



Published in final edited form as:

*Nat Biotechnol.* 2013 April ; 31(4): 350–356. doi:10.1038/nbt.2537.

## Efficient and specific gene knockdown by small interfering RNAs produced in bacteria

Linfeng Huang<sup>1</sup>, Jingmin Jin<sup>2</sup>, Padraig Deighan<sup>3,4</sup>, Evgeny Kiner<sup>1</sup>, Larry McReynolds<sup>2</sup>, and Judy Lieberman<sup>1</sup>

<sup>1</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston MA USA

<sup>2</sup>New England Biolabs, Ipswich MA USA

<sup>3</sup>Department of Microbiology and Immunobiology, Harvard Medical School, Boston MA USA

<sup>4</sup>Department of Biology, Emmanuel College, Boston MA USA

### Abstract

Synthetic small interfering RNAs (siRNAs) are an indispensable tool to investigate gene function in eukaryotic cells<sup>1,2</sup> and may be used for therapeutic purposes to knockdown genes implicated in disease<sup>3</sup>. Thus far, most synthetic siRNAs have been produced by chemical synthesis. Here we present a method to produce highly potent siRNAs in *E. coli*. This method relies on ectopic expression of p19, a siRNA-binding protein found in a plant RNA virus<sup>4,5</sup>. When expressed in *E. coli*, p19 stabilizes ~21 nt siRNA-like species produced by bacterial RNase III. Transfection of mammalian cells with siRNAs, generated in bacteria expressing p19 and a hairpin RNA encoding 200 or more nucleotides of a target gene, at low nanomolar concentrations reproducibly knocks down gene expression by ~90% without immunogenicity or off-target effects. Because bacterially produced siRNAs contain multiple sequences against a target gene, they may be especially useful for suppressing polymorphic cellular or viral genes.

---

RNA interference (RNAi) by double-stranded (ds) siRNAs induces the degradation of mRNAs bearing complementary sequences<sup>6,7</sup>. Although most synthetic siRNAs are designed by computer algorithms and produced by chemical synthesis, siRNAs can also be made from transcribed longer dsRNAs that are processed *in vitro* by RNase III family enzymes<sup>8,9</sup>. In the latter case, the resulting siRNAs contain many sequences against one target (rather than a single sequence as occurs with chemically-synthesized siRNA). A pool of several siRNAs can sometimes be more effective and have fewer off-target effects than any one single siRNA<sup>10,11</sup>. However, thus far functional siRNAs have not been produced in

---

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence to: Judy.Lieberman@childrens.harvard.edu.

**Competing financial interests** J.J. and L.M. are employees of New England BioLabs, a company that sells deep sequencing kits, p19 and other proteins for RNA and DNA research. L.H., P.D., E.K. and J.L. declare no competing financial interests.

**Author Contributions** L.H. and J.L. designed the experiments with advice from J.J., L.M., and P.D.. J.J. and L.M. prepared p19 beads and RNA deep sequencing libraries. P.D. constructed *E. coli* mutant strains. E.K. performed siRNA comparison and macrophage transfection experiments. L.H. performed all other experiments. L.H. and J.L. wrote the paper.

living cells. Here, we engineer bacterial cells to produce fully processed ready-to-use siRNAs specific for a target gene of interest.

The p19 protein encoded by the plant RNA virus *tombusvirus*<sup>4</sup> selectively binds to and inhibits the function of ~21 nt siRNAs, including those containing sequences complementary to virus RNA<sup>5</sup>. A p19 dimer binds to the ~19 nt duplex region of an siRNA in a sequence-independent manner<sup>12,13</sup>. In agreement with previous publications, we found that endogenous siRNAs in mammalian cells can be isolated using p19 coupled to magnetic beads (Fig. 1a)<sup>13</sup>. As a negative control because *E. coli* lacks canonical RNAi-processing machinery, we used p19 beads incubated with total RNA isolated from *E. coli* (a wild-type strain and a strain transformed with a pcDNA3.1+ plasmid encoding p19). Surprisingly, p19-coupled beads retrieved ~21 nt dsRNAs from the p19 plasmid-containing *E. coli* strain (Fig. 1a). Although the CMV promoter<sup>14</sup> driving expression from this plasmid is mostly used for efficient gene expression in mammalian cells, pcDNA3.1+ plasmids encoding FLAG-tagged *p19* or a FLAG-tagged control gene of a similar size (*TREX1*) drove detectable protein expression in *E. coli* (Fig. 1b). We detected small ~21 nt RNAs on SYBR Gold-stained denaturing polyacrylamide gels of total RNA harvested from p19-expressing bacteria, but not on gels of total RNA isolated from bacteria transformed with the empty vector or a vector encoding *TREX1* (Fig. 1b). These data suggest that p19 expression stabilizes a cryptic siRNA-like RNA species in *E. coli*. Similarly sized small RNAs were also detected in p19-expressing, but not WT, strains of the Gram-positive bacterium *Listeria monocytogenes* (Supplementary Fig. 1).

To determine if the small RNAs detected in *E. coli* depended on functional p19, RNA was isolated from *E. coli* expressing WT p19 or p19 containing mutations that disrupt siRNA binding<sup>12,15</sup> (Fig. 1c). The ~21 nt dsRNA band was more prominent in bacteria expressing WT than mutant p19. Thus siRNA-binding to p19 promotes the accumulation of siRNA-like RNAs in *E. coli*.

Next we looked for the nuclease responsible for making small RNAs. The most likely candidate was RNase III, an ancestor of eukaryotic Dicer, which is responsible for the final step of siRNA biogenesis<sup>16</sup>. *E. coli* RNase III can generate siRNA-sized dsRNAs from longer dsRNAs *in vitro*<sup>9</sup>. We transformed two RNase III mutant strains, *rnc14*<sup>17</sup> and *rnc38*<sup>18</sup>, with plasmids encoding p19 (Fig. 1d). In both RNase III mutant strains, p19 beads failed to pull down any visible small RNAs. Furthermore, restoration of RNase III expression in HT115(DE3) cells, a *rnc14* strain, restored the production of p19-dependent small RNAs (Fig. 1e). Thus, accumulation of these small RNAs in bacteria depends on ectopic p19 expression and endogenous RNase III expression.

We next asked whether small RNAs generated in p19-expressing *E. coli* exhibit properties similar to those of chemically synthesized siRNAs. We cloned *p19* into the pGEX-4T-1 plasmid to express a GST-p19 fusion protein with a C-terminal His tag (Fig. 2a). A T7 promoter driving expression of a hairpin RNA encoding the sequence of the target gene was inserted immediately after the His tag in this plasmid. We first used a hairpin encoding full-length *EGFP* (*EGFPFL*). The expression of the GST-p19-His fusion protein and hairpin RNA were both induced by IPTG. We used nickel (Ni) affinity chromatography to capture

the GST-p19-His protein, and 0.5% SDS to selectively elute p19-bound RNAs, which were predominantly ~21 nt long (Fig. 2b and Supplementary Fig. 2). Small RNAs were further purified from other longer RNAs by anion exchange high-performance liquid chromatography (HPLC). To verify that these bacterial small RNAs, which we refer to as pro-siRNAs for prokaryotic siRNAs, are double stranded, we treated them with a variety of nucleases. Like chemically synthesized siRNAs, bacterial small RNAs were sensitive to RNase A, but were insensitive to enzymes that digest ssRNA or DNA (Xrn1, RNase T1, exonuclease T (Exo T), exonuclease I (Exo I), or DNase Turbo) (Fig. 2c). Next we tested whether transfection of bacterial small RNAs that were purified from *E. coli* expressing p19 and *EGFPFL* hairpin into HeLa cells stably expressing *dIEGFP* (HeLa-dIEGFP) would load them into Argonaute (Ago), the central component of the RNA-induced silencing complex (RISC). To do this we performed immunoprecipitation with a pan-Ago antibody, and analyzed the ability of the associated RNAs to hybridize to an *EGFP* probe (Northern blot) (Fig. 2d). RNAs that co-immunoprecipitated with anti-Ago were ~21 nt long and hybridized to the *EGFP* probe only in pro-siRNA transfected cells; in contrast no small RNA co-immunoprecipitated with control mouse IgG. Thus bacterial small RNAs were similar to synthetic siRNAs in chemical composition and were incorporated into the RISC.

