Transferring a synthetic gene circuit from yeast to mammalian cells

Dmitry Nevozhay[1,2], **Tomasz Zal**[3], and **Gábor Balázsi**[1,†]

[1]Department of Systems Biology–Unit 950, The University of Texas MD Anderson Cancer Center, 7435 Fannin Street, Houston, TX 77030, USA

[2]School of Biomedicine, Far Eastern Federal University, 8 Sukhanova Street, Vladivostok, 690091, Russia

[3]Department of Immunology and Department of Molecular and Cellular Oncology–Unit 108, The University of Texas MD Anderson Cancer Center, 7435 Fannin Street, Houston, TX 77030, USA

Abstract

The emerging field of synthetic biology builds gene circuits for scientific, industrial, and therapeutic needs. Adaptability of synthetic gene circuits across different organisms could enable a synthetic biology pipeline, where circuits are designed *in silico*, characterized in microbes and reimplemented in mammalian settings for practical usage. However, the processes affecting gene circuit adaptability have not been systematically investigated. Here we construct a mammalian version of a negative feedback-based "linearizer" gene circuit previously developed in yeast. The first naïve mammalian prototype was non-functional, but a computational model suggested that we could recover function by improving gene expression and protein localization. After rationally developing and combining new parts as the model suggested, we regained function and could tune target gene expression in human cells linearly and precisely as in yeast. The steps we have taken should be generally relevant for transferring any gene circuit from yeast into mammalian cells.

Keywords

negative feedback; synthetic biology; mammalian; human; inducible; linearizer; adaptability

INTRODUCTION

The emerging field of synthetic biology applies engineering principles to design and build biological systems for predefined purposes[1,2]. Promising applications such as biofuel

[†]Correspondence should be addressed to: Gábor Balázsi, Ph.D., Associate Professor of Systems Biology - Unit 950, The University of Texas MD Anderson Cancer Center, 7435 Fannin Street, Houston, TX 77054, Tel. +1-713-834-6157, gbalazsi@mdanderson.org.

production[3] and drug synthesis[4] have stimulated rapid advances in this field. To date, most progress has occurred in microorganisms[5–11], while mammalian synthetic biology has lagged behind. Synthetic mammalian constructs, including inducible gene expression systems[12–14], logic gates[15,16], memory devices[17,18], and genetic oscillators[19] have employed molecular machinery specific to higher eukaryotes[20,21]. While these examples suggest that similar functions could be achieved in microbes and mammalian cells, the principles and methodology of gene circuit transfer into mammalian cells are unclear. Adaptable gene circuits could enable practical applications in the life sciences and health care[22]. For example, finely tunable mammalian gene expression control systems could precisely relate protein levels to function, or could enable novel approaches to gene therapy.

Most commonly, mammalian gene expression is tuned by gene expression systems. They consist of a regulator whose control over the expression of a target gene depends on the inducer level in the growth medium. For example, synthetic transactivator-based Tet-On/ Off[23] and similar[24] systems have been widely utilized for gene expression control[25]. Yet, synthetic transactivators contain virus-derived activation domains[26] that can be toxic for mammalian cells[27], interfering with normal cell physiology, and compromising the reliable assessment of gene function[28]. Repressor-based gene expression systems, such as the T-REx[29] and LacI systems[30] avoid this problem since they lack viral activation domains. Yet, repressor-based systems can have sigmoidal dose-response and highly variable gene expression at intermediate levels of induction[31], making it difficult to bring all cells uniformly to defined intermediate levels of induction. These limitations create an unmet need to develop novel strategies for precise mammalian gene expression control[32,33]. One alternative, the ProteoTuner system[34], relies on posttranslational regulation–but requires fusing a destabilization domain to the gene of interest, which is not guaranteed to function seamlessly for all genes.

We have previously developed a new, negative feedback-based "linearizer" gene circuit in yeast (Fig. 1a) with two attractive features: linear dependence of average gene expression on extralular inducer concentration and uniform gene expression (low variability) across the cell population at all induction levels[35]. These characteristics arose as negative feedback adjusted repressor protein expression to a level that was *just enough* to repress its own gene. A subsequent increase of inducer concentration allowed new protein synthesis, but *only up to the level* capable to overcome the sequestering effect of additional inducer. Consequently, the repressor level became proportional to the inducer concentration. Moreover, the expression of an additional gene from an identical promoter also depended linearly on the level of inducer[35].

We hypothesized that these beneficial properties of the linearizer system could also be reproduced in mammalian cells using an identical circuit design. This is a challenging, yet rewarding goal since no inducible mammalian systems exist for linearly dose-dependent tuning of gene expression. Based on this rationale, we set out to create a novel mammalian gene expression system for linearly tunable and uniform gene expression across the cell population. If successful, this effort should lead to greatly improved gene expression control, while revealing crucial steps applicable for moving synthetic gene circuits from yeast into mammalian cells.

RESULTS

Performance metrics

Synthetic gene circuits are built for predefined purposes, e.g., to function as switches, logical gates or oscillators. To measure how well they fulfill these functions, performance metrics are needed. Possible performance metrics for linearizer gene circuits are (Fig. 1b): (i) the *fold induction* defined as the ratio of expression levels in the fully induced state (*maximum expression*) versus the fully repressed state (*background expression*); (ii) the *range of linearity* defined as the inducer concentration range where linearity holds; (iii) the *coefficient of variation* (*CV*) that measures the non-genetic variability of gene expression in the cell population[36] and (iv) the *degree of linearity* that evaluates the straightness of the dose-response between two inducer concentrations, using parametric linear regression (R^2) or the L1-norm[35]. The latter metric estimates the deviation of the measured dose-dependence from a perfectly linear relationship, and falls within the [0 … 0.5] range, with lower values indicating better linearity[35]. We applied these metrics to evaluate the performance of mammalian linearizer gene circuits. Importantly, they are not entirely independent of each other. For example, measuring the degree of linearity requires some reasonable fold induction (Fig. 1b).

The naïve prototype is non-functional in mammalian cells

To facilitate troubleshooting, we first approached the problem of gene circuit transfer naïvely, using the design and components of the simplest yeast-based linearizer. In yeast, the gene circuit with fewest components that still had linear dose response and low noise was a fluorescent TetR::yEGFP fusion that could repress its own promoter[35]. To mimic this design, we built mammalian prototype TG1 (Fig. 2a) by expressing the bifunctional *tetR::eGFP* gene from pCMV-2xtetO, the *Cytomegalovirus* pCMV promoter[37] modified with two *tetO2* sites downstream of the TATA-box[29] in the widely used human breast adenocarcinoma MCF-7 cell line[38]. As in yeast, we expected *tetR::eGFP* self-repression to depend on the inducer concentration in the growth medium.

We measured the level of gene expression in repressed and fully induced conditions (at 0 ng/ml and 250 ng/ml doxycycline, respectively) by flow cytometry in cells harboring the TG1 circuit stably integrated into the genome. In all clones the fold induction was marginal, and maximum fluorescence was indistinguishable from the autofluorescence of untransfected MCF-7 cells (Fig. 2h). Considering that pCMV-2xtetO is a strong promoter[29], we suspected that poor fold induction was due to inappropriate expression of the *tetR::eGFP* fusion, supported by previous evidence for suboptimal *tetR* expression in mammalian cells[39,40]. We also observed that the dose-response of reporter gene expression in the T-REx system (incorporating the original prokaryotic *tetR* gene without feedback, Fig. S1c) had a very short plateau, followed by an upslope at relatively low inducer concentrations. This is a known hallmark of compromised TetR expression in gene expression systems without feedback[41].

