APPLICATION NOTE

horizon™
INSPIRED CELL SOLUTIONS

Development of the Dharmacon™ SMARTvector™ Lentiviral shRNA microRNA-adapted scaffold and algorithm for functional shRNAs

Annaleen Vermeulen, Maren Mayer Gross, Shawn McClelland, and Melissa L. Kelley
Horizon Discovery (formerly Dharmacon), Lafayette, CO, USA

Introduction

RNA interference (RNAi) is an endogenous, post-transcriptional gene silencing pathway that utilizes small, noncoding RNAs to target specific mRNA transcripts for degradation. Target-gene knockdown by RNAi can be mediated by synthetic short interfering RNAs (siRNAs) introduced into cells by transfection, or by short hairpin RNA (shRNA) expression constructs introduced into cells by plasmid transfection or viral transduction (Figure 1A).

When RNAi was gaining popularity for gene knockdown experiments, it became apparent that different sequences varied greatly in their potency to knock down gene expression. Computational algorithms were developed to predict siRNA function[1-4] and have increased the likelihood of selecting functional siRNA sequences. Although siRNA algorithms have been used to design shRNA for expression constructs, shRNA function cannot be accurately predicted based on siRNA data[5], presumably due to the additional processing requirements of the shRNA compared to no processing needed for the synthetic duplexed siRNA. While siRNA enters the RNAi pathway by binding the RNA-induced silencing complex (RISC) in the cytoplasm, shRNA requires expression in the nucleus, processing by the Microprocessor complex, export to the cytoplasm, and further processing by the Dicer-containing complex before being finally loaded into RISC (Figure 1A). Additionally, when shRNA is expressed from a vector, the effective intracellular concentration can be lower than when siRNA is transfected[6,7], which means that shRNA sequences must be highly functional at lower concentrations.

In the last decade there have been reports of shRNA functional algorithm development, with relatively limited training data sets and even smaller algorithm validation test data sets[8-10]. In 2006, the original Dharmacon shRNA algorithm was developed to select functional shRNAs for a microRNA-adapted scaffold, also with a smaller data set. More recently, a computational algorithm has been developed using a large data set based on reporter readout[11]. Here we describe the development of an advanced shRNA algorithm based directly on a large set of native mRNA knockdown data using the patented microRNA-based scaffold (US 9,284,554 B2), Dharmacon SMARTvector Lentiviral shRNA scaffold.