Since pro-siRNAs exhibit properties of siRNAs, we next tested whether pro-siRNAs can suppress expression of specific target genes. We transfected HeLa-dIEGFP cells with a chemically synthesized *EGFP* siRNA or with pro-siRNAs purified from *E. coli* expressing p19 and *EGFPFL* hairpin or a hairpin encoding a 100 nt fragment of *EGFP* that overlapped with the chemically synthesized siRNA sequence (*EGFP100*). As measured by quantitative RT-PCR (qRT-PCR) and flow cytometry, both *EGFPFL* and *EGFP100* pro-siRNAs knocked down *EGFP* expression more effectively than equimolar concentrations of chemically synthesized siRNA (Fig. 2e and Supplementary Fig. 3a). pro-siRNAs made from the p19-expressing plasmid lacking any *EGFP* sequence or expressing only the antisense half of the *EGFP* hairpin did not effectively knockdown *EGFP* (Supplementary Fig. 3b). As expected, silencing by pro-siRNA was Dicer-independent because *EGFPFL* pro-siRNA still functioned in Dicer-deficient HCT116 cells<sup>19</sup> and recombinant Dicer protein did not further process pro-siRNAs *in vitro* (Supplementary Fig. 4).

To test the effectiveness of pro-siRNA-mediated knockdown of endogenous and viral genes, we used convenient restriction sites to clone and express hairpins from the coding regions of *LMNA* (which encodes two splice variant products, lamin A and lamin C), *PLK1*, *TP53* and HIV *vif* (viral infectivity factor) and *gag* (capsid antigen). These hairpins contained 200–579 nt of each sense and antisense sequence (523 nt for *LMNA*, 299 nt for *PLK1*, 300 nt for *TP53*, 579 nt for *vif*, 200 and 500 nt for *gag*). The HPLC-purified pro-siRNAs for each gene contained a few different sized species that migrated close to the 21 nt marker on both native and denaturing polyacrylamide gels (Fig. 2f). For *LMNA* and *PLK1* pro-siRNAs, a minor RNA band migrated at ~25 nt. We transfected HeLa-dIEGFP and HCT116 with pro-siRNAs and commercially available chemically synthesized siRNAs (*LMNA* and *TP53* siRNAs were from a single sequence; *PLK1* siRNAs were a pool of 4 siRNAs and were chemically modified by proprietary methods for enhanced stability and reduced off-target effects<sup>20</sup>). The extent of gene knockdown was similar when siRNAs and pro-siRNAs were

transfected at 4 nM (Fig. 3a). To more closely evaluate the potency of pro-siRNAs, we performed dose response experiments comparing transfection of pro-siRNAs (0.2, 2, 20 nM) targeting *LMNA*, *TP53* and *PLK1* with five commercial siRNAs for each gene (four siRNAs from Dharmacon, of which the *PLK1* siRNAs were chemically modified<sup>20</sup>) (Supplementary Fig. 5). The potency of the commercial siRNAs varied, as best evaluated at the lowest concentration. The pro-siRNAs, whose sequences were not optimized, achieved similar gene knockdown as the commercially optimized siRNAs. At a concentration of 2 nM, each pro-siRNA achieved knockdown of ~90%. Because siRNA design algorithms are imperfect, identifying potent siRNAs often requires testing several sequences, which can be time consuming and costly. pro-siRNAs might circumvent the need to test multiple sequences to identify a single potent siRNA.

To examine potential toxicity of pro-siRNAs, we analyzed the growth of HeLa-d1GFP and HCT116 cells after they were transfected with either a negative control siRNA or *EGFP* pro-siRNA (Fig. 3b). Their growth curves were not significantly different. To compare the effectiveness of gene knockdown by pro-siRNAs and siRNAs, we examined cell proliferation after knocking down *PLK1*, which kills dividing cells<sup>21</sup>. *PLK1* siRNAs and pro-siRNAs markedly reduced viability with indistinguishable kinetics (Fig. 3b).

We next used pro-siRNAs to knockdown the HIV gene *vif*, which targets the host restriction factor APOBEC3G for ubiquitylation and degradation, thereby preventing APOBEC3G packaging into virions. *Vif* is therefore dispensable for the initial round of HIV replication but essential for spread of the infection to new cells. We compared the efficacy of *vif* pro-siRNAs with two validated chemically synthesized siRNAs<sup>22, 23</sup>. Neither siRNAs nor pro-siRNAs targeting *vif* altered the percentage of initially infected HeLa-CD4 cells (data not shown), but both suppressed *vif* gene expression and inhibited subsequent rounds of infection, as assessed in the TZM-bl luciferase reporter cell line (Fig. 3d). *vif* pro-siRNAs were more potent than chemically synthesized siRNAs.

One major obstacle to using RNAi to suppress HIV or other viruses is virus sequence diversity. Because pro-siRNAs target many sequences within a gene (see below), compared to chemically synthesized siRNAs, pro-siRNAs directed against a viral gene might have broader activity against diverse viral strains and might be less likely to incite generation of siRNA-resistant mutants. We previously tried unsuccessfully to identify an siRNA against HIV-1 clade B *gag* that could also inhibit viral isolates from other clades<sup>22</sup>. To investigate whether pro-siRNAs might have broader activity, we used hairpins encoding 200 and 500 nt from the *gag* coding region of clade B HIV-III<sub>B</sub> virus. Compared to a chemically synthesized *gag* siRNA, the *gagB200* and *gagB500* pro-siRNAs more potently suppressed HIV-III<sub>B</sub> (Fig. 2d). More importantly, both *gag* pro-siRNAs, but not the chemically synthesized *gag* siRNA, knocked down expression of *gag* mRNA and inhibited *in vitro* spread of viruses of different clades (UG29, clade A; IN22, clade C). However, not surprisingly, the *gag* pro-siRNAs more efficiently inhibited the III<sub>B</sub> virus than the UG29 or IN22 viruses. These data suggest that pro-siRNAs could be particularly useful for targeting heterogeneous and rapidly evolving viral genes.

Because mammalian cells are sensitive to bacterial endotoxin, which elicits production of inflammatory mediators by stimulating Toll-like receptor 4, we assessed whether purified pro-siRNAs are contaminated with endotoxin. Although SDS-eluted pro-siRNAs contained significant amounts of endotoxin as measured by Limulus amoebocyte lysate (LAL) assay, HPLC-purified pro-siRNAs, even at concentrations as high as 320 nM, contained no detectable endotoxin (Supplementary Table 1). We also tested for endotoxin contamination by measuring expression of proinflammatory cytokine and interferon-stimulated genes in monocyte-derived human macrophages incubated with (Supplementary Fig. 6a) or transfected with (Supplementary Fig. 6b) HPLC-purified pro-siRNAs. Neither pro-inflammatory nor interferon-stimulated genes were induced by pro-siRNAs in these highly endotoxin-sensitive cells.

To ascertain the sequence composition of pro-siRNAs, we cloned and deep sequenced pro-siRNAs using a method established for eukaryotic siRNAs (sequencing summary in Supplementary Table 2). Most reads were 20–22 nt in length (Fig. 4a and Supplementary Fig. 7). The majority of reads (on average ~75%) aligned to the target gene sequence, plasmid backbone or the *E. coli* genome and, consistent with the efficiency of target gene knockdown, the majority of aligned sequences (82–99%) originated from the target gene sequence (Fig. 4b). Reads spanned the entire target gene sequence, but some reads concentrated at specific sites ('hot spots') (Fig. 4c, Supplementary Fig. 7 and 8). We detected some sequence strand bias for most of the hot spots (Supplementary Fig. 8a). Because our data (Fig. 2c and f) strongly suggested that pro-siRNAs are double stranded, we suspected that strand bias may have been due to differences in ligation efficiency during cloning, a well-known problem<sup>24</sup>, rather than to the presence of many single-stranded RNAs. To evaluate this, we designed forward and reverse DNA oligonucleotide probes (26–27 nt) aligning to three *EGFPFL* pro-siRNA hot spots and performed solution hybridization and native gel electrophoresis (Supplementary Table 3 and Supplementary Fig. 8b). The relative intensity of hybridized bands was approximately equal for sense and antisense probes for each hot spot and loosely correlated with the number of reads from each hot spot (Supplementary Fig. 8c–e). Thus, the strand bias in the deep sequencing data likely reflects ligation bias introduced during cloning.

