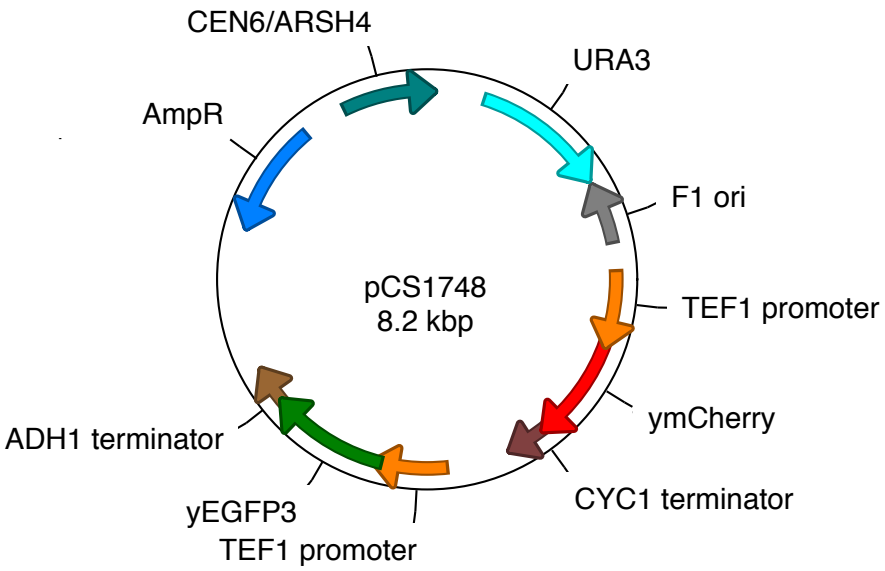
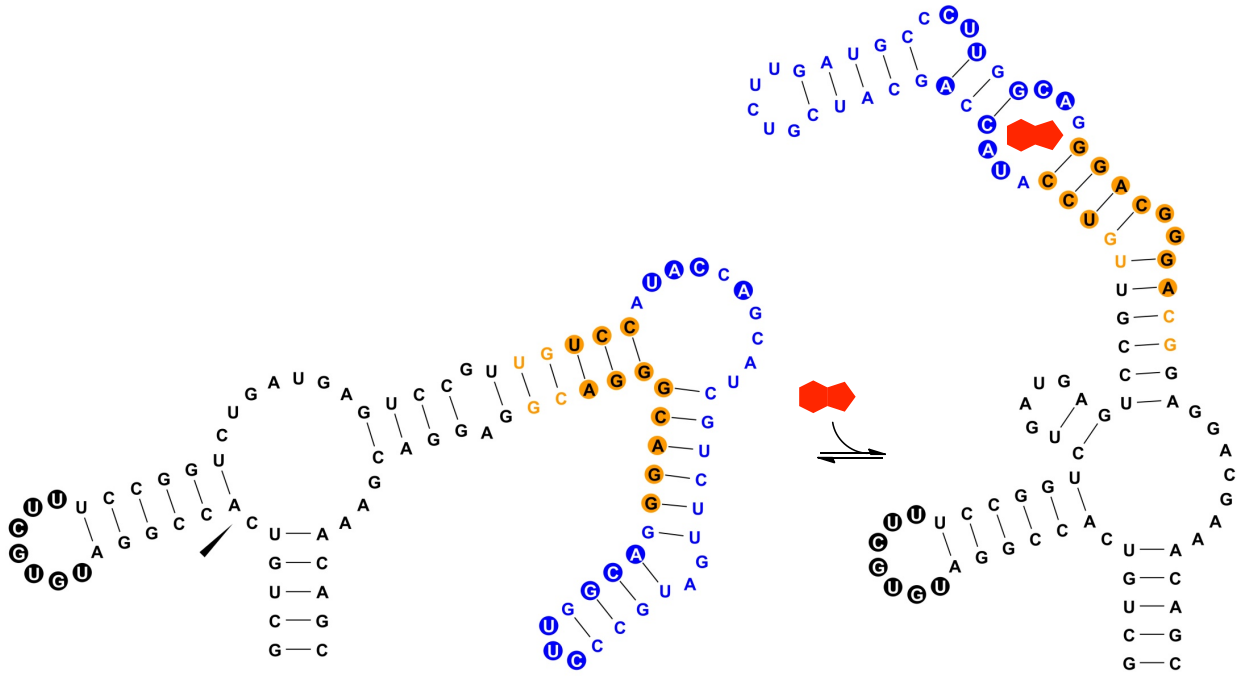
Liang J C, Chang A L, Kennedy A B, and Smolke C D

A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity

SUPPLEMENTARY DATA



Supplementary Figure S1. Plasmid map of pCS1748. RNA devices are inserted in the 3' UTR of yEGFP3.

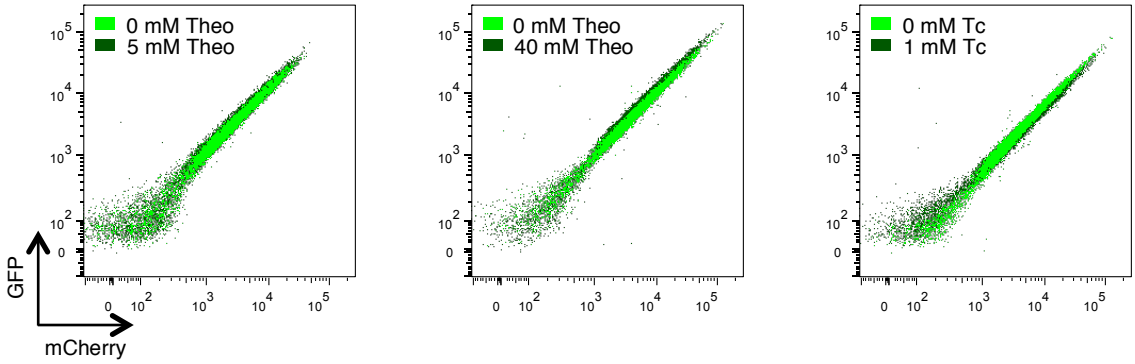


Supplementary Figure S2. Library design for sN10 sensor, aN7 actuator, and tN11 and a1-tN11 transmitter libraries. Predicted secondary structures of the active ribozyme conformation with unbound theophylline aptamer (left) and the inactive ribozyme conformation with bound theophylline aptamer (right) are shown. The sequence for the L2b8 parent device is shown. Nucleotides that comprise the sensor, actuator, and transmitter components are colored blue, black, and yellow, respectively. Theophylline is represented by the red polygon. Filled circles within each component indicate nucleotides randomized to generate respective device libraries. Secondary structures were predicted by RNAstructure folding software (43) and rendered using VARNA software (44).

