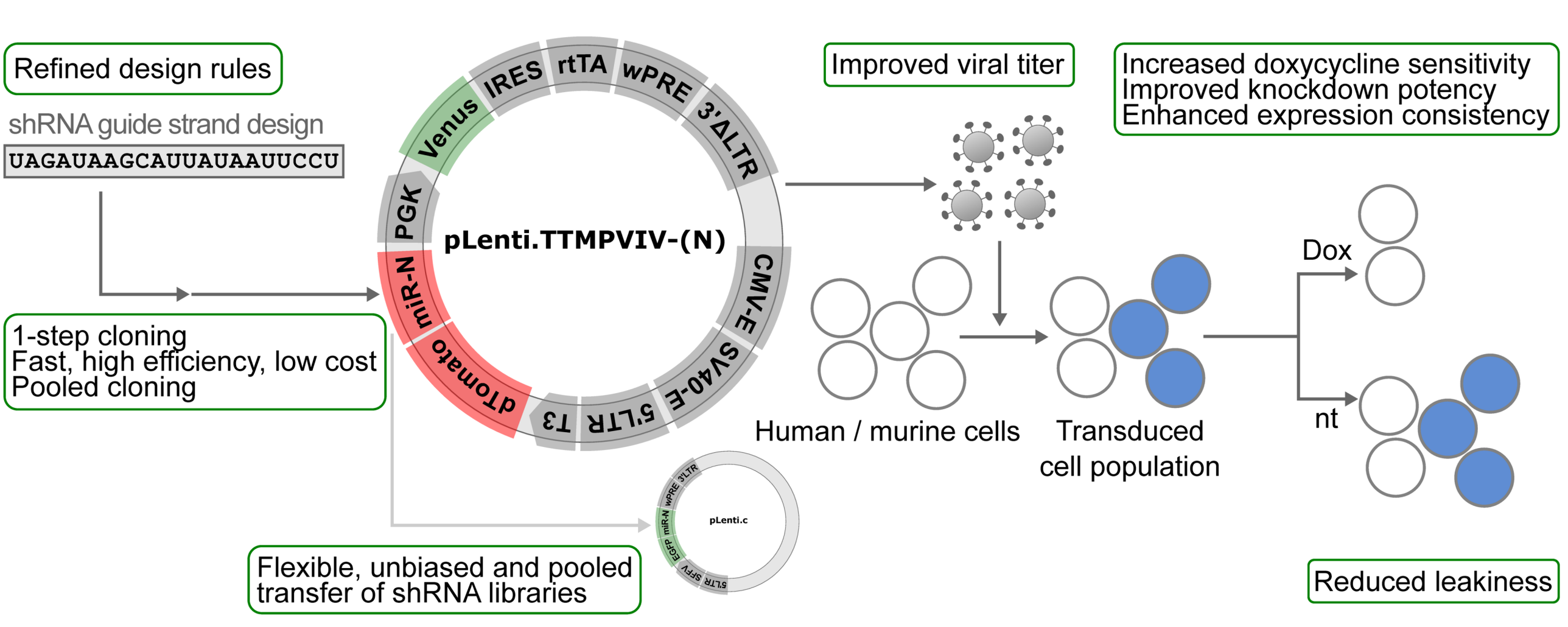

Refined design rules
shRNA guide strand design
UAGAUAAGCAUUAUAAUUCCU
1-step cloning
Fast, high efficiency, low cost
Pooled cloning
Venus
IRES
rtTA
wPRE
3'ΔLTR
CMV-E
SV40-E
5'LTR
T3
dTomato
miR-N
PGK
pLenti.TTMPVIV-(N)
Flexible, unbiased and pooled
transfer of shRNA libraries
pLenti.c
Improved viral titer
Human / murine cells
Transduced
cell population
Dox
nt
Increased doxycycline sensitivity
Improved knockdown potency
Enhanced expression consistency
Reduced leakiness

An Optimized Lentiviral Vector System for Conditional RNAi and Efficient Cloning of microRNA embedded short hairpin RNA Libraries

Felix F Adams[1], Dirk Heckl[2], Thomas Hoffmann[3], Steven R Talbot[4], Arnold Kloos[5], Felicitas Thol[5], Michael Heuser[5], Johannes Zuber[3], Axel Schambach[1,6,#] and Adrian Schwarzer[1,5,#,*]

[1] Institute of Experimental Hematology, Hannover Medical School, 30625 Hannover, Germany

[2] Pediatric Hematology and Oncology, Hannover Medical School, 30625 Hannover, Germany

[3] Research Institute of Molecular Pathology (IMP), 1030-Vienna, Austria

[4] Institute of Physiological Chemistry, Hannover Medical School, 30625 Hannover, Germany

[5] Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, 30625 Hannover, Germany

[6] Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, USA

# Authors contributed equally.

* Corresponding author: schwarzer.adrian@mh-hannover.de

Running title: miR30N

This is an accepted manuscript. Final article available at: http://dx.doi.org/10.1016/j.biomaterials.2017.05.032

Abstract

RNA interference (RNAi) and CRISPR-Cas9-based screening systems have emerged as powerful and complementary tools to unravel genetic dependencies through systematic gain- and loss-of-function studies. In recent years, a series of technical advances helped to enhance the performance of virally delivered RNAi. For instance, the incorporation of short hairpin RNAs (shRNAs) into endogenous microRNA contexts (shRNAmiRs) allows the use of Tet-regulated promoters for synchronous onset of gene knockdown and precise interrogation of gene dosage effects. However, remaining challenges include lack of efficient cloning strategies, inconsistent knockdown potencies and leaky expression. Here, we present a simple, one-step cloning approach for rapid and efficient cloning of miR-30 shRNAmiR libraries. We combined a human miR-30 backbone retaining native flanking sequences with an optimized all-in-one lentiviral vector system for conditional RNAi to generate a versatile toolbox characterized by higher doxycycline sensitivity, reduced leakiness and enhanced titer. Furthermore, refinement of existing shRNA design rules resulted in substantially improved prediction of powerful shRNAs. Our approach was validated by accurate quantification of the knockdown potency of over 250 single shRNAmiRs. To facilitate access and use by the scientific community, an online tool was developed for the automated design of refined shRNA-coding oligonucleotides ready for cloning into our system.