Computational modeling suggests optimization strategy

As mentioned above, measuring the degree of linearity of a gene circuit's dose-response requires some reasonable fold induction (Fig. 1b). To understand how we could improve *tetR::eGFP* expression and rescue fold induction, we adapted an earlier computational model of the linearizer gene circuit[35] (Fig. 3a), modifying its parameters to reflect possible differences between yeast and mammalian cells (Supplementary Note 1). The model still produced a linear relationship between the inducer concentration and TetR::EGFP expression in the rising portion of the dose-response curve (Fig. 3b). Next, we varied the parameters of the adapted model systematically to determine their individual effect on fold induction (Fig. 4a, b, c, d). The simulated dose-responses indicated that fold induction should be improvable through changes increasing *tetR::eGFP* expression and function, specifically (i) increasing the transcription rate (m, Fig. 4a); (ii) increasing the translation rate (p, Fig. 4b); (iii) decreasing mRNA degradation rate (μ, Fig. 4c); (iv) or increasing the repressor-DNA binding rate (r, Fig. 4d)

Following the suggestions from the model, we developed a three-stage strategy for restoring the function of the mammalian linearizer. First, at Stage 1 we planned to improve the efficiency of *tetR::eGFP* expression while minimizing background expression, to elevate fold induction. A summary of the changes applied to achieve this and the corresponding model parameters are listed in Table S1. A sufficiently large fold induction should allow testing the degree of linearity and level of noise at Stage 2. Finally, a target gene controlled by the same promoter could be introduced at Stage 3 to determine if linear dose-response and gene expression uniformity would transfer to additional genes as they did in yeast.

Intron insertion and codon optimization

We reasoned that *tetR::eGFP* expression was suboptimal because the gene was ill-adapted to its mammalian host cells. Considering that pCMV-2xtetO is a strong promoter[29], we decided to first concentrate our effort on improving translation (p) and mRNA stability (μ) of *tetR::eGFP* as suggested by the model. Seeking clues for gene circuit optimization, we looked for mammalian gene features less common in lower eukaryotes. One such feature is intron density[42]. Coincidentally, intron introduction into genes can increase their expression in mammalian cells[43,44]. The exact mechanism of this effect is not fully understood, but it is believed to be related to the enhancement of mRNA maturation and extranuclear transport by intron splicing[45,46]. This would increase mature mRNA levels, mimicking a decrease in the mRNA degradation rate μ in the computational model (Fig. 4c). Importantly, the same optimization was applied to *tetR* in the commercial T-REx system to improve its expression in mammalian cells. Considering all of the above, we introduced the rabbit β-globin intron II sequence[47] into the 5′ untranslated region (5′UTR) of the yeast-derived *tetR::eGFP* sequence, obtaining prototype TG2 (Fig. 2b). We confirmed by reverse transcription polymerase chain reaction (RT-PCR) that the intron was properly spliced out from the mRNA, and the *tetR::eGFP* coding sequence was intact (data not shown).

Another feature distinguishing mammalian and yeast gene sequences is synonymous codon usage bias. Adapting the codon bias to the host cell can improve heterologous gene expression in mammalian cells by improving translation efficiency[48]. Thus, we developed

prototype TG3 (Fig. 2c) from prototype TG2 by rebuilding the repressor-reporter fusion from the mammalian codon-optimized variants of *tetR* (*htetR*) and *eGFP* genes[23,49], with the aim of improving the translation rate *p* as suggested by the model (Fig. 4b).

To test the effect of these changes, we transiently transfected all three prototypes into MCF-7 cells and measured their fluorescence level after 2 days of incubation in saturating inducer concentrations (Fig. 5a). The TG3 prototype had the strongest maximal expression (median fluorescence level of 466 a.u. in transfected cells) compared to TG1 (76 a.u.) and TG2 (269 a.u.), respectively. This confirmed that the first pair of modifications improved *htetR::eGFP* expression.

Introducing a nuclear localization sequence

The computational model indicated that gene expression increase from altering mRNA content and translation rate occurred at the expense of increased background expression (Fig. 4b, c). However, the model also suggested a remedy, indicating that increasing the repressor-DNA binding rate *r* could compensate by lowering background expression while leaving the maximum expression unaltered (Fig. 4d). Since direct improvement of repressor-DNA binding affinity may require screening a large number of TetR mutants without a guarantee to find a variant with higher affinity, we decided to address this indirectly by introducing a nuclear localization sequence (NLS) into the *tetR::eGFP* coding sequence. While the addition of the NLS should not improve repressor-DNA binding affinity directly, it should increase nuclear repressor protein concentration. This amounts to altering the effective binding rate of TetR::EGFP to *tetO*2 sites in the promoter, mimicking an increase of the parameter *r* in the computational model.

To test this, we introduced the simian virus 40 (SV40) large-T-antigen NLS[50] into the middle of the TetR::EGFP protein, obtaining prototype TG4 (Fig. 2d). This should increase the effective binding rate of TetR::EGFP to the promoter by boosting the nuclear concentration of the repressor. Indeed, the NLS sequence caused preferential translocation of the TetR::NLS::EGFP protein into the nucleus in TG4 compared to the TG3 prototype (Fig. 5b) in bulk-selected MCF-7 cells stably expressing the genome-integrated TG3 and TG4 prototypes. Flow cytometry measurements in the absence and at saturating concentration of doxycycline (250 ng/ml) indicated that prototype TG4 had higher fold induction than TG3 (3.9 and 2.4, respectively), confirming the computational predictions on gene circuit performance (Fig. 5c).

Further changes in primary transcript sequence

We surveyed the literature, seeking additional modifications which could mimic the decrease in mRNA degradation rate *μ* suggested by the computational model. We found evidence that a particular sequence, the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) boosted heterologous gene expression in mammalian cells[51] when placed in the 3′ untranslated region (3′ UTR), presumably by enhanced mRNA polyadenylation, stabilization and extranuclear export. To test this possibility, we developed the TG5 prototype by introducing the WPRE into the 3′UTR of the *htetR::NLS::eGFP* transcript (Fig. 2e). Further, to optimize the translation rate *p* as suggested by the model, we

built prototype TG6 (Fig. 2f and Fig. S2) by converting the region around the ATG translation start codon to the consensus Kozak sequence, known to improve heterologous gene expression by enhancing translation in mammalian cells[52].

We tested how these changes influenced *htetR::NLS::eGFP* expression by transiently transfecting MCF-7 cells with the TG4, TG5, and TG6 prototypes using equal amounts of plasmid DNA (1 μg). Flow cytometry measurements after 2 days indicated a moderate improvement due to the WPRE sequence, while the Kozak sequence caused a dramatic gene expression increase (Fig. 5d; median fluorescence levels of 1124 a.u., 461 a.u., and 407 a.u., for TG6, TG5 and TG4, respectively).

Screening a novel library of TetR-repressible promoters

The modifications applied so far in prototypes TG1-TG6 most likely altered cellular mRNA content, translation and protein localization. At the same time, the transcription rate m remained unoptimized in earlier prototypes, despite computational predictions of its strong effect on fold induction (Fig. 4a). One way to optimize m would be to create novel promoter variants and replace the pCMV-2xtetO promoter, originally created by inserting two *tetO2* sites[29] between the TATA-box and the Initiator (Inr) motif of the widely used parental wild-type pCMV promoter[37] (Fig. 6a). To accommodate the two *tetO2* sites, the Inr motif was moved 54 bp downstream from its original position relative to the TATA box in the pCMV-2xtetO promoter[29] (Fig. 6a). We suspected that Inr displacement might have lowered maximum expression of the pCMV-2xtetO promoter, since in the wild-type pCMV promoter Inr was a transcription start site (TSS) crucial for efficient transcription[53]. To test this, we created a new promoter (pCMV-dInr), moving the Inr motif to exactly the same distance from the TATA-box as in the pCMV-2xtetO promoter, but replacing the two *tetO*2 sites with scrambled nucleotide sequences (Fig. S3a and Supplementary Note 2). Comparing *eGFP* expression from the wild-type pCMV, pCMV-2xtetO and pCMV-dInr promoters (Fig. S3b) in transiently transfected MCF-7 cells, we found that the latter two promoters (with displaced Inr motifs) had significantly lower eGFP expression than the wild-type pCMV promoter (Fig. S3c). In addition, we confirmed by reverse transcription polymerase chain reaction (RT-PCR, data not shown) that Inr ceased to be a TSS in pCMV-2xtetO, further confirming the suboptimality of the pCMV-2xtetO promoter structure.

