Supporting Information

Tan et al. 10.1073/pnas.1119873109

SI Experimental Procedures

Vectors, Library Construction, and Reporter Cell Lines. For vectors, library construction, and reporter cell lines, 40,458 210-mer oligonucleotides were synthesized on a 55,000 features oligonucleotide array (Agilent Technology). Each 210-mer contains a 101-nt miR30-shRNA fragment, a 16-nt sequence containing EcoRI and MluI restriction enzyme sites for inserting Venus-coding sequence, a 50-nt sensor cassette that is the target site for the shRNA, and a 25-mer unique barcode for microarray hybridization followed by the 18-nt primer binding site. A two-step pooled cloning strategy was used (1). In the first step, the 210-mer was cloned into the pSENSOR vector; in the second step, the 3′miR30-PGK-Venus fragment was inserted between the shRNA and the sensor cassette. The ERC reporter cell line was as reported. Thirteen shRNAs that target different genes adapted from Fellmann et al. (1) were used in the sensor screen and are as follows:

Weak shRNAs: Braf3750, Cdkn2a154, Cdkn2a218, Cebpa577, Cebpa 1063;
Medium shRNAs: Pten1524, Pten 1688, Cdkn2a157, Braf 3826;
Strong shRNAs: Braf5053, Trp53 1224, Pten 1523, Pten 2049.

Sensor Ping-Pong Assay. The FACS sortings were performed on a BD FACSAria II sorting system (BD Biosciences). Four-hundred millions cells were selected were infected with the virus library to ensure 1,000-fold representation of the library with single-copy integration of each virus. After selection by G418 (05 mg/mL), three cycles of Ping-Pong sorting were performed. In the "Ping" step, after treating the cells with Doxycycline for 7 d, the cells were sorted to select for the low Venus population. In the "Pong" step, after 7 d of Doxycycline withdrawal, cells were selected for high Venus population. At least a 1,000-fold representation was maintained after each sort and similar number of cells were frozen for extraction of genomic DNA. Five strong hairpins (top 5) and five weak hairpins (bottom 5) were used to determine the gating of the FACS sorting (1).

Genomic DNA Extraction, Amplification of the Libraries. Genomic DNA was extracted from at least 40 million cells before sorts and after each sort. Standard genomic DNA extraction protocol using Phase Lock Tubes (15 mL Light, 5 Prime) was performed (2). The genomic DNA was used as template for PCR using forward primer: Venus2: (TCCTGCTGGAGTTCGTGA) and reverse primer: TNL2 (TACTCTATAGGCTTCAGCTGG) to amplify the region from the end of the Venus cassette to the 25-mer barcode, which includes the sensor sequence. The PCR is 800 μL in total (eight tubes of 100 μL each), which contained the following components: 60 μg gDNA template, 200 μM each of the four dNTPs, 400 nM for each PCR primer, 2% DMSO, 1× Hotstart Taq buffer, and 40 U Hotstart Taq (Takara). The PCR program includes: 95 °C 5 min, 36 cycles of 94 °C 35 s, 52 °C 35 s, 72oC 1 min, and a final step of 72 °C 10 min.

Deep Sequencing of the Libraries. For deep sequencing, a second round PCR was performed to tag standard Illumina adaptors to the PCR product from genomic DNA. Forward primer is: CAAGCAGAAGACGGCATACGA GAATTGTTCAATTGACGCGT, reverse primer: AATGATACGGCGACCACCGACGTACGCAATTGTCCGCGTCG. Accuprime pfx polymerase (Invitrogen) was used for the PCR; 250 ng of product from the first-round PCR was used as template. In addition, the 50 μL reaction contains 1 μL each of forward/reverse primers (100 μM), 4 μL dNTP mix (25 mM for each of the four dNTPs), 35 μL DMSO, 1 μL of Accuprime pfx (25 U/μL), 5 μL 10× Accuprime reaction buffer. Cycling parameters are: 95 °C 5 min, 26 cycles of 95 °C 15 s, 50 °C 30 s, 68 °C 20 s, and then 68 °C 2 min. Next, a mixture of 5 μL of 10× Accuprime buffer, 1 μL each of forward/reverse primers (100 μM), 4 μL dNTP mix (25 mM for each of the four dNTPs), 35 μL DMSO, 1 μL of Accuprime pfx (25 U/μL), and 34.5 μL of distilled water was added to the reaction. Then one cycle of PCR was performed: 95 °C 5 min, 95 °C 15 s, 50 °C 30 s, 68 °C 20 s, and then 68 °C 2 min. This additional cycle of PCR helps clearing up smearing in the PCR product. Data from this sequencing are shown in Dataset S1.