To further investigate the hot spot pattern, we compared two independent preparations of *EGFPFL* pro-siRNAs cloned using different sets of adapters. Their potency, size profile and sequence content were similar, but not identical. The most abundant hot spots were consistent in the 2 samples, but the strand bias changed with the adapters, consistent with cloning bias (Supplementary Fig. 9a–d). Hot spots might be due to intrinsic sequence preferences for RNase III cleavage, or to differences in RNA stability or RNA binding to p19 after cleavage. To determine whether 'hot spots' are determined by sequence differences at or close to the hot spot, we constructed hairpins of equal sizes from the 5' and 3' ends of the full length *EGFP* sequence. The pro-siRNAs generated from the two half *EGFP* hairpins contained hot spots mostly identical to those generated from the full length *EGFPFL* hairpin (Supplementary Fig. 9e). Thus hot spots seem to be determined by local sequence differences. However a basic bioinformatic analysis searching for preferred sequence motifs or bases in the hot spots was inconclusive (data not shown). *E. coli* RNase III might process

dsRNA into siRNA-sized small RNAs *in vivo* through a mechanism that differs from Dicer<sup>25</sup>, whose cleavage of a long dsRNA results in phased and evenly distributed sequences along a target gene.

Because pro-siRNAs contained non-targeting sequences derived from the plasmid or *E. coli* genome, we were concerned about possible off-target effects<sup>26</sup>. To evaluate off-target effects, we compared by RNA deep sequencing the RNA expression profile of HeLa-d1EGFP cells transfected with 4 nM of negative control siRNA, chemically synthesized *EGFP* siRNA or *EGFPFL* or *EGFP100* pro-siRNAs (sequencing summary in Supplementary Table 2). We used Tophat and Cufflinks to analyze the data and generated volcano plots of all annotated transcripts<sup>27</sup> (fold change versus *p* value, Fig. 4d). Compared to chemically synthesized *EGFP* siRNA, *EGFPFL* and *EGFP100* pro-siRNAs induced significant changes in expression of a smaller and larger number of genes, respectively (Fig. 4e, Supplementary Fig. 10a and Supplementary Table 4). *EGFPFL* pro-siRNAs also produced the fewest changes in long non-coding RNA levels (Supplementary Fig. 10b,c). *EGFP100* pro-siRNAs, made from a shorter hairpin (100 bp), contained a higher proportion of sequences mapping to plasmid and *E. coli* genomic sequences, compared to pro-siRNAs made from longer hairpins (200 to 720 bp, Fig. 4b). Thus, longer target gene hairpins will likely generate pro-siRNA preparations that will have fewer off-target effects. We also compared by microarray the gene expression profiles of cells transfected with *LMNA* chemically synthesized siRNAs and pro-siRNAs. Consistent with the *EGFPFL* data, *LMNA* pro-siRNAs made from a longer hairpin (523 bp) led to fewer significantly changed genes compared with the chemically synthesized *LMNA* siRNA (Fig. 4e,f and Supplementary Fig. 10d and Supplementary Table 4). For both *EGFP* and *LMNA*, the intended target was always the most down-regulated gene, and pro-siRNAs consistently induced greater suppression than the siRNA. The significantly changed genes in each of these experiments were not enriched for innate immune genes<sup>28</sup> (Supplementary Table 4), confirming that the pro-siRNAs did not stimulate an innate immune response. Thus pro-siRNAs offer highly specific knockdown that is at least as good as synthetic siRNAs without the need to test multiple sequences.

Here we showed that bacteria can produce siRNAs that are not toxic and efficiently suppress expression of exogenous genes (*EGFP*), viral genes (*vif* and *gag*) and endogenous genes (*PLK1*, *TP53*, *LMNA*) in mammalian cells. Because pro-siRNAs are natural products of RNase III, they likely have favorable ends (e. g., 5'-phosphate, 3'-hydroxyl and 3' overhangs) for efficient loading by Ago into the RISC, and they do not activate cytosolic innate immune RNA sensors. Although we used one plasmid to express p19 and the target gene hairpin, it is possible to use two separate plasmids to express p19 and the sense and antisense strands of the target sequence (Supplementary Fig. 11). Like chemically synthesized siRNAs, but unlike short hairpin RNA-mediated stable gene knockdown, pro-siRNA-mediated gene silencing is temporary.

Without much optimization we achieved an average yield of ~4 nmol (~42 µg) pro-siRNA per liter of *E. coli* culture, which, albeit modest, is more than enough for most laboratory experiments, given the activity of pro-siRNAs at low nanomolar concentrations. The engineered plasmid or *E. coli* genome could potentially be further optimized to maximize

yield and improve effectiveness and/or specificity. For example, we doubled the yield of *EGFPFL* pro-siRNA by overexpressing *E. coli* RNase III (Supplementary Fig. 12).

Previous studies described use of *E. coli* to induce RNAi. For example, RNase III-deficient *E. coli* expressing dsRNAs can be fed to *C. elegans* to knockdown expression of worm genes<sup>17</sup>, and bacteria-derived dsRNAs can be applied to plants to induce specific gene knockdown<sup>29</sup>. In worms and plants, unlike mammalian cells, gene silencing is enhanced by RNA-dependent RNA polymerases that amplify small amounts of RNA. More recently, genetically engineered *E. coli* engineered to express an invasin (to induce bacterial uptake) and listeriolysin (to allow bacterial RNAs to escape from phagolysosomes), delivered dsRNAs into the cytoplasm of human cells through “trans-kingdom RNAi” technology<sup>30,31</sup>. In all these scenarios, target gene silencing requires host cell processing of long siRNA precursors by Dicer.

pro-siRNAs could become a valuable addition to existing RNAi techniques for both research and therapeutics. Use of pro-siRNAs would eliminate the need to purchase and test multiple individual chemically synthesized siRNAs. When generated from longer hairpins, pro-siRNA preparations containing multiple sequences might trigger fewer off-target effects than individual siRNAs and, in the cases of virus infection or cancer, might more effectively prevent target gene escape by mutation. The method we developed for producing pro-siRNAs was adapted from well-established techniques for making recombinant proteins from *E. coli* and could easily be adopted and scaled-up in an industrial setting. In addition, mammalian cDNA libraries might be used to generate pro-siRNA libraries for siRNA screening purposes. That said, chemical synthesis provides the opportunity for chemical modifications to increase potency, enhance stability and reduce off-target effects or to couple fluorophores or targeting moieties. More work is needed to determine if such modifications might also be possible for pro-siRNAs; this might be achieved by adding modified ribonucleotides to bacterial cultures during IPTG induction or by performing the same coupling reactions with purified pro-siRNAs as are used to modify chemically synthesized siRNAs.

A recent study found that yeast Ago protein expressed in *E. coli* binds small RNAs that are predominantly a mixture of sense and antisense strands derived from sequences of the expression plasmid<sup>32</sup>. That study confirms our finding that siRNA-like small RNAs are generated in bacteria and can be stabilized by binding to proteins like Ago or p19. Future studies should evaluate whether siRNAs generated in bacteria might have functional significance, for example in bacterial defense against foreign genetic elements and pathogens or in regulation of endogenous bacterial genes.

## Online Methods

### Bacterial strains and culture conditions

All *E. coli* strains used in this study are listed in Supplementary Table 5. *E. coli* strain DH5 $\alpha$  was used for cloning and for initial characterization of the siRNA-like RNA species. For recombinant protein expression and pro-siRNA production, we used T7 Express Iq (NEB), a BL21-derived *E. coli* strain. We utilized two mutants of RNase III, *rnc-14::DTn10* (Tet<sup>R</sup>)

and *Drnc-38* (Kan<sup>R</sup>). These were moved by P1 transduction from parent strains HT115(DE3)<sup>17</sup> and SK7622<sup>18</sup> into *E. coli* strain MG1655 *lacZYA* (also referred as MG1655 *lac*). All *E. coli* strains were cultured in LB broth, Lennox (BD) at 37°C with shaking at 250 rpm and antibiotics when required were used at the following concentrations; carbenicillin (100 µg/ml), kanamycin (50 µg/ml), spectinomycin (50 µg/ml), tetracycline (12.5 µg/ml).

