An endoribonuclease-based incoherent feedforward loop for decoupling resource-limited genetic modules

Ross D. Jones[1, 2 *], Yili Qian[2, 3 *], Breanna DiAndreth[1, 2 *], Velia Siciliano[1, 2 †], Jin Huh[1, 2], Ron Weiss[1, 2, 4 ‡], and Domitilla Del Vecchio[2, 3 ‡]

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

[2]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

[3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

[4]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA.

[*]These authors contributed equally to this work.

[†]Current address: Instituto Italiano di Tecnologia, Napoli, NA, 80125, Italy.

[‡]Correspondence should be addressed to R.W. (rweiss@mit.edu) or D.D.V. (ddv@mit.edu).

Abstract

A significant goal of synthetic biology is to develop genetic devices for accurate and robust control of gene expression. Lack of modularity, wherein a device output does not depend uniquely on its intended inputs but also on its context, leads to poorly predictable device behavior. One contributor to lack of modularity is competition for shared limited gene expression resources, which can induce 'coupling' between otherwise independently-regulated genes. Here we quantify the effects of resource competition on engineered genetic systems in mammalian cells and develop a feedfoward controller to make gene expression robust to changes in resource availability. In addition to mitigating resource competition, our feedforward controller also enables adaptation to multiple log-orders of DNA copy number variation and is predictably tunable with upstream open reading frames. Our resource competition characterization along with the feedforward control device will be critical for achieving robust and accurate control of gene expression.

1 Introduction

A promising strategy for engineering complex genetic devices is to compose together simpler systems that have been characterized in isolation[1–3]. A critical assumption of this modular design approach is that subsystems maintain their input/output (i/o) behavior when assembled into larger systems. However, this assumption often fails due to context dependence, *i.e.*, the behavior of a module depends on the surrounding systems[2,4]. There are many sources of context-dependence, including unexpected off-target interactions between regulators and promoters[5], transcription factor loading by DNA targets[6], gene orientation[7], and resource loading by expressed genes[8,9]. To date, much effort has gone into identifying gene regulators with unique binding specificity, *e.g.* between transcription factors (TFs) and their DNA binding sites. Unique binding specificity enables gene regulators to work orthogonally, since they do not directly interfere with each other's binding and regulation[5]. Nevertheless, even if subsystems are entirely composed of orthogonal regulators, they can become coupled with each other via competition for shared cellular resources[2,8–11]. For example, it has been demonstrated in prokaryotes that genes compete for the usage of ribosomes, such that increased expression from one gene decreases expression from others by sequestering (*i.e.* loading) the ribosome[8,9]. Little work has been done to understand how resource competition affects engineered genetic devices in eukaryotic cells. Furthermore, while solutions to the ribosome resource competition problem in bacterial cells have appeared recently[12–14], solutions to resource competition in mammalian cells are still missing.

In mammalian cells, loading of several types of cellular resources shared among multiple genes has been shown to affect gene expression, including splicing factors[15], miRNA processing factors[16], RISC complexes[17,18], and the proteasome[19]. A potent form of resource competition called 'squelching' occurs when transcriptional activators (TAs) or strong promoters sequester transcription coactivator proteins (CoAs) and/or general TFs (GTFs), reducing transcription of other genes[20–28]. At sufficiently high expression levels of a given TA, these transcriptional resources are sequestered even from the TA molecules bound to the target promoter, yielding a bell-like dose-response curve, where the expression of the TA's target gene peaks at an intermediate level of TA and then decreases as the TA concentration is further increased (often referred to as 'self-squelching')[22,24]. As many established synthetic eukaryotic gene regulation systems utilize TAs, squelching represents a potentially pervasive problem in the space of eukaryotic genetic engineering. Here, we consider competition for gene expression resources as a general problem, investigate the quantitative consequences of resource competition on mammalian genetic circuits, and introduce an engineering solution.

We first developed an experimental model system to recapitulate the effects of transcriptional resource competition and provide in-depth characterization of these effects. We then used this model system to evaluate the performance of a feedforward controller designed to cancel the effects of resource loading on gene expression. Specifically, our model system utilizes Gal4 TAs of varying strength to measure the extent to which TA expression sequesters

transcriptional resources from non-target genes. We developed a mathematical model that explains the effects of transcriptional resource sequestration by TAs on expression of target and non-target genes of the TA. We then measured the effect of different Gal4 TAs on commonly-used constitutive promoters in various mammalian cell lines to identify combinations of activators and promoters in each cell line where minimal effect on the non-target promoters is observed. Overall, our results provide extensive analysis for determining the extent to which transcriptional resource competition affects gene circuit behavior in mammalian cells.

Our ultimate goal is to make the output protein level of a given genetic device insensitive to changes in available gene expression resources, including CoAs and GTFs. By doing so, we can effectively decouple the behavior of resource-coupled genetic devices. To approach this problem, we regard resource availability as a disturbance input to a genetic device and design a controller that can be added to any device to 'cancel out' the effect of resource competition on the device's output. In prokaryotes, it has been shown that quasi-integral feedback control can make the output protein level of a genetic device insensitive to changes in ribosome availability[13]. In both prokaryotes and eukaryotes, incoherent feedforward loops (iFFLs) have been used to make gene expression levels insensitive to the copy number of a gene[29,30]. Here, we engineered an iFFL using CasE, an endoribonuclease (endoRNase) from a type I CRISPR system, to make a genetic device's output insensitive to changes in the availability of transcriptional resources (Figure 1a-d). Through experiments in mammalian cells, we show that iFFLs can make the output protein level of a genetic device insensitive to variations in availability of gene expression resources. We further found that our iFFL design performs well in combination with different activators and cell lines, demonstrating that our solution is general and applicable to a variety of contexts. Beyond resource competition, our iFFL design also makes a genetic device's output insensitive to multiple log-order changes in gene copy number, substantially improving upon previously published miRNA-based designs[29,31]. In addition, the iFFL reduces the dynamic effects of plasmid uptake and dilution on protein expression during transient transfection, thereby broadening the time window over which stable expression levels can be achieved. Overall, our iFFL design will find broad utility for engineering mammalian genetic devices which behave as predicted in a context-independent manner.

2 Results

2.1 Characterization of transcriptional resource competition

Cells provide a finite pool of resources for gene expression. To express any one gene, the cell must allocate resources to this gene, thereby reducing the availability of resources to other genes. Here, we consider the effect of this reduced availability of resources on the output of a genetic device. Specifically, we define a genetic device as a system composed of one gene that takes regulatory inputs (*e.g.* sequence-specific TFs) and gives the gene's expressed

RNA and/or protein as output. We further define a genetic module as one or more genetic devices that are linked together by regulatory interactions. Experimentally, we include fluorescent markers to measure the expression level of non-fluorescent proteins in some modules. Independently-regulated devices become implicitly coupled by competition for gene expression resources, wherein expression of a gene in one device 'loads' the pool of resources, thereby decreasing resource availability to other devices (Figure 1a). Because of this coupling, the i/o behavior of a genetic device or module becomes dependent on the presence of other devices and modules in the cell.

Previous studies have shown that competition for transcriptional resources including coactivators (CoAs) and general transcription factors (GTFs) can reduce gene expression levels[32]. Transcriptional activators (TAs) in eukaryotes are comprised of a DNA-binding domain (DBD) and an activation domain (AD), the latter of which recruits CoAs and/or GTFs to initiate transcription[33]. When a given TA is in excess, the binding between the TA and CoAs/GTFs in solution and at off-target DNA loci can form unproductive complexes that sequester these factors[24], a phenomenon referred to as squelching[20]. Importantly, ADs alone, without a DBD, can also cause squelching[23].

We recapitulated competition for transcriptional resources by different genetic devices using the genetic model system shown in Figure 1e. The Gal4 DBD was fused to several ADs of varying potency (Supplementary Figure 1), of which five were chosen for in-depth study: VP16, VPR, and the individual components of VPR (VP64, Rta, and p65). Our model system comprises two genetic modules, each with one or more genetic devices: (i) a device with a constitutive gene: CMV:Output$_1$ and (ii) Gal4 TA expression: hEF1a:Gal4-AD and a Gal4-activated gene: UAS:Output$_2$ (Figure 1e). To more precisely measure the delivery of Gal4 TAs to each cell, we included a fluorescent reporter (Gal4 Marker) that was co-titrated with the Gal4 TAs prior to transfection of all the plasmids into HEK-293FT cells. The resulting dose-response curves for knockdown of CMV:Output$_1$ and activation of UAS:Output$_2$ are shown in Figure 1f and Supplementary Figure 2, respectively. Full distributions for samples with varying Gal4 TA input levels and a comparison of the DNA input amount vs Gal4 Marker fluorescence are shown in Supplementary Figure 3. At the highest dosage tested, all five Gal4 TAs knocked down CMV:Output$_1$ by at least 2-fold, with Gal4-VPR causing nearly 8-fold knockdown (Figure 1f). Each curve was similar in shape, with the main difference being the amount of Gal4 TA needed to reduce CMV-driven expression by half, which varied by over 20-fold between Gal4-VP64 and Gal4-VPR.

Resource sequestration due to addition of Gal4 TAs can occur at different stages of gene expression: (a) the expression of Gal4 itself requires both transcriptional and translational resources, (b) the action of Gal4 activating its target causes additional sequestration of both types of resources due to expression of the target gene, and (c) Gal4 directly binds to and sequesters transcriptional resources in solution and/or at off-target DNA loci[24]. We validated that the Gal4 TAs repress CMV transcription by RT-qPCR measurement of CMV-driven mRNA levels (Supplementary Figure 4a-b). Indeed, CMV-driven mRNA levels were knocked down ~2-fold by Gal4-VP16 and ~16-fold by Gal4-VPR (Supplementary Figure 4b). In the same samples, a fraction of the cells were collected for

flow cytometry to measure protein expression levels. The magnitude of knockdown of protein levels closely matched that of the mRNA levels (Supplementary Figure 4c), suggesting that most of the knockdown was caused at the transcriptional level. Additional experiments showed that VPR alone and Gal4-VPR both knock down CMV expression, but neither the Gal4 DBD nor the luminescent protein Fluc2 do so (Supplementary Figure 4d). Because these proteins were expressed by the same promoter and thus place similar demands on gene expression resources, we concluded that (a) is negligible compared to (b) and (c). Furthermore, the knockdown of CMV expression by Gal4-VPR was similar regardless of whether the Gal4 target gene was present, indicating that in this system, (b) is small compared to (c). Thus, the AD, whether fused or not to the TA (Gal4), sequesters transcriptional resources from the CMV promoter and is the major player in the observed knock down of the CMV output expression.

We also characterized the dose-response curves of the Gal4 TAs activating the target gene (UAS:Output_2). Consistent with prior studies[22,24], the activation dose-response curve of some activators (Gal4-Rta, Gal4-p65, and Gal4-VPR) clearly showed decreasing output at high dosages of the activators (Supplementary Figure 2a-b). To provide a tool for predicting the effects of resource loading on gene circuit behavior, we developed a mathematical model, described in detail in Supplementary Note 1 and Supplementary Figure 5. This model can recapitulate both non-target gene knockdown (Figure 1f) and on-target self-squelching behavior by a TA (Supplementary Figure 2b). Interestingly, the relative UAS:Output_2 between each activator was strongly dose-dependent; for example, Gal4-p65 drove ~6-fold higher expression than Gal4-VP64 at the lowest DNA dosage, whereas Gal4-VP64 drove nearly 2-fold higher expression than Gal4-p65 at the highest DNA dosage (Supplementary Figure 3). For further discussion of this phenomenon, model fitting, and model validation, see Supplementary Note 2 and Supplementary Figures 6 & 7. From prior work, it is unclear whether the minimum concentration of TA necessary for maximal activation of on-target genes is sufficient to knock down non-target genes. We thus compared the expression of CMV:Output_1 to UAS:Output_2 at each level of each Gal4 TA to measure the trade-off in expression of both genes. We found that for each Gal4 TA, maximum UAS:Output_2 expression occurred at a concentration of Gal4 that knocked down CMV:Output_1 by at least 2-fold (Supplementary Figure 2c). Overall, these results indicate that for TAs to drive high levels of expression, significant knockdown of non-target genes is likely to be observed.

Finally, we validated that our results measured in transient transfection were consistent with the behavior of genetic devices integrated into the genome. To do so, we integrated into HEK-293FT cells one of two lentiviral constructs: (i) an rtTA activator and an rtTA-driven fluorescent reporter (Supplementary Figure 8) or (ii) a Gal4-driven fluorescent reporter (Supplementary Figure 9). Following lentiviral integration, we transfected both cell lines with Gal4 activators and found that both rtTA- and Gal4-driven expression were negatively affected by Gal4 activators at high activator dosages. The responses of the non-target tet-on system and on-target promoter were both well-predicted by the model fits from our transfection experiments (Supplementary Note 3). Thus, our resource competition results are extensible to genes located in various contexts.

2.2 Activator and non-target promoter combinations with minimal coupling

We extended the genetic model system of Figure 1e to a library of such systems with varying non-target promoters ({P}:Output$_1$) in Module 1 and varying ADs fused to Gal4 in Module 2, then transfected each combination of variants into different cell lines (Figure 2a). Figure 2b shows the nominal expression level of {P}:Output$_1$ in Module 1 for each combination of promoter and cell line tested. The nominal expression was measured in samples co-transfected with Gal4-None, which lacks an AD and thus does not appreciably load transcriptional resources (Supplementary Figures 1 & 4). Across this set of cell lines, the hEF1a promoter showed the most consistent nominal expression and the CMV-intron (CMVi) promoter generally showed the highest nominal expression level. The relative nominal expression level of each promoter was generally well-correlated between cell lines, with Vero cells showing higher expression overall (Supplementary Figure 10d). Figure 2c shows the fold-changes in {P}:Output$_1$ expression in each cell line when co-transfected with different Gal4 TAs, with the AD fused to Gal4 in each set of samples shown along the bottom of the heatmap. Decreased expression of Output$_1$ was observed in the majority of combinations, with the strongest knockdown of Output$_1$ observed for the CMV-based, RSV, and SV40 promoters, as well as in CHO-K1 and Vero 2.2 cells broadly. Some promoters had slightly increased expression in the presence of the Gal4 TAs. Possible explanations for these increases as well as how flow cytometry data analysis affects measurements in light of resource competition are discussed in Supplementary Note 4 & Supplementary Figure 11.

From the Output$_1$ fold-changes in Figure 2c, we can extract patterns that help guide design choices for specific combinations of promoters and TAs that minimize coupling between modules due to resource competition. Comparing the Gal4 TAs, we saw that across cell lines, Gal4-VP16 and -VP64 had relatively weak effects and Gal4-Rta, -p65, and -VPR had relatively strong effects. Gal4-VP64 and -VP16 caused less than a 20% change in expression of {P}:Output$_1$ on average across all promoters and cell lines, compared to a 30-40% reduction on average by Gal4-Rta, -p65, and -VPR (Supplementary Figure 12a). At the dosage of activators we tested, Gal4-VP64 and -VP16 also tended to give the highest level of UAS:Output$_2$ (Supplementary Figure 10).

Overall, the effects of Gal4 TAs on a given constitutive promoter were mildly correlated between cell lines, (Supplementary Figure 12b), with CHO-K1 and Vero 2.2 cells experiencing repressive effects across nearly all promoters (Supplementary Figure 12c). However, there were cases where different promoters were affected more or less strongly in different cell types. For example, in both HEK cell lines, the hEF1a promoter was less affected by Gal4 TA competition than the CMV promoter, whereas in CHO-K1 cells the opposite was seen (Figure 2c). The Gal4 TA with the strongest negative effect on a given promoter was also not necessarily the same between cell lines. For instance, Gal4-VPR typically showed the strongest knockdown in HEK cells, whereas Gal4-Rta and Gal4-p65 did so in HeLa and CHO cells, respectively. Thus, while many patterns were preserved between cell lines, the effects of resource competition in one cell line do not necessarily predict the effects in others.

The constitutive promoters we tested varied more than 2 orders of magnitude in strength and originated from both viral and human DNA. Some promoters drove expression that was nearly undetectable (Supplementary Figure 13), which we partially adjusted for by subtracting the median of the untransfected cells from that of the transfected cells prior to computing fold-changes (see Supplementary Note 4 & Supplementary Figure 14). In general, viral promoters in Module 1 were more strongly affected by resource competition than human promoters (Supplementary Figure 15a), suggesting that they have lower affinity for GTFs/CoAs and/or that they utilize a relatively high fraction of specific CoAs that are sequestered by the Gal4 TAs. Heirarchical clustering of the fold-changes in each promoter's expression identified that in our set, human promoters are more similarly affected by Gal4 TAs than viral promoters, possibly owing to their similar CpG-island based architectures (Supplementary Figure 15b-c).

While we saw widespread reductions and in some cases increases in Output_1 in response to the Gal4 TAs, there were some combinations of promoters and Gal4 TAs in each cell line that had little to no effect. The five promoter-TA combinations with either the least effect on or strongest knockdown of Output_1 are reported in Figure 2c (see Supplementary Figure 16 for all combinations). In HEK-293 and HEK-293FT cells, hPGK and the hUBC promoter variants were weakly affected in combination with most Gal4 TAs, together making up 8/10 of the most unchanged combinations across both cell lines. In HeLa cells, hEF1a and the hUBC promoter variants were generally the least-affected by competition; note that the TK promoter in HeLa was omitted from this analysis because its nominal expression level was undetectable (Supplementary Figure 13). In CHO-K1 and Vero 2.2 cells, RSV/hMDM2 variants and TK/hUBC were least affected by resource competition, respectively. Consistent with the general effects on promoters described above and in Supplementary Figure 15, the combinations of promoters and TAs with the strongest negative effects were nearly all with viral promoters. However, there were some notable exceptions; in particular, the hEF1a promoter was the most strongly knocked down promoter in CHO-K1 cells.