Microarray Hybridization. To label the PCR products from genomic DNA for microarray hybridization, 39.25 μL mixture of 2 μg of the PCR product and 2 μg/mL random nonamer primer (final concentration) was boiled for 5 min and chilled on ice for 2 min. The following reagents were then added to the mixture: 5 μL 10× NEB buffer 2, 1 μL of 10 mM dATP, dGTP, dTTP, 225 μL of Cy3 or Cy5 labeled-dCTP (Amersham), and 25 U Klenow (New England Biolabs). The reaction was allowed to proceed at 37 °C for 2 h and then stopped with 5 μL of 05 M EDTA pH 80 and cleaned using Microcon tubes (YM-30; Millipore). Cy3 dye was used to label PCR product of the genomic DNA of the presort cells, Cy5 dye was used to label the PCR product of genomic DNA of the cells after sorts 2, 4, and 6. The microarray was a high-density array custom ordered from Agilent that has four subarrays of 44,000 probes on each array. A hybridization mixture is consisting of 250 ng Cy5- and 250 ng Cy3-labeled probes, 55 μL 2× Agilent GEX hybridization buffer (Hi-RPM), 11 μL 10× Agilent blocking reagent, and water to a final volume of 110 μL 100 μL was added each subarray. The array was hybridized at 50 °C overnight, washed as per the manufacturer's wash protocol, and scanned using an Agilent microarray scanner. Data from these hybridizations are shown in Dataset S1.

HIV Infection Assay to Validate HIV-Targeting shRNAs. Individual shRNAs were cloned into murine stem cell virus (MSCV)-PM vector (2), which was packed into retroviral particles by cotransfecting 293 T cells with pVSV-G and pCG-*gag-pol* plasmids. The HeLa-CD4 cells (National Institutes of Health AIDS reagent program) were infected by the packed retrovirus at a multiplicity of infection (MOI) of ~0.1. After selection for puromycin resistance, 10^6 cells were seeded in each well of six-well plates and infected the next day by HIV NL43 virus (plasmid obtained from National Institutes of Health AIDS reagent program and amplified by transfecting 293T cells) at MOI = ~0.2. After 48 h, total RNA was extracted from cells using Qiagen RNeasy Plus kit. Reverse-transcription reaction was performed using the SuperScript III Reverse Transcriptase (Invitrogen) following the supplied protocol. Each 20-μL reaction contains 1 μg total RNA, 1 μL Oligo $dT_{12–18}$ (0.5 μg/μL), 1 μL of dNTP (10 mM each). Quantitative PCR reactions were performed on an Applied Biosystem-4500 Fast PCR machine using Platinum SYBR Green qPCR SuperMix-UDG kit from Invitrogen. Each reactions consists of 10-μL supermix reagent, 1 μL each of the forward/reverse primer (10 μM), 0.2–2 μL from the 20-μL reverse transcription reaction of 1 μg total RNA. Gene expression levels were normalized to the house-keeping gene *gapdh*. The primers used in the qPCR reactions are:

gap-pol: GGGGCAACTAAAGGAAGCTC/CCTCCAATTCCCCCTATCAT;

env: GAACGGATCCTTGGCACTTA/CGTCCCAGAAGTTCCACAAT;
tat: GACGGTACAGGCCAGACAAT/GATGCCCCAGACTGTGAGTT;
gapdh: AAGGTGAAGGTCGGAGTCAA/AATGAAGGGGTCATTGATGG.

