Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display

David M Shechner[1–3], Ezgi Hacisuleyman[1–3], Scott T Younger[1–3] & John L Rinn[1–4]

Noncoding RNAs play diverse roles throughout biology and exhibit broad functional capacity. To investigate and harness these capabilities, we developed clustered regularly interspaced short palindromic repeats (CRISPR)-Display (CRISP-Disp), a targeted localization method that uses Cas9 to deploy large RNA cargos to DNA loci. We demonstrate that functional RNA domains up to at least 4.8 kb long can be inserted in CRISPR guide RNA at multiple points, allowing the construction of Cas9 complexes with protein-binding cassettes, artificial aptamers, pools of random sequences and natural long noncoding RNAs. A unique feature of CRISP-Disp is the multiplexing of distinct functions at multiple targets, limited only by the availability of functional RNA motifs. We anticipate the use of CRISP-Disp for ectopically targeting functional RNAs and ribonucleoprotein (RNP) complexes to genomic loci.

© 2015 Nature America, Inc. All rights reserved.

Noncoding RNAs (ncRNAs) are central to diverse and fundamental processes in all kingdoms of life[1]. In eukaryotes, many ncRNAs are thought to help initiate or maintain regulatory processes within the nucleus[1–3]. However, the mechanistic dissection of these putative regulators *in vivo* is technically limited. For example, knock-in and knockout strategies[4,5] lack the throughput required for high-resolution structure-function analysis, and cannot easily separate roles performed by an RNA transcript from those performed by a functional DNA element or by a cryptically encoded peptide[5,6].

A method for relocating ncRNA transcripts to an ectopic site would thus be an enabling tool for the study of ncRNA function. Such a method could also provide a powerful engine for synthetic biology. Many RNA domains—both natural and artificial—have been adapted as components in synthetic regulators, reporters and scaffolds[7–13]. Targeting RNA devices to specific DNA loci would allow a wide range of novel synthetic biological methods.

We reasoned that ncRNA targeting could be achieved using an artificial protein 'conduit' that is programmed to bind a ncRNA and localize it to a specific DNA locus (**Supplementary Fig. 1a**). An excellent candidate conduit is the *Streptococcus pyogenes* Cas9 nuclease (*Sp.* Cas9), a high-affinity programmable DNA-binding protein isolated from a type II CRISPR–associated system[14–16]. In a multitude of CRISPR-based applications[17–22], Cas9 is targeted to DNA by a so-called short guide RNA (sgRNA), which consists of the two natural Cas9 cofactors (crRNA and tracrRNA)[14,15] fused together via an engineered loop (**Supplementary Fig. 1b**). Despite recent work on Cas9 RNA recognition[23–26], it remained unclear whether and where large structured RNA domains could be added to CRISPR complexes while maintaining RNA-directed localization.

Here, we demonstrate that a nuclease-deficient *Sp.* Cas9 mutant (dCas9)[17,27–29] can deploy a large RNA cargo to targeted DNA loci by incorporating the cargo into the sgRNA. We term this strategy for displaying RNA domains on dCas9 CRISPR-Display, or 'CRISP-Disp'. With the appropriate expression system and insertion point, CRISP-Disp does not appear to be inherently limited by the size or sequence composition of its RNA cargo. This allows dCas9 complexes to be functionalized with structured RNA domains, natural long noncoding RNAs (lncRNAs) several kilobases in length, artificial RNA modules and pools of random sequences. Furthermore, these RNA-based functions can be multiplexed using a shared pool of dCas9. This work provides initial insights into the utility of CRISP-Disp for the study of natural ncRNAs and the construction of novel RNA-based devices.

RESULTS

Adapting CRISPR-Cas9 as an RNA display device

To assess potential *in vivo* localization methods for ncRNA, we implemented a highly sensitive reporter system in HEK293FT cells. In this system, the efficacy of an ectopically targeted transcriptional regulator (either a protein or an RNP) is reported by luciferase and fluorescent protein expression. RNA-targeting strategies were surveyed using both transiently transfected and stably integrated reporters (**Supplementary Fig. 1c–f** and **Supplementary Table 1**)[27,30].

Our initial strategy used transcription activator–like effectors (TALEs) as conduits for targeting RNA to the reporters[30]. However, we were unable to coax TALEs into recruiting ncRNAs to a chromatin-integrated locus (**Supplementary Figs. 2** and **3**). We ascribed this problem to the separable DNA- and RNA-binding activities of our conduit, which could

[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA. [3]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA. [4]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. Correspondence should be addressed to J.L.R. (johnrinn@fas.harvard.edu).

RECEIVED 3 NOVEMBER 2014; ACCEPTED 8 MAY 2015; PUBLISHED ONLINE 1 JUNE 2015; DOI:10.1038/NMETH.3433

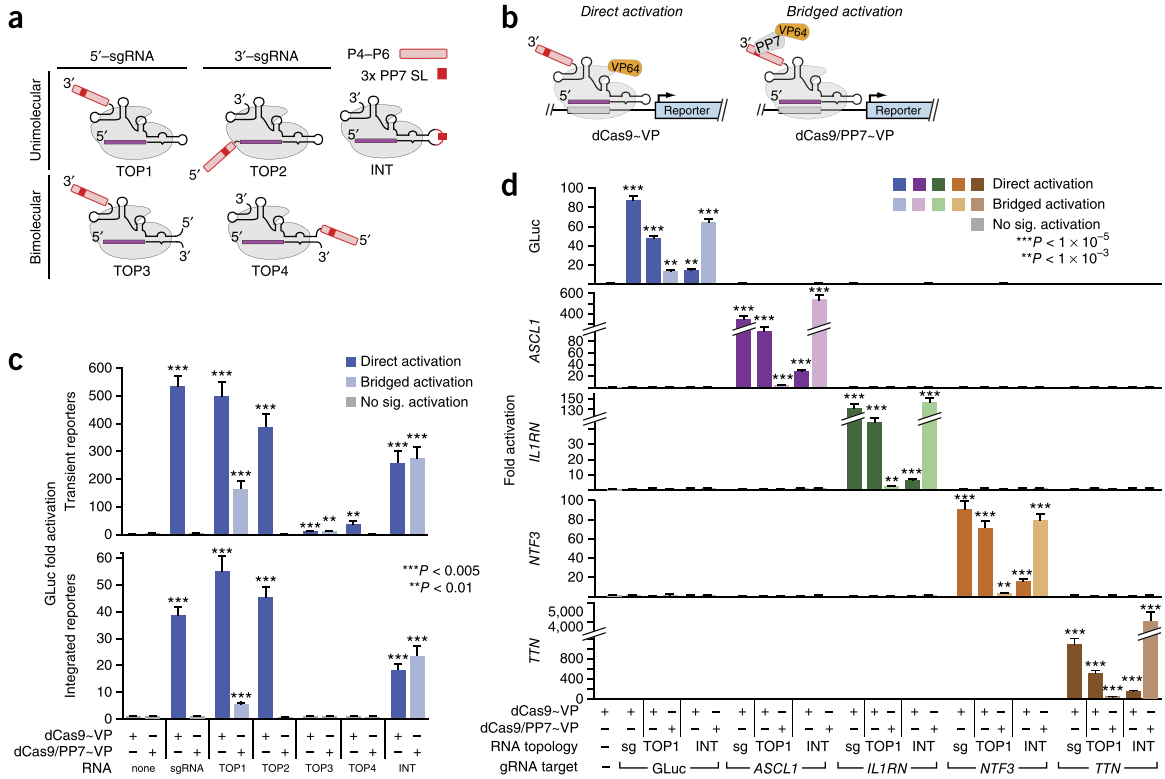


Figure 1 | Large, structured RNA domains can be functionally appended onto the sgRNA scaffold at multiple points. (**a**) Design of "TOP" topology constructs. Accessory RNA domains are detailed in (**Supplementary Fig. 4a**,**b**); expression constructs in (**Supplementary Fig. 1c**). (**b**) Schematics summarizing direct activation (left) and bridged activation (right) assays. (**c**) Luciferase reporter assays of the five topology constructs. (**d**) Targeting minimal ("sg") and expanded ("TOP1" and "INT") sgRNAs to endogenous loci. *ASCL1*, *IL1RN*, *NTF3* and *TTN* were each targeted using mixed pools of four guide RNAs[28,29,33] (**Supplementary Table 2**). GLuc activation was measured by luciferase assays; endogenous gene activation was measured using qRT-PCR (**Supplementary Table 3**). Values in **c** and **d** are means ± s.d.; ($n = 3$, biological replicates) for luciferase assays; ($n = 4$, technical replicates) for qRT-PCR. Student's one-tailed *t*-test relative to cells expressing dCas9~VP alone.

be independently saturated without forming DNA–TALE–RNA ternary complexes (**Supplementary Note 1**).

To circumvent this issue, we turned to the *Sp.* CRISPR–dCas9 system, which intrinsically couples its DNA- and RNA-binding activities (**Supplementary Fig. 1b–f**)[15,16]. However, it was unclear a priori where insertions within the dCas9-bound RNAs would be tolerated and how large they could be. To examine this, we devised five model cofactor RNAs (TOP1–TOP4 and INT) in which structured, 81- to 250-nt 'accessory domains' were either inserted within the sgRNA or tracrRNA or appended to their termini[15]. Each accessory domain carried a cassette of stem-loops recognized by the PP7 phage coat protein[31] (**Fig. 1a** and **Supplementary Fig. 4**). The largest of these constructs (357 nt) adds three-fold more sequence than the longest modified sgRNA previously reported[20,25,26].