Keywords

RNAi; knockdown; shRNA; miR-30; lentiviral vector; Tet-regulation

Introduction

The ability to rapidly and systematically probe the knockdown of genes using RNA interference (RNAi) has been transformative for the field of biological research. In mammalian cells, transient gene silencing is typically accomplished by transfection of synthetic short, double-stranded siRNAs [1]. Stable and heritable sequence-specific gene silencing can be achieved by (retro)viral delivery of siRNA precursors, such as short hairpin RNAs (shRNAs) [2]. First generation vector-based RNAi systems used strong Polymerase III promoters to drive expression of short, simple stem-loop shRNAs that were further processed into RNAi triggers by Dicer [2,3]. Although this strategy elicited potent

knockdown in a number of instances, ineffective and imprecise processing caused accumulation of unprocessed transcripts resulting in severe off-target effects due to obstruction of the endogenous miRNA pathway [4]. A recent head-to-head comparison of different RNAi platforms revealed that commonly used Polymerase III driven stem-loop designs, such as the popular TRC-design [5], perform poorly under limiting conditions. More importantly, Polymerase III driven shRNAs were not processed in the intended way, as the vast majority of produced guide RNAs were found to be shifted ~4 nt downstream of the expected start site [6].

Second generation RNAi-systems resemble natural microRNAs and were obtained by embedding shRNAs into endogenous miRNA backbones, such as human miR-30a, that are processed through the complete RNAi biogenesis pathway [7–9]. Such shRNAmiRs can be generated from Polymerase II promoters, enabling constitutive [8], tissue specific [10] or inducible expression [9]. shRNAmiRs can easily be multiplexed [11] and directly linked to reporter genes in polycistronic transcripts to track cells with functional expression of the shRNAmiR [12]. Despite these advantages, second generation shRNAmiR systems frequently displayed insufficient knockdown efficacy when expressed at single-copy level [13]. This was addressed in third generation systems through new insights into the processing of small RNAs in mammalian cells. Biology-guided optimization of the miR-30a backbone greatly enhanced knockdown potency by restoration of a conserved motif 3' of the basal stem that had been destroyed by the introduction of restriction sites ("miR-E"; [14]). In parallel, large scale efforts developed rules predictive for optimal shRNA sequences embedded in the miR-30 backbone [15–17].

A highly controlled, synchronous and reversible onset of gene knockdown is desired in many experimental conditions. For these purposes, the Tet-ON RNAi system was developed, where the reverse tetracycline transactivator (rtTA) drives the expression of the shRNAmiR downstream of a tetracycline-responsive element (TRE) promoter [8,9]. A major advance was the introduction of a dual color, all-in-one TRMPVIR system (TRE.dsRed.miR.PGK.Venus.IRES.rtTA3), which coupled TRE-driven expression of an shRNAmiR linked to a fluorescent reporter with constitutive expression of the rtTA coupled to a second reporter gene [12]. This design allowed isolation and tracking of those cells that were transduced and expressed the functional shRNA construct upon doxycycline addition.
However, in conjunction with the greatly enhanced potency of the miR-E system, the intrinsic leakiness of the TRMPVIR system and several other shortcomings limit practical applications.

Here, we systematically improved several aspects of previous systems for vector based shRNAmiR expression, to produce a complete and versatile lentiviral all-in-one system for conditional RNAi.

Based on the natural human miR-30a backbone, we designed the third generation miR backbone "miR-N", which facilitates simplified one-step, PCR-free cloning of high quality shRNAmiR libraries in a minimal amount of time. Optimization of the different components of the TRMPVIR system resulted in a new lentiviral all-in-one vector that combines desired practicalities of enhanced doxycycline sensitivity, improved titers and substantially reduced leakiness. To further enhance the power of our system, we reanalyzed a published large scale shRNA efficacy data set [16] and discovered new sequence features refining previously established shRNA design rules for better prediction of potent RNAi triggers as validated on our independent set of more than 250 shRNAs measured in a FACS-based reporter assay. For automated design of oligonucleotides ready to clone into the miR-N system, we developed an online tool that includes the Sensor criteria and our refined shRNA design rules for improved shRNAs (www.mh-hannover.de/mir-n.html).

Materials and Methods

Vector construction, shRNAmiR cloning and library

The miR-N cassette for cloning of overlapping oligonucleotides via BsmBI was designed according to the human endogenous MIR30A and subsequently cloned into pLenti.c [18] via PCR amplification. For construction of an inducible, lentiviral vector for miR-N cloning, BsmBI restriction sites were removed from the pGamma.TTMPVIR-(E) (T3.TMPVIR-(E), **T3**.d**T**omato.**m**iR**E**.**P**GK.**V**enus.**I**RES.**r**tTA3) expression cassette [12,14] by point mutation (PGK*) inserted via a gBlock (IDT). For a transactivator-optimized vector, PGK* was cloned into pGamma.TTMPVIV-(E) with rtTA-V10 as a codon optimized transactivator (T3.TMP*VIV-(E), **T3**.d**T**omato.**m**iR**E**.**P**GK.**V**enus.**I**RES.rtTA-**V**10). The miR-N cloning cassette from pLenti.c (pLKO5d.SFFV.EGFP.miR-N.Spacer) was inserted into pGamma.TTMPVIV-(E) before transferring the complete expression cassette (T3.TMP*VIV-(N), **T3**.d**T**omato.**m**iR**N**.**P**GK.**V**enus.**I**RES.rtTA-**V**10) to the lentiviral pLKO5d vector backbone with an SV40- and CMV IE-enhancer ([18]) via previously included multiple cloning sites, yielding pLenti.TTMPVIV-(N) (pL40C.**T**3.d**T**omato.**m**iR**N**.**P**GK.**V**enus.**I**RES.rtTA-**V**10.WPRE). The T3.TMP*VIV-(N) cassette was also transferred to the pLenti backbone without SV40- and CMV IE-enhancer, yielding pLenti.TTMPVIV-(N)$^{-SV40E-CMVIEE}$. Codon optimized dTomato, which is used in all dTomato expressing vectors, and rtTA-V10 were generated using an online tool

that utilizes the GeneOptimizer algorithm and accounts for several factors that may influence gene expression (e.g. transcription, splicing, translation, mRNA degradation) [19,20].