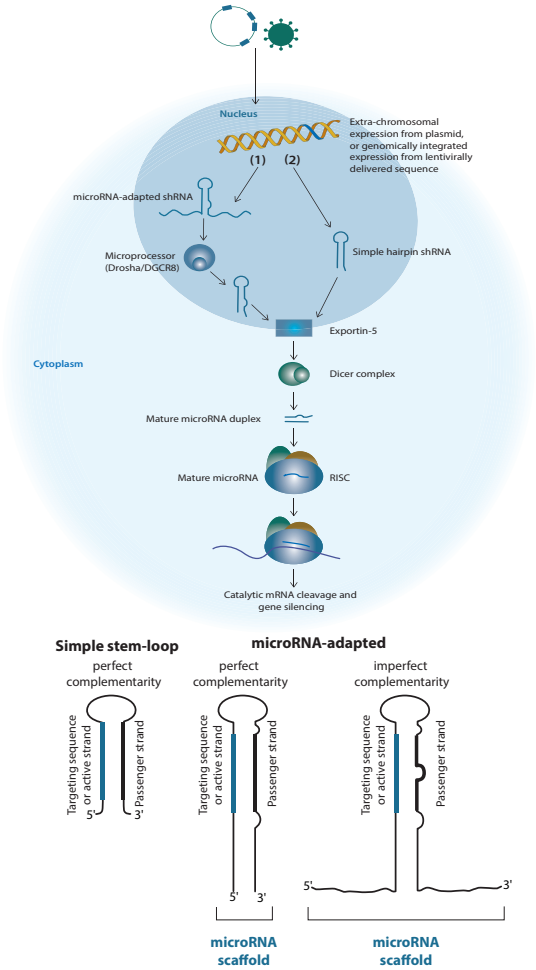


Figure 1. Borrowing the endogenous microRNA pathway to induce RNAi. **A.** shRNA approaches include the introduction of plasmid-based or viral-based vectors expressing the silencing sequences embedded in a microRNA-adapted scaffold (1) or simple hairpin shRNA (2). Expressed shRNAs (blue) enter the endogenous pathway at an early stage and are efficiently processed into potent silencing molecules using the endogenous microRNA mechanim, which leads to target mRNA cleavage (purple) and gene knockdown. **B.** Types and structures of expressed shRNAs used for targeted gene knockdown (see text for details).

Results & Discussion

microRNA-adapted scaffold development

There are two types of shRNA scaffolds, simple hairpin or microRNA-based. The simple hairpin has a stem-loop structure where the active and passenger strands are complementary and expressed from a Pol III promoter, such as U6 and H1 (Figure 1B). The microRNA-based, or microRNA-adapted, scaffold has a longer stem-loop structure based on a naturally occurring microRNA where the active and passenger strands have perfect or imperfect complementarity (*e.g.*, mismatches, G:U wobbles), and may also contain essential single-strand sequence, or flanking sequence, upstream and downstream of the stem-loop (Figure 1B). Because the microRNA-adapted scaffold sequence is longer than the simple hairpin, it is expressed from Pol II promoters, such as CMV and EF1α; this further enables the expression of a fluorescent reporter and/or antibiotic resistance marker on the same transcript as the shRNA, simplifying experimental optimization. It has been demonstrated that simple hairpin scaffolds using PolI III promoters can lead to cellular toxicity while the same targeting sequence expressed from a microRNA-based scaffold can reduce this toxicity[12]. In addition, the over-expression of simple-hairpin scaffolds have been shown to impact microRNA processing and result in off-target effects[13-15]. We therefore decided to proceed with a microRNA-based scaffold approach.

The ideal microRNA-based scaffold is required to be processed efficiently, to accept foreign gene-targeting sequences while maintaining accurate processing, and to maintain good functionality from the active strand while minimizing passenger strand activity. Ten endogenous microRNAs were tested to determine which had the most efficient processing of the active strand and highest knockdown function. It was found that different microRNAs resulted in different levels of knockdown, presumably because of varying levels of processing efficiency (Figure 2A). Five of the highly functional microRNA-based scaffolds were selected to determine the relative function of the active and passenger strands. The data showed that miR-126 and miR-30a had high function for both active and passenger strands, which could lead to off-target effects by the passenger strand, while miR-26b and miR-196a-2 had high functionality for only the active strand and miR-204 had the highest active strand functionality and moderate passenger strand functionality (Figure 2B). These three microRNAs were selected to examine their ability to accept foreign sequence while maintaining high efficiency knockdown.

To assess acceptance of foreign sequence, a "walk" was performed by designing shRNAs for miR-26b, miR-196a-2 and miR-204 targeting every two nucleotides along human GAPDH while maintaining secondary structure (imperfect complementarity) between the active and passenger strands similar to the respective endogenous microRNA. Additionally, these same GAPDH shRNA sequences were designed for miR-196a-2 where the hairpin contained perfect complementarity between the active strand and passenger strand (no secondary structure). For each scaffold, 48 shRNAs were cloned, sequence confirmed and transfected into HeLa cells to measure knockdown of GAPDH in a dual-luciferase reporter construct. miR-196a-2 scaffold with shRNAs having imperfect complementarity was found to have a greater number of functional shRNAs than with perfect complementarity and compared to the miR-26b and miR-204 scaffolds (Figure 3). Importantly, the same shRNA sequences in the context of different microRNA-adapted scaffolds did not result in the same level of functionality, indicating that a design algorithm must be developed for a specific microRNA scaffold. Therefore, the final SMARTvector scaffold based on miR-196a-2, which includes high functionality of the active strand, minimal function of the passenger strand, and proprietary mismatches and wobbles between the active and passenger strands, was used for algorithm development.