In addition to these ambiguities regarding the Inr sequence, there was no clear justification for the number and position of the *tetO2* sites in the pCMV-2xtetO promoter. This is important, because increasing the number of *tetO2* sites could lower background expression in yeast[54], raising another possibility to improve fold induction.

Considering these uncertainties about the pCMV-2xtetO promoter structure and the computational suggestion to improve fold-change by increasing transcription efficiency (Fig. 4a), we set out to generate a library of newly synthesized promoters (Supplementary Note 2), which could be screened for fold induction improvements. First, to determine how the number and position of additional *tetO2* sites may affect fold induction in the context of the already characterized pCMV-2xtetO promoter, we inserted one or two additional *tetO2* sites either upstream (Fig. 6a, pCMV-C3) or downstream (Fig. 6a, pCMV-C4) of the TATA-box, regardless of the position of the Inr motif. Second, to determine whether restoring the

wild-type position of the Inr motif relative to the TATA box increased maximum expression, while also reducing background expression depending on the number of *tetO2* sites, we developed another set of promoters (Fig. 6a; pCMV-D2i, pCMV-D2t, pCMV-D3, pCMV-D4, and pCMV-D5) by introducing increasing numbers of *tetO2* sites into the wild-type pCMV, spaced such that they left the relative position of the TATA-box and Inr motifs intact.

We tested the full set of novel promoters, replacing the pCMV-2xtetO promoter with each of the 8 reengineered promoter versions in the context of the TG6 prototype. This resulted in 8 new intermediate prototypes that we transformed into MCF-7 cells. For each promoter variant we selected cells in bulk to obtain polyclonal populations stably expressing the chromosomally integrated prototype and then measured by flow cytometry the background and maximum expression of the selected populations in 0 and 250 ng/ml doxycycline, respectively. Contrary to expectations, we found that fold induction and maximum expression generally decreased with the number of *tetO2* sites in the promoter (Fig. 6b, c). In fact, the pCMV-2xtetO, pCMV-D2i, and pCMV-D2t promoters with only two *tetO2* sites conferred the highest fold induction (22.1 ± 6.5, 21.8 ± 5.4, and 25.5 ± 5.1, respectively). In addition, these promoters also had the highest maximum expression among all gene circuit variants tested so far (Fig. 6c), and therefore we selected them for in-depth assessment.

Next, we studied a set of clonal MCF-7 cell lines with stably genome-integrated constructs using the three promoters selected by preliminary screening. We expanded individual clones for each prototype (7 clones for the pCMV-2xtetO promoter, 12 clones for the pCMV-D2t promoter and 12 clones for the pCMV-D2i promoter) and assessed their fluorescence in both repressed and fully induced conditions. Flow cytometry measurements indicated that the pCMV-D2i promoter gave higher fold induction (34.0 ± 12.0) than either of the pCMV-2xtetO and pCMV-D2t promoters (Fig. 6d; 22.9 ± 7.2; $p = 0.042$ and 20.6 ± 6.2; $p = 0.004$ respectively, based on ANOVA). However, the maximal expression of these promoters did not differ significantly (Fig. 6c and Fig. S4), indicating that, at least in our settings, the pCMV-D2i promoter had optimal fold induction because of lower background expression rather than higher maximum expression.

The promoter screen concluded Stage 1, throughout which fold induction had gradually improved, from negligible induction in the naïvely built TG1 prototype to 46-fold induction in the best clones expressing the TG7 prototype (Fig. S5), approaching the yeast linearizer's performance[35]. The computational model guided these alterations causing gradual fold-change improvements. Finally, we selected the circuit with the pCMV-D2i promoter as the TG7 prototype (Fig. 2g) for further testing at Stage 2, due to its highest fold induction among all tested prototypes (Fig. 6d).

Linear dose-response of gene expression in prototype TG7

To start assessing further performance characteristics at Stage 2, we concentrated on the dose-response of a clonal MCF-7 cell line stably expressing the chromosomally integrated prototype TG7. Flow cytometry measurements at increasing inducer concentrations (0–25 ng/ml doxycycline) after 5 days of induction revealed a nearly linear dose-response (Fig. 7a, $R^2 = 0.99$, L1-norm $= 4.0 \times 10^{-2}$) up to approximately 71% (6 ng/ml doxycycline) of the

maximum expression (Fig. S6a). In addition, the gene expression distributions were remarkably narrow (Fig. 7b), barring a few non-reacting cells, indicating uniform, linearly tunable gene expression over most inducer concentrations (Fig. 7c). These findings echo the performance of the yeast linearizer[35], confirming its successful transfer from yeast to mammalian cells.

Linearized regulation of a second target gene

Finally, at Stage 3 we tested if dose-response linearity and gene expression uniformity can be transferred to another gene over a regulatory cascade as in yeast[35]. Thus, we introduced into the circuit the red fluorescent reporter *mCherry*, controlled by the same pCMV-D2i promoter and containing the same translational regulatory elements as the *htetR::NLS::eGFP* gene (Fig. 8a).

Flow cytometry measurements of an MCF-7 clonal cell line stably expressing this genome-integrated two-color linearizer after 5 days of induction at increasing concentrations of doxycycline (0–25 ng/ml) indicated a high fold induction in both parts of the circuit (30.8 and 38.5 for the *htetR::NLS::eGFP* and *mCherry* genes, respectively). Both dose-responses were nearly linear up to 60% and 63% of the maximum expression of *htetR::NLS::eGFP* ($R^2 = 0.99$, L1-norm = 3.1×10^{-2}) and *mCherry* ($R^2 = 0.99$, L1-norm = 3.1×10^{-2}), respectively (Fig. 8b and Fig. S6b), in contrast with the higher L1-norm values for the gene expression system without feedback (Fig. S6c). The dose-responses of average *htetR::NLS::eGFP* and *mCherry* expression were highly correlated (Fig. S6d; Pearson correlation coefficient r = 0.999, p < 0.0001). Finally, gene expression distributions at different levels of induction were almost uniformly narrow (Fig. 8c, d, e) over the range of inducer concentrations. These findings demonstrate that this new mammalian gene expression system can impart dose-response linearity and gene expression uniformity to another arbitrary gene of choice, making gene expression precisely tunable.

DISCUSSION

We achieved linearly inducer-dependent gene expression control and low gene expression variability in mammalian cells using a negative feedback-based gene circuit design identical to the one in yeast. These results confirm the adaptability of the yeast linearizer to mammalian cells without introducing any extra design features, solely by systematically optimizing parts responsible for efficient *gene expression* and *protein localization*. These two processes are at the heart of any synthetic gene circuit that employs regulators to control the expression of a target gene. Gene circuits will not function if a constituent gene is not expressed, or if a regulator lacks activity. Consequently, the steps we described should be relevant to transfer any synthetic gene circuit from microbes into mammalian cells. Even for gene circuits with more complex dynamics (such as bistable systems or oscillators) our strategy should enable adjustments to regain function lost in transfer across organisms. While some of these modifications were known to improve gene expression, their combined effect had not been tested quantitatively–especially in the context of synthetic gene networks, where the effects of modifications can interact in nontrivial manner, creating unpredictable behavior[54]. Overall, the novel parts we developed should greatly improve the

performance of already existing mammalian gene constructs. For example, since TetR is widely used in mammalian synthetic biology[29,55], our optimization steps can directly benefit other gene circuits employing this repressor, including T-REx.

Compared to the existing T-REx gene expression system[29], the main advantage of our design consists in allowing consistently precise and uniform gene expression control across a wide induction range, especially at intermediate levels of induction. For example, negative feedback-based systems such as the TG7 (Fig. 2g) or the two-gene linearizer (Fig. 8a) had very low L1-norm values (< 0.1) for almost the entire rising portion of their dose response curves (up to 70–80% of saturation, Fig. S6a, b), indicating linearity in that region. By contrast, the feedback-devoid T-REx system (Fig. S1a, b, c, d) started from higher L1-norm values and continued to rise, with the exception of an intermediate portion (Fig. S6c), indicating significant deviations from linearity. Moreover, contrary to the Tet-On/Off system[23], the mammalian linearizer does not require a viral activation domain with toxic effects in mammalian cells that could be mistakenly attributed to the studied transgene[56].