For miR-N shRNAmiR cloning, 67 bp oligonucleotides used for miR-N design were purchased from Integrated DNA Technologies (IDT) or Eurofins Genomics. Oligonucleotides for different genes were pooled to clone libraries. Oligonucleotides were phosphorylated by a polynucleotide kinase at 37 °C for 45 min, subsequently heated to 95 °C for 2.5 min and annealed by cooling to 22 °C at 0.1 °C/s. Phosphorylated and annealed oligonucleotides were diluted 1:500 in water and ligated into the BsmBI digested vector backbone by standard cloning techniques. For library cloning, the ligation reaction was scaled up to achieve a proper representation of clones (evaluated by counting bacterial colonies with OpenCFU 3.9.0) [21]. For miR-E shRNAmiR cloning, 97 bp oligonucleotides were purchased from IDT or Eurofins Genomics and amplified by polymerase chain reaction (PCR) in a 25 µL reaction with 2.5 µL PCR buffer, 0.75 µL 10 mM dNTP, 0.5 µL $MgSO_4$, 0.75 µL 10 mM primer mix, 0.5 µL miR-E oligonucleotide, 0.2 µL Platinum Pfx (Invitrogen Platinum Pfx Kit) and 2.5 µL Enhancer solution. PCR conditions were 94 °C for 4 min, 95 °C for 20 s, 56 °C for 30 s, 72 °C for 25 s, 35x repeats of step 2-4, 72 °C for 3 min. The reaction was purified with a PCR purification kit (Qiagen), digested with EcoRI/XhoI, purified on a 1.5 % agarose gel using a gel elution kit (Qiagen) and ligated into a digested vector. Oligonucleotide forward (fw) sequences for shRNAs used here are given in Table S1.

Next generation sequencing

Sequencing of shRNAmiR libraries followed a strategy by Sims et al [22]. Briefly, plasmid libraries were amplified by PCR in a 50 µL reaction with 1 ng template, 5 µL 10x PCR buffer, 4 µL 25 mM magnesium chloride, 1 µL 10 mM dNTP, 3 µL forward and reverse primer (1.5 µL each, 10 µM) and 0.5 µL Amplitaq Gold polymerase (Invitrogen). Primer sequences are given in Table S2. PCR reaction conditions were 95 °C for 10 min, 95 °C for 30 s, 50 °C for 45 s, 72 °C for 60 s, 31x to step 2, 72 °C for 7 min. PCR products were purified in an agarose gel using a gel elution kit (Qiagen). After DNA quantification with the Qubit 2.0 Fluorometer (Invitrogen), different samples (libraries) were mixed and overall DNA quantified by qPCR (KAPA Library Quantification Kit Illumina platforms, Kapa Biosystems). Library sequencing was performed on a MiSeq platform according to manufacturers' protocols (Illumina). Quality of each run was controlled with the FastQC software (version 0.11.4, [23]). Trimming and filtering of the data, as well as quantitative and graphical analysis was performed using R/Bioconductor scripts.

Reporter assay for quantifying protein knockdown levels

The reporter assay for quantifying protein knockdown (KD) levels (referred to as “KD reporter assay”) was performed as previously described [14]. To obtain a stable reporter cell line for KD reporter assays, up to 2 kb gBlocks (IDT) were designed according to Fellmann et al [14]. 100 ng of a gBlock in TE were digested by XhoI/EcoRI, purified using the Qiagen PCR purification kit and ligated into digested and dephosphorylated pRSF91.mTagBFP2.HP-Linker.wPRE (reporter vector). In general, viral particles carrying the reporter or shRNAmiR expressing constructs were produced by PEI transfection of 3.5E+05 293T cells in 12-wells (or 6-/24-wells). Transduction of 2E+04 cells in 12-wells (or 6-/24-wells) was performed using spin inoculation at 632 x g for 1 h at 37 °C. Reporter cell lines were generated by transduction of MEFs (murine embryonic fibroblasts) or 32D cells by transduction with ecotropic viral particles of the reporter vector. Transduced cells were sorted for BFP to obtain stable reporter cell lines.

For the KD reporter assay, stable reporter cell lines (previous paragraph) were transduced with ecotropic or VSVg pseudotyped viral particles of shRNAmiR expressing constructs. Reporter cells containing inducible vectors were cultured with or without 1 µg/mL doxycycline two days after transduction. Cells were analyzed by flow cytometry (BD LSR II) five or six days after transduction. If available, knockdown values were normalized to shPten.1 and otherwise normalized to the mean knockdown of shPten.1 which is 95 %.

Western Blot

Western blots for assessment of protein knockdown were performed as previously described [24]. In short, 40 µg denatured and quantified proteins (Pierce BCA Protein Assay Kit, ThermoFischer Scientific) were separated on SDS gels and blotted onto nitrocellulose membranes. After 1 h blocking with 2.5 % milk powder in TBS with 0.1 % Tween 20 (TBST), the primary antibody was applied for 1 h at room temperature or 4 °C overnight in 2.5 % BSA in TBST. HRP-conjugated secondary antibodies were incubated for 1 h at room temperature in 2.5 % milk powder in TBST. HRP-conjugated primary antibodies were incubated for 1 h at room temperature in 2.5 % milk powder in TBST. After incubation with the ECL Western Blot Substrate (Thermo Scientific), the chemiluminescence signal was measured. Antibodies used are listed in Table S3.

