Molecular Cell, Volume *41*

Supplemental Information

Functional Identification of Optimized RNAi Triggers Using a Massively Parallel Sensor Assay

Christof Fellmann, Johannes Zuber, Katherine McJunkin, Kenneth Chang, Colin D. Malone, Ross A. Dickins, Qikai Xu, Michael O. Hengartner, Stephen J. Elledge, Gregory J. Hannon, and Scott W. Lowe

Figure S1

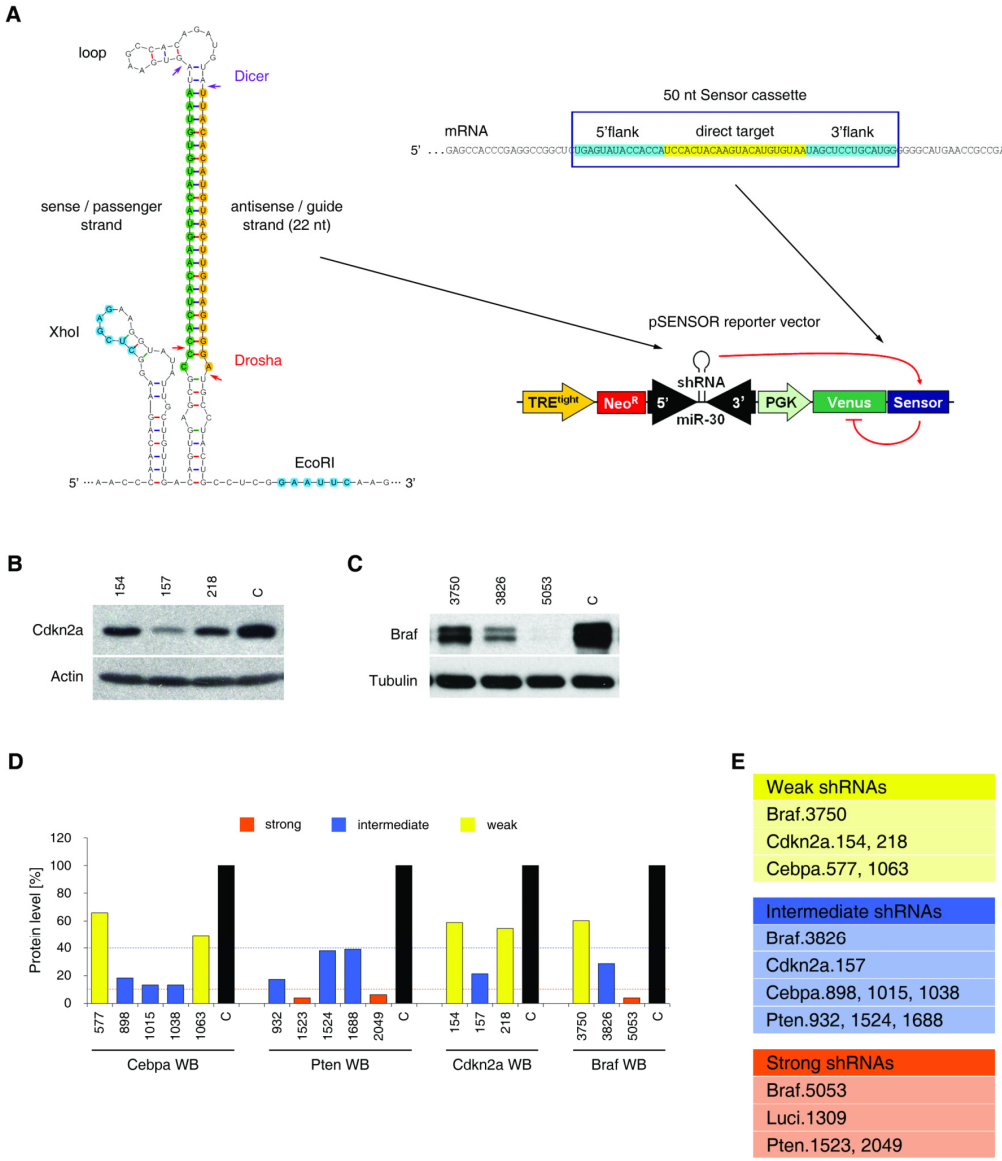

Figure S1. Design of the pSENSOR vector and classification of control shRNAs

(A) Schematic depicting the structure and assembly of the pSENSOR vector. In the miR30 shRNA cassette (left, containing sh.p53.814, as an example), the 22 nt guide strand (orange) and its complementary variable region (green) are highlighted. Restriction sites used for cloning (XhoI, EcoRI) are marked in blue. Mapped canonical cleavage sites for Drosha and Dicer are indicated by red and blue arrows, respectively (Silva et al., 2005). The 50 nt Sensor cassette (top right, here containing the target of sh.p53.814, as an example) is composed of a 5'flanking region, the direct target site and a 3'flanking region. Assembly of the pSENSOR reporter vector (bottom right) genetically links each shRNA to its cognate Sensor. The RNA structure was generated using UNAfold (Markham and Zuker, 2008) at dinamelt.bioinfo.rpi.edu/twostate-fold.php.
(B) Western blot analysis of Cdkn2a in *p53*$^{-/-}$ MEFs expressing indicated Cdkn2a shRNAs. C, sh.Luci.1309. Actin served as loading control.
(C) Western blot analysis of Braf in NIH3T3 cells expressing indicated Braf shRNAs. C, control shRNA targeting Map2k1 (sh.Map2k1.1200). Tubulin served as loading control.
(D) Comparison of control shRNAs. Quantification of shRNA-mediated knockdown efficiency by protein densitometry of the Cebpa and Pten western blots (Figure 1B), as well as the Cdkn2a and Braf western blots (B, C).
(E) Classification of control shRNAs based on western blot data. The 17 shRNAs used to set up and monitor the system were classified into three groups according to their target knockdown efficiencies (D). The knockdown level of sh.Luci.1309 was assessed previously (Silva et al., 2005). Weak shRNAs: <60% knockdown; intermediate shRNAs: 60-90% knockdown; strong shRNAs: >90% knockdown.

Figure S2

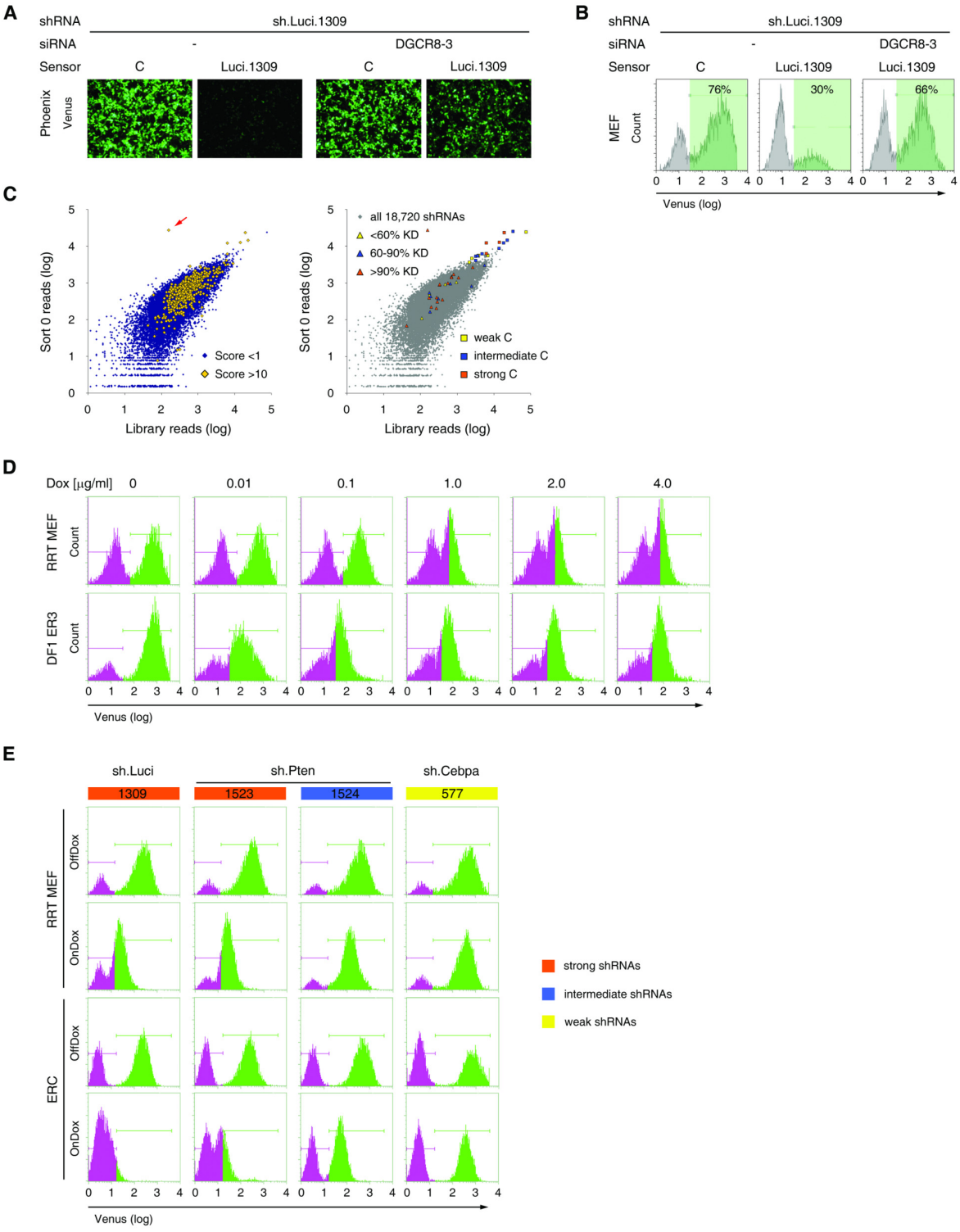
A
shRNA
sh.Luci.1309
siRNA
-
DGCR8-3
Sensor
C
Luci.1309
C
Luci.1309
Phoenix
Venus
B
shRNA
sh.Luci.1309
siRNA
-
DGCR8-3
Sensor
C
Luci.1309
Luci.1309
MEF
Count
76%
30%
66%
Venus (log)
C
Sort 0 reads (log)
Score <1
Score >10
Library reads (log)
all 18,720 shRNAs
<60% KD
60-90% KD
>90% KD
weak C
intermediate C
strong C
Sort 0 reads (log)
Library reads (log)
D
Dox [μg/ml]
0
0.01
0.1
1.0
2.0
4.0
RRT MEF
DF1 ER3
Count
Venus (log)
E
sh.Luci
sh.Pten
sh.Cebpa
1309
1523
1524
577
RRT MEF
ERC
OffDox
OnDox
Venus (log)
strong shRNAs
intermediate shRNAs
weak shRNAs