*Listeria monocytogenes* strain 10403S was cultured in brain-heart infusion medium (BD Biosciences) at 30°C. Transformation of bacterial cells was performed as previously described<sup>34</sup>.

## Genes and plasmids

The p19 gene used in this study was cloned from *Tomato bushy stunt virus* (gift of James Carrington, Donald Danforth Plant Science Center). All plasmids are listed in Supplementary Table 6 and will be made available through Addgene. To produce p19 in *E. coli*, we used pcDNA3.1+ (Invitrogen) to express the p19 protein with a C-terminal FLAG tag (pcDNA3.1-p19-FLAG) or an N-terminal His tag (pcDNA3.1-His-p19). Plasmid pcDNA3.1-TREX1-FLAG encodes a C-terminal FLAG-tagged TREX1 protein. To express p19 in *L. monocytogenes*, pLIV-1-His-p19 plasmid was used, which encodes for p19 with an N-terminal His tag in the pLIV-1 plasmid (gift of Darren Higgins, Harvard Medical School). *E. coli* RNase III with an N-terminal FLAG was cloned into pcDNA3.1+ and pCDF-1b (Novagen) plasmids.

We used two strategies for pro-siRNA production in *E. coli*. In one approach p19-His was fused to GST in pGEX-4T-1 (to express GST-p19-His fusion protein). On the same plasmid we cloned a hairpin RNA expressing cassette consisting of an inverted repeat separated by a 32 bp linker downstream of a T7 promoter (Fig. 2a). The hairpin RNA sequences were: *EGFPFL*, the entire 720-bp *EGFP* coding sequence (from pEGFP-N1, Clontech); *EGFP100*, 100 bp from nt 219 to 318; *EGFP Hotspot-1* 360 bp from nt 1 to 360; *EGFP Hotspot-2* 360 bp from nt 361 to 720; *LMNA* (NM\_005572.3), 523 bp from nt 267 to 789; *TP53* (NM\_000546.5), 301 bp from nt 376 to 676; *PLK1* (NM\_005030.3), 299 bp from nt 92 to 390; *vif* (K03455), the entire 579-bp; *gag* (K03455), *gagB200*: 200 bp from nt 1183 to 1382, *gagB500*: 500 bp from nt 1004 to 1503. (Genbank entries listed; numbers refer to position with respect to the translation start site).

In another approach we used two compatible plasmids for pro-siRNA production. The GST-p19-His protein was cloned under the control of the T7 promoter in pRSF-1b (Novagen) or pCDF-1b to generate pRSF-GST-p19-His and pCDF-GST-p19-His. The second plasmid is a L4440 plasmid encoding the entire *EGFP* coding sequence (L4440-EGFP).

All cloning was performed using PCR and standard techniques. All primers (with information for restriction enzyme sites) are listed in Supplementary Table 7.

## Cells

HeLa-d1EGFP, HCT116, HCT116 *Dicer*<sup>-/-</sup>, HeLa-CD4, TZM-bl, U87.CD4.CXCR4 and U87.CD4.CCR5<sup>33</sup> were cultured in DMEM medium (Invitrogen) supplemented with 10%



heat-inactivated fetal bovine serum (FBS). ACH2 cells (human leukemia T cell line CEM latently infected with HIV-1) were cultured in RPMI medium (Invitrogen) supplemented with 10% heat-inactivated FBS. For assays using primary monocyte-derived human macrophages, monocytes were isolated from the blood of a healthy donor by Ficoll-Paque Plus (GE Healthcare) density separation. Human samples were obtained with approval by the Boston Children's Hospital Investigation Review Board. Monocytes were plated on PRIMARIA plates (FALCON) in RPMI medium (Invitrogen) supplemented with 10% heat-inactivated human serum and adherent cells were cultured for 5 d to allow differentiation into macrophages.

### RNA isolation and qRT-PCR

Total RNA was isolated from 3 ml of *E. coli* stationary phase culture with 1 ml Trizol reagent (Invitrogen) following the manufacturer's protocol. RNA from human cells was collected in Trizol and extracted according to the manufacturer's protocol. Total RNA (1 µg) was converted to cDNA using SuperScript III Reverse Transcriptase (Invitrogen). For qRT-PCR, 10 µl reaction, containing SsoFast EvaGreen mastermix (Bio-Rad), appropriate primers (Supplementary Table 5), and template cDNAs made from 10 ng RNA, was amplified on a Bio-Rad CFX 96 Thermal Cycler. All qRT-PCR data were normalized to the human *GAPDH* gene. qRT-PCR primers for human genes (Supplementary Table 7) were selected from PrimerBank (<http://pga.mgh.harvard.edu/primerbank/>).

### siRNA isolation from total RNA using p19 magnetic beads

p19 magnetic beads were prepared at NEB as previously described<sup>13</sup>. To pull down siRNAs, 50 µg of total RNA (isolated from human or *E. coli* cells) was used following the manufacturer's protocol<sup>13</sup>.

### His-tag purification of GST-p19-His and bound pro-siRNA

GST-p19-His was purified as follows. A fresh single transformant of T7 Express Iq containing pGEX-4T-1-p19-T7 was used to inoculate 300 ml LB medium in a 1.5 L flask. When the OD<sub>600</sub> reached 0.3-0.6, protein and pro-siRNA expression were induced by adding 0.5 mM IPTG for 1 hr. Cells were centrifuged and lysed in 10 ml lysis buffer (50 mM Phosphate buffer pH 7.0, 300 mM NaCl, 10 mM imidazole, 1% Triton X-100, 1 mg/ml lysozyme) at 4°C for ~30 min followed by sonication (Misonix S-4000) until the lysate was non-viscous. Following centrifugation the lysate was incubated with rotation with 1 ml Ni-NTA resin (Thermo Scientific) overnight at 4°C. The resin was washed with lysis buffer 4 times, each time for 10 min at 4°C with rotation. Bound GST-p19-His was eluted in lysis buffer containing 300 mM imidazole at room temperature.

To purify p19-bound pro-siRNA the procedure was as above until the final elution step when 500 µl 0.5% SDS was added for 10 min at room temperature with rotation. This step was repeated and both SDS eluates were combined and passed through a 0.22 µm centrifuge filter (Corning) before HPLC purification on a Bio WAX NP5 anion exchange column (Agilent Technologies). The HPLC buffers were: Buffer A, 25 mM Tris-HCl, 2 mM EDTA; Buffer B, 25mM Tris-HCl, 2 mM EDTA, 5 M NaCl. HPLC was initiated with a flow rate of 1 ml/min at 25°C. Elution was performed using a linear gradient of 0–10% Buffer B over 4

min, followed by 10% Buffer B for 6 min, and a second linear gradient of 10–25% Buffer B over 15 min at a reduced flow rate of 0.5 ml/min. pro-siRNA eluted in the second gradient was collected by isopropanol precipitation.

### **Polyacrylamide gel electrophoresis (PAGE) of RNA**

For denaturing electrophoresis of RNA, mini-sized pre-cast 15% polyacrylamide TBE-Urea gels (Invitrogen) were used. RNA samples were heated to 95°C for 5 min in Gel Loading Buffer II (Ambion) and then immediately placed on ice until gel loading. Electrophoresis was performed in a 70°C water bath (to ensure complete denaturation of siRNA) and gels were stained with SYBR Gold (Invitrogen). For analysis of *E. coli* total RNA, 20 µg samples of Trizol-isolated RNA were loaded. RNA size standards (miRNA marker, siRNA marker and Low Range ssRNA Ladder) were from NEB.