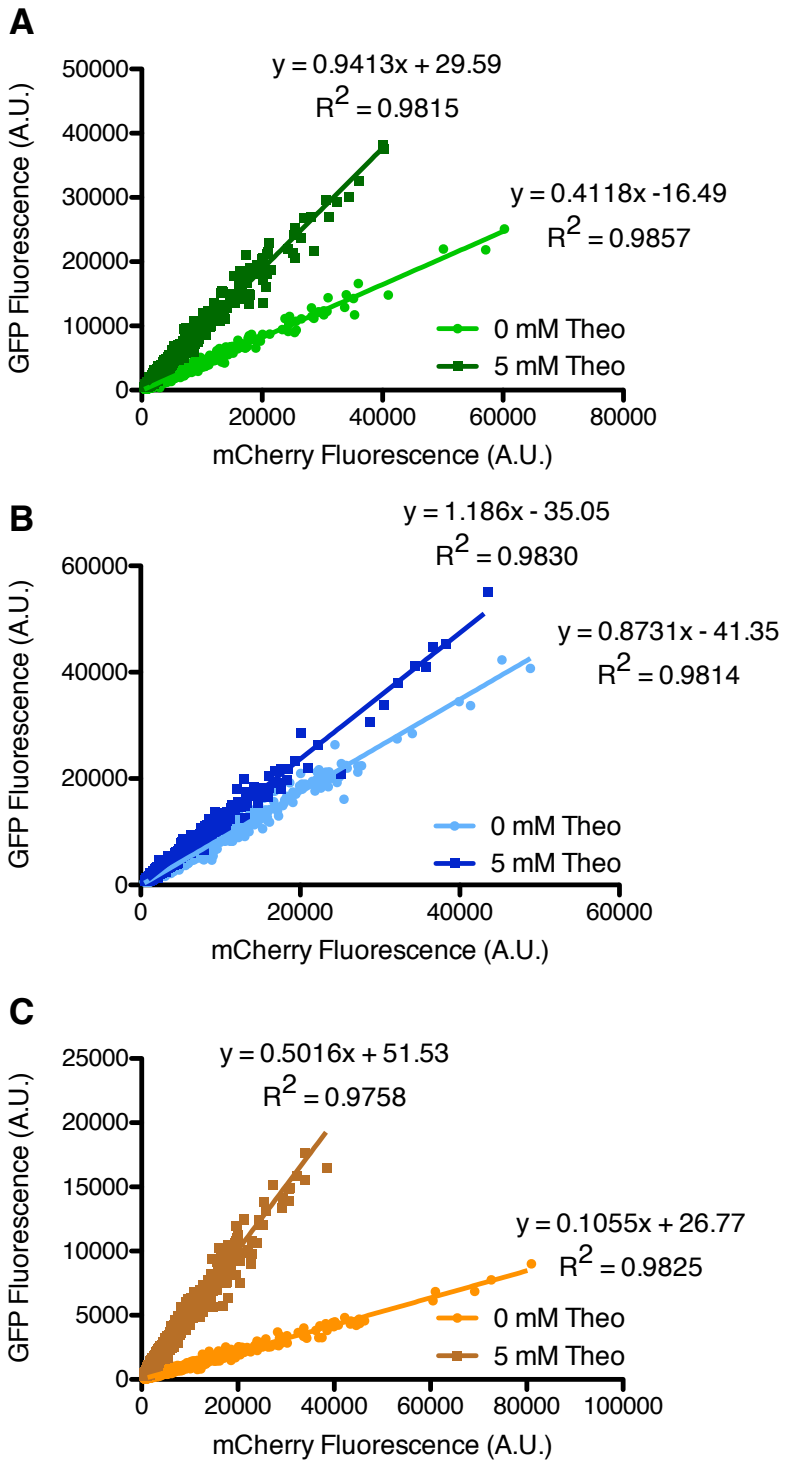
(**A**)

	GFP (A.U.)			mCherry (A.U.)			Viability (%)		
	0mM Theo	**5mM Theo**	**Ratio**	**0mM Theo**	**5mM Theo**	**Ratio**	**0mM Theo**	**5mM Theo**	**Ratio**
Replicate 1	3969.31	4532.12	1.14	2097.92	2339.26	1.12	98.24	97.16	0.99
Replicate 2	6282.93	6868.85	1.09	2699.57	2589.77	0.96	97.51	96.60	0.99
Replicate 3	2960.27	3129.06	1.06	2505.69	2496.95	1.00	96.47	96.73	1.00
Mean ± S.D.			1.10 ± 0.04			1.02 ± 0.08	97.41 ± 0.89	96.83 ± 0.29	0.99 ± 0.01
	0mM Theo	**40mM Theo**	**Ratio**	**0mM Theo**	**40mM Theo**	**Ratio**	**0mM Theo**	**40mM Theo**	**Ratio**
Replicate 1	4687.84	6380.01	1.36	5653.65	5504.26	0.97	97.63	96.62	0.99
Replicate 2	4777.71	6249.37	1.31	5137.39	5502.14	1.07	98.39	96.59	0.98
Replicate 3	4878.88	5977.69	1.23	4722.92	4921.71	1.04	97.99	94.83	0.97
Mean ± S.D.			1.30 ± 0.07			1.03 ± 0.05	98.00 ± 0.38	96.01 ± 1.02	0.98 ± 0.01
	0mM Tc	**1mM Tc**	**Ratio**	**0mM Tc**	**1mM Tc**	**Ratio**	**0mM Tc**	**1mM Tc**	**Ratio**
Replicate 1	2585.21	2382.28	0.92	2791.13	2278.18	0.82	98.05	98.72	1.01
Replicate 2	2310.21	2036.61	0.88	2896.15	2628.35	0.91	97.73	98.74	1.01
Replicate 3	2326.70	2203.89	0.95	3694.26	3691.82	1.00	97.75	98.22	1.00
Mean ± S.D.			0.92 ± 0.03			0.91 ± 0.09	97.84 ± 0.18	98.56 ± 0.29	1.01 ± 0.00

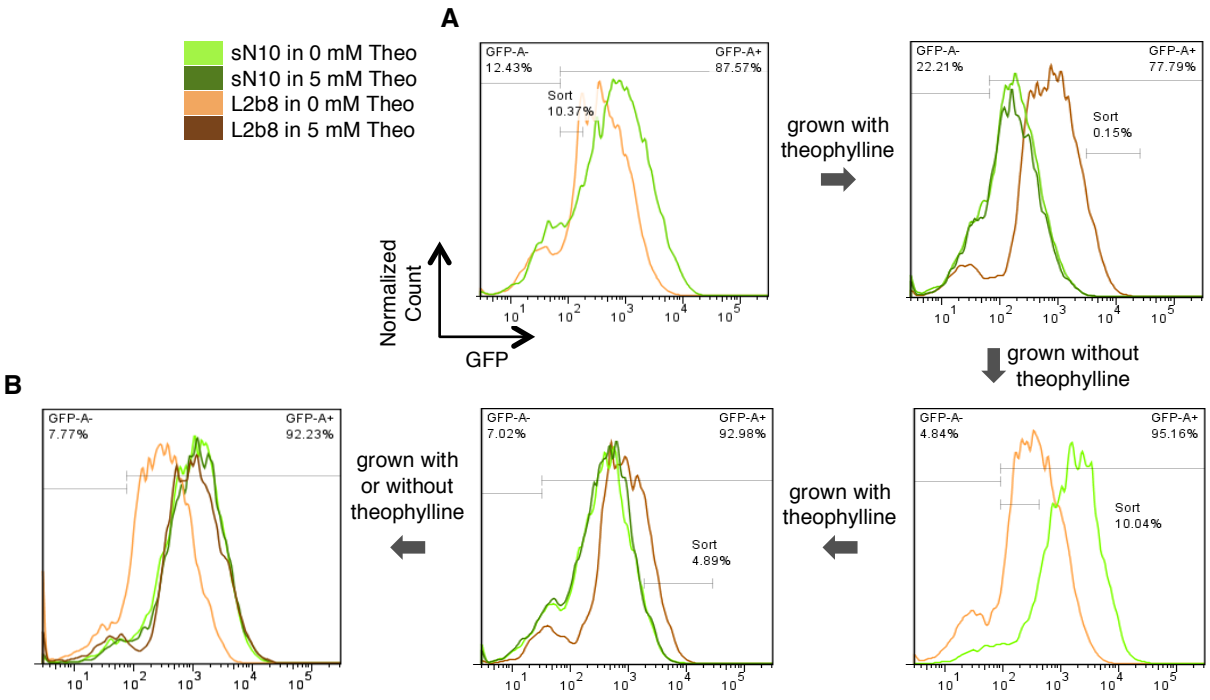
(**B**)