Algorithm development

To design an algorithm that will reliably predict shRNA function, we generated a large training data set by measuring knockdown at the mRNA level. Multiple targeting sequences to 18 genes were cloned into the SMARTvector scaffold and sequence verified. shRNA-expressing plasmids were transfected into HeLa cells. After 24 hours, resazurin was added to examine cellular toxicity and then puromycin was added to select for transfected cells. Cells were harvested 72 hours after transfection and mRNA was isolated for expression analysis using a branched DNA method.

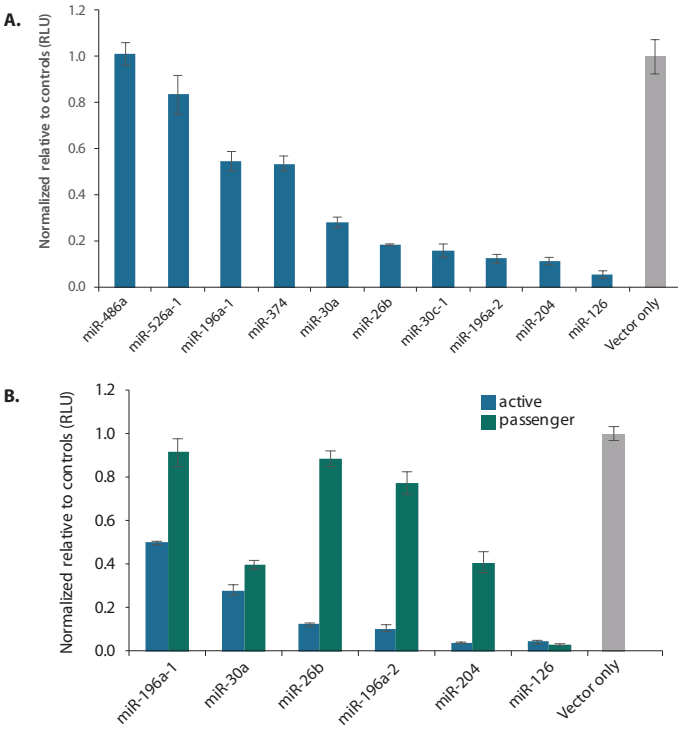


Figure 2. Functionality of endogenous microRNAs is different, including the activity of the passenger strand. A. Ten constructs encoding endogenous microRNA sequences targeting a luciferase reporter gene demonstrate a wide range of knockdown functionality. **B.** Five of these microRNAs were also tested for functionality of the passenger strand, where various levels of gene targeting were observed. The ideal microRNA scaffold has high knockdown efficiency of the active strand and low levels of functionality from the passenger strand. Endogenous microRNAs, including 200-300 bp flanking the stem-loop region, were PCR amplified from genomic DNA isolated from HeLa cells. microRNAs were cloned into a vector for co-expression of the microRNA and GFP under a CMV promoter and sequenced verified. HeLa cells were plated at 10,000 cells/well in 96-well format one day prior to transfection. microRNA-expressing plasmids (60 ng) were co-transfected with a dual luciferase reporter plasmid [psiCheck2 (Promega); 40 ng] containing the active strand-targeting sequence or the passenger strand-targeting sequence in the 3' UTR of Renilla luciferase. Data were normalized to the cloning vector without a microRNA insert (vector only) co-transfected with a psiCheck2 plasmid with the respective targeting sequence, and further normalized to the Firefly luciferase expression to calculate knockdown efficiency.

Knockdown data was filtered to only include data points where viability was not affected and transfection controls passed an appropriate threshold. After filtering, the training set contained 548 shRNA constructs with 18 genes having 9 to 69 shRNAs per gene (Figure 4A) that spanned a range of knockdown efficiencies (Figure 4B). Approximately two thirds of these sequences gave poor knockdown (with greater than 40% residual gene expression) that indicates the necessity for developing an shRNA algorithm (Figure 4B).

