Supplementary Information

Transferring a synthetic gene circuit from yeast to mammalian cells

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

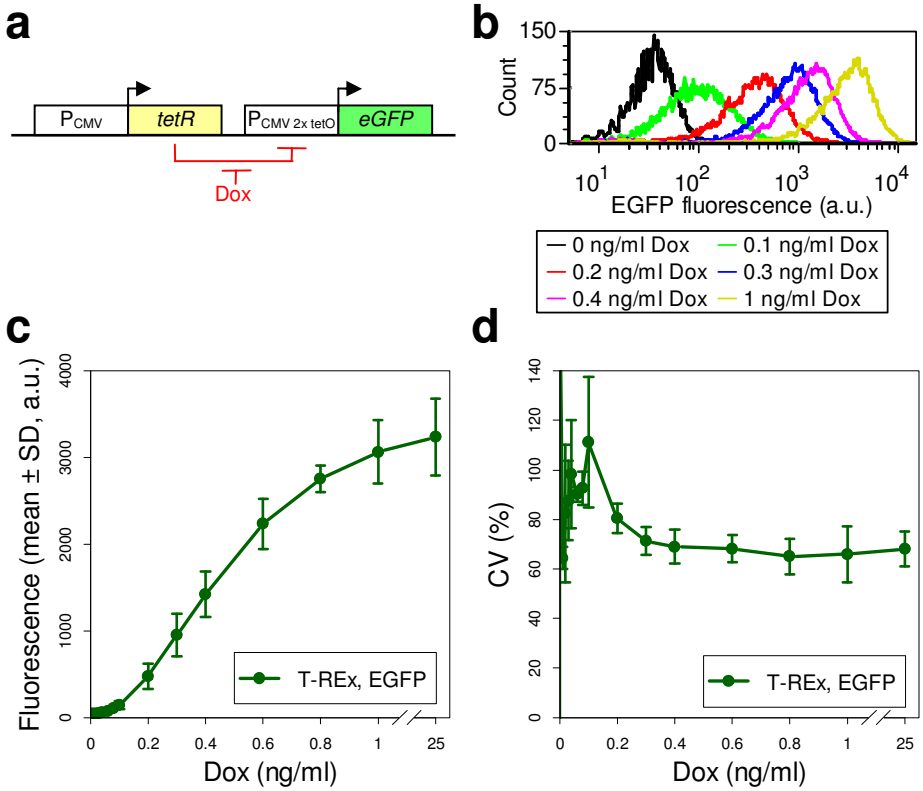
[1]Department of Systems Biology – Unit 950, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

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

[3]Department of Immunology – Unit 901 and Department of Molecular and Cellular Oncology – Unit 108, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[†] **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

SUPPLEMENTARY FIGURES



Supplementary Figure S1. The T-REx inducible gene expression system. (**A**) Schematic representation of the T-REx system with constitutively expressed TetR repressor protein, controlling the expression of a gene of interest (*eGFP* in this case) through a TetR-repressible promoter. TetR-mediated repression can be tuned by the level of external inducer concentration (tetracycline derivatives). (**B**) Gene expression distributions in MCF-7 cells carrying the stably integrated T-REx system. (**C**) Dose-response curve of the MCF-7 cells carrying the stably integrated T-REx system at different levels of induction. (**D**) Variability of gene expression (CV, %) of the MCF-7 cells carrying the stably integrated T-REx system at different levels of induction.

a -3..+9 region of *htetR* gene in the prototype TG5 (original)

GATATGTCTAGA

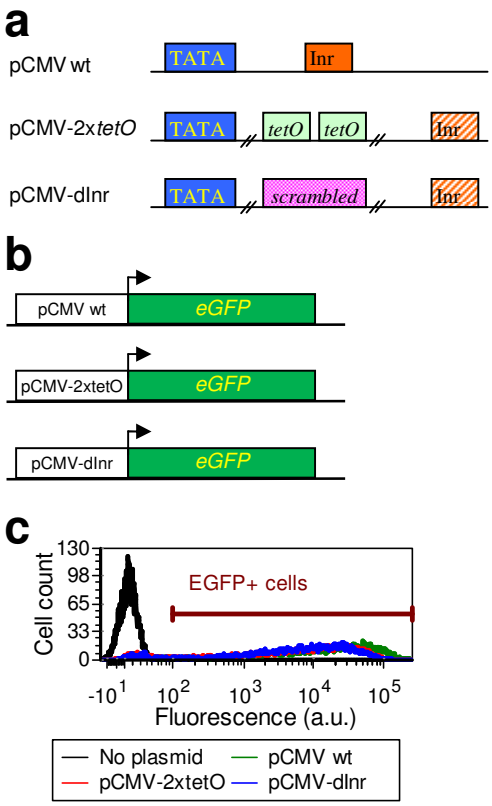
b Kozak consensus sequence

AccATGG

c Altered *htetR* gene (with introduced Kozak sequence), prototype TG6 and higher

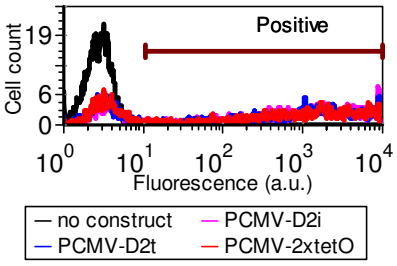
ACCATGGGTTCTAGA

Supplementary Figure S2. **Introduction of the Kozak sequence in the *htetR* gene.** (**A**) The original -3..+9 sequence of the *htetR* gene (blue) in prototype TG5 (**Fig. 2e**). (**B**) Mammalian Kozak consensus sequence[52]. (**C**) Altered translational start region of the *htetR* gene, indicating the nucleotides modified (red) to resemble the Kozak consensus sequence in prototype TG6 (**Fig. 2f**).



