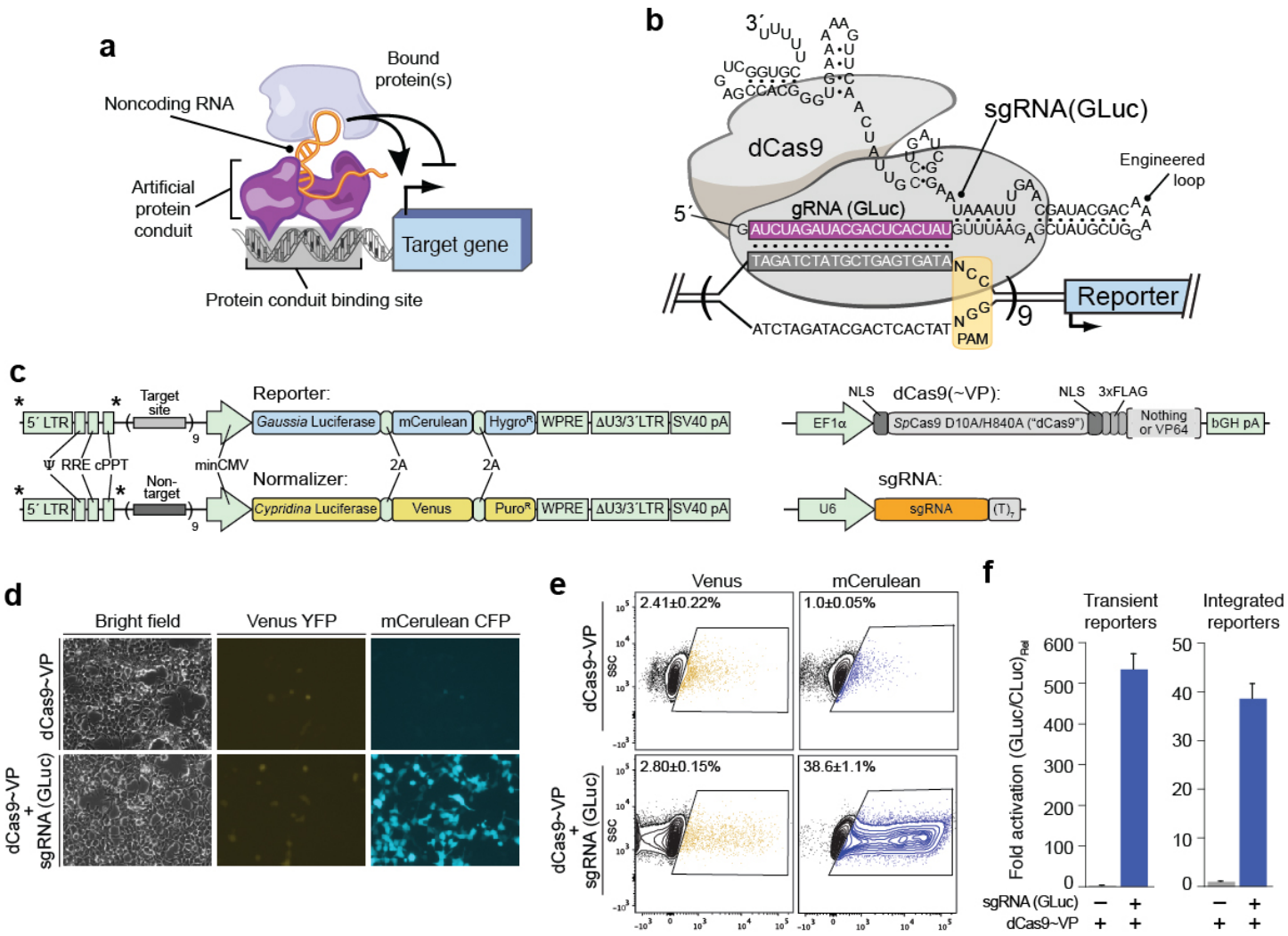
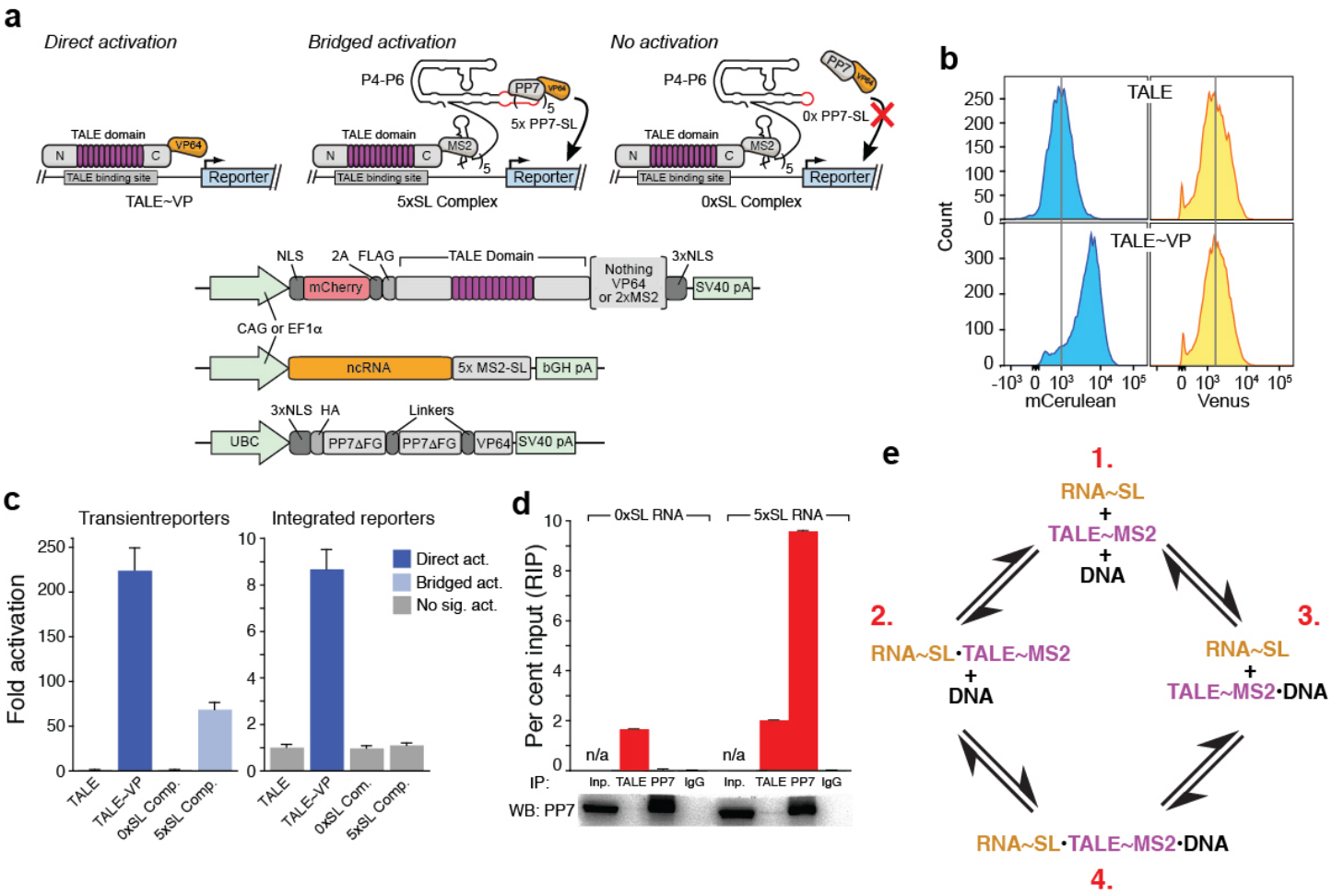


Supplementary Figure 1

Project Overview, design and controls.

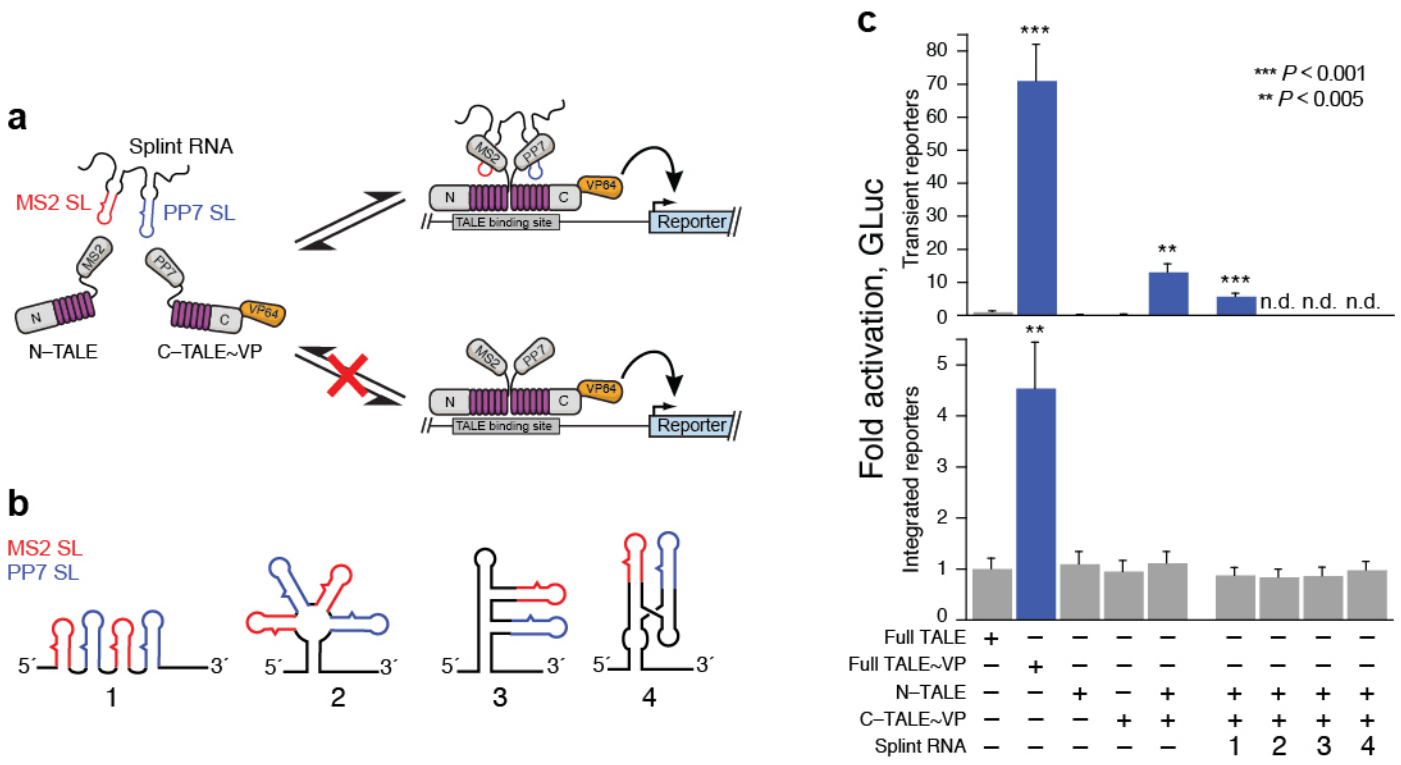
(**a**) Schematic of a general, targeted ncRNA localization system: a ncRNA cargo (and potential associated proteins) is ectopically targeted to a DNA locus via a programmable protein conduit. (**b**) Overview DNA targeting by *S. pyogenes* dCas9, directed by a minimal sgRNA that targets the GLuc reporter[18]. (**c**) Expression constructs in the dual reporter transcription activation system. Target and non-target sites in the Reporter and Normalizer are absent from the human genome (**Supplementary Table 1**) and are multimerized to exploit the additivity of artificial transcription activators[28,29]. minCMV: a 60-nt minimal cytomegalovirus promoter; 2A: a 2A 'self-cleaving' peptide. Lentiviral variants are shown; in transient reporter assays, regions bracketed by asterisks are absent. CRISPR construct design is modeled after reference 18. (**d**–**f**) Reporter system visualization of locus-specific RNA-guided transcription activation in HEK293FT cells. (**d**) Live-cell fluorescence microscopy using transient reporters. All images are 10× magnification (**e**) FACS analysis of the same experiment. Cells were gated with an mCherry cotransfection control (*data not shown*). The percentages of Venus and mCerulean positive cells (mean, ± standard deviation, n = 3) are quoted. (**f**) Luciferase assays. Values are means ± standard deviation, n = 3.



Supplementary Figure 2

A proposed ncRNA ectopic localization system based on TALE two-hybrids.