The sequences of the guide strands of the shRNAs used the assay are [firefly luciferase (FF)]:

N1: CCTGCTATGTCACTTCCCCTTG;
N2: TCATTCTTGCATACTTTCCTGT;
N3: TGTGAAGCTTGCTCGGCTCTTA;
N4: ATGAAACAAACTTGGCAATGAA;
N5: TGACTTTGGGGATTGTAGGGAA;
N6: TGTATCATTATGGTAGCTGGAT;
N7: GTGCTGATACTTCTCCTTCACT;
N8: CTCCGCTTCTTCCTGCCATAGT;
N9: CTGTATCATCTGCTCCTGTATT;
FF: CCCGCCTGAAGTCTCTGATTAA.

Hepatitis C Virus Infection Assay to Validate Hepatitis C Virus-Targeting shRNAs. MSCV-PM vector harboring each hepatitis C virus (HCV)-targeting shRNA was packed into retroviral particles by cotransfecting 293T cells with p-VSV-G and pCG-Gag-Pol plasmids. Huh 7.5.1 cells were infected by MSCV-PM viruses and selected by puromycin (1 μg/mL) selection. These cells were then plated on 96-well plates (2,000 cells per well) and infected by JFH1 virus the next day at MOI = ~0.2. After 48 h, cells were fixed by 4% paraformaldehyde and permeabilized by 0.2% Triton X-100 and stained for HCV core protein using anti-HCV core monoclonal antibody produced from the anti-HCV core 6G7 hybridoma cells (a gift from Harry Greenberg and Xiaosong He, Stanford University, Palo Alto, CA). Hoechst 33342 was used to stain the nuclei for quantification of total cell number. The cells were then imaged using ImageXpress Micro microscope (Molecular Devices) and quantitated using the MetaXpress software (Molecular Devices).

The sequences of the guide strands of the shRNAs used the assay are:

J1: AGAAAATCAGGTGTCTCCCTCC;
J2: AGGCACTTCCACATGGCGTCCC;
J3: ATCCAAGAAAGGACCCAGTCTT;
J4: TAAGCAGTGATGGGAGCAAGGA;
J5: GATGCCGCTAATGAAGTTCCAC;
J6: GTGGAGTCTTCTGAGTAGGCTG;
J7: TACCTGATCAGACTCCAGGTCC;
J8: ACTACTACCGGGCTGGGGGTGA;
J9: TTGAGTGGCAGAAAATCAGGTG;
J10: TTGAAGCTCTACCTGATCAGAC;
J11: CCCACGTGGCTGGGACCGCTCC;
J12: CCAGTGGAAACATACCTATAAG;
FF: CCCGCCTGAAGTCTCTGATTAA.

siRNA Transfection. siRNAs were synthesized based on four shRNAs targeting HCV-JFH1 from Dharmacon. UU dinucleotides overhangs were added to the 3′ of each strand. The sequences are:

J6: CAGCCUACUCAGAAGACUCCACUU/GUGGAGUCUUCUGAGUAGGCUGUU;
J7: GGACCUGGAGUCUGAUCAGGUAUU/UACCUGAUCAGACUCCAGGUCCUU;
J9: CACCUGAUUUUCUGCCACUCAAUU/UUGAGUGGCAGAAAAUCAGGUGUU;
J11: GGAGCGGUCCCAGCCACGUGGGUU/CCCACGUGGCUGGGACCGCUCCUU;
FF: CGUACGCGGAAUACUUCGAUU/AATCGAAGTATTCCGCGTACG.

For transfection, a 96-well plate format was used. For each well, 0.01 nM to 20 nM final concentration of siRNA, 0.2 μL of Lipofectamine RNAiMAX (Invitrogen) were mixed with 20 μL of OptiMEM media (Invitrogen) and incubated for 20 min at room temperature. Six-thousand Huh 7.5.1 cells in 100 μL DMEM supplemented with L-glutamine and penicillin-steptomycin were added the siRNA-lipid mixture and plated in each well. Twenty-four hours later, JFH1 virus was added at a MOI ~0.2. Forty-eight hours after infection, the plates were imaged as describe above.