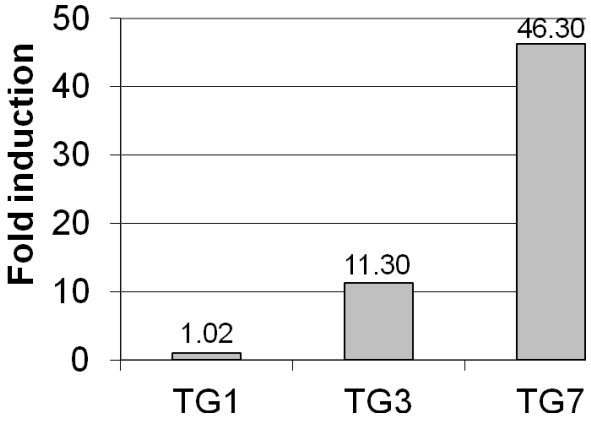
Promoter	Median fluorescence (a.u.) of positive cell
no construct	15
pCMV wt	17670
pCMV-2x*tetO*	9642
pCMV-dInr	8418

Supplementary Figure S3. **The influence of Inr motif displacement on expression from pCMV-based promoters**. (**A**) Schematic representation of the wild-type pCMV, pCMV-2xtetO (Invitrogen) and pCMV-dInr promoters. The latter has scrambled DNA sequence replacing the *tetO*2 sites in pCMV-2xtetO, such that the Inr motif is at the same distance from the TATA-box as in the pCMV-2xtetO promoter. (**B**) Schematic representation of three genetic constructs with *eGFP* expression controlled by each of the three promoters from (**A**). (**C**) Gene expression distributions measured by flow cytometry 2 days after MCF-7 cells were transiently nucleofected with plasmids bearing constructs from (**B**). Median statistics of gene expression were calculated for EGFP positive cells bearing the plasmid DNA (indicated by the horizontal bar). For control cells the same statistic was calculated from the entire population of cells lacking the plasmid.

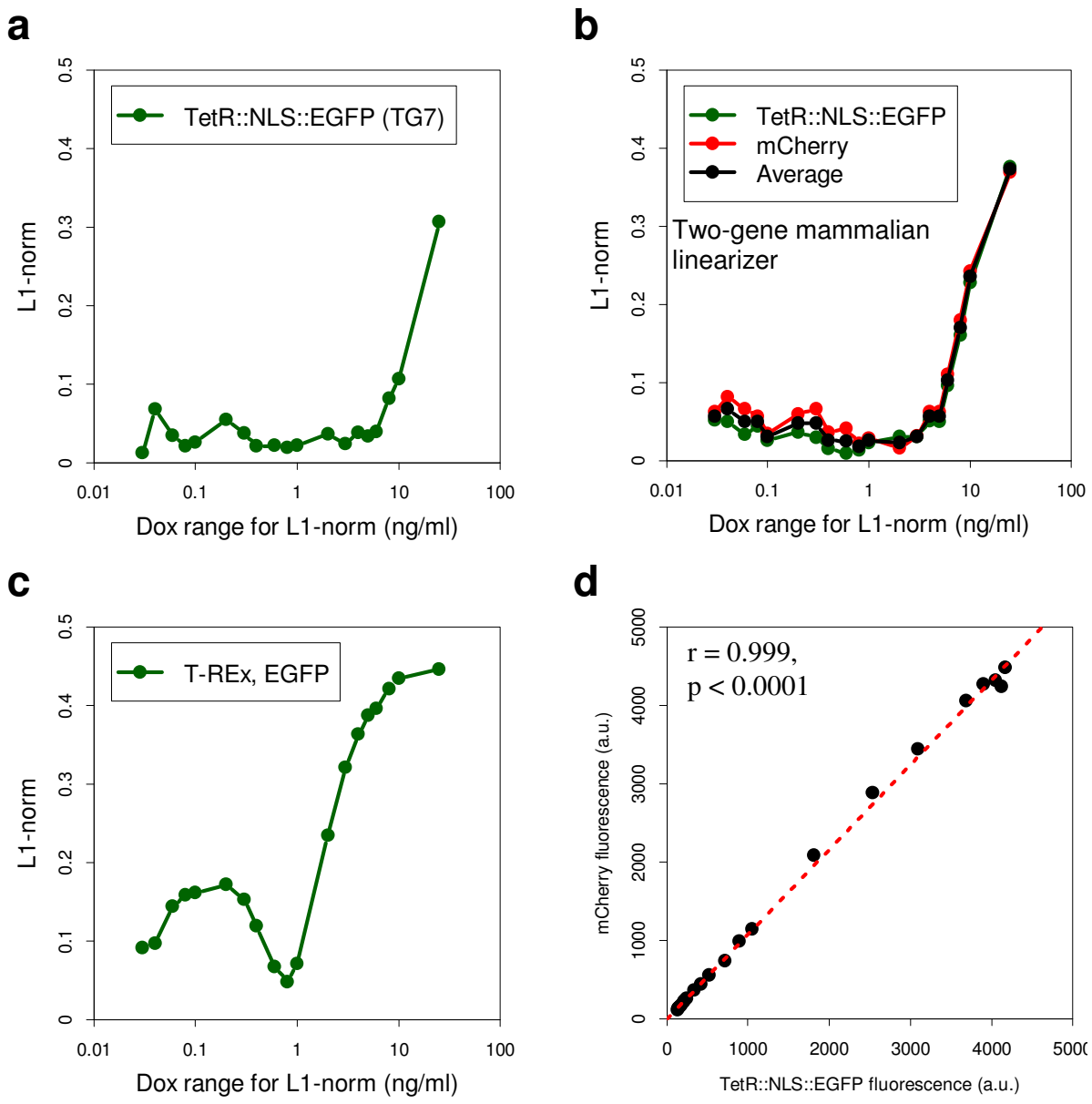


Promoter	Median fluorescence (a.u.) of positive cell
no construct	3
pCMV-D2i	1192
pCMV-D2t	1165
pCMV-2x*tetO*	1186