Besides its advantages, the mammalian linearizer system also has some limitations that should be considered for practical usage. Due to negative feedback regulation, it is expected to have somewhat higher background expression compared to systems based on repressor (T-REx) or transactivator (Tet-On) genes without feedback. This can limit the usage of the linearizer system for regulating strongly toxic genes. One way to decrease background expression in linearizer system could be to introduce a second repressor protein (such as LacI) or translational repression using the siRNA machinery[12]. We are currently working on implementing these improved linearizers with decreased background expression.

Our results support the 'abstraction principle' in synthetic biology, illustrating how the different parts of a biological circuit can be optimized for better functionality in new biological settings, while leaving the overall design of the system intact. They also suggest the possibility to implement a synthetic biology pipeline, in which circuits are designed *in silico*, tested and characterized in lower eukaryotes (benefiting from their relatively easy genetic modification) and finally reimplemented and optimized for functionality in mammalian settings for practical usage. For example, the mammalian linearizer could be utilized in advanced vectors for highly-needed[32,33], precisely controlled gene expression in individualized gene therapy. This promising direction for treating genetic disorders and cancer could benefit from precisely tuning the expression of genes with narrow therapeutic window such as rhodopsin[57] according the patient's condition and disease progression[22,32,33]. Moreover, linearizers could tune the expression of specific genes to reveal their effect on development, immune response, or nervous system response. Finally, precisely tunable gene expression circuits and their constituent parts can be building blocks for more complex mammalian synthetic gene systems[58], fostering progress in mammalian synthetic biology.

METHODS

Construction of plasmids

The plasmids used in this study were constructed using the pcDNA4/TO and pcDNA6/TR plasmids from the T-REx system (Invitrogen, Carlsbad, CA) and the pDN-G1TGt yeast plasmid carrying the original *tetR::eGFP* gene. The oligonucleotides we used can be found in Table S2, while the detailed description of plasmid construction can be found in the Supplementary Methods.

Cell lines and transfection

MCF-7 (human breast adenocarcinoma cell line) and HEK 293 (human embryonic kidney cell line) were obtained from the American Type Culture Collection. MCF-7 and HEK 293 cell lines were maintained in RPMI 1640 and DMEM (Mediatech, Manassas, VA) media, respectively, each supplemented with 5% certified tetracycline-free fetal bovine serum (Clontech, Mountain View, CA). Cells were cultured at 37°C, saturated with 5% CO_2. For dose-response experiments, growth media were supplemented with different concentrations of Doxycycline hydrochloride (Acros Organics, Geel, Belgium) and incubated with the inducer for 24–48 hours (transiently transfected cells) or 120 hours (stably expressing clones).

Plasmid DNA into the cells was introduced using the Amaxa Nucleofector device (Lonza, Walkersville, MD), according to manufacturer protocol, using $5–10 \times 10^6$ cells, 1–5μg of plasmid DNA, Solution V, and program P-20. In transient transfection experiments cells grew for 1–2 days before assessing them by flow cytometry. Plasmid DNA was linearized to obtain cell lines with stably genome-integrated gene circuits and cells were then selected using Zeocin™ (1000 μg/ml) or Blasticidin (6 μg/ml) drugs (Invitrogen, Carlsbad, CA) for 2–3 weeks.

For lentivirus bulk infection, virions were packaged into the HEK 293 cell line and then used for infection of the target cells following the manufacturer's protocol for the Lentiphos HT™ system and packaging plasmid mix (Clontech, Mountain View, CA). Target MCF-7 cells were then selected using 2 μg/ml of Puromycin (Clontech, Mountain View, CA) for 3 weeks.

Flow cytometry and cell sorting

Before flow cytometry cells were trypsinized and resuspended in fresh media. Then cells were either read on a BD FACScan flow cytometer (BD Biosciences, San Jose, CA) using the 488-nm argon excitation laser and 530/30 emission filter (EGFP) or read/sorted on a BD FACSAria II (BD Biosciences, San Jose, CA) using the 488-nm blue excitation laser and 530/30 emission filter for EGFP and the 561-nm yellow-green excitation laser and 610/20 emission filter for mCherry. At least 5000–6000 cells were typically collected. Control experiments showed lack of significant spillover between the EGFP and mCherry channels in two-color flow cytometry experiments (Fig. S7).

Fluorescence microscopy

For fluorescence microscopy cells were grown on Poly-D-Lysine coated coverslips for 2 days and fixed with 4% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA) for 30 min. The coverslips were washed twice with PBS for 5 min and stained with DAPI (1ug/ml) for 1 min. Images were acquired on a Nikon TE2000-E inverted fluorescence microscope (Nikon, Melville, NY) equipped with a CoolSNAP HQ^2 camera (Photometrics, Tucson, AZ), using a Nikon Plan Fluor 40x/1.30 Oil objective and B-2E/C filter (EX 465–495 nm, DM 505 nm, BA 515–555 nm) for EGFP and a UV-2E/C filter (EX 340–380 nm, DM 400 nm, BA 435–485 nm) for DAPI (both from Nikon, Melville, NY). Composite images with scale bars were prepared in NIS Elements (Nikon, Melville, NY). Finally, all images were cropped and brightened uniformly in Adobe Photoshop (Adobe Systems, San Jose, CA).

RT-PCR

Total mRNA was isolated from the cells using the QIAGEN RNEasy Mini Kit (QIAGEN, Germantown, MD). Reverse transcription was performed using the GoScript™ Reverse Transcription System (Promega, Madison, WI), according to the manufacturer's protocol. The cDNA of interest was then amplified using the primers: CMV-TSS+75-f, CMV-TSS-f and BGH-close-r (see Table S2 for oligonucleotide sequences).

Data processing and statistical analysis

Flow cytometry data was analyzed with FCS Express 3 (De Novo Software, Los Angeles, CA) and/or the flowCore package[59] in the R Project for Statistical Computing 2.13.1. Forward-scatter and side-scatter gates were used to minimize variation due to cell size, and a fluorescence-based gate was imposed to eliminate cells lacking gene circuits. One-way analysis of variance (ANOVA) and post-hoc Tukey HSD test in STATISTICA 9.1 (StatSoft Inc., Tulsa, OK) were used for statistical comparison of clonal cell lines with different promoters. Only inducible cell sublines (fold induction   2), derived from the clonal populations with at least 90% of expressing cells and without multiple plasmid integrations (double peaks) were selected for analysis.

We assessed linearity using the L1-norm[35] that varies from 0 (perfectly linear) to 0.5 (least linear). To determine the ranges of linearity for prototype TG7 (Fig. 2g), the two-gene mammalian linearizer (Fig. 8a) and the T-REx system (Fig. S1a), we calculated the L1-norm for increasing ranges of inducer concentration starting from 0 ng/ml of doxycycline up to the maximum concentration used in the experiments (Fig. S6a, b, c). Dose-response curves with less than 4 data points were not considered for L1-norm estimation. For each inducer dose range we obtained the L1-norm in three steps: (i) we rescaled the relevant inducer concentrations and fluorescence values to the [0..1] range; (ii) we interpolated this rescaled dose-response curve using the function *interp1* (piecewise cubic Hermite interpolating polynomial) from the **signal** package in the R Project for Statistical Computing 2.13.1 (http://www.r-project.org/); (iii) we calculated and reported the L1-norm as the area enclosed by the rescaled dose-response curve and a straight line connecting the coordinates (0,0) to (1,1), using the *trapz* function from the **caTools** package in R. In addition to

calculating the L1-norm, we also performed simple parametric linear regression to calculate R^2 as an alternative linearity metric using the R Project for Statistical Computing 2.13.1.