For native electrophoresis of RNA, mini-sized homemade 10–20% polyacrylamide TBE gels were used with the Bio-Rad Mini-PROTEAN Tetra Cell. RNA samples were prepared in Gel Loading Buffer II (Ambion) without heat denaturation and electrophoresis was performed at room temperature.

### **Nuclease sensitivity assay**

The nucleases tested were: RNase A, RNase T1, and Turbo DNase (all from Ambion), Xrn1, exonuclease T, and exonuclease I (all from NEB). For each assay, 200 ng of an unmodified synthetic negative control siRNA (GenePharma) and *vif* pro-siRNA were used and assays were incubated in a 20 µl reaction volume using standard amounts of enzymes at 37°C for 1 hr. Treated RNAs were purified by phenol/chloroform extraction followed by isopropanol precipitation.

### **Tests for endotoxin activity and immune activation of primary human monocyte-derived macrophages**

RNA samples diluted in ddH<sub>2</sub>O to the indicated concentration were analyzed by the single vial Gel Clot LAL assay (detection limit 0.25 EU/ml, Lonza) following the manufacturer's protocol. Lipopolysaccharide (LPS) from *E. coli* O111:B4 (Sigma-Aldrich) was used as a positive control.

To test for cytokine gene activation, monocyte-derived macrophages plated in 24 well plates ( $1 \times 10^5$  cells/well) were incubated with medium containing RNA or LPS at the indicated concentration for 4 hr before harvesting RNA. siRNAs and pro-siRNAs (20 nM) were also transfected into cells using Lipofectamine 2000 (Invitrogen) and total RNA was harvested 24 hr after transfection.

### **5<sup>32</sup>P labeling of RNA**

RNA samples were dephosphorylated by Antarctic Phosphatase (NEB) for 30 min at 37°C in the presence of Murine RNase Inhibitor (NEB). The Antarctic Phosphatase was deactivated by incubation at 65°C for 5 min and the RNA was end-labeled with  $\gamma$ -<sup>32</sup>P ATP (PerkinElmer) and T4 Polynucleotide Kinase (NEB). Gels were exposed using a phosphorimager screen and visualized using a FLA-9000 Image Scanner (Fujifilm).

### Small RNA northern blot

Northern blot for small RNAs was performed as previously described<sup>35</sup>. The *EGFP* specific sense probe was a <sup>32</sup>P-UTP-internally labeled RNA prepared by *in vitro* transcription using T7 RNA polymerase (NEB) and a PCR-generated DNA template of the full-length *EGFP* gene that incorporated a T7 promoter.

### siRNA transfection for testing RNA silencing efficiency

All siRNA transfections were performed using Lipofectamine 2000 following the manufacturer's protocol. Briefly, cells were plated in 24 well plates ( $1 \times 10^5$  per well) and the transfection complex (containing 1.0 ml Lipofectamine 2000 and siRNAs) was added directly to the medium. RNA and protein samples were isolated from cells 24 hr post-transfection. For the PLK1 cell killing experiment, cells were counted using a TC-10 automatic cell counter (Bio-Rad). The following siRNAs were used: ON-TARGETplus Non-targeting siRNA #4 (D-001810-04-05, Dharmacon), siGENOME Lamin A/C Control siRNA (D-001050-01-20, Dharmacon), Set of 4: siGENOME LMNA siRNAs (MQ-004978-01-0002, Dharmacon), ON-TARGETplus SMARTpool - Human PLK1 (L-003290-00-0005, Dharmacon), Set of 4 Upgrade: ON-TARGETplus PLK1 siRNA (LU-003290-00-0002, Dharmacon), Set of 4: siGENOME TP53 siRNA (MQ-003329-03-0002, Dharmacon), Negative control siRNA (NC siRNA, B01001, GenePharma), Positive control siRNA TP53 (B03001, GenePharma), custom *EGFP* siRNA (sense, GGCUACGUCCAGGAGCGCACC; antisense, UGCGCUCCUGGACGUAGCCUU), custom *vif* siRNA-1<sup>22</sup> (sense, GUUCAGAAGUACACAUCCCT; antisense, GGGAUGUGUACUUCUGAACTT) and custom siRNA-2<sup>23</sup> (sense, CAGAUGGCAGGUGAUGAUUGT; antisense, AAUCAGCACCUGCCAUCUGTT), custom *gag* siRNA<sup>24</sup>: (sense, GAUUGUACUGAGAGACAGGCU; antisense, CCUGUCUCUCAGUACAAUCUU).

### RISC Immunoprecipitation

Cells ( $3 \times 10^6$ ) were transfected with 4 nM NC siRNA or *EGFPFL* pro-siRNAs. After 24 hr cells were scraped from the plate in 2 ml lysis buffer (150 mM KCl, 25 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.5 mM DTT, 1% NP-40 and Roche Complete Protease Inhibitor Cocktail). Cells were then mechanically disrupted for 1 min using a micro-MiniBeadbeater (BioSpec). The cell lysate was incubated at 4°C with rotation for 1 hr to ensure complete lysis. IP was performed by adding anti-Ago (2A8) antibody (Millipore, MABE56) or mouse total IgG (Jackson Labs) at 1:100 dilution together with 30  $\mu$ l protein G Dynabeads (Invitrogen) and samples were rotated at 4°C overnight. After washing 4 times in lysis buffer, precipitated RNAs were isolated using Trizol reagent from 90% of the reaction mix, while 10% was saved for immunoblot input.

### Western Immunoblot

Protein samples were prepared by heating cells to 95°C for 5 min in  $1 \times$  SDS loading buffer before SDS-PAGE. Immunoblot was performed using SNAP i.d. Protein Detection System (Millipore) following the manufacturer's protocol. Antibodies and their dilutions were: anti-FLAG (M2) 1:1,000 (Sigma-Aldrich, F1804), anti-His tag 1:500 (Covance, MMS-156P),

anti-PLK1 1:100 (Santa Cruz, sc-17783), anti-LaminA/C 1:1,000 (Santa Cruz, sc-7292), anti-p53 (DO-1) 1:500, (Santa Cruz, sc-126), anti-beta-Tubulin 1:10,000 (Sigma-Aldrich, T5168), anti-Ago (2A8) 1:1,000 (Millipore, MABE56). Horseradish peroxidase conjugated anti-mouse or anti-rabbit IgG secondary antibodies were used at 1:5,000 dilution followed by incubating the membranes in SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

### **Solution hybridization and native gel electrophoresis assay**

DNA oligonucleotides purchased from IDT were PAGE purified. Purified DNA oligonucleotides (10 pmol) were end-labeled with  $\gamma$ -<sup>32</sup>P ATP by T4 Polynucleotide Kinase (NEB) and 2 pmol was then mixed with 5 ng of pro-siRNAs in buffer containing 20 mM Tris-HCl pH 7.9, 100 mM NaCl and 2 mM EDTA. Samples were heated to 80°C for 10 min and allowed to cool to room temperature. A fraction of the sample was separated on a native 15% polyacrylamide gel. The gel was directly exposed to a phosphorimager screen. Multi-gauge software (Fujifilm) was used for image quantification.

### **siRNA library preparation, deep sequencing, and data analysis**

siRNAs were cloned according to the Illumina small RNA sample preparation guide v1.5 with the following exceptions. Custom 5' RNA ligation adapters were synthesized with a 4 nt nucleotide barcode sequence (Supplementary Table 8). Small RNA libraries were pooled and sequenced on one sequencing lane of an Illumina GAII sequencer (Genome Technology Core, Whitehead Institute or NEB). Novocraft software ([www.novocraft.com](http://www.novocraft.com)) was used for sequence alignment. Reference genome was *E. coli* K12 substr. MG1655. We wrote Perl software scripts for data analysis. All small RNA deep sequencing data are available in the BioProject database, NCBI (PRJNA188722, <http://www.ncbi.nlm.nih.gov/>).