Algorithm training was performed using a Naïve Bayes classifier. Features of the shRNA constructs (nucleotide identity by position, dinucleotide identity by position, GC content, binding energy between shRNA strands, among others) and mRNA target (targeting position within the transcript among others) were examined for association with good knockdown. The algorithm was found to have a positive predictive power of 0.48 and true positive rate of 0.60 where 108 shRNA were predicted to be functional and 117 predicted to be nonfunctional. This is a significant improvement over random picking where roughly one in five constructs has acceptable functionality.

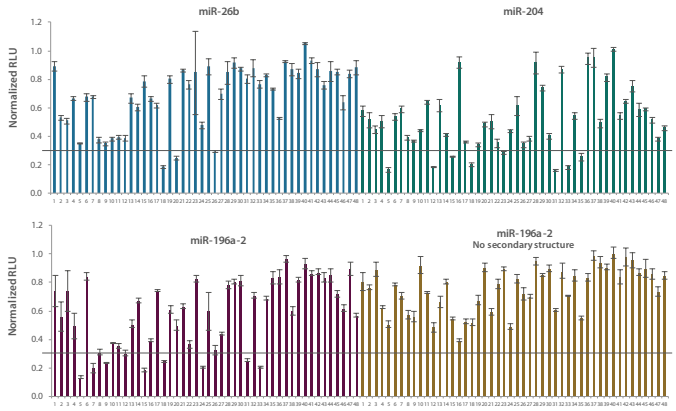


Figure 3. Certain microRNA-adapted scaffolds can accept foreign targeting sequences to consistently result in greater gene knockdown. Three different microRNA-adapted scaffolds (maintaining secondary structure between targeting and passenger strand similar to microRNA structure) showed differences in knockdown when accepting foreign targeting sequences. Additionally, removing secondary structure between the active and passenger strands for miR-196a-2 results in loss of knockdown function compared to maintaining secondary structure. Line indicates 70% knockdown. shRNA sequences were designed every two nucleotides spanning 142 bp of GAPDH and cloned into a vector for co-expression of the shRNA and GFP under a CMV promoter and sequenced verified. HeLa cells were plated at 10,000 cells/well in 96-well format one day prior to transfection. shRNA-expressing plasmids (60 ng) were co-transfected with a dual luciferase reporter plasmid [psiCheck2 (Promega); 40 ng] containing ~ 150 bp of human GAPDH in the 3' UTR of Renilla luciferase. Data were normalized to firefly luciferase in the same sample and further normalized to control sample (the cloning vector without an shRNA insert co-transfected with a psiCheck2 plasmid) to calculate knockdown efficiency.

A.

Gene target	# shRNA constructs per gene
ACTB	31
AKT1	33
ARAF	12
ARHGAP1	17
B2M	21
CDK2	22
GAPDH	31
GSK3A	29
HPRT1	19
MAP2K2	26
PPIA	9
PPP2CB	20
PPP2R1A	44
PTEN	67
RAC1	17
RB1	69
RHOA	39
XPO5	42
Total	**548**

B.

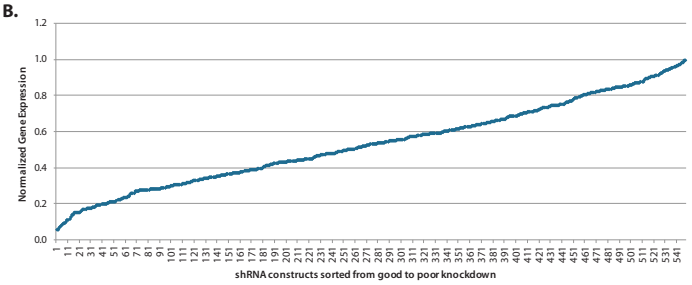


Figure 4. Endogenous mRNA expression data with a wide range of knockdown efficiencies for multiple genes were selected for algorithm development. A. Data collected from 18 target genes and the corresponding number of shRNA constructs per gene that passed quality control were utilized for algorithm development. **B.** Gene knockdown data shown as normalized gene expression sorted from lowest to highest for all 548 shRNA constructs used for algorithm development.