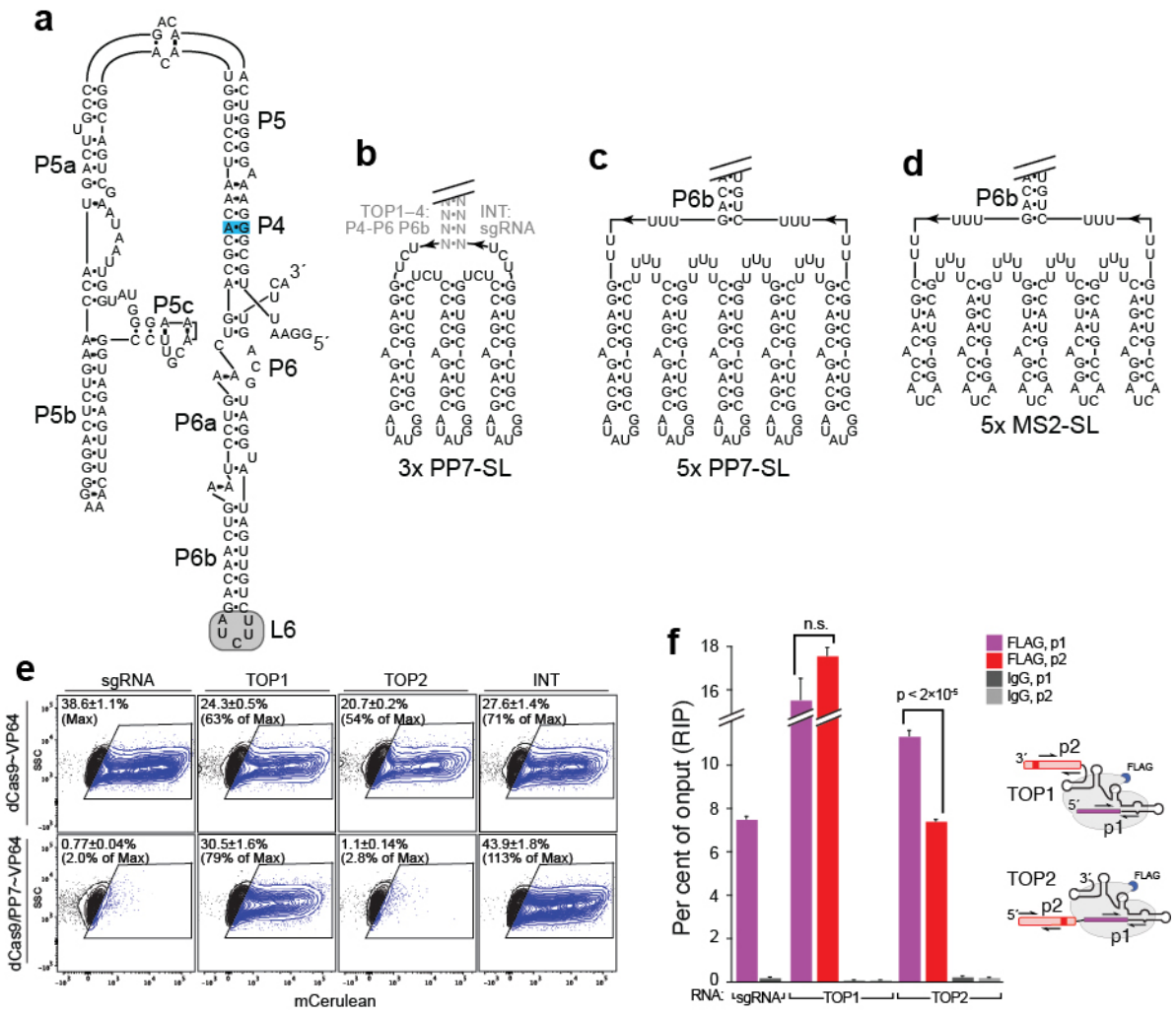
(**a**) *Top*: System schematic and controls. *Top*, *left*: Direct activation by TALE~VP, a chimera of an optimized TALE domain[30,39] and the VP64 transcription activator. *Top*, *middle*: Positive control for bridged activation in the TALE two-hybrid system. VP64 is replaced with the MS2 phage coat protein generating TALE~MS2. This is coexpressed with PP7~VP and a synthetic control RNA, derived from the *T. thermophila* Group I intron P4–P6 domain (**Supplementary Fig. 4a**), appended at its 3´ terminus with five copies of the MS2 stem-loop (5x MS2-SL), and internally with a cassette of five PP7 stem-loops (5x PP7-SL). The resulting ternary complex, TALE~MS2•RNA•PP7~VP, is termed the "5xSL Complex." *Top*, *right*: Negative control for the TALE two-hybrid system. In the "0xSL Complex" the internal cassette of PP7 stem-loops is ablated. *Bottom*: Expression constructs used for this system. Reporter constructs are identical to those outlined in **Supplementary Figure 1c**. (**b**) FACS plots demonstrating direct activation by the TALE system, using integrated reporters. *Top*: baseline reporter expression in the presence of the unmodified TALE domain. *Bottom*: direct activation by TALE~VP. (**c**) Luciferase assays, as in **Supplementary Figure 1f**, using transient (*left*) and integrated (*right*) reporters. Values are means ± standard deviation; *n* = 3. (**d**) Components of the TALE system form the expected binary and ternary subcomplexes. Each control RNA (0xSL and 5xSL) was coexpressed with TALE~MS2 and PP7~VP, and immunoprecipitated with anti-FLAG (TALE) or anti-HA (PP7~VP) antibodies. Lane markers are shared for RIP (*top*) and western blot (*bottom*) panels. Inp, input. *Top*: RIP demonstrates the formation of TALE~MS2•[0xSL RNA], TALE~MS2•[5xSL RNA] and PP7•[5xSL RNA] subcomplexes. n/a, not applicable. *Bottom*: western blot on the same samples, demonstrating the formation of TALE~MS2•[5xSL RNA]•PP7~VP ternary complexes, but, as expected, not the corresponding complex with the 0xSL RNA. Inputs correspond to 2.5% of the starting sample; immunoprecipitates correspond to 12.5% of the recovered material. (**e**) Thermodynamic scheme summarizing the binding events within the TALE two-hybrid system, including the formation of binary complexes (2 and 3) and ternary complex (4) with the target DNA locus. "SL" refers to the 5xMS2-SL cassette. For clarity, reactions involving PP7 are omitted. When using integrated reporters, the concentration of target DNA is expected to be low (~1–2 copies/cell), far below those of TALE~MS2 and RNA~SL. When the concentration of TALE~MS2 furthermore exceeds that of RNA~SL, complex 3 is expected to saturate the target locus, outcompeting formation of complex 4.



Supplementary Figure 3

A split TALE approach to couple DNA-binding to RNA-binding in the TALE two-hybrid system.

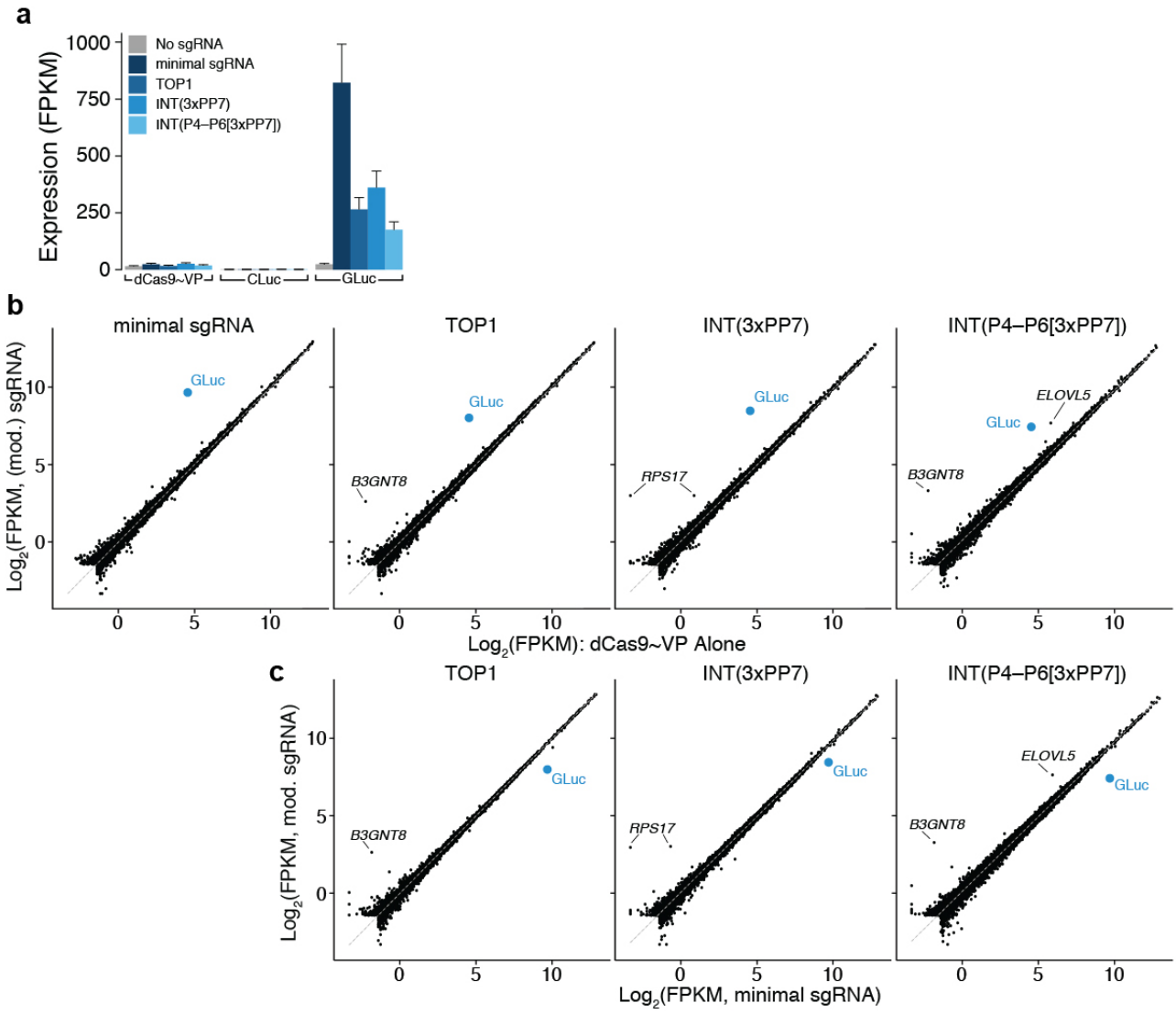
(**a**) System schematic. The optimized TALE domain, which recognizes a 12-nt target[30], is split into two smaller six-nt binding domains, N–TALE (fused to MS2) and C–TALE~VP (fused to PP7). Alone, these proteins should inefficiently recognize the full 12-nt target motif (*bottom right*). Coexpression of a "Splint RNA" bearing both cognate stem-loops should template assembly of a native-like TALE domain with higher competency (*top right*). (**b**) Cartoons summarizing splint RNAs tested. Each construct was under control of a U6 promoter; the splinting cassette (red and blue) preceded a 30-nt unstructured RNA common to all constructs. (**c**) Luciferase assays, demonstrating failure of the split TALE approach in its current design. Using transient reporters (*top*), the TALE domains induce reporter activation in the absence of a splint RNA. Coexpression of splint 1 reduces the apparent activity. n.d., no data. Using integrated reporters (*bottom*), no significant activation is seen with any of the splints. Values are means ± standard deviation; $n = 3$.



Supplementary Figure 4

Secondary structures of TOP1–4, INT, and Double TOP0–2 accessory domains; additional biochemical data.

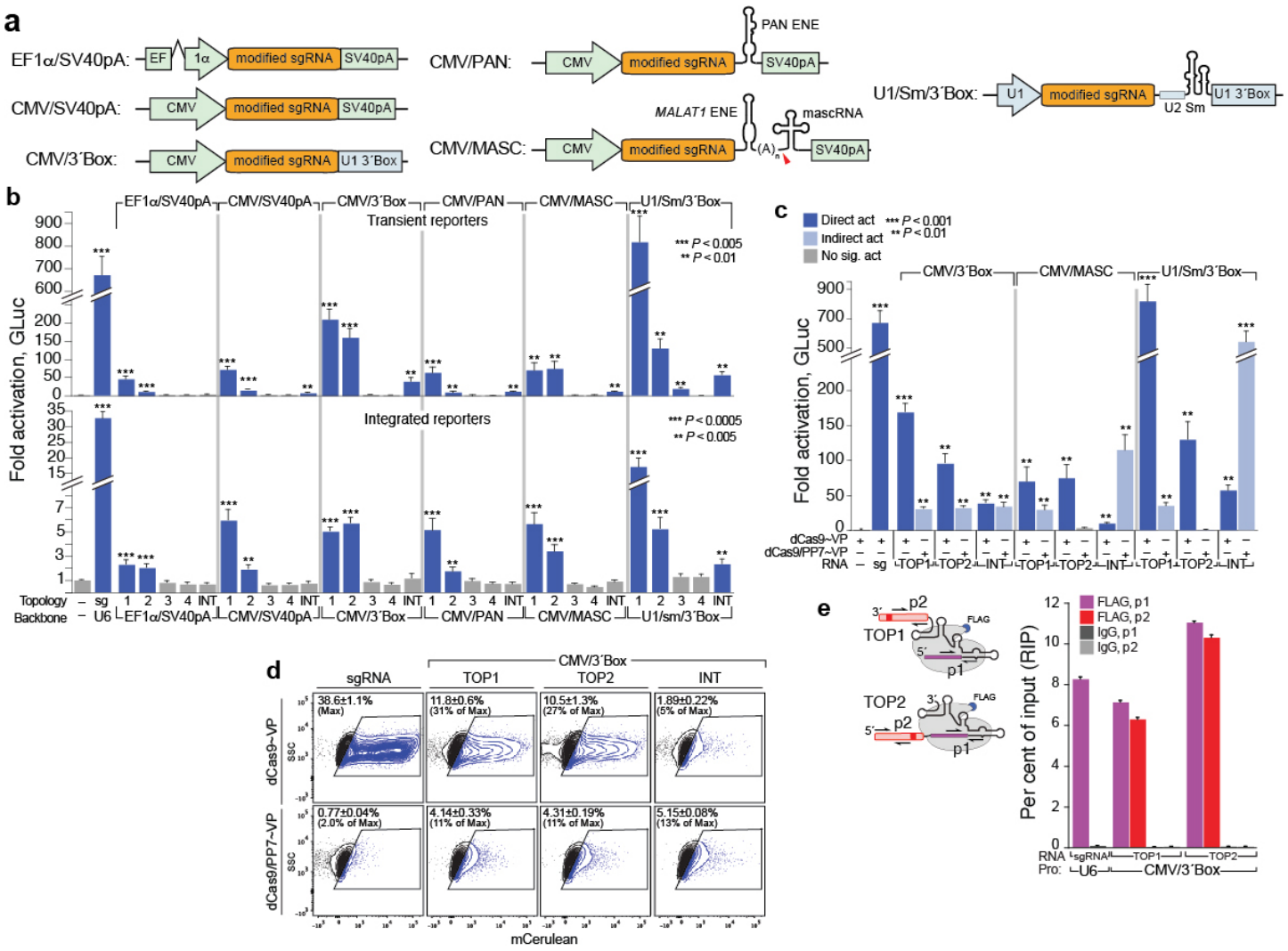
(**a**–**d**) Secondary structures of accessory domain components. (**a**) Secondary structure of the thermostable ΔC209 mutant of the *T. thermophila* group I intron P4–P6 domain, from which the accessory domains of TOP1–4, and their Double TOP analogs, were derived. Base pairing geometries are indicated using Leontis and Westhof nomenclature; the Watson-Crick G•A base pair introduced by the ΔC209 mutation is highlighted in blue. PP7-SL and MS2-SL cassettes (**b**–**d**) were grafted into L6, boxed in gray. (**b**) Secondary structure of the 3xPP7 cassette used in U6-driven topology constructs. In TOP1–4, this cassette is inserted within P4–P6 stem P6b (*top, left*); the identical construct comprises the accessory domain of INT(P4–P6[3xPP7-SL]) (**Fig. 3b,e,f**). In INT, the cassette it is inserted within the sgRNA core (*top, right*). (**c**) Secondary structure of the 5xPP7 cassette used in Pol II–driven topology and double topology constructs (**d**) Secondary structure of the 5xMS2 cassette used in Double TOP0–2. (**e**,**f**) Additional biochemical data, in support of **Figure 1**. (**e**) FACS analyses on transient reporter assays using U6-driven TOP1, TOP2 and INT, as in **Supplementary Figure 1e**. Means ± standard deviation; *n* = 3. (**e**) RIP–qRT-PCR of dCas9•TOP1 and dCas9•TOP2. qPCR primers target the core sgRNA and the accessory domain (p1 and p2, respectively, *right*, **Supplementary Table 3**). Recovery of the sgRNA core and accessory domains from dCas9•TOP1 complexes was quantitative; from dCas9•TOP2 complexes the yield of accessory domain was nearly half that of the sgRNA core. Although this did not indicate a complete loss of the accessory domain, we hypothesize that other factors—such as RNA folding or the geometry of the dCas9•TOP2 complex—may contribute to the loss of TOP2-bridged activation. Furthermore, this result implies that structured 5′ additions might partially stabilize a modified sgRNA from degradation, as has been reported elsewhere[20]. Values are means ± standard deviation. *n* = 4. Student's one-tailed t-test.



Supplementary Figure 5

CRISPR targeting specificity is not significantly altered by appending the sgRNA core with accessory RNA domains.