Statistical analysis

Graphical and statistical analysis was performed with R version 3.1.0 or higher. For linear regression, P-values were calculated using the lm function (lm). For categorical data, depending on the results of Shapiro-Wilk normality test and graphical data distribution assessment, P-values were calculated using the t-test (one sample or Welch corrected two sample t-test) or Wilcoxon test (W-test). P-values for multiple comparisons were calculated with a One-way Analysis of Variance (ANOVA) and paired t-test or Kruskal-Wallis-test with Bonferroni's correction as post-hoc tests (TTB, KWB). Applied P-value ranges indicated by asterisks were: *** < 0.001 < ** < 0.01 < * < 0.05 < ns. Boxplots are drawn on the first, second and third quartiles. Boxplot whiskers mark the lowest or highest value within 1.5 * inter-quartile range (IQR) of the second quartile, while outliers are indicated by points. Barplots show the mean with whiskers indicating mean + standard deviation (SD). The root mean square error (RMSE) was calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(KD_i - pred_i)^2}$$

Online tool

The online tool was written with R/Shiny, a web application framework that allows programming of websites in R [25]. For predesigned shRNAs, a list of previously predicted shRNAs [14] was converted to the miR-N format using custom R scripts. shRNAs were reordered according to the prediction model described in the results. The shRNA design algorithm for custom sequences was written in R based on previous work [16,26]. The online tool is freely available at www.mh-hannover.de/mir-n.html.

Results

An improved microRNA platform for efficient shRNAmiR cloning and optimized knockdown potency

Following a series of improvements in shRNAmiR design [2,6,7,12,14–17,27–30], we sought to generate and evaluate a modified miR-30 backbone that fully resembles flanking sequences of endogenous human MIR30A and at the same time simplifies the cloning of synthetic stem-loop structures. To this end, we generated two miR-30 backbones, the first one comprised of 135 bp

natural sequences (termed “miR-N”) flanking the stem-loop in the human MIR30A on either side (Fig. 1A). In the second scaffold, “miR-S”, a point mutation (A➔G at position -12 upstream of the 5p Drosha cleavage site) was introduced in the basal stem-loop to eliminate the bulge adjacent to the basal UG-motif, which was proposed to enhance Drosha processing [31]. For cloning purposes, the MIR30A stem-loop was replaced by a short DNA sequence (“spacer”) with two opposing BsmBI restriction sites, which enables excision of the spacer to yield a miR-N backbone without residual restriction sites but with overhangs consisting of the adjacent sequence (Fig. 1B). 67mer oligonucleotides encoding the passenger (22 bp), loop (19 bp), guide (22 bp) and a 4 bp overhang can be inserted into the backbone as follows: Phosphorylation and annealing of forward and reverse oligonucleotides, followed by ligation into the BsmBI digested miR backbone (Fig. 1B). The miR-N/S cassettes with the BsmBI flanked spacer were linked to the 3’ UTR of different fluorochromes (EFGP/dTomato/tagRFP657/mTagBFP2) and cloned into a lentiviral expression vector (pLenti.c [18]) under the control of a powerful SFFV (spleen focus forming virus) promoter. In contrast to recently described similar approaches [28], all BsmBI sites from our vector backbones were mutated to allow direct insertion of shRNAmiRs into the final expression vectors without any cloning intermediates.

First, we sought to test the knockdown potency of the miR-N and miR-S scaffolds in comparison to miR-E and the original miR-30. We cloned four shRNAmiRs (shPten.1, shPten.2, shRptor.1, sh4ebp2) of different strengths into the four different miR-30-derived backbones and measured protein knockdown. To accurately quantify knockdown potencies, we applied the previously established fluorochrome-based KD reporter assay [14], in which multiple shRNA target sequences are cloned into the 3’ UTR of a blue fluorescent protein (BFP) reporter gene (Fig. 1C). We concatenated up to 65 shRNA target sequences (cloned via gBlocks, Integrated DNA Technologies) in the 3’ UTR of the fluorescence reporter gene (mTagBFP2). Stable reporter cell lines were created by transducing the reporter construct into mouse embryonic fibroblast (MEF) cells, which were subsequently transduced with the shRNAmiR constructs. Importantly, the knockdown potency of 242 individually measured shRNAs with their target sequences inserted at different positions within the reporters showed high knockdown consistency independent of the spatial position of the target sequence in the reporter (lm, $P = 0.93$, Fig. S1).

Consistent with previous results from Fellmann et al, miR-E substantially enhanced knockdown potency compared to miR-30 for all hairpins except shPten.1 [14], which displayed maximum knockdown efficacy also in the miR-30 scaffold. No major differences between miR-N, miR-S and miR-

E were detectable (Fig. 1D). Since removal of the lower bulge in the stem-loop (miR-S) did not lead to increased knockdown, we continued with the miR-N design. To validate our findings, miR-N and miR-E were tested again with five shRNAs, each with a minimum of three values from independent experiments. These tests confirmed that both backbones confer a high knockdown potency with no significant differences (Wilcoxon test (W-test), P = 0.87, Fig. 1E). Consistent with these findings, comparison of miR-E and miR-N by western blot revealed a highly potent knockdown for both miR backbones (Fig. 1F). Altogether, we designed a miR-30-based backbone ("miR-N"), which displays an optimal knockdown potency and enables efficient shRNAmiR cloning via fast and cost-effective 67mer duplexed oligonucleotides into optimized lentiviral expression vectors.