Figure S2 (continued)

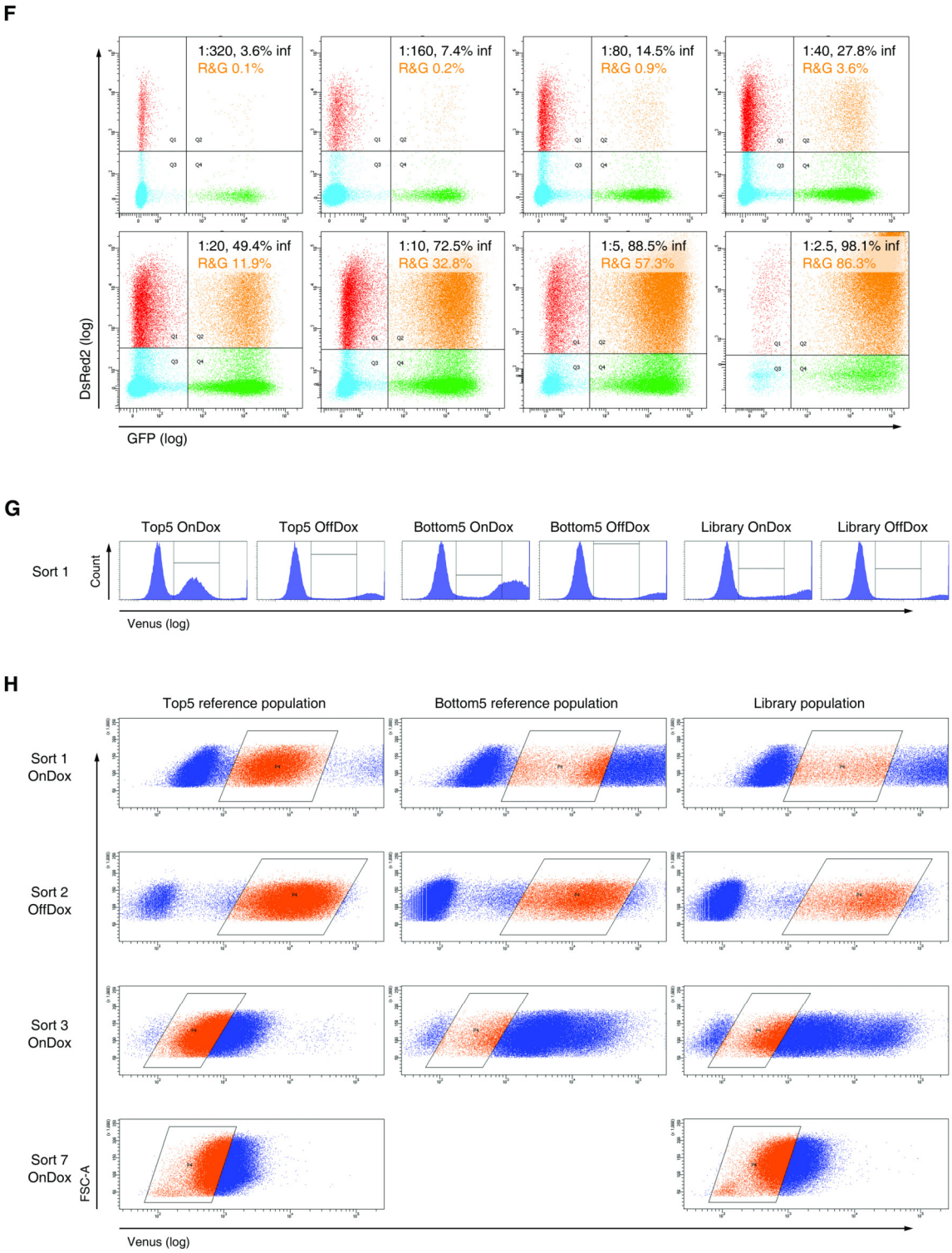

Figure S2. Optimization of the Sensor assay and reference guided Ping-Pong sorting of complex libraries

(A) Epifluorescence images of Phoenix packaging cells 3 days after transfection of pSENSOR constructs harboring sh.Luci.1309 and its cognate (Luci.1309) or a mismatched (C) target Sensor; both without and with co-transfection of a potent siRNA targeting DGCR8 (DGCR8-3).
(B) Flow cytometry analysis of *p53*$^{-/-}$ MEFs infected with viral supernatants harvested from packaging cells show in A. Venus fluorescence was measured 48 h post-infection. pSENSOR constructs harboring a potent shRNA and its cognate target Sensor show reduced viral titers, likely due to RNAi-mediated suppression of proviral transcripts during packaging. Co-transfection of a DGCR8 siRNA almost completely restores normal packaging.
(C) Deep-sequencing based analysis of packaging efficiency. Shown are correlation plots of reads per shRNA between the plasmid library and the initial pool of infected cells (Sort 0). Values represent normalized means of technical (plasmid library) or biological (Sort 0) duplicates. The left panel compares the correlation of all non-scoring (Score <1; 17,909 shRNAs) and scoring (Score >10; 453 shRNAs) shRNAs, while the right panel shows the distribution of all validated (triangles) and control (squares) shRNAs of indicated potency (see Figure S1E for details). In all comparisons, the strong correlation between plasmid library and infected target cells is independent of shRNA potency, indicating that retroviral packaging in the presence of DGCR8 siRNAs is equally efficient for all shRNAs. Control shRNAs were designed to be 15-fold overrepresented in the pool, which is reflected in both plasmid and Sort 0 samples. An outlier, sh.p53.703, is highlighted (red arrow).
(D) Flow cytometry plots of immortalized Rosa26-rtTA-M2 (RRT) MEFs and DF1-ER3 reporter cells transduced with a pSENSOR construct carrying an sh.Luci.1309 shRNA-Sensor cassette and treated with various concentrations of Dox for 4 days. In both reporter cell lines, Dox treatment induced repression of the Venus reporter in a dose-dependent manner. DF1-ER3 cells expressing optimized rtTA3 show higher Dox sensitivity than the rtTA-M2 line.
(E) Flow cytometry plots of RRT MEFs and ERC reporter cells (a single cell clone derived from DF1-ER3 cells) transduced with the indicated shRNA-Sensor constructs and treated for 7 days +/- Dox. Gating in OffDox samples shows infected subpopulations (green) and allows tracking of shRNA-Sensor mediated Venus repression OnDox. Levels of Venus repression in both reporter cell lines accurately reflect the potency of different shRNAs (see Figure S1E for details). Overall, ERC cells show a broader dynamic range resulting from stronger Venus repression by potent shRNAs (compare sh.Luci.1309 in both cell lines).
(F) Flow cytometry dot plots of ERC cells 2 days after co-infection with ecotropic MSCV-DsRed2 and MSCV-GFP viral supernatants (always mixed in equal proportions; diluted at rates ranging from 1:2.5 to 1:320). Overall infection rate (inf) and percentage of DsRed2/GFP double positive cells (R&G) are indicated in each plot. ERC cells can be homogeneously (>98%) infected, while infection rates <15% almost exclusively result in single copy genomic integration of the provirus, as indicated by the absence of double-

positive cells. The rate of double-positive cells (orange, R&G) is a function of the infection efficiency and approximates to a theoretical Poisson distribution.
(G) Reference guided gating in Ping-Pong sorting of complex shRNA-Sensor libraries. Shown are representative Venus intensity histograms of the two reference populations (Top5, a pool of five highly potent shRNAs, and Bottom5, a pool of five weak shRNAs) and the library population (Library), with or without Dox treatment (On- and OffDox, respectively).
(H) Venus/FSC-A dot plots used for gating of Sensor Ping-Pong sorts. Shown are representative samples of reference (Top5, Bottom5) and library populations. Parallelograms indicate the sorting gates, with sorted cells highlighted in orange. All pools (Top5, Bottom5, Library in duplicates) were always cultured and analyzed both On- and OffDox. Shown is a subset of these populations at the indicated stages (Sort 1, 2, 3 and 7). Note that the Library population progressively resembles the Top5 reference population over the course of the experiment.

Figure S3

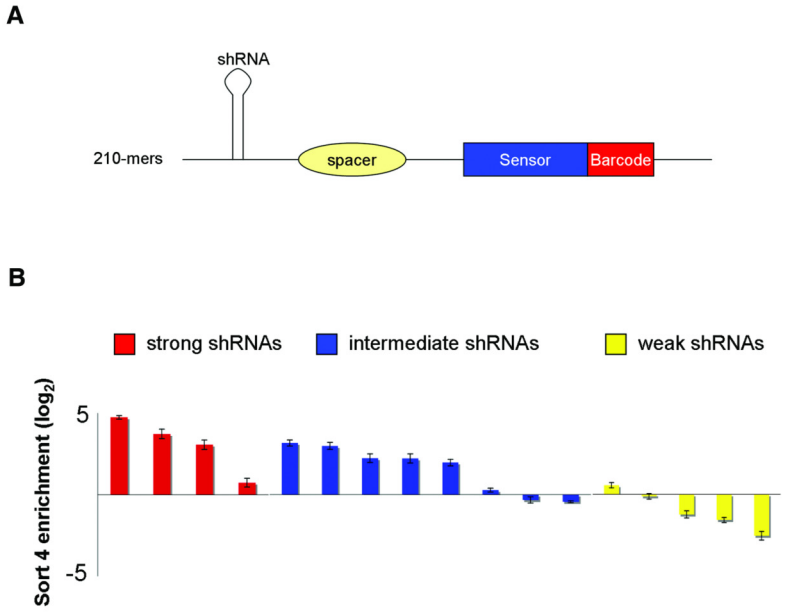

Figure S3. Microarray based monitoring of barcoded shRNA-Sensor libraries

(A) Schematic of the design of barcoded shRNA-Sensor libraries.
(B) Enrichment or depletion of 17 barcoded control shRNA-Sensor constructs in transduced reporter cells after Sort 4. The change in representation was quantified through hybridization of amplified barcodes to custom microarrays. Color coding corresponds to Figure S1E. Error bars represent the standard error of the mean.