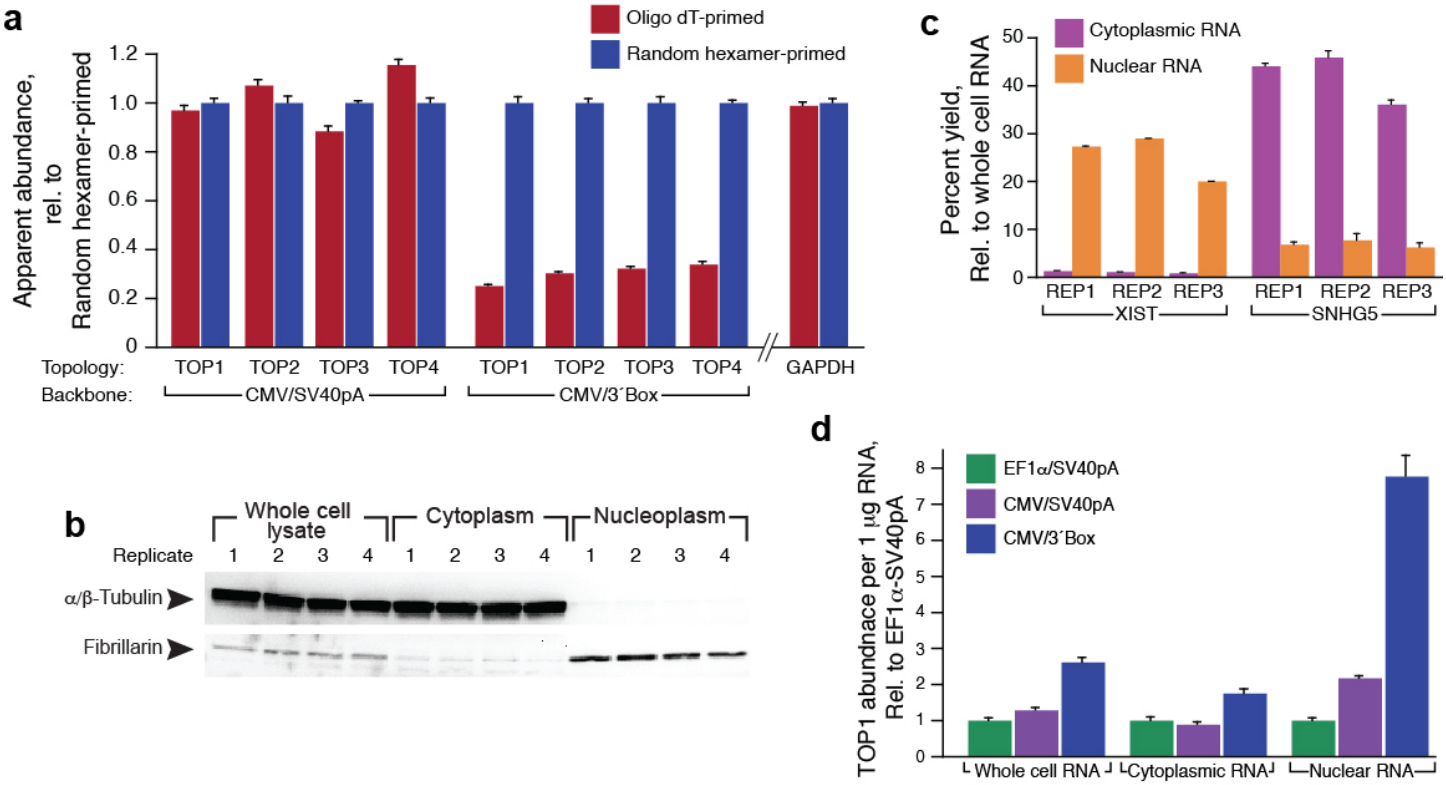
Global gene expression data, measured by poly(A)$^+$ mRNA-Seq, from GLuc reporter HEK293FT cells transiently expressing dCas9~VP alone, or concomitantly with GLuc-targeting minimal sgRNA, TOP1, INT(3xPP7) or INT(P4–P6[3xPP7]) constructs (**Supplementary Fig. 4**). All data are averaged from two biological replicates. (**a**) Specific activation of the GLuc-Reporter by sgRNA and modified sgRNA constructs, observed by mRNA-seq, corroborates observations from luciferase and FACS assays. FPKM, fragments per kilobase million reads, as quantified with CuffDiff[44]. For comparison, values for the CLuc normalizer and dCas9~VP are shown. (**b**) Differential gene expression plots for the minimal and each modified sgRNA construct, relative to cells expressing dCas9~VP alone. All axes are on the same scale. The large blue circles correspond to the GLuc reporter locus; reads mapping to the (modified) sgRNAs–which were not observed in all experiments–have been masked from view. (**c**) Differential gene expression plots for each modified sgRNA, relative to the minimal sgRNA. In **b** and **c**, off-target genes with the highest fold changes in gene expression (*B3GNT8*, *ELOVL5* and *RPS17*) are indicated; these were not measured as statistically significant. Moreover, no PAM-adjacent off-target binding sites complementary to the GLuc gRNA (defined by the 3′-most 10 nt of the guide) were observed within these loci and the adjoining 10 kb. From this, we hypothesize that sporadic activation of these genes may have resulted in response to expression of the modified sgRNA constructs, and not from guide-directed off-target activation by dCas9~VP.



Supplementary Figure 6

Surveying Pol II expression systems for CRISPR-Display.

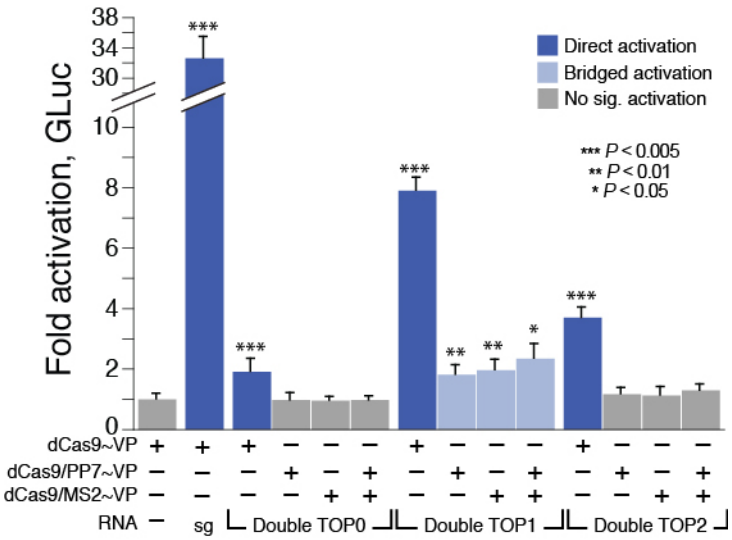
(**a**) Schematics of RNA polymerase II (Pol II) expression systems for modified sgRNA constructs. Promoters and terminator elements are described in (**Supplementary Note 2**). For the CMV/MASC system, the RNase P cleavage point is indicated (red arrow). Modified sgRNA chimeras are as defined in **Figure 1a** and **Supplementary Figure 4a**,**b**. (**b**) Direct activation by Pol II–driven TOP constructs, measured via luciferase reporter assays. Values are means ± standard deviation. *n* = 3, Student's one-tailed t-test, relative to negative controls (*far left*). "sg," minimal sgRNA. (**c**–**e**) Expression from the CMV/3´Box backbone appears to restore function to the TOP2 accessory domain. (**c**) Direct and bridged activation by the most effective constructs, using the most effective Pol II expression systems tested. Transient reporter assays are shown. Values are means ± standard deviation, *n* = *3*; "sg," minimal sgRNA, driven from a U6 promoter. (**d**) FACS analyses on transient reporter assays with CMV/3´Box constructs, as in **Supplementary Figure 1e** and **4e**. Data for the sgRNA controls are the same as in **Supplementary Figure 4e**. Values are the means ± standard deviation. *n* = 3. (**e**) RIP/qRT–PCR of dCas9 complexed with CMV/3´-Box TOP1 or TOP2, as in **Supplementary Fig. 4f**. Positions of qPCR primer pairs are indicated in the schemas (*left*). Unlike U6-driven TOP2, under CMV/3´-Box expression, recovery of the sgRNA core and accessory domains was nearly quantitative. Values are means ± standard deviation, *n* = *4*.



Supplementary Figure 7

The CMV/3´Box system generates non-polyadenylated, nuclear-localized transcripts.

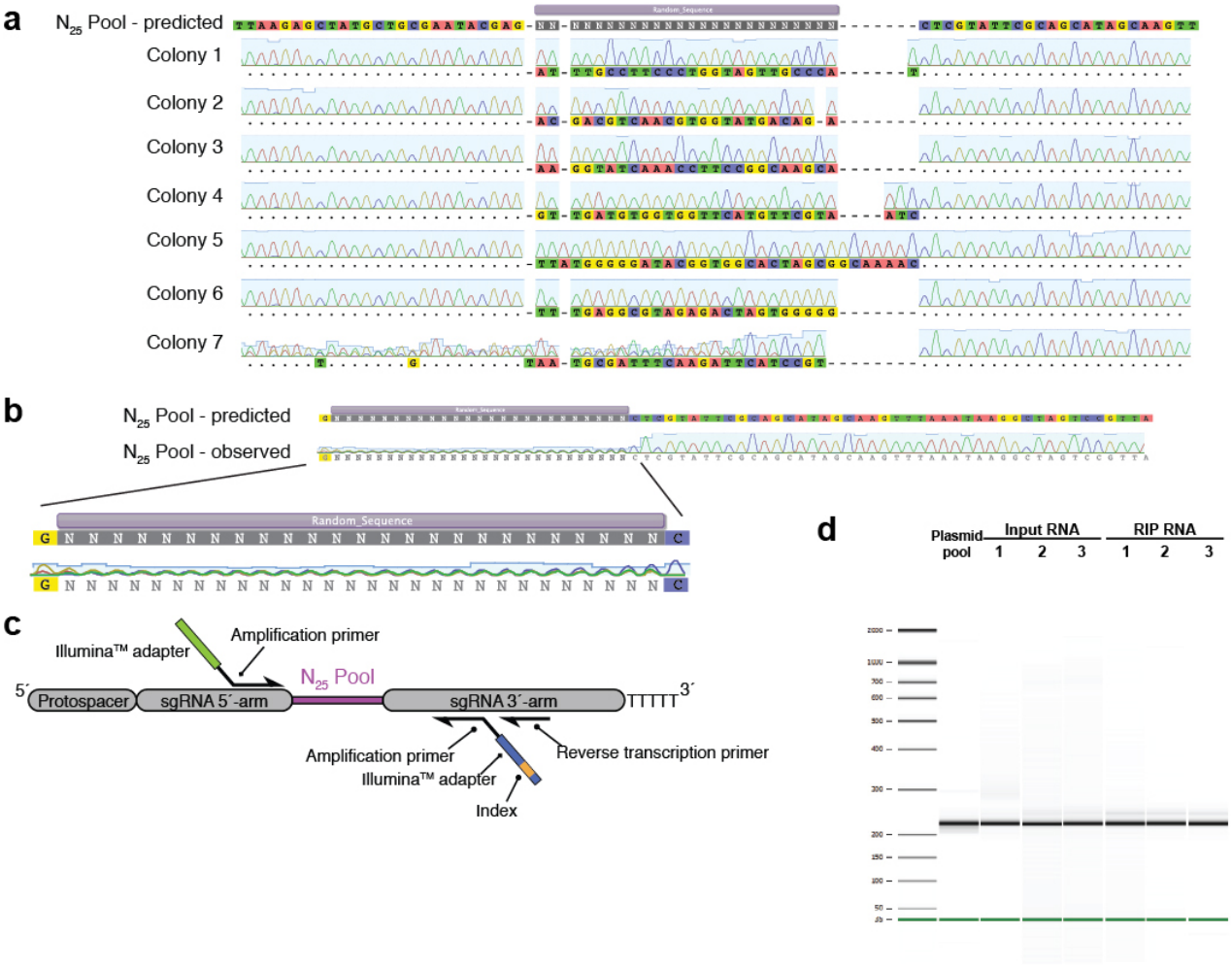
(**a**) The majority of CMV/3´Box transcripts are nonpolyadenylated. Whole cell RNA was isolated from HEK293FT cells expressing TOP1–4 from CMV/SV40pA or CMV/3´Box backbones, and cDNA was synthesized in parallel reactions using random hexamer or oligo-dT primers. The apparent abundance of each construct was measured by qPCR using primer pair p2 (**Supplementary Figs. 4f** and **6e** and **Supplementary Table 3**), and normalized to the signal observed with random hexamers. As a control, endogenous GAPDH was measured using the same protocol. (**b**,**c**) Subcellular fractionation quality controls. (**b**) Western blot analysis of four replicate samples. α/β-Tubulin and fibrillarin are cytoplasmic and nuclear markers, respectively. Whole cell lysates were generated from 5% of the initial samples, prior to fractionation, by boiling cells in RIPA buffer. 1 μg of protein was loaded onto each lane. (**c**) qRT-PCR analysis of three replicates. *XIST* and *SNHG5* are nuclear and cytoplasmic ncRNAs, respectively. Data were processed as in RIP experiments, renormalizing observed C_t values to the total mass of RNA isolated from each subcellular compartment. Whole cell RNA was isolated from 5% of cells prior to fractionation. The percent yield, relative to whole cell RNA, reflects the abundance of transcript in each compartment. Values are means ± standard deviations; four technical replicates. Primers are listed in **Supplementary Table 3**. (**d**) CMV/3´Box transcripts are preferentially nuclear retained. Cells expressing TOP1 from each Pol II backbone were fractionated and analyzed as in **c**. The abundance in each compartment, relative to that observed with EF1α-SV40pA-TOP1-expressing cells, is shown. cDNA was primed with random hexamers; abundances were measured by qPCR using primer pair p2 (**Supplementary Figs. 4f** and **6e** and **Supplementary Table 3**). Values are means ± standard deviations, *n* = 3.



Supplementary Figure 8

Integrated reporter luciferase assays with "Double TOP" constructs.

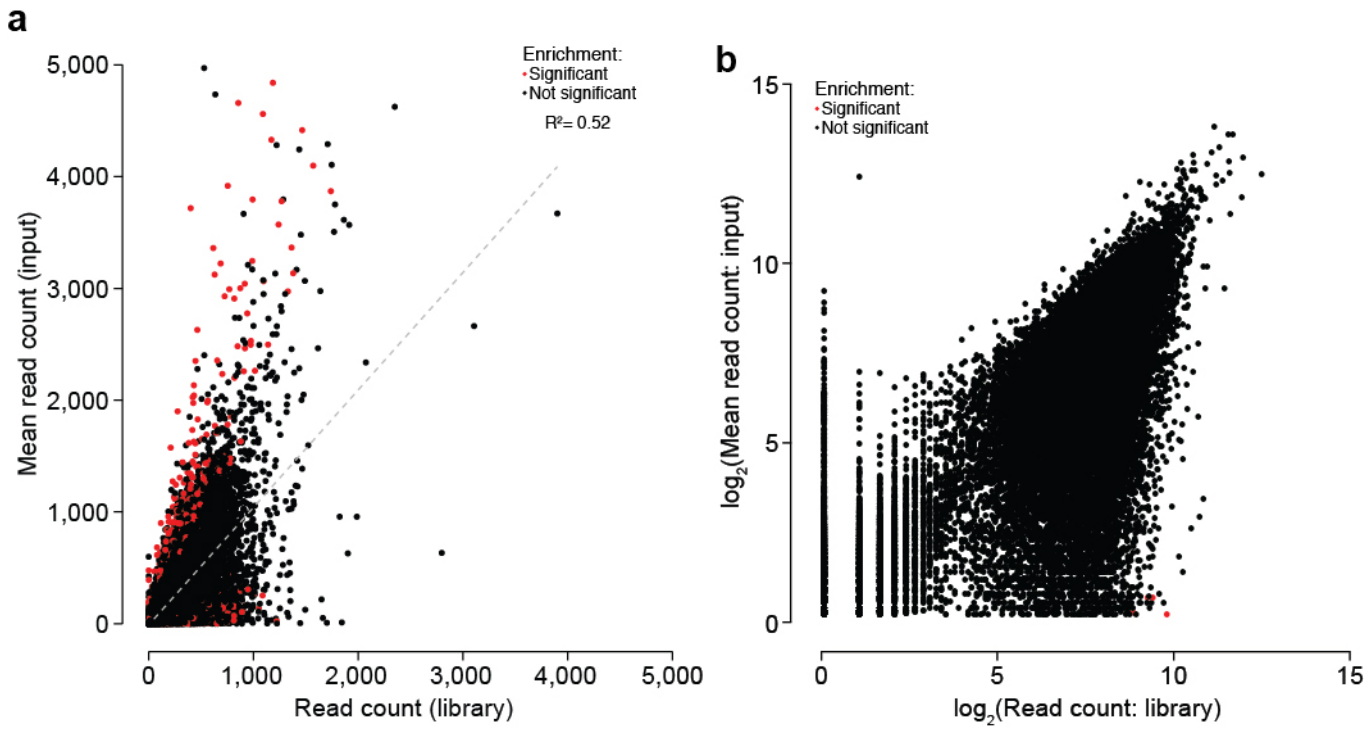
"Double TOP" constructs (defined as in **Figure 2a** and **Supplementary Fig. 4**) were expressed using the CMV/3´Box system; "sg," minimal sgRNA, driven from a U6 promoter. Values are means ± standard deviations, *n* = 3. Significance was measured using Student's one-tailed t-test, relative to negative control cells expressing dCas9~VP alone (*far left*).