miR-N enables flexible and consistent pooled cloning of shRNAmiR libraries

To evaluate the improved cloning procedure of miR-N, we manually cloned >350 shRNAmiRs and verified the stem-loop by Sanger sequencing (Fig. 2A). We found an overall error rate of around 10 % mutated shRNAmiRs, with most mutations (25 %, 2.5 % of total) as a single nucleotide deletion in the passenger strand, reflecting contamination of the non-purified oligonucleotides with n-1 products, rather than cloning artifacts (Fig. 2A). We routinely select only one clone for sequencing, which is correct in nine out of ten cases. Next, we determined whether the method can be used to generate pooled shRNAmiR libraries. For screening purposes, the resulting shRNAmiR library should ideally represent all hairpins equally to provide maximum sensitivity. As a prototype, we constructed a library with 525 hairpins against ~140 target genes of the Rac pathway, including controls, by pooling, phosphorylating, annealing and ligating either all 1050 oligonucleotides in one tube (pLenti.c_1050) or by using three pools of about 350 oligonucleotides each (pLenti.c_3x350) (Fig. 2B). With increasing numbers of pooled oligonucleotides, we observed a decrease in bacterial colonies obtained after transformation (Fig. 2C), likely due to non-annealed oligonucleotides in the ligation mix. However, our data demonstrates that pools of ~1000 oligonucleotides can be efficiently cloned in this one-step approach. In addition, we demonstrate the modularity of the system by transferring the pLenti.c_3x350 library to a gammaretroviral backbone via the NotI/MluI restriction sites flanking the miR-N cassette (pGamma.TTMPVIV-(N)_RE). Next, we picked 94 single clones from different libraries and analyzed them by Sanger sequencing (Fig. S2A). Consistent with the results from individual cloning, only few clones (<10 %) were religated or harbored a mutation.

A common approach for quantifying the abundance of individual shRNAmiRs in multiplexed screens relies on next generation sequencing of shRNAmiR guide strands [12]. Since this approach only allows for detecting and excluding guide strand mutations, we were interested to which degree shRNAmiRs harboring mutations in the passenger strand retain their knockdown activity as previously suggested [32–35]. Measurement of 13 mutated shRNAmiRs with the KD reporter assay showed 70 - 100 % remaining activity for the more frequently occuring 1 bp deletions in the loop or passenger strand, 20 - 30 % activity for 1 bp guide strand deletions and 0 - 20 % activity for mutations affecting more base pairs (activities normalized to wildtype shRNAmiR, Fig. S2B). This suggests that the more severe mutations are detected during half-hairpin sequencing or are less likely to occur, while the most frequently observed mutations retain most of their knockdown activity and are therefore unlikely to distort the results of drop out screens.

For in-depth analysis of shRNA distribution in the library, we next amplified the guide strands of all libraries by PCR and performed deep sequencing. The distribution of read counts showed a very uniform representation of guides for all libraries with more than 90 % of shRNAmiRs scattering in a narrow range of 1.4 logs (Fig. 2D,E). As an “ideal” reference, we pooled equimolar amounts of plasmid DNA of individually cloned hairpins, which resulted in a slightly narrower distribution of hairpins, but still did not result in a completely unbiased library (pGamma.c_Manual, Fig. 2D,E). Interestingly, the read count distribution of pLenti.c_1050 was as uniform as pLenti.c_3x350, which was compiled from three subpools, indicating that mixing of individually cloned libraries does not introduce a substantial bias into shRNA representation. Overall, there was a high concordance between the reads of the two independently cloned libraries ($R^2 = 0.69$, $p < 2E\text{-}16$, Fig. 2F). Notably, upon transfer of the complete library from one backbone to another backbone by cloning, the read count distribution was fully conserved ($R^2 = 0.92$, $p < 2E\text{-}16$, Fig. 2G). In summary, we showed that parallel and pooled cloning of custom shRNAmiR libraries via miR-N is reliable and that libraries can be easily transferred to other vectors by pooled cloning.

An improved lentiviral vector for inducible RNAi

Studying genes involved in proliferation or survival with RNAi leads to a strong proliferative advantage for cells that fail to express the RNAi trigger and thereby confound the screen [12]. Furthermore, for conditional RNAi, reliable and detectable expression of a transactivator is needed. This has been addressed in the functional all-in-one TRMPVIR

(P**T**ET.ds**R**ed.**m**iR30.**P**GK.**V**enus.**I**RES.**r**tTA3) vector design that provides constitutive expression of a reverse transactivator (rtTA3) and a fluorescent reporter gene (Venus) connected via an IRES element, as well as Tet-regulated expression of the miR-30 shRNAmiR coupled to dsRed (TRMPVIR) or dTomato (TTMPVIR) [12]. This design has later been adapted in miR-E vectors that also contained the T3 promoter which provides tighter control of Tet-regulated genes [14]. An overview of vectors used here can be found in Table 1 and Fig. S3A.

To test the performance of these vectors in conjunction with a potent third generation miR backbone, we transduced a murine T-ALL cell line [36] with pGamma.TTMPVIR-(E) harboring hairpins against the mTOR components Rptor (alternatively named Raptor) and Rictor. In conjunction with the powerful third generation miR backbone, we observed a substantial protein knockdown despite the absence of doxycycline (Fig. 3A), indicating significant leakiness of the shRNA expression. KD reporter assays conducted on multiple different hairpins confirmed substantial knockdown of the fluorescence of 20 – 60 %, depending on shRNA potency when the target site was present in the 3' UTR of the reporter in the non-induced state, suggesting leakiness of the vector (Fig. S3B, pGamma.T3TMPVIR-(E)). To overcome leakiness of expression, we aimed at optimizing TTMPVIR by exchanging the T3 promoter with the T11 promoter [37], which provides an even tighter control of Tet-regulated genes [38]. However, the exchange of promoters did not reduce the leakiness of the system (Fig. S3B). Next, we tested different transactivators, but even in a construct with no transactivator, the same amount of leakiness was detected (Fig. S3B,C). To test if leakiness was derived from the TRE driven minimal promoter at all, we removed the T3 promoter and dTomato from the construct without rtTA3. Unexpectedly, this promoterless vector yielded even greater knockdown rates of the fluorescence reporter (Fig. S3C)**,** suggesting that functional expression of the miR cassette might be initiated from elements in the vector backbone upstream of the T3 promoter. Therefore, we transferred the TTMPVIV-(N) expression cassette (harboring the rtTA-**V**10 transactivator) to a lentiviral self-inactivating (SIN) backbone. The resulting vector pLenti.TTMPVIV-(N) displayed significantly reduced leakiness (60 % reduction, W-test, P = 0.0005) compared to pGamma.TTMPVIV-(N) measured by KD reporter assays and confirmed by western blots (Fig. 3B, Fig. 3C upper panel non-treated (nt), Fig. 3D). We previously reported that "tighter" doxycycline-inducible systems often showed reduced maximum expression levels [39]. Importantly, despite reduced leakiness, the knockdown potency of pLenti.TTMPVIV-(N) was fully maintained compared to pGamma.TTMPVIV-(N) (W-test, p = 1, Fig. 3C middle panel and Fig. 3E). Furthermore, higher consistency of the knockdown was achieved with the