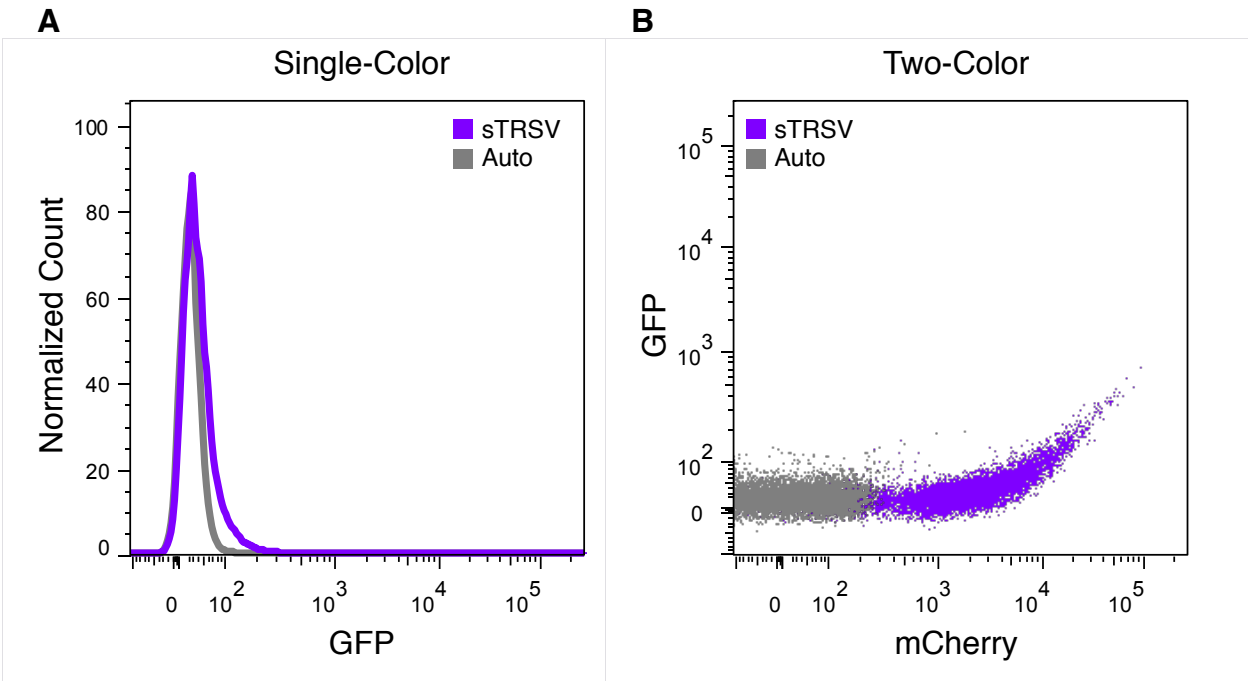
Supplementary Figure S3. Quantification of nonspecific ligand effect on fluorescence intensity and cell viability. (**A**) Unnormalized data for GFP fluorescence, mCherry fluorescence, and cell viability for the positive control construct (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) in all ligand conditions tested. Fluorescence values are reported as the geometric mean of the GFP or mCherry fluorescence of sTRSV Contl. Cell viability is reported as the percentage of cells included in the DAPI-(-) gate. Mean and standard deviation of three independent experiments are indicated. The ratio of fluorescence levels or cell viability percentages in the presence and absence of theophylline (theo) or tetracycline (tc) is indicated for each ligand condition. Variability in absolute GFP or mCherry signal among replicates for an individual ligand condition is due to differences in instrument calibration between experiments. Values for each replicate pair in the absence and presence of ligand were obtained in the same experiment with identical instrument calibration. (**B**) Scatter plots for sTRSV Contl under each ligand condition from a single representative experiment.



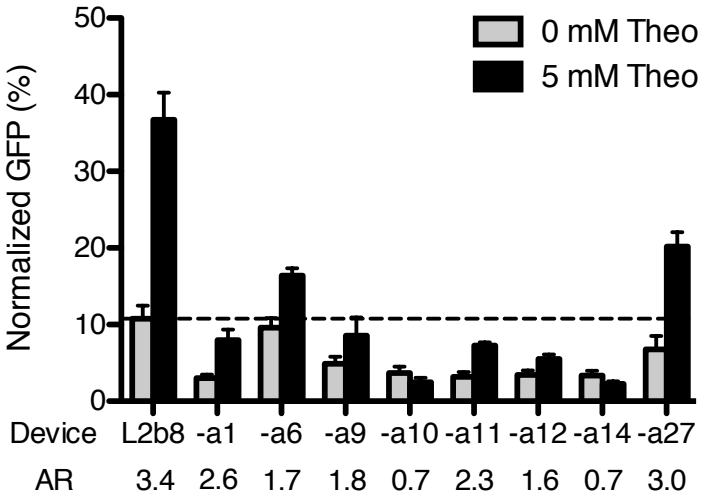
Supplementary Figure S4. Linear correlation between mCherry and GFP fluorescence for L2b1 (**A**), L2b5 (**B**), and L2b8 (**C**) devices within the two-color construct. For each device, 2,000 cells are plotted from a single representative experiment at 0 and 5 mM theophylline (theo). Linear regression and R^2 values are reported and support a distinct slope for different device gene-regulatory activities.



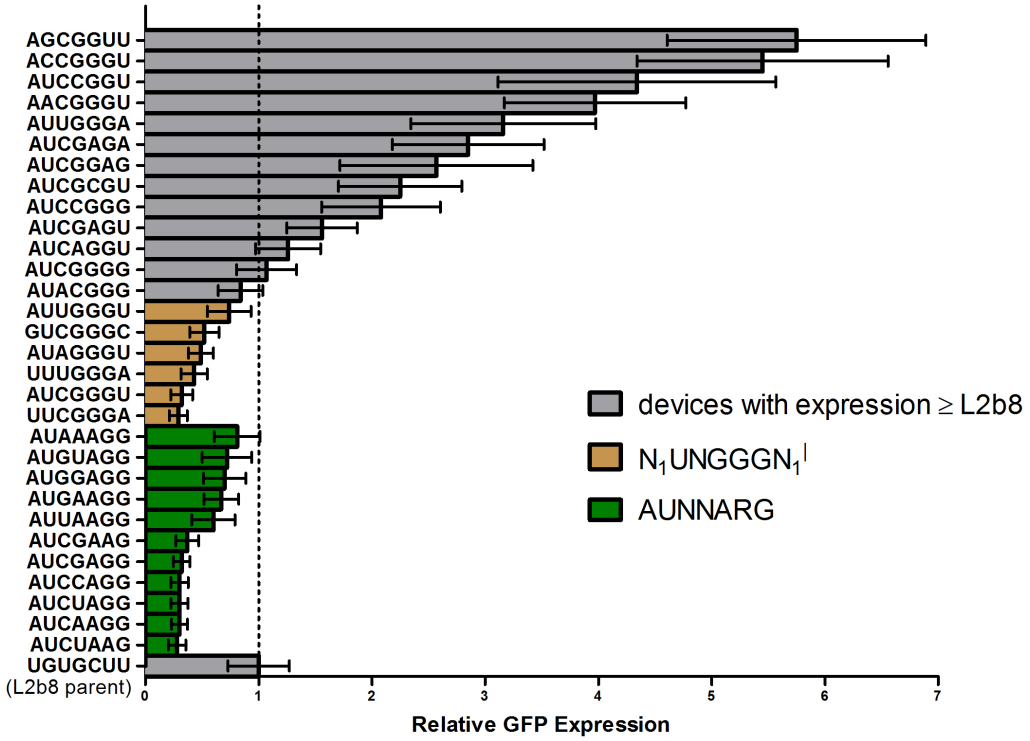
Supplementary Figure S5. Screening of the sN10 sensor library based on a single-color output shows no enrichment of theophylline-responsive population after two sorting rounds. (**A**) The sN10 sensor library as described in Figure 3 is subjected to two sorting rounds based on the GFP output only. Each round consists of one negative sort in the absence of theophylline, followed by one positive sort in the presence of theophylline. For the negative sorts, the sorting gates are set by collecting the bottom 10% of the cell population after gating out the GFP-(-) population set by the negative-control construct, pCS4. For the positive sorts, the sorting gates are determined by overlapping the two GFP histograms of the L2b8 population in the absence and presence of theophylline. The gate was roughly set to enclose the non-overlaid part of the L2b8 population in the presence of theophylline. Percentage of cells collected in the sorting gate is indicated on each plot. At least 10 times the estimated library size was screened in each sort to ensure that the full library diversity was covered. (**B**) The enriched sensor library exhibits no population shift in response to theophylline, in contrast to the same library enriched from the two-color screen. This result suggests that the enrichment efficiency of the single-color screen is poor and additional sorting rounds would be necessary to recover the functional sequence.