We subjected these RNA chimeras to two variations of our CRISPR transcription activator assay (**Fig. 1b**). In direct activation assays, they were coexpressed with dCas9 fused to the VP64 transcriptional activator (dCas~VP). In this format, reporter gene activation indicates that the sgRNA variant competently binds and targets dCas9. In bridged activation assays, RNA chimeras were coexpressed with dCas9 and a PP7 fusion to VP64 (PP7~VP). Bridged activation should only occur if the accessory domain remains intact in the mature dCas9 complexes.

Using transient reporters, we observed direct activation with all five topologies (**Fig. 1c**, top). However, whereas the activities of TOP1, TOP2 and INT were less than two-fold lower than that of the minimal sgRNA, TOP3 and TOP4 were ~9–30-fold less proficient. Critically, bridged activation was only observed with TOP1, TOP3 and INT, indicating that the other constructs do not retain functional accessory domains in mature dCas9 complexes. These results were corroborated by cell sorting (**Supplementary Fig. 4e**) and showed a similar pattern using integrated reporters (**Fig. 1c**, bottom). We attributed the lack of TOP2-bridged activation to partial degradation of that construct's accessory domain, as indicated using RNA immunoprecipitation (RIP) and qRT-PCR (**Supplementary Fig. 4f**)[20,25,32].

Targeting long RNAs to endogenous genomic loci

To illustrate the applicability of CRISP-Disp to endogenous loci, we generated TOP1 and INT sgRNAs targeting the human *ASCL1*, *IL1RN*, *NTF3* and *TTN* promoters[28,29,33] (**Supplementary Table 2**) and surveyed direct and bridged activation of these genes by qRT-PCR. Activation of each locus paralleled the results obtained using our GLuc reporter, demonstrating CRISP-Disp enables deployment of large RNA domains to genomic loci (**Fig. 1d**).

To confirm that dCas9 targeting fidelity is not perturbed by the addition of the accessory RNA domains, we performed RNA sequencing from reporter cells expressing dCas9~VP and GLuc-targeting sgRNA, TOP1 or INT constructs. As predicted, changes in global gene expression induced by each RNA construct—a proxy for off-target dCas9 activity—were essentially indistinguishable,

© 2015 Nature America, Inc. All rights reserved.

indicating that TOP1- and INT-like accessory domains do not substantially alter dCas9 fidelity (**Supplementary Fig. 5**).

CRISP-Disp with artificial lncRNA scaffolds

We next sought to engineer CRISP-Disp for use with even longer ncRNAs, which required replacing the conventional RNA polymerase III (Pol III) promoter used above with an RNA polymerase II (Pol II) promoter and terminator. We therefore surveyed a variety of standard and noncanonical Pol II expression systems for the ability to generate nuclear-localized CRISP-Disp RNAs *de novo* (**Supplementary Figs. 6** and **7** and **Supplementary Note 2**). Although Pol II transcripts were generally less effective than their Pol III–driven counterparts[19], expression from several different backbones partially restored their activity (**Supplementary Fig. 6b**). Surprisingly, one backbone (CMV/3′-Box) also enabled bridged activation with TOP2, which failed under Pol III expression (**Supplementary Fig. 6c–e**).

Next, we attempted to build dCas9 complexes with model RNAs approaching the size of natural lncRNAs. We expanded our CMV/3′-Box TOP1 and TOP2 constructs with a second P4–P6 domain, bearing stem-loops recognized by the MS2 phage coat protein[31], to generate a series of 'artificial lncRNA' scaffolds with ~650-nt accessory domains (Double TOP0–2; **Fig. 2a** and **Supplementary Fig. 4a–d**). Although all three of these constructs induced measurable direct activation, Double TOP1 and Double TOP2 were more proficient, with activities nearly rivaling those of their single-domain counterparts (**Fig. 2b**). Moreover, all three constructs exhibited notable bridged activation. In transient reporter assays, activity monotonically increased upon coexpression with PP7~VP, MS2~VP or both activator proteins, indicating that each RNA construct retained both P4–P6 domains in mature dCas9 complexes (**Fig. 2b**). A qualitatively similar trend was observed with Double TOP1 using integrated reporters, despite the assay's limited dynamic range (**Supplementary Fig. 8**). We also confirmed the integrity of the Double TOP1 and Double TOP2 accessory domains by dCas9 RIP; for each construct, we obtained essentially stoichiometric yields of the sgRNA core and double P4–P6 accessory domains (**Fig. 2c**).

CRISP-Disp with natural lncRNA domains

Reconstitution of natural lncRNA activity at an ectopic site is unattainable by existing methods[5,6] and would be a powerful use of CRISP-Disp. We first established that natural lncRNAs could be incorporated into CRISP-Disp complexes. We generated Pol II–driven TOP1- and INT-like constructs appended with human lncRNA domains: the repressive NoRC-binding pRNA stem-loop[34] and *Xist* A-repeat ('RepA') domains[4], three enhancer-transcribed RNAs (eRNAs[35]) and the 4,799-nt transcriptional activator *HOTTIP*[36] (**Supplementary Table 4**).

Each construct induced substantial direct activation, indicating functional dCas9-targeting complexes (**Fig. 2d**). This was further supported by RNA immunoprecipitation qRT-PCR (RIP-qPCR). Although overall dCas9 complexation appeared to decline monotonically with increasing sgRNA-lncRNA length, in all cases nearly quantitative yields of intact lncRNA domains were recovered relative to the corresponding sgRNA core (**Fig. 2e**). We propose that the observed minor variation in construct integrity may be contingent on several factors, including length and RNA structure[25].

We next tested if the lncRNA components of CRISP-Disp complexes could themselves regulate our reporter. Encouragingly, most of the lncRNA constructs repressed or activated GLuc expression as would be predicted from their previously characterized activities[4,34–36]: pRNA and RepA diminished GLuc expression, while two eRNAs (*TRERNA1* and ncRNA-a3) and *HOTTIP* induced moderate activation (**Fig. 2f**). Although these effects were quite modest, they may have been limited by technical aspects of our assay (for example, the promoter used or its placement relative to the targeting cassette (**Supplementary Fig. 1c**)) or by the absence of particular factors in HEK293FT cells. Regardless, these initial results demonstrated the plausibility of larger-scale lncRNA functional studies with CRISP-Disp.

CRISPR-Display of a diverse array of RNA species

Given the remarkable proficiency of U6-driven INT (**Fig. 1a**,**c**), we reasoned that the sgRNA-engineered loop might provide a universal insertion point for exogenous RNA domains. To explore this possibility, we first examined the influence of internal insert size on targeting and activity. INT-like constructs bearing 25–137 nt cassettes of PP7 stem-loops each induced robust activation in all assay formats (**Fig. 3a**), as did a construct bearing an ~250-nt domain equivalent to those of TOP1–4 (**Fig. 3b** and **Supplementary Table 5**). Hence, dCas9 can easily accommodate even large, structurally discontinuous inserts within the sgRNA core.

To examine the effect of insert sequence, we synthesized a pool of ~1.2×10^6 unique sgRNA variants displaying internal cassettes of 25 random nucleotides (INT-N_{25}Pool, **Supplementary Fig. 9**). In aggregate, the activity of this pool rivaled that of the minimal sgRNA (**Fig. 3c**), implying that many of the variants formed productive CRISP-Disp complexes. To confirm this, we immunoprecipitated dCas9–INT-N_{25}Pool complexes and analyzed the copurified sgRNA sequences by deep sequencing (RIP-Seq, **Fig. 3d** and **Supplementary Fig. 10**). Fewer than 0.01% and 0.02% of the sequence variants were significantly enriched or de-enriched, respectively (Benjamini-Hochberg–corrected $P < 0.05$); and motif analysis of these variants revealed no apparent sequence constraints influencing sgRNA-dCas9 complexation. Although this pool represents only a small sampling of the total 25-mer sequence space, these results suggested that CRISP-Disp is not intrinsically limited by the sequence of an internal insert.

To explore the potential utility of INT-like inserts, we next generated a series of INT-like constructs displaying functional RNA domains comprising natural protein-binding motifs[31,37,38] or artificial aptamers that bind proteins[7,8] or small molecules[13] (**Supplementary Fig. 11** and **Supplementary Table 5**). All constructs were viable CRISP-Disp substrates (**Fig. 3e**). Variation in construct efficacy did not appear to result from limiting RNA expression levels (**Supplementary Fig. 12**) implying that design of a high-efficiency CRISP-Disp construct for a given motif may require some structural or sequence optimization.

Next, we used INT-like constructs to deploy structured RNA domains to endogenous loci. Constructs bearing the S1 streptavidin aptamer, a potentially useful RNA device[7], and P4–P6[3xPP7], the largest INT-like construct tested, each activated *ASCL1*, *IL1RN*, *NTF3* and *TTN*[28,29,33] as predicted (**Fig. 3f**). Furthermore, INT-P4–P6[3xPP7] did not appear to perturb dCas9 targeting

© 2015 Nature America, Inc. All rights reserved.