lentiviral system by loss of a population expressing dTomato but not the shRNAmiR (dTomato^{+}BFP^{+}) (Fig. 3C lower panel doxycycline treated). Western blot analysis confirmed absence of leakiness – despite usage of the potent shRNA shPten.1 – and potent knockdown upon doxycycline induction in cells transduced with pLenti.TTMPVIV-(N) (Fig. 3F). To further improve signal-to-noise ratio of the system, we exchanged the rtTA3 transactivator for the optimized transactivator rtTA-**V**10 (TTMPVI**V**-(E/N)) [40], resulting in enhanced doxycycline sensitivity indicated by a higher percentage of inducible (dTomato^{+}) cells and dTomato mean fluorescence intensity (MFI) (Fig. 3G,H). Estimation of the area under the curve as an assessment for doxycycline sensitivity showed a significant increase from rtTA3 (pGamma.TTMPVI**R**-(E)) to rtTA-V10 (pGamma.TTMPVI**V**-(E)) (W-test, $p < 2.6E-4$ for MFI and $p < 2E-5$ for percentage).

To achieve one-step cloning capability, we inserted a BsmBI-devoid expression cassette into the lentiviral backbone, resulting in pLenti.TTMPVIV-(N) that has two BsmBI sites exclusively in the spacer for cloning.

In order to enhance the versatility of the vector system for difficult to transduce cells, we sought to improve the titer of the system. First, dTomato and rtTA-V10 were codon-optimized to remove cryptic splice sites, secondary structures and for maximal expression (Fig. 3I). Furthermore, we inserted additional cytomegalovirus immediate-early (CMV IE)- and simian virus 40 (SV40)-enhancer elements 5' of the hybrid RSV/LTR in order to increase expression of the genomic viral RNA during packaging. As a result, pLenti.TTMPVIV-(N) had a more than two-fold increased titer over pGamma.TTMPVIV-(N) as measured by titration of VSVg pseudotyped vectors on the human HT1080 cell line (t-test, P = 0.032, Fig. 3J). When the two vectors were pseudotyped with the envelope proteins most suitable for each vector type, VSVg pseudotyped pLenti.TTMPVIV-(N) again displayed a significantly higher titer over ecotropic pseudotyped pGamma.TTMPVIV-(N) as measured on murine SC-1 cells (t-test, P = 0.006, Fig. S3D). Direct comparison of titers between pLenti.TTMPVIV-(N) with or without additional enhancer elements revealed that the increased titer was largely due to the insertion of the SV40- and CMV IE-enhancer elements into the vector backbone (fold change = 1.8, t-test, P = 1.3E-06, Fig. S3E). In our hands, titers of $>1x10^{7}$ TU/mL are easily attainable by seventy-fold concentration of the supernatants via a 2 h ultracentrifugation step. Altogether, we generated an optimized, modular lentiviral vector system for Tet-regulated RNAi with enhanced doxycycline sensitivity, improved titer, more consistent knockdown, substantially reduced leakiness and the possibility to directly insert shRNAs via the duplexed oligonucleotide method.

Functional validation of the optimized inducible lentiviral vector system

Having established this versatile all-in-one system, we wanted to assess its performance by probing genes necessary for survival of human acute myeloid leukemia (AML). We first tested the capacity of pLenti.TTMPVIV-(N) to consistently silence target genes over a longer time period. We performed immunoblots of the human chronic myeloid leukemia cell line K-562 on days 10 and 15 after doxycycline induction. In the presence of doxycycline, a stable and consistent knockdown of PTEN was observed with shPten.1 when compared to mock-treated and shLuc-control shRNAmiR (Fig. 4A). Next, we performed a sensitive KD reporter assay in RagMEFs showing effective silencing of the fluorescent reporter over a period of 27 days (Fig. 4B).