Algorithm validation

To test the algorithm's ability to select functional sequences, the measurement of shRNA function for an independent set of genes is required. We selected a new set of genes with high expression in our test cells. shRNAs were designed for 107 genes with up to eight constructs per gene (Figure 5A). Plasmid transfections and data analysis were performed as for the training data set. A large validation set of 514 constructs had high quality data and 269 showed good functionality (representative data set shown in Figure 5B). This test set has a positive predictive power of 0.52, which confirms the ability of the algorithm to select functional sequences.

A.

Gene target	# of shRNA constructs	Gene target	# of shRNA constructs	Gene target	# of shRNA constructs
CDK9	8	*CDC42*	7	*CASP3*	4
DICER1	8	*AKT1*	7	*SRC*	4
GAPDH	8	*GSK3B*	7	*CSNK2A1*	4
HIF1A	8	*PARP1*	7	*PINK1*	4
BRD2	8	*CDKN1B*	7	*PPP2CA*	4
ILK	8	*CHEK2*	7	*RAF1*	4
MCL1	8	*EIF2AK2*	7	*ROCK2*	4
PRKDC	8	*AURKA*	7	*BRAF*	3
CSK	8	*MAP2K2*	7	*GSK3A*	3
ERBB2	8	*MYC*	7	*MAPK8*	3
MDM2	8	*PLK1*	7	*PIK3C3*	3
ROCK1	8	*CDKN1A*	7	*ADAM17*	3
CDK2	8	*ULK1*	7	*ATM*	3
CTNNB1	8	*CDC20*	6	*TBK1*	3
DNMT1	8	*ITGB1*	6	*MAP3K14*	3
PTEN	8	*MAP2K1*	6	*TNFRSF1A*	3
RPS6KA3	8	*PPARG*	6	*AKT2*	2
CDC2	8	*PRKCA*	6	*FRAP1*	2
TRIM28	8	*AXL*	6	*FYN*	2
MAP4K4	8	*PIK3R1*	6	*MET*	2
RAC1	8	*ATR*	6	*CXCR4*	2
CHEK1	8	*MAPK3*	6	*MAP3K7*	2
IGF1R	8	*TP53*	6	*RB1*	2
APOA1	8	*ABL1*	6	*LYN*	2
EPHA2	8	*AURKB*	6	*PPIB*	2
MAPK14	8	*PLK4*	6	*RHOA*	2
PRKCZ	8	*PRKAA1*	6	*STK33*	1
PTK2	8	*TGFBR1*	5	*RNASEN*	1
CDK5	7	*PAK1*	5	*TRIM24*	1
CHUK	7	*PRKCD*	5	*WEE1*	1
MAPK1	7				

B.

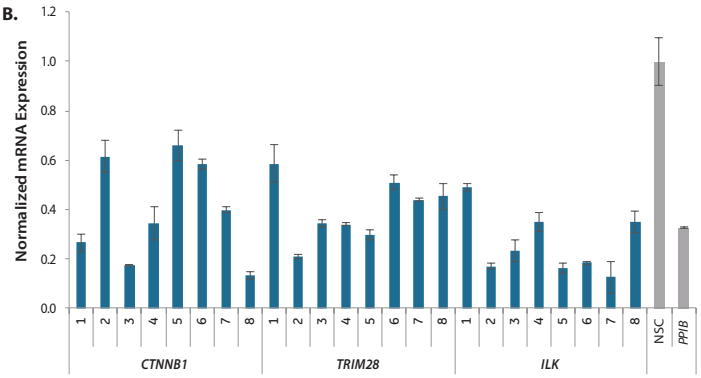


Figure 5. Endogenous mRNA knockdown data for a large, independent target gene set was utilized for algorithm validation. A. The algorithm was used to design multiple shRNA constructs for 91 target genes. **B.** Knockdown data for plasmid transfections are shown for representative target genes (*CTNNBI*, *TRIM28* and *ILK*) as normalized mRNA expression data. Gene expression was measured by branched DNA assay and normalized to the non-silencing control (NSC). Knockdown of a positive control (*PPIB*) was included to verify good transfection efficiency.