Supplementary Figure S6. The two-color sorting strategy supports separation of autofluorescent and low expression cell populations. sTRSV (the active ribozyme control) exhibits ~1.2% normalized GFP fluorescence, relative to a noncleaving ribozyme positive control (sTRSV Contl) with a scrambled core. (**A**) A cell population harboring the sTRSV construct exhibits low fluorescence levels that overlap significantly with a cell population harboring no fluorescence reporter gene in the single-color histogram. (**B**) The sTRSV and autofluorescent populations separate based on mCherry expression in the two-color plot, allowing cells exhibiting low GFP expression due to stringent device gene-regulatory activity to be effectively enriched through two-color sorting.



Supplementary Figure S7. An actuator sort preserves switching activity in 6 out of 8 characterized devices. Basal activity of the parent device, L2b8, is indicated by a dashed line. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR) is determined as the ratio of gene expression levels in the presence and absence of theophylline (theo).

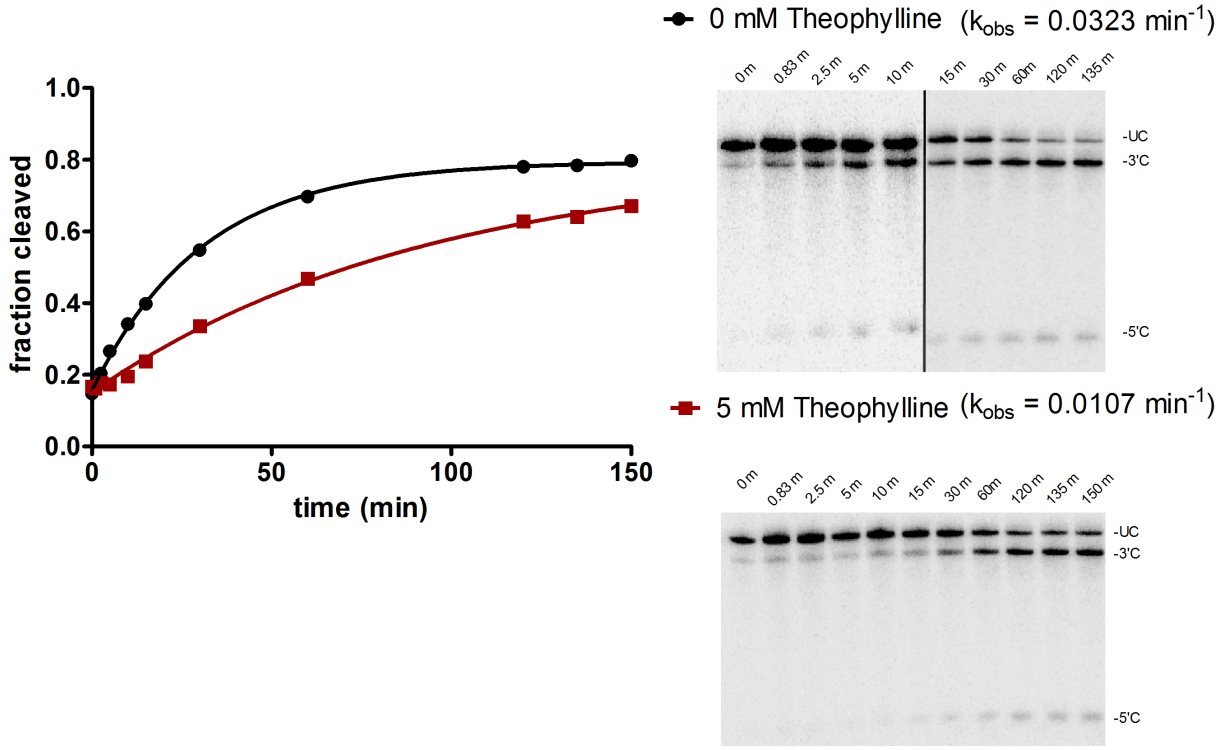


Supplementary Figure S8. Point mutation analyses of variant loop I sequences identify consensus sequences of ribozyme variants supporting improved gene-regulatory activities. The impact of point mutations to the recovered loop I sequences on basal activity was measured by flow cytometry. Relative expression levels are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of the L2b8 parent device. Reported values are the mean and standard deviation of at least three independent experiments. Indicated consensus sequences support decreased basal activities relative to the L2b8 parent as discussed in the main text (AUNNARG and $N_1UNGGGN_1^{|}$).

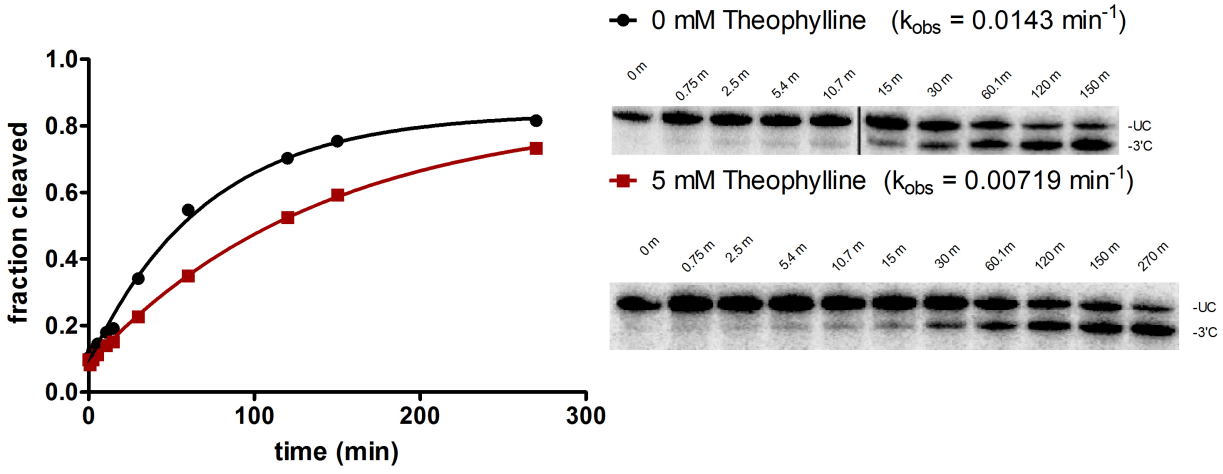
Two putative loop I consensus sequences were supported from the sequences recovered from the actuator library screen: a heptaloop (AUYNRRG) and a triloop ($NUYGGGN^{|}$). Analysis of the point mutants made to interrogate nucleotide base constraints at each position of the heptaloop sequence found that any base (N) is tolerable at the third and fourth positions, but specific identities at the third position result in improved basal activity (C > A/U > G). The fifth and sixth positions accept both purine bases (A, G) and retain activity similar to that of the L2b8 parent; however, an adenine base (A) in the fifth position results in improved basal activity. Point mutants verified the requirement of a guanine base (G) in the seventh position.