Figure S4

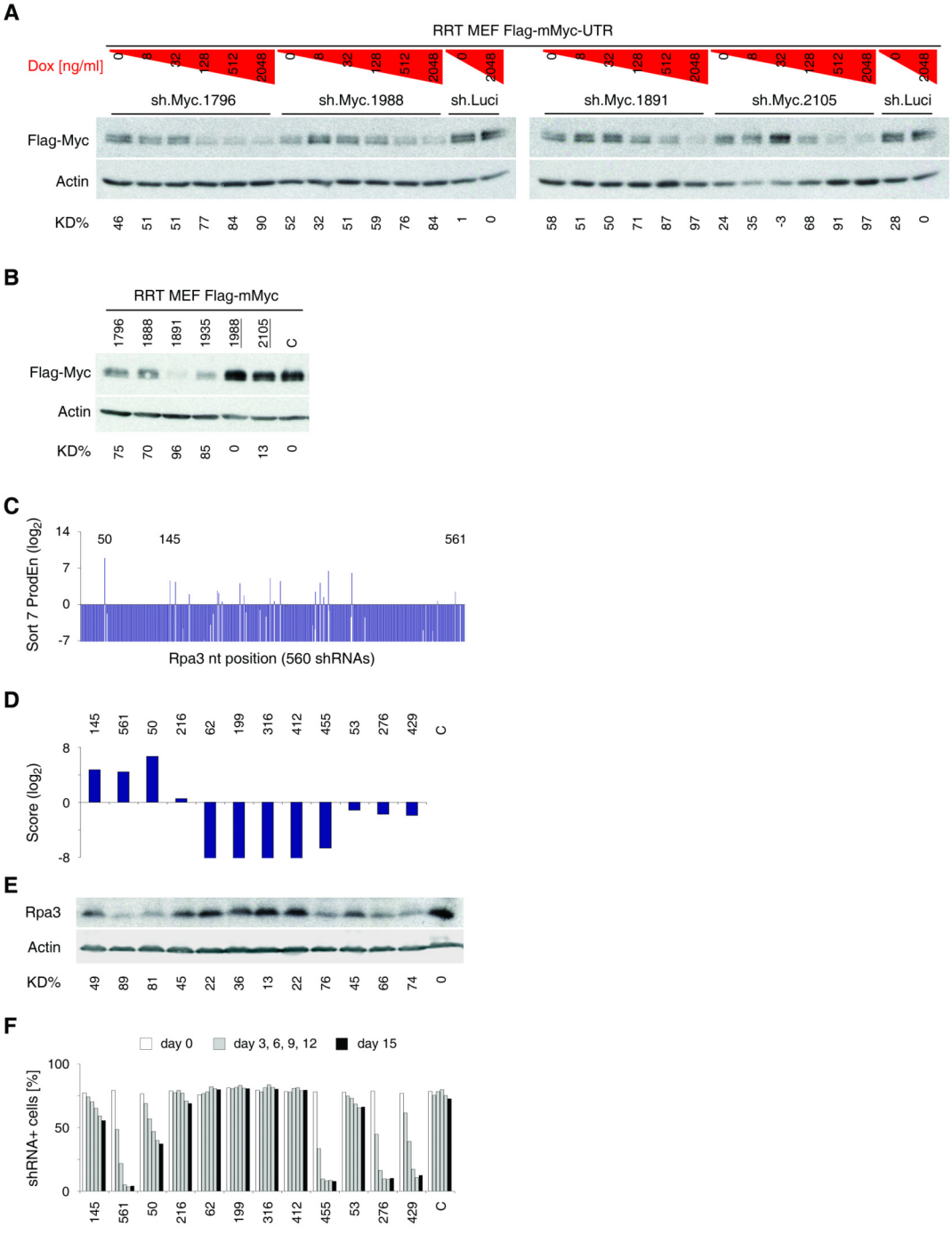
A
RRT MEF Flag-mMyc-UTR
Dox [ng/ml]
0 8 32 128 512 2048 0 8 32 128 512 2048 0 2048 0 8 32 128 512 2048 0 8 32 128 512 2048 0 2048
sh.Myc.1796
sh.Myc.1988
sh.Luci
sh.Myc.1891
sh.Myc.2105
sh.Luci
Flag-Myc
Actin
KD%
46 51 51 77 84 90 52 32 51 59 76 84 1 0 58 51 50 71 87 97 24 35 -3 68 91 97 28 0
B
RRT MEF Flag-mMyc
1796 1888 1891 1935 1988 2105 C
Flag-Myc
Actin
KD%
75 70 96 85 0 13 0
C
Sort 7 ProdEn (log2)
14
7
0
-7
50
145
561
Rpa3 nt position (560 shRNAs)
D
145 561 50 216 62 199 316 412 455 53 276 429 C
Score (log2)
8
0
-8
E
Rpa3
Actin
KD%
49 89 81 45 22 36 13 22 76 45 66 74 0
F
day 0
day 3, 6, 9, 12
day 15
shRNA+ cells [%]
100
50
0
145 561 50 216 62 199 316 412 455 53 276 429 C

Figure S4. Doxycycline titration of Tet-regulated shRNAs and evaluation of shRNAs targeting essential genes

(A) Flag western blot analysis in RRT MEFs co-expressing Flag-tagged Myc from a transcript containing the *Myc* 3'UTR and indicated Tet-regulated shRNAs. Cells were treated with various concentrations of Dox for 4 days resulting in dose-dependent Myc repression. Actin served as loading control. sh.Luci, sh.Luci.1309 control shRNA. KD%, quantification of knockdown efficiencies by protein densitometry. Flag-Myc expression levels are normalized to actin and relative to sh.Luci.1309, 2048 ng/ml Dox of the respective immunoblot.
(B) Flag western blot analysis in RRT MEFs co-expressing the indicated shRNAs and Flag-tagged Myc from a transcript lacking the *Myc* 3'UTR. C, sh.Luci.1309. Underlined numbers (1988, 2105) indicate shRNAs targeting the 3'UTR of *Myc*. KD%, quantification of knockdown efficiencies relative to C and normalized to actin. A complementary western blot in cells overexpressing Flag-mMyc-UTR is shown in Figure 5C.
(C) Product enrichment scores (ProdEn) of 560 shRNA-Sensor constructs covering the mouse *Rpa3* transcript. Numbers highlight enriched shRNAs that were selected for subsequent validation.
(D) Integrated Score for selected Rpa3 shRNAs tested in western blot analysis and proliferation assays, including shRNAs strongly scoring in the Sensor assay, non-scoring controls and 3 functional shRNAs (276, 429, 455) that were previously identified through empirical testing (Zuber et al., 2011).
(E) Western blot analysis of Rpa3 in RRT MEFs expressing shRNAs indicated above at single copy. C, control shRNA targeting Bcl2 (sh.Bcl2.1132). Actin served as loading control. KD%, quantification of knockdown efficiencies relative to C and normalized to actin.
(H) Competitive proliferation assay of RRT MEFs expressing indicated shRNAs from a single genomic integration. The bar graph shows the relative percentage of shRNA expressing cells at indicated days after shRNA induction. C, sh.Luci.1309.