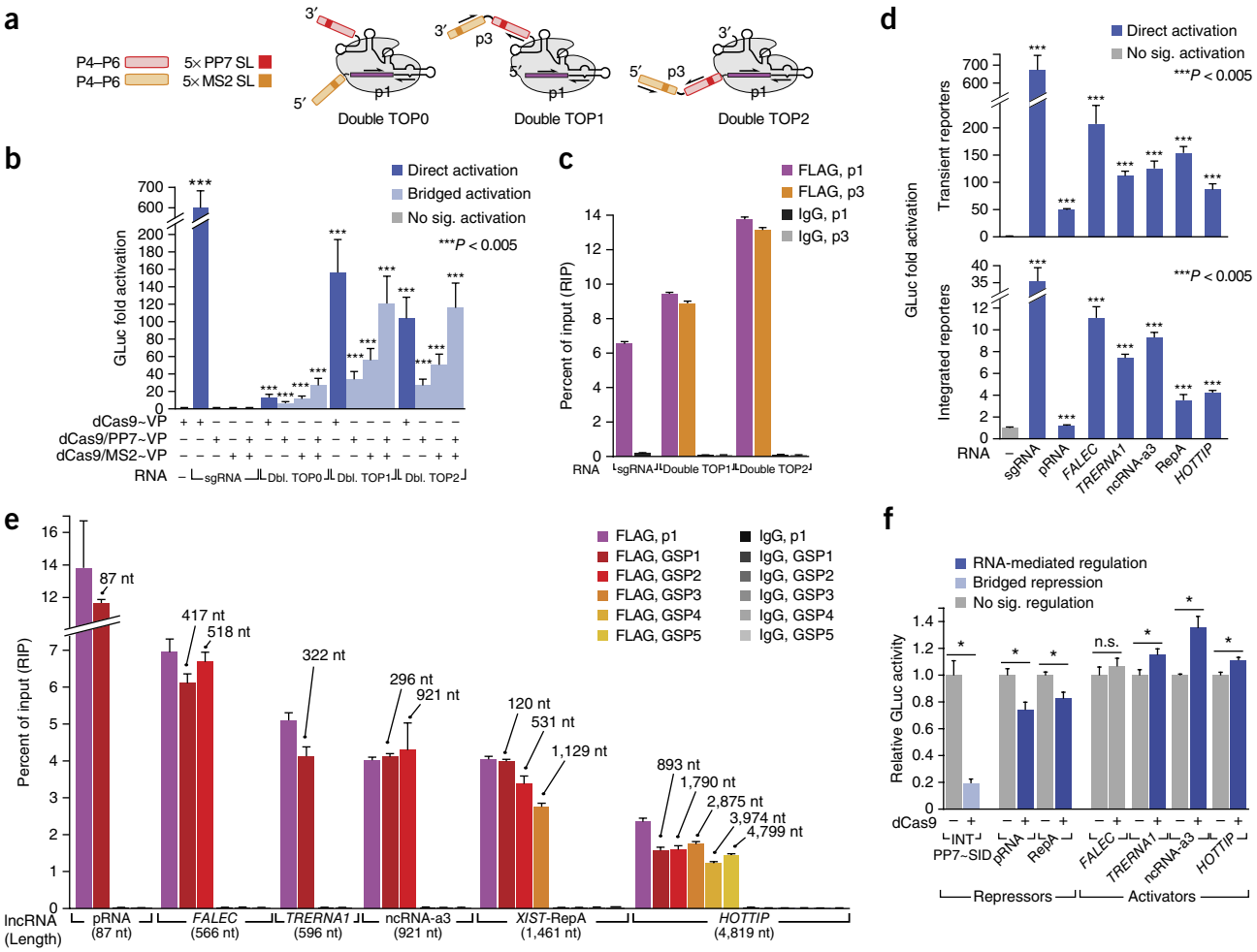


Figure 2 | CRISPR-Display supports artificial and natural lncRNAs. (**a**) Design of "Double TOP" artificial lncRNA constructs. Accessory domains are detailed in (**Supplementary Fig. 4a–d**). The sgRNA core "p1" and domain-spanning "p3" qPCR primer pairs are indicated (**Supplementary Table 3**). (**b**) Direct and bridged activation assays using Double TOP ("Dbl TOP") constructs, expressed from the CMV/3′ Box backbone (**Supplementary Fig. 6a**). Transient reporter assays are shown. (**c**) RIP–qRT-PCR of dCas9–Double TOP1 and dCas9–Double TOP2. (**d**) sgRNAs modified with a battery of natural lncRNA domains (**Supplementary Table 4**) form functional complexes with dCas9~VP. Direct activation assays are shown. The minimal "pRNA" stem-loop[34] was displayed internally; all other domains were appended on the sgRNA 3′ terminus. RNAs were expressed using the CMV/MASC system (**Supplementary Fig. 6a**). (**e**) LncRNA accessory domains remain intact in CRISP-Disp complexes. Immunopurified RNA was analyzed using qPCR primers targeting the sgRNA core (p1) and intervals along the each lncRNA domain (GSP1–GSP5, **Supplementary Table 3**). Above each primer set, the maximum distance from the sgRNA core is indicated. (**f**) Transient reporter assays with CRISP-Disp lncRNA constructs. Values are normalized relative to those of control cells expressing each lncRNA alone. For comparison, bridged repression with U6-driven INT complexed with dCas9 and PP7~SID[39] is shown (left, light blue). Values in **b**–**f** are means ± s.d.; (n = 3, biological replicates) for luciferase assays; (n = 4, technical replicates) for qPCR. Student's one-tailed t-test relative to cells expressing dCas9~VP alone (**b**,**d**) or lncRNA alone (**f**); *P < 0.05; n.s., not significant.

fidelity (**Supplementary Fig. 5**). Our results suggested that INT-like constructs could display large RNA domains at specific loci genome wide.

CRISP-Disp enables multiple concurrent functions

One theoretical advantage of CRISP-Disp is that it enables different functions to be performed simultaneously at diverse loci in the same cell, using a single toolset. This modularity could be achieved using a set of orthogonal RNA-binding proteins, each fused to a unique functional domain and targeted by sgRNAs bearing its cognate RNA motif[25]. We first established a preliminary set of such RNA-binding proteins (*Archaeoglobus fulgidus* L7Ae, MS2 and PP7)[31,37] for display on dCas9. As predicted, when fused to VP64, each protein was only competent to induce bridged reporter activation when coexpressed with an sgRNA bearing its cognate RNA motif (**Fig. 4a**).

As a first demonstration of CRISP-Disp modularity, we bound dCas9 to multiple genomic targets, but activated only one (**Fig. 4b**). We coexpressed complementary pairs of GLuc- and *NTF3*-targeting INT-like sgRNA variants bearing PP7 and MS2 stem-loops (GLuc[3xPP7] with *NTF3*[3xMS2], or GLuc[3xMS2] with *NTF3*[3xPP7]) in integrated reporter cells. In the presence of dCas9~VP, we observed robust activation of both target genes regardless of the sgRNA pair used (**Fig. 4b**), indicating that dCas9 had bound both loci under all conditions. However, in the presence of dCas9 and PP7~VP, only the gene targeted by sgRNAs bearing PP7 stem-loops was activated; the converse was observed upon coexpression of dCas9 and MS2~VP (**Fig. 4b**).

© 2015 Nature America, Inc. All rights reserved.