Supplementary Figure 9

Synthesizing and sequencing the INT-N_{25} pool.

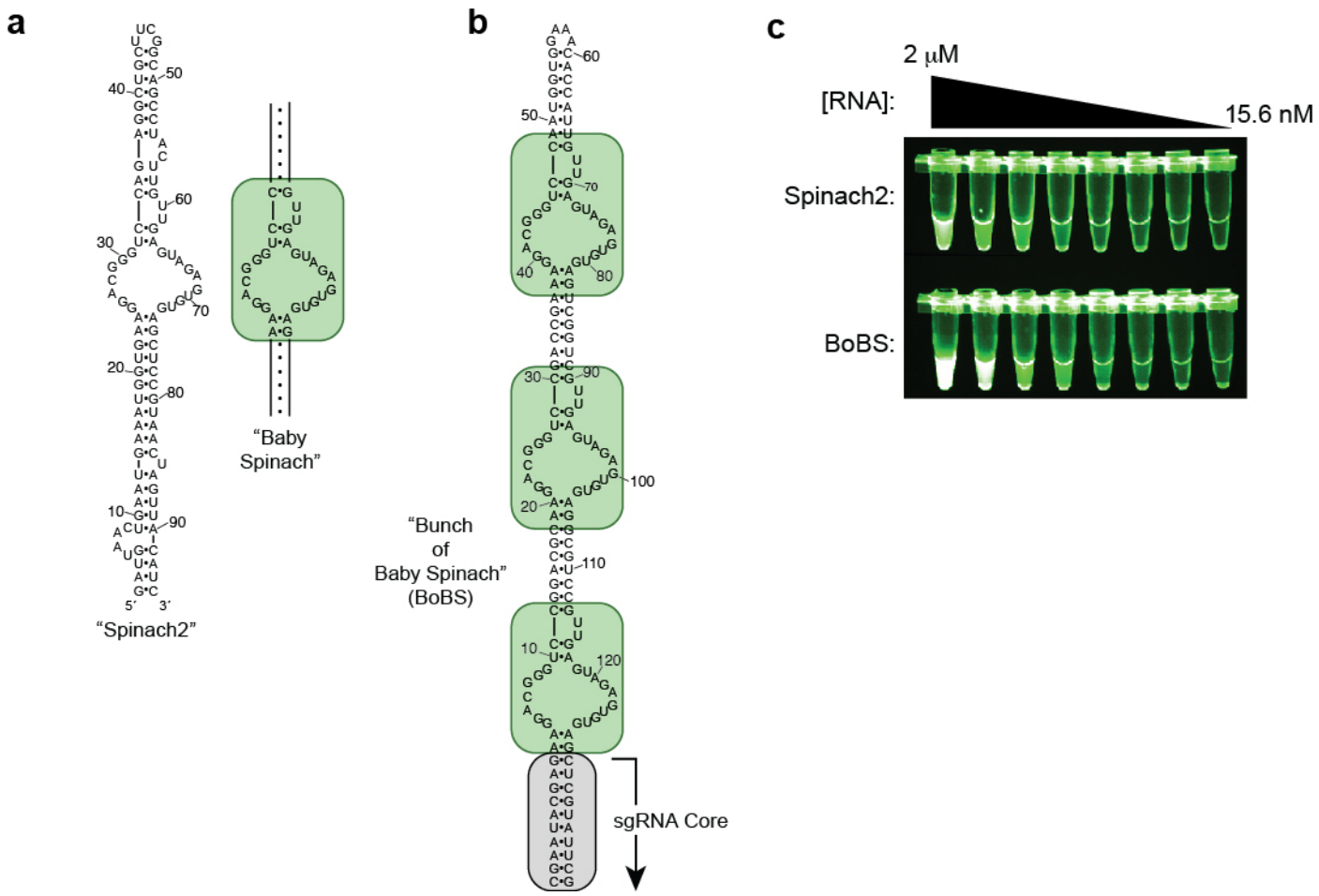
(**a**) Sequences and chromatograms of seven individual clones isolated during initial synthesis of the INT-N_{25} plasmid pool, aligned to the consensus (*top*). Nucleotides that match the consensus are notated as dots. Analysis and alignments were performed in Geneious (Biomatters Ltd). (**b**) Sequencing chromatogram of the aggregate INT-N_{25} pool, aligned to the consensus. Priming was initiated from a M15 reverse site located downstream of the sgRNA 3´ terminus (not shown). Heterogeneity 5´ of the insert is likely due to the small population of molecules containing inserts greater than 25 nucleotides in length (e.g., colony 5, above). (**c**) Schematic summarizing the primer design used to generate targeted deep sequencing libraries. Complete primer sequences are listed in **Supplementary Table 6**. (**d**) Bioanalyzer traces of the final sequencing libraries. For each, the expected length is 220 nt. Green and purple bands correspond to lane markers.



Supplementary Figure 10

Sequence diversity and expression of the INT-N_{25} pool.

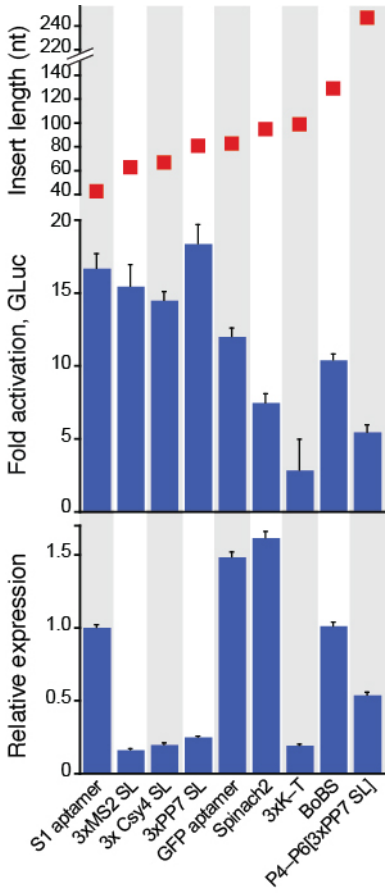
(**a**) Read counts of 25-mers observed in the plasmid pool versus mean read counts observed in input (not immunoprecipitated) RNA libraries. Of the 783,612 unique sequences observed, 524 (0.07%) and 7,011 (0.9%) were significantly enriched or depleted in the input RNA libraries, respectively (red), as determined using DESeq2 (Ref. 48). Sequences containing more than five consecutive uridines, which act as Pol III termination signals, are among the depleted sequences. A zoomed view, representing the majority of data, is shown. (**b**) All data from the same experiment, plotted on a logarithmic scale.



Supplementary Figure 11

Design of the "Bunch of Baby Spinach" (BoBS) construct.

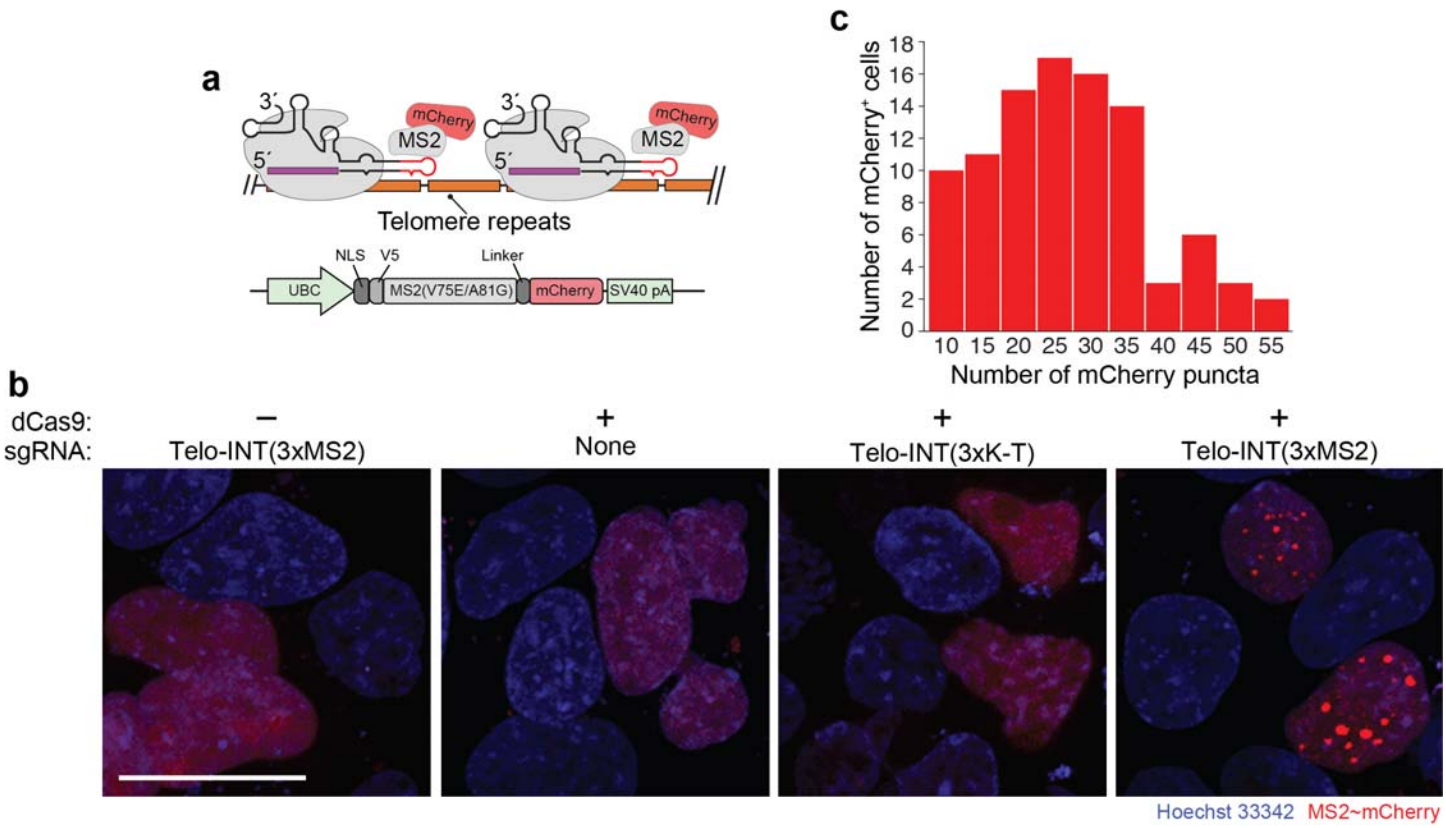
(**a**) *Left*: secondary structure of the 94-nt Spinach2 aptamer[13], as observed in crystal structures. *Right*: predicted secondary structure of a minimal Spinach aptamer, "Baby Spinach." A 29-nt double G-quadruplex core, which is responsible for binding DFHBI–1T, the fluorophore ligand[13], is boxed in green. The core is abutted on both sides by base-paired stems, for which absolute sequence identity is thought to be inconsequential to activity. (**b**) Design of a 129-nt "BoBS" construct. Three tandem copies of the Baby Spinach core (green boxes) are embedded in a single, extended stem-loop, contiguous with the sgRNA core (gray box). (**c**) BoBS is brighter than Spinach2. INT-Spinach2 and INT-BoBS RNAs were prepared by *in vitro* transcription and folded by heat denaturation and buffer addition, as described[13]. Folded RNAs were serially diluted two-fold, incubated with 20 μM DFHBI-1T (Lucerna Technologies) and simultaneously imaged using a conventional UV trans-illuminator (Bio-Rad).



Supplementary Figure 12

Efficacy of INT-like CRISP-Disp constructs partially varies with length and expression level.

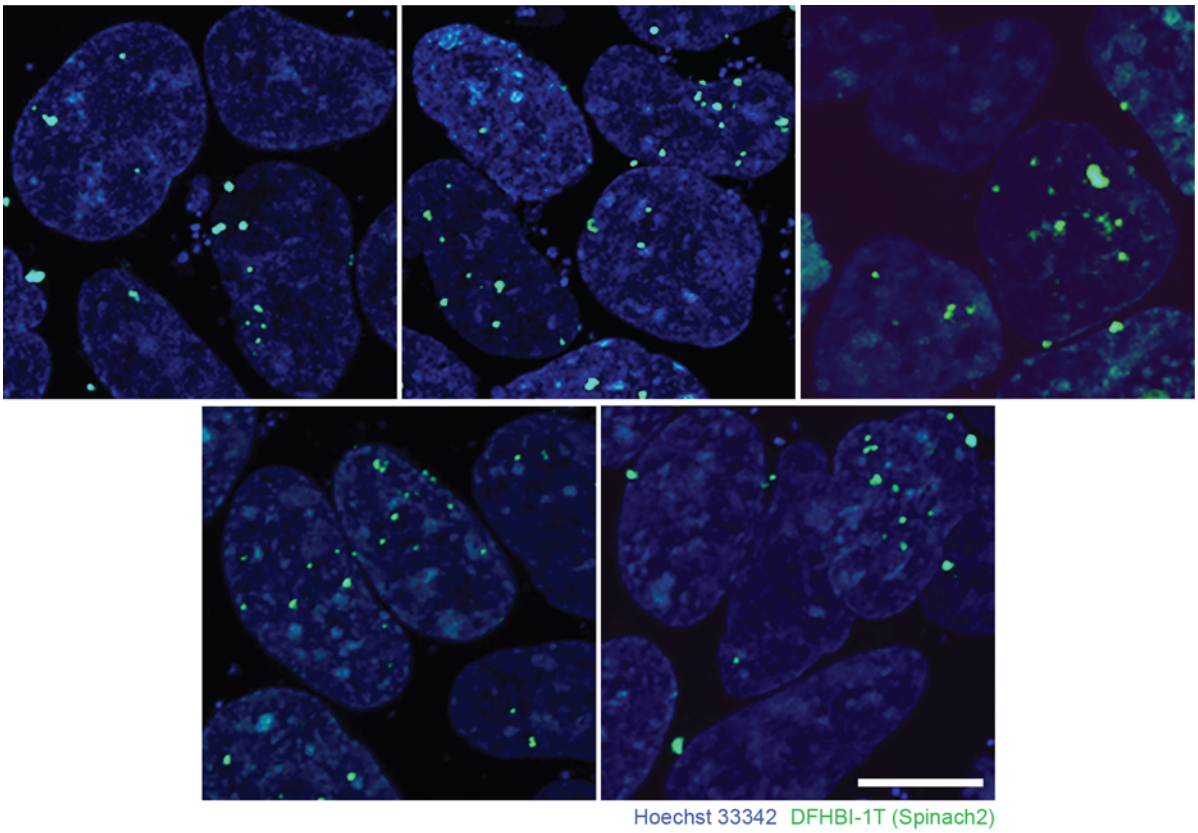
Direct activation activities (using integrated reporters, *middle*) and relative expression levels (*bottom*) for all INT-like constructs examined in **Figure 3e**, sorted according to insert length (*top*). Expression was quantified using RT-qPCR using universal primer pair pINT, which targets the sgRNA core outside of the INT insertion site (**Supplementary Table 3**). Values were normalized to internal *GAPDH* controls, and thereafter relative to the expression of the INT-S1 aptamer construct (*far left*); n = 4. Luciferase values are the same as plotted in **Figure 3e**. The most lowly expressed construct (INT-3xMS2 SL, expressed at 16% the level of INT-S1 Aptamer) exhibited activity within 90% of that of the INT-S1 aptamer construct. Conversely, the most highly expressed construct (INT-Spinach2, expressed at 160% the level of INT-S1 aptamer), exhibited 44% of the activity of the INT-S1 aptamer construct.