Statistical Analysis. We used Pearson χ^2 method to test the significance of specific nucleoid use for top-scoring shRNAs. The shRNAs having sensor values larger than 4 were used to test against those having sensor values less than 1. Calculation of the Pearson correlation coefficient (r) between SHAPE activity and sensor score sequence features of the four viral genomes were based on previously published features (3) and compared using Wilcoxon rank-sum test, and those having P values < 0.05 between HCV and the other three genomes were reported in Table S1. The sequence alignment and consensus sequence of HIV and HCV sequences were retrieved from HIV Databases (http://www.hiv.lanl.gov/) and HCV Databases (4).

1. Fellmann C, et al. (2011) Functional identification of optimized RNAi triggers using a massively parallel sensor assay. *Mol Cell* 41:733–746.
2. Schlabach MR, et al. (2008) Cancer proliferation gene discovery through functional genomics. *Science* 319:620–624.
3. Lu ZJ, Mathews DH (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res* 36:640–647.
4. Kuiken C, Yusim K, Boykin L, Richardson R (2005) The Los Alamos hepatitis C sequence database. *Bioinformatics* 21:379–384.

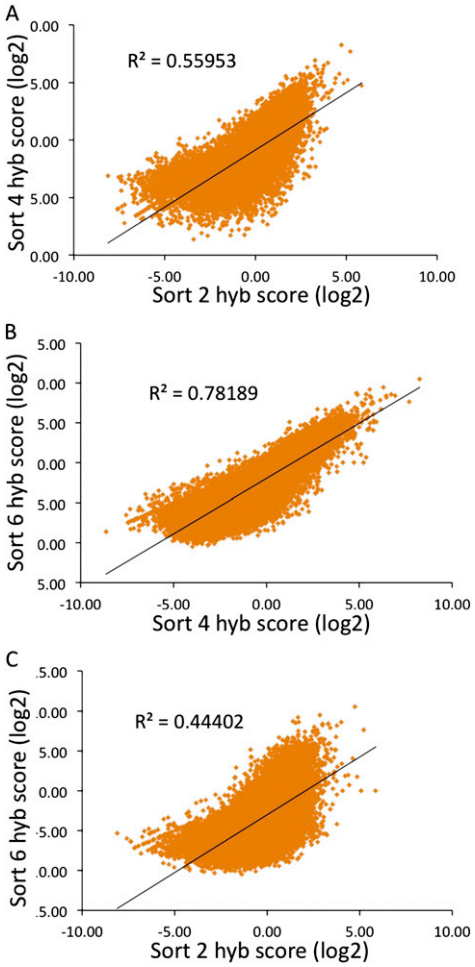

Fig. S1. Correlations between the enrichment scores of different sorts showing that in later sorts more shRNAs were dropping out. (*A*) Correlation between the log2(sort 4/sort 0) and log2 (sort2/sort0) scores of the microarray hybridization. (*B*) Correlation between the log2(sort 6/sort 0) and log2 (sort 4/sort 0) scores of the microarray hybridization. (*C*) Correlation between the log2(sort 6/sort 0) and log2 (sort 2/sort 0) scores of the microarray hybridization.