Figure S5

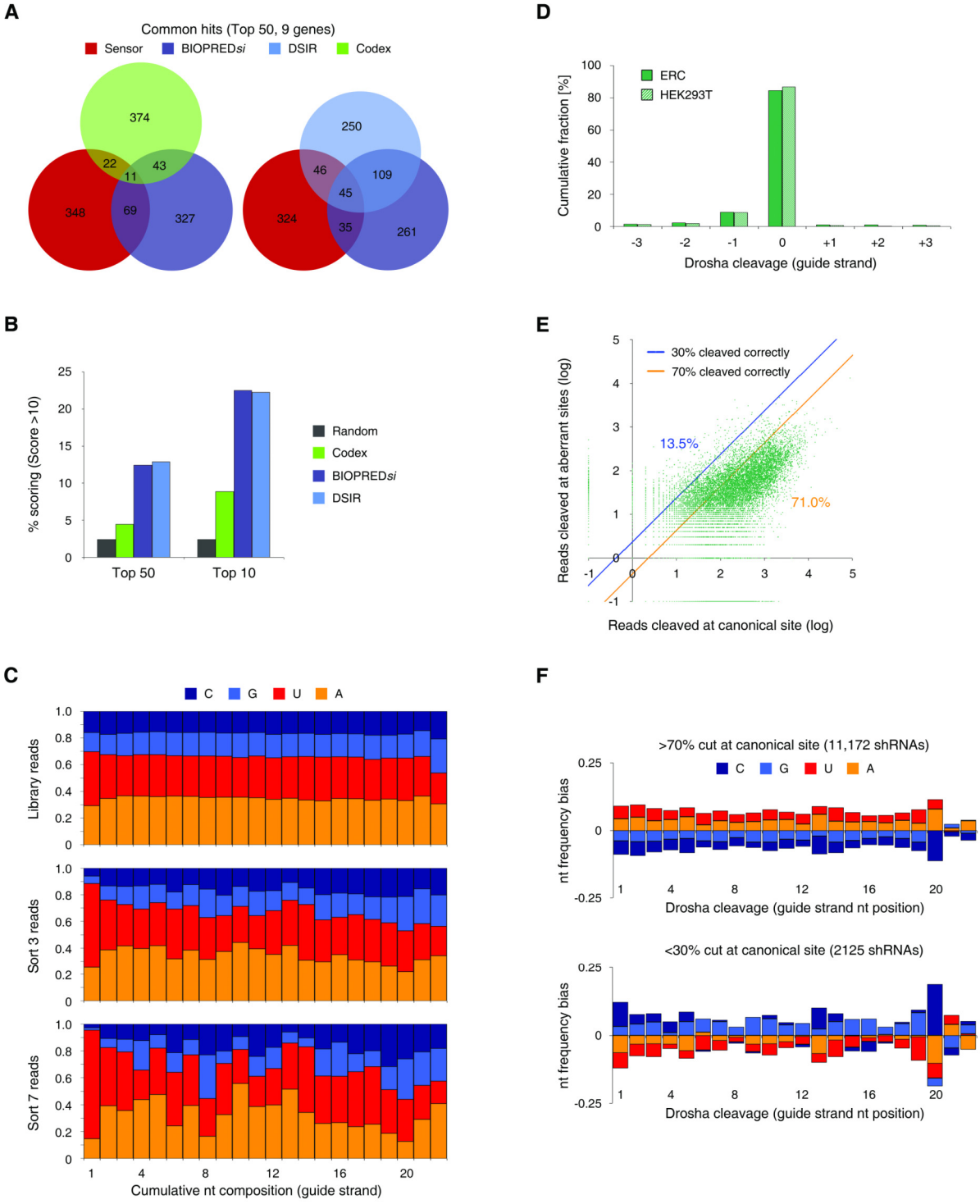
A
Common hits (Top 50, 9 genes)
Sensor
BIOPREDsi
DSIR
Codex
374
22
43
11
348
69
327
250
46
109
45
324
35
261
B
% scoring (Score >10)
Random
Codex
BIOPREDsi
DSIR
Top 50
Top 10
C
C
G
U
A
Library reads
Sort 3 reads
Sort 7 reads
Cumulative nt composition (guide strand)
D
Cumulative fraction [%]
ERC
HEK293T
Drosha cleavage (guide strand)
E
30% cleaved correctly
70% cleaved correctly
13.5%
71.0%
Reads cleaved at aberrant sites (log)
Reads cleaved at canonical site (log)
F
>70% cut at canonical site (11,172 shRNAs)
nt frequency bias
Drosha cleavage (guide strand nt position)
<30% cut at canonical site (2125 shRNAs)

Figure S5 (continued)

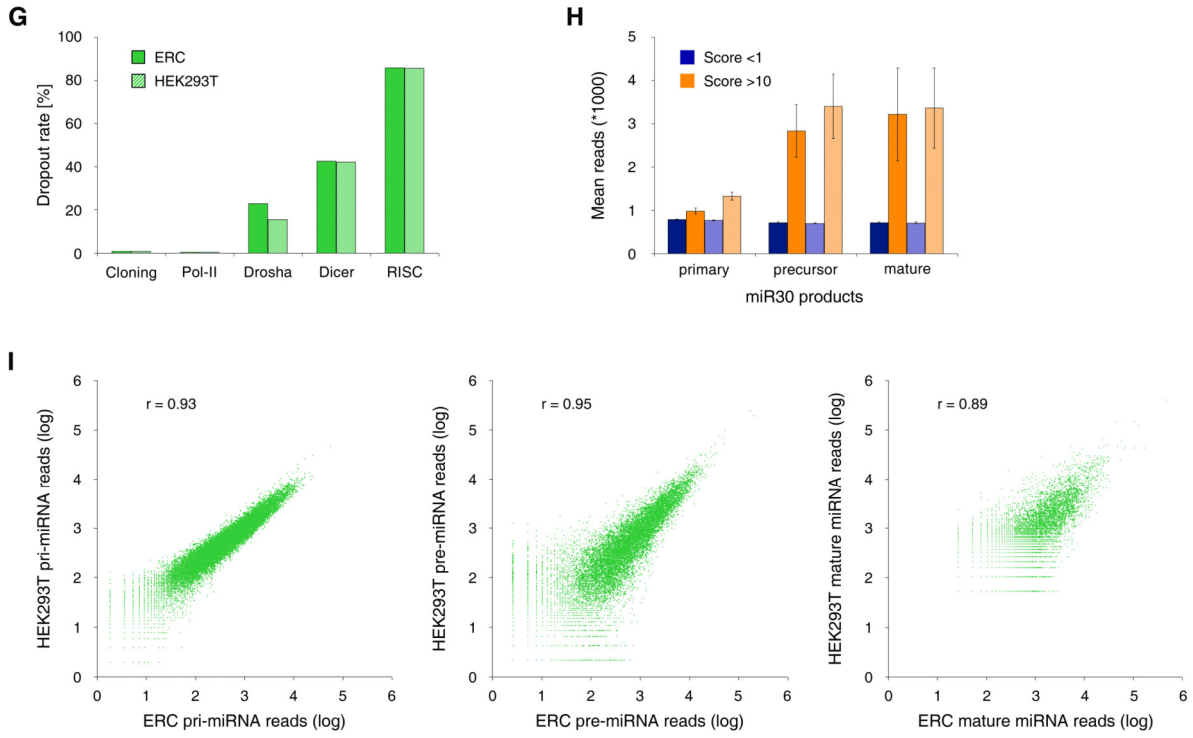

Figure S5. Comparison to prediction algorithms and analysis of features associated with efficient shRNA biogenesis and function

(A) Venn diagram illustrating the overlap between predictions from different algorithms and hits from the Sensor assay. The 50 highest ranking predictions and the top 50 Sensor hits for each of the 9 tiled transcripts are compared.
(B) Performance of existing prediction algorithms. For each algorithm, the percentage of top predictions that validated as effective shRNAs in the Sensor assay (Score >10) is plotted. The overall frequency of scoring shRNAs (Score >10), reflecting an unbiased (Random) shRNA design, is indicated as a reference. Notably, the scoring rates observed for the top 10 BIOPRED*si* (Huesken et al., 2005) and DSIR (Vert et al., 2006) predictions (both ~22%) closely resemble validation rates observed in empirical testing of shRNAs produced using these design tools (J.Z. and S.W.L., unpublished data).
(C) Cumulative nucleotide composition of all sequence reads obtained for the initial plasmid library (Library) and library pools of transduced reporter cells after the indicated sorts (Sort 3, Sort 7), showing the progressive emergence of specific sequence features. Bars represent means of technical (Library) or biological (Sort 3, Sort 7) duplicates. The

general A/U bias found in the plasmid pool is likely due to PCR amplification biases during library preparation and/or sequencing.
(D) Cumulative distribution of Drosha cleavage sites in ERC and HEK293T cells transduced with the Sensor library. Cleavage was assessed for each individual pre-miRNA sequence and is indicated with respect to the canonical site (see Figure S1A for details). Negative values (-3, -2, -1) mark cleavage closer to the 5' end of the guide strand, while positive values (+1, +2, +3) indicate a shift in the opposite direction.
(E) Scatter plot of aberrantly and canonically cleaved pre-miRNA reads for each individual shRNA in HEK293T cells transduced with the Sensor library, showing that the majority of shRNAs (71.0%) are usually cut at the canonical cleavage site (>70% of reads). For better visualization, lines indicate the thresholds for 30% (blue) and 70% (orange) correctly cleaved sequences. Comparable results were obtained with ERC cells.
(F) Nucleotide bias of correctly (>70% cut at canonical site) and aberrantly (<30% cut at canonical site) cleaved pre-miRNAs. Bars indicate the deviation of the nucleotide frequency from the unbiased expectation (0.25) at each position of the guide strand. Data is shown for HEK293T cells; comparable patterns were observed in ERC cells.
(G) Step-specific shRNA dropout rate in ERC and HEK293T cells transduced with the Sensor library, indicating that processing is selective at each step of shRNA biogenesis. Individual shRNAs were counted as dropouts if they were significantly represented (>100 reads) at the previous stage, but not present (0 reads) after the indicated step. Drosha symbolizes the transition from pri- to pre-miRNAs. Dicer indicates the transition from pre- to mature miRNAs. RISC marks the transition from mature miRNAs to scoring Sensor-identified shRNAs (Score >10). Dropout rates for cloning and RNA polymerase II (Pol-II) expression are indicated as reference.
(H) Mean read number of small RNA precursors of scoring (Score >10) and non-scoring (Score <1) shRNAs in ERC (solid colors) and HEK293T (striped colors) cells transduced with the Sensor library, indicating that Sensor-identified potent shRNAs lead to higher amounts of mature small RNAs and suggesting that efficient processing is a prerequisite for shRNA potency. Error bars represent the standard error of the mean.
(I) Correlation of reads per shRNA precursor molecule (pri-, pre- and mature miRNA) in ERC and HEK293T cells transduced with the Sensor library. The high correlation at all levels indicate that i) shRNA processing is similar between human and avian cells, ii) drop-out of shRNAs at each level is due to a specific, evolutionarily conserved process, and iii) ERC cells are valid reporter cells for evaluating mammalian shRNA processing and function. r, Pearson correlation coefficient.