Supplementary Figure 13

Bridged imaging of genomic loci with CRISPR-Display.

(**a**) (*Top*) Experimental design. A telomere-targeting sgRNA[18] internally appended with three MS2 stem-loops ("Telo-INT(3xMS2)") binds an MS2~mCherry fusion, and is localized to the telomeric repeats by dCas9. (*Bottom*) Schematic of the MS2~mCherry expression construct. UBC: the human Ubiquitin Chain C promoter; MS2 (V75E/A81G) is a non-aggregating variant[40]; V5, a V5 epitope tag. dCas9 and modified sgRNA expression constructs are as defined in (**Supplementary Fig. 1c**). (**b**) Bridged telomere imaging requires a cognate sgRNA and dCas9. All cells express MS2~mCherry, in addition to the indicated constructs. "Telo-INT(3xK-T)," a telomere-targeting INT-like sgRNA appended with a cassette of three kink-turns (**Supplementary Table 5**). Images are merged z-stacks at 63× magnification, akin to **Figure 4c**. (**c**) Histogram of observed fluorescent puncta in 97 mCherry$^+$ cells. Scale bar, 15 μm.



Supplementary Figure 14

Representative aptamer-based live cell images.

All cells are dCas9^{+}, and express a telomere-targeting sgRNA^{18} internally appended with Spinach2 (Ref. 13), akin to the lower-right field in **Figure 4d**. Images are merged z-stacks at 63× magnification. Scale bar, 15 μm.

Supplementary Note 1

TALE-based two-hybrid approaches appear ineffective at targeting RNAs to individual DNA loci.

In our initial strategy the chromatin-targeting protein "conduits" were based upon transcription activator-like effectors (TALEs, **Supplementary Fig. 2a**), a versatile class of customizable DNA-binding repeat proteins derived from *Xanthomonas sp.*[1, 2]. Synthetic TALE domains are attractive because of the relative ease and modularity of their design: a TALE that specifically targets a given DNA sequence can be designed by altering the repeat-variable diresidues (RVDs) harbored within each of its repeats, according to a simple code[3, 4]. As such, TALEs have been exploited to precisely modulate the genome, transcriptome and epigenome in a wide variety of biological systems. In our design, the TALE was fused to a non-aggregating mutant of the MS2 phage-coat protein, a high-affinity, highly specific RNA-binding domain that has been extensively characterized[5]. Likewise, the ncRNA of interest was fused a cassette of cognate stem-loops recognized by the MS2 protein (MS2-SL). Hence, the TALE~MS2 chimera should serve as a "molecular bridge" that targets the ncRNA~MS2-SL chimera to a DNA locus, in a scheme reminiscent of a classic two-hybrid system (**Supplementary Fig. 2a**).

We confirmed that a TALE domain could regulate our dual reporter system (**Supplementary Fig. 2b**,**c**), through direct activation. In mixed populations of integrated reporter cells, expression of an optimized GLuc-targeting TALE domain fused to the VP64 transcription activator (TALE~VP, **Supplementary Fig. 2a**) dramatically increased levels of mCerulean and *Gluc*, activating them 4.1-fold relative to control cells expressing an unmodified TALE (**Supplementary Fig. 2b**). Some isolated clonal cell lines exhibited even higher direct activation: in one line, mCerulean and *GLuc* expression increased ten-fold in response to

TALE~VP (*data not shown*). Next, we tested if noncovalent TALE•RNA complexes could also modulate reporter function. However, since it is unclear *a priori* which, if any, natural ncRNA can be functionally reconstituted ectopically, we first employed a synthetic ncRNA scaffold designed to serve as a positive control, in a scheme reminiscent of a classic three-hybrid system (**Supplementary Fig. 2a**), and analogous to the "bridged activation" assay we employed in the development of CRISP-Disp (**Fig. 1b**). As in that system, the RNA scaffold (analogous to the CRISP-Disp TOP1–4"accessory domain," **Fig. 1a**) was derived from the *T. thermophila* Group I intron P4–P6 domain, a well studied ~250 nucleotide, autonomously folding motif[6] (**Supplementary Fig. 4a**,**b**). We replaced a dispensable internal loop within P4–P6 (L6b, Ref. 6) with a GAAA tetraloop, or with a cassette of five high-affinity stem-loops for the bacteriophage PP7 coat protein[7], generating "0xSL" and "5xSL" constructs, respectively. These constructs were coexpressed with the TALE~MS2 chimera and with PP7~VP. Hence, TALE recruitment of the 5xSL complex, but not of the 0xSL complex, should concomitantly recruit the PP7~VP protein to the reporter site and activate expression (**Supplementary Fig. 2a**).

Using transient reporters, we observed robust bridged activation by the 5xSL complex (~75-fold activation, $p < 0.005$, Student's one-tailed t-test, relative to cells expressing an unmodified TALE; n=3), corresponding to approximately 31% of direct activation exhibited by the TALE~VP fusion under equivalent conditions (**Supplementary Fig. 2c**). As expected, no bridged activation was observed with the 0xSL complex. In stably integrated reporter cells, however, we did not observe bridged activation using the three-hybrid system: cells expressing the 5xSL complex exhibited *Gluc* levels indistinguishable from those expressing the 0xSL complex or the TALE domain alone (**Supplementary Fig. 2c**, $P > 0.1$, Student's one-tailed t-test relative to cells expressing an unmodified TALE; n=9). This is in stark contrast with the results

Nature Methods: doi:10.1038/nmeth.3433

of the analogous CRISP-Disp experiments. There, bridged activation by U6-driven TOP1 and INT, scaled relative to the direct activation exhibited by each construct, was comparable using transient or integrated reporters (**Fig. 1c**). We confirmed by qRT-PCR that all components of the TALE 5xSL complex were expressed at comparable levels (*data not shown*). Furthermore, RNA Immunoprecipitation (RIP) against the TALE~MS2 and PP7~VP proteins demonstrated that the predicted TALE~MS2•RNA and 5xSL•PP7~VP binary complexes were formed (**Supplementary Fig. 2d**). Co-IP/western –blotting likewise indicated measurable formation of the expected TALE~MS2•5xSL•PP7~VP ternary complex (**Supplementary Fig. 2d**). Hence, the lack of reporter induction cannot be explained by the absence of a critical component or complex.

While the sizable RNA yields in the TALE RIP experiments (approximately 2.0% of inputs, **Supplementary Fig. 2d**) imply that a substantial fraction of the Scaffold RNA is bound by the protein chimera, the converse quantity–the fraction of TALE protein bound by RNA–is opaque to this method. Since translation of the TALE mRNA likely amplifies TALE protein abundance by several orders of magnitude, relative to that of the Scaffold RNA[8], we anticipate that this fraction may be very low. Moreover, because binding of TALE~MS2 chimera to its DNA and RNA targets are essentially independent events, we hypothesized that the lack of bridged activation in integrated reporter cells resulted from the pool of free TALE protein outcompeting the comparably smaller pool of TALE•RNA complexes for the target DNA locus (**Supplementary Fig. 2e**). Hence, the limiting concentrations of target DNA and excess of conduit protein preclude the formation of DNA•Protein•RNA ternary complexes[9]. Our observations with transient reporters are consistent with this hypothesis, since these conditions essentially increase the concentration of available DNA target sites, facilitating population of the

Nature Methods: doi:10.1038/nmeth.3433

ternary complex and, consequently, bridged activation. However, given the difficulty of simultaneously measuring the absolute abundances of unrelated RNA and protein species[8]–as well as their effective K_Ds–*in vivo*, we have not exhaustively tested this model.

In theory, several strategies might be employed to rescue the fundamental stoichiometric imbalance that we suspected of hampering the TALE two-hybrid system. First, we could elevate the ncRNA concentration to approximate that of the TALE component. However, this approach is technically challenging, may require re-optimization for each new ncRNA construct tested, and would move the ncRNA into an expression regime far above physiological levels populated by its natural counterparts[10, 11]. An alternate approach would be to diminish the TALE concentration to approximate that of the ncRNA component. This too is technically challenging: in pilot experiments, titrating the TALE concentration downward–either by decreasing the transfected plasmid mass, or by decreasing the TALE promoter strength–caused a loss of direct activation without facilitating bridged activation (*data not shown*). We suspect that, while lowering the TALE fusion concentration would cause a larger proportion of TALEs to be bound by its cognate ncRNA (for which the K_D is in the low-nM to high-pM range[5]), it would also weaken the binding of the TALE to its DNA target. Hence, the efficacy of this approach would be ultimately constrained by the TALE•DNA K_D, a parameter for which optimization is expected to be cumbersome and target-dependent[12].

A final approach toward circumventing the TALE-ncRNA stoichiometric imbalance entails reengineering the system so as to directly couple the TALE's DNA-binding and RNA-binding activities. Were these two activities intrinsically linked–as is naturally accomplished with the CRISPR system (**Supplementary Fig. 1b**)–then a naked TALE~MS2 protomer would be incapable of competing with TALE~MS2•ncRNA complexes for the DNA target locus

Nature Methods: doi:10.1038/nmeth.3433

(**Supplementary Fig. 2e**). Towards this end, we devised as "split TALE" approach, analogous to split-fluorescent protein methods[13], in which a full-length TALE was divided into two smaller domains that are assembled by approximation on a "Splint RNA" scaffold (**Supplementary Fig. 3a**). In our design, the optimized 12 nt-binding TALE domain[2] was split at the boundary between the sixth and seventh repeats (between GLY34 and LEU1), generating six nt-binding N–TALE and 6.5 nt-binding[2] C–TALE constructs (**Supplementary Fig. 3a**). The N–TALE construct was appended at its C-terminus with an MS2 domain; C–TALE was likewise appended at its N–terminus with PP7. To facilitate RNA-templated assembly, these domains were coexpressed with Splint RNAs bearing cassettes of MS2 and PP7 stem-loops (**Supplementary Fig. 3b**).

To test the viability of our approach, we devised a variation in our Bridged Activation assay, in which C–TALE was appended with a C-terminal VP64 transcription activator (**Supplementary Fig. 3a**). Templated assembly of the N–TALE•RNA•C–TALE ternary complex should therefore recruit VP64 to the reporter locus, mimicking activation by the full TALE~VP construct (**Supplementary Fig. 2a–c**). We surveyed four Splint RNAs–designed in reference to established constructs[13-15] –for the ability to nucleate split TALE assembly (**Supplementary Fig. 3b**). Unfortunately, our design proved unviable in several regards. With transient reporters, coexpression of the split TALE domains alone, in absence of a Splint RNA, was sufficient to induce measurable reporter activity to ~20% of the level observed with the intact TALE~VP construct (**Supplementary Fig. 3c**). This activation was diminished upon Splint RNA coexpression, impling both that high copy-number target DNA was sufficient to template the formation of N–TALE•C–TALE dimers, and that Splint RNA–contrary to its design–diminished this activity. Furthermore, with integrated reporters, we did not observe bridged activation in any context (**Supplementary Fig. 3c**). As observed the full TALE two-hybrid system

Nature Methods: doi:10.1038/nmeth.3433

(**Supplementary Fig. 2c**), we could not explain these results through an absence of any individual component or subcomplex (*data not shown*). We concede that a more systematic exploration of the experimental variables (*e.g.*, testing additional Splint RNA designs, additional TALE~phage coat linkers, different TALE domain cut sites or mutagenizing residues that might stabilize untemplated interdomain complexation) might render the split TALE approach more viable. However, such these experiments appear cumbersome and, given the comparable success of our CRISPR-based system, unnecessary.

Taken together, these data imply that under common experimental conditions, TALE-based two-hybrid approaches–and potentially, mammalian two-hybrid approaches in general–appear to be markedly less robust than CRISP-Disp. This, combined with the simplicity, flexibility and multiplexability of CRISPR-based technologies makes CRISP-Disp the superior method for ectopic ncRNA localization at individual target loci.

Supplementary Note 2

Overview of CRISPR-Display Pol II expression systems

In the initial implementation of CRISPR-Display (**Fig. 1a**), RNA constructs were driven from a human U6 promoter, the standard used in many CRISPR-based applications in mammalian cells (**Supplementary Fig. 1c**). Although extremely powerful, like other RNA Polymerase III (Pol III) promoters, U6 is limited in the length and sequence of its products: Pol III cannot produce transcripts longer than few hundred nucleotides, and is often terminated by stretches of consecutive uridines. Since this would preclude the use of CRISP-Disp with natural long noncoding RNAs (lncRNAs) or synthetic lncRNA-like devices, we sought a method to express functional sgRNAs driven by RNA Pol II.

Nature Methods: doi:10.1038/nmeth.3433

The majority of Pol II transcripts are spliced and polyadenylated. This enhances export into the cytoplasm, which we reasoned would render them unsuitable for CRISP-Disp. Recently, methods have been described that generate unspliced, non-adenylated sgRNAs from Pol II-products[16, 17]. These methods excise mature sgRNAs from longer primary transcripts, using self-cleaving ribozymes[16], by co-expressing them with a sequence-specific ribonuclease, or by exploiting endogenous nucleolytic RNA processing pathways[17]. However, we feared that the RNAs generated via these methods would resemble spontaneous hydrolysis products, which might diminish the stability or complex formation of the sgRNAs' accessory domains. Therefore, we instead sought a biogenesis scheme that generates nuclear-retained sgRNAs *de novo*.

Towards this end, we tested the efficacy of our five extended sgRNA topology constructs (**Fig. 1a**), expressed from six different Pol II backbones (**Supplementary Fig. 6a**). Two ("EF1α/SV40pA", and "CMV/SV40pA") were canonical expression systems, pairing the EF1α promoter (which contains an internal ~1 kb intron) or the Cytomegalovirus (CMV) immediate-early promoter-enhancer, with an SV40 polyadenylation site. Transcripts arising from each system should be capped and polyadenylated, though only those driven by the EF1α-SV40pA backbone should be spliced.

Three additional systems ("CMV/3′Box," "CMV/PAN" and "CMV/MASC") paired the CMV promoter-enhancer with nonstandard Pol II terminators (**Supplementary Fig. 6a**). The "CMV/3′Box" system employed the U1 3′-Box, an ~80 nt element, derived from the U1 small nuclear RNA (snRNA) gene, that directs specialized Pol II transcription termination without polyadenylation[18]. The 3′Box is a modular motif that can functionally replace the terminators of other snRNA genes[18] and of noncanonical Pol II transcripts[19]. Although synthesis of a *bona fide* snRNA 3′ terminus is greatly enhanced by coupling the 3′-Box with an snRNA promoter[20, 21], it

Nature Methods: doi:10.1038/nmeth.3433

has been shown that transcripts synthesized from strong viral promoters and terminated by the 3′-Box are preferentially enriched in the nucleus[22]. This effect is enhanced with longer transcripts[22].