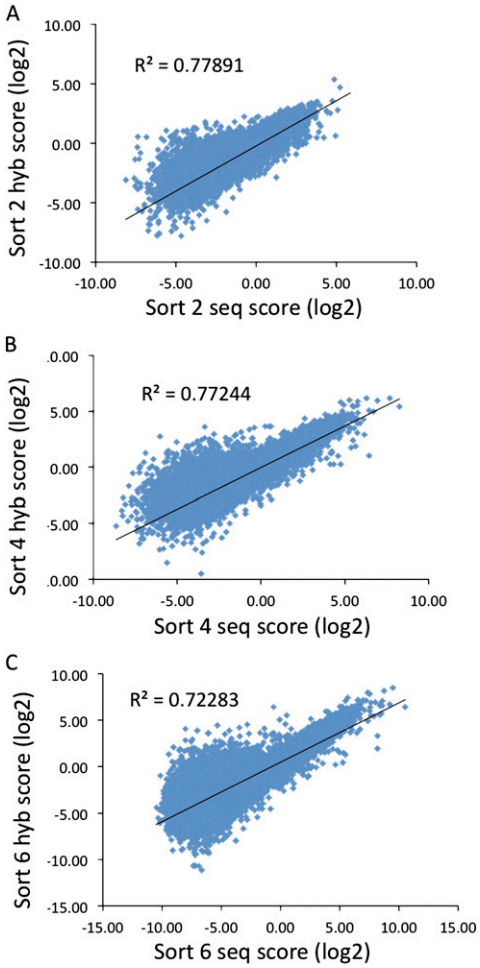

Fig. S2. High correlation between the results from deep sequencing and microarray hybridization. (*A*) Correlation between the log2(sort 2/sort 0) ratios from deep sequencing and microarray hybridization. (*B*) Correlation between the log2(sort 4/sort 0) ratios from deep sequencing and microarray hybridization. (*C*) Correlation between the log2(sort 6/sort 0) ratios from deep sequencing and microarray hybridization.

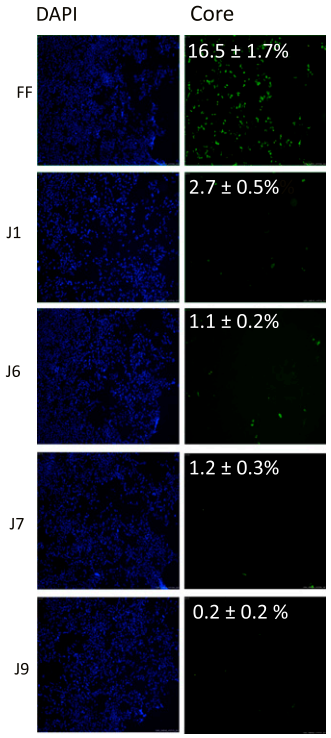

Fig. S3. Immunostaining images of four top scoring shRNAs (J1, J6, J7, J9) targeting HCV-JFH1 and shRNA targeting firefly luciferase were shown. (*Left*) DAPI staining showing nuclei of Huh 7.5.1 cells; (*Right*) Staining of HCV core protein. The numbers are the percentages of cells infected by HCV. (Magnification: 20×.)

Table S1. Sequence features that distinguish HCV from the other three viruses

Sequence features	HCV-JFH1	HIV-1084i	HIV-NL43	Flu- PR8	*P* value between HCV and the other three viruses*
Frequency of UU bases at position 1 and 2†	0.04	0.12	0.12	0.11	8.5×10^{-107}
Frequency of GG bases at position 1 and 2	0.10	0.04	0.04	0.04	7.2×10^{-98}
Frequency of GC bases at position 1 and 2	0.08	0.05	0.04	0.04	5.1×10^{-37}
Frequency of C base at position 1	0.28	0.24	0.24	0.24	6.7×10^{-13}
Frequency of CC base at position 1 and 2	0.09	0.07	0.07	0.06	6.6×10^{-17}

*The *P* values were calculated using Wilcoxon rank-sum test.
†The position is the location on the complementary siRNA guide strand from 5′ to 3′.

Table S2. Genomic base compositions of 55 single stranded RNA virus and the top shRNAs from sensor screen