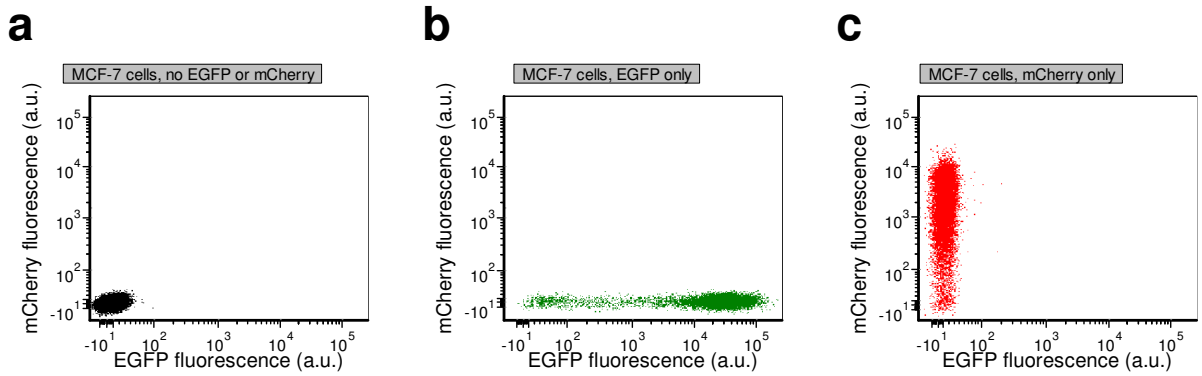
Supplementary Figure S4. **Comparison of expression efficiency from three TetR-repressible promoters in MCF-7 cells**. Gene expression distributions of MCF-7 cells transiently nucleofected with plasmids bearing TG prototypes based on the pCMV-2xtetO, pCMV-D2t, and pCMV-D2i promoters in the presence of saturating inducer concentration (Dox, 250 ng/ml). Median statistics of gene expression were calculated for positive cells bearing the plasmid DNA (indicated by the horizontal bar). For control cells the same statistic was calculated from the entire population of cells without the plasmid DNA.



Supplementary Figure S5. **Increase of fold induction in prototypes as the result of their optimization.** Three clonal MCF-7 cell lines with stably integrated TG1 (**Fig. 2a**), TG3 (**Fig. 2c**), and TG7 (**Fig. 2g**) prototypes were established and *tetR::eGFP* expression was assessed in repressed and fully induced states. Fold induction was calculated as the ratio of maximum to minimum expression. The results illustrate the increase of fold induction from prototype TG1 to TG7 after subsequent rounds of optimization.



Supplementary Figure S6. Linearity assessment and correlation of gene expression in the circuits. See the Materials and Methods section in the main text on how the L1-norm values were calculated for this figure. (**A**) L1-norm values calculated for all possible Dox ranges in prototype TG7 (**Fig. 2g**). (**B**) L1-norm values calculated for all possible Dox ranges in the two-gene mammalian linearizer (**Fig. 8a**). (**C**) L1-norm values calculated for all possible Dox ranges in the T-REx system (**Fig. S1a**). (**D**) The *mCherry* expression mean plotted against the *htetR::NLS::eGFP* expression mean in the two-gene mammalian linearizer (**Fig. 8a**) shows high correlation between protein synthesis in these two parts of the circuit on the population level (Pearson correlation coefficient r = 0.999, p < 0.0001).



Supplementary Figure S7. **Lack of spillover between the EGFP and mCherry fluorescence channels in flow cytometry experiments.** (**A**) MCF-7 cells with no EGFP or mCherry fluorescent protein expression. Fluorescence in both channels is at background. (**B**) MCF-7 cells expressing EGFP fluorescent protein only. EGFP-positive cells show strong signal in the EGFP channel, but stay at the same background level in the mCherry channel as EGFP-negative cells, illustrating lack of significant spillover of EGFP fluorescence into the mCherry channel. (**C**) MCF-7 cells expressing mCherry fluorescent protein only. mCherry-positive cells show strong signal in the mCherry channel, but stay at the same background level in the EGFP channel as mCherry-negative cells, illustrating lack of significant spillover of mCherry fluorescence into the EGFP channel.

SUPPLEMENTARY TABLES

Prototype	Genetic modification	Anticipated effect	Model parameter which suggested the modification
TG2	intron introduction	increased mature mRNA content	μ – mRNA degradation rate
TG3	gene codon optimization	improved translation	p – translation rate
TG4	nuclear localization sequence introduction	protein localization to the nucleus*	r – TetR-*tetO*2 binding rate
TG5	Woodchuck hepatitis virus post-transcriptional regulatory element	increased mRNA stability	μ – mRNA degradation rate
TG6	Kozak sequence introduction	improved translation	p – translation rate
TG7	novel pCMV-D2i promoter	improved transcription in the unrepressed state (maximum expression) and/or lower background expression in the repressed state**	m – transcription rate

* - as explained in the main text, the idea behind this modification was that, while the nuclear localization sequence would not increase the TetR-*tetO*2 binding rate *per se*, it would still concentrate the repressor predominantly in the nucleus, increasing the likelihood of its binding to the promoter in the repressed state.

** - as explained in the main text, the observed effects of this modification manifested mainly in lower background expression, resulting in higher fold induction of the pCMV-D2i promoter. However, in terms of maximum expression, the pCMV-D2i and pCMV-2xtetO promoters were comparable.