To demonstrate that our vector enables the study of stable loss-of-function phenotypes in a biologically relevant context, we used the human AML cell line ML-2, which contains a t(6;11)(q27;q23) translocation (MLL-AF6). ML-2 cells were transduced with pLenti.TTMPVIV-(N) harboring a control shRNAmiR (shLuc) or KD reporter assay-validated shRNAmiRs against MYB and DOT1L [41]. The transcription factor MYB and the H3K79 methyltransferase DOT1L were previously shown to be essential for MLL-driven AML [42]. We monitored the percentage of $Venus^{+}dTomato^{+}$ ($shRNAmiR^{+}$) cells after addition of doxycycline as well as the percentage of $Venus^{+}$ (transduced) cells in the absence of doxycycline (Fig. 4C). Upon doxycycline induction, the cells expressing shRNAmiRs against MYB and DOT1L were completely eliminated from the culture, while cells expressing a non-targeting control shRNAmiR continued to proliferate (Fig. 4D and Fig. S4A). Continuous depletion of cells expressing shDOT1L between each time point (> 30 days) demonstrates stable long-term knockdown in the presence of doxycycline. Importantly, in the absence of doxycycline, percentages of transduced cells (Venus+) were stable over more than 30 days (Fig. 4E and Fig. S4B), even with shRNAmiRs that strongly deplete when expressed, demonstrating sufficiently tight control of shRNAmiR expression. Similar to the original TRMPVIR vector system [12], a population (12 - 15 %) of $Venus^{+}dTomato^{-}$ cells, which failed to functionally express the shRNAmiR, was detected in this experimental setting (Fig. 4F). However, this population was clearly identified by the lack of dTomato expression and could be accounted for accordingly. To compare depleting and control shRNAmiRs simultaneously under exactly the same conditions, we performed a competitive growth assay, in which shRNAmiRs were distinguished within one culture well of ML-2 cells by dTomato (pLenti.T**T**MPVIV-(N)) or E2-Crimson (pLenti.T**C**MPVIV-(N)) expression (Fig. 4G,H). Upon doxycycline induction, shMYB.1 was depleted in comparison to shLuc independently of the fluorescent proteins as indicated

by similar results obtained when fluorescent proteins were exchanged (Fig. 4G,H). In agreement with Fig. 4E and Fig. S4B, quantification of shLuc and shMYB.1 expressing cells without long-term doxycycline induction demonstrated stable maintenance of shRNAmiRs over more than 20 days (Fig. S4C,D). In summary, pLenti.TTMPVIV-(N) provided stable maintenance of transduced cells without doxycycline and reliable shRNAmiR expression coupled to the fluorescent reporter upon doxycycline induction.

Improved shRNAmiR design to achieve optimal knockdown potencies

To refine selection of highly active shRNAs, we prospectively tested the Sensor rules in a third generation miR backbone. Assessment of 277 individually cloned and analyzed shRNAmiRs fully complying with the existing Sensor rules in the KD reporter assay resulted in knockdown rates of at least 70 % in four out of five tested hairpins (Fig. 5A). Hence, the combination of Sensor rules and third generation shRNAmiRs led to a remarkable efficacy in identifying powerful RNAi triggers. However, for hairpins that fulfilled the Sensor rules, there was no further correlation between the actual measured knockdown potencies and the predicted ranking by the Sensor algorithm [14,16] (ANOVA, $P = 0.91$) or by another widely used shRNA prediction score [15] (W-test, $P = 0.68$) (Fig. S5A,B). This is partially explained by the substantial overlap of these scores with the Sensor criteria. Using the data from our KD reporter assays, we considered whether additional design rules could be added to the Sensor rules in order to further increase their potency. First, by using sequence logos [43], we analyzed sequence features from the most potent 20 % and weakest 20 % of shRNAmiRs (Fig. 5B). The sequence logos indicated that more potent shRNAmiRs contained a higher AT content and a strong preference for T at position 1 and AT at position 13 [16]. However, although differences between potent and weak shRNAs could be identified by this approach, there was a strong dependence on the threshold used for the analysis. Furthermore, although our set of single knockdown assays is among the largest reported to date, the analysis of 277 datapoints might not be sensitive enough for exhaustive feature discovery. Therefore, we reanalyzed the Fellmann dataset since we and others already validated that it can predict potent design features [16]. For reanalysis, we used the least absolute shrinkage and selection operator (LASSO) regression, which was shown to be very powerful for derivation of design rules for siRNAs potency [26,44,45] and had not been applied to the dataset before.

To derive a model that is as exact as possible, we used the majority of the Fellmann dataset (90 %, n = 18720 shRNAs) for model building and set aside the remaining 10 % for validation in addition to our KD reporter assay dataset. For each sequence feature (T1, AT1-14, …) in 90 % of the Fellmann dataset, we computed the predicted knockdown. As a measure for the predictive quality of each feature, we calculated the root mean square error (RMSE) between the Fellmann dataset Score and the predicted knockdown (Fig. 5C). Discrimination of the most potent features (with lowest RMSE) was done by applying a threshold at RMSE at half maximum (Fig. 5D). Importantly, matching the analysis from Fellmann et al, we were able to reproduce the Sensor criteria, which validates our approach (Fig. 5D, Table 2). Notably, besides the Sensor rules obtained from [49], the LASSO algorithm identified additional features, e.g. a strong favor for T at position 1 instead of A or T at position 1, which corresponds to the data derived from the sequence logos (Fig. 5B). To build a prediction model on the resulting additional features, we filtered the Fellmann dataset for the Sensor criteria (n = 2648 shRNAs) and calculated the RMSE for every possible model composed of a varying number of sequence features (Fig. 5E). An increasing number of features resulted in a decreased RMSE, indicating a better fit between the prediction and the Fellmann Score by the model (Fig. 5F). For further analysis, we chose the most predictive model consisting of 9 features (Fig. 5G, Table 2). Ranking the most important features in the model showed that, in addition to the Sensor rules, the best shRNAs have an even higher AT content in the first 14 nucleotides than originally proposed, with a strong preference for T at position 1, 10, 14 and 17 and A or T at 13. Validation on the remaining 10 % of the Fellmann dataset, which was filtered to comply with the original Sensor criteria, resulted in a significant increase in the measured knockdown score for shRNAs that ranked high vs. low in our model (W-test, P = 0.006, high = shRNAs within upper third of score, low = shRNAs within lower third of score, Fig. 5H). Importantly, validation of the model on our completely independent dataset, consisting of 277 individually measured shRNAs that match the Sensor criteria, yielded a highly significantly increased true knockdown potency for shRNAs scoring high in our refined model (high = median knockdown = 91 % vs. low = median knockdown = 79 %, W-test, P = 1.1E-4 , Fig. 5I). The refined model is particularly good at identifying the most powerful shRNAs as demonstrated by high abundance (52 %) of ultra-potent hairpins (>90 % protein knockdown) in the set of shRNAs with a high prediction score versus only 13 % ultra-potent shRNAs in the low score set.