Name	Polarity	GC (%)	G (%)	C (%)	A (%)	U(%)	A/U
Astrovirus	+	46.18	23.18	22.99	29.67	24.14	1.23
Coronavirus	+	38.42	21.5	16.91	26.8	34.77	0.77
Coxsackie virusA9	+	46.97	24.47	22.49	29.18	23.84	1.22
Dengue virus type 1	+	46.38	25.88	20.49	31.96	21.65	1.48
Dengue virus type 2	+	45.8	25.21	20.58	33.21	20.98	1.58
Dengue virus type 3	+	46.47	25.91	20.55	32.12	21.4	1.50
Dengue virus type 4	+	46.91	26.31	20.59	31.03	22.06	1.41
Enterovirus 71	+	47.99	24.37	23.61	27.55	24.46	1.13
Eastern equine encephalitis	+	50.33	24.34	25.98	27.69	21.97	1.26
Hepatitis A virus	+	37.15	21.74	15.4	30.08	32.76	0.92
Hepatitis C virus	+	57.88	28.13	29.74	20.54	21.57	0.95
Hepatitis E virus	+	57.29	26.19	31.09	17.32	25.38	0.68
Hepatitis G virus	+	58.83	32.18	26.64	17.81	23.36	0.76
Japanese encephalitis virus	+	51.48	28.42	23.05	27.46	21.05	1.30
Norwalk virus	+	48.54	23.62	24.91	26.5	24.95	1.06
O'Nyong Nyong virus	+	48.56	24.42	24.13	30.89	20.54	1.50
Poliovirus type 3	+	45.96	23.28	22.68	30.18	23.85	1.27
Rhinovirus type 89	+	38.29	19.44	18.84	32.14	29.56	1.09
Ross river virus	+	52.18	26.6	25.57	27.37	20.44	1.34
Rubella virus	+	69.59	30.87	38.71	14.9	15.5	0.96
SARS coronavirus	+	41.02	21.08	19.93	28.25	30.72	0.92
Sindbis virus	+	51.05	25.18	25.87	27.98	20.96	1.33
Venezuelan encephalitis virus	+	50.12	25.49	24.62	28.08	21.79	1.29
West Nile virus	+	51.2	28.79	22.4	27.23	21.56	1.26
Yellow fever virus	+	49.73	28.58	21.13	27.06	23.21	1.17
HIV-1	Retro	43.28	24.93	18.34	34.66	22.05	1.57
HIV-2	Retro	45.89	25.19	20.69	33.34	20.76	1.61
HTLV-1	Retro	52.68	18.16	34.51	23.06	24.25	0.95
HTLV-2	Retro	53.62	17.75	35.86	24.4	21.96	1.11
Borna virus	–	50.65	25.06	25.58	25.06	24.21	1.04
Bunyamwera virus	–	35.97	19.25	16.71	35.31	28.71	1.23
Crimean-Congo virus	–	43.59	22.44	21.14	31.55	24.85	1.27
Ebola virus	–	44.36	21.54	22.81	30.62	25.02	1.22
Hantaan virus	–	40.44	22.59	17.84	31.82	27.74	1.15
Hendra virus	–	42.38	22.97	19.4	32.53	25.08	1.30
Hepatitis D virus	–	58.8	28.83	29.96	20.15	21.05	0.96
Influenza A virus (H3N2)	–	43.57	24.31	19.25	32.43	23.98	1.35
Influenza B virus	–	41.13	22.47	18.65	35.23	23.63	1.49
Influenza C virus	–	38.6	20.58	18.02	35.78	25.6	1.40
La Crosse virus	–	37.64	20.39	17.24	34.6	27.74	1.25
Marburg virus	–	40.71	19.66	21.04	31.94	27.34	1.17
Measles virus	–	47.19	24.34	22.84	28.45	24.35	1.17
Metapneumovirus	–	39.09	20.97	18.11	36.94	23.96	1.54
Mokola virus	–	45.28	25.34	19.94	30.15	24.56	1.23
Mumps virus	–	41.98	19.24	22.73	29.78	28.23	1.05
Nipah virus	–	40.36	21.5	18.85	33.21	26.43	1.26
Parainfluenzavirus 1	–	38.53	20.1	18.42	36.39	25.06	1.45
Parainfluenzavirus 2	–	39.9	18.43	21.46	31.71	28.39	1.12
Parainfluenzavirus 3	–	36.52	18.94	17.56	37.99	25.48	1.49
Rabies virus	–	46.05	24.35	21.69	28.67	25.27	1.13
Respiratory syncytial virus	–	35.32	15.17	20.14	38.98	25.7	1.52
Rift Valley fever virus	–	45.61	25.27	20.33	28.12	26.26	1.07
Sendai virus	–	46.5	24.9	21.59	29.4	24.09	1.22
Sin Nombre virus	–	39.13	22.29	16.84	31.31	29.54	1.06
Vesicular stomatitis virus	–	41.76	22.27	19.48	31.82	26.42	1.20
Top-scoring shRNAs in sensor (498)		43.3	22.49	20.84	34.81	21.85	1.59

Partly adapted from Auewarakul (1).

1. Auewarakul (2005) Composition bias and genome polarity of RNA viruses. *Virus Res* 109:33–37.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (PDF)