For the triloop sequence, point mutants indicate that any combination of bases in the first and seventh positions resulting in canonical Watson-Crick base pairing (AU, GC; order independent) results in improved basal activity. The bases in the second and sixth positions must result in a Watson-Crick GU wobble pair, where this requirement is sensitive to identity and order (i.e., standard Watson-Crick pairing abolishes improved activity). The point mutants indicate that the third position is able to accept any nucleotide base (N), whereas a guanine base (G) is required at the fourth and fifth positions to retain improved activity relative to the L2b8 parent device.

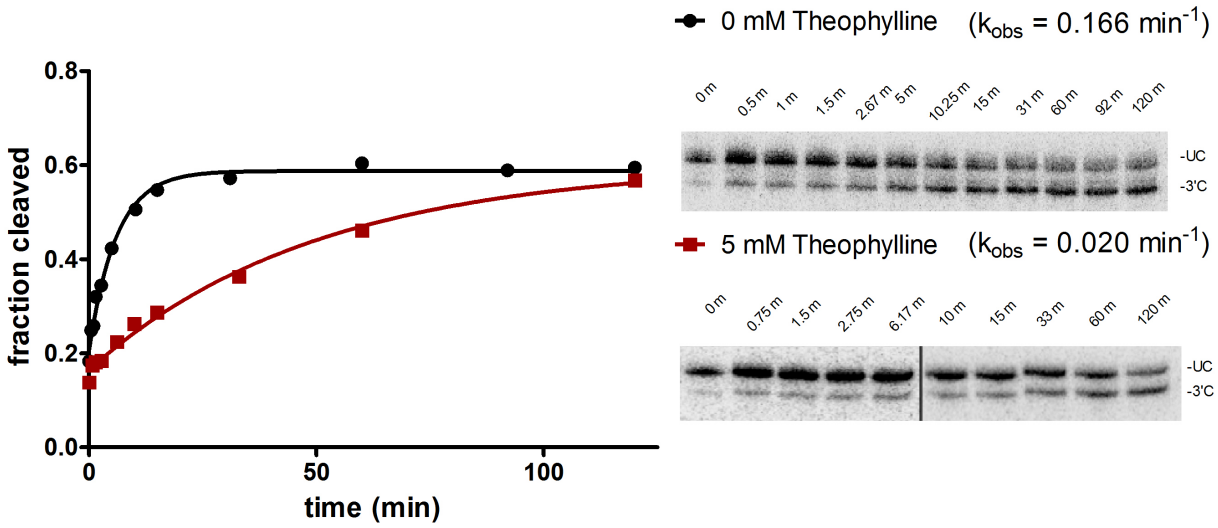
(**A**) L2b1:


0 mM Theophylline (k_{obs} = 0.0323 min^{-1})
fraction cleaved
time (min)
0 m
0.83 m
2.5 m
5 m
10 m
15 m
30 m
60m
120 m
135 m
-UC
-3'C
-5'C
5 mM Theophylline (k_{obs} = 0.0107 min^{-1})
150 m

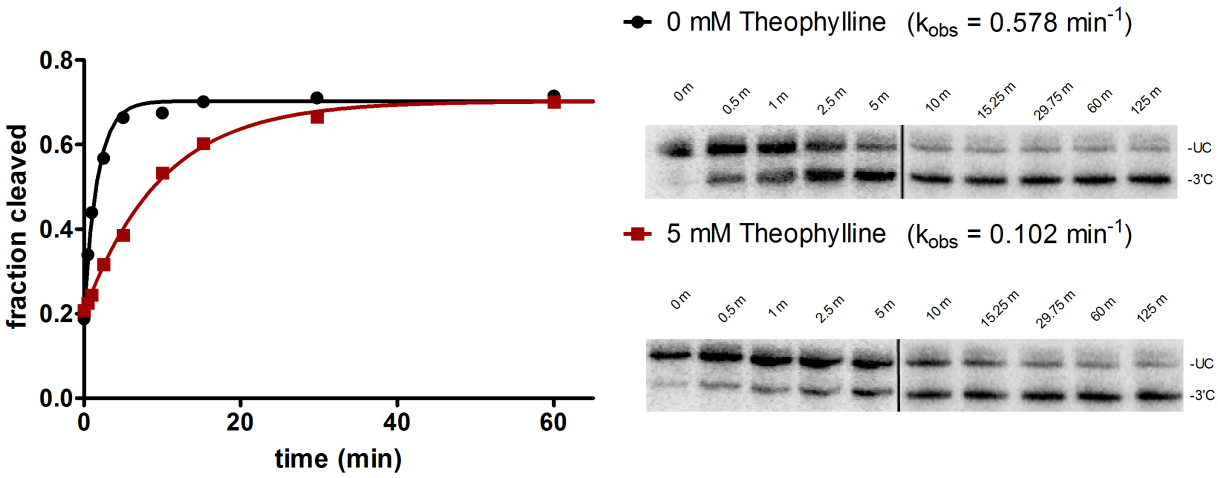
(**B**) L2b5:


0 mM Theophylline (k_{obs} = 0.0143 min^{-1})
fraction cleaved
time (min)
0 m
0.75 m
2.5 m
5.4 m
10.7 m
15 m
30 m
60.1m
120 m
150 m
-UC
-3'C
5 mM Theophylline (k_{obs} = 0.00719 min^{-1})
270 m

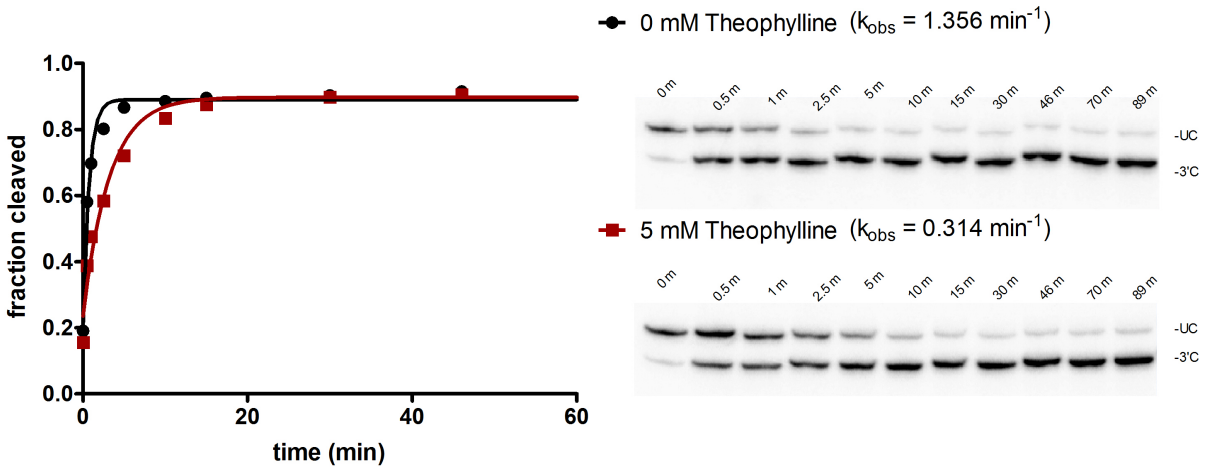
(**C**) L2b8:



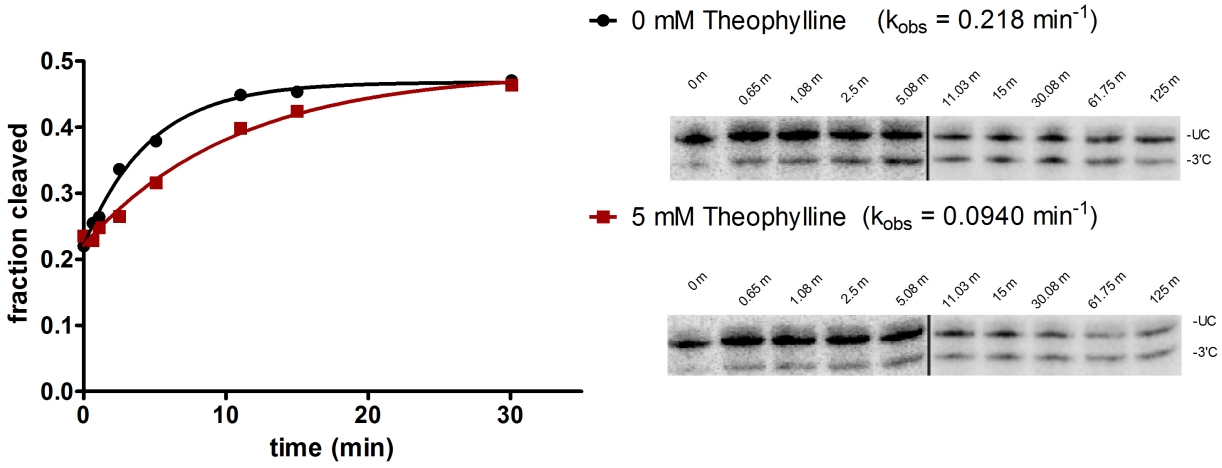
(**D**) L2b8-a1:



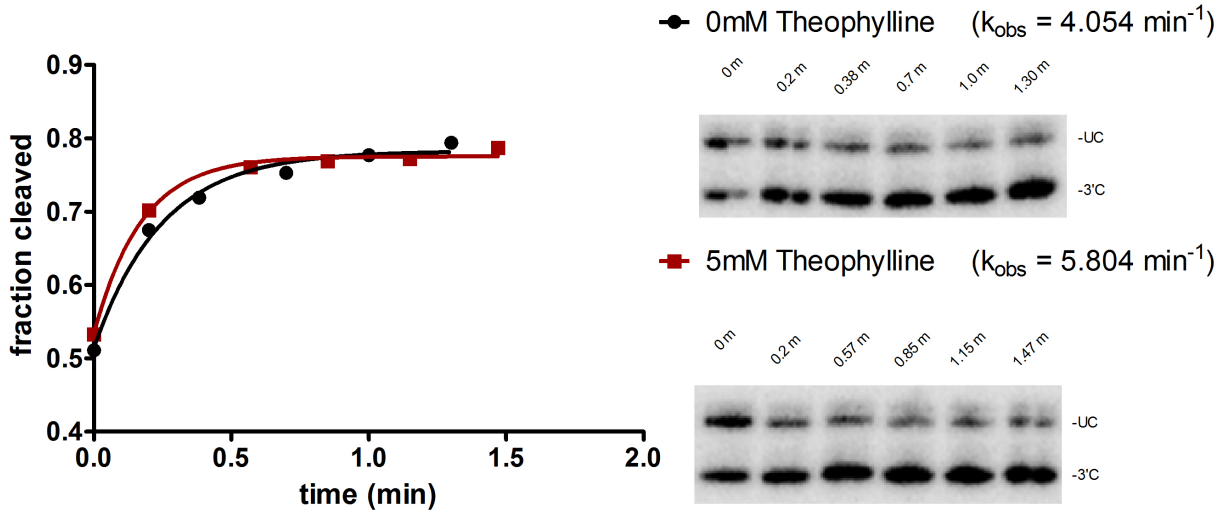
(**E**) L2b8-a14:



(**F**) L2b8-a1-t41:



(**G**) sTRSV ribozyme:



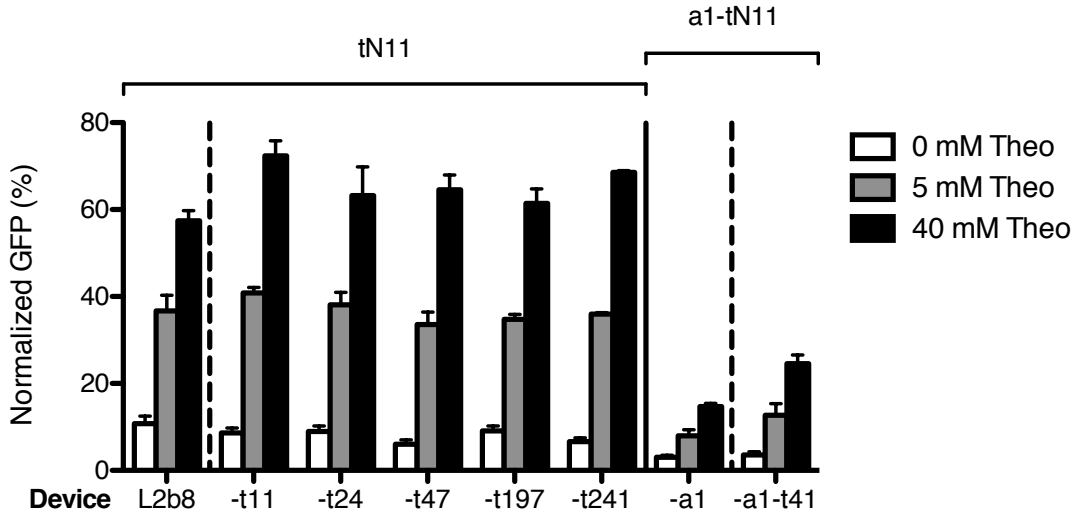
(**H**)

RNA Device	Cleavage Rate (k, min^{-1})	
	0 mM theophylline	**5 mM theophylline**
L2b1	0.033 ± 0.008	0.011 ± 0.003
L2b5	0.013 ± 0.002	0.007 ± 0.001
L2b8	0.14 ± 0.02	0.025 ± 0.009
L2b8-a1	0.70 ± 0.3	0.10 ± 0.03
L2b8-a14	1.4 ± 0.1	0.33 ± 0.01
L2b8-a1-t41	0.16 ± 0.05	0.096 ± 0.02
sTRSV ribozyme	4.3 ± 0.8	4.3 ± 1.4