So far we have focused on how coupling between the constitutive promoters (Module 1) and the Gal4 TAs (Module 2) affects expression of the promoters (Output_1); however, this coupling can also work in reverse such that Gal4-driven expression (Ouptut_2) is affected by Module 1. We thus examined the effects of resource sequestration by the constitutive reporters on Gal4-driven activation of UAS:Output_2. We found that the expression UAS:Output_2 was largely the same between samples with the same Gal4 TA but different promoters (Supplementary Figure 10), but with notable exceptions in each cell line (Supplementary Figure 17). For example, samples with the CMV or hEF1a promoters generally had higher or lower UAS:Output_2 expression, respectively. In addition, samples with the hMDM2c promoter showed reduced UAS:Output_2 expression in CHO-K1 cells, and samples with hUBC promoter variants showed increased UAS:Output_2 expression in Vero 2.2 cells.

While we have assumed thus far that effects from resource competition are entirely derived from the AD of a TA such as Gal4, it has been shown that reducing the strength of or eliminating DNA-DBD binding may relieve the effects of transcriptional resource loading by TAs[27]. We therefore compared the effect that VPR alone, Gal4-VPR,

and VPR fused to a zinc finger protein (ZFP), dCas9, and rTetR had on expression of each of the constitutive promoters tested above. We found that the effects on each promoter followed a similar trend, with rTetR-VPR showing the most similar effects to Gal4-VPR, dCas9-VPR showing the least strong effects, and ZFP-VPR being the only variant to not knock down the CMV-based promoters (Supplementary Figure 18a-c). The effects on each constitutive promoter were similar for dCas9-VPR +/- gRNA as well as rTetR-VPR +/- Dox (Supplementary Figure 18d), indicating that expression by these specific promoters did not significantly load gene expression resources and that Dox or gRNA binding are not required for rTetR-VPR or dCas9-VPR to sequester resources, respectively.

Overall, the characterization results in Figure 2c-d will guide choices of combinations of promoters and Gal4 TAs that minimize resource competition in engineered genetic circuits in the given mammalian cell lines.

2.3 Model-guided design of a resource-decoupled genetic module using an endoRNase-based iFFL

Though we have identified some combinations of promoters and TAs in commonly-used cell lines with minimal resource coupling, a more general solution to transcriptional resource competition would enable the use of any combination of promoters and TAs in any cell line. We thus designed a resource-decoupled genetic module comprised of the genetic device of interest and a feedforward controller that makes the genetic device's output insensitive to perturbations in resource availability, thereby decoupling expression of the module's output from resource usage by other modules (Figure 1b-c). The feedforward controller manifests as an iFFL: an endoRNase protein is expressed from the identical promoter as the output, then binds and cuts a specific target site in the 5'UTR of the output mRNA, leading to its degradation and preventing translation of the output protein. Using a simple mathematical model of the iFFL module with gene expression resources as an input, we predicted that resource loading would proportionally affect both the endoRNase and the output expression, such that changes in endoRNase levels could offset changes in output production and make the output expression insensitive to resource loading (Figure 1d). We have recently developed a toolkit for creating post-transcriptional genetic circuits using Cas6- and Cas13-family CRISPR endoRNases (DiAndreth *et al.*, manuscript in preparation). These endoRNases show very strong repression of protein expression (~50- to 250-fold knockdown) and good orthogonality in terms of targeting specific RNA hairpins. To implement our iFFL design, we utilized one of the stronger endoRNases in this toolkit, CasE[34]. The endoRNase and output mRNAs are transcribed as separate RNA species but from identical promoters (P) on the same strand of DNA with no insulator in between, helping to couple their transcriptional inputs and kinetics[35]. We placed the CasE target sites in the 5'UTR because Cas6-family endoRNases more strongly knock down gene expression when target sites are in the 5'UTR rather than the 3'UTR[36] (DiAndreth *et al.*, manuscript in preparation).

The ability of our iFFL to make output expression level insensitive to resource availability is revealed through a

mathematical model of the iFFL. The model also predicts that the robustness of the iFFL-regulated output can be tuned through variable numbers of short upstream open reading frames (uORFs)[37] in the 5'UTR of the endoRNase transcription unit. With reference to Figure 3a, the iFFL module consists of an endoRNase (x) that targets the mRNA m_y of the output protein (y) for cleavage. The two proteins are encoded on the same DNA plasmid and driven by identical promoters. This ensures that the two genes share the same pool of transcriptional resources (i.e., CoAs). We assume that the endoRNase x enzymatically degrades the output's mRNA following Michaelis-Menten kinetics. Under these assumptions, the steady state output protein concentration can be written as (see Supplementary Note 5 for derivations):

$$y = \frac{\alpha_y \beta_y u R}{\gamma_y k \kappa_y \delta_y} \cdot \left[1 + \theta \frac{\alpha_x \beta_x u R}{\gamma_x k \kappa_x \delta_x \delta_y K_M} \right]^{-1}, \tag{1}$$

where $R := R_{TX} \cdot R_{TL}$ lumps the free concentrations of the transcriptional resource R_{TX} and the translational resource R_{TL} and u is the concentration of the DNA plasmid that encodes both genes. For i =x,y, parameter α_i is the transcription initiation rate constant of gene i; δ_i is the decay rate constant of the mRNA transcript m_i; γ_i is the decay rate constant of protein i; β_i is the translation initiation rate constant, and κ_i is the dissociation constant describing the binding between translational resource (*i.e.*, ribosome) and the mRNA transcript m_i and thus governs translation initiation. The parameter θ is the catalytic rate constant of the endoRNase cleaving m_y; K_M is the Michaelis-Menten constant describing the binding of the endoRNase with m_y, and k is the dissociation constant describing binding of transcriptional resource with the identical promoters driving the expression of both x and y. To determine how each parameter changes the iFFL module's response to variations in resource availability, we simplify the expression of y in (1) by introducing the following lumped parameters:

$$V_y := \frac{\alpha_y \beta_y}{\gamma_y k \kappa_y \delta_y}, \qquad \text{and} \qquad \epsilon := \frac{\gamma_x k \delta_x \delta_y K_M}{\alpha_x \beta_x \theta} \cdot \kappa_x, \tag{2}$$

and re-write (1) as:

$$y = V_y \cdot \frac{u \cdot R}{1 + u \cdot R/\epsilon}. \tag{3}$$

Note that by (2), for a fixed output gene, the parameter V_y is fixed and does not change with any physical parameter of the endoRNase. On the other hand, changing the physical parameters governing the production, decay, and enzymatic reactions of the endoRNase only changes the lumped parameter ϵ. According to (3), for $u \cdot R/\epsilon \gg 1$, we have $y \approx Y_{\max} := V_y \cdot \epsilon$, which is independent of R, and therefore independent of the free concentrations of both transcriptional and translational resources. This implies that if we design the parameter ϵ to be sufficiently small, the iFFL module's output can adapt to variations in resource availability.

To experimentally quantify the iFFL module's robustness to resource availability, we use the fluorescence level of a co-transfected transfection marker (TX Marker) protein z as a proxy for the free amount of resources R. This is because the steady state of z can be written as $z = V_z \cdot u \cdot R$, where V_z is a lumped parameter independent of u and R

and defined similarly to V_y in (2) (see Supplementary Note 5 for more details). This enables us to re-write y in (3) as a function of the experimentally measurable quantity z:

$$y = V_y \cdot \frac{z/V_z}{1 + z/(V_z\epsilon)}. \tag{4}$$

We thus introduce an experimentally quantifiable inverse measure of robustness, Z_{50}, which is the TX Marker's fluorescence level at which the iFFL module's output is half of its maximum value (see Figure 3b) (*i.e.* $y \geq Y_{\max}/2$ for all $z \geq Z_{50}$). By substituting $y = Y_{\max}/2$ into equation (3), we find $Z_{50} = V_z\epsilon$, implying that robustness increases as the parameter ϵ decreases.

We therefore constructed a library of resource-decoupled device modules with different ϵ parameters. To construct this library, we increased the number of uROFs (n) in the 5'UTR of the endoRNase's transcript m_x to effectively increase the dissociation constant κ_x between the ribosome and m_x[37], thus increasing ϵ. With reference to Figure 3c, the relationship between n and κ_x has been experimentally characterized in[38], where the authors measured expression of a constitutive fluorescent protein p with different numbers of uORFs in the 5'UTR of its transcript. Since the expression level of a constitutive gene is inversely proportional to the dissociation constant between ribosome and its transcript (i.e., $p \propto 1/\kappa_x$, see Supplementary Note 5), we have

$$\text{relative } \kappa_x = (\text{relative } \kappa_x)(n) := \frac{\kappa_x(n)}{\kappa_x(0)} = \frac{p(0)}{p(n)}, \tag{5}$$

where $p(n)$ and $\kappa_x(n)$ are the steady state expression of p and the dissociation constant between ribosome and protein p's mRNA transcript in the presence of n uORFs, respectively. Since we have derived from equation (4) that (i) $Y_{\max}$ and Z_{50} are both proportional to ϵ and hence proportional to κ_x and that (ii) $\kappa_x(n) = (\text{relative } \kappa_x)(n) \times \kappa_x(0)$ according to (5), our model predicts that $Y_{\max} = Y_{\max}(n)$ and $Z_{50} = Z_{50}(n)$ are both proportional to relative κ_x.

To verify this model prediction, for $n = 0, 1, 2, 4, 8$ and 12, we plot the iFFL modules' output (y) for different levels of TX Marker (z). The shape of the experimentally measured TX Marker vs output dose response curves (see Figure 3d for select samples and Figure 6b for all data) matches well with the model prediction in Figure 3b, suggesting that Z_{50} is a reasonable inverse measure of the module's robustness. We therefore fit the experimental data with (4) and evaluate the fitting function to describe $Y_{\max}$ and Z_{50} for different n in the experimental data. In Figure 3e, we plot $Y_{\max}$ and Z_{50} against the relative κ_x values listed in Figure 3c, which we excerpted from Figure. 21 in the Supplementary data of Gam *et al.*[38]. We observe that $Y_{\max}$ and Z_{50} are both linearly related to relative κ_x, indicating that our model (4) can capture the salient steady state behavior of the iFFL module. For a fixed output gene (*i.e.* given V_y), since Z_{50} and $Y_{\max}$ only depend on ϵ according to (3), our model also highlights a key design trade-off for an iFFL module: increasing maximum output $Y_{\max}$ via tuning ϵ necessarily increases Z_{50}, which indicates a decrease in robustness. The number of uORFs on endoRNase's transcript can thus serve as a convenient knob to balance this trade-off between robustness and maximum expression level. To increase $Y_{\max}$ without affecting Z_{50}, the relative

promoter copy number of the output can be increased relative to the endoRNase, as we demonstrate with poly-transfection[38] in Supplementary Figure 19.

In addition to robustness to variations in free transcriptional and translational resource concentrations, the iFFL can also attenuate the effect of DNA plasmid variation (*i.e.* changes in u) on the module's output. In fact, since u and R are clustered together in (3), our analysis on the module's robustness to R carries over directly when analyzing its robustness to u: when $uR \gg \epsilon$, we have $y \approx V_y\epsilon$ according to (3), which is independent of u. Robustness to variations in u also includes temporal variability of DNA concentration, which is present in transient transfection experiments due to dilution of DNA plasmids as cells grow and divide. As one decreases the number of uORFs in the endoRNase's transcript, our model predicts that the iFFL module becomes more robust to DNA copy number variability in the sense that it's output remains the same for a wider range of DNA copy numbers (*i.e.* smaller Z_{50}). This allows the module's output to maintain $Y_{\max}$ for a longer period of time as DNA concentration gradually decreases, a phenomenon we observed both experimentally (see Supplementary Figure 33) and numerically (see Supplementary Figure 37).

2.4 The resource-decoupled module's output is robust to resource loading by Gal4 TAs

To determine the extent to which the output expression of the iFFL design is insensitive to resource loading, we constructed the CasE-based iFFL shown in Figure 4a, which uses the CMVi promoter to drive expression of both CasE and the output. We chose the CMVi promoter because of all the combinations of activators, promoters, and cell lines we have tested, the largest fold-decrease in promoter expression was caused by Gal4-VPR on CMVi in HEK-293FT cells (Figure 2, Supplementary Figure 16). We thus transfected the CMVi iFFL plasmid along with plasmids bearing a hEF1a-driven transfection marker and hEF1a:Gal4-VPR into HEK-293FT cells to measure the response of the iFFL to resource loading. We placed 0, 1, 2, 4, 8, or 12 uORFs in front of CasE to reduce its translation rate by between 2- and 200-fold[38] and thereby proportionally tune key ϵ parameter. As a control, we made an unregulated (UR) variant of the iFFL module which replaced CasE with the luminescent protein Fluc2, and thus does not form an iFFL. To account for differences in protein expression levels between the UR and iFFL modules, we transfected cells with equimolar, 1:4, 1:16, or 1:64 dilutions of the UR plasmid relative to the iFFL plasmid in samples transfected with iFFL variants. To quantify the degree to which an iFFL or UR module is sensitive to resource loading by Gal4-VPR, we measured the fold-changes relative to nominal output (*i.e.* the median output in the absence of Gal4-VPR) and from those computed robustness scores:

$$\text{Fold-}\Delta(\text{Gal4-VPR} = x) = \frac{\text{Output}(\text{Gal4-VPR} = x)}{\text{Output}(\text{Gal4-VPR} = 0)} \tag{6}$$

$$\text{Robustness}(\text{Gal4-VPR} = x) = 100 \cdot (1 - \left|1 - \text{Fold-}\Delta(\text{Gal4-VPR} = x)\right|); \tag{7}$$

Our results from co-transfecting the iFFL and UR plasmids with increasing amounts of Gal4-VPR show that variants of the iFFL with 4 or fewer uORFs in front of CasE are significantly less affected by Gal4-VPR than the UR controls (Figure 4b-d). At the highest dosage of Gal4-VPR tested (30 ng), the output of the UR samples decreased between 2- and 3-fold, whereas the iFFL variants with 4x or 2x uORFs changed by less than 1.5-fold (Figure 4b). In terms of robustness, most UR samples ranged between 30% and 60% regardless of the nominal output level. The iFFL samples with lower nominal output (higher CasE levels obtained via fewer uORFs) showed high robustness (70-90%), as predicted from the model due to a lower ϵ (see Section 2.3). iFFL samples with higher nominal output (lower CasE level obtained via more uORFs) showed lower robustness also as expected by the model due to a larger ϵ (see Section 2.3). The samples with 12x-uORFs show comparable robustness to the UR samples. Direct comparisons of the full output distributions of UR and iFFL samples with comparable nominal outputs, in the presence and absence of Gal4-VPR, are shown in Figure 4d (distributions are shown for each experimental repeat). From these histograms, we see that Gal4-VPR competition causes the entire distributions of the UR sample (UR 1:64 diluted) to shift down in expression, whereas the distributions of the iFFL sample (4x uORFs) retained approximately the same median with comparable variance. Taken together, this data validates the design demonstrating increased robustness for the iFFL module and that as predicted from theory (see Section 2.3), we can tune this robustness by tuning the parameter ϵ with uORFs.

We next tested whether the iFFL module functions in other cell lines and whether its output expression is insensitive to resource loading by different Gal4 TAs (Figure 5a). Overall, we found that the fold-changes in CMVi iFFL output in response to resource competition are much lower than the UR controls for all Gal4 TAs and cell lines tested (Figure 5b-c & Supplementary Figure 22). On average, the UR variants were knocked down between 30% and 40% whereas the iFFL variants changed by 15% or less (Figure 5d & Supplementary Figure 23). As with the experiment shown in Figure 4, we saw larger fold-changes in iFFL variants with 8 uORFs compared to those with fewer. The robustness scores of UR-activator combinations across cell lines were as low as ~30% and less than a third (30.7%) of the combinations had robustness above 80% (Figure 5e). By contrast, the iFFL variants had robustness above 80% for the large majority of combinations (81.5%). On average, the robustness scores of the UR variants ranged between 60% and 75%, whereas those of the iFFL variants ranged between 85% and 90%.

The distribution of robustness scores per cell line for all iFFL and UR variants, as well as constitutive promoters from Figure 2, are compared in Figure 5f. We found that the iFFLs variants showed the highest robustness in HeLa and U2OS cells (100% and 93.3% of combinations over 80% robustness averaging 94.6% and 92.3%, respectively). The high robustness of the iFFL in CHO-K1 cells (84.4% above 80%, averaging 90.9%) is particularly striking in comparison to both the low robustness of the UR variants in that cell line (8.90% above 80%, averaging 59.7%) and our earlier observation that nearly all combinations of constitutive promoters and activators tested in CHO-K1 cells led to decreased expression (Figure 2 & Supplementary Figure 16). The iFFL variants in HEK-293FT cells showed

relatively low robustness (46.7% above 80%, averaging 77.2%) compared to other cell lines, though still higher than the UR variants (37.8% above 80%, averaging 69.2%). This relatively poor performance in HEK-293FT cells appears to result from a slight increase in expression of the iFFL output in response to the Gal4 TAs. Overall, these results demonstrate that the iFFL output is robust to resource loading by different resource competitors across varying cell line contexts, suggesting it is a general solution easily applicable across cell types.

To ensure that our results were not specific to the CMVi promoter, we repeated the experiments in Figures 4 & 5 with a version of the iFFL which replaces the CMVi promoters with hEF1a (Supplementary Figures 24-28). As with the CMVi iFFL, variants of the hEF1a iFFL with fewer uORFs (smaller ϵ) showed reduced fold-changes and higher robustness scores in response to Gal4 TAs than UR variants with comparable nominal outputs (Supplementary Figures 24 & 26). Compared to the CMVi iFFL, the hEF1a iFFL generally showed higher fold-changes and lower robustness scores, especially in U2OS and HeLa cells co-transfected with Gal4-Rta (Supplementary Figure 26). Interestingly, the hEF1a iFFL output for variants with 4 or fewer uORFs was slightly increased (<1.5-fold) by the Gal4 TAs in HEK-293 and HEK-293FT cells. This increase apparently results from toxicity by the Gal4 TAs being exacerbated by the use of several hEF1a promoters in the circuit (Supplementary Figure 29). Toxicity may reduce the growth rate of the cells and thus alter the circuit's dynamics. In support of this notion, transfecting the hEF1a iFFL variants into HEK-293FT cells with a less toxic reagent Viafect (rather than Lipofectamine 3000) eliminates Gal4-VPR-dependent toxicity as well as the increase in output expression (Supplementary Figure 30). For further discussion, see Supplementary Note 6 & Supplementary Figure 31. Overall, like the CMVi iFFL, the hEF1a iFFL module is much more robust than comparable UR variants to resource loading by Gal4 TAs.