The "CMV/PAN" and "CMV/MASC" systems employed so-called expression and nuclear retention elements (ENEs). The "CMV/PAN" ENE is derived from the Kaposi's sarcoma-associated herpesvirus (KSHV) polyadenylated nuclear (PAN) lncRNA. This ENE facilitates stabilization and nuclear retention of the lncRNA by sequestering its poly(A) tail into an extended RNA triplex structure[23-25]. The *MALAT1* ENE used in our "CMV/MASC" system forms a similar triple-helical structure[26, 27], though it is derived from an endogenous human gene and is generated through a fundamentally different biogenesis pathway. This ENE precedes a genomically encoded adenosine-rich sequence and a tRNA-like structure termed the *MALAT1*-associated small cytoplasmic RNA (mascRNA)[27, 28]. RNase P cleaves the mascRNA from the primary transcript, liberating the poly(A)-like *MALAT1* 3′-terminus[28], which is sequestered within the ENE[26]. Our constructs included all three elements required for this process: the ENE, the A-rich tract and the mascRNA motif (**Supplementary Fig. 6a**).

Our final expression system ("U1/Sm/3′Box") was designed to mimic the biogenesis of most small nuclear RNAs (snRNAs), which are transcribed by a noncanonical Pol II complex, exported into the cytoplasm and subsequently reimported into the nucleus[29, 30]. Reimport, via snurportin[30], requires two elements on the snRNA: a nonstandard 5′-(2,2,7) trimethyl guanosine cap, and an "Sm" box motif, onto which the SM protein complex assembles[29]. To express sgRNAs modified in this manner, we drove transcription from the U1 snRNA promoter, appended the 3′ terminus with the U2 Sm Box, and terminated transcription with the U1 3′Box. The U2 Sm box was chosen because, unlike its U1 counterpart, it is isolated within a modular

Nature Methods: doi:10.1038/nmeth.3433

structural domain at the RNA's 3´-terminus, and can functionally replace the analogous motifs of other snRNAs[18, 29].

Direct activation activities of all constructs, driven from each expression system, are summarized in (**Supplementary Fig. 6b**). As predicted, the capped, spliced and polyadenylated transcripts arising from the EF1α-SV40pA backbone were the least effective, exhibiting activities reduced to ~0.6–6.8% that of a minimal (U6-driven) sgRNA. Activities of the unspliced transcripts arising from the CMV/SV40pA were improved ~2–fold relative to their spliced EF1α/SV40pA counterparts, even though overall RNA expression levels were equivalent (**Supplementary Fig. 7d**). Activities of the CMV/PAN products were essentially indistinguishable from those of CMV/SV40pA transcripts.

However, we observed notable improvements when constructs were expressed from the CMV/MASC, CMV/3´Box and U1/Sm/3´Box systems. Under CMV/MASC expression, direct activation by TOP2 was improved ~10–fold relative to its EF1α/SV40pA, CMV/SV40pA and CMV/PAN counterparts. Using transient reporters, the direct activation activities of CMV/3´Box-driven TOP1, TOP2 and INT were ~25–31% that of the minimal (U6-driven) sgRNA. Overall, however, the most robust system tested was U1/Sm/3´Box, under which the activity of TOP1 exceeded that of the U6-driven minimal sgRNA, and those of TOP2, TOP3 and INT rivaled or exceeded the efficacies of their CMV/3´Box counterparts (**Supplementary Fig. 6b**).

Bridged activation assays for the best-performing constructs (TOP1, TOP2, INT), expressed from the most proficient backbones (CMV/3´Box, CMV/MASC, U1/Sm/3´Box) are summarized in (**Supplementary Fig. 6c**). From these data, two points are notable. First, as discussed in the text, expression from the CMV/3´Box system appears to restore bridged

Nature Methods: doi:10.1038/nmeth.3433

activation to TOP2 (**Supplementary Fig. 6c–e**). Second, the comparable inefficiency of Pol II-driven INT in direct activation assays, relative to its Pol III-driven counterpart (compare **Supplementary Fig. 6b–d** and **Figure 1c–d**), is rescued in bridged activation assays when INT is expressed from the CMV/MASC and U1-Sm-3′Box backbones. This was not observed with INT species expressed from CMV/3′Box (**Supplementary Fig. 6c–d**), EF1α/SV40pA, CMV/SV40pA or CMV/PAN backbones (*not shown*). We hypothesize, but have not tested, that this restoration of activity may be due to an RNA folding issue, wherein misfolded RNAs are funneled back into a viable conformation by the binding of PP7~VP64. This would correlate with the broader observation that, although dCas9 accommodates single insertions at the INT site or on the sgRNA termini, it appears less tolerant to the addition of multiple large accessory domains (as in Double TOP0, **Fig. 2a,b** and **Supplementary Fig. 8**). All of our Pol II expression systems append sequences to the 3′ terminus (poly(A) tails, ENEs, Sm boxes and or 3′Box elements). Of these, the 3′-terminal domains added by the CMV/MASC and U1/Sm/3′Box systems are the most structured and compact: the isolated 3′Box (as in CMV/3′Box constructs) is relatively unstructured in absence of the Sm Box, and all other systems append long poly(A) tails, which may preclude folding and dCas9 complexation.

While the U1/Sm/3′Box system may be of general use for the expression of unmodified sgRNAs and of artificial CRISP-Disp devices, we disfavor its use with natural lncRNAs. Transcripts generated from this backbone would lack a conventional 5′-cap and would be assembled into complexes with Sm proteins; this may influence their subnuclear localization and overall complex formation[29]. For that reason, in our own experiments, we have focused on the CMV/3′Box and CMV/MASC systems.

Nature Methods: doi:10.1038/nmeth.3433

Supplementary Note 3

"Best Practices" for CRISRP-Display construct design

Based on the results presented here, we propose a preliminary set of "best practices" for CRISP-Disp construct design. In general, the conventional U6 expression system was the most robust we employed. Surprisingly, although the U6 promoter naturally drives a product of only ~100 nt, we were able to generate transcripts several times that length at high levels (**Supplementary Fig. 12**); the overall length limitation of this system remains unclear. Under U6 control, by far the most robust insertion site for exogenous RNA domains is the "engineered loop," (**Supplementary Fig. 1b**), which can tolerate diverse, large (>250 nt) and structurally varied inserts. This site is therefore ideal if the RNA domain of interest does not require an exposed terminus, and would not be functionally constrained by placement within a stem-loop. Although we have not tested if other internal insertion points can accommodate exogenous sequences with this degree of modularity, we anticipate that other such amenable points exist, given the structural plasticity of the sgRNA core[31, 32], and its ability to accommodate short stem-loops at other positions[33, 34]. We furthermore hypothesize that certain internal inserts might perturb sgRNA folding or dCas9-binding, potentially in a guide-dependent manner, though we have not examined this exhaustively. We note, however, that problems of this sort appear more pronounced when INT-like constructs are expressed from a Pol II system (**Supplementary Fig. 6b–d**). In lieu of an INT-like design, display on the 3′ terminus, as in TOP1, is viable, although overall efficacy may be limited by the local structure near the attachment point[34].

For substantially larger structures, however, or those containing stretches of poly(uridine), several expression systems are available (Refs. 16, 17, 35, and the CMV/3′-Box,

Nature Methods: doi:10.1038/nmeth.3433

CMV/MASC and U1/Sm/U2 constructs outlined in **Supplementary Fig. 6a**). For example, the activities of constructs expressed from the U1 promoter (a "nonstandard" Pol II promoter that lacks the sequence and length limitations of U6, Ref. 19) nearly rivaled those driven by U6 (**Supplementary Fig. 6b**,**c**). Our U1/Sm/U2 backbone is intended to mimic the biogenesis of natural snRNAs, including their nuclear reimport[29, 30]. While the Sm box is also expected to influence protein complexation and RNA subnuclear localization[29] (and therefore might diminish general CRISPR-Display efficacy), for the purpose of CRISP-Disp it appears dispensable and should be eliminated (*data not shown*). If a canonical 5´-cap is required, expression from the CMV/3´Box or CMV/MASC systems should suffice. If the ncRNA domain requires display on the sgRNA 5´-terminus to function, then only the CMV/3´Box system appears sufficient (**Supplementary Fig. 6c**–**e**). For these larger constructs, we generally prefer TOP1-type (and in the case of CMV/3´Box-transcripts, TOP2-type) chimeras, since INT-like construct efficacy appears to diminish somewhat with increasing insert size (**Fig. 3a**,**b, e** and **Supplementary Fig. 12**), and since Pol II-driven INT-like constructs appear less efficient overall. It should be noted that, for some of the TOP1-type lncRNA constructs tested here, some accessory domain loss was evident (**Fig. 2e**). However, this apparent degradation was not simply contingent upon RNA length, suggesting that RNA structure might contribute to a particular sgRNA~lncRNA's overall stability, as has been observed with smaller accessory domains[34]. Therefore, when implementing a CRISPR-Display lncRNA experiment, we anticipate that one may need to screen several constructs, surveying subdomains within the lncRNA, and or altering the structure at the sgRNA attachment point, so as to maximize construct efficacy and stability.

Nature Methods: doi:10.1038/nmeth.3433

Supplementary Tables

Supplementary Table 1: Sequences of the Target and Non-target cassettes in the "Reporter" and "Normalizer" constructs, respectively. Target/Non-Target motifs are boxed in gray; PAM sequences in beige; the minimal CMV Promoter in green.

Construct	Sequence
Reporter (GLuc/ mCerulean/ Hygro^{R})	5´–ATCTAGATACGACTCACTAT AGG CAAAGCTCTA– ATCTAGATACGACTCACTAT AGG AAAGGAAGCAGCCAC– ATCTAGATACGACTCACTAT AGG ATAACGTTAG– ATCTAGATACGACTCACTAT AGG GTGAAGTAGTCTTTGCGGTA– ATCTAGATACGACTCACTAT AGG CAAAGCTCTA– ATCTAGATACGACTCACTAT AGG AAAGGAAGCAGCCAC– ATCTAGATACGACTCACTAT AGG ATAACGTTAG– ATCTAGATACGACTCACTAT AGG GTGAAGTAGT– ATCTAGATACGACTCACTAT AGG ATCCACGTATGTCGAGG– TAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGC–3´
Normalizer (CLuc/ Venus/ Puro^{R})	5´–ATCTAGATCGCCCGTCCCCT AGG CAAAGCTCTA– ATCTAGATCGCCCGTCCCCT AGG AAAGGAAGCAGCCAC– ATCTAGATCGCCCGTCCCCT AGG ATAACGTTAG– ATCTAGATCGCCCGTCCCCT AGG GTGAAGTAGTCTTTGCGGTA– ATCTAGATCGCCCGTCCCCT AGG CAAAGCTCTA– ATCTAGATCGCCCGTCCCCT AGG AAAGGAAGCAGCCAC– ATCTAGATCGCCCGTCCCCT AGG ATAACGTTAG– ATCTAGATCGCCCGTCCCCT AGG GTGAAGTAGT– ATCTAGATCGCCCGTCCCCT AGG CCGGATCCACGTATGTCGAGG– TAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGC–3´

Supplementary Table 2: gRNA sequences used in this work.

Target	Sequence	Reference
GLuc	GAUCUAGAUACGACUCACUAU	This work
ASCL1–1	GCUGGGUGUCCCAUUGAAA	Ref. 39
ASCL1–2	GCAGCCGCUCGCUGCAGCAG	Ref. 39
ASCL1–3	GUGGAGAGUUUGCAAGGAGC	Ref. 39
ASCL1–4	GUUUAUUCAGCCGGGAGUC	Ref. 39
IL1RN–1	GUGUACUCUCUGAGGUGCUC	Ref. 39
IL1RN–2	GACGCAGAUAAGAACCAGUU	Ref. 39
IL1RN–3	GCAUCAAGUCAGCCAUCAGC	Ref. 39
IL1RN–4	GAGUCACCCUCCUGGAAAC	Ref. 39
NTF3–1	GAGCGCGGAGCCAUCUGGCC	Ref. 37
NTF3–2	GCGCGGCGCGGAAGGGGUUA	Ref. 37
NTF3–3	GCGGCGCGGCGCGGGCCGGC	Ref. 37
NTF3–5	GCGGUUAUAACCAGCCAACC	Ref. 37
TTN–1	GCCUUGGUGAAGUCUCCUUUG	Ref. 57
TTN–2	GAUGUUAAAAUCCGAAAAUGC	Ref. 57
TTN–3	GGGCACAGUCCUCAGGUUUG	Ref. 57
TTN–4	GAUGAGCUCUCUUCAACGUUA	Ref. 57
Telomere	GUUAGGGUUAGGGUUAGGGUUA	Ref. 36

Supplementary Table 3: qPCR primers used in this work, in the order cited

Target	Forward Primer	Reverse Primer	Name[a]	Efficiency[b]
sgRNA(GLuc)	AGATACGACTCACTATGTTTAAGAGC	TCAAGTTGATAACGGACTAGCCT	p1	0.848
P4–P6	CAGCCGTTCAGTACCAAGTCT	GGACCATGTCCGTCAGCTT	p2	0.858
GAPDH	TTCGACAGTCAGCCGCATCTTCTT	GCCCAATACGACCAAATCCGTTGA		0.847
ASCL1[c]	GGAGCTTCTCGACTTCACCA	AACGCCACTGACAAGAAAGC		0.848
IL1RN[c]	GGAATCCATGGAGGGAAGAT	TGTTCTCGCTCAGGTCAGTG		0.827
NTF3[d]	GATAAACACTGGAACTCTCAGTGCAA	GCCAGCCCACGAGTTTATTGT		0.851
TTN[e]	TGTTGCCACTGGTGCTAAAG	ACAGCAGTCTTCTCCGCTTC		0.851
Double P4–P6	GGATGCAGTTCACACCTCCA	CCTTTCCCGCAATTCCGAAG	p3	0.840
pRNA_GSP1[f]	GTCGGTGACGCGACCT	*TAACTTGCTACGAATACGAG*TCC		0.892
*FALEC*_GSP1	GCAGGTTTCACAGAGGGAAGA	CCACTGAGGACACCGACTAC		0.852
*FALEC*_GSP2	AGTCGGTGTCCTCAGTGGTA	AAGAGCAGGCTACAAGTGCT		0.859
*TRERNA1*_GSP1	GTGGTTTTACGTGGCCGATT	GCCTGACGTGAAGTAGCTTT		0.873
ncRNA-a3_GSP1	AGTACCCGACGAGCGTTATG	AGGCTGGTACAGATGGGTCT		0.814
ncRNA-a3_GSP2	GGAGTTTGCAGTGAGCCAAG	ACGAATCGAGAAAGAGCCTCA		0.852
RepA_GSP1[f]	*AGTCGGTGCTTCATT*CACTCT	GCCCCGATGGGCGAATAA		0.815
RepA_GSP2	GGGTTGTTGCACTCTCTGGA	TCATTCTCTGCCAAAGCGGT		0.884
RepA_GSP3	AAGGTCTTGCCGCAGTGTAA	CAACGCCTGCCATATTGTCC		0.845
*HOTTIP*_GSP1	ATGGTAGGGTGTTGGTGCTG	CCCAGAACCCCTCGACAAAA		0.861
*HOTTIP*_GSP2	TCTCGCCTCTGACTCTGTTC	GAAGAGTCGGTAAACACCGC		0.888
*HOTTIP*_GSP3	TTACGCCCGCAACAAAACAG	CCCTCCTTCCTTCAAACGCT		0.886
*HOTTIP*_GSP4	TTCCACCTTTGCCCGATACA	GGAGATGGGTACCTAGGGGT		0.871
*HOTTIP*_GSP5	GCTTGGCAACTTCAGAAAGCA	AGCAGCCGGGTAGTGTAAAA		0.871
XIST	CCCTACTAGCTCCTCGGACA	ACACATGCAGCGTGGTATCT		0.850
SNHG5	GTGGACGAGTAGCCAGTGAA	GCCTCTATCAATGGGCAGACA		0.844
INT(GLuc)	CTCGTATTCGCAGCATAGCAA	TTCAAGTTGATAACGGACTAGCCT	pINT	0.891

[a] As referred to in the main text, **Fig. 2a,c** and **Supplementary Figs. 4f**, **6e**, and **12**

[b] As calculated using Realtime qPCR Miner (Ref 58.)