Figure S6

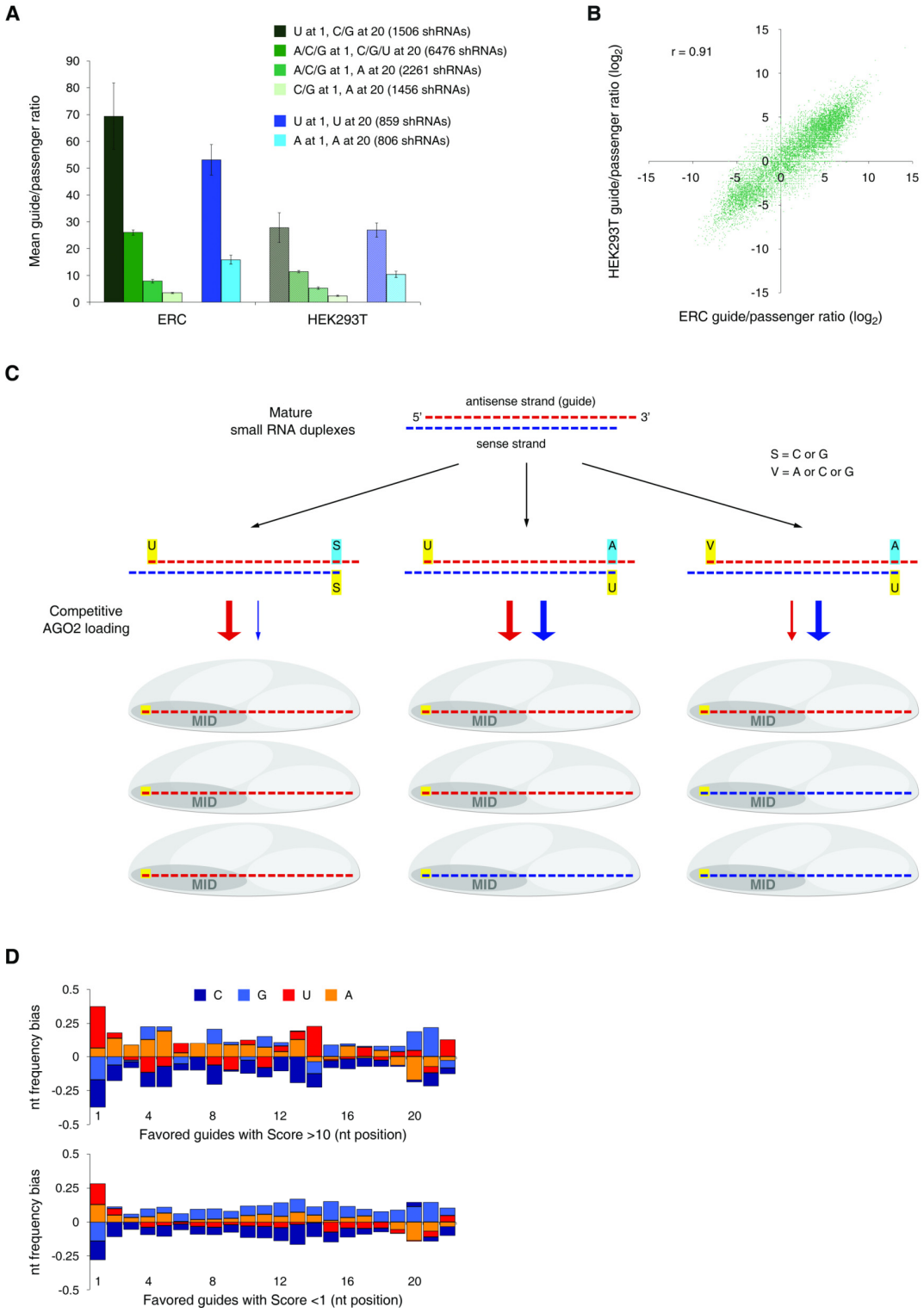
A
U at 1, C/G at 20 (1506 shRNAs)
A/C/G at 1, C/G/U at 20 (6476 shRNAs)
A/C/G at 1, A at 20 (2261 shRNAs)
C/G at 1, A at 20 (1456 shRNAs)
U at 1, U at 20 (859 shRNAs)
A at 1, A at 20 (806 shRNAs)
Mean guide/passenger ratio
ERC
HEK293T
B
r = 0.91
HEK293T guide/passenger ratio (log2)
ERC guide/passenger ratio (log2)
C
Mature
small RNA duplexes
antisense strand (guide)
5'
3'
sense strand
S = C or G
V = A or C or G
Competitive
AGO2 loading
MID
D
C
G
U
A
nt frequency bias
Favored guides with Score >10 (nt position)
Favored guides with Score <1 (nt position)

Figure S6. Sensor-identified nucleotide biases suggest a model of competitive AGO2-mediated strand selection

(A) Mean guide/passenger ratios for different subsets of shRNAs with specific sequence features in ERC (solid colors) and HEK293T (striped colors) cells transduced with the Sensor library. In both cell lines, the relative nucleotides at position 1 and 20 of the guide strand heavily influence the guide/passenger ratio (green bars). Even under conditions where the thermodynamic stability at both 5'-ends is equal, the strand bias was strongly affected by the specific nucleotides at position 1 and 20 (blue bars). Only shRNAs with at least one read for either the guide or passenger strand in both cell lines were considered (11,868 shRNAs in total). Error bars represent the standard error of the mean.
(B) Correlation of guide/passenger ratios in ERC and HEK293T cells transduced with the Sensor library, indicating that strand selection is based on a highly conserved and specific mechanism. r, Pearson correlation coefficient.
(C) Model for AGO2-mediated competitive strand selection. Both the sense and antisense strand of mature small RNA duplexes associate with AGO2; however, only one strand is selected for final RISC assembly and target recognition, while the other is cleaved and degraded. Our results suggest that the nucleotide at the 5'-end of both strands strongly impacts strand selection, presumably due to binding affinities of the MID domain of AGO2. Generally, strands harboring U at their 5'-end are strongly preferred. In a situation of rate limiting amounts of small RNA duplexes, a clear bias for selecting one strand will affect the overall RNAi response. Specifically, sense strands harboring U in position 1 can outcompete target-specific antisense strands for RISC loading, thereby diminishing the specific RNAi effect. Conversely, a strong selection for loading of the intended guide strand will strengthen the target-specific RNAi response, while minimizing passenger-mediated off-target effects.
(D) Nucleotide bias of different subsets of favored (guide/passenger >50) guide strands in ERC cells transduced with the Sensor library (see Figure 7C for comparison). While the subset of non-scoring (Score <1) shRNAs does not show any marked thermodynamic asymmetry beyond nucleotide 1 and 20, the subset of scoring (Score >10) shRNAs shows a clear asymmetry reminiscent of the profile at the genomic level after Sort 7 (see Figure 6C for details). Comparable results were obtained with HEK293T cells.

Table S1 – Control shRNAs

Braf.3750
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCTAGCATAATGACAATTATTTATAGTGAAG
CCACAGATGTATAAATAATTGTCATTATGCTAATGCCTACTGCCTCGGAATTC

Braf.3826
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCCCATTGTTTCTTCCAACTTATAGTGAAG
CCACAGATGTATAAGTTGGAAGAAACAATGGGATGCCTACTGCCTCGGAATTC

Braf.5053
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCTAGGGTGATGTCTCACTTGAATAGTGAAG
CCACAGATGTATTCAAGTGAGACATCACCCTATTGCCTACTGCCTCGGAATTC

Cdkn2a.154
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCGCGCTCTGGCTTTCGTGAACATAGTGAAG
CCACAGATGTATGTTCACGAAAGCCAGAGCGCATGCCTACTGCCTCGGAATTC

Cdkn2a.157
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACGCTCTGGCTTTCGTGAACATTAGTGAAG
CCACAGATGTAATGTTCACGAAAGCCAGAGCGCTGCCTACTGCCTCGGAATTC

Cdkn2a.218
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACGGAATCCTGGACCAGGTGATTAGTGAA
GCCACAGATGTAATCACCTGGTCCAGGATTCCGGTGCCTACTGCCTCGGAATTC

Cebpa.577
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCGGCCGCTGGTGATCAAACAATAGTGAA
GCCACAGATGTATTGTTTGATCACCAGCGGCCGTTGCCTACTGCCTCGGAATTC

Cebpa.898
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACCGAGATAAAGCCAAACAACGTAGTGAA
GCCACAGATGTACGTTGTTTGGCTTTATCTCGGCTGCCTACTGCCTCGGAATTC

Cebpa.1015
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAAGCCGAGATAAAGCCAAACAATAGTGAA
GCCACAGATGTATTGTTTGGCTTTATCTCGGCTCTGCCTACTGCCTCGGAATTC

Cebpa.1038
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACAACGTGGAGACGCAACAGAATAGTGAA
GCCACAGATGTATTCTGTTGCGTCTCCACGTTGCTGCCTACTGCCTCGGAATTC

Cebpa.1063
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACTGGAGTTGACCAGTGACAATTAGTGAA
GCCACAGATGTAATTGTCACTGGTCAACTCCAGCTGCCTACTGCCTCGGAATTC

Luci.1309
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCCGCCTGAAGTCTCTGATTAATAGTGAAG
CCACAGATGTATTAATCAGAGACTTCAGGCGGTTGCCTACTGCCTCGGAATTC

Pten.932
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCGACTTAGACTTGACCTATATTAGTGAAG
CCACAGATGTAATATAGGTCAAGTCTAAGTCGATGCCTACTGCCTCGGAATTC

Pten.1523
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACCAGCTAAAGGTGAAGATATATAGTGAA
GCCACAGATGTATATATCTTCACCTTTAGCTGGCTGCCTACTGCCTCGGAATTC

Pten.1524
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACAGCTAAAGGTGAAGATATATTAGTGAA
GCCACAGATGTAATATATCTTCACCTTTAGCTGGTGCCTACTGCCTCGGAATTC

Pten.1688
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCTTGGGTAAATACGTTCTTCATTAGTGAAG
CCACAGATGTAATGAAGAACGTATTTACCCAAATGCCTACTGCCTCGGAATTC

Pten.2049
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAAAGATCAGCATTCACAAATTATAGTGAA
GCCACAGATGTATAATTTGTGAATGCTGATCTTCTGCCTACTGCCTCGGAATTC

Table S1. Nucleotide sequences of control shRNAs

Sequences of validated control shRNAs with different potency (see Figure 1 and Figure S1B-E for details). Provided are 116 nt XhoI/EcoRI miR30-shRNA fragments. Restriction sites are highlighted in grey, while 97-mer oligonucleotides used for PCR-based single shRNA cloning are underlined.