2.5 The resource-decoupled module output adapts to plasmid DNA copy number variation

From the model of our endoRNase-based iFFL design, we expect that when $z/\epsilon \gg 1$, the iFFL output will also be robust to variation in the DNA copy number of the iFFL (see Section 2.3). We thus tested whether the output expression of the hEF1a iFFL could adapt to the multiple log decades of variation in plasmid uptake between individual cells seen in transient transfections (Figure 6a). We again used uORFs to tune ϵ, smaller values of which are predicted by the model to make iFFL output adapt to larger ranges of DNA copy numbers (see Section 2.3). Fits of the 2D transfection marker vs iFFL output curves given by equation (4) (see Section 2.3 and Figure 3a-b) are shown overlaid on the cell scatterplots in Figure 6b. For the UR samples, the output is proportional to the transfection marker, so we fit with a simple linear formula: Output = $m \cdot$ TX Marker. The histograms below the scatterplots illustrate that the distribution of expression for cells in different transfection marker-delineated bins is unaffected by changes in DNA copy number over multiple log decades of copy number variation. Similar bin histograms for UR outputs diluted 1:1, 1:4, 1:16, or 1:64 compared to the iFFL output plasmid show that decreasing expression of the

output does not itself cause adaptation to DNA copy number (Supplementary Figure 32a).

To quantify the extent of adaptation of iFFL output expression to DNA copy number, we compared the median expression of cells in finely-sampled transfection marker-delineated bins to the fit value of $Y_{\max}$, and considered a bin to be 'adapted' to copy number variation if $\log_{10}$(output) was within 5% of $\log_{10}(Y_{\max})$ (*i.e.* the log-scale robustness score was above 95% - Figure 6c). As expected based on the model, increasing ϵ by increasing the number of uORFs decreases the range over which the iFFL output adapts to DNA copy number (Figure 6d). Since $Y_{\max}$ is also correlated to ϵ, fit values of $Y_{\max}$ are also highly correlated with the adaptation range (Supplementary Figure 32b). We repeated these experiments and analyses with the CMVi-driven CasE iFFL and found similar results (Supplementary Figure 32c-f).

Since the iFFL can adapt to DNA copy number variations between cells, we investigated whether the iFFL can also adapt to DNA copy number variations within a single cell. In particular, we measured iFFL output over time following transient transfection, during which the dynamics of plasmid DNA expressing the iFFL module are characterized by a step-increase followed by dilution due to cell division. Thus, the dynamics of any constitutively-expressed gene transiently transfected into dividing cells resembles a pulse. According to our modeling results, the iFFL output should be robust to slow temporal changes of DNA copy number. Indeed, variants of the iFFL with fewer uORFs (and thus smaller ϵ) exhibited smaller changes in median expression over the time period of 120 hours post-transfection compared to a UR control as well as a miRNA-based iFFL (Figure 6e, see Supplementary Figure 33 for full distributions at each timepoint). The miRNA-based iFFL used miR-FF4 and was similar to previous designs[29,31,39], with the miR-FF4-expressing intron moved to the 3'UTR and the miRNA target site moved to the 5'UTR. In both the CasE and miR-FF4 iFFLs, placing the target sites for either FF4 or CasE in the 5'UTR rather than the 3'UTR significantly improves adaption to DNA copy number (Supplementary Figure 34). During the time courses, the CasE iFFL showed less change over time than the miR-FF4 iFFL: at 120 hours, the miR-FF4 and 2x-uORFs CasE iFFLs both had similar median output levels, but the maximum miR-FF4 iFFL output was ~2-fold higher (Figure 6e). Finally, we saw that the fit parameters $Y_{\max}$ and Z_{50} of the CasE iFFL variants remain largely unchanged between 48 and 120 hours, whereas those of the miR-FF4 iFFL decreased 5-10-fold over the same period (Figure 6f). Overall, these data demonstrate that the CasE iFFL can also accurately set gene expression levels regardless of DNA dosage to cells and in the face of dynamic transcriptional disturbances such as plasmid dilution.

3 Discussion

The development of sophisticated synthetic genetic circuits will be enabled through improved understanding of how the function of a genetic component is affected by its context[2]. Competition between genes and their products for shared gene expression resources is an important factor causing context-dependence. Several biomolecules at all

levels of gene expression in mammalian cells are known to be shared among many genes and known to induce couplings in their expression[15–20,26,28,32,33]. However, the effects of competition for these resources on engineered genetic systems has not been systematically investigated. We thus focused on understanding the extent to which competition between genetic devices for shared transcriptional resources affects gene expression in synthetic genetic systems. While improved characterization of resource competition effects can improve genetic circuit design[9], it has also been shown that gene expression controllers can be used to automatically mitigate the effects of resource competition in bacterial systems[12–14]. To provide a solution to the resource competition problem in mammalian cells, we thus developed a simple genetic controller that can be added to any genetic device to make its expression level robust to resource availability. This iFFL effectively makes gene expression robust to resource loading by different Gal4 TAs across cell lines, demonstrating the applicability of the system to variable transcriptional contexts.

Transcriptional resource competition

From our characterization of resource competition, we found that constitutive promoters are affected disparately by different Gal4 TAs (Figure 2). The differences in promoter responses may result from the promoters utilizing different subsets of transcriptional resources[40]: there are hundreds of transcriptional cofactors (including CoAs and subunits of the mediator complex) that interact with native and synthetic TFs[41,42]. It was recently shown that TATA-box based and CpG island-based core promoters are activated by different subsets of CoAs[43]. Consistent with these results, we found that the responses of promoters with large CpG islands to Gal4 TAs were more similar to each other than to promoters without CpG islands (Supplementary Figure 15).

In each cell line, we identified several combinations of promoters and Gal4 TAs for which the TA minimally affected promoter expression. These 'non-coupled' combinations may result from the TA recruiting different specific CoAs than those used by the constitutive promoter, and/or from the promoter having relatively high affinity for CoAs. The combinations with minimal or reduced coupling will be useful for choosing parts in synthetic genetic circuits that, when combined together, enable more accurate prediction of circuit behavior. Additionally, relatively strong constitutive promoters that are less affected by resource loading may be utilized as more reliable transfection markers.

A previous study called into question whether squelching by TAs was an artifact that only affected episomal genes by showing that self-squelching only occurs if the TA-driven promoter is in a plasmid, but not in the genome[44]. However, other experiments have demonstrated coupling between signal-responsive genes in the genome[45] and there is growing evidence for the role of squelching in natural gene regulation[32]. Our results show that both non-target squelching and self-squelching can indeed affect integrated genes (see Supplementary Note 3 & Supplementary Figures 8-9). Thus, our results are extendable to various contexts, including episomal and genomic genetic systems.

EndoRNase-based iFFLs decouple genetic device expression levels from resource inputs

In various cell lines and in combination with different Gal4 TAs, we showed that our endoRNase-based iFFL design can effectively cancel-out the effects of transcriptional resource competition on gene expression output

(Figures 4-5). From the model of the the iFFL design (see Section 2.3), we identified a parameter inversely proportional to the robustness of the iFFL, ϵ. We tuned ϵ by placing variable numbers of uORFs in the 5'UTR of the iFFL endoRNase (CasE). Our experiments validated the model prediction that smaller values of ϵ lead to higher robustness of the iFFL output level to resource competition (Figures 4-5) and a larger range over which the iFFL output adapts to DNA copy numbers (Figure 6). In our iFFL design, tuning ϵ also affects the iFFL's output expression level, yielding an inherent trade-off between the set-point of the iFFL and its robustness to transcriptonal perturbations. The output level can be independently tuned from the robustness via the parameter V_y (see Section 2.3), which is proportional to the transcription, translation, mRNA degradation, and protein degradation rates of the output protein. To validate this relationship, we used poly-transfection[38] to sample the iFFL at various ratios of CasE and the output plasmids, finding that indeed, increasing the output DNA dosage (and thus transcription rate) relative to that of CasE enabled higher $Y_{\max}$ at equivalent values of Z_{50} (Supplementary Figure 19). Tuning the relative gene dosage is not possible in all genetic systems and tuning relative transcription rates by changing promoters may break the coupling of resource inputs between the output and endoRNase transcription units. More generally, RNA aptazymes[46] and inducible protein degradation domains[47–49] could be utilized to independently tune RNA or protein stability of the output without affecting the ϵ parameter.

Notably, we found that the nominal output levels (median expression in the absence of resource competition) and fit $Y_{\max}$ parameters for both the CMVi and hEF1a iFFLs were highly correlated across cell lines (Supplementary Figures 35-36). To a simple approximation, differences in gene expression between cell lines can be attributed to the differences in gene expression resources in the cells. If this approximation is valid, then the model of our iFFL (see Section 2.3) predicts that the iFFL output expression will be similar in different cell lines. Indeed, the median output and $Y_{\max}$ values in HEK-293FT cells well-predicted those in other cell lines, with $R^2 > 0.8$ in all cases except for the CMVi iFFL in CHO-K1 cells, which had both values consistently ~4-fold lower than those in HEK-293FT cells (see Supplementary Figure 35 for further discussion). Overall, the iFFL/uORF strategy of setting and tuning gene expression levels enables much higher cell-to-cell consistency in expression levels than using unregulated devices or different constitutive promoters of varying strength (Supplementary Figure 10).

Many existing genetic devices are easily amenable to augmentation by our iFFL design. Augmentation can be achieved by adding (i) an endoRNase driven by the identical promoter as the output of the original device and (ii) an endoRNase target site in the 5'UTR of the device's output mRNA (Figure 1). In bacterial genetic circuits, the endoRNase target site can instead be placed in-frame between the RBS and coding sequence, as previously shown with Cys4[50]. We have recently identified several strong endoRNases that orthogonally cut unique target sites (DiAndreth *et al.*, manuscript in preparation). It is thus feasible that many such iFFLs can be created in a single cell to independently control expression of different genes. Our iFFL may be useful in many applications, including controlling the ratios of genetic device components to maximize device performance[38], making dCas9-based circuits

robust to shared dCas9 resources[51,52], and precisely setting the levels of signaling receptors to achieve unique input functions[53].

To our knowledge, we are the first to show that iFFLs can be used to mitigate the effects of resource competition. For over a decade it has been known that the iFFL topology may enable a genetic device's output to adapt to perturbations[54]. This property have been exploited to create iFFLs that can adapt to DNA copy number variation[29,30] or inducer input levels[31]. Previous solutions to the ribosome competition problem in bacteria utilized negative feedback loops (NFBLs)[12,13]. An advantage of NFBLs is that they can tolerate unknown dynamics in the output gene expression process, whereas iFFLs require the effect of the disturbance to exactly be canceled out by the controller species. Conversely, iFFLs are generally much simpler to design and operate. Recent work showed that combining miRNA-based iFFLs and NFBLs yielded circuits with some properties of both control mechanisms[39], indicating that iFFLs and NFBLs are not mutually exclusive and can synergize when used together.

In bacteria where competition for ribosomes is most prominent[8], it has been proposed that centralized controllers for ribosome levels can ensure a constant supply regardless of loads placed on the ribosome[14]. Conversely, the number of CoAs and GTFs competed for by eukaryotic TAs and promoters is likely far too large to build such a centralized controller[41]. Even if the mediator complex were the only shared transcriptional resource, it is comprised of dozens of domains[55], each of which would need to be controlled.

Adaptation of iFFL expression to its DNA copy number

In addition to resource loading, we found that our iFFL design is also robust to static and dynamic variability in its DNA copy number (Figure 6). Our endoRNase-based iFFL output adapted to copy number variation over ~1-2 log decades, depending on the value of ϵ. This range of adaptation is comparable to the TALER-based iFFL implemented by Segall-Shapiro *et al.* in bacteria[30] and is a major improvement compared to the current standard of miRNA-based iFFLs in mammalian cells[29,31]. EndoRNase-based iFFLs have several advantages over miRNA-based designs, including independence from accessory factors such as RISC, the ability to make gene expression robust to changes in translational resources, and apparently faster dynamics (see Supplementary Note 5 for a more detailed comparison).

Measurements of the CasE iFFL output dynamics during transient transfection showed improved stability over time compared to a UR device and a miRNA-based iFFL (Figure 6). Simulations with an ODE model of the CasE iFFL indicated that changes in iFFL output over time during a transient transfection can be minimized by making the dynamics of CasE expression fast compared to cell division (see Supplementary Note 5 & Supplementary Figure 37). The median degradation rate of native miRNAs (half-life ~20 hours[56]) is comparable to the typical rate of cell division for a mammalian cell (doubling time ~20 hours). Because the CasE iFFL output varies much less than the miR-FF4 iFFL output during transient transfection, our modeling thus suggests that the CasE protein may have a relatively fast degradation rate compared to miR-FF4 and cell division. This data and modeling points to an interesting new mechanism to control the temporal dynamics of iFFLs.

Conclusions

Overall, we have presented characterization of transcriptional resource competition in mammalian cells and the design and performance testing of an endoRNase-based iFFL design which mitigates the effects of resource competition on gene expression. Our characterization of resource competition will be useful both for designing genetic circuits with minimal competition among genetic devices composing the circuits, as well as for predicting the behavior of complex circuits composed of genetic devices which compete for shared resources. Our iFFL is a simple and accurate controller of gene expression that will find many uses in engineering mammalian cells. Altogether, this work will enable more accurate bottom-up design of genetic systems in mammalian cells, facilitating the development of more complex and reliable circuits for applications in cell therapy, tissue/organoid engineering, and cellular bioproduction.

4 Acknowledgements

We would like to acknowledge Douglas Lauffenburger and Ahmad Khalil for discussion as well as Olga Parkin, Cameron Haase-Pettingell, and Darlene Ray for administrative support.

MD2.G and psPax2 were gifts from Didier Trono (Addgene plasmid #12259 & #12260). pFUGW was a gift from David Baltimore (Addgene plasmid #14883). Vero 2.2 cells were a gift from Xandra Breakefield.

5 Funding

This work was supported by the National Institutes of Health (NIH P50GM098792), the National Science Foundation (MCB-1840257), and the United States Air Force Office of Scientific Research (FA9550-14-1-0060).

6 Conflict of Interest

The authors declare no conflicts of interest.

References

1. Purnick, P. E. & Weiss, R. The second wave of synthetic biology: From modules to systems. *Nature Reviews Molecular Cell Biology* **10,** 410–422 (2009).
2. Del Vecchio, D. Modularity, context-dependence, and insulation in engineered biological circuits. *Trends in Biotechnology* **33,** 111–119 (2015).
3. Nielsen, A. K. *et al.* Genetic circuit design automation. *Science* **352,** aac7341 (2016).
4. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal* **7,** 856–866 (2012).
5. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology* **15,** 196–204 (2019).
6. Jayanthi, S., Nilgiriwala, K. S. & Del Vecchio, D. Retroactivity controls the temporal dynamics of gene transcription. *ACS Synthetic Biology* **2,** 431–441 (2013).
7. Yeung, E. *et al.* Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Systems* **5,** 11–24.e12 (2017).
8. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophysical Journal* **109,** 639–646 (2015).
9. Qian, Y., Huang, H. H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synthetic Biology* **6** (2017).
10. Sabi, R. & Tuller, T. Modelling and measuring intracellular competition for finite resources during gene expression. *Journal of the Royal Society Interface* **16,** 20180887 (2019).
11. Kim, J., Darlington, A., Salvador, M. & Jime, I. Trade-offs between gene expression, growth and phenotypic diversity in microbial populations. *Current Opinion in Biotechnology* **62,** 29–37 (2020).
12. Shopera, T., He, L., Oyetunde, T., Tang, Y. J. & Moon, T. S. Decoupling Resource-Coupled Gene Expression in Living Cells. *ACS Synthetic Biology* **6,** 1596–1604 (2017).
13. Huang, H.-H., Qian, Y. & Del Vecchio, D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nature Communications* **9,** 5415 (2018).
14. Darlington, A. P., Kim, J., Jiménez, J. I. & Bates, D. G. Engineering Translational Resource Allocation Controllers: Mechanistic Models, Design Guidelines, and Potential Biological Implementations. *ACS Synthetic Biology* **7,** 2485–2496 (2018).

15. Munding, E. M., Shiue, L., Katzman, S., Donohue, J. & Ares, M. Competition between Pre-mRNAs for the splicing machinery drives global regulation of splicing. *Molecular Cell* **51,** 338–348 (2013).

16. Boudreau, R. L., Martins, I. & Davidson, B. L. Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. *Molecular Therapy* **17,** 169–75 (2009).

17. Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441,** 537–541 (2006).

18. Castanotto, D. *et al.* Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic Acids Research* **35,** 5154–5164 (2007).

19. Lobanova, E. S., Finkelstein, S., Skiba, N. P. & Arshavsky, V. Y. Proteasome overload is a common stress factor in multiple forms of inherited retinal degeneration. *PNAS* **110,** 9986–9991 (2013).

20. Gill, G. & Ptashne, M. Negative effect of the transcriptional activator GAL4. *Nature* **334,** 721–724 (Aug. 1988).

21. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes & Development* **2,** 718–729 (1988).

22. Kelleher, R. J., Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. English. *Cell* **61,** 1209–1215 (June 1990).

23. Tasset, D., Tora, L., Fromental, C., Scheer, E. & Chambon, P. Distinct classes of transcriptional activating domains function by different mechanisms. *Cell* **62,** 1177–1187 (1990).

24. Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: Evidence for transcriptional adaptors. English. *Cell* **61,** 1199–1208 (1990).

25. Flanagan, P. M., Kelleher, R. J., Sayre, M. H., Tschochner, H. & Kornberg, R. D. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **354,** 436–438 (1991).

26. Farr, A. & Roman, A. A pitfall of using a second plasmid to determine transfection efficiency. *Nucleic Acids Research* **20,** 920 (1992).

27. Gilbert, D. M., Heery, D. M., Losson, R., Chambon, P. & Lemoine, Y. Estradiol-inducible squelching and cell growth arrest by a chimeric VP16-estrogen receptor expressed in Saccharomyces cerevisiae: suppression by an allele of PDR1. *Molecular and Cellular Biology* **13,** 462–72 (1993).

28. Fuhrer, D., Han, S. & Ludgate, M. Enhancement of glycoprotein hormone alpha subunit promoter reporter gene activity in co-transfection studies - A cautionary reminder. *Hormone and Metabolic Research* **40,** 787–793 (2008).

29. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology* **7,** 519 (2011).

30. Segall-shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nature Biotechnology* **36** (2018).

31. Strovas, T. J., Rosenberg, A. B., Kuypers, B. E., Muscat, R. A. & Seelig, G. MicroRNA-Based Single-Gene Circuits Buffer Protein Synthesis Rates against Perturbations. *ACS Synthetic Biology* **3,** 324–331 (2014).

32. Schmidt, S. F., Larsen, B. D., Loft, A. & Mandrup, S. Cofactor squelching: Artifact or fact? *BioEssays* **38,** 618–626 (2016).

33. Ptashne, M. How eukaryotic transcriptional activators work. *Nature* **335,** 683–689 (1988).

34. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321,** 960–965 (2008).

35. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).

36. Borchardt, E. K. *et al.* Controlling mRNA stability and translation with the CRISPR endoribonuclease Csy4. *RNA* **21,** 1921–1930 (2015).

37. Ferreira, J. P., Overton, K. W. & Wang, C. L. Tuning gene expression with synthetic upstream open reading frames. *PNAS* **110,** 11284–11289 (2013).

38. Gam, J. J., DiAndreth, B., Jones, R. D., Huh, J. & Weiss, R. One-pot transfection method for rapid characterization and optimization of genetic systems. *Nucleic Acids Research* **47,** e106 (2019).

39. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Research* **46,** 9855–9863 (2018).

40. Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528,** 147–151 (2015).

41. Schaefer, U., Schmeier, S. & Bajic, V. B. TcoF-DB: Dragon database for human transcription co-factors and transcription factor interacting proteins. *Nucleic Acids Research* **39,** 106–110 (2011).

42. Poss, Z. C., Ebmeier, C. C. & Taatjes, D. J. The Mediator complex and transcription regulation. *Critical Reviews in Biochemistry and Molecular Biology* **48,** 575–608 (2013).

43. Haberle, V. *et al.* Transcriptional cofactors display specificity for distinct types of core promoters. *Nature* **4** (2019).

44. Natesan, S., Rivera, V. M., Molinari, E. & Gilman, M. Transcriptional squelching re-examined. en. *Nature* **390,** 349–350 (1997).

45. Webster, G. A. & Perkins, N. D. Transcriptional Cross Talk between NF-κB and p53. *Molecular and Cellular Biology* **19,** 3485–3495 (1999).

46. Yokobayashi, Y. Aptamer-based and aptazyme-based riboswitches in mammalian cells. *Current Opinion in Chemical Biology* **52,** 72–78 (2019).

47. Chung, H. K. *et al.* Tunable and reversible drug control of protein production via a self-excising degron. *Nature Chemical Biology* **11,** 713–720 (2015).

48. Lai, A. C. & Crews, C. M. Induced protein degradation: an emerging drug discovery paradigm. *Nature Reviews Drug Discovery* **16,** 101–114 (2017).

49. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron system with low basal degradation in human cells. *Nature Methods* **16** (2019).

50. Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A. & Arkin, A. P. RNA processing enables predictable programming of gene expression. *Nature Biotechnology* **30,** 1002–1006 (2012).

51. Nielsen, A. A. K. & Voigt, C. A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular Systems Biology* **10,** 763–763 (2014).

52. Chen, P. Y., Qian, Y. & Del Vecchio, D. A Model for Resource Competition in CRISPR-Mediated Gene Repression. *Proceedings of the IEEE Conference on Decision and Control (CDC),* 4333–4338 (2018).

53. Antebi, Y. E. *et al.* Combinatorial Signal Perception in the BMP Pathway. *Cell* **170,** 1184–1196.e24 (2017).

54. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* **138,** 760–773 (2009).

55. Allen, B. L. & Taatjes, D. J. The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology* **16,** 155–166 (2015).

56. Marzi, M. J. *et al.* Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Research* **26,** 554–565 (2016).

7 Methods

Modular plasmid cloning scheme

Plasmids were constructed using a modular Golden Gate strategy similar to previous work in our lab[38,57]. Briefly, basic parts (insulators, promoters, 5'UTRs, coding sequences, 3'UTRs, and terminators – termed level 0s (pL0s)) were assembled into transcription units (TUs – termed level 1s (pL1s)) using BsaI Golden Gate reactions. TUs were assembled into multi-TU plasmids using SapI Golden Gate reactions. To make lentivirus transfer plasmids, pL0s or pL1s were cloned into a vector derived from pFUGW (AddGene plasmid #14883) using either BsaI or SapI Golden Gate, respectively. Genbank files for each plasmid and vector backbone used in this study are described in Supplementary Table 6 and available in Supplementary Data 1.

Cell culture

HEK-293 cells (ATCC), HEK-293FT cells (Thermo Fisher), HeLa cells (ATCC), and Vero 2.2 cells (Massachusetts General Hospital) were maintained in Dulbecco's modified Eagle media (DMEM) containing 4.5 g/L glucose, L-glutamine, and sodium pyruvate (Corning) supplemented with 10% fetal bovine serum (FBS, from VWR). CHO-K1 cells (ATCC) were grown in F12-K media containing 2 mM L-glutamine and 1500 mn/L sodium bicarbonate (ATCC) supplemented with 10% FBS. U2OS cells (ATCC) were grown in McCoy's 5A media with high glucose, L-glutamine, and bacto-peptone (Gibco) supplemented with 10% FBS. All cell lines used in the study were grown in a humidified incubator at 37deg and 5% CO_2. All cell lines tested negative for mycoplasma.

Transfections

Cells were cultured to 90% confluency on the day of transfection, trypsinized, and added to new plates simultaneously with the addition of plasmid-transfection reagent mixtures (reverse transfection). Transfections were performed in 24-well or 96-well pre-treated tissue culture plates (Costar). Following are the volumes, number of cells, and concentrations of reagents used for 96-well transfections; for 24-well transfections, all values were scaled up by a factor of 5. 120 ng total DNA was diluted into 10 μL Opti-MEM (Gibco) and lightly vortexed. The transfection regent was then added and samples were lightly vortexed again. The DNA-reagent mixtures were incubated for 10-30 minutes while cells were trypsinized and counted. After depositing the transfection mixtures into appropriate wells, 40,000 HEK-293, 40,000 HEK-293FT, 10,000 HeLa, 20,000 CHO-K1, 20,000 Vero 2.2, or 10,000 U2OS cells suspended in 100 μL media were added. The reagent used in each experiment along with plasmid quantities per sample and other experimental details are shown in Supplementary Tables 4 & 5. Lipofectamine LTX (ThermoFischer) was used at a ratio of 1 μL PLUS reagent and 4 μL LTX per 1 μg DNA. PEI MAX (Polysciences VWR) was used at a ratio of 3 μL PEI per 1 μg DNA. Viafect (Promega) was used at a ratio of 3 μL Viafect per 1 μg DNA. Lipofectamine 3000 was used at a ratio of 2 μL P3000 and 2 μL Lipo 300 per 1 μg DNA. Attractene (Qiagen) was used at a ratio of 5 μL Attractene per 1 μg DNA. For experiments with measurement windows between 12-72

hours (as indicated on the figures or in their captions), the media of the transfected cells was not replaced between trasnfection and data collection. For experiments with measurements at longer time points, the transfected cells were passaged at 72 hours in fresh media on a new plate. In order to maintain a similar number of cells for data collection at longer time points, transfected cells were split at ratios of 1:2 or 1:4 for samples being collected at 96 or 120 hours, respectively. For all transfections with Doxycycline (Dox, Sigma-Aldrich), Dox was added immediately after transfection; an exception is the experiment shown in Supplementary Figure 8, in which Dox was added 24 hours after transfection.

In each transfection sample, we included a hEF1a-driven transfection marker to indicate the dosage of DNA delivered to each cell and to facilitate consistent gating of transfected cells. Of the strong promoters we tested (CMV, CMVi, and hEF1a), the hEF1a promoter gave the most consistent expression across cell lines and was generally less affected by resource loading by Gal4 TAs (Supplementary Figures 1, 10, 15, & 18). The data in Supplementary Figure 19 used CMV promoters for all transcription units (including the transfection marker).

Lentivirus production and infection

Lentivirus production was performed using HEK-293FT cells and second-generation helper plasmids MD2.G (Addgene plasmid #12259) and psPax2 (Addgene plasmid #12260). HEK-293FT cells were grown to 90% confluency, trypsinized, and added to new pre-treated 10 cm tissue culture plates (Falcon) simultaneously with addition of plasmid-transfection reagent mixtures. Four hours before transfection, the media on the HEK-293FT cells was replaced. To make the mixtures, first 3 μg psPax2, 3 μg pMD2.g, and 6 μg of the transfer vector were diluted into 600 μL Opti-MEM and lightly vortexed. 72 μL of FuGENE6 (Promega) was then added and the solution was lightly vortexed again. The DNA-FuGENE mixtures were incubated for 30 minutes while cells were trypsinized and counted. After depositing the transfection mixtures into appropriate plates, 1×10^6 HEK-293FT cells suspended in 10 mL media were added. 16 hours after transfection, the media was replaced. 48 hours after transfection, the supernatant was collected and filtered through a 0.45 PES filter (VWR). HEK-293FT cells were grown to 90% confluency, trypsinized, and 1×10^6 cells were resuspended in 2 mL of viral supernatant and together added to a pre-treated 6-well tissue culture plate (Costar). To facilitate viral uptake, we added polybrene (Millipore-Sigma) to a final concentration of 8 μg/mL. Cells infected by lentiviruses were expanded and cultured for at least two weeks before use in experiments using the conditions for culturing HEK-293FT cells described above.

Flow cytometry

To prepare samples in 96-well plates for flow cytometry, the following process was followed: media was aspirated, 50 μL PBS (Corning) was added to wash the cells and remove FBS, the PBS was aspirated, and 40 μL Trypsin-EDTA (Corning) was added. The cells incubated for 5-10 minutes at 37deg C to allow for detachment and separation. Following incubation, 80 μL of DMEM without phenol red (Gibco) with 10% FBS was added to inactivate the trypsin. Cells were thoroughly mixed to separate and suspend individual cells. The plate(s) were then spun down at

400 × g for 4 minutes, and the leftover media was aspirated. Cells were resuspended in 170 μL of PBS supplemented with 1% BSA (Thermo Fisher), 5 mM EDTA (VWR), and 0.1% sodium azide (Sigma-Aldrich) to prevent clumping. For prepping larger plates, all volumes were scaled up in proportion to surface area and samples were transferred to 5 mL polystyrene FACS tubes (Falcon) after trypsinization. For standard co-transfections, 10,000-50,000 cells were collected per sample. For the poly-transfection experiment and transfections into cells harboring an existing lentiviral integration, 100,000-200,000 cells were collected per sample.

For the experiments shown in Figure 1 and Supplementary Figure 2-3, samples were collected on a BD LSR II cytometer equipped with a 405nm laser with 450/50nm filter ('Pacific Blue') for measuring TagBFP or EBFP2, 488 laser with 515/20 filter ('FITC') for measuring EYFP or mNeonGreen, 561nm laser with 582/42nm filter ('PE') or 610/20nm filter ('PE-Texas Red') for measuring mKate2 or mKO2, and 640 laser with 780/60nm filter ('APC-Cy7') for measuring iRFP720. For all other experiments, samples were collected on a BD LSR Fortessa equipped with a 405nm laser with 450/50nm filter ('Pacific Blue') for measuring TagBFP or EBFP2, 488 laser with 530/30 filter ('FITC') for measuring EYFP or mNeonGreen, 561nm laser with 582/15nm filter ('PE') or 610/20nm filter ('PE-Texas Red') for measuring mKate2 or mKO2, and 640 laser with 780/60nm filter ('APC-Cy7') for measuring iRFP720. 500-2000 events/s were collected either in tubes via the collection port or in 96-well plates via the high-throughput sampler (HTS). All events were recorded and compensation was not applied until data processing (see below).

Flow cytometry data analysis

Analysis of flow cytometry data was performed using our MATLAB-based flow cytometry analysis pipeline (`https://github.com/Weiss-Lab/MATLAB_Flow_Analysis`). Arbitrary fluorescence units were converted to standardized molecules of equivalent fluorescein (MEFL) units using RCP-30-5A beads (Spherotech) and the TASBE pipeline process[58]. Briefly, fluorescence compensation was performed by subtracting autofluorescence (computed from wild-type cells), computing linear fits between channels in single-color transfected cells, then using the fit slopes as matrix coefficients for matrix-based signal de-convolution. Single cells were isolated by drawing morphological gates based on cellular side-scatter and forward-scatter. Threshold gates were manually drawn for each channel based on the fluorescence of untransfected cells. Generally, transfected cells within a population were gated by selecting cells that pass either the gate for the output of interest (Output$^+$) or pass the gate for the transfection marker (TX Marker$^+$).

In Figure 2, our library of constitutive promoters had different nominal expression levels and were variably affected by competition. We thus include a discussion and examples of how fluorescent gating strategies affect the measurements of expression and fold-changes in Supplementary Note 4 and Supplementary Figure 11. Additionally, in order to limit the bias in our reporting of minimally-affected promoters by the the proximity of {P}:Output$_1$ expression to autofluorescence (Supplementary Figure 13), our analysis of just this dataset incorporates an additional

autofluorescence subtraction step described in Supplementary Note 4. A comparison of the differences in fold-changes with and without this additional autofluorescence subtraction is shown in Supplementary Figure 14a. This step reduced the correlation between the nominal output levels of {P}:Output$_1$ and the fold-changes in response to resource loading by Gal4 TAs (Supplementary Figure 14b).

When first analyzing the data in Figure 4, we found that the measurements of fold-changes and robustness for the UR variants with diluted output plasmid DNA were sensitive to the fluorescent gating strategy used in the analysis. Our typical gating routine of selecting cells positive for either the output or the transfection marker yielded fold-changes of the diluted UR variants that were much larger than when gating on cells positive for just the output. Conversely, both gating strategies yielded similar fold-changes for the iFFL variants regardless of their nominal output. We suspect that the difference in measurements for the diluted UR variants may result from (i) reduced UR plasmid uptake when forming lipid-DNA complexes for co-transfection with the Gal4-VPR plasmid (which is larger than the DNA-mass-offsetting plasmid Gal4-None) and/or (ii) repression of UR output expression below the autofluorescence threshold. Since these confounding factors could not be distinguished, we report the results for the cells gated positive for just the output (which more conservatively estimates fold-changes in the output of the UR system) in the main figures and include results for gating cells positive for either the output or the transfection marker in Supplementary Figures 20 & 22 for comparison. For the hEF1a iFFL, we also include comparisons of results with both gating strategies in Supplementary Figures 24-27.

Estimation of cell concentration by flow cytometry

When collecting flow cytometry data, we typically constrained the number of events collected, making the count of cells per sample not representative of the total number of cells per well. We instead estimated the concentration of cells in a given sample by the following formula:

$$[\text{Cells}](cells/\mu L) = \frac{\text{Event rate}(\text{cells}/s)}{\text{Flow rate}(\mu L/s)}$$

To compute the event rate, we estimated the number of cells (*i.e.* events passing morphological gating) per second in each sample. The length of time between the measurements of individual cells in flow cytometry approximately follows an exponential distribution. We thus fit an exponential distribution using the MATLAB function 'fitdist()' (`https://www.mathworks.com/help/stats/fitdist.html`) to the differences between time-stamps of collected cells. Before fitting, we removed inter-cell times larger than the 99.9th percentile to prevent biasing by large outliers. The characteristic parameter of the exponential distribution (λ) is the inverse of the average time between events. Thus, the event rate is given by $\frac{1}{\lambda}$ (*i.e.* the mean of the exponential distribution).

To ensure a known and controlled flow rate, any sample for which we would measure the concentration was collected via the HTS attached to the flow cytometer. The flow rate of the HTS was can be set through the FACSDiva

Software (BD) controlling the instrument. The flow rate of each sample was recorded and input into the calculation.

RT-qPCR

Transfections for qPCR were conducted in 24-well plates (Costar). RNA was collected 48 hours after transfection with the RNeasy Mini kit (Qiagen). Reverse-transcription was performed using the Superscript III kit (Invitrogen) follwoing the manufacturer's recommendations. Real-time qPCR was performed using the KAPA SYBR FAST qPCR 2X master mix (Kapa Biosystems) on a Mastercycler ep Realplex (Eppendorf) following the manufacturer's recommended protocol. Primers for the CMV-driven output (mKate) targeted the coding sequence. Primers for 18S rRNA were used as an internal control for normalization. The qPCR calculations are shown in Supplementary Table 7.