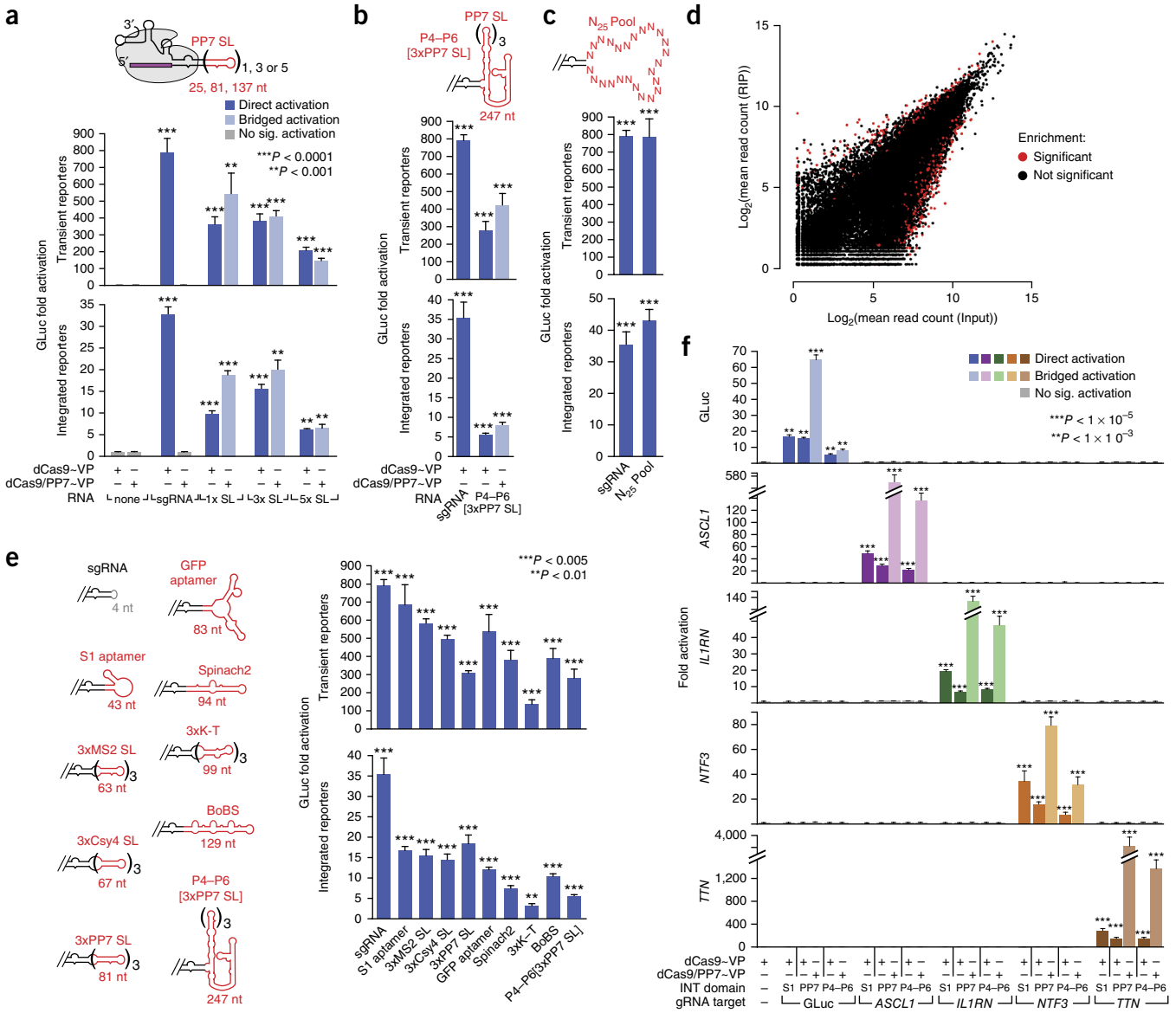


Figure 3 | CRISPR-Display is compatible with structurally diverse RNA domains. (**a**) INT insert size has a modest effect on CRISP-Disp efficacy. "1, 3 or 5" refers to the number of internal PP7 stem-loops. (**b**) Functional INT inserts can be large and structurally discontinuous with the sgRNA core. Accessory domain is detailed in (**Supplementary Fig. 4a**,**b**); $***P < 0.005$. (**c**) Assembly of functional CRISP-Disp INT complexes is independent of the inserted sequence. Direct activation assays with a pool of ~1.2×10^6 unique INT–N_{25} Pool variants (**Supplementary Fig. 9**); $***P < 0.005$. (**d**) dCas9 binds to nearly all expressed *INT*–N_{25}Pool variants. Significance testing: Benjamini-Hochberg corrected $P < 0.05$. See also (**Supplementary Fig. 10**). (**e**) Assembling functional CRISP-Disp complexes bearing an assortment of natural and artificial RNA domains. Left: cartoons depicting INT constructs. Insert lengths are indicated in red, details in **Supplementary Table 5**. BoBS, "Bunch of Baby Spinach," (**Supplementary Fig. 11**). Right: direct activation activities, sorted by construct length. (**f**) Targeting INT-like constructs bearing an RNA device (S1 aptamer, "S1") and a large domain (P4–P6[3xPP7], "P4–P6") to endogenous loci. Constructs were targeted to *ASCL1*, *IL1RN*, *NTF3* and *TTN*, as in **Figure 1d**. Data for the original INT ("PP7") are included for comparison. Values in **a**–**e** are means ± s.d. ($n = 3$, biological replicates) for luciferase assays; ($n = 4$, technical replicates) for qPCR. Student's one-tailed *t*-test relative to cells expressing dCas9~VP alone. RNA constructs were expressed from the human U6 promoter.

Hence, CRISP-Disp enables modular control of gene expression, as has been recently demonstrated with TOP1-like constructs[25].

As a second demonstration of CRISP-Disp modularity, we sought to simultaneously activate one locus and image another (**Fig. 4c**). We implemented a bridged imaging approach, in which targeted loci (in this case, telomeres[18]) are visualized using a ternary complex comprising dCas9, MS2~mCherry and INT-like [3xMS2] constructs (**Supplementary Fig. 13a**). Using bridged telomere imaging in HEK293FT cells, we observed numerous (8–55; average of 26.6, in ~97 mCherry$^+$ cells) fluorescent nuclear foci, which required expression of dCas9 and the proper cognate sgRNA (**Supplementary Fig. 13b**,**c**). Next, to simultaneously activate one locus and image another, we employed our integrated reporter cells and performed bridged activation (targeted by INT[3xPP7]) at the GLuc reporter and bridged imaging at telomeres (**Fig. 4c**). As predicted, each function could be individually or concomitantly controlled by expression with PP7~VP64, MS2~mCherry or both (**Fig. 4c**).

© 2015 Nature America, Inc. All rights reserved.

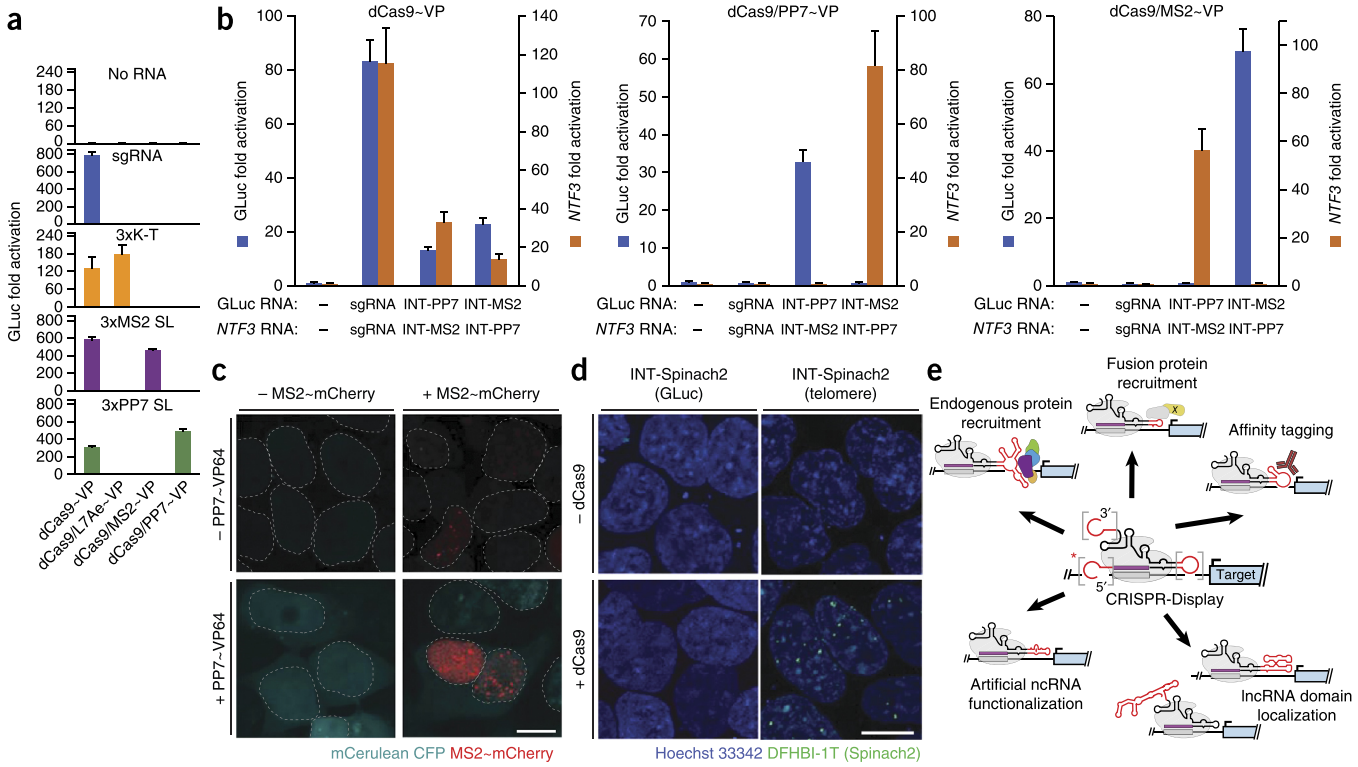


Figure 4 | CRISPR-Display expands the functional repertoire of CRISPR-based methods. (**a**–**c**) CRISP-Disp enables modular, simultaneous control of multiple functions using a shared pool of dCas9. (**a**) Bridged activation assays using an orthogonal set of RNA-binding proteins. K-T: a "kink-turn" motif, recognized by L7Ae (**Supplementary Table 5**)[37]. (**b**) Simultaneous binding and transcriptional activation of distinct genomic loci. Indicated sgRNAs and INT derivatives ("INT-PP7" and "INT-MS2") were coexpressed in direct and bridged activation assays (left, direct activation; middle, bridged activation with PP7~VP; right, bridged activation with MS2~VP). Values in **a**,**b** are means ± s.d. (n = 3) for luciferase assays; (n = 4) for qPCR. (**c**) Simultaneous activation and imaging of distinct genomic loci, using the indicated constructs, in integrated GLuc reporter cells. GLuc- and telomere-targeting INT derivatives were coexpressed in all experiments. Dotted lines denote nuclear membranes. Scale bar, 15 µm. (**d**) CRISPR-Display of novel RNA-based functions. Top: schematic of the experimental design. Middle and bottom: aptamer-based imaging of DNA loci using the artificial fluorescent aptamer "Spinach2" (ref. 13). Scale bar, 15 µm. (**e**) Novel, multiplexable functions made possible by CRISPR-Display. Accessory domains on the sgRNA 5′ end (red asterisk) are not tolerated by most expression systems (**Supplementary Fig. 6**). Confocal fluorescence images (**c**,**d**) are at 63× magnification. Fields are representative of 6–10 images, each encompassing ~20–30 cells.