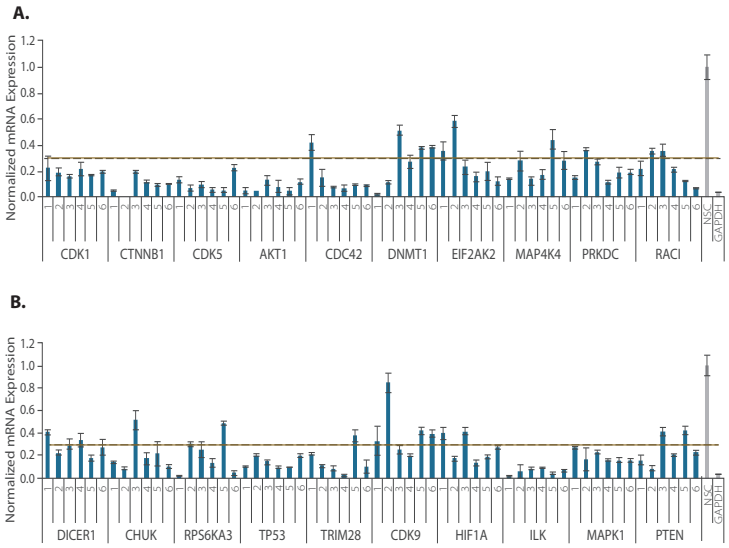


Figure 6. High levels of gene knockdown using SMARTvector Lentiviral shRNA
Lentiviral transductions of algorithm-designed shRNAs (6 per gene) targeting 20 different genes show very good gene knockdown. Gene expression was measured by branched DNA assays and normalized to the non-silencing control (NSC). Knockdown of a positive control (*GAPDH*) was included to verify good transduction efficiency. 70% knockdown is indicated by the dark brown line.

Product validation

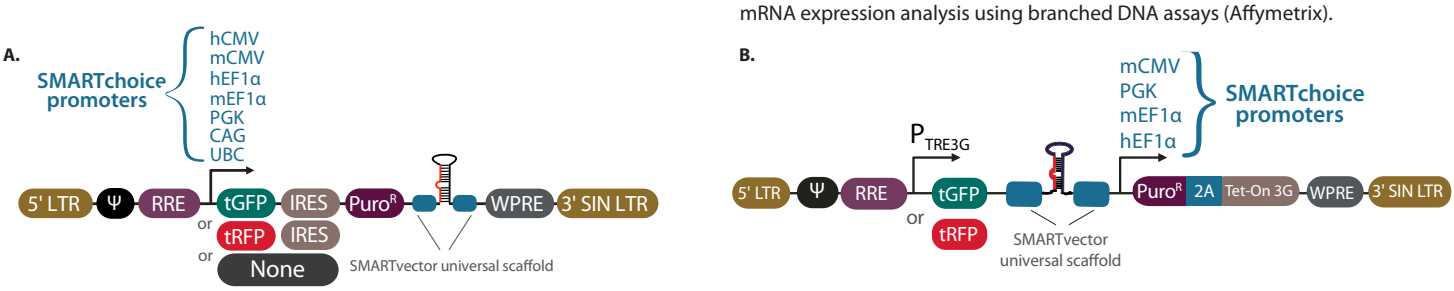
This SMARTvector algorithm was used to generate shRNA product designs. RefSeq version 63 was used to design six shRNAs per gene for 20 genes. Since the SMARTvector shRNA vector is optimized for lentiviral transduction, for product validation the lentiviral particles were transduced into A549 cells and mRNA knockdown was analyzed 72 hours post-transduction using branched DNA analysis. Of 119 shRNA constructs tested, 98 (82%) resulted in greater than 70% gene knockdown (Figure 6).