### **mRNA profiling by microarray and deep sequencing**

siRNAs and pro-siRNAs (4 nM) were transfected into HeLa-d1EGFP cells and RNA was isolated 24 hr post-transfection. Non-targeting siRNA #4 (Dharmacon) was used as negative control siRNA. Data from biological duplicates were analyzed at the Microarray Core, Dana Farber Cancer Institute for microarray analysis using GeneChip 1.0 ST (Affymetrix). Microarray data was analyzed using dChip software and *p* values of gene expression changes were calculated using paired T-test method<sup>36</sup>. Microarray data can be retrieved from the Gene Expression Omnibus database at NCBI (GSE44105, <http://www.ncbi.nlm.nih.gov/>).

For RNA deep sequencing, a Ribo-Zero rRNA Removal Kit (Epicentre) was used to remove large ribosomal RNAs from total RNA following the manufacturer's protocol. For each condition, two libraries were constructed and sequenced independently from duplicate samples. rRNA-depleted RNA (from 500 ng total RNA) was used to construct a deep sequencing library using NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530) according to the manufacturer's protocol. Illumina GAII was used for sequencing (NEB). Tophat and Cufflinks software suites were used to analyze the RNA deep sequencing data from biological duplicates according to<sup>27</sup>. Reference genome was Human genome GRCh37/hg19 and annotations of lincRNA transcripts were downloaded from UCSD genome

browser. Deep sequencing data can be retrieved from the Gene Expression Omnibus database at NCBI (ID number pending, <http://www.ncbi.nlm.nih.gov/>).

### Flow cytometry

For EGFP, cells were removed from plates by trypsin digestion and re-suspended in FACS buffer, DPBS (Invitrogen) containing 2% heat-inactivated FBS. Intracellular staining of p24 antigen was performed using an Intracellular Staining Kit (Invitrogen) according to the manufacturer's protocol and fluorescein-labeled p24 antibody (1:200, Beckman Coulter, cat#KC57-FITC). Fluorescence was analyzed on a FACSCalibur (BD) using FlowJo software (Tree Star).

### HIV infection and TZM-bl assay

HeLa-CD4 cells were transfected with 4 nM siRNA or pro-siRNA in 24 well plates ( $1 \times 10^5$  cells/well). Cells were infected 12 hr post-transfection with HIV<sub>IIIIB</sub> (~400 ng/ml p24) and culture medium was changed 12 hr post-infection. For HIV<sub>UG29</sub> U87.CD4.CXCR4 cells were used and for HIV<sub>IN22</sub> U87.CD4.CCR5 cells were used. Culture medium was collected for TZM-bl assay and RNA was extracted for qRT-PCR 36 hr post-infection. TZM-bl cells, plated in 24 well plates ( $1 \times 10^5$  cells/well) 12 hr before, were analyzed 24 hr after adding culture supernatants by luciferase assay performed using a Luciferase Assay System kit (Promega) following the manufacturer's protocol.

### RNase A digestion assay for *E. coli* total RNA

~2 ug of total *E. coli* RNA were incubated with 1.0 unit of RNase A for 15 min at 37°C in  $1 \times$  DNase I reaction buffer (NBE) supplemented with 400 mM NaCl. The resulting products were analyzed on a 0.8% agarose gel containing EtBr.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

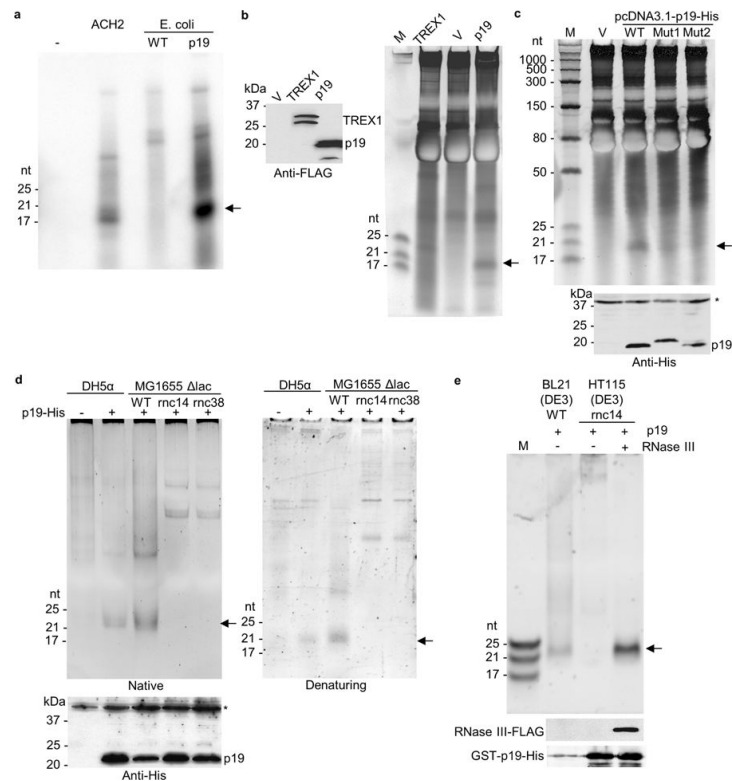
The authors thank James Carrington (Donald Danforth Plant Science Center) for the p19 clone, Chi Zhang and Gary Ruvkun (Massachusetts General Hospital) for L4440 plasmid and HT115(DE3) strain, Sidney Kushner (University of Georgia) for SK7622 strain, Darren Higgins (Harvard Medical School) for pLIV-1 plasmid, Shahin Ranjbar (Boston Children's Hospital) for HIV strains and cell lines. We thank Laurie M. Mazzola, Joanna M. Bybee and Daniela B. Munafo from NEB for assistance with RNA deep sequencing and Zander Ansara for technical help. We thank Ann Hochschild (Harvard Medical School) for suggestions and critical reading of the manuscript and Lieberman Lab members for technical assistance, helpful discussions and comments on the manuscript. This work was supported by NIH grant AI087431 (J.L.) and a GSK-IDI Alliance fellowship (L.H.).

### References

1. Elbashir SM, et al. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*. 2001; 411:494–498. [PubMed: 11373684]
2. Caplen NJ, Parrish S, Imani F, Fire A, Morgan RA. Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. *Proc Natl Acad Sci U S A*. 2001; 98:9742–9747. [PubMed: 11481446]

3. Rettig GR, Behlke MA. Progress toward in vivo use of siRNAs-II. *Mol Ther.* 2012; 20:483–512. [PubMed: 22186795]
4. Voinnet O, Pinto YM, Baulcombe DC. Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proc Natl Acad Sci U S A.* 1999; 96:14147–14152. [PubMed: 10570213]
5. Silhavy D, et al. A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *Embo J.* 2002; 21:3070–3080. [PubMed: 12065420]
6. Fire A, et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature.* 1998; 391:806–811. [PubMed: 9486653]
7. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science.* 1999; 286:950–952. [PubMed: 10542148]
8. Myers JW, Jones JT, Meyer T, Ferrell JE Jr. Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nat Biotechnol.* 2003; 21:324–328. [PubMed: 12592410]
9. Yang D, et al. Short RNA duplexes produced by hydrolysis with *Escherichia coli* RNase III mediate effective RNA interference in mammalian cells. *Proc Natl Acad Sci U S A.* 2002; 99:9942–9947. [PubMed: 12096193]
10. Morlighem JE, Petit C, Tzertzinis G. Determination of silencing potency of synthetic and RNase III-generated siRNA using a secreted luciferase assay. *Biotechniques.* 2007; 42:599–605. [PubMed: 17515198]
11. Semizarov D, et al. Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci U S A.* 2003; 100:6347–6352. [PubMed: 12746500]
12. Vargason JM, Szittyá G, Burgyan J, Hall TM. Size selective recognition of siRNA by an RNA silencing suppressor. *Cell.* 2003; 115:799–811. [PubMed: 14697199]
13. Jin J, Cid M, Poole CB, McReynolds LA. Protein mediated miRNA detection and siRNA enrichment using p19. *Biotechniques.* 2010; 48:xvii–xxiii. [PubMed: 20569217]
14. Davis MG, Huang ES. Transfer and expression of plasmids containing human cytomegalovirus immediate-early gene 1 promoter-enhancer sequences in eukaryotic and prokaryotic cells. *Biotechnol Appl Biochem.* 1988; 10:6–12. [PubMed: 2841949]
15. Chu M, Desvoyes B, Turina M, Noad R, Scholthof HB. Genetic dissection of tomato bushy stunt virus p19-protein-mediated host-dependent symptom induction and systemic invasion. *Virology.* 2000; 266:79–87. [PubMed: 10612662]
16. Knight SW, Bass BL. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science.* 2001; 293:2269–2271. [PubMed: 11486053]
17. Timmons L, Court DL, Fire A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene.* 2001; 263:103–112. [PubMed: 11223248]
18. Babitzke P, Granger L, Olszewski J, Kushner SR. Analysis of mRNA decay and rRNA processing in *Escherichia coli* multiple mutants carrying a deletion in RNase III. *J Bacteriol.* 1993; 175:229–239. [PubMed: 8416898]
19. Cummins JM, et al. The colorectal microRNAome. *Proc Natl Acad Sci U S A.* 2006; 103:3687–3692. [PubMed: 16505370]
20. Jackson AL, et al. Position-specific chemical modification of siRNAs reduces “off-target” transcript silencing. *Rna.* 2006; 12:1197–1205. [PubMed: 16682562]
21. Spankuch B, et al. Cancer inhibition in nude mice after systemic application of U6 promoter-driven short hairpin RNAs against PLK1. *J Natl Cancer Inst.* 2004; 96:862–872. [PubMed: 15173270]
22. Lee SK, et al. Lentiviral delivery of short hairpin RNAs protects CD4 T cells from multiple clades and primary isolates of HIV. *Blood.* 2005; 106:818–826. [PubMed: 15831707]
23. Sugiyama R, Habu Y, Ohnari A, Miyano-Kurosaki N, Takaku H. RNA interference targeted to the conserved dimerization initiation site (DIS) of HIV-1 restricts virus escape mutation. *J Biochem.* 2009; 146:481–489. [PubMed: 19556221]
24. Jayaprakash AD, Jabado O, Brown BD, Sachidanandam R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* 2011; 39:e141. [PubMed: 21890899]