Computational modeling and parameter scans

We adapted an earlier computational model[35] changing parameters to account for biological differences between yeast and mammalian cells and implemented it in the software Dizzy[60]. We then used the computational model to study the effect of altering different parameters (the transcription rate m, the translation rate p, the mRNA degradation rate μ, and the effective promoter-TetR::EGFP binding rate of r) on the predicted values of fold induction, seeking clues on the biological processes that should be optimized first in linearizer circuit. The effect of these parameter scans was estimated for doxycycline concentrations ranging from 0 to 2000 ng/ml. The detailed description of the model, Dizzy code and parameters scans can be found in Supplementary Note 1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nat Rev Genet. 2010; 11:367–379. [PubMed: 20395970]

2. Purnick PE, Weiss R. The second wave of synthetic biology: from modules to systems. Nat Rev Mol Cell Biol. 2009; 10:410–422. [PubMed: 19461664]

3. Stephanopoulos G. Challenges in engineering microbes for biofuels production. Science. 2007; 315:801–804. [PubMed: 17289987]

4. Ro DK, et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 2006; 440:940–943. [PubMed: 16612385]

5. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R. A synthetic multicellular system for programmed pattern formation. Nature. 2005; 434:1130–1134. [PubMed: 15858574]

6. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000; 403:335–338. [PubMed: 10659856]

7. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000; 403:339–342. [PubMed: 10659857]

8. Stricker J, et al. A fast, robust and tunable synthetic gene oscillator. Nature. 2008; 456:516–519. [PubMed: 18971928]

9. Tabor JJ, et al. A synthetic genetic edge detection program. Cell. 2009; 137:1272–1281. [PubMed: 19563759]

10. Win MN, Smolke CD. Higher-order cellular information processing with synthetic RNA devices. Science. 2008; 322:456–460. [PubMed: 18927397]

11. Song H, Payne S, Gray M, You L. Spatiotemporal modulation of biodiversity in a synthetic chemical-mediated ecosystem. Nat Chem Biol. 2009; 5:929–935. [PubMed: 19915540]

12. Deans TL, Cantor CR, Collins JJ. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. Cell. 2007; 130:363–372. [PubMed: 17662949]

13. Karlsson M, Weber W, Fussenegger M. De novo design and construction of an inducible gene expression system in mammalian cells. Methods Enzymol. 2011; 497:239–253. [PubMed: 21601090]

14. Bleris L, et al. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. Mol Syst Biol. 2011; 7:519. [PubMed: 21811230]

15. Rinaudo K, et al. A universal RNAi-based logic evaluator that operates in mammalian cells. Nature Biotechnology. 2007; 25:795–801.

16. Leisner M, Bleris L, Lohmueller J, Xie Z, Benenson Y. Rationally designed logic integration of regulatory signals in mammalian cells. Nat Nanotechnol. 2010; 5:666–670. [PubMed: 20622866]

17. Burrill DR, Inniss MC, Boyle PM, Silver PA. Synthetic memory circuits for tracking human cell fate. Genes Dev. 2012; 26:1486–1497. [PubMed: 22751502]

18. Kramer BP, et al. An engineered epigenetic transgene switch in mammalian cells. Nat Biotechnol. 2004; 22:867–870. [PubMed: 15184906]

19. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M. A tunable synthetic mammalian oscillator. Nature. 2009; 457:309–312. [PubMed: 19148099]

20. Culler SJ, Hoff KG, Smolke CD. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science. 2010; 330:1251–1255. [PubMed: 21109673]

21. Xie Z, Wroblewska L, Prochazka L, Weiss R, Benenson Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. Science. 2011; 333:1307–1311. [PubMed: 21885784]

22. Ruder WC, Lu T, Collins JJ. Synthetic biology moving into the clinic. Science. 2011; 333:1248–1252. [PubMed: 21885773]

23. Urlinger S, et al. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. Proc Natl Acad Sci U S A. 2000; 97:7963–7968. [PubMed: 10859354]

24. Weber W, et al. Macrolide-based transgene control in mammalian cells and mice. Nat Biotechnol. 2002; 20:901–907. [PubMed: 12205509]

25. Lewandoski M. Conditional control of gene expression in the mouse. Nat Rev Genet. 2001; 2:743–755. [PubMed: 11584291]

26. Triezenberg SJ, Kingsbury RC, McKnight SL. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. Genes Dev. 1988; 2:718–729. [PubMed: 2843425]

27. Gossen M, Bujard H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A. 1992; 89:5547–5551. [PubMed: 1319065]

28. Morimoto M, Kopan R. rtTA toxicity limits the usefulness of the SP-C-rtTA transgenic mouse. Dev Biol. 2009; 325:171–178. [PubMed: 19013447]

29. Yao F, et al. Tetracycline repressor, tetR, rather than the tetR-mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. Hum Gene Ther. 1998; 9:1939–1950. [PubMed: 9741432]

30. Hu MC, Davidson N. The inducible lac operator-repressor system is functional in mammalian cells. Cell. 1987; 48:555–566. [PubMed: 3028641]

31. Kramer BP, Weber W, Fussenegger M. Artificial regulatory networks and cascades for discrete multilevel transgene control in mammalian cells. Biotechnol Bioeng. 2003; 83:810–820. [PubMed: 12889021]

32. Curtin JF, Candolfi M, Xiong W, Lowenstein PR, Castro MG. Turning the gene tap off; implications of regulating gene expression for cancer therapeutics. Mol Cancer Ther. 2008; 7:439–448. [PubMed: 18347132]

33. Guo ZS, Li Q, Bartlett DL, Yang JY, Fang B. Gene transfer: the challenge of regulated gene expression. Trends Mol Med. 2008; 14:410–418. [PubMed: 18692441]

34. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell. 2006; 126:995–1004. [PubMed: 16959577]

35. Nevozhay D, Adams RM, Murphy KF, Josić K, Balázsi G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. Proc Natl Acad Sci U S A. 2009; 106:5123–5128. [PubMed: 19279212]

36. Blake WJ, et al. Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell. 2006; 24:853–865. [PubMed: 17189188]

37. Boshart M, et al. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell. 1985; 41:521–530. [PubMed: 2985280]

38. Soule HD, Vazguez J, Long A, Albert S, Brennan M. A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst. 1973; 51:1409–1416. [PubMed: 4357757]

39. Gossen M, Bonin AL, Bujard H. Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. Trends Biochem Sci. 1993; 18:471–475. [PubMed: 8108860]

40. Zhang J, et al. A more efficient RNAi inducible system for tight regulation of gene expression in mammalian cells and xenograft animals. RNA. 2007; 13:1375–1383. [PubMed: 17616554]

41. Murphy KF, Adams RM, Wang X, Balázsi G, Collins JJ. Tuning and controlling gene expression noise in synthetic gene networks. Nucleic Acids Res. 2010; 38:2712–2726. [PubMed: 20211838]

42. Mourier T, Jeffares DC. Eukaryotic intron loss. Science. 2003; 300:1393. [PubMed: 12775832]

43. Brinster RL, Allen JM, Behringer RR, Gelinas RE, Palmiter RD. Introns increase transcriptional efficiency in transgenic mice. Proc Natl Acad Sci U S A. 1988; 85:836–840. [PubMed: 3422466]

44. Choi T, Huang M, Gorman C, Jaenisch R. A generic intron increases gene expression in transgenic mice. Mol Cell Biol. 1991; 11:3070–3074. [PubMed: 2038318]

45. Rose AB. Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol. 2008; 326:277–290. [PubMed: 18630758]

46. Huang MT, Gorman CM. Intervening sequences increase efficiency of RNA 3′ processing and accumulation of cytoplasmic RNA. Nucleic Acids Res. 1990; 18:937–947. [PubMed: 1690394]

47. van Ooyen A, van den Berg J, Mantei N, Weissmann C. Comparison of total sequence of a cloned rabbit beta-globin gene and its flanking regions with a homologous mouse sequence. Science. 1979; 206:337–344. [PubMed: 482942]

48. Nagata T, Uchijima M, Yoshida A, Kawashima M, Koide Y. Codon optimization effect on translational efficiency of DNA vaccine in mammalian cells: analysis of plasmid DNA encoding a CTL epitope derived from microorganisms. Biochem Biophys Res Commun. 1999; 261:445–451. [PubMed: 10425204]