Supplementary Table S1. Computational model parameters varied in computer simulations and respective genetic modifications in the circuit effect of which was predicted by the model.

Primer name	Sequence (5' -> 3')
2602-Add-XhoI-f	TGTCCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCG
BGH-close-r	AGATGGCTGGCAACTAGAAGGCACAG
CMV-AgeI-f	GCGCACCGGTGTTGACATTGATTATTGAC
CMV-AflII-mamma-f	TAAACTTAAGCTTGGTACCCGGGGATCCTC
CMV-AflII-r	GCGCCTTAAGTTTAAACGCTAGAGTCCG
CMV-AseI-f	AGTTATTAATAGTAATCAATTACGGGGTCATTAG
CMV-dInr-f	GTGTGCGTGTGTGGCTGTGGTCTTTCTGGGCGCGGTCGCTTCCTCCTG
CMV-dInr-Inr-f	TCTGGGCGCGGTCGCTTCCTCCTGGAGCTCGTTTAGTGAACCGTCAGATC
CMV-dInr-r	CAGGAGGAAGCGACCGCGCCCAGAAAGACCACAGCCACACACGCACAC
CMV-dInr-TATA-r	AAGACCACAGCCACACACGCACACGAGCTCTGCTTATATAGACCTCCC
CMV-Ench-TATA-f	CTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
CMV-Ench-TATA-r	CCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAG
CMV-SpeI-f	GCGCACTAGTTATTAATAGTAATCAATTACG
CMV-TSS+75-f	CCACGCTGTTTTGACCTCCATAGAAG
CMV-TSS-f	CCTATCAGTGATAGAGATCGTCGACG
CMVwt-TATA-f	TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATC
CMVwt-TATA-r	AGGCGATCTGACGGTTCACTAAACGAGCTCTGCTTATATAGACCTCCC
es-WPRE-r	TGATTGCGGCCGCACCGGTTTACTTGTACAGCTCGTCCATGCCG
es-WPRE-f	AAGTAAACCGGTGCGGCCGCAATCAACCTCTGGATTACAAAATTTG
hEGFP-BamHI-f	GCGCGGATCCATGGTGAGCAAGGGCGAGGAG
hEGFP-NheI-r	GCGCGCTAGCTTACTTGTACAGCTCGTCCATGC
hEGFP-PacI-link-f	TGGTTTAATTAACATGGTGAGCAAGGGCGAGGAGCTG
hEGFP-WPRE-f	GGCATGGACGAGCTGTACAAGTAAAATCAACCTCTGGATTACAAAATTTG
hEGFP-XhoI-r	GCGCCTCGAGTTACTTGTACAGCTCGTCCATGC
hmCherry-AgeI-r	GCGCACCGGTCTACTTGTACAGCTCGTCCATGC
hmCherry-IVS-r	GTTATCCTCCTCGCCCTTGCTCACCATGGTGGCGCCTGCAGGACCTGTAG
htetR-IVS-f	CGACTCACTATAGGGCGAATTGATATGTCTAGACTGGACAAGAGCAAAG
htetR-NLS-EcoRI-r	GCGCGAATTCTTACTTTCTCTTCTTTTTTGGCCCGCCGCTTTCGC
htetR-NLS-NheI-r	CACTGCTAGCTTTCTCTTCTTTTTTGGCCCGCCGCTTTCGCACTTTAGC
htetR-NLS-PacI-link-r	CATGTTAATTAAACCAGCACCGTCACCCTTTCTCTTCTTTTTTGGCC
htetR-PacI-linker-r	CATGTTAATTAAACCAGCACCGTCACCCCCGCCGCTTTCGCACTTTAGC
htetR-XhoI-r	GCGCCTCGAGTTAATAAGATCTGAATTCCCGGGA
IVS-hmCherry-f	CTACAGGTCCTGCAGGCGCCACCATGGTGAGCAAGGGCGAGGAGGATAAC
IVS-htetR-r	CTTTGCTCTTGTCCAGTCTAGACATATCAATTCGCCCTATAGTGAGTCG
Kid-NheI-f	GAAAGCTAGCAGTGTCAGTGACATTTGAAG
Kid-EcoRI-r	ATCTGAATTCCCGGGATCCTTATTACCAGGGATCCTCTCCTTGCTGCAAC
rtTA-BamHI2-f	GCGCGGATCCATGTCTAGATTAGATAAAAGTAAAG
sk-htetR-r	GAACCCATGGTGGCGCCTGCAGGACCTGTAGGAAAAAGAAGAAGGCATG
sk-htetR-f	ACAGGTCCTGCAGGCGCCACCATGGGTTCTAGACTGGACAAGAGCAAAG
SV40-Seq-r	AATGCAGTGAAAAAAATGCTTTATTTGTG
WPRE-XhoI-r	GCGCCTCGAGGACAACACCACGGAATTGTCAG

Supplementary Table S2. Oligonucleotides used in this study.

SUPPLEMENTARY NOTE 1

Computational model of the negative feedback-based single-gene circuit. We used a computational model previously developed for the negative feedback-based synthetic gene circuit in yeast[35]. We modified several parameters to adapt this model for the mammalian circuit, based on the literature as follows (**Fig. 3a**).

Contrary to earlier yeast studies we used the inducer doxycycline (Dox) in mammalian cells rather than anhydrotetracycline (aTc) in all dose-response experiments. Both are tetracycline analogs that regulate the DNA binding affinity of the TetR protein, but Dox has ~7 times smaller binding constant to TetR as compared to aTc[68]. Therefore, we decreased the inducer-repressor binding rate (b) 7-fold in the mammalian model.