Primers:

mKate (CMV:Output) forward: GGTGTCTAAGGGCGAAGAGC

mKate (CMV:Output) reverse: GCTGGTAGCCAGGATGTCGA

18S forward: GTAACCCGTTGAACCCCATT

18S reverse: CCATCCAATCGGTAGTAGCG

Model fitting

Where possible, fluorescent reporters were used to estimate the concentration of a molecular species for the purpose of model fitting. For fitting the Gal4 activator dose response curves (both activation and competition) in Figure 1 and Supplementary Figure 2, we used a fluorescent marker co-titrated with the Gal4 activators to estimate the amount of Gal4 delivered per cell. The Gal4 marker correlated with the DNA dosage with an R^2 value of 0.86 or better for each experimental repeat (Supplementary Figure 3a). However, the sensitivity of activation to Gal4 levels made the measurements as a function of Gal4 DNA dosage relatively noisy between experimental repeats (Supplementary Figure 3b-e). Thus, the marker levels could more accurately estimate the amount of Gal4 expressed in the median cell than the DNA dosages.

For fitting both the resource competition and iFFL models, we used the MATLAB function 'lsqcurvefit()' (`https://www.mathworks.com/help/optim/ug/lsqcurvefit.html`), which minimizes the sum of the squares of the residuals between the model and the data. As the function input values we used the level of either the Gal4 TA (in the case of resource competition – as measured by Gal4 Marker) or the transfection marker (in the case of the iFFL). For fitting the Gal4 TA dose-response data, the residuals were computed between the median CMV:Output$_1$ or UAS:Output$_2$ levels and function outputs directly. In addition, all median values computed from different experimental repeats were pooled together before fitting. For fitting iFFL and UR models, the residuals were computed between the $\log_1 0$- and biexponentially-transformed levels of the output protein of interest and the $\log_1 0$- and biexponentially-transformed function outputs, respectively. In experiments with the hEF1a iFFL being tested only in HEK-293FT cells, the entire morphologically-gated population of cells was used for fitting. In hEF1a iFFL

experiments containing multiple cell types, to prevent the model from over-fitting the untransfected population in more difficult-to-transfect cells, the cells in each sample were analytically binned into half-log-decade-width bins based on the transfection marker, and an equivalent number of cells from each bin were extracted, combined, and used for fitting. In samples with the CMVi iFFL, the relatively high expression of the CMVi promoter compared to the hEF1a promoter (which is used as a transfection marker and proxy for DNA/resource input level z) in most cell lines imposes non-linearity in the transfection marker vs output curve at low plasmid DNA copy numbers per cell. This non-linearity led us to gate cells positive for either the iFFL output or the transfection marker for fitting. For the resource competition models, all parameters for all Gal4 TAs were fit simultaneously using a custom function, 'lsqmultifit()', that was created based on 'nlinmultifit()' on the MATLAB file exchange (`https://www.mathworks.com/matlabcentral/fileexchange/40613-multiple-curve-fitting-with-common-parameters-using-nlinfit`).

Goodness of fit was measured by computing the normalized root-mean-square error CV(RMSE). CV(RMSE) was computed with the following equation:

$$\text{CV(RMSE)} = \frac{\sqrt{\frac{1}{\bar{y}} \sum_i (y(x_i) - f(x_i))^2}}{\bar{y}}$$

Where $y(x_i)$ is the value of the data at the input value x_i, $\bar{y}$ is the mean of y for all values of x, and $f(x_i)$ is the function output at input value x_i.

References

38. Gam, J. J., DiAndreth, B., Jones, R. D., Huh, J. & Weiss, R. One-pot transfection method for rapid characterization and optimization of genetic systems. *Nucleic Acids Research* **47,** e106 (2019).

57. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Research* **42,** 13440–13451 (Nov. 2014).

58. Beal, J., Weiss, R., Yaman, F., Adler, A. & Davidsohn, N. A Method for Fast, High-Precision Characterization of Synthetic Biology Devices. *MIT CSAIL Technical Report 2012–008* (2012).

Figures

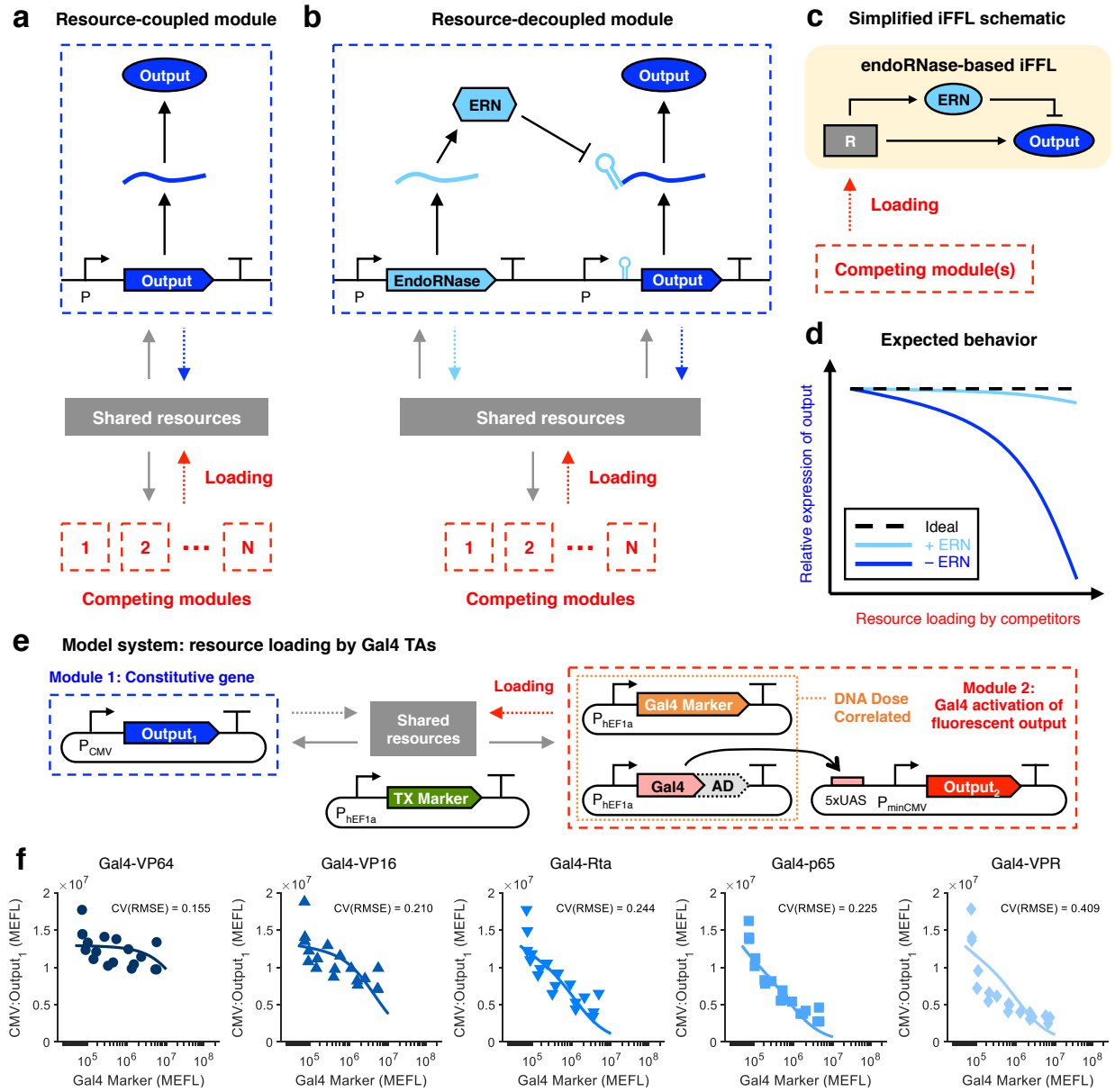


Figure 1

iFFL-based approach for decoupling modules with shared limited resources. (a) A genetic module comprising a single constitutive transcription unit. Other competing modules place a load (indicated by the dashed red arrow) on the free cellular resources, affecting expression of the module of interest (resource-coupled module). The module of interest also applies a load to the resources (indicated by the dashed blue arrow). (b) An incoherent feedforward loop (iFFL) device within the module of interest decouples the module's output from resource variability. An

endoribonuclease (endoRNase/ERN) produced by the identical promoter as the output represses the output by binding to a specific target site in its 5'UTR and cutting the mRNA. (c) A simplified schematic of the iFFL showing how cellular resources directly contribute to the production of both the endoRNase and the output. (d) The expected behavior of the output of the resource-coupled and resource-decoupled modules in response to resource loading by other modules. (e) Experimental model system of resource loading. The module of interest comprises a constitutively expressed protein (Output_1). In a competitor module, Gal4 transcriptional activators (TAs) drive expression of an another protein (Output_2). Different activation domains (ADs) were fused to the DNA binding domain (DBD) of Gal4. A reporter (Gal4 Marker) was titrated together with the Gal4 TAs to mark their delivery per cell. Note that the transfection marker (TX Marker) also competes for resources, but that process is omitted from schematic for simplicity. (f) Dose-dependent effect of Gal4 TAs on Output_1. The markers indicate median expression levels from three experimental repeats. The lines represent fits of our steady-state resource competition model (equation (35a), see Supplementary Note 1 and Supplementary Figure 5). Dose response curves and model fits for Output_2 are shown in Supplementary Figure 2. The CV(RMSE) is the root-mean-square error between the model and data, normalized by the mean of the data. All data were measured by flow cytometry at 48 hours post-transfection in HEK-293FT cells. Median values for each sample are shown in Supplementary Table 2. All measurements were made on cells gated positive for the TX Marker *or* Output_1. Fit parameters for each sample are shown in Supplementary Table 3.

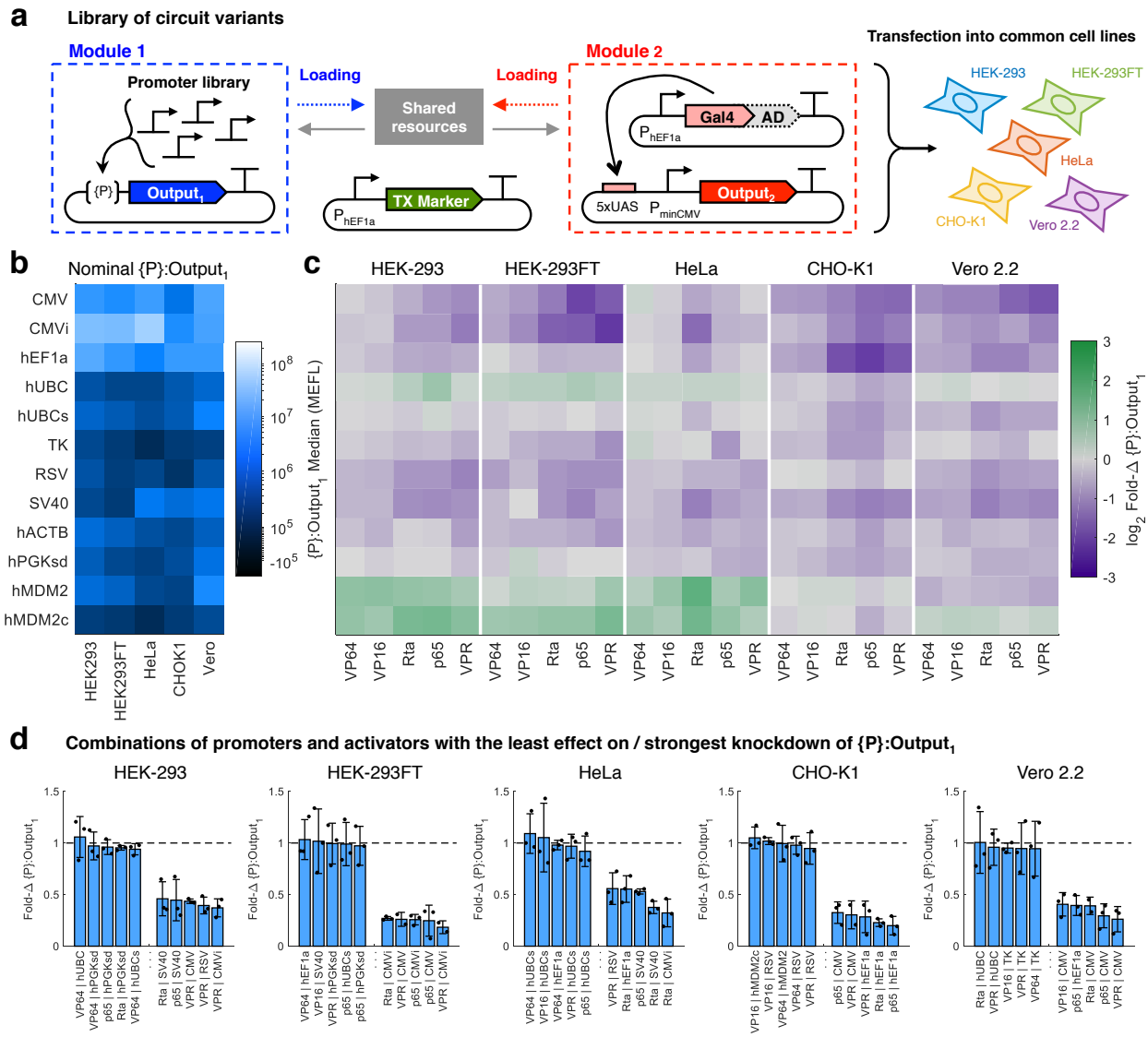


Figure 2

Effect of resource competition between promoters and activators across cell lines. (a) Genetic model system to study competition for resources between different combinations of promoters, Gal4 transcriptional activators (TAs), and cell lines. The specific promoters, activation domains (ADs) fused to Gal4, and cell lines are shown alongside the data in b & c. (b) Nominal Outputs are the median expression levels of each promoter in Module 1 ({P}:Ouptut$_1$) in each cell line when co-transfected with Gal4-None (*i.e.* the Gal4 DNA binding domain), which does not load resources (Supplementary Figure 4). (c) Fold-changes (Fold-Δs) in the level of {P}:Output$_1$ in response to Gal4 TAs. The Fold-Δs are computed independently for each promoter and cell line by dividing the median level of {P}:Output$_1$ for each sample co-transfected with different Gal4 TAs by the Nominal Output. (d) The five promoter-activator combinations in each cell line with the smallest effect or largest negative effect on the level of

$Output_1$. The plots show the mean and standard deviation of three experimental repeats (represented by the individual points). All data were measured by flow cytometry at 48 hours post-transfection in the cell lines indicated. Panels b and c show the geomean of median measurements and mean of fold-changes from three experimental repeats, respectively. Median values for each sample are shown in Supplementary Figure 10 and Supplementary Table 2. All measurements were made on cells gated positive for TX Marker *or* {P}:$Output_1$. See Supplementary Figure 11 for a comparison with other gating strategies.

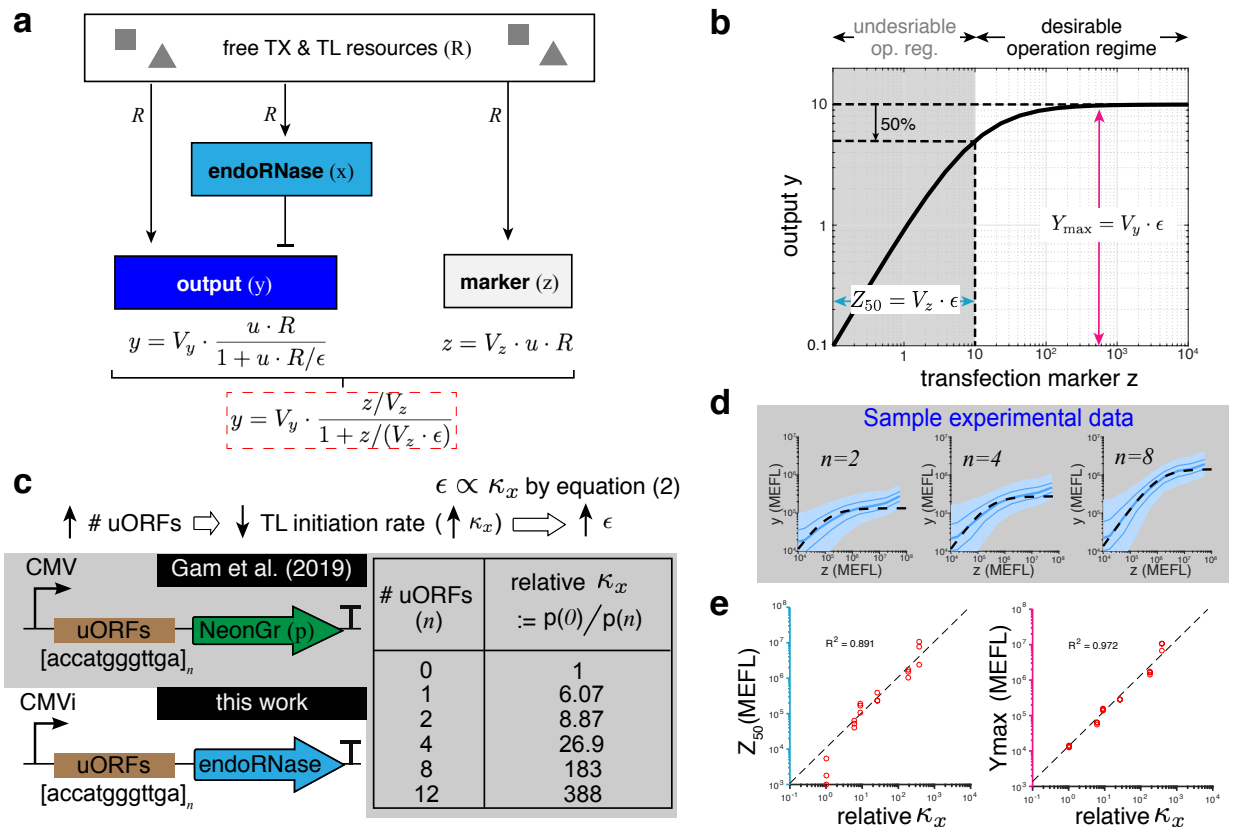


Figure 3

Adaptation to resource perturbations of an endoRNase-based iFFL module. (a) A schematic of the iFFL module. The expression of the output gene (y) may change due to variations in the availability of free transcriptional (TX) and translational (TL) resources (R). When the gene is regulated by an endoRNase (x), an unintended increase (decrease) in R increases (decreases) the amount of endoRNase to reduce (increase) the amount of the output by enhancing its mRNA degradation. This action compensates for the unintended increase (decrease) in the regulated gene's production rate due to variations in R. Since the same pool of TX and TL resources is also used to express the transfection marker z in a transient transfection experiment, we use the marker's concentration z as a proxy to quantify R experimentally. (b) The steady state output level (y) of the iFFL can be written as a function of the marker level (z). We evaluate the performance of an iFFL by (i) its maximum output ($Y_{\max}$) and (ii) its robustness to variation in R and therefore z, characterized by (Z_{50}). In our model, both $Y_{\max}$ and Z_{50} are linear functions of ϵ, which can be used as a design parameter (see equation (2)). ϵ is proportional to the decay rates of the endoRNase and output mRNA and is inverseley proportional to the production rate and catalytic efficiency of the endoRNase. (c) An increase in the number of uORFs in the 5' UTR of the endoRNase's transcript leads to a decrease in its TL initiation rate. We model it as an increase in the dissociation constant between the ribosome and the endoRNase's mRNA

transcript (κ_x), which increases ϵ. The relationship between the number of uORFs and the fold decrease in TL initiation (*i.e.*, parameter κ_x in the model) is summarized in the table using previously-published experimental data[38]. (d) Sample experimental data corresponding to the theoretical plot in b. n indicates the number of uORFs in the 5'UTR of CasE. The shroud indicates the 5th to 95th percentiles of the output in each half-log-decade TX Marker bin. The thin lines mark the 25th and 75th percentiles, and the thick line marks the median Output in each bin. (e) Comparison between experimentally measured inverse robustness metric (Z_{50}) and maximum output ($Y_{\max}$) and the relative difference in ribosome-mRNA dissociation constant (κ_x) for different numbers of uORFs in the 5'UTR of the endoRNase. Fit parameters for each sample are shown in Supplementary Table 3.