Summary

SMARTvector shRNA designs incorporate several features to enable good experimental design and robust knockdown. SMARTvector shRNAs are based on an optimized microRNA-adapted scaffold that ensures preferential loading of the active strand for reduced off-target effects and efficient processing for improved gene silencing. This patented SMARTvector shRNA scaffold was used to develop a design algorithm to predict functional constructs, which was derived from over 500 endogenous mRNA knockdown data points and selects shRNAs with robust function based on proprietary criteria, such as position-dependent nucleotide preferences, secondary structure, thermodynamic stability, and transcript target features. Our algorithm allows for design of multiple shRNA designs per gene for redundancy and confidence in experimental results. The SMARTvector shRNA sequences are designed to target all gene transcripts and are rigorously filtered to minimize off-targeting of other coding and noncoding genes. The SMARTvector shRNA platform includes customizable options such as several constitutive and inducible promoter options ensuring functional knockdown in most human, mouse and rat cell types, as well as fluorescent reporter options for compatibility with various phenotypic assays (Figure 7). These constructs are also available as pooled lentiviral shRNA libraries for screening hundreds to thousands of genes. The SMARTvector shRNA products carry a functional guarantee. When you purchase three SMARTvector lentiviral shRNAs to the same protein-coding gene target using the optimal SMARTchoice promoter for your cell type, at least one of the shRNA constructs will reduce target mRNA levels by 70% or more when used with the vector matched non-targeting control and GAPD or PPIB positive control.

Materials and methods

Algorithm development: Targeting sequences (1024) were cloned into the SMARTvector plasmid and sequence verified. Glycerol stocks were generated and plasmid DNA was isolated and quantified for each construct. Plasmid DNA was transfected into HeLa cells using 0.4 µL TurboFect transfection reagent (Thermo Scientific) and 200 ng/well plasmid in a 96-well plate. Resazurin was added 24 hours post-transfection to examine toxicity and 3 µg/mL puromycin was added for selection. Cells were harvested 72 hours after transfection for mRNA expression analysis using branched DNA assays (Affymetrix).



Vector Element	Utility
5' LTR	5' Long Terminal Repeat is necessary for lentiviral particle production and integration of the construct into the host cell genome
Ψ	Psi packaging sequence allows lentiviral genome packaging using lentiviral packaging systems
RRE	Rev Response Element enhances titer by increasing packaging efficiency of full-length lentiviral genomes
TRE3G	Inducible promoter with Tetracycline Response Elements which is activated by the Tet-On 3G protein in the presence of doxycycline
tGFP or tRFP	TurboGFP or TurboRFP reporter for visual tracking of transduction. Expression upon doxycycline induction in the inducible format
IRES	Internal Ribosomal Entry Site allows expression of fluorescent reporter and puromycin resistance gene in a single transcript
SMARTvector universal scaffold	Optimized proprietary scaffold based on native primary microRNA in which gene-targeting sequence is embedded
Puro^R	Puromycin resistance gene permits antibiotic selection of transduced cells
2A	Self-cleaving peptide that enables the expression of both Puro^R and Tet-On 3G transactivator from a single RNA pol II promoter
Tet-On 3G	Encodes the doxycycline-regulated transactivator protein, which binds to TRE3G only in the presence of doxycycline
WPRE	Woodchuck Hepatitis Post-transcriptional Regulatory Element enhances transgene expression in target cells
3' SIN LTR	3' Self-inactivating Long Terminal Repeat for generation of replication-incompetent lentiviral particles

Figure 7. Vector elements for the constitutive and inducible expression of shRNA. A. SMARTvector Lentiviral shRNA vector elements. **B.** SMARTvector Inducible shRNA Lentiviral vector elements.

Data was retained where the viability was more than 90% relative to untransfected cells at 24 hours and less than 150% relative to untransfected cells after 72 hours. Plates were removed from analysis where the control gene knockdown (*PPIB*) was less than 65%, indicating inefficient transfection. In addition, data were removed from analysis where relative expression data was not between 0 and 1. As a result, 548 shRNA sequences had high quality data on which to train the algorithm.

Algorithm selection and training: Several rounds of Classifier performance were assessed in order to choose the most predictive machine learning system. Different functional knockout definitions were also evaluated during this process. Bayesian classifiers gave the top performance versus random. Algorithm training was performed using a Naïve Bayes classifier with functional knockout defined as 60% knockdown. Feature selection of our chosen classifier (and most of the top performing classifiers that we assessed) included chi-squared-selected attributes from a long list of potential features derived from both the shRNA constructs and mRNA targets. These features and weightings from the final algorithm training allow the algorithm to select the shRNA that are most likely to be functional against any given mRNA target.