[c–e] After Refs. 39, 37 and 57, respectively

[f] Italicized nucleotides correspond to linker regions joining the lncRNA core to the sgRNA backbone

Supplementary Table 4: Sequences of natural lncRNA constructs used in this study

pRNA

CGAUGGUGGCGUUUUUGGGGACAGGUGUCCGUGUCGCGUGUCGCGCGUCGCCUGGGCCGGCGGCGUGGUCGGUGACGCGACCUCCCGGCCCCGGGGA

FALEC **(ncRNA-a1)**

GCGCAUCUCCUACGGCCUCCAGGACAGAGGAACCGGGGGAGGCAGGGGGAAAAGGCCGGCCCAGCAAUUCCCCUACCCCCCGGUCCCACGUGUACCCUCCUGGCCU
GGGUCGCCCCAGCCCACGGGGAGCGGGCGGAGUCCUGGCCCACGAAGCCUUGUCACCUGGCGGGCGAAUCCGCAAGCGGAGACUUGUCUUUAAAGGGCUUUGGGCC
GGGCGCGGUGGCUCAUGCCUGGAAUCCCAGCACUUUGGGAGGCCGAGGCGGUGGAUCACGAGGUCAGGAGUUCAAGACCAGCCUGGCCAAGAAGCUCAUACUGACU
AAGGCAGCAGAACAUACAGGAGGAAGAGGAGCAGGUUUCACAGAGGGAAGACAUGAGUUCAAUUUUGGACUUCUCAGUAGUCGGUGUCCUCAGUGGUAGCAACUUC
AAACGGAAGGUGUCAAAAGUCAAAUUCUGGAGAGUUGAGUAUGAAUGGGAGAUGAAGAAAAGGAGGCAGCACUUGUAGCCUGCUCUUAAUGUAUUUCUGCACUCUA
CACUAGCAGCCUAUUACACAGGACACUUGGAUGUCU

TRERNA1 **(ncRNA-a7)**

CCGUUCCUGCCUCCCACAGACACCUAUUAAGUGCCUCCAGUUUUAGGAACUGGGUAUAGAUAUGGCUGGGAACAAAAUGGAAGAAAACAAAAACAAAAAUCCCUGC
CCUCAUGGUGCUGACUCUCCUGUGGCAGAGACGGAGAAGAUGAACAGGGAUUUUAUACCAGGCGUCAGAAGGGAACCAGUGCUAAAGAAAAUGAAAACACCAGGCC
GGGAGAGGCAGCUGGCAUGCGGGCCGUGGUGGUUUUACGUGGCCGAUUUGAGAGAGUGAGACCCCUGGGGUCUUGGAGCCAGGCCUGGGAAAAGCUACUUCACGUC
AGGCCAGGGGCUGUAGCCCUGGCAACCUCCACUCCGCCUGGAAAUCCUCCACCUCGGGGCCUCUCUUUGCCCAGACCUGGCCCAGGAGGAGCACAUGGGAGCCGGG
ACCUUCCCAACAAUCCUUGCCGUUGGCUCCACAAACCUCAGCCAGUCCUGCAACCUGGGAUGCCUUUUCCACCAGGAUGCCUGCUACUGUCACUGUUGUCAUUAGA
UAAUUAAUGAACUAUAAUUAGAAAUCAUAUCAAUAAAAUUUCACAGUCUAAGGCUGUUGAAAUAGG

ncRNA-a3

GAAGUUGAGCUUCAGGCGCGGCUCUUCCCCGUCACACUGGGACCGGACGCAUUUCCAUGGCGUGGUCCCAGGAACCUCUCAGAGUGAACUGAAUUGGAUGCAAGAU
CACGGUGCGUCAGAGCUAGCAAGAUCCUUAGGAUCAUUUAGCCUGGUUUACUAAUAUUACACUAUGGAAUUUAAGCCCAAGGAAUGGAGAGUACCCGACGAGCGUU
AUGUAAGGAGUCGAGUGAGAAGUAAGCUGGAUGCUCUGCUUGGCUGGCAGGUACAGAAUGUGCCAGACCCAUCUGUACCAGCCUGGAUCUCUUGAGGCAUCAGCAC
AAUGGACCUGGCCACACCAGUUUAUUCCACACCGCUGAGGCUGGUCUUUGAGGAAUCACCACACUGUCUUCCACAAUGCACCAUGGAAUACUAUGCAGACAUAAAA
AGGAAUGAGAUCAUGUCGUUUGCAGGGACAUGGAUGGAGCUAGAAGCCAUUAUCCUCAGAAAACUAACACAGGAACAGAAAACCAAAUACCACAUGUUCUCACUUA
UAAGUGGGAGCUGAAUGAUGAGAACACAUGGACACGUGGUGCGGGAACAAUACACCUGGGGCCUGUUGGAGGGUGGGGGCUGGGAGGAGGGAGAGCAUCAAGAAUA
GCUAAUGAGGCCAGGCACAGUGGCUCACGCCUGUAAUCCUAGCAUUUUGGGAGGCUGAGGCGGGCAGAUCAUUUGAGGUCAGGAGUUUGAGACCAGCCUGGCCAAU
AUGGUGAAACCCCGUCUCUAUUAAAAAUACAAAAAUAUUAGCCAGGCAUGGUGGCAAUGCCCGUAGUCCCUGCAACUUGGGAGGCUGAGGCAGGAGAAUCGUUUGA
ACCUGGGAGGUGGAGUUUGCAGUGAGCCAAGAUCGCGCCACUGCACUCCAGCCUGGGCGACAGAGUGAGGCUC

RepA

CACUCUCUUUUCUAUAUUUUGCCCAUCGGGGCUGCGGAUACCUGGUUUUAUUAUUUUUUCUUUGCCCAACGGGGCCGUGGAUACCUGCCUUUUAAUUCUUUUUUAU
UCGCCCAUCGGGGCCGCGGAUACCUGCUUUUUAUUUUUUUUUCCUUAGCCCAUCGGGGUAUCGGAUACCUGCUGAUUCCCUUCCCCCUCUGAACCCCCAACACUCUG
GCCCAUCGGGGUGACGGAUAUCUGCUUUUUAAAAAUUUUCUUUUUUUGGCCCAUCGGGGCUUCGGAUACCUGCUUUUUUUUUUUUUAUUUUUCCUUGCCCAUCGGG
GCCUCGGAUACCUGCUUUAAUUUUUGUUUUUCUGGCCCAUCGGGGCCGCGGAUACCUGCUUUGAUUUUUUUUUUUUCAUCGCCCAUCGGUGCUUUUUAUGGAUGAAA
AAAUGUUGGUUUUGUGGGUUGUUGCACUCUCUGGAAUAUCUACACUUUUUUUUGCUGCUGAUCAUUUGGUGGUGUGUGAGUGUACCUACCGCUUUGGCAGAGAAUG
ACUCUGCAGUUAAGCUAAGGGCGUGUUCAGAUUGUGGAGGAAAAGUGGCCGCCAUUUUAGACUUGCCGCAUAACUCGGCUUAGGGCUAGUCGUUUGUGCUAAGUUA
AACUAGGGAGGCAAGAUGGAUGAUAGCAGGUCAGGCAGAGGAAGUCAUGUGCAUUGCAUGAGCUAAACCUAUCUGAAUGAAUUGAUUUGGGGCUUGUUAGGAGCUU
UGCGUGAUUGUUGUAUCGGGAGGCAGUAAGAAUCAUCUUUUAUCAGUACAAGGGACUAGUUAAAAAUGGAAGGUUAGGAAAGACUAAGGUGCAGGGCUUAAAAUGG
CGAUUUUGACAUUGCGGCAUUGCUCAGCAUGGCGGGCUGUGCUUUGUUAGGUUGUCCAAAAUGGCGGAUCCAGUUCUGUCGCAGUGUUCAAGUGGCGGGAAGGCCA
CAUCAUGAUGGGCGAGGCUUUGUUAAGUGGUUAGCAUGGUGGUGGACAUGUGCGGUCACACAGGAAAAGAUGGCGGCUGAAGGUCUUGCCGCAGUGUAAAACAUGG
CGGGCCUCUUUGUCUUUGCUGUGUGCUUUUCGUGUUGGGUUUUGCCGCAGGGACAAUAUGGCAGGCGUUGUCAUAUGUAUAUCAUGGCUUUUGUCACGUGGACAUC
AUGGCGGGCUUGCCGCAUUGUUAAAGAUGGCGGGUUUUGCCGCCUAGUGCCACGCAGAGCGGGAGAAAAGGUGGGAUGGACAGUGCUGGAUUGCUGCAUAACCCAA
CCAAUUAGAAAUGGGGGUGGAAUUGAUCACAGCCAAUUAGAGCAGAAGAUGGAAUUAGACUGAUGACACACUGUCCAGCUACUCAGCGAAGACCUGGGUGAAUUAG
CAUGGCACUUCGCAGCUGUCUUUAGCCAGUCAGGAGAAAGAAGUGGAGGGGCCACGUGUAUGUCUCCCAGUGGGCGGUACACC

(*continued on next page*)

Supplementary Table 4 (*continued*):

HOTTIP

GUGGGGCCCAGACCCGCGCAACCAGGCGGGGAGGGGAGGUGGGCGCGCGAUUGGGUUGCGAUCUGGAGCAGUGGGGACAGGUCAGGAACCGGCGCGUAUUUCUGCA
GUGAGACCACAGGACGGACAUCGGCGCCUUCGGCUUCGAUGGAGUUGCGAUUUUGCUCUUUCCAGGGAAACAGUGGCAGGGUGUUUGCUGCUUAUCGGUUCCUGCG
GAUAUGCCUGGGUCCCAGGACAUUCCACUGGAGGCUUGGACUGCAUUUAGGAGCCCCUAUCCCUUCCCUGUCCACACUGUUAGUGAGCAAUUUCAUAUGUUUGCAU
UUAGACCCAUAGACUCAGAACGACUCAUCACACACACACACAGUGUACACUGACACACUCACAUUCGCACACUUAGGUAUACAGCCUGAUCCUUGCUCUGACCUGG
UAACAACGCUUCCUCCUCCAGAGACUUUGAGAUAGAGCGAGCGAUCCCUGUGCACCAUUCAUCCAUGCUCCCACCUCGCCAGUAUGGCUGGCUUAGUUCUGGAAGG
GGCUUAAGAGGAACAAGCCCCAGCUGUGCUUCUGGCUGGGACUUAAACCCCCCUUCUGGGCCCUAAAGCCACGCUUCUUUGUGGACCGGACCUGACUCUCCAGGAA
UCUGGGAACCCGCUAUUUCACUCUAUUUUGGGACAAGAAAAAGGGGCUCUUUGGGGCCACUUCCUGCCUUCCCCUCAAGUAGGAUCUCCAGCCUGCAGAGGGUGCC
UAGUCCUUCUUUGCCCAAGAACCAGUCCAAGAAGCCUUUCCUCUGUGCCUGGGAAAUGCAACCUUUUCUUGGGAGCAUGGUAGGGUGUUGGUGCUGAAGAACCAAG
CAGCGACCCGUCUUGUAGCUGCCAUGUUUUGUCGAGGGGUUCUGGGGGUCCUGCUGCUUUAGAGCCACAUACUUCCACUUCCUGAUUCACUACUGUGAGCUGGUCA
GAUGCCUAGAAGAGGAACAAGCGUUCAAAGUGAAAGUGGGCACAUUACCGGAAUAGUGCUGGGGAGAGUGCUGGAUUCUUUUCCACCCCAGGCGGACUGGUGAGAA
GCCAGGCUUGGACCUGUCCUCUGCUCCUAGCUUGCACACUCAGCCCUAAACUCAGAGCAGCACGCAUACCACCCCUCACACACACCCCACCAUCUGCUGUCUAAGG
CCCCUGGGCUUCCUGCAGGAUCCAGACCAAUGUGGCUGGGCUUGGGCUUUUAUCUGUCCUGAUCCUGGAUUUGUCCUGACCAAUGUAAGUGUCGCCCAAUAAAACC
UUCUAUGACCCCCACACCAGCCACCCCCCCACCAAGUGUGCCCUUUCCUUCUUGACUUUUUAGCAGUUCUGGGUAAAUAUUGAUUUGCCCCCAGUUUACCUUCUCC
CUGACUGGCCAUUUGCAGACUCAGGAACUAGCCUCUGUAGGGACUUGAUUUUUCUGUUACUUUCUGGCCGUUUCACCACCCCCCUUCCUCCCCUCCAAGUGGCAUUG
UAAAACUCACAGUGACAAAGAGACAGAGUAGGGUUCUAGGCCCCUGUUCCUGGGGACUUGAAGGCGGUUUUACAUACUGGUCAGACACGGCUGGAGGCCAAGGUCA
AGUUGAAAGUUGCAGUCCAGCCAGCAUGAGAACUGCCAUGCGAGCGUAGAGACACAGGCAGCAGCAAAAGGCCCAUUGCCCACAUCCCCCUCACUCUUAAUUUUCUC
UCUCUUUUUAAAAUUCUCGCCUCUGACUCUGUUCGGCUGCCCAGAAUUUUUUGGUGCCUUCGUGGGGUUUUUGGGGCGGUGUUUACCGACUCUUCUCUGCCUCCGC
CCUGCUCAGCCAGGGCUUUGAGCCUCUUCGGUUUUCCGGCCAGACCCGGAAAAACGAAAACACAGCUUGGGGAGCCCCCACUAGCCGGCGCCUGUGCCAGCUCACC
UCUGGCCAUGGCGCAGCUGCCGGUGCACACGGCGGCCAAGGCCAGCUCCACAUUCUUCCCUCCCCCUCCCACUUCACCGUAGCCCCGAACCCUGCGCGCAGAGAAA
GGGUCUCAGCUCCACAGACGACUGGGUCCCUCCUCACCAAAAAUGGUGAGACAAGAUUUCAUCUGUCGGCCGAGGAGCCACAAGCAGGUACCACAAAGCCACUAGU
GCACAGGGACUCAGAAAAGACGGCAGGAGCCCAAGGAAAACUCCAAUUUGAGUACAGCCCUGCCUUGUUUCCCCCAGAGAGUCCCUGAGCAAGGAGACCUCCACCC
CACACACACCAUUUCAGAACAACCAGGUUCCAGACUCCCAUGAGGAGCAUCUCCCACUGCAGAGCCUUGGCCAGCCGCGCCCGGACUCCUCAGAGCUGGCGCAAAC
UCCGUCCUCCAAAACUCGGCUCUGGGAGGCCUAAGUGACUCCGAAGCCGGCGGCAGCCGCGGCAGCGGCCGUGGUGGUGGAAGAGCUCUUUUCCCCGACAGUGCCA
CUGAUCGCUCUUCACUGGAGCUGGAAACAGCCUUCGCGGAAAGGACCGGAGCAUGCGUUAGAAGCAGAGGGAGCUUGGUGAAGGGCUCGGCUGGAAGGAGGAAACG
CCUUCUCGCAGUGCGCGGCCAGCCCGCGGGGGACACCGGCUUGCUGGACUGCAGGGGCCCGUGCCACCCAGGAAGUGACCUGCGGGUCACUCAGCCGGGGCGCUGG
GCGAGCGCGGGACGGCCCGGAGAAUUCCGUGCGGCUGCGACGGGAAAAGGACGAGGGGUCUCUGUACCCGACGCUGCCACUGGCCCAAAGGAAUUUUACCCGCGAG
CGCCCACCCCACCCUAGCUUGAUGCUUACGCCCGCAACAAAACAGGAAACCAGGACUGGGCAGUGCAUUCUUUAAGUCAACAAAUACACUGAAGACUUCGAGCGUU
UGAAGGAAGGAGGGGGUUUGCACGUAAGCCUGGCCCCGCCGGGCUCGGCUUUCUCGCUGAGAAAGCGGCGCAGGCAGCCAGGCGGCCUGGGCCCGCGGGGGUCCAU
CUCGCCCUAGACUCCUAAGAACUCCCACGGCCCUGUUCCCAGCUGCGAAUUCUUAAUGCACAACGCGACGGAGGGAAGGAAAUUCACCAGCGCAGCGACGAGGAAG
GGGAACUCAGGACCCCUUCAAGUACACACUGAGGUGUGAUCAGAGUUUUAUGGGCACUUUAUAUGCUGUAAUCAUAACGAUGUGUGUGCCUUGAUAUGCACGCAUA
UUCACGCAUCAAACGUGCAUACACACACAGAGUGAAUGUGCGCAUCCAAUGUCAUGUGGGUGAAAUACAAGCAUCAUACCCAGCCCUACGAAAAAAAAAAUUCACCC
UGUCGGACCAGGCUGGUGACAUACUUCGCUGGCGCAUCUCCUUACUCACUCUUACUUUUCCGACCCCUCACCAUUCCCUCUCCUGUGGCUUGGUAAAUACACCUGC
CCUCCGUGGAAGGUGAGUCCUGGACUGGCGUUGCCAGGUUCGCAUGUCCUCCCCAGAACCUCCGUCUGGCUCCAGGGACUCUCACUGAGCGGGUCUAGAGCACCCA
GCACUUUUCAAGGAACAGCCGCGGUUCCUUUGUCCCGCGGCUCCAGCCCCGUUCGGCCCAGCUCUCAGGGAAACGAAGCGCUCAGUAAGAACUUUUGAUAUUAGUU
UGUAUGGGUAUUUACACUCUGGUGAGGGGAGCUGAGUACGGAAGUUCCAUUAAUCAUACUCCAACCUUGGGUUUAGAUAUUCAGUUUAUGGGUUGGGAGAGGGAGU
UUGCCGGAAAGAAAGCAUCAAGGUUGGCCGCUGACUCCAGAGAAAUGAAAAGGGAGCAAGGUCGUUUUCUGUUUCUGGAAAUCAAGAAUUAGGAAUGGGCAACUAC
AGGUGCUAACCAACAGACCACUUUUUUGUUUUUUGGUAGCCCUUUGGCAGGGAUAGUUUUUCCACCUUUGCCCGAUACAAUUUAAAAAAAAAAAUCCUUUUAUUAU
GGAAUUUGUCAAACACACACACAAGCAUAACAAACCCCUAGGUACCCAUCUCCAAGUUUUGACCCCUAUUAUAAUUUCAUCUUCAGUGUUUUAUUAUCCACUUCCU
CUCUCUCUAUCUUUAGUAUUUUAAAGUAAAUCCCAGAUAGCAUCACAUCAUUUCACCCCCACCAUAGGAUUUCAAAGAUCUGUUAUAUUUCAAGAUUGAGUAAAAG
GGCUUGAAAUUGGGUUAUUGCAAUGAAACUCUAGAAAAAGCUUGAGGGUUCACCCAGGAGUAAGCUGGACAAAAAAGGGGUUUGAGGGGUGGACCCAUCUUGCCUA
AAAAUCUUGUCUCAUCUUUCUAAAAAUUACAUAUGAAAGAGGAAGAUUUAUGUUACUUUUUUAUAUGAGAGAAUCGUCCUUUAAUAGAAAAUUUCUAUUGCUGCAU
CAGAAUUAUGGAGGAACACAAAAAACAUACCUCAGUCCUUAGUGUGUCCUAAAUUAACACAUAUUCACUUAUUAGUGGGUAAAUGACUAUAUUUCAUUUCAGCACA
ACUUCUCCCCUGGUAGAAACACAAAAGAAAUUUCUAAUGAUUAAACUAGGAAAGUUUGCACUGAAUUGAUGGCUUAUCAGAGCAACCGCAGUUUUCAGGAAGAAAU
UCAAUGCCAUGCGUUGAAAAUAUCCCCCUAGCAAUAAGGGAUUAUUUUUAAAAAAGAAUGAAUAAAGAUGUUCUGGUUUCUUUUGUUUUAAUCUGGUAGUCUCAUU
UACAACGAGCAUGAUUCUCCCUGUCGAACUCUGAAAGUGACUUAACUGAAAGGCUUGGCAACUUCAGAAAGCAAAAAGGUAAAAACAGAAAAUAGCACACGGUUGA
AUUUGACAACUUUUACACUACCCGGCUGCUUAAUAAAUUCUAACCCCACUU