CRISP-Disp with autonomous RNA domains

Another potential advantage of CRISP-Disp is that it allows autonomously functional RNA domains, such as ribozymes, aptamers and regulatory devices[7–13], to be targeted to individual loci. As a preliminary illustration, we targeted the Spinach2 aptamer[13] to telomeres[18]. Spinach2 fluoresces upon binding the cell-permeable dye DFHBI-1T[13]. When we coexpressed a Spinach2-appended telomere-targeting sgRNA with dCas9 and treated cells with DFHBI-1T, we observed numerous (10–20; average of 12, in ~20% of cells) nuclear fluorescent foci (**Fig. 4d** and **Supplementary Fig. 14**). Critically, no fluorescent foci were observed in the absence of dCas9 or when Spinach2 was targeted to the Gluc reporter. Although the Spinach2 signal was less robust than that observed using conventional CRISPR-based imaging[18] or our own bridged imaging approach, this experiment is an important proof-of-principle that artificial RNA domains can imbue dCas9 with novel properties.

DISCUSSION

In the short time since its initial characterization[14,15], *Sp.* CRISPR–Cas9 has already been coopted for a host of powerful genome modification and regulatory technologies[17–22]. We envision that CRISP-Disp, which limits the function of dCas9 to DNA targeting and 'outsources' all other roles to RNA domains, will provide the basis for an even wider array of methods (**Fig. 4e**). The present studies establish a preliminary framework for the implementation of such experiments, including 'best practices' for RNA construct design (**Supplementary Note 3**).

CRISP-Disp has several implications. First, it allows cargo-specific functions to manifest at a number of target sites using a common toolkit. In theory, the breadth of distinct functions accessible in a single experiment is limited only by the number of orthogonal RNA domains or RNA–binding-protein pairs available (as in **Fig. 4a**–**c**). For example, one might simultaneously activate and repress transcription, epigenetically mark histones and DNA, induce double-stranded breaks and image discrete sets of target loci. Our preliminary experiments targeting discrete, unrelated genomic loci for binding, activation and imaging (**Fig. 4a**–**c**) suggest the broad utility of the approach.

Second, because CRISP-Disp is not intrinsically limited by RNA length (**Figs. 2** and **3a**,**b**), it may provide a method for the locus-targeted reconstitution of natural regulatory RNAs, which could advance the study of lncRNA mechanism[1–6]. For example, CRISP-Disp could be used to bring a lncRNA domain to a target locus in order to assess if its activity can be decoupled from its transcription. Our preliminary data (**Fig. 2d**–**f**) demonstrate the plausibility of such experiments, but more thorough, *bona fide*

© 2015 Nature America, Inc. All rights reserved.

lncRNA reconstitution is likely to require consideration of many variables, such as the chromatin state at the targeted locus.

Third, CRISP-Disp brings the broad functional repertoire of artificial RNA devices to the Cas9 platform. We demonstrated an imaging application (**Fig. 4d**), and others based on aptamers, ribozymes, sensors, processors and scaffolds[7–13] are possible. For example, CRISP-Disp with RNA scaffolds might allow enzymatic activities to be targeted to discrete subnuclear sites[12]. RNA processors might enable complex regulatory functions to be gated in response to external stimuli or small molecules[9–11,19].

Finally, CRISP-Disp may be used to isolate novel functional RNA domains. The sgRNA scaffold can accommodate a wide range of internally inserted sequences (**Fig. 3c**,**d**), suggesting that it could be used as the backbone for SELEX. Such selections might yield, for example, aptamers that sequester or recruit endogenous protein complexes to target loci.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus SuperSeries GSE66756 (encompasses the RIP-Seq data in SubSeries GSE62305 and total mRNA sequencing in SubSeries GSE66755).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank L. Cong and F. Zhang for TALE plasmids and advice on reporter construct design; Y. Sancak, C. Fulco, S. Donovan and M. Morse for their general technical assistance; G. Kenty for his help with luminometry; D. Hendrickson and D. Tenen for their RIP expertise; C. Gerhardinger and C. Daly for their support with deep sequencing; and M. Tabebordbar, J. LaVecchio, S. Ionescu and M. Sauvageau for their assistance with FACS. We are grateful to all members of the Rinn laboratory for their thoughtful discussions and critiques. This work was supported by US National Institutes of Health grant P01GM099117.

AUTHOR CONTRIBUTIONS

D.M.S. designed and performed experiments; E.H. designed and performed microscopy experiments (**Figs. 4c**,**d** and **Supplementary Figs. 13** and **14**); S.T.Y. assisted with computational analysis (**Fig. 3d** and **Supplementary Figs. 5** and **10**); and J.L.R. directed research. D.M.S. and J.L.R. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cech, T.R. & Steitz, J.A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* **81**, 145–166 (2012).
3. Ulitsky, I. & Bartel, D.P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
4. Minks, J., Baldry, S.E., Yang, C., Cotton, A.M. & Brown, C.J. XIST-induced silencing of flanking genes is achieved by additive action of repeat a monomers in human somatic cells. *Epigenetics Chromatin* **6**, 23 (2013).
5. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749 (2013).
6. Bassett, A.R. *et al.* Considerations when investigating lncRNA function *in vivo*. *eLife* **3**, e03058 (2014).
7. 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 Mol. Biol.* **714**, 423–444 (2011).
8. Tome, J.M. *et al.* Comprehensive analysis of RNA-protein interactions by high-throughput sequencing–RNA affinity profiling. *Nat. Methods* **11**, 683–688 (2014).
9. Ausländer, S. *et al.* A general design strategy for protein-responsive riboswitches in mammalian cells. *Nat. Methods* **11**, 1154–1160 (2014).
10. Liang, J.C., Bloom, R.J. & Smolke, C.D. Engineering biological systems with synthetic RNA molecules. *Mol. Cell* **43**, 915–926 (2011).
11. Carothers, J.M., Goler, J.A., Juminaga, D. & Keasling, J.D. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716–1719 (2011).
12. 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).
13. Song, W., Strack, R.L., Svensen, N. & Jaffrey, S.R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).
14. Garneau, J.E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
15. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
16. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
17. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
18. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
19. 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. *Mol. Cell* **54**, 698–710 (2014).
20. Ryan, O.W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife* **3**, e03703 (2014).
21. Gilbert, L.A. *et al.* Genome-Scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
22. Citorik, R.J., Mimee, M. & Lu, T.K. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* **32**, 1141–1145 (2014).
23. Briner, A.E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56**, 333–339 (2014).
24. Wright, A.V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. USA* **112**, 2984–2989 (2015).
25. Zalatan, J.G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
26. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
27. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA–guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
28. Maeder, M.L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013).
29. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
30. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).
31. Chao, J.A., Patskovsky, Y., Almo, S.C. & Singer, R.H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
32. Ran, F.A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
33. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
34. Mayer, C., Neubert, M. & Grummt, I. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep.* **9**, 774–780 (2008).
35. Ørom, U.A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
36. Wang, K.C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
37. Saito, H. *et al.* Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat. Chem. Biol.* **6**, 71–78 (2010).
38. Sternberg, S.H., Haurwitz, R.E. & Doudna, J.A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**, 661–672 (2012).
39. Cong, L., Zhou, R., Kuo, Y.C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.* **3**, 968 (2012).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Plasmid synthesis. Mammalian expression and reporter constructs were generated using standard restriction enzyme-based and ligation-independent cloning methods. Components were acquired as follows: the T7 promoter-targeting TALE[30] was the generous gift of F. Zhang (Broad Institute); *Gaussia* and *Cypridina* luciferases were derived from pGLuc-Basic and pCLuc-Basic, respectively (New England BioLabs); dCas9 (*S. pyogenes* D10A/H841A Cas9) was isolated from Addgene plasmid 47754, the EF1α promoter from Addgene plasmid 11154, mCerulean from Addgene plasmid 23244, Venus from Addgene 15753; and the human Ubiquitin C promoter (hUBCPro) used to drive expression of L7Ae~VP, PP7~VP (**Supplementary Fig. 2a**, bottom) and MS2~mCherry (**Supplementary Fig. 13a**) from Addgene plasmid 17627. All other components were synthesized *de novo* from small synthetic oligonucleotides or from gBlocks (Integrated DNA Technologies).

The backbone for lentiviral reporter constructs was derived from pLenti6.3/TO/V5-DEST (Life Technologies), from which the Tet-reponsive promoter and Gateway cloning sites were removed. The backbone for the T7 TALE and MS2~VP constructs was derived from pcDNA3.1(+) (Life Technologies) in which the neomycin expression cassette was removed. All other constructs were cloned into pNEB193 (New England BioLabs).