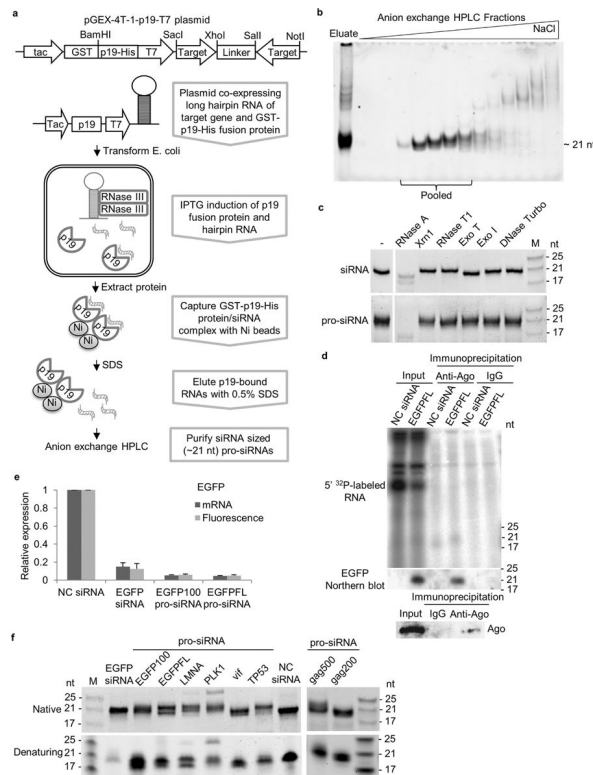
25. Weinberg DE, Nakanishi K, Patel DJ, Bartel DP. The inside-out mechanism of Dicers from budding yeasts. *Cell*. 2011; 146:262–276. [PubMed: 21784247]
26. Jackson AL, et al. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol*. 2003; 21:635–637. [PubMed: 12754523]
27. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7:562–578. [PubMed: 22383036]
28. Korb M, et al. The Innate Immune Database (IIDB). *BMC Immunol*. 2008; 9:7. [PubMed: 18321385]
29. Tenllado F, Martinez-Garcia B, Vargas M, Diaz-Ruiz JR. Crude extracts of bacterially expressed dsRNA can be used to protect plants against virus infections. *BMC Biotechnol*. 2003; 3:3. [PubMed: 12659646]
30. Zhao HF, et al. High-throughput screening of effective siRNAs from RNAi libraries delivered via bacterial invasion. *Nat Methods*. 2005; 2:967–973. [PubMed: 16299483]
31. Xiang S, Fruehauf J, Li CJ. Short hairpin RNA-expressing bacteria elicit RNA interference in mammals. *Nat Biotechnol*. 2006; 24:697–702. [PubMed: 16699500]
32. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. Structure of yeast Argonaute with guide RNA. *Nature*. 2012; 486:368–374. [PubMed: 22722195]
33. Princen K, Hatse S, Vermeire K, De Clercq E, Schols D. Establishment of a novel CCR5 and CXCR4 expressing CD4+ cell line which is highly sensitive to HIV and suitable for high-throughput evaluation of CCR5 and CXCR4 antagonists. *Retrovirology*. 2004; 1:2. [PubMed: 15169555]
34. Dancz CE, Haraga A, Portnoy DA, Higgins DE. Inducible control of virulence gene expression in *Listeria monocytogenes*: temporal requirement of listeriolysin O during intracellular infection. *J Bacteriol*. 2002; 184:5935–5945. [PubMed: 12374827]
35. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. *Nat Protoc*. 2008; 3:1077–1084. [PubMed: 18536652]
36. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*. 2001; 98:31–36. [PubMed: 11134512]



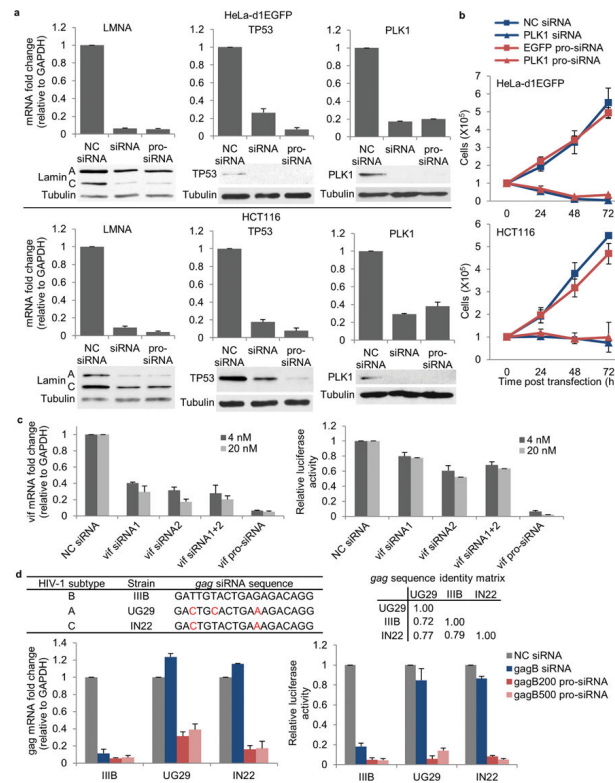
**Figure 1.**

Ectopic p19 expression captures small RNAs in *E. coli*. (a) p19-coupled magnetic beads were incubated with total RNA isolated from mammalian ACH2 cells, or from *E. coli* cells that were either wild-type (WT) or transformed with a pcDNA3.1-p19 expression plasmid. Captured RNAs were 5' <sup>32</sup>P-labeled, separated on a native polyacrylamide gel and detected by autoradiography. (b) FLAG-tagged p19 or TREX1, or empty vector (V), were expressed in *E. coli*. Left, anti-FLAG immunoblot. Right, total RNAs were separated on a denaturing polyacrylamide gel and stained with SYBR Gold. (c) Top, total RNAs purified from *E. coli* expressing an empty vector (V), or WT or mutant (Mut1<sup>12</sup>: W39G, W42G and Mut2<sup>15</sup>: K71A, R72G; mutants were defective in RNA binding) His-tagged p19 proteins were separated on a denaturing polyacrylamide gel and stained with SYBR Gold. Bottom, anti-His immunoblot. (d) p19-coupled magnetic beads were incubated with total RNA extracted from WT *E. coli* (DH5α or MG1655  $\Delta$ lac) or RNase III mutant strains (*rnc14* and *rnc38* in MG1655  $\Delta$ lac background) expressing or not His-tagged p19. p19-captured RNAs were separated on native (left) or denaturing (right) gels and stained with SYBR Gold. Bottom, anti-His immunoblot. The asterisk (\*) indicates equal loading of a background band. (e) p19-coupled magnetic beads were incubated with total RNA extracted from *E. coli* WT BL21(DE3) cells or *rnc14* mutant HT115(DE3) cells that were co-transfected with p19 and a vector encoding Flag-tagged *E. coli* RNase III (or empty vector). p19-bound RNAs were separated on a native polyacrylamide gel and stained with SYBR Gold. Bottom, anti-FLAG and anti-His immunoblots. Arrows indicate the ~21 nt small RNA band. M, markers. Data are representative of at least 2 independent experiments.