Second, we increased the rate of Dox entry into the cell by increasing the effective inducer concentration constant modifier parameter (c), considering that inducer diffusion is most likely higher in mammalian cells than yeast, due to the lack of a cell wall in the former. We have not changed the rate of inducer diffusion into the cells (f) explicitly because it is a compound parameter influenced not only by permeability of the cell membrane, but also by the area-to-volume ratio of the cell[69]. The influence of these different factors could be approximated by this relationship:

$$f \sim P\frac{A}{V}, \qquad [S1]$$

where P is the permeability of the cell membrane, A is the cell membrane area and V is the cell volume[69] . The parameter P is expected to increase in mammalian cells due to the lack of cell wall, while the A/V ratio is most likely decreasing, because of the generally larger size of mammalian cells compared to yeast. Due to the lack of data on the relative magnitudes of these effects and considering that the rate of inducer diffusion into the cells is critical for linearity[35], we decided to leave f unchanged.

Finally, we changed the parameter accounting for protein degradation and dilution, considering that mammalian cells divide much slower than yeast. In the yeast gene circuit protein removal was mainly attributed to dilution by cell growth, considering that the cell division time in yeast (~ 2 h) is much shorter than the EGFP half-life (~26 h)[35,70]. In contrast, MCF-7 cells have a doubling time of ~36h[71,72]. Consequently, the combined degradation/dilution rates (δ) of the TetR::EGFP and EGFP proteins in mammalian cells should be much lower, and can be calculated using the formula:

$$\delta = \frac{\ln(2)}{G} + \frac{\ln(2)}{D}\ , \qquad [S2]$$

where G and D are the cell doubling time and EGFP protein half-life, respectively. For simplicity, we also assumed that TetR::EGFP protein half-life was approximately equal to EGFP half-life. With G = 36 h and D = 26 h, we obtained and used $\delta = 0.046\ h^{-1}$ in the model.

To obtain estimations of average gene expression in the population, we implemented a computational model in the software Dizzy[60] and then performed deterministic simulations (to obtain the steady state solution of the ODE corresponding to the chemical reaction system) using ODEtoJava-dopr54-adaptive algorithm. The results from the simulations were then processed in

the R Project for Statistical Computing 2.13.1 (R Development Core Team, 2011). The inset below shows the full Dizzy code with parameter descriptions and initial values.

After adapting this model for mammalian cells, we simulated the inducer dose-responses of the mean in the gene circuit by varying the parameter I_e (extracellular inducer concentration) from 0 to 100 ng/ml of Dox and plotting the mean of simulated TetR::EGFP expression. As expected, the model showed linear relationship between the inducer concentration and TetR::EGFP expression in the rising portion of the dose-response curve (**Fig. 3b**).

We used the model to check the effect of different genetic alterations in the circuit on its fold induction. The simulations showed that increasing either the transcription (m) or translation rate (p), or decreasing the mRNA degradation rate (μ) should increase the fold induction due to the higher TetR::EGFP expression levels (**Fig. 4a,b,c**). The effect of these parameter changes was estimated for Dox concentrations spanning a range from 0 ng/ml to 2000 ng/ml. These insights from the computational simulations suggested biological processes that should be optimized in linearizer circuit.

In addition, we simulated the effect of adding nuclear localization sequence (NLS) to TetR::EGFP protein by varying the effective binding rate of TetR::EGFP to the promoter (r). The assumption was that by adding NLS to the TetR::EGFP protein we are not increasing its actual DNA binding rate *per se*, but rather preferentially concentrating the repressor in the nucleus, where it has higher chance to bind its target *tetO2* sites, therefore indirectly increasing the effective binding rate. The simulations showed that fold induction grew with the increase of r, due to lower gene expression leakage in the fully repressed state (**Fig. 4d**). These results suggested that adding a NLS to the TetR::EGFP protein should have beneficial effect on fold induction in linearizer gene circuit.

Finally, we also performed stochastic simulations based on the Gillespie algorithm in Dizzy[60] using the computational model to estimate noise of gene expression in linearizer circuit. Noise was not significantly affected by changes in any parameters (data not shown). The Dizzy code for the computational model of the negative feedback-based single-gene circuit is shown on the next page:

```
b = 15;             // binding rate of inducer to TetR::EGFP
lk = 2;             // TetR::EGFP basal transcriptional rate (leakage)
m = 200;            // TetR::EGFP max transcriptional rate
mu = 3.5;           // TetR::EGFP mRNA degradation
p = 75/2;           // TetR::EGFP translational rate (/2 due to TetR dimers)
delta = 0.046;      // TetR::EGFP degradation + dilution
a = 3.7;            // rate of transcription machinery assembling
alpha = 0.28;       // rate of transcription machinery disassembling
rho = 20.8;         // rate of TetR::EGFP unbinding from the promoter
r = 0.2;            // rate of TetR::EGFP binding to the promoter
f = 0.9242;         // rate of inducer entering the cells [0..infinity]
c = 140;            // effective inducer concentration constant modifier
                    // (increase it for cells which are easily penetrated)
basal = 1;
M = 0;              // TetR::EGFP mRNA
D = 0;              // free TetR::EGFP (total TetR::EGFP will be D+H+B)
H = 0;              // 1 inducer-bound TetR::EGFP
B = 0;              // 2 inducer-bound TetR::EGFP

                    // Promoter states:
R01 = 0;            // one tetO2 site occupied (single repressed state)
R02 = 1;            // two tetO2 sites occupied (double repressed state)
A = 0;              // unrepressed state
Aup = 1;            // promoter state with transcription machinery present
Aup0 = 0;           // promoter state w/o transcription machinery (inactive)
Ii = 0;             // intracellular concentration of inducer
Ie = 0;             // extracellular concentration of inducer

Iinflux,      Ie->Ie+Ii,      f*c;
Ioutflux,     Ii->,           f;
r3,           Aup->Aup0,      alpha;
r4,           Aup0->Aup,      a;
r5,           Aup+A->Aup+A+M, m;
r5_1,         basal->basal+M, lk;
r6,           M->,            mu;
r7,           M->M+D,         p;
r8,           D->,            delta;
r9,           D+Ii->H,        2*b;
r10,          H->,            delta;
r11,          H+Ii->B,        b;
r12,          B->,            delta;
s01,          D+A->R01,       r;
s02,          R01->D+A,       rho;
s03,          D+R01->R02,     r;
s04,          R02->D+R01,     rho;
```