Supplementary Table 5: Sequences of the internal insertion constructs in **Figures 3a, b**, **e**, **f** and **4**. All constructs had the constant backbone:

GAUCUAGAUACGACUCACUAUGUUUAAGAGCUAUGCUGCGAAUACGAG**XXXX**CUCGUAUUCGCAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU

...wherein **XXXX** is replaced with the sequence below.

Construct	Insert Sequence
1xPP7 SL	GGAGCAGACGAUAUGGCGUCGCTCC
S1 Aptamer	CCGACCAGAAUCAUGCAAGUGCGUAAGAUAGUCGCGGGCCGGC
3xMS2 SL	CGUACACCAUCAGGGUACGUCUCAGACACCAUCAGGGUCUGUCUGGUACAGCAUCAGCGUACC
3xCsy4 SL	UCUUACUGCUGUAUAAGCAGCUCUUACUGCCGUGUAGGCAGCUUCUACUUCUGUAUAAGAAGCUUUC
3xPP7 SL (“INT”)	GGAGCAGACGAUAUGGCGUCGCUCCUCUCCACGAGAGCAUAUGGGCUCCGUGGUCUCCAGCAGACGAUAUGGCGUCGCUGG
GFP Aptamer	GCUUCUGGACUGCGAUGGGAGCACGAAACGUCGUGGCGCAAUUGGGUGGGGAAAGUCCUUAAAAGAGGGCCACCACAGAAGCC
Spinach2	GAUGUAACUGAAUGAAAUGGUGAAGGACGGGUCCAGUAGGCUGCUUCGGCAGCCUACUUGUUGAGUAGAGUGUGAGCUCCGUAA–CUAGUUACAUC
3xK–T	UCUGCUCCCGUGAUGGCGAAAGCCUGAGGAGCUCUCUGGCCGUGAUGGCGAAAGCCUGAGCCAGUCUCUGCCCGUGAUGGCGAA–AGCCUGAGGCAGUCU
Bunch of baby Spinach (BoBS)	AAGGACGGGUCCGGACGCAAGGACGGGUCCGACCGAAAGGACGGGUCCAAUGGUGGAAACACCAUUGUUGAGUAGAGUGUGAGU–CGGUCGUUGAGUAGAGUGUGAGGCGUCCGUUGAGUAGAGUGUGAG
5xPP7 SL	GGAGCAGACGAUAUGGCGUCGCUCCUCUCCACGAGAGCAUAUGGGCUCCGUGGUCUGCAGCAGACGAUAUGGCGUCGCUGCUCU–CGUAGAUGCCAUAUGGGGCACUACGUCUCCAGCAGACGAUAUGGCGUCGCUGG
P4–P6[3xPP7 SL]	UCUGGAAUUGCGGGAAAGGGGUCAACAGCCGUUCAGUACCAAGUCUCAGGGGAAACUUUGAGAUGGCCUUGCAAAGGGUAUGGU–AAUAAGCUGACGGACAUGGUCCUAACACGCAGCCAAGUCCUAAGUCAACAGUCUGGAGCAGACGAUAUGGCGUCGCUCCUCUCCA–CGAGAGCAUAUGGGCUCCGUGGUCUCCAGCAGACGAUAUGGCGUCGCUGGUCUCUGUUGAUAUGGAUGCAGUUCAUCU

PP7 SL: cognate stem-loop recognized by the PP7 phage coat protein[7]. S1: an artificial streptavidin aptamer[53]. MS2 SL: cognate stem-loop recognized by the MS2 phage coat protein[5]. Csy4 SL: cognate stem-loop recognized by the *P. aeruginosa* Csy4 protein[54]. GFP aptamer: an artificial GFP/YFP/CFP aptamer[55]. Spinach2: an artificial, small-molecule-binding fluorescent aptamer[51]. K–T: a cognate “kink-turn” motif recognized by the *A. fulgidus* L7Ae protein[56]. BoBS: “Bunch of Baby Spinach” (**Supplementary Fig. 11**). P4–P6: the thermostable ΔC209 mutant of the *T. thermophila* group I intron P4–P6 domain[40], as in (**Supplementary Fig. 4a**–**b**).

Supplementary Table 6: Amplification primers used to generate INT–N_{25}Pool sequencing libraries

Primer	Sequence
RT Primer	CGACTCGGTGCCACTTTT
Forward Library[a,b]	CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT–TCAAGTTGATAACGGACTAGC
Reverse Library[a]	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCTATG–CTGCGAATACGAG

[a]Illumina™ adapters are highlighted in blue; indexes in yellow
[b]X's correspond to Illumina™ indexes 1–7.

Supplementary References

1. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature communications* **3**, 968 (2012).

2. Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature biotechnology* **29**, 149-153 (2011).

3. Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512 (2009).

4. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).

5. LeCuyer, K.A., Behlen, L.S. & Uhlenbeck, O.C. Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry* **34**, 10600-10606 (1995).

6. Cate, J.H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685 (1996).

7. Chao, J.A., Patskovsky, Y., Almo, S.C. & Singer, R.H. Structural basis for the coevolution of a viral RNA-protein complex. *Nature structural & molecular biology* **15**, 103-105 (2008).

8. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

9. Douglass, E.F., Jr., Miller, C.J., Sparer, G., Shapiro, H. & Spiegel, D.A. A comprehensive mathematical model for three-body binding equilibria. *Journal of the American Chemical Society* **135**, 6092-6099 (2013).

10. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**, 145-166 (2012).

11. Ulitsky, I. & Bartel, D.P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26-46 (2013).

12. Meckler, J.F. et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic acids research* **41**, 4118-4128 (2013).

13. Wu, B., Chen, J. & Singer, R.H. Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Scientific reports* **4**, 3615 (2014).

14. Delebecque, C.J., Lindner, A.B., Silver, P.A. & Aldaye, F.A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science* **333**, 470-474 (2011).

15. Geary, C., Chworos, A. & Jaeger, L. Promoting RNA helical stacking via A-minor junctions. *Nucleic acids research* **39**, 1066-1080 (2011).

16. Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of integrative plant biology* **56**, 343-349 (2014).

17. Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. & Lu, T.K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Molecular cell* **54**, 698-710 (2014).

18. Cuello, P., Boyd, D.C., Dye, M.J., Proudfoot, N.J. & Murphy, S. Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *The EMBO journal* **18**, 2867-2877 (1999).

19. Bartlett, J.S. et al. Efficient expression of protein coding genes from the murine U1 small nuclear RNA promoters. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 8852-8857 (1996).

20. de Vegvar, H.E., Lund, E. & Dahlberg, J.E. 3' end formation of U1 snRNA precursors is coupled to transcription from snRNA promoters. *Cell* **47**, 259-266 (1986).

21. Hernandez, N. & Weiner, A.M. Formation of the 3' end of U1 snRNA requires compatible snRNA promoter elements. *Cell* **47**, 249-258 (1986).

22. Fuke, H. & Ohno, M. Role of poly (A) tail as an identity element for mRNA nuclear export. *Nucleic acids research* **36**, 1037-1049 (2008).

23. Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A. & Steitz, J.A. Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* **330**, 1244-1247 (2010).

24. Conrad, N.K., Mili, S., Marshall, E.L., Shu, M.D. & Steitz, J.A. Identification of a rapid mammalian deadenylation-dependent decay pathway and its inhibition by a viral RNA element. *Molecular cell* **24**, 943-953 (2006).

25. Conrad, N.K. & Steitz, J.A. A Kaposi's sarcoma virus RNA element that increases the nuclear abundance of intronless transcripts. *The EMBO journal* **24**, 1831-1841 (2005).

26. Brown, J.A. et al. Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nature structural & molecular biology* **21**, 633-640 (2014).

27. Wilusz, J.E. et al. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes & development* **26**, 2392-2407 (2012).

28. Wilusz, J.E., Freier, S.M. & Spector, D.L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-932 (2008).

29. Battle, D.J. et al. The SMN complex: an assembly machine for RNPs. *Cold Spring Harbor symposia on quantitative biology* **71**, 313-320 (2006).

30. Huber, J. et al. Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *The EMBO journal* **17**, 4114-4126 (1998).

31. Briner, Alexandra E. et al. Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Molecular cell* **56**, 333-339 (2014).

32. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2984-2989 (2015).

33. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).

34. Zalatan, J.G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350 (2015).

35. Tsai, S.Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology* **32**, 569-576 (2014).

36. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).

37. Maeder, M.L. et al. CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977-979 (2013).

38. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology* **31**, 833-838 (2013).

39. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**, 973-976 (2013).

40. Juneau, K., Podell, E., Harrington, D.J. & Cech, T.R. Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions. *Structure* **9**, 221-231 (2001).

41. Leontis, N.B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *Rna* **7**, 499-512 (2001).

42. Ryan, O.W. et al. Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife* **3** (2014).

43. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* **31**, 46-53 (2013).

44. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832 (2013).

45. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature biotechnology* (2014).

46. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology* (2014).

47. Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature biotechnology* **33**, 187-197 (2015).

48. Strack, R.L., Disney, M.D. & Jaffrey, S.R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nature methods* **10**, 1219-1224 (2013).

49. Huang, H. et al. A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore. *Nature chemical biology* **10**, 686-691 (2014).

50. Warner, K.D. et al. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature structural & molecular biology* **21**, 658-663 (2014).

51. Song, W., Strack, R.L., Svensen, N. & Jaffrey, S.R. Plug-and-play fluorophores extend the spectral properties of Spinach. *Journal of the American Chemical Society* **136**, 1198-1201 (2014).

52. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA mimics of green fluorescent protein. *Science* **333**, 642-646 (2011).

53. Walker, S.C., Good, P.D., Gipson, T.A. & Engelke, D.R. The dual use of RNA aptamer sequences for affinity purification and localization studies of RNAs and RNA-protein complexes. *Methods in molecular biology* **714**, 423-444 (2011).

54. Sternberg, S.H., Haurwitz, R.E. & Doudna, J.A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *Rna* **18**, 661-672 (2012).

55. Tome, J.M. et al. Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nature methods* **11**, 683-688 (2014).

56. Saito, H. et al. Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nature chemical biology* **6**, 71-78 (2010).

57. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nature methods* (2015).

58. Zhao, S. & Fernald, R.D. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of computational biology : a journal of computational molecular cell biology* **12**, 1047-1064 (2005).