49. Yang TT, Cheng L, Kain SR. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res. 1996; 24:4592–4593. [PubMed: 8948654]

50. Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. Cell. 1984; 39:499–509. [PubMed: 6096007]

51. Schambach A, et al. Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. Gene Therapy. 2006; 13:641–645. [PubMed: 16355114]

52. Kozak M. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 1987; 15:8125–8148. [PubMed: 3313277]

53. Juven-Gershon T, Cheng S, Kadonaga JT. Rational design of a super core promoter that enhances gene expression. Nat Methods. 2006; 3:917–922. [PubMed: 17124735]

54. Murphy KF, Balázsi G, Collins JJ. Combinatorial promoter design for engineering noisy gene expression. Proc Natl Acad Sci U S A. 2007; 104:12726–12731. [PubMed: 17652177]

55. Jones J, et al. Optimization of tetracycline-responsive recombinant protein production and effect on cell growth and ER stress in mammalian cells. Biotechnology and Bioengineering. 2005; 91:722–732. [PubMed: 15981277]

56. Morimoto M, Kopan R. rtTA toxicity limits the usefulness of the SP-C-rtTA transgenic mouse. Developmental Biology. 2009; 325:171–178. [PubMed: 19013447]

57. Olsson JE, et al. Transgenic mice with a rhodopsin mutation (Pro23His): a mouse model of autosomal dominant retinitis pigmentosa. Neuron. 1992; 9:815–830. [PubMed: 1418997]

58. Canton B, Labno A, Endy D. Refinement and standardization of synthetic biological parts and devices. Nat Biotechnol. 2008; 26:787–793. [PubMed: 18612302]

59. Hahne F, et al. flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics. 2009; 10:106. [PubMed: 19358741]

60. Ramsey S, Orrell D, Bolouri H. Dizzy: stochastic simulation of large-scale genetic regulatory networks (supplementary material). J Bioinform Comput Biol. 2005; 3:437–454. [PubMed: 15852514]
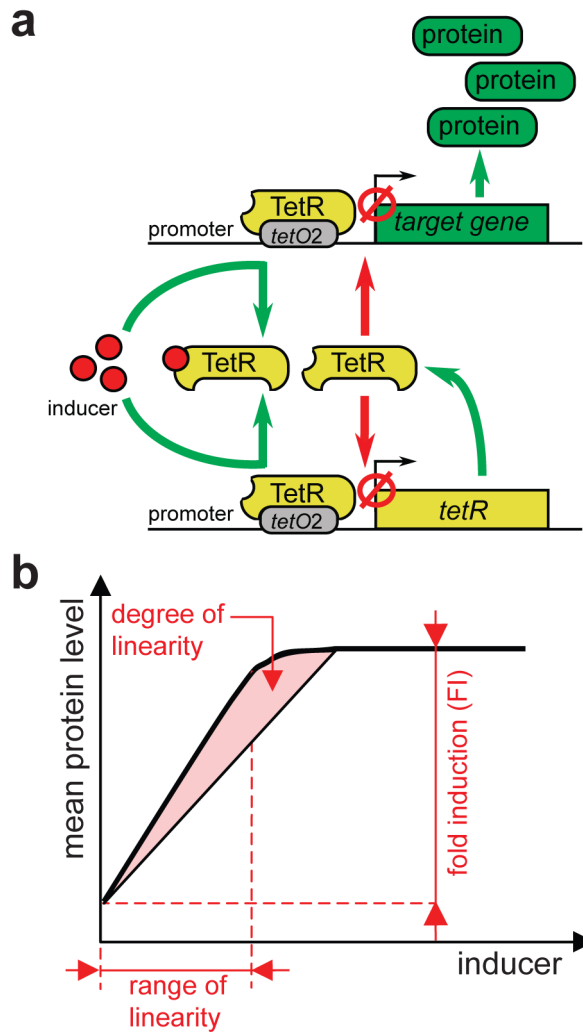
Figure 1. Linearizer gene circuits and their performance characteristics

(**a**) A linearizer gene circuit consists of the TetR repressor and an arbitrary target gene, both controlled by the same promoter. In the absence of inducer (tetracycline or its derivatives), the TetR protein binds to *tetO2* site(s) and physically blocks transcription from both promoters (red arrows). If inducer is added to the growth medium, it diffuses into cells and binds TetR, which dissociates from the *tetO2* sites (green arrows). Protein levels start to increase (green arrows) until TetR synthesis exceeds inducer influx and TetR blocks both promoters once again.

(**b**) Performance metrics of linearizer gene circuits based on the dose-response curve, defined as the average gene expression versus the inducer level (thick black line). The *fold induction* is the ratio of maximal and minimal (background) expression. The *range of linearity* covers the inducer concentrations where dose-response appears linear. The *degree of linearity* measures the straightness of dose-response between two inducer concentrations by the L1-norm (based on the shaded area). In addition, *gene expression variability* is measured by the coefficient of variation (CV, not shown).
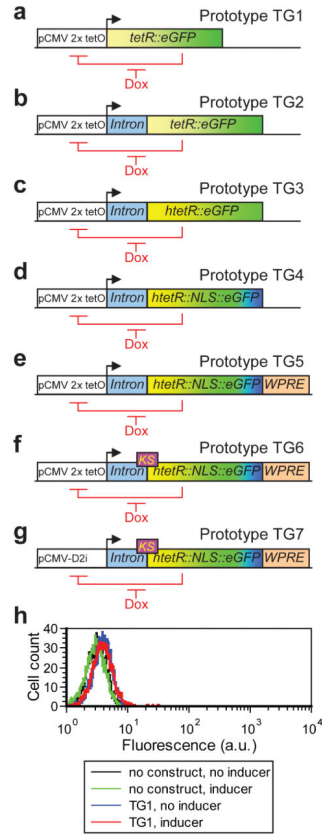
Figure 2. Initial deficiency of expression and subsequent linearizer prototypes

(**a**) Mammalian linearizer prototype TG1, naïvely constructed based on a bifunctional repressor-reporter fusion *tetR::eGFP* from the original yeast circuit.

(**b**) Prototype TG2, with an intron sequence introduced upstream of *tetR::eGFP*.

(**c**) Prototype TG3, with the human codon-optimized *htetR::eGFP* gene

(**d**) Prototype TG4, with a nuclear localization sequence (NLS) added between *htetR* and *eGFP*.

(**e**) Prototype TG5, with the Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) sequence introduced into the 3′ UTR of the mRNA.

(**f**) Prototype TG6, with the Kozak sequence (KS) introduced in the beginning of the *htetR::NLS::eGFP* gene.

(**g**) Prototype TG7, using the novel pCMV-D2i promoter.