SUPPLEMENTARY NOTE 2

Synthetic CMV-derived promoters used in the study. Below are the detailed sequences of the original pCMV promoter[37,53,73], the pCMV-2xtetO promoter[29], the pCMV-dInr promoter, and all engineered TetR-repressible promoters that we constructed. The following parts are highlighted:

CGC..TTG	– Immediate early promoter enhancer
TATATAA	– TATA-box motif
TCAGATC	– Inr motif
TCCCTATCAGTGATAGAGA	– *tetO*2 site(s)
GAGCTC	– restriction site(s)
GTG..CTG	– scrambled DNA sequence used to move the Inr motif in the pCMV-dInr promoter to the same distance from the TATA-box as in the pCMV-2xtetO promoter, but without introducing *tetO*2 sites and keeping the region as far as possible from the Inr consensus sequence YYANWYY[66,67] (**Supplementary Fig. S3a**).

Original promoters:

WT pCMV (shown in full, the underlined region is modified in the engineered variants):

ACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCG
TTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTC
AATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTA
TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTT
TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG
TACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG
TCAATGGGAGTTTGTTTTGGAACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGC
CCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA
GTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGG
ACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-2xtetO:

CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTC
CCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACC
GTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
CAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-dInr:

CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGT
GTGCGTGTGTGGCTGTGGTCTTTCTGGGCGCGGTCGCTTCCTCCTGGAGCTCGTTTAGTGAACC
GTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
CAGCCTCCGGACTCTAGCGTTTAAACTTAAG

Engineered promoters:

pCMV-C3:

CGCCCCATTGACGCAAATGGGCGGTTCCCTATCAGTGATAGAGATCTATATAAGCAGAGCTCTC
CCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACC
GTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
CAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-C4:

CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTC
CCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGAGATCC
CTATCAGTGATAGAGAACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
CAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-D2i:

CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCATCCCTATC
AGTGATAGAGATCAGATCTCCCTATCAGTGATAGAGAGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-D2t:

CGCCCCATTGACGCAAATGGGCGGTAGTCCCTATCAGTGATAGAGATATATAAGCATCCCTATC
AGTGATAGAGATCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-D3:

CGCCCCATTGACGCAAATGGGCGGTAGTCCCTATCAGTGATAGAGATATATAAGCATCCCTATC
AGTGATAGAGATCAGATCTCCCTATCAGTGATAGAGAGCTGTTTTGACCTCCATAGAAGACACC
GGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAG

pCMV-D4:

CGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCATCCCTATC
AGTGATAGAGATCAGATCTCCCTATCAGTGATAGAGAGCTCCCTATCAGTGATAGAGAGATCCC
TATCAGTGATAGAGATCCGGACTCTAGCGTTTAAACTTAAG

pCMV-D5:

CGCCCCATTGACGCAAATGGGCGGTAGTCCCTATCAGTGATAGAGATATATAAGCATCCCTATC
AGTGATAGAGATCAGATCTCCCTATCAGTGATAGAGAGCTCCCTATCAGTGATAGAGAGATCCC
TATCAGTGATAGAGATCCGGACTCTAGCGTTTAAACTTAAG

SUPPLEMENTARY METHODS

Detailed description of plasmid construction in this study. The oligonucleotides used during the construction of the plasmids can be found in **Supplementary Table S2.** The pcDNA4/TO-EGFP plasmid was produced by amplifying the mammalian codon optimized variant of *eGFP* gene from the pIRES2-EGFP plasmid (Clontech, Mountain View, CA) using the hEGFP-BamHI-f and hEGFP-XhoI-r primers and inserting it between the *BamH*I and *Xho*I sites of the pcDNA4/TO plasmid (Invitrogen, Carlsbad, CA). The plasmid pDN-VTG4 containing the prototype TG1 was created by digesting the *tetR::eGFP* gene out from the plasmid pDN-G1TGt[35] using the *BamH*I and *Xho*I restriction enzymes and inserting it between the same sites into the pcDNA4/TO-EGFP plasmid. The bifunctional *tetR::eGFP* fusion gene was constructed from the *tetR* repressor[61] and *eGFP* fluorescent reporter[62] genes fused by a linker with the GDGAGLIN amino acid sequence[63]. The *Afl*II-*Xba*I fragment from the plasmid pcDNA6/TR (Invitrogen, Carlsbad, CA) containing the rabbit β-globin intron II sequence[47] and the *Xba*I-*Pac*I fragment from the plasmid pDN-VTG4 were inserted into the pDN-VTG4 between *Afl*II and *Pac*I sites, resulting in the plasmid pDN-VrTG4 bearing the prototype TG2. The mammalian codon optimized fusion *htetR::eGFP* gene was created as following. First, the fragment containing the rabbit β-globin intron II sequence was PCR amplified from the pDN-VrTG4 plasmid using the CMV-AflII-mamma-f and IVS-htetR-r primers and fragment containing *htetR* gene was PCR amplified from the pTet-OFF Advanced plasmid (Clontech, Mountain View, CA) using the htetR-IVS-f and htetR-PacI-linker-r primers. Next, these two fragments were fused together and amplified in the subsequent PCR using the primers CMV-AflII-mamma-f and htetR-PacI-linker-r producing the fragment *Afl*II-IVS-*htetR*-*Pac*I. In addition, the *Pac*I-heGFP-*Xho*I fragment was amplified from the pcDNA4/TO-EGFP using the heGFP-PacI-link-f and heGFP-XhoI-r primers. The *Afl*II-IVS-*htetR*-*Pac*I and *Pac*I-heGFP-*Xho*I fragments were digested with respective restriction enzymes and inserted into the pDN-VrTG4 plasmid between *Afl*II and *Xho*I sites, resulting in the plasmid pDN-VrTG4h bearing the prototype TG3 with mammalian codon optimized *htetR::eGFP* gene.