Algorithm validation: The validation shRNA test set was generated by cloning eight shRNA per gene for 107 genes (Figure 5A). Cloning, sequencing, and transfection were performed as in Algorithm development.

Product validation: shRNAs were designed against human NCBI Entrez genes of RefSeq 63. shRNAs for each gene are common to all transcript variants. Protein coding shRNAs were targeted to the ORF and 3' UTR. Design features for specificity were incorporated such that all 19-mer designs were analyzed for homology against coding and noncoding transcripts; shRNA designs aligning with 18 or 19 nucleotides perfectly complimentary to an unintended target transcript were discarded, shRNA designs aligning to an unintended target transcript with 17 mismatches were disfavored in the algorithm, and shRNA designs that contain a known microRNA seed or high frequency seeds [16] were filtered out when possible.

Designs for 20 genes (six shRNA per gene) were cloned into the SMARTvector scaffold containing the mouse CMV promoter and TurboGFP™ (Evrogen, Moscow, Russia) reporter options. These shRNAs were packaged into lentiviral particles and transduced at a multiplicity of infection (MOI) of 20 TU/mL into A549 cells seeded at 7,500 cells per well one day prior to transduction in a 96-well plate. Gene knockdown was analyzed 72 hours post-transduction using branched DNA analysis. No antibiotic selection was applied.

References

1. A. Reynolds *et al*., Rational siRNA design for RNA interference. *Nature Biotechnology* **22**, 326-330 (2004).
2. S. Takasaki, *et al*., An effective method for selecting siRNA target sequences in mammalian cells. *Cell Cycle* **3**, 790-795 (2004).
3. K. Ui-Tei *et al.*, Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* **32**, 936-948 (2004).
4. A. C. Hsieh *et al.*, A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res* **32**, 893-901 (2004).
5. D. J. Taxman *et al.*, Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnol* **6**, 7 (2006).
6. K. Berns *et al*., A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-437 (2004).
7. P. Paddison *et al*., A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427-431 (2004).
8. L. Li, *et al*., Defining the optimal parameters for hairpin-based knockdown constructs. *RNA* **13**, 1765-1774 (2007).
9. H. Zhou, X. Zeng, Energy profile and secondary structure impact shRNA efficacy. *BMC Genomics 10 Suppl 1*, **S9** (2009).
10. O. Matveeva *et al.*, Comparison of approaches for rational siRNA design leading to a new efficient and transparent method. *Nucleic Acids Res* **35**, e63 (2007).
11. S. R. Knott *et al.*, A computational algorithm to predict shRNA potency. *Molecular cell* **56**, 796-807 (2014).
12. J. L. McBride *et al.*, Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5868-5873 (2008).
13. S. T. Baek *et al.*, Off-target effect of doublecortin family shRNA on neuronal migration associated with endogenous microRNA dysregulation. *Neuron* **82**, 1255-1262 (2014).
14. S. Beer *et al.*, Low-level shRNA cytotoxicity can contribute to MYC-induced hepatocellular carcinoma in adult mice. *Molecular therapy : The journal of the American Society of Gene Therapy* **18**, 161-170 (2010).
15. Q. Pan *et al.*, Disturbance of the microRNA pathway by commonly used lentiviral shRNA libraries limits the application for screening host factors involved in hepatitis C virus infection. *FEBS letters* **585**, 1025-1030 (2011).
16. E. M. Anderson *et al.*, Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA* **14**, 853-861 (2008).

If you have any questions, contact

t +44 (0) 1223 976 000 (UK) **or** +1 800 235 9880 (USA); +1 303 604 9499 (USA)
f + 44 (0)1223 655 581
w horizondiscovery.com/contact-us **or** dharmacon.horizondiscovery.com/service-and-support
Horizon Discovery, 8100 Cambridge Research Park, Waterbeach, Cambridge, CB25 9TL, United Kingdom

horizon™
INSPIRED CELL SOLUTIONS

V3-1018