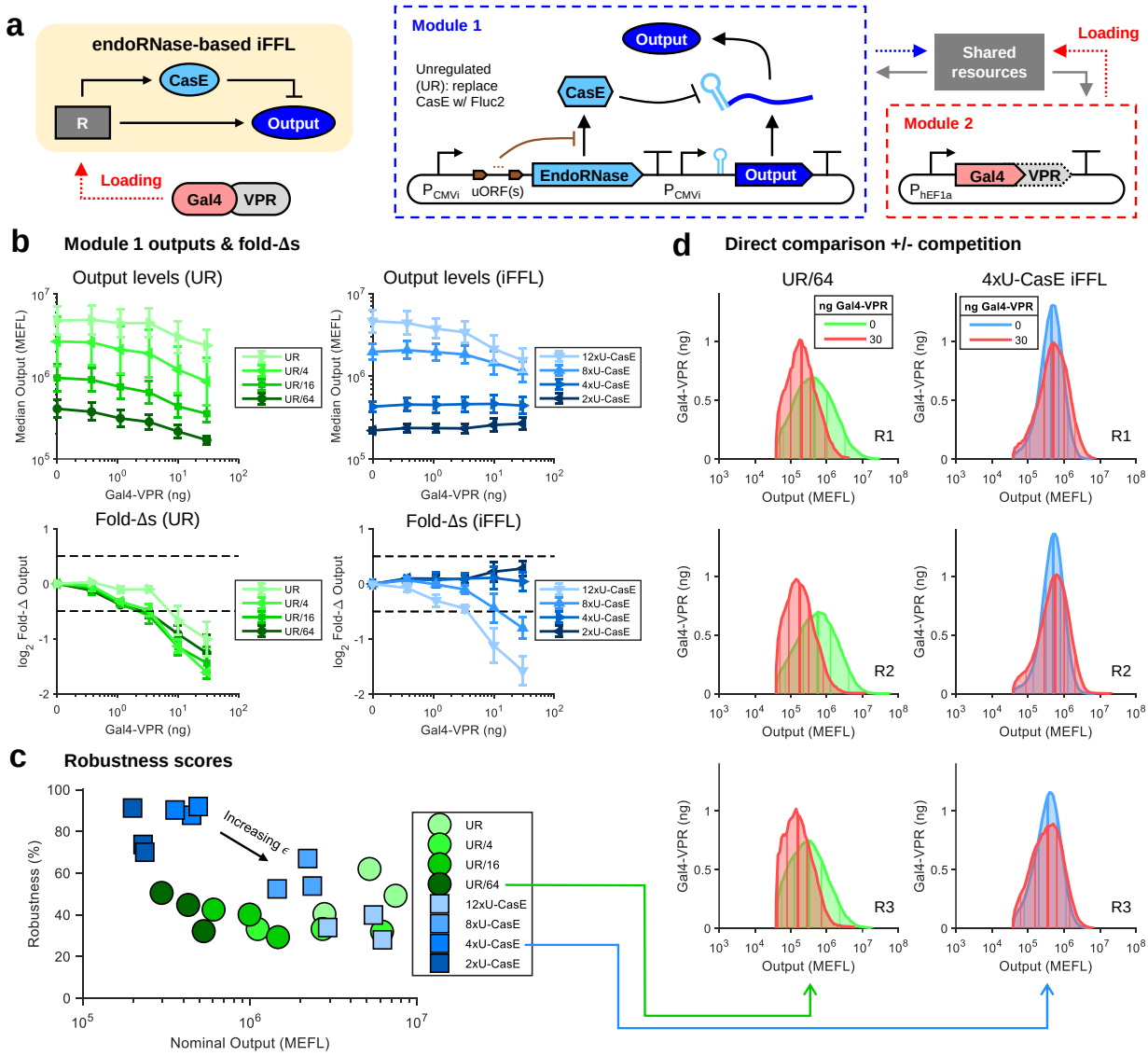


Figure 4

Robustness of the iFFL output level to resource loading by Gal4-VPR. (a) Schematic and genetic diagram of the endoRNase-based iFFL module. 0-12 uORFs are placed in front of CasE to reduce its translation. CasE binds and cleaves to a specific target site in the 5'UTR of the output mRNA. The unregulated (UR) control is the same overall design but replaces CasE with Fluc2. The UR control has no uORFs and retains the CasE target site in the 5'UTR of the output mRNA. (b) Comparison of UR and iFFL responses to resource loading by Gal4-VPR. The median output plots show the mean (μ) ± relative error ($\frac{1}{ln(10)} \cdot \frac{\sigma}{\mu}$) of medians for three experimental repeats. We use relative error rather than standard deviation (σ) to more accurately represent error on the log-scale of the y-axis[59]. Fold-changes (Fold-Δs were computed by dividing the median level of Output at a given concentration of Gal4-VPR by the Nominal Output level (the median level of Output at 0 ng Gal4-VPR). The Fold-Δ plots show the mean ± standard

deviation of Fold-Δs of individual repeats. (c) Robustness scores (see equation (7) in the text) for each UR and iFFL variant as a function of the Nominal Output. The robustness of the iFFL decreases as ϵ is increased. Each box or square represents an individual experimental replicate. (d) Direct comparison of the distributions of Output at 0 and 30 ng Gal4-VPR for UR and iFFL variants with similar Nominal Outputs, for all three experimental repeats. The lines on the histograms denote the 5th, 25th, 50th, 75th, and 95th percentiles. All data were measured by flow cytometry at 72 hours post-transfection in HEK-293FT cells. Median values for each sample are shown in Supplementary Table 2. All measurements were made on cells gated positive for Output. Measurements on cells gated positive for either Output or TX Marker are shown in Supplementary Figure 20. iFFL samples with 0 or 1 uORFs are not shown because most or all of the cells in those samples did not express Output above the autofluorescence background and their median expression levels were much lower than that of any UR variants.

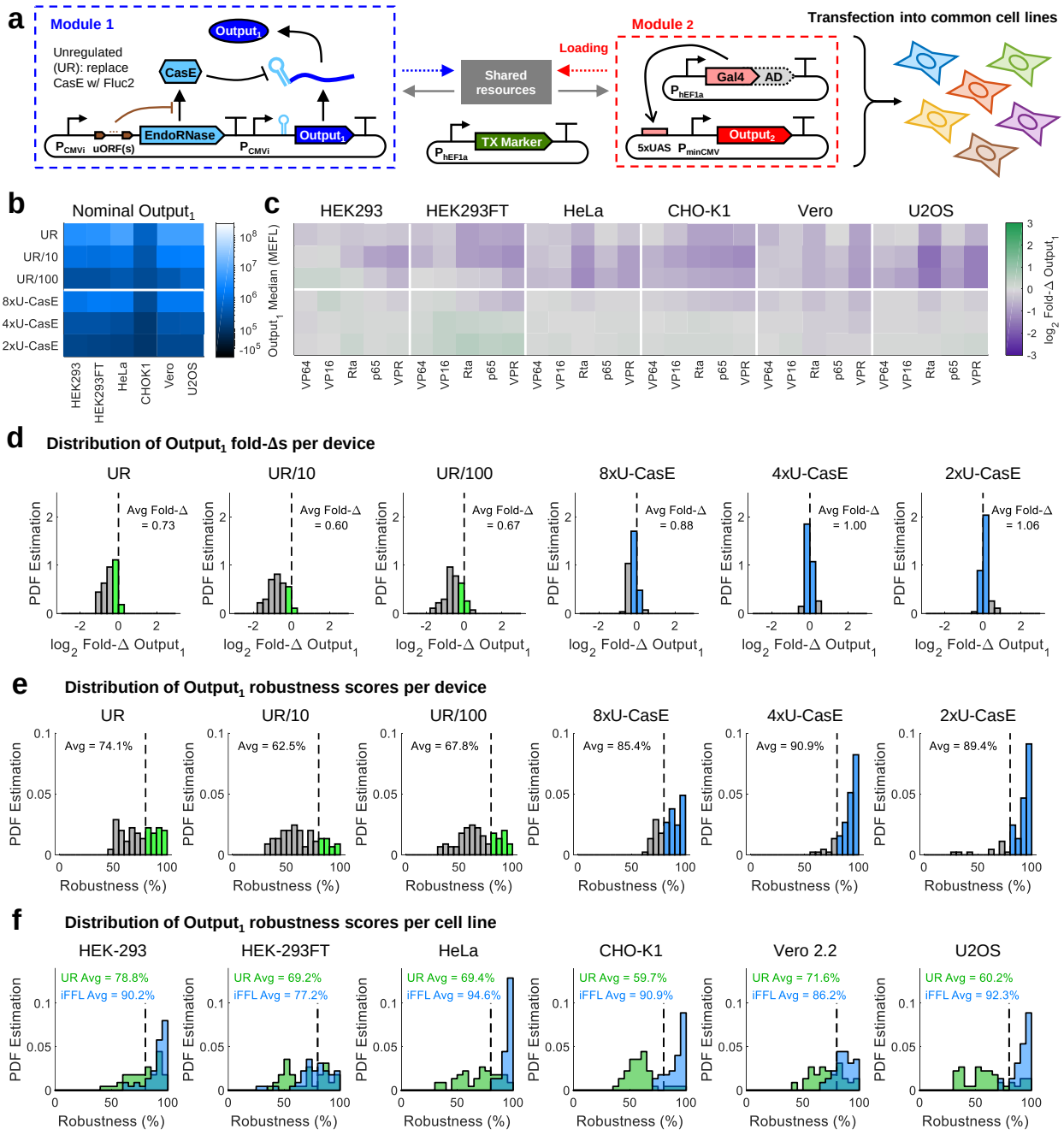


Figure 5

Robustness of the iFFL output level to resource loading across cell lines. (a) Schematic of experiment to test the performance of the iFFL in different cell lines with different Gal4 TAs loading resources. (b) Nominal Outputs are the median expression levels of each UR or iFFL variant ($Ouptut_1$) in each cell line when co-transfected with Gal4-None (*i.e.* the Gal4 DNA binding domain), which does not load resources (Supplementary Figure 4). (c) Fold-changes (Fold-Δs) in the level of $Output_1$ in response to Gal4 TAs. The Fold-Δs are computed independently for

each UR and iFFL variant and cell line by dividing the median level of $Output_1$ for each sample co-transfected with different Gal4 TAs by the Nominal Output. (d) Distribution of Fold-Δs for each UR and iFFL variant. Histogram bins with minimal Fold-Δs are highlighted green for UR variants and blue for iFFL variants. The average (mean) of all Fold-Δs for a given UR or iFFL variant is shown inset in each plot. (e) Distribution of robustness scores for each UR and iFFL variant. Histogram bins with > 80% robustness are highlighted green for UR variants and blue for iFFL variants. (f) Comparison of distributions of robustness scores for all UR vs iFFL variants in each cell line. The average (mean) robustness scores for each group are shown inset in each plot. All data were measured by flow cytometry at 72 hours post-transfection in the cell lines indicated. Panels b and c show the geomean of median measurements and mean of fold-changes from three experimental repeats, respectively. Median values for each sample are shown in Supplementary Figure 21 and Supplementary Table 2. Panels d, e, and f combine data from each replicate. All measurements were made on cells gated positive for $Output_1$ only. Measurements on cells gated positive for either $Output_1$ or TX Marker are shown in Supplementary Figures 21 & 22.

