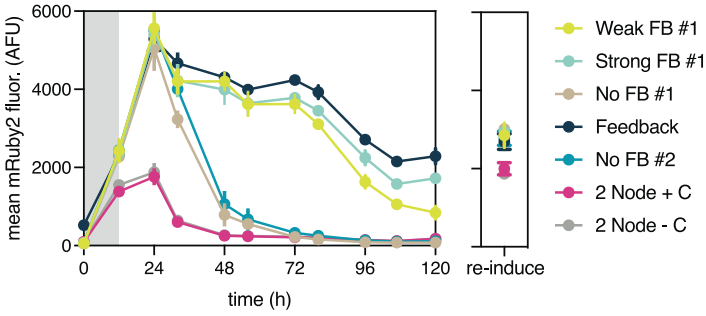
B Growth rate traces for Feedback circuits



C Circuit output for Feedback circuits



D SynTF Expression for Feedback circuits



(legend on next page)

Figure S7. Design of 3-node positive feedback circuit and testing of design variants, related to Figure 6

(A) Design details of positive feedback strains. The table lists synTF-CRM architecture at the autoregulatory node ("node 2 promoter"), PDZ ligand fused to the second synTF (ZF 42-10), and clamp valency. The first synTF species ("trigger") uses 4X mut ZF 43-8 ($K_d \sim 13.6$ nM); the second synTF species uses 4X mut ZF 42-10 ($K_d \sim 15$ nM). The clamp uses syntrophin PDZ domains: PDZ <> VKESLV ligand $K_d \sim 1.9$ μM; PDZ <> IRETIL $K_d \sim 0.18$ μM.

(B) Growth rate traces for feedback circuit variants throughout the eVOLVER continuous culture experiment.

(C) Reporter circuit output for feedback circuit variants throughout the eVOLVER continuous culture experiment. Samples were measured by flow cytometry. Points represent mean values for four biological replicates ± SD.

(D) Autoregulated synTF concentration for feedback circuit variants throughout the eVOLVER continuous culture experiment. Samples were measured by flow cytometry. Points represent mean values for four biological replicates ± SD.