Table S2 – Tiled transcripts

Gene (*species*)	**Gene ID**	**Tiled accession**	**Length**	**Comment**
Bcl2 (*Mus musculus*)	12043	NM_009741.3, NM_177410.2	1986	Common region of transcript variant 1 and 2
Hras1 (*Mus musculus*)	15461	NM_008284.1	570	Now new, substantially different transcripts
Kras (*Mus musculus*)	16653	NM_021284.4	4662	
Mcl1 (*Mus musculus*)	17210	NM_008562.3	3498	
MYC (*Homo sapiens*)	4609	NM_002467.3	2377	
Myc (*Mus musculus*)	17869	NM_010849.4	2399	Now multiple transcript variants
Pcna (*Mus musculus*)	18538	NM_011045.2	1260	
Rpa3 (*Mus musculus*)	68240	NM_026632.2	609	
Trp53 (*Mus musculus*)	22059	NM_011640.1	1782	Now multiple transcript variants

Table S2. Tiled transcripts

Summary of NCBI Entrez Gene IDs, accession numbers and length of tiled mRNAs. For each of the indicated transcripts, every possible shRNA was designed and incorporated into the shRNA-Sensor library. Comment column includes notes on changes in transcript annotation that have occurred since the design of the tiling constructs.

Table S3 – Sensor scoring system

eScore: shRNA enrichment Score for individual replicates

Behavior of shRNA	Criteria	eScore
Strongly enriched	S7 / V >5	10
Enriched	S7 / V >2	7
Slightly enriched	S7 / V >1	3
Retained	S7 / V >0	1
Strongly depleted	S3 = 0 and S0 >100	0.01
Depleted	All remaining cases	0.1

Sensor Score: Integration of eScores over all replicates

$$\text{Score} = \sqrt{(\text{eScore S7A}_1 * \text{eScore S7A}_2)} * \sqrt{(\text{eScore S7B}_1 * \text{eScore S7B}_2)}$$

If S9 = 0 in any replicate, Score = Score / 10

Table S3. Sensor scoring system

Scores attributed to each shRNA are based on their read number in the plasmid pool (V) or the library pool at Sort 0 (S0), 3 (S3), 7 (S7) and 9 (S9). Read numbers for the plasmid pool are means of technical duplicates. After transduction into reporter cells two entirely independent biological replicates (A and B) were analyzed throughout the experiment. For Sort 7, each of the two biological replicates was sequenced in technical duplicates (A_1/A_2 and B_1/B_2). Intermediate enrichment scores (eScore) are calculated according to the given criteria, separately for all four S7 replicates. The final integrated Sensor Score is calculated based on these 4 eScores using the given formula, and divided by 10 if an shRNA is absent in any replicate at Sort 9.

Table S4 – Sensor-identified potent shRNAs

Trp53.393
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACCCTGTCATCTTTTGTCCCTTTAGTGAAGC
CACAGATGTAAAGGGACAAAAGATGACAGGGGTGCCTACTGCCTCGGAATTC

Trp53.703
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCCGGGTGGAAGGAAATTTGTATAGTGAA
GCCACAGATGTATACAAATTTCCTTCCACCCGGATGCCTACTGCCTCGGAATTC

Trp53.814
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCCACTACAAGTACATGTGTAATAGTGAAG
CCACAGATGTATTACACATGTACTTGTAGTGGATGCCTACTGCCTCGGAATTC

Bcl2.906
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCGCACAGGAATTTTGTTTAATATAGTGAAG
CCACAGATGTATATTAAACAAAATTCCTGTGCATGCCTACTGCCTCGGAATTC

Bcl2.1132
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCGACTGATATTAACAAAGCTTATAGTGAAG
CCACAGATGTATAAGCTTTGTTAATATCAGTCTTGCCTACTGCCTCGGAATTC

Bcl2.1420
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGACCGGGAGAACAGGGTATGATATAGTGAA
GCCACAGATGTATATCATACCCTGTTCTCCCGGCTGCCTACTGCCTCGGAATTC

Mcl1.1334
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAAAGAGTCACTGTCTGAATGAATAGTGAA
GCCACAGATGTATTCATTCAGACAGTGACTCTTCTGCCTACTGCCTCGGAATTC

Mcl1.1792
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAAACAGCCTCGATTTTTAAGAATAGTGAAG
CCACAGATGTATTCTTAAAAATCGAGGCTGTTCTGCCTACTGCCTCGGAATTC

Mcl1.2018
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCGGACTGGTTATAGATTTATAATAGTGAAG
CCACAGATGTATTATAAATCTATAACCAGTCCATGCCTACTGCCTCGGAATTC

Myc.1888
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAGAAACGACGAGAACAGTTGAATAGTGAA
GCCACAGATGTATTCAACTGTTCTCGTCGTTTCCTGCCTACTGCCTCGGAATTC

Myc.1891
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCACGACGAGAACAGTTGAAACATAGTGAA
GCCACAGATGTATGTTTCAACTGTTCTCGTCGTTTGCCTACTGCCTCGGAATTC

Myc.2105
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCTGCCTCAAACTTAAATAGTATAGTGAAG
CCACAGATGTATACTATTTAAGTTTGAGGCAGTTGCCTACTGCCTCGGAATTC

hMYC.1702
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGATGACCAGATCCCGGAGTTGGATAGTGAA
GCCACAGATGTATCCAACTCCGGGATCTGGTCACTGCCTACTGCCTCGGAATTC

hMYC.1831
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGAGAAACGACGAGAACAGTTGAATAGTGAA
GCCACAGATGTATTCAACTGTTCTCGTCGTTTCCTGCCTACTGCCTCGGAATTC

hMYC.1834
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCACGACGAGAACAGTTGAAACATAGTGAA
GCCACAGATGTATGTTTCAACTGTTCTCGTCGTTTGCCTACTGCCTCGGAATTC

Rpa3.50
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCAACGCCAGCATGTTACCACATAGTGAAG
CCACAGATGTATGTGGTAACATGCTGGCGTTGATGCCTACTGCCTCGGAATTC

Rpa3.145
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCCAGATGGAGAAGGAAAAAATGTAGTGAA
GCCACAGATGTACATTTTTTCCTTCTCCATCTGATGCCTACTGCCTCGGAATTC

Rpa3.455 (*not identified by the Sensor assay, but validated as very potent)
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCGCGACTCCTATAATTTCTAATTAGTGAAG
CCACAGATGTAATTAGAAATTATAGGAGTCGCTTGCCTACTGCCTCGGAATTC

Rpa3.561
CTCGAGAAGGTATATTGCTGTTGACAGTGAGCGCAAAAGTGATACTTCAATATATTAGTGAAG
CCACAGATGTAATATATTGAAGTATCACTTTTATGCCTACTGCCTCGGAATTC

Table S4. Nucleotide sequences of Sensor-identified potent shRNAs

Sequences of the most potent validated Sensor-identified shRNAs (see Figure 3, 4, 5 and Figure S4 for details). Provided are 116 nt XhoI/EcoRI miR30-shRNA fragments. Restriction sites are highlighted in grey, while 97-mer oligonucleotides used for PCR-based single shRNA cloning are underlined. Data for all 18,720 shRNAs analyzed in the Sensor assay is provided in Table S6.

Table S5 – Sensor rules

Key features of shRNA processing and potency (Score >10; 453 shRNAs in total)

Position	Feature	Relevance	Frequency
1-22	9-18 A/U total	Overall A/U content	98%
1-14	Total A/U >7	A/U-rich 5'-side	83%
1-22	(1-14 A/U%)/(15-22 A/U%) >1	Thermodynamic asymmetry	75%
1	U/A	Strand selection	88%
20	No A	Strand selection	89%
13/14	13 A/U or 14 U	Drosha cleavage efficiency	86%
20/21	G	Drosha cleavage efficiency*	45%
20	No C	Drosha cleavage accuracy*	69%
10	A/U	Target cleavage**	67%
1-22	No AAAAAA pattern	Overall RNAi efficiency	99%
1-22	No TTTTT pattern	Overall RNAi efficiency	96%
1-22	No CCCC pattern	Overall RNAi efficiency	98%
1-22	No GGGG pattern	Overall RNAi efficiency	98%

Table S5. Sensor rules

Key features of Sensor-identified potent shRNAs. Provided are prominent features relevant for efficient shRNA processing and/or overall RNAi potency. Frequency, ratio of potent Sensor-identified shRNAs (Score >10; 453 shRNAs) that display the indicated feature. Asterisk, features associated with processing accuracy or efficiency during early steps of shRNA biogenesis, but not enriched in the profile of overall RNAi potency. Double asterisk, target cleavage occurs between nucleotide 10 and 11 of the guide strand (Elbashir et al., 2001).

Supplemental Experimental Procedures

Vectors and library construction

The pSENSOR reporter vector (TRE^{tight}-Neo^{R}-miR30-PGK-Venus-Sensor) was constructed in the pQCXIX retroviral backbone (Clontech) containing a self-inactivating 3'LTR. Using standard cloning techniques, we replaced the CMV-promoter/IRES fragment with the following elements: i) an improved Tet-responsive element promoter (TRE^{tight}, Clontech), ii) a neomycin resistance (Neo^{R}) marker harboring the human miR30 scaffold in its 3'UTR [cloned from pPRIME-CMV-Neo (Stegmeier et al., 2005)], and iii) a PGK promoter driving the expression of yellow-fluorescent Venus (Nagai et al., 2002). We chose Venus because it is a strongly fluorescent, fast maturing YFP molecule that is sufficiently expressed from the PGK promoter. Single shRNAs were cloned into the miR30 context as 110 nt XhoI/EcoRI fragments obtained from existing shRNA vectors or generated by PCR amplification of 97-mer oligonucleotides (Sigma-Aldrich) using 5'miR-XhoI (TACAATACTCGAGAAGGTATATTGCTGTTGACAGTGAGCG) and 3'miR-EcoRI (ACTTAGAAGAATTCCGAGGCAGTAGGCA) primers, 0.05 ng oligonucleotide template and the Platinum Pfx Kit (Invitrogen). Single target sensors were generated by annealing complementary oligonucleotides (Sigma-Aldrich) and cloned into a MluI/BsiWI cloning site in the Venus 3'UTR. Detailed vector maps, sequence information and cloning protocols are available on request.