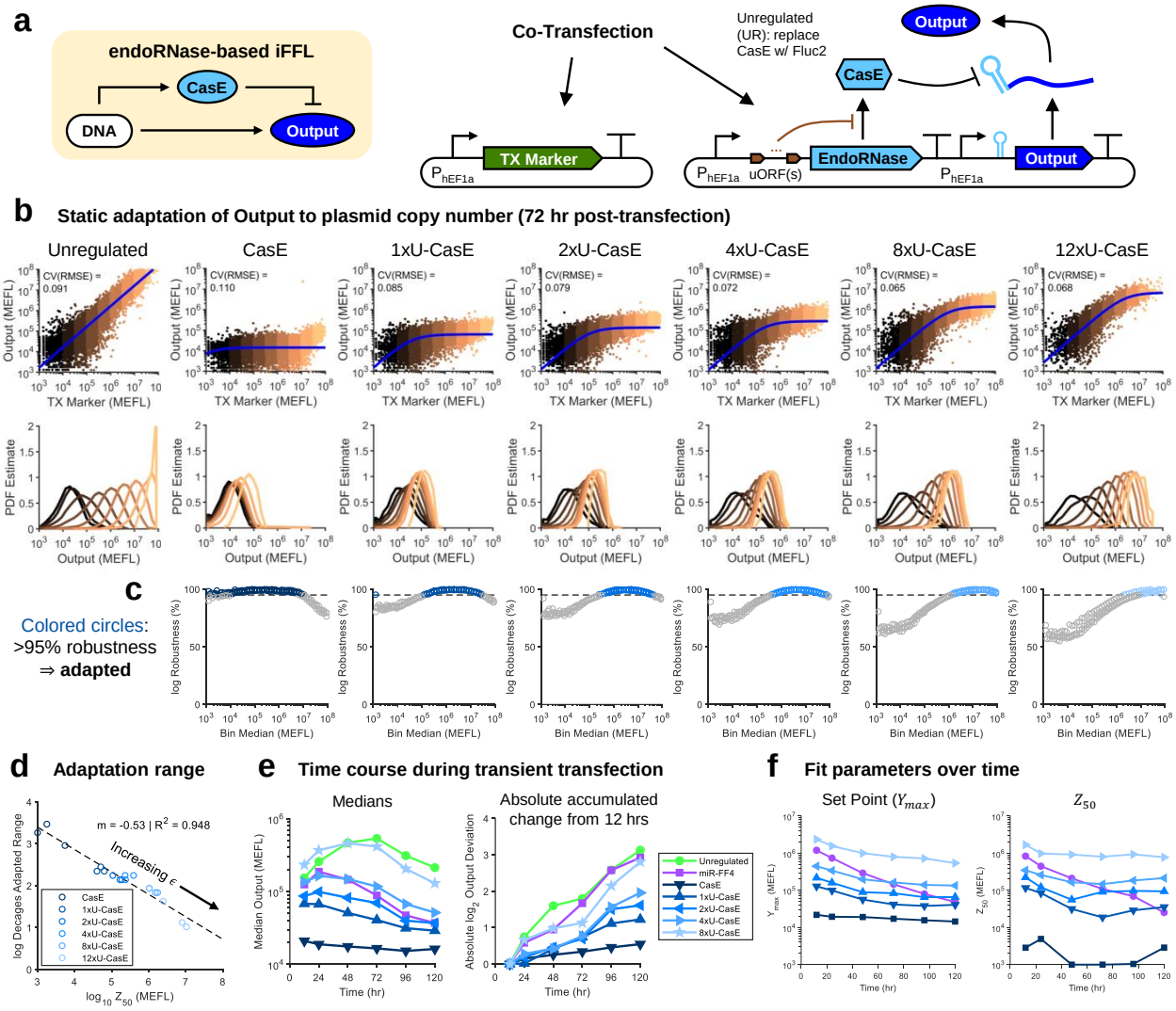
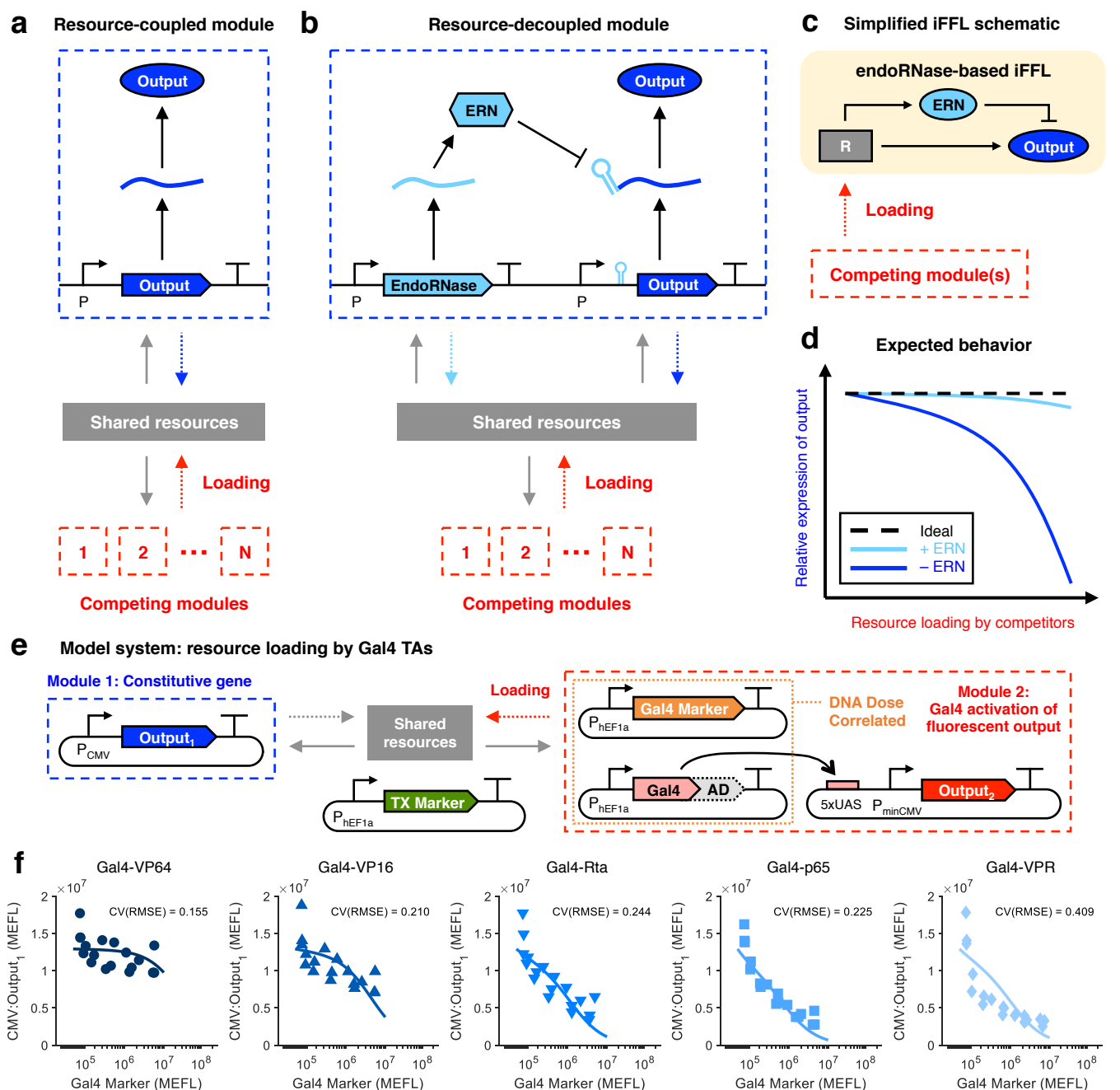
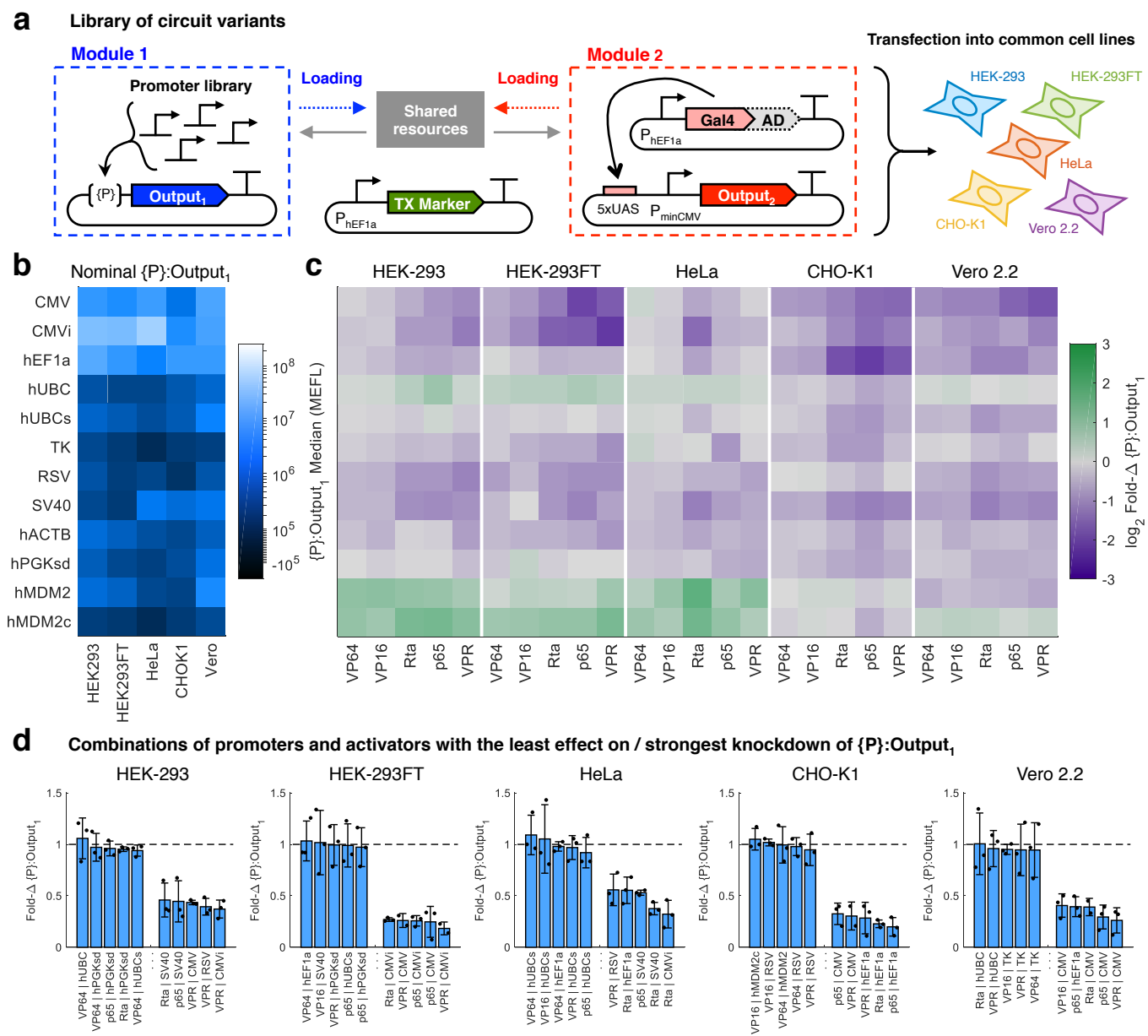


Figure 6

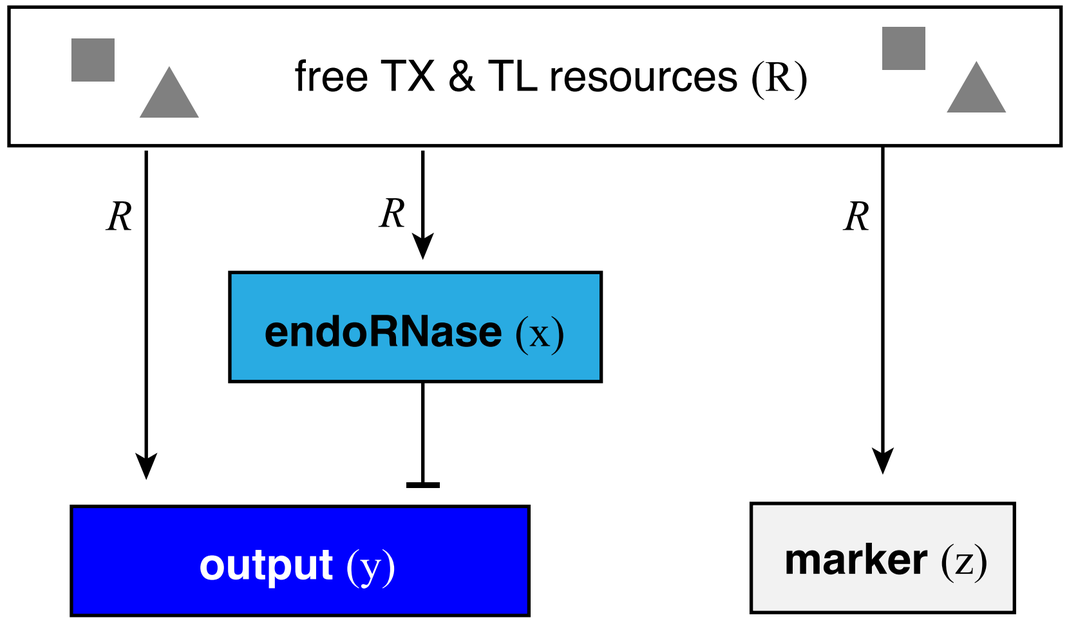
Adaptatation of the iFFL output level to DNA copy number variation. (a) Genetic diagram of the iFFL and a constitutive TX Marker to report plasmid dosage delivered to each cell. The hEF1a promoter is used to drive transcription of all genes. (b) Top row: TX Marker vs Output levels for each sample, overlaid with fits of the iFFL model. The CV(RMSE) is the root-mean-square error between the model and non-binned data, normalized by the mean of the data ($\log_{10}$-transformed first since the cell-to-cell variance is approximately log-normally distributed). Bottom row: histograms of the Output levels for cells within each color-coded bin (as indicated in the scatters). Data is representative and taken from the first experimental replicate. (c) Robustness of iFFL Output levels in more finely-sampled bins, computed in reference to the fit parameter $Y_{\max}$. The values were log-transformed before the calculation. Bins with robustness scores over 95% (which we define as adapted to DNA copy number) are highlighted in shades of blue. Individual replicates are shown separately. (d) Correlation between the range of DNA

copy numbers over which an iFFL variant is adapted and Z_{50}, which is proportional ϵ. The adaptation range is defined as the largest sum of the log-widths of contiguous adapted bins in a sample. (e) Median expression over time for UR and iFFL variants (including a miRNA-based iFFL for comparison). The absolute accumulated change is the sum of the absolute values of the $\log_2$ changes in median expression between time points, summed from 12 to 120 hours. (f) Comparison of fit parameters $Y_{\max}$ and Z_{50} over time for the iFFL variants (including a miRNA-based iFFL for comparison). Median values for each sample are shown in Supplementary Table 2. Fit parameters for each sample are shown in Supplementary Table 3. The results are from populations of cells gated positive for either Output or TX Marker.


a
Resource-coupled module
Output
Output
P
Shared resources
Loading
1
2
N
Competing modules
b
Resource-decoupled module
ERN
Output
EndoRNase
Output
P
P
Shared resources
Loading
1
2
N
Competing modules
c
Simplified iFFL schematic
endoRNase-based iFFL
ERN
R
Output
Loading
Competing module(s)
d
Expected behavior
Relative expression of output
Ideal
+ ERN
– ERN
Resource loading by competitors
e
Model system: resource loading by Gal4 TAs
Module 1: Constitutive gene
P CMV
Output1
Shared resources
Loading
P hEF1a
TX Marker
P hEF1a
Gal4 Marker
DNA Dose Correlated
Module 2: Gal4 activation of fluorescent output
P hEF1a
Gal4
AD
5xUAS
P minCMV
Output2
f
Gal4-VP64
CV(RMSE) = 0.155
Gal4-VP16
CV(RMSE) = 0.210
Gal4-Rta
CV(RMSE) = 0.244
Gal4-p65
CV(RMSE) = 0.225
Gal4-VPR
CV(RMSE) = 0.409
CMV:Output1 (MEFL)
Gal4 Marker (MEFL)


a
Library of circuit variants
Module 1
Promoter library
{P}
Output1
Loading
Shared resources
Loading
Module 2
Gal4
AD
PhEF1a
5xUAS
PminCMV
Output2
PhEF1a
TX Marker
Transfection into common cell lines
HEK-293
HEK-293FT
HeLa
CHO-K1
Vero 2.2
b
Nominal {P}:Output1
CMV
CMVi
hEF1a
hUBC
hUBCs
TK
RSV
SV40
hACTB
hPGKsd
hMDM2
hMDM2c
HEK293
HEK293FT
HeLa
CHOK1
Vero
c
HEK-293
HEK-293FT
HeLa
CHO-K1
Vero 2.2
{P}:Output1 Median (MEFL)
log2 Fold-Δ {P}:Output1
VP64
VP16
Rta
p65
VPR
d
Combinations of promoters and activators with the least effect on / strongest knockdown of {P}:Output1
HEK-293
HEK-293FT
HeLa
CHO-K1
Vero 2.2
Fold-Δ {P}:Output1

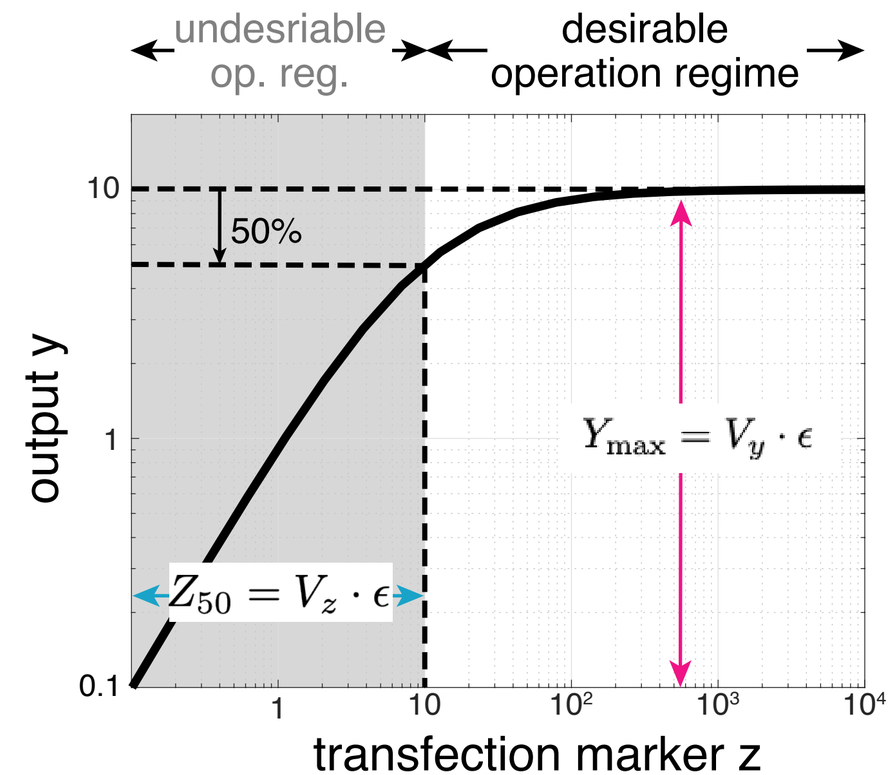
a



$$y = V_y \cdot \frac{u \cdot R}{1 + u \cdot R/\epsilon} \qquad z = V_z \cdot u \cdot R$$

$$y = V_y \cdot \frac{z/V_z}{1 + z/(V_z \cdot \epsilon)}$$

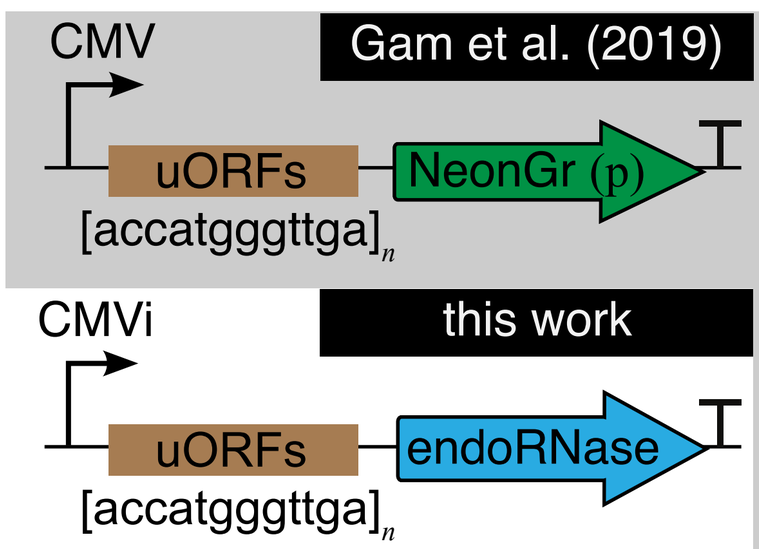
b



c

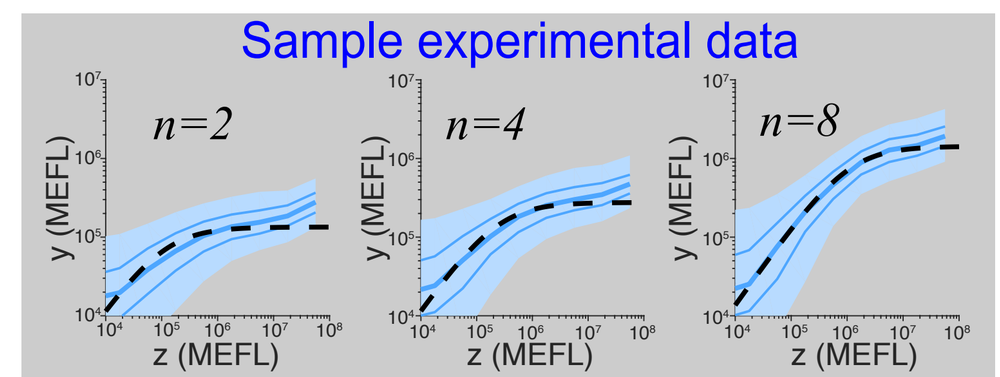
$\epsilon \propto \kappa_x$ by equation (2)