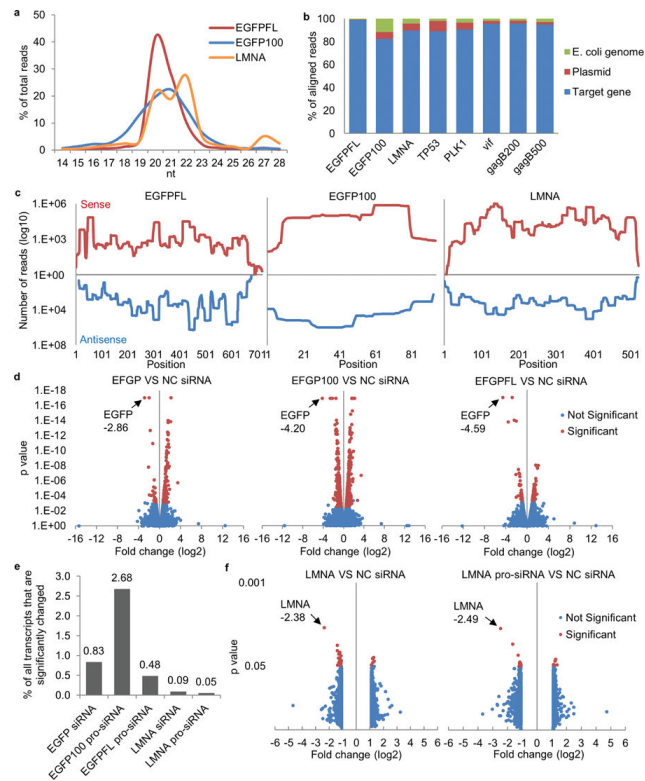




**Figure 2.** pro-siRNAs knockdown target gene expression. **(a)** Schematic of pGEX-4T-1-p19-T7 plasmid and the method to produce pro-siRNAs from *E. coli*. **(b)** Anion exchange HPLC fractions of SDS-eluted RNAs (isolated from *E. coli* engineered as in **(a)** to express pro-siRNAs) were separated on a native polyacrylamide gel and stained with SYBR Gold. HPLC fractions containing only ~21 nt small RNAs were combined (“Pooled”) for use in subsequent experiments. **(c)** Synthetic siRNAs (top) and HPLC purified pro-siRNAs (bottom) were incubated with indicated nucleases and separated on a native polyacrylamide gel stained with SYBR Gold. -, no nuclease. **(d)** Anti-Ago mAb 2A8 or mouse total IgG were used to immunoprecipitate RNAs from HeLa-d1EGFP cells transfected with negative control (NC) siRNA or *EGFPFL* pro-siRNAs. Immunoprecipitated RNAs were analyzed by 5' <sup>32</sup>P end-labeling (top) or by Northern blot using a probe complementary to the full length *EGFP* coding sequence (middle). Bottom, anti-Ago immunoblot. **(e)** HeLa-d1EGFP cells were transfected with 4 nM of synthetic *EGFP* siRNA or with pro-siRNAs. *EGFP* expression was measured by qRT-PCR (mRNA) and flow cytometry (fluorescence). Data are normalized to cells treated with negative control (NC) siRNA. mRNA level is relative to *GAPDH*. Data are mean±SD of 2 (qRT-PCR) and 3 (*EGFP* fluorescence) independent experiments. **(f)** HPLC-purified pro-siRNAs or synthetic siRNAs (*EGFP* and NC) were separated on native (top) or denaturing (bottom) polyacrylamide gels stained with SYBR Gold. M, markers. Data are representative of at least 2 independent experiments.



**Figure 3.** pro-siRNA-mediated knockdown of endogenous and viral gene expression in human cells. (a) Synthetic siRNAs or pro-siRNAs specific for the indicated target genes or negative control (NC) siRNA were transfected (4 nM) into HeLa-d1EGFP (top) or HCT116 (bottom) cells. Target gene expression was measured by qRT-PCR (bar graphs) or immunoblots (tubulin is loading control). (b) HeLa-d1EGFP (top) or HCT116 (bottom) cells were transfected with *PLK1* siRNA or pro-siRNA (4 nM) or NC siRNA or *EGFP* pro-siRNA as nontargeting controls, respectively. Cells were counted at indicated time points. (c) HeLa-CD4 cells were transfected with *vif* siRNAs (*vif* siRNA-1 and *vif* siRNA-2, alone or together) or *vif* pro-siRNA 24 hr before HIV-1 infection. Left, *vif* mRNA was measured by qRT-PCR 36 hr after infection. Right, infectivity of culture supernatants from transfected HeLa-CD4 cells was measured by HIV-1 *tat*-driven expression of a luciferase reporter gene in TZM-bl assay. (d) Top, sequence of the *gag* siRNA (complementary to HIV-1 clade B) used here; corresponding sequences in clade A UG29 and clade C IN22 strains are shown together (left) with the proportion of identical bases in the entire *gag* coding region for the 3 isolates tested (right). Bottom, HeLa-CD4 cells were infected with HIV-1 clade B (III<sub>B</sub>), U87.CD4.CXCR4<sup>33</sup> cells were infected with HIV-1 clade A (UG29) and U87.CD4.CCR5 cells<sup>33</sup> were infected with HIV-1 clade C (IN22). Cells were transfected with 20 nM of the indicated synthetic or pro-siRNA 12 hr before HIV-1 infection. Left, *gag* mRNA was measured by qRT-PCR 36 hr after infection. Right, infectivity of culture supernatants was measured by TZM-bl assay. Data are mean±SD of 3 (a–c) and 2 (d) independent experiments. mRNA expression and TZM-bl luciferase data are normalized to cells transfected with NC siRNA.



**Figure 4.** pro-siRNA sequences and assessment of off-target effects. **(a)** Length distribution of *EGFPFL*, *EGFP100* and *LMNA* pro-siRNAs assessed by deep sequencing. **(b)** Percentage of deep sequencing reads aligning to the target gene hairpin, the *E. coli* genome or the plasmid backbone. **(c)** pro-siRNA deep sequencing read alignment to indicated positions of the *EGFPFL*, *EGFP100* and *LMNA* target genes (sense and antisense). **(d)** Volcano plots of expression changes versus  $p$  value of all annotated transcripts detected by RNA deep sequencing in HeLa-d1EGFP cells transfected with *EGFP* synthetic siRNA or *EGFP100* or *EGFPFL* pro-siRNAs relative to expression in cells transfected with a negative control (NC) siRNA. Arrows indicate *EGFP* and the numbers indicate its fold change (log2). Cut-off for significance is  $q$  value < 0.05 (the default in Cufflinks). **(e)** Percent of significantly changed transcripts in (d and f). Numbers on top of bars indicate exact percentages. Significantly changed genes are listed in Supplementary Table 4. **(f)** Volcano plots of expression changes (1.2 fold less or more) versus  $p$  value detected by microarray in HeLa-d1EGFP cells transfected with *LMNA* synthetic siRNA or pro-siRNAs relative to expression in cells transfected with a NC siRNA. Arrows indicate *LMNA* and the numbers indicate its fold change (log2). Cut-off for significance is  $p$  < 0.05 (paired T-test by dChip).