Supplementary Figure S9. Representative cleavage assays for measuring cleavage rate constants (k) for ribozyme-based devices and controls. A representative assay is shown for each device in the absence and presence (0 and 5 mM, respectively) of theophylline: L2b1 (**A**), L2b5 (**B**), L2b8 (**C**), L2b8-a1 (**D**), L2b8-a14 (**E**), L2b8-a1-t41 (**F**), and sTRSV ribozyme (**G**). Bands for both the cleaved products (5'C, 3'C) and the full-length uncleaved substrate (UC) are shown for L2b1. For subsequent devices, the shorter 5'C product is omitted from the inset image for clarity. Methods used to prepare full-length, uncleaved RNA transcripts and conditions of the cleavage assays are detailed in the Materials and Methods section of the main text. Briefly, RNA was heated to 95°C for 5 min and cooled to 37°C in a secondary structure refolding buffer (100 mM NaCl, 50 mM Tris-HCl (pH 7.5)). A zero time point aliquot was removed prior to initiating the reaction with addition of $MgCl_2$ to a final concentration of 500 μM. Reactions were quenched at the indicated time points. At least seven time points were taken in each cleavage assay to capture the cleavage dynamics of RNA devices exhibiting different cleavage kinetics. A black vertical bar was used to denote time points run on different gels. Phosphorimaging analysis of relative levels of the UC, 5'C, and 3'C bands was used to determine the fraction cleaved at each time point (F_t). The fraction cleaved at the beginning (F_0) and end of reaction (F_∞) varied between assays, but all assays were well-fit to the single exponential equation ($R^2 > 0.95$):

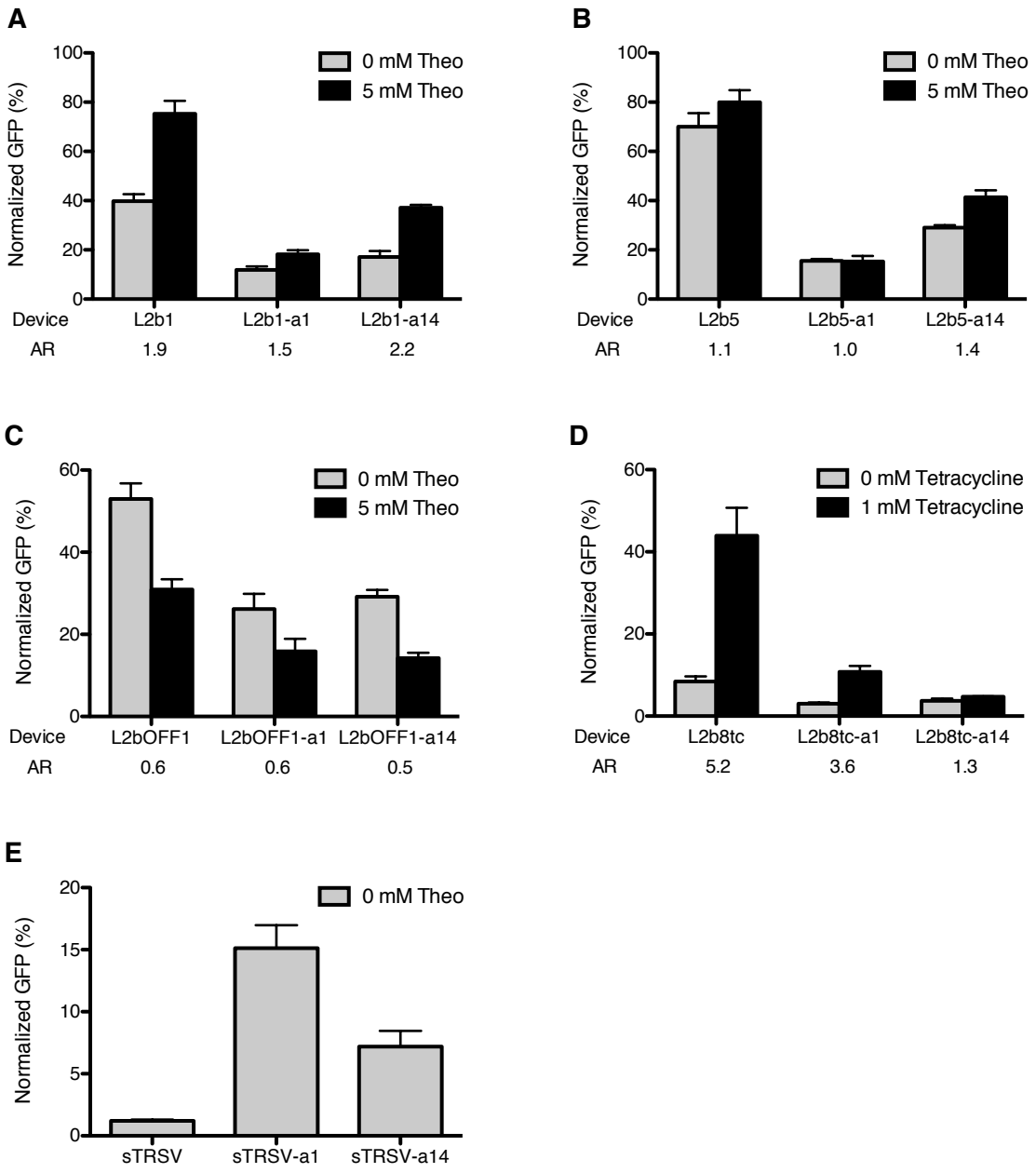
$$F_t = F_0 + (F_\infty - F_0) \times (1 - e^{-kt})$$

The black and red fit lines represent assays performed at 0 and 5 mM theophylline, respectively. The cleavage rate constant value (k) was determined for each assay. The reported k for each device and theophylline assay condition is the mean and standard deviation of at least three independent experiments (**H**).



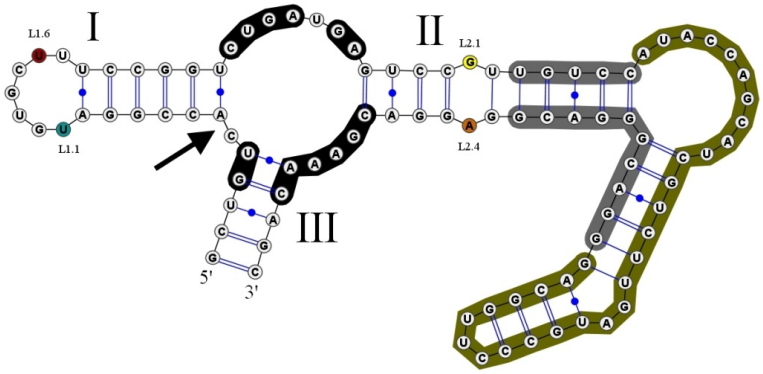
Device	L2b8	-t11	-t24	-t47	-t197	-t241	-a1	-a1-t41
AR at 5 mM Theo	3.4	4.7	4.2	5.5	3.8	5.4	2.6	3.6
AR at 40 mM Theo	5.3	8.4	7.0	10.7	6.8	10.4	4.8	6.9

Supplementary Figure S10. Increased input ligand concentration results in higher device ON states. Increasing ligand concentration to 40 mM theophylline (theo) increases device activation ratio (AR) compared to 5 mM theophylline. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR) is determined as the ratio of gene expression levels in the presence and absence of theophylline (theo).

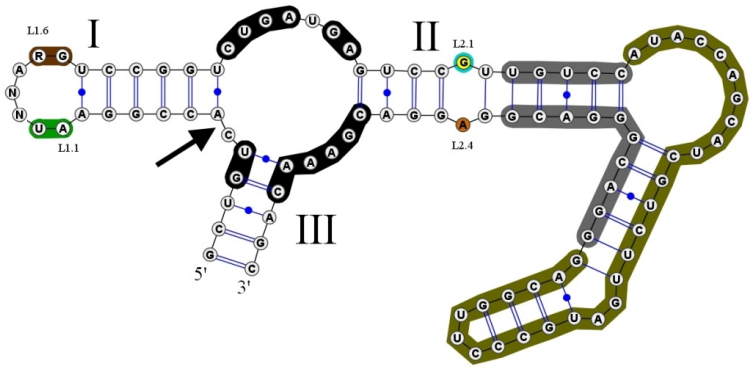


Supplementary Figure S11. Gene-regulatory activities of the optimized actuator components are maintained in the context of different sensor and transmitter components. Replacement of the loop I sequence with the a1 or a14 sequence decreases basal activity in all devices tested, including theophylline (theo)-responsive ON (**A**, **B**) and OFF (**C**) switches and a tetracycline-responsive ON switch (**D**). Switching activity is maintained in most devices. Notably, replacement of the loop I sequence in the sTRSV ribozyme (**E**) significantly increases basal activity. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR) is determined as the ratio of gene expression levels in the presence and absence of theophylline (theo) (**A**, **B**, **C**) or of tetracycline (**D**).