(**h**) Gene expression distributions of MCF-7 cells stably expressing the chromosomally integrated prototype TG1 (see panel **c**) in 0 ng/ml doxycycline (blue) and 250 ng/ml doxycycline (red), relative to controls (black and green).
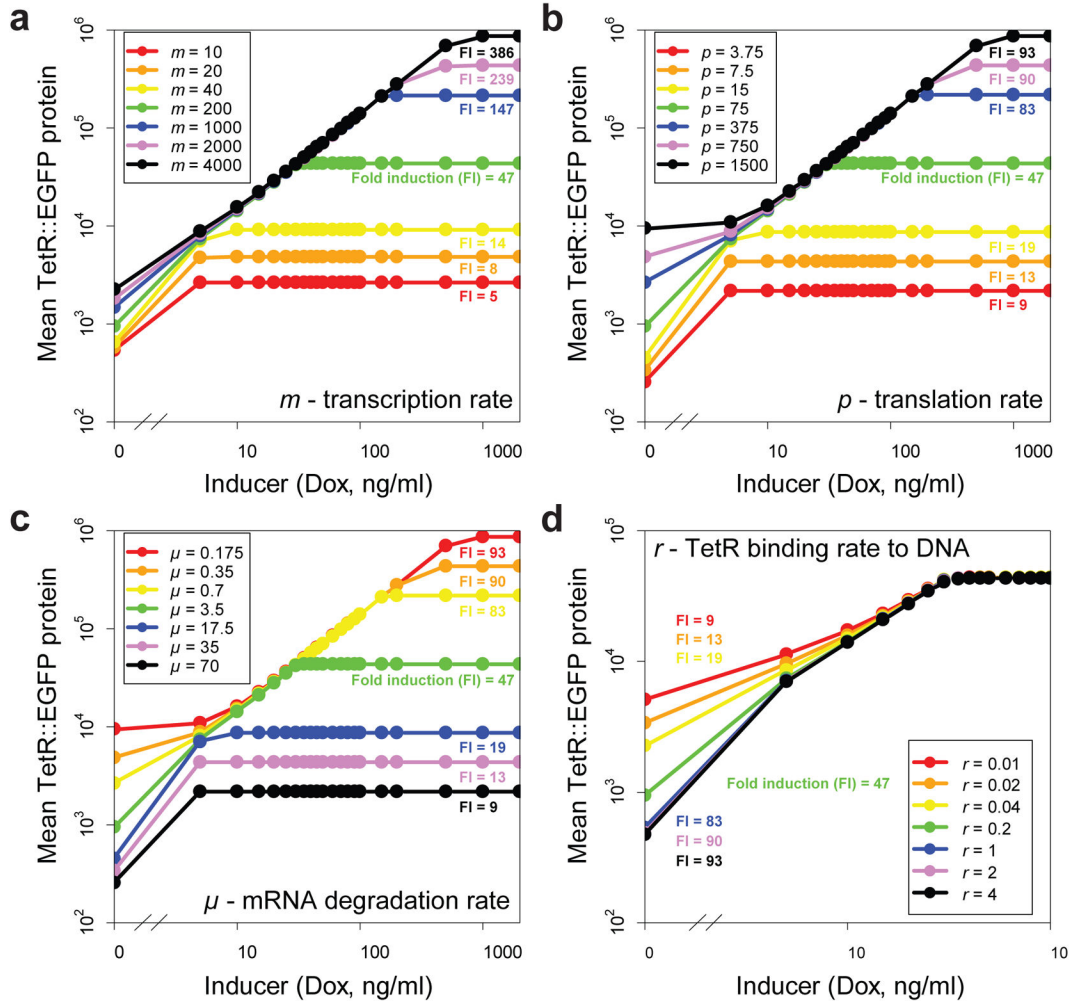
Figure 3. Computational model and simulated dose-response of the single-gene linearizer circuit
(**a**) Schematic representation of all chemical species, rates and reactions in the computational model.

(**b**) Simulated mean of *tetR::eGFP* gene expression at different levels of inducer (I_e, doxycycline concentration, 0 to 100 ng/ml).

Figure 4. The effect of parameter changes on the dose-response of mean TetR::EGFP protein levels

Green curves represent nominal values of the parameters in the simulations. Each parameter was varied 5, 10 and 20 times up and down from the nominal value to investigate how it affects fold induction. Simulated inducer (I_e, doxycycline concentration) levels were between 0 and 1000 ng/ml.

(**a**) Increasing the transcription rate (*m)* improves fold induction.

(**b**) Increasing the translation rate (*p*) improves fold induction.

(**c**) Decreasing the mRNA degradation rate (μ) improves fold induction.

(**d**) Increasing the TetR-DNA binding rate (*r*) lowers minimum expression, while leaving the maximum expression unaffected, overall improving fold induction.
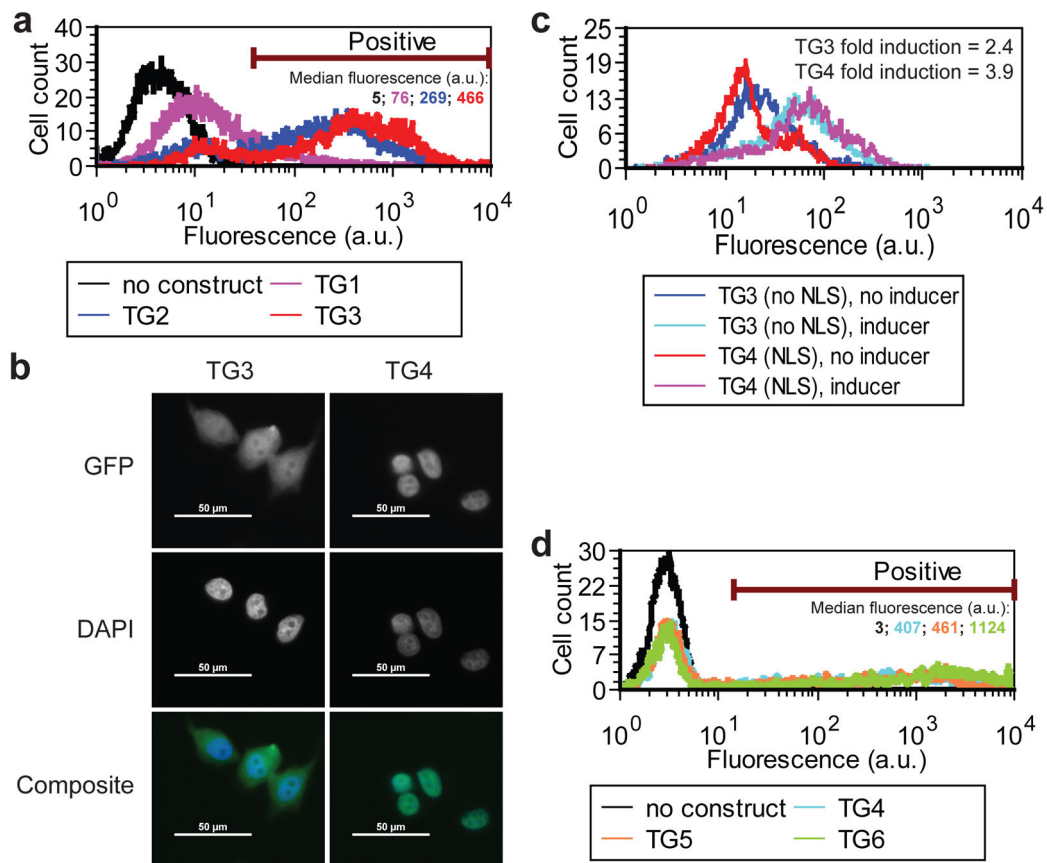
Figure 5. Improving fold induction in prototypes TG2 to TG6

(**a**) Gene expression distributions of MCF-7 cells transiently nucleofected with plasmids harboring prototypes TG1, TG2, and TG3 in saturating concentration of inducer (1000 ng/ml anhydrotetracycline). Median fluorescence was calculated for the cells carrying plasmid DNA. For control cells lacking plasmid DNA the same statistics was calculated based on the entire population.

(**b**) Fluorescent images of MCF-7 cells harboring genome-integrated prototypes TG3 (no NLS, panels on the left) and TG4 (with NLS, panels on the right). EGFP fluorescence overlaid with DAPI nuclear staining shows preferentially nuclear localization of TetR::NLS::eGFP in prototype TG4 compared to prototype TG3. The scale bar represents 50 μm.

(**c**) Gene expression distributions of MCF-7 cells stably expressing genome-integrated prototypes TG3 and TG4 in 0 ng/ml doxycycline (blue and red) and in 250 ng/ml doxycycline (cyan and magenta), indicating the corresponding fold induction.