For constructing the shRNA-Sensor library we designed ~20,000 185-mer oligonucleotides each containing a 101-mer miR30-shRNA fragment (every possible shRNA targeting 9 mammalian transcripts, Table S2), an EcoRI/MluI cloning site, the cognate 50 nt Sensor cassette (Figure S1A) and an 18 nt priming site (CCTAGGATCGACGCGGAC). The library was synthesized in duplicate on customized 55,000 features oligonucleotide arrays (Agilent Technologies), which also contained 17 control shRNA-Sensor oligonucleotides (each synthesized in 30 chip positions to ensure strong representation in the final library). 10 pmol oligonucleotides were provided in a single tube and resuspended in 200 µl H_2O. The final shRNA-Sensor library was constructed in a two-step cloning procedure. In step one, shRNA-Sensor fragments were amplified using the Platinum Pfx Kit (Invitrogen) in 20 parallel 50 µl PCR reactions, each containing 0.2 ng oligonucleotide template, 5 µl 10x Pfx PCR-Buffer, 1.5 µl dNTP (10 mM), 1 µl MgSO4 (50 mM), 0.5 µl Pfx DNA polymerase, 1.5 µl Sens5'Xho primer (TACAATACTCGAGAAGGTATATT-GCTGTTGACAGTGAGCG, 10 mM) and 1.5 µl Sens3'Mfe primer (ATTCATCACAATTGTCC-GCGTCGATCCTAGG, 10 mM). Cycling parameters were 94°C for 5'; 31 cycles of 94°C for 20", 56°C for 20" and 68°C for 25"; 68°C for 5'. PCR products were cloned into XhoI/MfeI of TtNL - a pSENSOR vector lacking the 3'miR30-PGK-Venus fragment. In step two, the amplified intermediate plasmid pool was converted into the final shRNA-Sensor library by cloning of the missing 3'miR30-PGK-Venus fragment into the EcoRI/MluI site between each shRNA and its cognate Sensor. During each cloning step a representation of at least 1000 fold the complexity of the library was maintained. A more detailed cloning protocol is available upon request.

For validation experiments, shRNAs were cloned into constitutive or Tet-regulated shRNA expression vectors. Mouse Myc was PCR amplified from the IMAGE clone 8861953 (Open Biosystems) and BamHI/MluI cloned into MSCV-5xFlag-PGK-Hygro. A second version of this vector with a partial 3'UTR was generated by annealing 200-mer oligonucleotides (IDT) of the immediately adjacent 3'UTR and XhoI/MluI ligation into the aforementioned vector. Detailed cloning strategies and primer sequences are available upon request.

Cell culture and reporter cell lines

NIH3T3s were grown in Dulbecco's Modified Eagle Medium (DMEM, Gibco-Invitrogen) supplemented with 10% bovine calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin (100-Pen-Strep) at 37°C with 5% CO_2. Primary and immortalized mouse embryonic fibroblasts (MEFs), HEK293Ts and amphotropic or ecotropic Phoenix HEK293T cells were grown in DMEM supplemented with 10% fetal bovine serum (FBS) and 100-Pen-Strep at 37°C with 5% CO_2. Mouse Eµ-Myc *p53*$^{-/-}$ lymphoma cells were cultured in 45% DMEM, 45% Iscove's Modified Dulbecco's Medium (IMDM, Gibco-Invitrogen), 10% FBS, 4 mM L-glutamine, 100 mM β-mercaptoethanol and 100-Pen-Strep at 37°C with 7.5% CO_2. Rosa26-rtTA-M2 wild-type (wt) MEFs were isolated from Rosa26-rtTA-M2 transgenic mice (Hochedlinger et al., 2005). Rosa26-rtTA-M2; *p53*$^{-/-}$ double-transgenic MEFs were isolated from embryos obtained after serial cross-breeding of Rosa26-rtTA-M2 mice with *p53*$^{-/-}$ mice (Donehower et al., 1992). RRT MEFs were generated by immortalizing Rosa26-rtTA-M2 wt MEFs through infection with lentiviruses expressing a CMV-driven SV40 large T antigen.
ERC chicken reporter cells were generated through co-infection of DF-1 chicken (*Gallus gallus*) embryonic fibroblasts (Himly et al., 1998) (generously provided by H. Varmus, Memorial Sloan-Kettering Cancer Center) with two VSV-G pseudotyped retroviruses, MSCV-rtTA3-PGK-Puro and MSCV-EcoReceptor-PGK-Hygro. After puromycin (2.5 µg/ml) and hygromycin B (400 µg/ml) double-selection, rtTA3 function of this polyclonal cell line (DF1-ER3) was tested. Subsequently, 25 clones were isolated and individually assessed for i) growth and FACS characteristics, ii) stable and homogeneous morphology, iii) stability in confluent states, iv) stability in freeze/thaw cycles, v) ecotropic infectability and vi) rtTA3 function. ERC cells are derived from the clone that performed best in these tests. ERCs were grown in DMEM supplemented with 10% FBS, 1 mM sodium pyruvate and 100-Pen-Strep at 37°C with 5% CO_2, and frozen in 5% DMSO, 30% FBS and 65% culture medium.
The human leukemia cell lines K-562 and MOLM-13 were cultured in RPMI 1640 (Gibco-Invitrogen) supplemented with 20% FBS and 100-Pen-Strep at 37°C with 7.5% CO_2. Ecotropically infectable Tet-On competent K-562 and MOLM-13 derivatives were generated by transduction of an MSCV-based retrovirus expressing an improved reverse Tet-transactivator (rtTA3) and the ecotropic receptor (MSCV-EcoR-IRES-rtTA3-PGK-Puro) and subsequent puromycin selection. Tet-regulatable shRNAs were induced at Dox concentrations of 1.0-2.0 µg/ml in MEFs, 0.5 µg/ml in ERCs and 1.0 µg/ml in human leukemia cell lines. Dox was replaced every 3 days in culture media. ABT-737 was synthesized by WuXi PharmaTech and dissolved in DMSO to 100x stock solutions.

Retroviral transduction

Cells were transduced as previously described (Serrano et al., 1997); however, target cells were usually infected only once without centrifugation. Chloroquine (25 µM, Sigma-Aldrich) was added to enhance transfection efficiency; polybrene (4 µg/ml, Sigma-Aldrich) was only used when required for higher infection efficiency. For each calcium-phosphate transfection 16 µg plasmid DNA and 6.5 µg helper plasmid or 15 µg plasmid DNA, 5 µg helper plasmid and 5 µg siRNA were used. For virus production of pSENSOR vectors, DGCR8 siRNAs were co-transfected to suppress shRNA biogenesis in the packaging cells and allow for more efficient virus production. Custom-designed DGCR8 siRNAs (21 bp target sequence: CGGGUGGAUCAUGACAUUCCA, Qiagen) were prepared and applied according to the

manufacturer's recommendations. Retroviral co-transduction was carried out as previously described (Zuber et al., 2009). Efficiency of retroviral transduction of fluorescent reporter constructs was assessed 48 h post infection by flow cytometry (Guava EasyCyte, Guava Technologies). Where a specific infection rate was desired, test infections were carried out at different dilution rates and ideal infection ratios deduced. Transduced cell populations were usually selected 48 h after infection, using 2.0-2.5 μg/ml puromycin (Sigma-Aldrich), 100-400 μg/ml hygromycin B (Roche) or 500-1500 μg/ml G418 (Geneticin, Gibco-Invitrogen). To obtain single-copy integrations, cell lines (ERC, RRT MEF, NIH3T3) were initially tested to be homogeneously infectable up to approximately 100% and then infected with an efficiency of 5-20%, guaranteeing <2% cells with multiple integrations (Figure S2F).

Protein expression analysis

Validation of Sensor-identified shRNA knockdown efficiency was assessed at single copy genomic integration in all cases. Unintended changes in protein levels due to Dox toxicity or shRNA expression were controlled for by the use of neutral shRNAs that potently suppress their target (sh.Luci.1309 or other suitable controls), instead of potentially non-processed scrambled control sequences. For immunoblotting, samples were lysed in Laemmli buffer, separated by SDS-PAGE and transferred to PVDF (Immobilon-P, Millipore) membranes. Adriamycin was used at 50 μg/ml to induce p53 expression 4-12 h prior to cell harvest. Cebpa expression was assessed in NIH3T3s stably overexpressing Cebpa (MSCV-Cebpa-PGK-Hygro). We used antibodies against Bcl2 (BCL/10C4, 1:1000; BioLegend), Braf (F-7/sc-5284, 1:1000; Santa Cruz Biotechnology), Cebpa (sc-61, 1:1000; Santa Cruz Biotechnology), Cdkn2a (5-C3-1, 1:500; Upstate), Mcl1 (600-401-394, 1:5000; Rockland), Pten (6H2.1, 1:1000; Cascade BioScience), Rpa3 (M-18/sc-46508, 1:100; Santa Cruz Biotechnology), Trp53 (IMX25, 1:1000; Vector Laboratories), β-actin (AC-15, 1:20,000; Sigma-Aldrich) and α-tubulin (B-5-1-2, 1:5000; Sigma-Aldrich). Flag-tag labeled proteins were detected with an anti-FLAG M2 antibody (F1804, 1:1000; Sigma-Aldrich). Secondary antibodies were anti-goat (705-035-003; Jackson ImmunoResearch), anti-mouse (NA931V; GE Healthcare UK), anti-rabbit (NA934V; GE Healthcare UK) and anti-rat (NA935V; GE Healthcare UK). Protein densitometry was performed using ImageJ software according to method 1 outlined at www.lukemiller.org/journal/2007/08/quantifying-western-blots-without.html, and knockdown levels normalized to the respective loading control.