L7Ae, MS2 and PP7 were codon-optimized for expression in human cells and synthesized as gBlocks (Integrated DNA Technologies). The PP7 construct consists of two tandem copies of the non-aggregating ΔFG mutant[31] joined by a flexible seven amino acid linker with the sequence GSTSGSG (**Supplementary Fig. 2a**, bottom). Similarly, the MS2 construct consists of two tandem copies of the non-aggregating V75E/A81G mutant[40] joined by the same linker. L7Ae was designed according to a published sequence[37].

INT-like constructs (**Figs. 3** and **4**, and **Supplementary Table 5**) were cloned as follows. We first cloned an INT general-purpose cloning vector, "sgINTgpc," containing the following pertinent sequence:

*GATCTAGATACGACTCACTAT*GTTTAAGAGCTATGCTGC
GAATACGAGAA**GTCTTC**TTTTTT**GAAGAC**AATCGTATTC
GCAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAA
CTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

wherein italicized nucleotides denote the GLuc-targeting protospacer sequence (**Supplementary Table 2**), underlined nucleotides denote an extended sgRNA stem1 (ref. 18) and bold nucleotides denote two outward-facing *BbsI* restriction sites. This cassette is under expression of a human U6 promoter (not shown). Inserts cloned into this backbone had the general format: 5′ CGAG–[Insert]–CTCGT 3′, wherein underlined nucleotides denote the sticky ends used for cloning and the additional "C" following the insert restores base-pairing at the end of stem1. These inserts were generated by PCR and restriction digestion with *BbsI*, or by annealing synthetic, 5′-phosphorylated oligonucleotides (following the protocol used for the N_{25} pool, see below). Inserts were ligated into *BbsI*-digested, gel-purified sgINTgpc using the Quick Ligation Kit (New England BioLabs).

All sgRNAs and derivatives were initially cloned bearing a GLuc-targeting protospacer (**Supplementary Table 2**). *ASCL1*-, *IL1RN*-, *NTF3*-, *TTN*- and telomere-targeting constructs (**Figs. 1d**, **3f**, **4b–d** and **Supplementary Table 2**) were derived from these parental constructs using an inverse-PCR method, using a forward primer that anneals downstream of the protospacer and a reverse primer that anneals to the 3′ end of the U6 promoter. Namely, PCR products were amplified with primers of the general format:

Forward: TAGTA**GAAGAC**AAXXXXXXXXXXXXXGTTTA
AGAGCTATGCTGCGAATACG

Reverse: TAGTA**GAAGAC**AAYYYYYYYYYYYYYGGTGTTT
CGTCCTTTCCAC

The bold nucleotides denote *BbsI* restriction sites; X's denote nucleotides 9–21 of the new protospacer sequence; Y's denote the reverse complement of nucleotides 1–9 of the new protospacer; and underlined nucleotides are reverse complementary to one another. PCR products were purified using the QIAgen PCR cleanup kit, digested with *BbsI* and *DpnI*, purified again and quantified by UV-vis spectroscopy. Products (25 ng, in 11 μL final) were self-ligated using the Quick Ligation Kit (New England BioLabs).

All plasmid sequences were confirmed by Sanger sequencing (GeneWiz) before use.

Cloning the N_{25} pool. Pool oligonucleotides (Integrated DNA Technologies) were as follows:

5′ [P]–CGAGNNNNNNNNNNNNNNNNNNNNNNC–3′
5′ [P]–ACGAGNNNNNNNNNNNNNNNNNNNNNNN–3′

"5′ [P]" denotes a 5′ phosphate, and N denotes an equimolar mixture of all four nucleotides. Oligonucleotides were each resuspended in annealing buffer (10 mM Tris, pH 7.0, 50 mM NaCl) to 100 μM. 10 μL of each oligo were mixed in a 0.2 mL PCR tube; this mixture was heated to 95 °C for 10 min and slowly annealed to 25 °C over the course of 2 h in a thermocycler. The reaction was snap-cooled on ice and diluted 100 fold with ice-cold annealing buffer. 1 μL of this diluted duplex mix was ligated into 25 ng of *BbsI*-cut sgINTgpc, in 12 μL final volume, using the Quick Ligation Kit (New England BioLabs). The entire reaction was transformed into 120 μL of XL10-Gold ultracompetent cells (Agilent), plated onto 12 LB ampicillin LB(Amp) plates and grown overnight at 37 °C. Seven bacterial colonies were picked for Sanger sequencing (**Supplementary Fig. 9**), and the remainder were pooled by scraping the plates into 100 mL of liquid LB(Amp). Bacteria were pelleted by ultracentrifugation, and the plasmid pool was harvested in a single plasmid maxi-prep (QIAgen) (**Supplementary Figs. 9** and **10**).

Cell culture, stable and **transient transgene expression.** HEK293FT cells were acquired from ATCC, and used without further validation. Cells were maintained on gelatinized plates in high glucose Dulbecco's modified Eagle's medium (DMEM, Gibco), supplemented with 10% FBS, 1× penicillin-streptomycin and 2 mM L-Glutamine (Gibco). Cells were grown at 37 °C and 5% CO_2 in a humidified incubator. Cells were routinely checked for mycoplasma contamination every 3–6 months, using the PCR Mycoplasma Test Kit I/C (PromoKine).

Lentiviral particles were generated using standard second generation packaging plasmids, in 293T cells. Integrated reporter cells were generated as follows: 250,000 HEK293FT cells were plated per well of a gelatinized 6-well dish and incubated overnight. Growth media was thereafter removed, cells were washed

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nmeth.3433

once in warmed PBS and supplied with 1.7 mL fresh warmed media supplemented with 200 µL CLuc reporter lentivirus and 8 µg/mL polybrene. After 24 h this process was repeated with a second dose of CLuc virus. Cells were subsequently passaged onto 10-cm gelatinized plates and selected with 2 µg/mL puromycin. CLuc reporter cells were then plated onto gelatinized 6-well dishes and transduced with GLuc reporter lentivirus following the same transduction protocol. GLuc-transduced cells were not selected with hygromycin before use. A lentiviral variant of our EF1α-dCas9 construct (**Supplementary Fig. 1c**) was also used for aptamer-based imaging (**Fig. 4d**) following the same transduction protocol (without antibiotic selection). To enrich for cells that expressed low levels of dCas9 (ref. 18), we transiently transfected GLuc reporter, U6-INT and PP7~VP plasmids (as in analytical luciferase assays, see below) and collected $GLuc^+$ cells by FACS.

Transient transfections were performed using Lipofectamine 2000 (Life Technologies), following the manufacturer's protocol. For luciferase assays, 125,000 cells in 0.6 mL media were plated per well of gelatinized 12-well dishes and incubated overnight. Transfection mixes contained 33 ng of each luciferase reporter plasmid (where appropriate), 59 ng of dCas9 or dCas9~VP plasmid, 66 ng of PP7~VP, L7Ae~VP or MS2~VP (where appropriate), 11.6 ng of U6-driven or 542 ng of Pol II–driven sgRNA variants. For experiments using *TOP3* and *TOP4* (**Fig. 1a**,**c** and **Supplementary Fig. 6b**), 11.6 ng of a separate U6-driven gRNA plasmid was also included. For FACS, (**Supplementary Figs. 1e**, **2b**, **4e** and **6d**) transfection mixes also contained 10 ng of an mCherry cotransfection control. In all cases, the total transfected plasmid mass was brought to 750 ng per well using pNEB193 (New England BioLabs) in 18 µL final volume, with 2.25 µL Lipofectamine 2000.

For RNA immunoprecipitation (RIP) and RIP-Seq experiments, 2.1 million cells in 10 mL growth media were plated onto gelatinized 10-cm dishes and grown overnight. Transfection mixes were as described above, but all masses and volumes were scaled 15.7 fold to account for the increase in growth area and cell number. RIP transfection mixes included each luciferase reporter to independently monitor CRISP-Disp function.

To test CRISP-Disp function at endogenous loci (**Figs. 1d** and **3f**), cells were plated in gelatinized 12-well dishes as in standard luciferase assays. Transfection mixes were similar to those described[28] and contained 500 ng dCas9 or dCas9~VP plasmid, 500 ng GLuc-Targeting sgRNA construct or 500 ng of a mix containing equal masses (125 ng each) of four constructs targeting the particular locus under examination (*ASCL1*, *IL1RN*, *NTF3* or *TTN*[28,29,33]; **Supplementary Table 2**). Where appropriate, 556 ng of PP7~VP plasmid was also included. All mixes were brought to 1556 ng per well using pNEB193, in 38 uL final volume, with 4.7 µL Lipofectamine 2000.

For multiplexing experiments (**Fig. 4b**) cells were plated in gelatinized 12-well dishes as above. Transfection mixes contained 250 ng dCas9 or dCas9~VP, 250 ng GLuc-targeting sgRNA variant, 250 ng of a mix containing equal masses (62.5 ng each) of four *NTF3*-targeting constructs (**Supplementary Table 2**) and 278 ng of PP7~VP or MS2~VP, where appropriate. In all cases the total transfected mass was brought to 1028 ng using pNEB193, in 30 µL volume, with 3.1 µL Lipofectamine 2000.