The first intermediate plasmid pcDNA6/TR-NF was created by PCR amplifying the *tetR* fragment from the pcDNA6/TR plasmid using the rtTA-BamHI2-f and htetR-XhoI-r primers; combining it with the *Spe*I-*BamH*I fragment from the pcDNA4/TO, digesting them with respective enzymes and inserting both between *Spe*I and *EcoR*I sites of the pcDNA6/TR plasmid. The second intermediate plasmid pDN-VrTS6h was created as following. First, the simian virus 40 (SV40) large-T-antigen nuclear localization sequence[50] was added to *htetR* gene by PCR amplifying its region from the pDN-VrTG4h using the CMV-AflII-mamma-f and htetR-NLS-NheI-r primers. Second, the KRAB-AB silencing domain of the Kid-1 protein[64] was PCR amplified from the pTet-tTS plasmid (Clontech, Mountain View, CA) using the Kid-NheI-f and Kid-EcoRI-r primers. Then both fragments were digested with respective restriction enzymes and inserted into the pcDNA6/TR-NF plasmid between *Afl*II and *EcoR*I sites. The third intermediate plasmid pDN-VrTN6h was created by PCR amplifying *Afl*II-IVS-htetR-NLS-*EcoR*I fragment from the pDN-VrTS6h plasmid using the primers CMV-AseI-f and htetR-NLS-EcoRI-r and inserting replacing with it into the pDN-VrTS6h plasmid fragment between the *Afl*II and *EcoR*I sites.

The lentivirus-based vector variants with for comparison of prototypes TG3 and TG4 were created as following. The entire prototype TG3 (promoter and *htetR::eGFP* gene) was PCR amplified from the pDN-VrTG4h plasmid using the primers CMV-AgeI-f and hEGFP-NheI-r.

The prototype TG4 was created by combining two fragments PCR amplified from the pDN-VrTN6h using the primers CMV-AgeI-f and htetR-NLS-PacI-link-r and from the pDN-VrTG4h using the primers hEGFP-PacI-link-f and hEGFP-NheI-r, respectively. The above fragments were then inserted into the pLVX-DD-tdTomato plasmid (Clontech, Mountain View, CA) between *Xma*I and *Xba*I sites resulting in plasmids pDN-VrTGhv and pDN-VrTNGhv (lentivirus-based prototypes TG3 and TG4, respectively). The plasmid-based version of prototype TG4 was created by inserting the *BamH*I-*Pac*I fragment from the pDN-VrTNGhv between the *BamH*I and *Pac*I sites of the pDN-VrTG4h plasmid.

Prototype TG5 was created as follows. First, the fragment containing the *heGFP* was PCR-amplified from the pDN-VrTG4h plasmid using the hEGFP-PacI-link-f and hEGFP-XhoI-r primers and fragment containing the woodchuck hepatitis virus post-transcriptional regulatory element (WPRE)[51,65] was PCR amplified from the pDN-VrTNGhv plasmid using the hEGFP-WPRE-f and WPRE-XhoI-r primers. Next, these two fragments were fused together and amplified subsequently by PCR using the primers hEGFP-PacI-link-f and WPRE-XhoI-r producing the fragment *Pac*I-*heGFP*-WPRE-*Xho*I. This fusion fragment, together with the *BamH*I-*Pac*I fragment from the pDN-VrTNGhv, was digested with respective restriction enzymes and inserted between *BamH*I and *Xho*I sites of the pDN-VrTG4h plasmid resulting in the plasmid pDN-VrTNG4wh bearing the prototype TG5. Subsequently, the three fragments were PCR-amplified from the pDN-VrTNG4wh plasmid, using the following three pairs of primers: CMV-AflII-mamma-f and skhtetR-r; sk-htetR-f and es-WPRE-r; es-WPRE-f and WPRE-XhoI-r. These three fragments were then fused together in a single PCR reaction with the CMV-AflII-mamma-f and WPRE-XhoI-r primers. The reaction product was digested with the respective restriction enzymes and inserted between *Afl*II and *Xho*I sites of the pDN-VrTNG4wh plasmid resulting in plasmid pDN-VrTNG4kwh bearing prototype TG6.

The seven parts of the modified CMV-based synthetic promoters (pCMV-C3, pCMV-C4, pCMV-D2i, pCMV-D2t, pCMV-D3, pCMV-D4, and pCMV-D5) limited by the 46 bp sequence upstream of the TATA-box and 116-170 bp downstream of it (including the *Afl*II site) with introduced changes were synthesized *de novo* (Bio Basic Inc., Markham, Ontario, Canada). The promoter structures are outlined on **Fig. 6a** of the main text and the exact sequences can be found in the **Supplementary Note 2**. The intermediate PCR fragment designated as UP* was PCR amplified from the pDN-VrTNG4kwh plasmid using the primers CMV-SpeI-f and CMV-Ench-TATA-r. Next, synthetic promoter fragments were PCR amplified from the plasmids provided by the DNA synthesis vendor using the CMV-Ench-TATA-f and CMV-AflII-r primers. Subsequently, each PCR product containing synthetic promoter fragment was fused to the UP* fragment and amplified in another PCR reaction using the CMV-SpeI-f and CMV-AflII-r primers; digested with respective restriction enzymes and inserted between the *Spe*I and *Afl*II sites in the pDN-VrTNG4kwh plasmid, resulted in series of plasmids: pDN-C3rTNG4kwh, pDN-C4rTNG4kwh, pDN-D2irTNG4kwh, pDN-D2itTNG4kwh, pDN-D3rTNG4kwh, pDN-D4rTNG4kwh, and pDN-D5rTNG4kwh, bearing the pCMV-C3, pCMV-C4, pCMV-D2i, pCMV-D2t, pCMV-D3, pCMV-D4, and pCMV-D5 promoters respectively.