Flow cytometry

All fluorescence activated cell sorting (FACS) procedures were carried out on a FACSAria II (BD Biosciences). To functionally deconvolute complex shRNA-Sensor libraries, we developed a Sensor Ping-Pong sorting strategy based on sorting cells according to reference populations. The Top5 gating reference contained the shRNAs Braf.5053, Luci.1309, Pten.1523, Pten.2049 and Trp53.1224. The Bottom5 gating reference contained the shRNAs Braf.3750, Cdkn2a.218, Cebpa.577, Cebpa.898 and Trp53.1647. ERC reporter cells were infected with the pSENSOR shRNA-Sensor libraries at single copy and sorted either after treatment with Dox and G418 (500 μg/ml) for 6-7 days or after Dox and G418 withdrawal for 6-7 days. For each sort a representation of 1000 fold the complexity of the sorted pool was maintained. Hence, at Sort 1 $20*10^6$ infected ERC cells were sorted for each replicate. Representative sorting gates are shown in Figure S2H; a detailed protocol of the Sensor Ping-Pong assay is available upon request.

Deep sequencing analysis and quantification of shRNA representation

To profile the library composition over time, genomic DNA was isolated from >20 million cultured cells (>1000 fold library representation) using standard techniques. Deep sequencing template libraries were generated by PCR amplification of shRNA guide strands using primers that tag standard Solexa/Illumina adaptors to the product (p7+loop, CAAGCAGAAGACGGCATACGATAGTGAAGCCACAGATGTA; p5+miR3', AATGATACGGCGACCACCGACTAAAGTAGCCCCTTGAATTC). For each sample, DNA from at least 10 million cells was used as template in multiple parallel 50 µl PCRs, each containing 1-2 µg template, 5 µl 10x buffer, 5 µl dNTP (2 mM each), 1.5 µl of each primer (10 µM) and 0.5 µl AmpliTaq Gold (Applied Biosystems). Reactions were run using the following cycling parameters: 95°C for 10'; 35 cycles of 95°C for 20", 52°C for 30" and 72°C for 25"; 72°C for 7'. 117 nt PCR products were combined for each sample, precipitated and purified on a 2% agarose gel (QIAquick gel extraction kit, Qiagen). Approximately 8 pM of each sample were loaded on a separate flow cell lane and sequenced using a custom primer reading reverse into the guide strand (miR30EcoRISeq, TAGCCCCTTG-AATTCCGAGGCAGTAGGCA). Individual lanes (each corresponding to 1 sample) yielded ~5-15 million reads. Sequence processing was performed using a customized Galaxy platform (Taylor et al., 2007) available at CSHL. For each shRNA and condition the number of completely matching sequences was determined, normalized to the total read number per lane and imported into a database for further analysis (Access 2003, Microsoft). As final readout for shRNA potency we applied either a directly quantitative measurement of shRNA enrichment or a semi-quantitative integrated score. The quantitative readout is defined as the enrichment in reads for a given shRNA in Sort 7 compared to the plasmid pool (reads $shRNA_i^{Sort\ 7}$/reads $shRNA_i^{Library}$). This value multiplied between the two biological duplicates gives the product enrichment (ProdEn). The integrated readout (Score) takes initial representation, consistency between replicates and representation of shRNAs during Sensor Ping-Pong sorts into account. This logistic function plateaus at 100 for highly potent shRNAs and at 0 for non-functional shRNAs. The calculation of the Score is defined in Table S3.

Cloning and sequencing of small RNA libraries

RNA was purified, cloned and sequenced on the Solexa/Illumina platform, as previously described (Malone et al., 2009). In brief, approximately $10*10^6$ human HEK293T or chicken ERC cells transduced with the pSENSOR library were homogenized in TRIzol (Invitrogen), followed by chloroform and phenol:chloroform:isoamyl alcohol (125:24:1, pH 4.5, Ambion) purification, precipitation in isopropanol and resuspension in RNase-free water (Gibco-Invitrogen). Samples for pri-miRNA quantification were then DNase I (Roche) treated and again phenol:chloroform:isoamyl alcohol purified, precipitated in isopropanol and resuspended in water. Approximately 40 µg of total RNA was run on a 12% urea polyacrylamide gel (SequaGel, National Diagnostics), where either siRNA/miRNA sized small RNAs (18-26 nt) or pre-miRNAs (50-70 nt) were selected for cloning and sequencing. Mature small RNAs were ligated to the 3' adaptor 3link_moD and the 5' adaptor Solexa_linker, while pre-miRNAs were ligated to the 3' adaptor 3link_moD only. The 3' adaptor 3link_moD was a fully activated, 5'-adenylated and 3'-blocked oligonucleotide (5'-rAppCTGTAGGCACCATCAATCGT/3ddA-3', IDT), while the 5' adaptor Solexa_linker was an RNA oligonucleotide (5'-rArCrArCrUrCrUrUrUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrC-3', IDT). Mature small RNA libraries were generated using the p7+5link (5'-CAAGCAGAAGACGGCATACGACACTCTTTCCCTACACG-3') and p5+moD (5'-AATGATA-CGGCGACCACCGATAGCCCCTTACGATTGATGGTGCCTACAG-3') primers, precursor miRNA libraries were obtained using the miR30-specific p7+loop (5'-CAAGCAGAAGACGGCA-TACGATAGTGAAGCCACAGATGTA3') and p5+moD primers, and primary miRNA libraries were produced using the miR30-specific p7+loop and miR30-specific p5+miR3 (5'-AATGAT-

ACGGCGACCACCGACTAAAGTAGCCCCTTGAATTC-3') primers. All small RNA libraries were quantified by real-time PCR and sequenced using custom primers on the Solexa/Illumina Gene Analyzer II platform. Mature miRNA libraries were sequenced using the SBS3 primer (5'-ACACT-CTTTCCCTACACGACGCTCTTCCGATC-3') reading forward into the siRNA/miRNA sequence, while pre-miRNA libraries and pri-miRNA libraries were sequenced with the p5moDSeq (5'-TAGCCCCTTACGATTGATGGTGCCTACAG-3') and miR30EcoRISeq (see above) primers, respectively, reading reverse into the guide strand. Individual lanes (each corresponding to 1 sample) yielded ~15 million reads. Sequence processing, including trimming of the 5' adapter (mature miRNA sequences) or loop (pri- and pre-miRNA sequences) and collapsing of unique sequences was performed using a customized Galaxy platform (Taylor et al., 2007) available at CSHL, followed by alignment to the Sensor library using blastn (Camacho et al., 2009). For each shRNA and condition the number of completely matching sequences was determined, normalized to the total read number per lane, and imported into a database for further analysis (Access 2003, Microsoft). Analysis of processing features was carried out considering only correctly cut 22 nt guide and passenger strands, while for the examination of Drosha cleavage accuracy, sequences of 19-25 nt (+/- 3 nt around the canonical cleavage site of the respective shRNA) were considered. A detailed cloning protocol is available upon request.

Statistical analysis

Linear dependence of variables was evaluated using the Pearson product-moment correlation coefficient r. The significance of biases in nucleotide frequency were assessed by comparing the distribution of all four nucleotides in the various samples with the theoretical distribution in the total population, thus normalizing for nucleotide frequency deviations within the total set of 18,720 shRNAs. For each nucleotide position, the goodness-of-fit was calculated using Pearson's χ^2 test and p-values were derived (chitest, Excel 2003, Microsoft). The results were adjusted for inflation of the α level due to large sample sets using the Šidák correction, $\alpha_{tw} = 1 - (1 - \alpha_{fw})^{1/n}$, where the *familywise* $\alpha_{fw} = 0.05$ or 0.01 (as noted in the text), and n = the number of shRNAs in the sample. For example, for n = 453 (see Figure 6C) and $\alpha_{fw} = 0.01$, the Šidák corrected *testwise* $\alpha_{tw} = 2.22*10^{-5}$. Hence, in this example, only nucleotide biases with $p < 2.22*10^{-5}$ for Pearson's χ^2 test are judged significant for an α level of 0.01.

Supplemental References

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. Nature *356*, 215-221.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

Himly, M., Foster, D.N., Bottoli, I., Iacovoni, J.S., and Vogt, P.K. (1998). The DF-1 chicken fibroblast cell line: transformation induced by diverse oncogenes and cell death resulting from infection by avian leukosis viruses. Virology *248*, 295-304.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic Expression of Oct-4 Blocks Progenitor-Cell Differentiation and Causes Dysplasia in Epithelial Tissues. Cell *121*, 465-477.

Huesken, D., Lange, J., Mickanin, C., Weiler, J., Asselbergs, F., Warner, J., Meloon, B., Engel, S., Rosenberg, A., Cohen, D*., et al.* (2005). Design of a genome-wide siRNA library using an artificial neural network. Nature Biotechnology *23*, 995-1001.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Markham, N.R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol *453*, 3-31.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nature Biotechnology *20*, 87-90.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593-602.

Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R*., et al.* (2005). Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet *37*, 1281-1288.

Stegmeier, F., Hu, G., Rickles, R.J., Hannon, G.J., and Elledge, S.J. (2005). A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. PNAS *102*, 13212-13217.

Taylor, J., Schenck, I., Blankenberg, D., and Nekrutenko, A. (2007). Using galaxy to perform large-scale interactive data analyses. Curr Protoc Bioinformatics *Chapter 10*, Unit 10 15.

Vert, J.P., Foveau, N., Lajaunie, C., and Vandenbrouck, Y. (2006). An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinformatics *7*, 520.

Zuber, J., McJunkin, K., Fellmann, C., Dow, L.E., Taylor, M.J., Hannon, G.J., and Lowe, S.W. (2011). Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. Nat Biotechnol *29*, 79-83.

Zuber, J., Radtke, I., Pardee, T.S., Zhao, Z., Rappaport, A.R., Luo, W., McCurrach, M.E., Yang, M.M., Dolan, M.E., Kogan, S.C.*, et al.* (2009). Mouse models of human AML accurately predict chemotherapy response. Genes Dev *23*, 877-889.