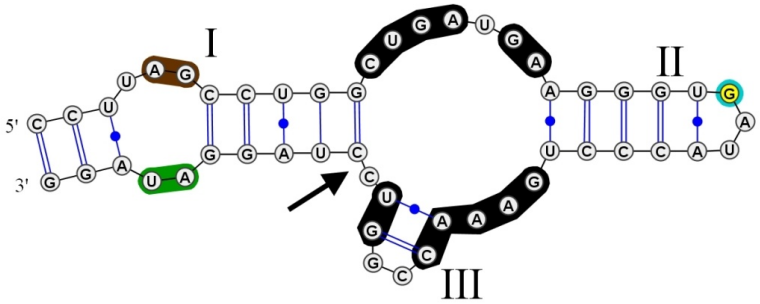
(**A**) L2b8 parent (Loop I: UGUGCUU)



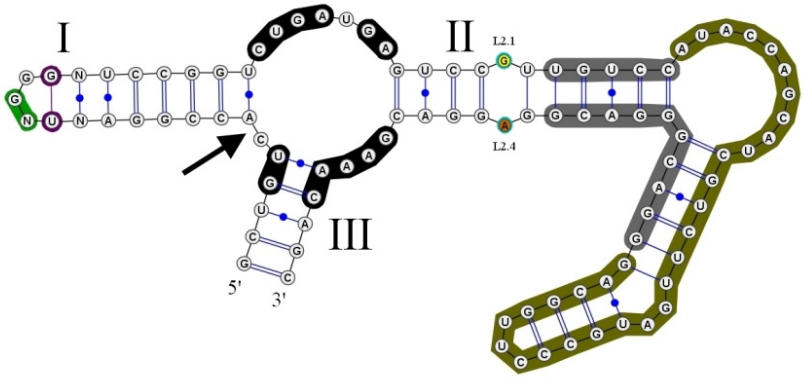
(**B**) aN7 heptaloop consensus (AUNNARG)



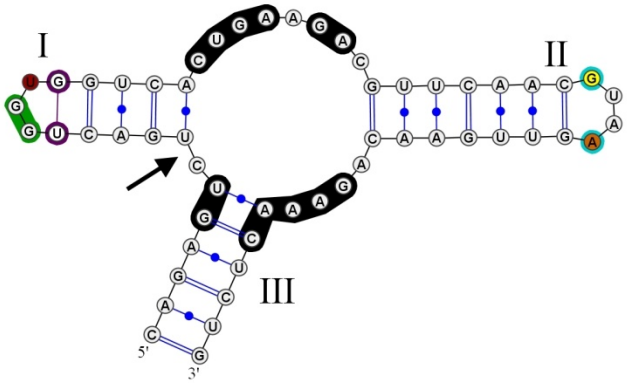
(**C**) Newt hammerhead ribozyme



(**D**) aN7 triloop consensus ($N_1UN_2GGGN_1$)



(**E**) CarSVRNA(-) hammerhead ribozyme



Supplementary Figure S12. Secondary structure analysis indicates similarities between engineered RNA devices and natural hammerhead ribozymes (HHRzs). The HHRz invariant nucleotide bases are boxed in black, and the cleavage site is denoted by an arrow in all structures. In RNA device structures (**A**, **B**, **D**), nucleotide bases of the transmitter and aptamer domains are boxed in grey and olive green respectively. Using the common numbering system for HHRz with apical loops (45), nucleotide bases highlighted in cyan (L1.1 U) and red (L1.6 U) are those conforming to the common HHRz loop I motif, UN_mYN, where m is typically 3 or 4 bases. Common loop II motif nucleotide bases, (RN_nA), are highlighted in yellow (L2.1 G) and orange (L2.1 A), where n is usually 2 or 4 bases (36). R and Y denote purine and pyrimidine bases respectively. (**A**) The HHRz core of the L2b8 parent device forms a predicted GU base pair at L2.2-L2.3 due to transmitter-sensor integration into loop II when analyzed by RNAstructure folding software (43). HHRz loop sequence similarities exist between a device containing the variant heptaloop consensus sequence selected in this study (**B**) and the newt HHRz (**C**). In loop I, both HHRz structures have AU in the first two positions 5' from the catalytic core (boxed in green) and RG in last two positions (boxed in brown). In loop II, both structures have a G nucleotide not constrained in a Watson-Crick base pair (circled in cyan). Loop sequence similarities also exist between a device containing the variant triloop consensus sequence (**D**) and the CarSVRNA(-) HHRz (**E**). Both loop I sequences in these HHRzs contain a triloop with a GU base pair closing the loop (circled in purple) and accommodate GG in the first two positions 5' from the catalytic core (boxed in green). In loop II, both HHRz structures have L2.1 G and L2.4 A nucleotides not constrained in Watson-Crick base pairs, available for loop I-loop II interaction, circled in cyan (36). Secondary structures rendered using VARNA software (44).

Supplementary Table S1. Summary of plasmids constructed.

Plasmid	Color Controls
pCS1748	two-color screening plasmid: TEF1p-GFP-ADH1t and TEF1p-mCherry-CYC1t
pCS1585	single-color plasmid: TEF1p-GFP-ADH1t
pCS1749	single-color plasmid: TEF1p-mCherry-CYC1t
	Ribozyme-based Devices and Controls
pCS1750	pCS1748+sTRSV
pCS1751	pCS1748+sTRSV Contl
pCS1752	pCS1748+L2b1
pCS1753	pCS1748+L2b8
pCS2260	pCS1748+L2b5
pCS2261	pCS1748+L2b8-a1
pCS2262	pCS1748+L2b8-a6
pCS2263	pCS1748+L2b8-a9
pCS2264	pCS1748+L2b8-a10
pCS2265	pCS1748+L2b8-a11
pCS2266	pCS1748+L2b8-a12
pCS2267	pCS1748+L2b8-a14
pCS2268	pCS1748+L2b8-a27
pCS2269	pCS1748+L2b8-t11
pCS2270	pCS1748+L2b8-t24
pCS2271	pCS1748+L2b8-t47
pCS2272	pCS1748+L2b8-t197
pCS2273	pCS1748+L2b8-t241
pCS2274	pCS1748+L2b8-a1-t41
pCS2275	pCS1748+L2b8-a1-t55
pCS2276	pCS1748+L2b8-a1-t64
pCS2277	pCS1748+L2b1-a1
pCS2278	pCS1748+L2b1-a14
pCS2279	pCS1748+L2b5-a1
pCS2280	pCS1748+L2b5-a14
pCS2281	pCS1748+L2bOFF1
pCS2282	pCS1748+L2bOFF1-a1
pCS2283	pCS1748+L2bOFF1-a14
pCS2284	pCS1748+L2b8tc
pCS2285	pCS1748+L2b8tc-a1
pCS2286	pCS1748+L2b8tc-a14
pCS2287	pCS1748+sTRSV-a1

pCS2288	pCS1748+sTRSV-a14
pCS2289	pCS1748+L2b8-a1-t11
pCS2290	pCS1748+L2b8-a1-t47
pCS2291	pCS1748+L2b8-t41