In bridged imaging experiments (**Fig. 4c** and **Supplementary Fig. 13**), 80,000 cells in 1 mL growth media were plated per well of untreated Nunc Lab-Tek glass two-chamber slides (Thermo Scientific). 24 h thereafter, growth media was changed and cells were transfected with 440 ng dCas9 and 235 ng of each modified sgRNA. Where appropriate, 440 ng of PP7~VP64 and/or 100 ng of MS2~mCherry were included. The total transfected mass was brought to 1,500 ng with pNEB193, in 11.4 µL, with 4.5 µL Lipofectamine 2000, according to the manufacturer's protocol. For aptamer-based imaging (**Fig. 4d**), 80,000 dCas9-transduced ("+dCas9," see above) or untransduced ("–dCas9") cells in 1 mL growth media were plated per well of Nunc Lab-Tek glass two-chamber slides that had been treated as follows. Wells were coated with 100 µg/mL poly-L-lysine (Millipore) overnight at 4 °C. The next day, wells were washed twice with ddH_2O, UV sterilized for five minutes in a biosafety cabinet, coated with 100 µg/mL rat collagen-I (Corning) and 50 µg/mL laminin (Life Technologies) for 2 h at 37 °C, and dried before plating cells. Transfections were performed 24 h thereafter, with 600 ng (telomere- or GLuc-targeting) INT-Spinach2 construct, 600 ng of pNEB193 and 4.5 ng of an mCherry cotransfection control, in a total volume of 11.4 µL, with 3.8 µL Lipofectamine 2000, according to the manufacturer's protocol. All live-cell imaging experiments were performed 48–72 h post-transfection (see below).

Luciferase and FACS assays. Luciferase assays were performed using the BioLux *Gaussia* and *Cypridina* Luciferase Assay kits (New England BioLabs), following the manufacturer's protocols. Growth media (200 µL) was harvested 3 d after transfection and, if not used immediately, was stored in the dark at 4 °C in parafilm-sealed 96-well dishes. 20 µL of each experimental sample was manually pipetted into black-walled 96-well plates (Corning) and assayed using a FLUOstar OPTIMA Luminometer equipped with automatic injectors (BMG Labtech). *Gaussia* and *Cypridina* assays were performed in parallel. For each, a single empirically determined gain was applied to all samples within an experimental series. Each sample was injected with 50 µL of luciferase assay buffer and mixed for two seconds before data acquisition. Signal was integrated over 20 s using an open (unfiltered) top-down optic.

For each sample, experimental raw luciferase signals were background subtracted, and the ratio of luciferase values (GLuc/CLuc) was calculated. Biological replicates (at least three per experiment) were used to calculate a mean value, <GLuc/Cluc>. Fold activation was then calculated relative to a control sample in which dCas9~VP was expressed in the absence of an sgRNA construct:

$$\text{Fold activation} = \frac{\left\langle \frac{\text{GLuc}}{\text{CLuc}} \right\rangle_{(\text{experimental sample})}}{\left\langle \frac{\text{GLuc}}{\text{CLuc}} \right\rangle_{(\text{dCas9~VP alone})}}$$

Statistical significance testing likewise used this dCas9~VP control as the basis for comparison.

For FACS assays, cells were propagated and transfected in gelatinized 12-well dishes, as described for luciferase assays, and analyzed 3 d after transfection. Cells were harvested by

© 2015 Nature America, Inc. All rights reserved.

trypsinization, quenched by the addition of chilled growth media, diluted threefold in chilled staining media (Hank's Balanced Salt Solution (HBSS, Gibco), supplemented with 2% Donor Bovine Serum (DBS, Atlanta Biologicals)), and pelleted at 200 g in a swinging bucket rotor. Cells were resuspended in chilled staining media and analyzed on a BD LSR II Flow Cytometer (BD Biosciences), equipped with HcRed CFP and YFP filters. Voltages, compensations and gates were empirically determined using unstained and single color controls, according to standard methods. 100,000 $mCherry^+$ cells were recorded from each sample.

RNA immunoprecipitation (RIP). Cells were propagated on gelatinized 10-cm dishes, transfected as described above and harvested 3 d after transfection. Thereafter, RIP was performed essentially as described previously[41]. Growth media was aspirated and cells were washed twice with 10 mL room temperature PBS (Gibco). Cells were cross-linked by incubation in 0.1% (v/v) formaldehyde in PBS for 10 min at room temperature, under very gentle agitation. Cross-linking was quenched by the addition of glycine to 133 mM and gentle agitation for an additional 5 min at room temperature, after which the liquid phase was aspirated. Cross-linked cells were washed twice with room temperature PBS, harvested by scraping, allotted into samples of 1×10^7 cells (typically three samples per 10-cm dish) and pelleted at 200*g* in a swinging bucket rotor. PBS was aspirated and cell pellets were flash-frozen in liquid nitrogen and stored at −80 °C until use.

Cell pellets were thawed on ice, gently resuspended into 1 mL of ice-cold RIPA(+) buffer (standard RIPA supplemented with 0.1 U/μL RNAseOUT (Life Technologies), 1× EDTA-free Proteinase Inhibitor Cocktail (Thermo Scientific) and 0.5 mM DTT) and lysed for 10 min at 4 °C with end-over-end agitation. Samples were then sheared using a Branson Digital Sonifier 250 (Emerson Industrial Automation) at 10% amplitude for three 30-s intervals (0.7 s on + 1.3 s off), with 30-s resting steps between intervals. Samples were held in ice-cold metal thermal blocks throughout sonication. Sheared samples were then clarified by ultracentrifugation and diluted with 1 mL each of ice-cold Native Lysis Buffer(+) (25 mM Tris, pH 7.4, 150 mM KCl, 5 mM EDTA, 0.5% (v/v) NP-40, supplemented with inhibitors and DTT as above), filtered through a 0.45 μm syringe-mounted filter, and flash-frozen in liquid nitrogen before use.

Clarified lysates were thawed on ice and pre-cleared by incubation with buffer-equilibrated magnetic Protein G beads (Life Technologies) for 30 min at 4 °C with end-over-end rotation. 100 μL aliquots were removed and frozen, to serve as "input" normalization controls. Cleared lysates corresponding to 5×10^6 cells were then incubated with 6 μg rabbit anti-FLAG (Sigma) or rabbit normal IgG (Cell Signaling Technology), for 2 h at 4 °C with end-over-end rotation. Buffer-equilibrated magnetic Protein G beads were then added and the samples were again rotated end-over-end for 1 h at 4 °C. Beads were collected and washed twice with Native Lysis Buffer(+) for 10 min at 4 °C with end-over-end rotation. Immunoprecipitated RNA was thereafter isolated as described below.

RNA isolation, quantitative RT–PCR and mRNA-seq. Whole cell RNA (**Figs. 1d**, **3f** and **4b** and **Supplementary Figs. 5** and **12**) and RNA from subcellular fractions (**Supplementary Fig. 7**) were isolated by extraction with Trizol and Trizol-LS Reagent (Life Technologies), respectively, following the manufacturer's protocols. RNA was precipitated with isopropanol using GlycoBlue (Life Technologies) as a carrier, and subsequently purified using RNEasy spin columns (QIAgen), following the manufacturer's "RNA Cleanup" protocol, with on-column DNase treatment.

RNA from RIP and RIP-seq experiments (**Figs. 2c**,**e** and **3d** and **Supplementary Figs. 2d**, **4f** and **6e**) was isolated as follows. Following RIP (see above), protein G beads were suspended in 56 μL nuclease-free water and processed alongside input samples (56 μL; 5.6% of the total). All samples were brought to 100 μL with 3× Reverse Cross-linking Buffer (final concentrations: 1× PBS, 2% N-Lauroyl Sarcosine, 10 mM EDTA, 5 mM DTT, 0.4 U/μL RNaseOUT and 2 mg/mL proteinase K (Ambion)). Formaldehyde cross-links were reversed by incubation in a thermocycler at 42 °C for 1 h, and then 55 °C for 1 h. RNA was thereafter purified using 400 μL Agencourt RNAClean XP Beads (Beckman Coulter), following the manufacturer's protocol, and eluted into 30 μL nuclease-free water. Residual DNA was removed by treatment with 5 U RNase-free DNase (RQ1, Promega) in 50 μL, following the manufacturer's protocol. RNA was subsequently purified using 200 μL Agencourt RNAClean XP beads, eluted into 20 μL nuclease-free water and stored at −20 °C until use.

cDNA was synthesized using SuperScript III reverse transcriptase (Life Technologies), according to the manufacturer's protocol, priming from anchored oligo-dT_{21}, random hexamers (Life Technologies) or a gene specific primer (Integrated DNA Technologies), where appropriate. Target RNA abundance was quantified by qRT-PCR on a 7900HT Fast Real-Time PCR System (Applied Biosystems), using Rox-normalized FastStart Universal SYBR Green Master Mix (Roche) and gene-specific primers (**Supplementary Table 3**), in quadruplicate. Nonreverse-transcribed RNA was used as a negative control. "Clipped" data were processed using Realtime PCR Miner[42], to calculate C_T and efficiency values (**Supplementary Table 3**). Bulk gene expression measurements (**Figs. 1c**, **3f** and **4b** and **Supplementary Fig. 12**) were normalized to a GAPDH internal control (**Supplementary Table 3**) and RIP measurements were normalized to input RNA levels. In subcellular fractionation experiments (**Supplementary Fig. 7**), the yield of RNA in each compartment was quantified relative to the unfractionated input level, as in RIP experiments. Data analysis was performed using standard methods.