↑ # uORFs ⇨ ↓ TL initiation rate (↑ κ_x) ⇨ ↑ ϵ

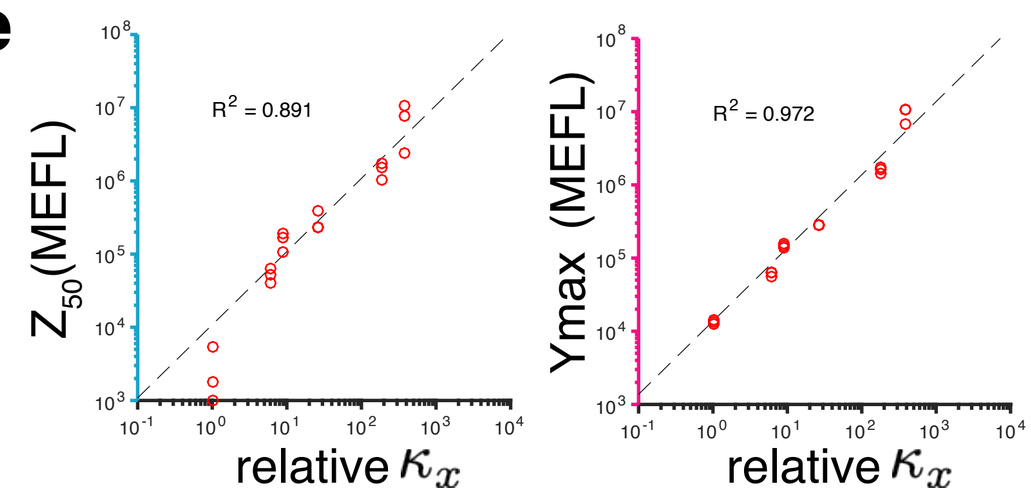


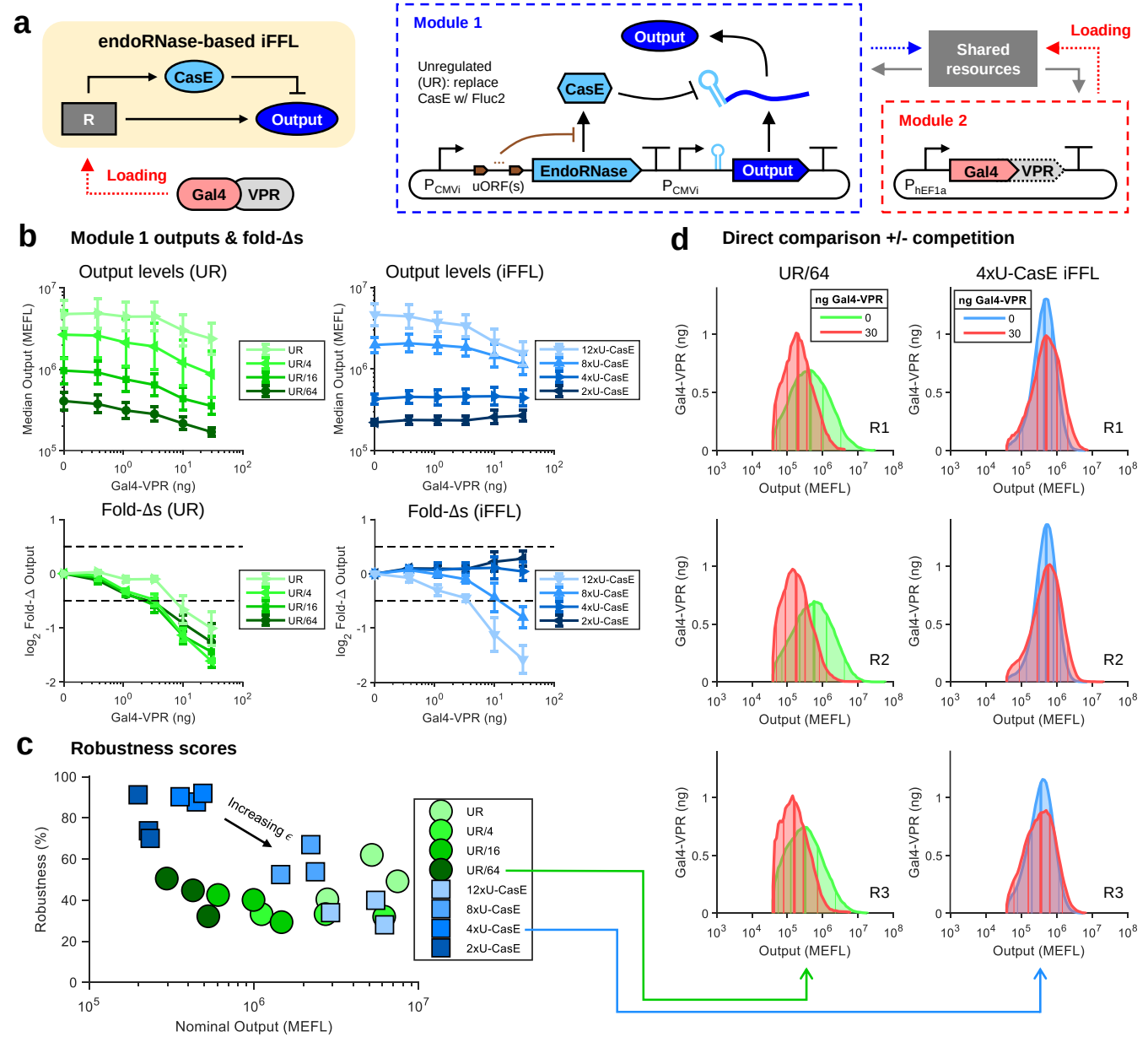
# uORFs (n)	relative κ_x := p(0)/p(n)
0	1
1	6.07
2	8.87
4	26.9
8	183
12	388

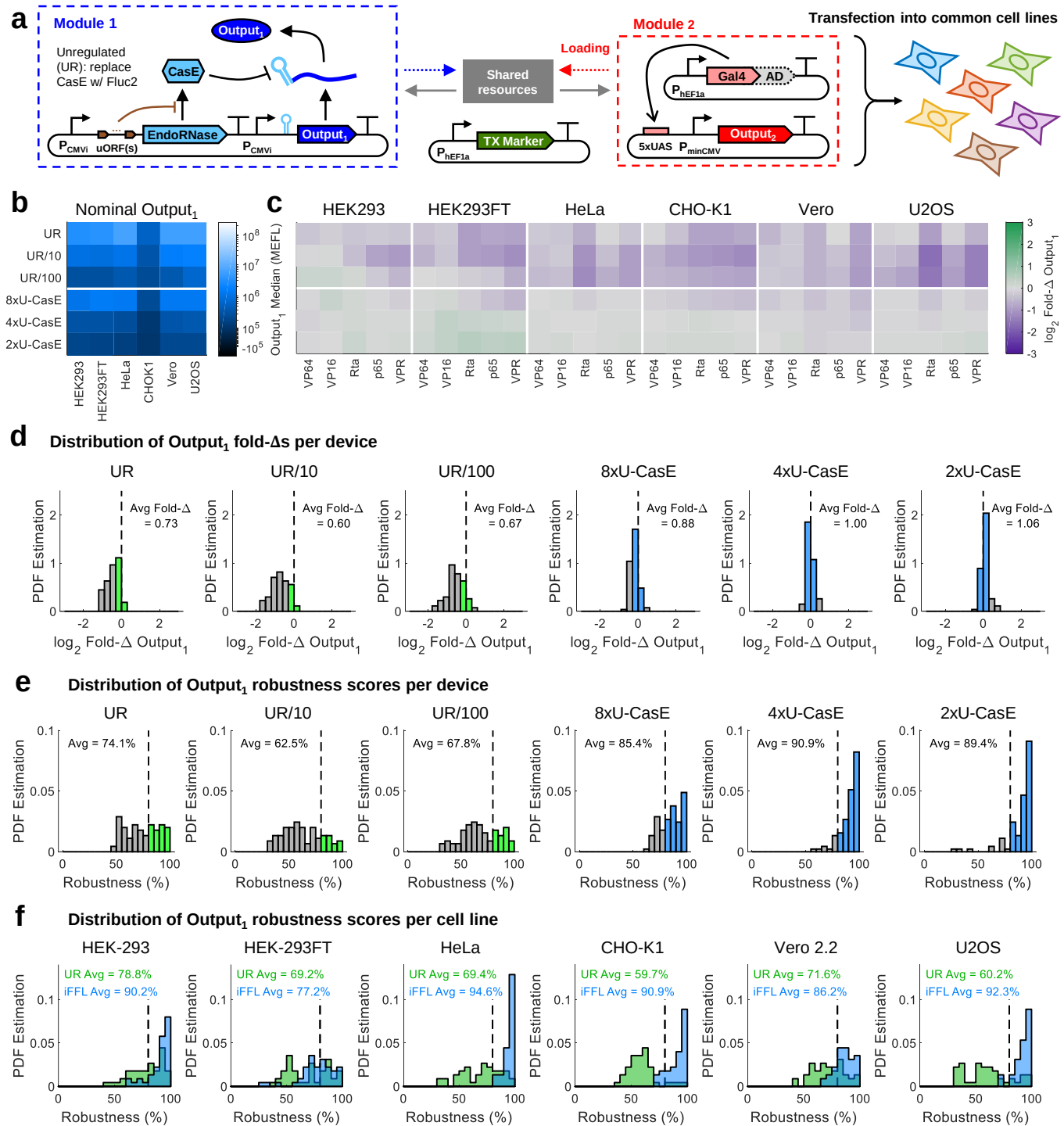
d

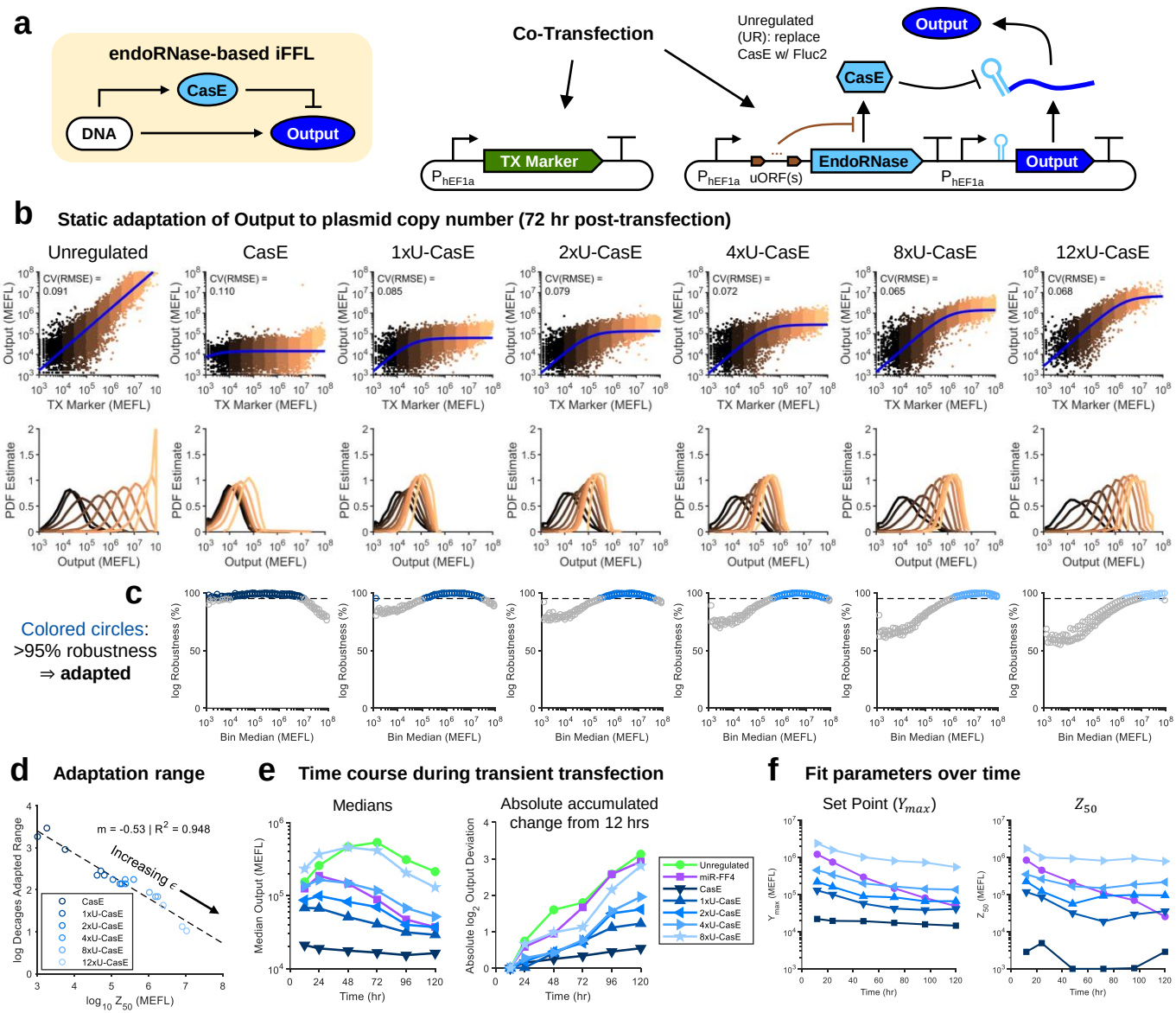


e




a
endoRNase-based iFFL
CasE
R
Output
Loading
Gal4
VPR
Module 1
Output
Unregulated (UR): replace CasE w/ Fluc2
CasE
EndoRNase
Output
P_CMVi
uORF(s)
P_CMVi
Shared resources
Loading
Module 2
Gal4
VPR
P_hEF1a
b
Module 1 outputs & fold-Δs
Output levels (UR)
Output levels (iFFL)
Median Output (MEFL)
Gal4-VPR (ng)
UR
UR/4
UR/16
UR/64
12xU-CasE
8xU-CasE
4xU-CasE
2xU-CasE
Fold-Δs (UR)
Fold-Δs (iFFL)
log2 Fold-Δ Output
d
Direct comparison +/- competition
UR/64
4xU-CasE iFFL
ng Gal4-VPR
0
30
Gal4-VPR (ng)
Output (MEFL)
R1
R2
R3
c
Robustness scores
Robustness (%)
Nominal Output (MEFL)
Increasing ϵ
UR
UR/4
UR/16
UR/64
12xU-CasE
8xU-CasE
4xU-CasE
2xU-CasE


a
Module 1
Unregulated (UR): replace CasE w/ Fluc2
CasE
Output1
PCMVi
uORF(s)
EndoRNase
PCMVi
Output1
Shared resources
Loading
TX Marker
PhEF1a
Module 2
Gal4
AD
PhEF1a
5xUAS
PminCMV
Output2
Transfection into common cell lines
b
Nominal Output1
UR
UR/10
UR/100
8xU-CasE
4xU-CasE
2xU-CasE
HEK293
HEK293FT
HeLa
CHOK1
Vero
U2OS
c
HEK293
HEK293FT
HeLa
CHO-K1
Vero
U2OS
Output1 Median (MEFL)
VP64
VP16
Rta
p65
VPR
log2 Fold-Δ Output1
d
Distribution of Output1 fold-Δs per device
UR
UR/10
UR/100
8xU-CasE
4xU-CasE
2xU-CasE
Avg Fold-Δ = 0.73
Avg Fold-Δ = 0.60
Avg Fold-Δ = 0.67
Avg Fold-Δ = 0.88
Avg Fold-Δ = 1.00
Avg Fold-Δ = 1.06
PDF Estimation
log2 Fold-Δ Output1
e
Distribution of Output1 robustness scores per device
UR
UR/10
UR/100
8xU-CasE
4xU-CasE
2xU-CasE
Avg = 74.1%
Avg = 62.5%
Avg = 67.8%
Avg = 85.4%
Avg = 90.9%
Avg = 89.4%
PDF Estimation
Robustness (%)
f
Distribution of Output1 robustness scores per cell line
HEK-293
HEK-293FT
HeLa
CHO-K1
Vero 2.2
U2OS
UR Avg = 78.8%
iFFL Avg = 90.2%
UR Avg = 69.2%
iFFL Avg = 77.2%
UR Avg = 69.4%
iFFL Avg = 94.6%
UR Avg = 59.7%
iFFL Avg = 90.9%
UR Avg = 71.6%
iFFL Avg = 86.2%
UR Avg = 60.2%
iFFL Avg = 92.3%
PDF Estimation
Robustness (%)


a
endoRNase-based iFFL
CasE
DNA
Output
Co-Transfection
Unregulated (UR): replace CasE w/ Fluc2
Output
CasE
TX Marker
P hEF1a
uORF(s)
EndoRNase
Output
b
Static adaptation of Output to plasmid copy number (72 hr post-transfection)
Unregulated
CasE
1xU-CasE
2xU-CasE
4xU-CasE
8xU-CasE
12xU-CasE
CV(RMSE) = 0.091
CV(RMSE) = 0.110
CV(RMSE) = 0.085
CV(RMSE) = 0.079
CV(RMSE) = 0.072
CV(RMSE) = 0.065
CV(RMSE) = 0.068
Output (MEFL)
TX Marker (MEFL)
PDF Estimate
Output (MEFL)
c
Colored circles: >95% robustness ⇒ adapted
log Robustness (%)
Bin Median (MEFL)
d
Adaptation range
m = -0.53 | R² = 0.948
Increasing ε
log Decages Adapted Range
log10 Z50 (MEFL)
CasE
1xU-CasE
2xU-CasE
4xU-CasE
8xU-CasE
12xU-CasE
e
Time course during transient transfection
Medians
Median Output (MEFL)
Time (hr)
Absolute accumulated change from 12 hrs
Absolute log2 Output Deviation
Unregulated
miR-FF4
CasE
1xU-CasE
2xU-CasE
4xU-CasE
8xU-CasE
f
Fit parameters over time
Set Point (Ymax)
Z50
Ymax (MEFL)
Z50 (MEFL)
Time (hr)