The plasmids for two-gene mammalian linearizer were created as following. First, the fragment containing the *mCherry* gene was PCR amplified from the pTRE-Dual2 plasmid (Clontech, Mountain View, CA) using the IVS-hmCherry-f and hmCherry-AgeI-r primers and the second fragment was PCR amplified from the pDN-D2irTNG4kwh plasmid using the CMV-SpeI-f and hmCherry-IVS-r primers. Next, these two fragments were fused together and amplified in the subsequent PCR using the primers CMV-SpeI-f and hmCherry-AgeI-r

producing the fragment *Spe*I-IVS-mCherry-*Age*I. This fusion fragment was digested with respective restriction enzymes and inserted between *Spe*I and *Age*I sites of the pDN-D2irTNG4kwh plasmid resulting in the plasmid pDN-D2irC4kwh bearing the reporter part. Second, the fragment containing selection marker was PCR amplified from the pcDNA6/TR plasmid using the 2602-Add-XhoI-f and SV40-Seq-r primers and digested with the *Bpl*I and *Xho*I enzymes. Next, fragment containing the regulatory part was cut from the pDN-D2irTNG4kwh plasmid using the *Spe*I and *Xho*I enzymes. These two fragments were inserted between *Spe*I and *Bpl*I sites of the pcDNA6/TR plasmid resulting in the plasmid pDN-D2irTNG6kwh bearing the regulatory part.

The additional plasmids with the wild-type pCMV promoter and the pCMV-dInr promoter with displaced Inr motif were created as follows. First, we used sequential PCR and the primers CMV-Spe-f, CMVwt-TATA-r, CMVwt-TATA-f, and CMV-AflII-r to remove the *tetO*2 sites from the pCMV-2xtetO promoter, reverting its sequence back to the original pCMV wild-type promoter. Second, the wild-type pCMV promoter was inserted into the pcDNA4/TO-EGFP plasmid instead of pCMV-2xtetO between the *Spe*I and *Afl*II sites, resulting in the pDN-MG4 plasmid. Third, we used sequential PCR and primers CMV-Spe-f, CMV-dInr-TATA-r, CMV-dInr-r, CMV-dInr-f, CMV-dInr-Inr-f, and CMV-AflII-r to replace the *Sac*I-*Sac*I fragment of the pCMV-2xtetO promoter containing the two *tetO*2 sites with the scrambled DNA sequence of exactly the same length, displacing the Inr motif to the same distance from the TATA-box as in the pCMV-2xtetO promoter, keeping the region as far as possible from the Inr motif (YYANWYY)[66,67] and *tetO*2 sites' consensus sequence. The modified pCMV-dInr promoter was then inserted into the pcDNA4/TO-EGFP plasmid instead of pCMV-2xtetO, between the *Spe*I and *Afl*II restriction sites, resulting in the pDN-MdG4 plasmid.

SUPPLEMENTARY REFERENCES

61 Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K. & Meier, I. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. J Mol Biol 169, 707-721 (1983).

62 Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173, 33-38 (1996).

63 Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. Yeast 21, 661-670 (2004).

64 Freundlieb, S., Schirra-Muller, C. & Bujard, H. A tetracycline controlled activation/repression system with increased potential for gene transfer into mammalian cells. J Gene Med 1, 4-12 (1999).

65 Brun, S., Faucon-Biguet, N. & Mallet, J. Optimization of transgene expression at the posttranscriptional level in neural cells: implications for gene therapy. Mol Ther 7, 782-789 (2003).

66 Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. & Smale, S. T. DNA sequence requirements for transcriptional initiator activity in mammalian cells. Mol Cell Biol 14, 116-127 (1994).

67 Jin, V. X., Singer, G. A., Agosto-Perez, F. J., Liyanarachchi, S. & Davuluri, R. V. Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. BMC Bioinformatics 7, 114 (2006).

68 Degenkolb, J., Takahashi, M., Ellestad, G. A. & Hillen, W. Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. Antimicrob Agents Chemother 35, 1591-1595 (1991).

69 Sigler, A., Schubert, P., Hillen, W. & Niederweis, M. Permeation of tetracyclines through membranes of liposomes and Escherichia coli. Eur J Biochem 267, 527-534 (2000).

70 Corish, P. & Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng 12, 1035-1040 (1999).

71 Lammers, T. et al. Role of PP2Calpha in cell growth, in radio- and chemosensitivity, and in tumorigenicity. Mol Cancer 6, 65 (2007).

72 Fjelldal, R., Moe, B. T., Orbo, A. & Sager, G. MCF-7 cell apoptosis and cell cycle arrest: non-genomic effects of progesterone and mifepristone (RU-486). Anticancer Res 30, 4835-4840 (2010).

73 Davis, M. G. & Huang, E. S. Transfer and Expression of Plasmids Containing Human Cytomegalo-Virus Immediate Early Gene-1 Promoter Enhancer Sequences in Eukaryotic and Prokaryotic Cells. Biotechnol Appl Bioc 10, 6-12 (1988).