For global gene expression analysis (**Supplementary Fig. 5**), poly(A)$^+$ mRNA-seq libraries were prepared using the TruSeq RNA sample preparation kit, v2 (Illumina) as described[41]. Indexed libraries were pooled and subjected to 50 cycles of paired end sequencing, followed by 25 cycles of indexing, on two lanes of an Illumina HiSeq 2500 (FAS Center for Systems Biology, Harvard). For characterization of gene expression, sequencing reads were mapped to a custom gene set comprising UCSC known human genes (hg19) appended with dCas9, GLuc, CLuc and sgRNA constructs, using TopHat2 with default options[43]. Differential analysis of gene expression was assessed using Cuffdiff2 with default options[44]. Genes plotted in **Supplementary Fig. 5** were restricted to the top 75% of expressed genes, based on FPKM values.

Error propagation and reproducibility. For luciferase and qRT-PCR assays, experimental uncertainties were propagated as

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nmeth.3433

described previously[45]. Given S, the sum or difference of values A and B, uncertainly was calculated using the formula:

$$\sigma_S = \sqrt{(\sigma_A)^2 + (\sigma_B)^2}$$

wherein σ_A and σ_B are the measurement errors of A and B, respectively. For P, the product or quotient of values A and B, uncertainty was calculated using the formula:

$$\sigma_P = P \times \sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2}$$

The uncertainty of other functions, F(x), was calculated using the first derivative approximation:

$$\sigma_{f(x)} = \sigma_x \times f'(x)$$

Sample sizes were determined in accordance with standard practices used in similar experiments in the literature; no sample-size estimates were performed to ensure adequate power to detect a prespecified effect size. Experiments were neither randomized nor blinded to experimental conditions. No samples were excluded from analysis.

Subcellular fractionation. Cytoplasmic and nuclear fractions (**Supplementary Fig. 7**) were isolated as described[46,47]. Briefly, cells were grown and transfected in gelatinized 10-cm dishes, as described above for RIP experiments. 3 d after transfection, cells were harvested by trypsinization, quenched with growth media, pelleted and washed thrice with ice-cold PBS. Cells were gently resuspended in five packed cell pellet volumes ("cv's") of ice-cold Cyto Extract Buffer(+) (20 mM Tris, pH 7.6, 0.1 mM EDTA, 2 mM $MgCl_2$, supplemented with 0.5 U/μL RNAseOUT and 1× EDTA-free Proteinase Inhibitor Cocktail), and swollen by incubation at room temperature for 2 min, and on ice for 10 min more. Cells were then lysed by addition of CHAPS to 0.6% final, gentle pipetting, and two passages through a syringe equipped with a 20G needle. Lysate was clarified by centrifugation at 500*g* in a tabletop microcentrifuge at 4 °C; 70% of the resulting supernatant was retrieved as the cytoplasmic fraction. The pellet, corresponding to nuclei and cell debris, was washed twice by gentle resuspension into five cv's of Nuclear Wash Buffer(+) (Cyto Extract Buffer, supplemented to 0.6% CHAPS and with inhibitors, as above) followed by centrifugation at 500*g*. Washed nuclei were gently resuspended into 2 cv's of Nuclei Resuspension Buffer(+) (10 mM Tris, pH 7.5, 150 mM NaCl, 0.15% (v/v) NP-40, supplemented with inhibitors, as above), layered onto a cushion of 5 cv's Sucrose Buffer(+) (10 mM Tris, pH 7.5, 150 mM NaCl, 24% (w/v) Sucrose, plus inhibitors) and pelleted at 14,000 rpm in a tabletop microcentrifuge at 4 °C. The resulting pelleted nuclei were resuspended into 2 cv's of ice-cold PBS and pelleted at 500g before harvesting RNA. We confirmed the success of our fractionations by two methods: western blotting and qRT-PCR (**Supplementary Fig. 7b**,**c**). In western blots, aliquots of whole cell lysate, the cytoplasmic fraction and PBS-suspended nuclei were probed using antibodies against (α/β)-Tubulin and Fibrillarin (Cell Signaling Technology, product numbers 2148 and 2639, respectively) at 1:1,000 dilution. For qPCR, extracted RNA (see above) was quantified using primers against the cytoplasmic ncRNA *SNHG5* and the nuclear ncRNA *XIST* (**Supplementary Table 3**).

N_{25} RNA library preparation, sequencing and analysis. For the N_{25} RIP-seq experiment (**Fig. 3d**), cell growth, transfection, RIP and RNA preparation were performed as described above, in triplicate. Seven deep sequencing libraries were prepared: one from the starting plasmid pool, three from replicates of the input RNA and three from replicates of the immunoprecipitated RNA. The plasmid pool library was generated directly via PCR using 5 ng of plasmid template in a 50 μL reaction and amplified through 19 cycles of PCR with *Pfu* Ultra II HS polymerase (Agilent), according to the manufacturer's protocol. Gene-specific PCR primers that bracketed the N_{25} insertion site, appended with standard Illumina adaptors and indexes, were used (**Supplementary Table 6** and **Supplementary Fig. 9**). For each input and RIP library, 10 ng RNA was reverse-transcribed in 20 μL as described above, using a gene-specific primer (**Supplementary Table 6** and **Supplementary Fig. 9**). Each cDNA reaction was used in its entirety as a PCR template, using the same primer design as was used for the plasmid pool, but with different Illumina indexes. The pools were amplified in 200 μL, through 26 cycles of PCR with *Pfu* Ultra II HS polymerase (Agilent), according to the manufacturer's protocol. The resulting deep sequencing libraries were purified twice with one volume of Agencourt AMPure XP Beads (Beckman Coulter), according the manufacturer's protocol, and eluted in EB Buffer (QIAgen). The plasmid pool library contained traces of high–molecular weight contaminants (not shown) that were removed by "reverse selection:" the sample was treated with 0.65 volumes of AMPure XP Beads and the unbound fraction was retained. The integrity and concentration of each final library was measured using a "DNA High Sensitivity" assay on an Agilent 2100 model Bioanalyzer (**Supplementary Fig. 9**).

Libraries were denatured in 50 mM NaOH, diluted in buffer HT1 (Illumina) and combined to yield a 20 pM pool, according to standard protocols. This pool was doped with TailorMix Indexed PhiX Control Library (SeqMatic), at a ratio of 7:3 N_{25}Pool:PhiX, and sequenced on two lanes of an Illumina HiSeq 2500 (FAS Center for Systems Biology, Harvard) for 150 cycles, followed by 25 cycles of indexing.

Random insert sequences were extracted from raw sequencing reads by removing the constant sequences abutting each side of the insertion point. The number of occurrences of each random sequence within each individual sample was then tabulated. Sequence counts were used to calculate enrichment using DESeq2 (ref. 48).

Live-cell imaging. Images in (**Supplementary Fig. 1d**) were collected on an Axio Observer D1 system (Zeiss) equipped with eYFP and eCFP filters.

Live fluorescence images (**Fig. 4c**,**d** and **Supplementary Figs. 13** and **14**) were taken with an LSM 700 Inverted Confocal Microscope (Harvard Center for Biological Imaging), with an aperture setting of 1 A.U., using the DAPI filter for Hoechst 33342, the CFP filter for mCerulean, the mCherry filter for mCherry and the FITC filter for DFHBI-1T, where appropriate. In bridged imaging experiments (**Fig. 4c** and **Supplementary Fig. 13**),

© 2015 Nature America, Inc. All rights reserved.

cells were imaged 2 d post-transfection, in their growth media. Images are max-merges of 37–47 Z-stacks, taken with a step size of 0.33 µm, at 63× magnification. For aptamer-based imaging (**Fig. 4d** and **Supplementary Fig. 14**), growth media was replaced with imaging media (Fluorobrite DMEM (Life Technologies), 25 mM HEPES, 5 mM $MgSO_4$, 1 µg/ml Hoechst 33342 (Life Technologies) and 20 µM DFHBI-1T (Lucerna)) for 30 min at 37 °C. Images in **Figure 4c**,**d** are max-merges of 20–30 Z-stacks, taken with a step size 0.35 µm, at 63× magnification.

40. 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).
41. Kelley, D.R., Hendrickson, D.G., Tenen, D. & Rinn, J.L. Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. *Genome Biol.* **15**, 537 (2014).
42. Zhao, S. & Fernald, R.D. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J. Comput. Biol.* **12**, 1047–1064 (2005).
43. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
44. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
45. Shechner, D.M. & Bartel, D.P. The structural basis of RNA-catalyzed RNA polymerization. *Nat. Struct. Mol. Biol.* **18**, 1036–1042 (2011).
46. Rosner, M. & Hengstschlager, M. Detection of cytoplasmic and nuclear functions of mTOR by fractionation. *Methods Mol. Biol.* **821**, 105–124 (2012).
47. Bhatt, D.M. *et al.* Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290 (2012).
48. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nmeth.3433