(**d**) Gene expression distributions of MCF-7 cells transiently nucleofected with plasmids harboring prototypes TG4, TG5, and TG6 in saturating inducer concentration (250 ng/ml doxycycline). Median statistics was calculated as in (a).
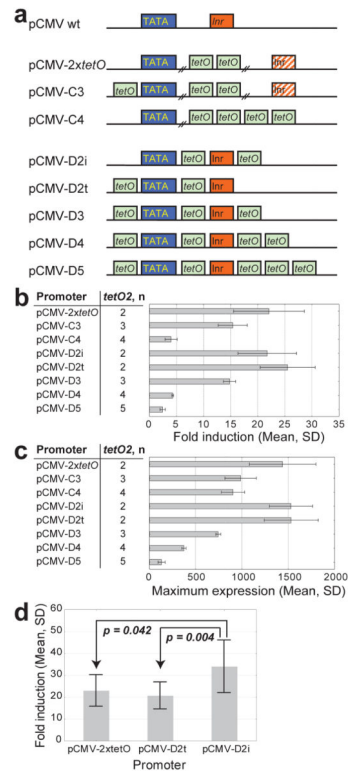
Figure 6. A library of novel TetR-repressible promoters

(**a**) A schematic representation of wild-type pCMV, pCMV-2xtetO (Invitrogen) and two sets of novel TetR-repressible promoters with different numbers and positions of *tetO2* sites and with (pCMV-D2i, pCMV-D2t, pCMV-D3, pCMV-D4, and pCMV-D5) or without (pCMV-C3 and pCMV-C4) wild-type distance between the TATA box and the Inr motif.

(**b**) Fold induction (mean ± SD, n=5) in MCF-7 cells bulk-transfected and stably expressing genome-integrated prototypes with the newly engineered promoters from (**a**).

(**c**) Maximum expression (mean ± SD, n=5, a.u., at 250 ng/ml doxycycline) in the same MCF-7 cells as in (**b**).

(**d**) Difference in fold induction between sets of clonal MCF-7 cell lines stably expressing TG prototypes based on the pCMV-2xtetO (mean ± SD, n = 7), pCMV-D2t (mean ± SD, n = 12), and pCMV-D2i (mean ± SD, n = 12) promoters (ANOVA, overall p = 0.003, followed by a Tukey HSD test: pCMV-D2i vs. pCMV-2xtetO, p = 0.042; pCMV-D2i vs. pCMV-D2t, p = 0.004).
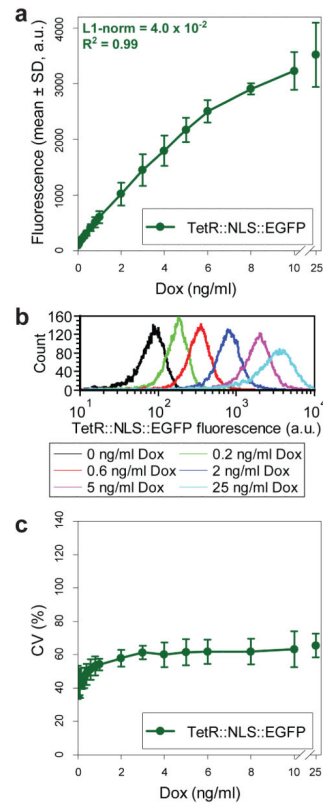
Figure 7. Selection and assessment of prototype TG7

(**a**) Dose-response curve averaged for three independent assessments of MCF-7 cells stably expressing the genome-integrated prototype TG7 at increasing concentrations of doxycycline inducer (mean ± SD).

(**b**) Representative gene expression distributions of MCF-7 cells stably expressing the genome-integrated prototype TG7 at different levels of induction.

(**c**) Variability of gene expression (CV) averaged for three independent assessments of MCF-7 cells stably expressing genome-integrated prototype TG7 at increasing concentrations of doxycycline inducer (mean ± SD).
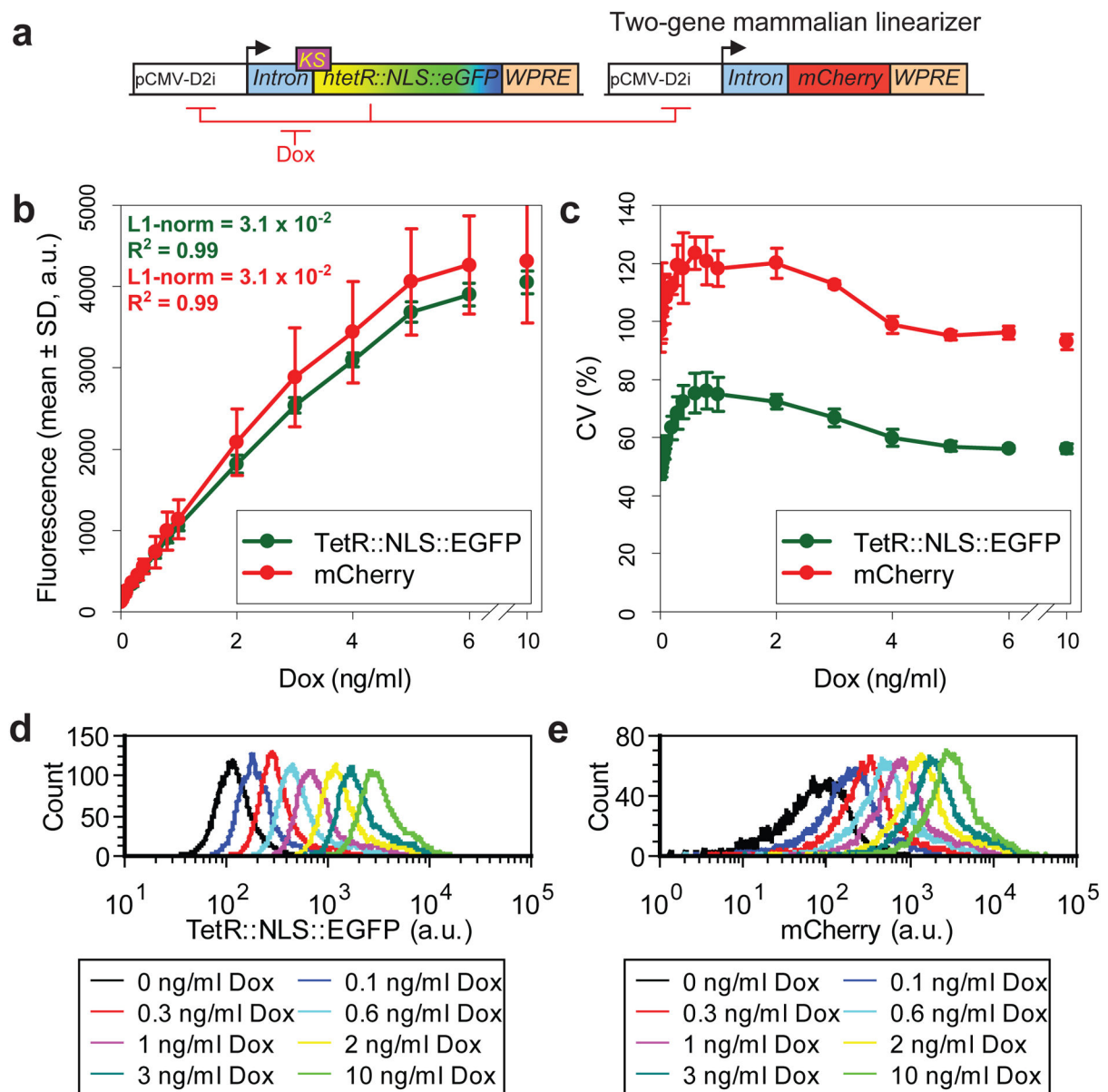
Figure 8. Two-gene mammalian linearizer system

(**a**) Two-gene mammalian linearizer based on the TG7 prototype driving the expression of the fluorescent reporter gene *mCherry*.

(**b**) Dose-response curves of *htetR::NLS::eGFP* and *mCherry* expression averaged for three independent measurements of MCF-7 cells stably expressing genome-integrated two-gene linearizer system at increasing concentrations of doxycycline inducer (mean ± SD).

(**c**) Variability (CV) of gene expression for *htetR::NLS::eGFP* and *mCherry* genes measured as described for panel (**b**) (mean ± SD).

(**d**) Representative distributions of hTetR::NLS::EGFP measured by flow cytometry for the same cells as in panels (**b**) and (**c**).

(**e**) Representative distributions of mCherry measured by flow cytometry for the same cells as in panels (**b**) and (**c**)