In summary, we show that the Sensor criteria in combination with miR-E or miR-N yield a high predictive capacity for powerful shRNA triggers. By employing a different approach to reinvestigate

shRNA properties of the Fellmann dataset, we identified additional shRNA features, which further increase the likelihood to find potent shRNAmiRs, as validated on an independent dataset. For practical reasons, we design our guides for a given gene with the Sensor rules and subsequent ranking using our prediction score. To provide an easily accessible source for shRNA design, we designed an online tool (www.mh-hannover.de/mir-n.html) that facilitates shRNA oligonucleotide design for custom sequences as well as selection of predesigned shRNAmiRs.

Discussion

Probing genetic liabilities through knockdown or knockout of genes is crucial for biological research and can be achieved by RNAi and CRISPR-Cas9, respectively. A recent comparison of the two methodologies demonstrated high power and precision of both systems to detect essential genes, but, unexpectedly, revealed more mutually exclusive than common hits, likely owing to inherent biological differences of both systems [46]. Therefore, advanced RNAi and CRISPR-Cas9 should be regarded as complementary rather than competitive technologies, with each system presenting unique strengths and weaknesses. For instance, the size of the Cas9 gene makes viral delivery into primary cells challenging, whereas small shRNA cassettes are efficiently delivered into many different cell types *in vitro* and *in vivo* [28].

Earlier versions of RNAi systems suffered from poor knockdown performance and substantial off-target effects caused by insufficient processing of the hairpins, resulting in low production of RISC-loadable guides and occurrence of RISC-loadable passengers [6,47]. Guided by better understanding of the cellular miRNA processing machinery and by the development of sophisticated shRNA design rules, the dramatic technical improvements over the last decade have made RNAi a mature and reliable technology. Here, we systematically improved several shortcomings of conditional vector-based RNAi by developing the miR-N backbone, improving lentiviral delivery and refining shRNA design rules. As expected, our natural miR-30 based backbone ("miR-N") performed similarly to the potent miR-E backbone. Although BsmBI-mediated cloning was described before [28], the system was not widely adopted since it required subsequent shuffling of the cloned hairpins from cloning vectors into the final viral expression vectors. Importantly, and in contrast to previous systems, miR-N offers a simple, cheap and efficient, PCR-free, one-step cloning process via duplexed oligonucleotides directly into our optimized lentiviral expression vectors. We show that at least 1000 shRNA-coding oligonucleotides can be easily pooled, annealed and cloned directly into the final expression vectors.

The resulting libraries have very low mutation rates and are free of PCR-introduced bias. This results in very even hairpin distribution and, theoretically, in improved statistical power of screening applications. The modularity of the system ensures that the libraries can be readily transferred to other vectors in a pooled manner. Our optimized, Tet-regulated lentiviral expression vector pLenti.TTMPVIV-(N) produces increased viral titers and exhibits a substantially reduced leakiness (i.e. 60 % less leaky than pGamma.TTMPVIV-(N)). The improved doxycycline sensitivity becomes particularly important for animal experiments, since doxycycline concentrations are not evenly distributed throughout the body and doxycycline intake can be variable or impaired in diseased animals. Despite the large size of the pLenti.TTMPVIV-(N) construct, optimized titers facilitated transduction of primary human AML samples with ~20 % transduced cells (Fig. S6, MOI ~25 titrated on the HT1080 cell line). Comparison to a previously published all-in-one vector (FKBP12-HA.FLPs-ERT2, [48]) showed major differences regarding vector components and properties (Table 1). Due to its tightness achieved by utilization of recombinases, FKBP12-HA.FLPs-ERT2 should be used when low leakiness is absolutely critical. However, low titers, absent possibilities to control dose and the lack of reversibility limit the use of FKBP12-HA.FLPs-ERT2 for many experimental settings and underline the advantages of pLenti.TTMPVIV-(N).

Another approach to compensate for limited potency of shRNAs was the creation of high coverage shRNA libraries with ~ 30 shRNAs per gene [27]. However, these ultra-complex libraries have a complicated readout and a high cost per target gene. Most importantly, the application of these libraries in *in vivo* studies is challenging, given the limited number of tumor-initiating / propagating-(adult)-stem-cells in many models. For these reasons, instead of increasing the number of shRNAs per target, we sought to improve the power of RNAi screens by enhancing the efficacy of each shRNA reagent. Having shown that the knockdown potency of third generation shRNAmiR systems is difficult to further improve by changing the architecture of the miR backbone, we explored possibilities to improve the prediction of potent RNAi-triggers to be cloned into these backbones. The widely used Sensor rules were used as the basis for our work [16]. We accurately determined the knockdown potency of more than 250 shRNAmiRs designed according to the Sensor rules and showed that 4 out of 5 shRNAmiRs conferred a knockdown of at least 70 %. Application of another shRNA-design algorithm proposed by Kampmann [15] to our dataset, or the ranking provided by the Sensor rules [14,16] gave no further correlation with the observed knockdown potencies. However, analysis of sequence logos indicated additional features linked to shRNA efficacy. By applying LASSO-regression

to the dataset from Fellmann et al [16], we independently retrieved the Sensor rules and identified additional sequence features predictive for powerful knockdown efficacy. We combined these individual sequence features into a refined model for shRNAmiR design and validated the model on our completely independent, individually measured dataset prefiltered for the Sensor criteria. Hence, extension of the Sensor rules by our new prediction model identified the most potent shRNAs with every second shRNA being highly potent (90 % knockdown), thus paving the way for design of focused shRNA libraries with minimal complexity that can be used for *in vitro* or *in vivo* loss-of-function screens.

Conclusions

The combination of the newly developed miR-N backbone with the lentiviral vectors pLenti.c and pLenti.TTMPVIV-(N) provides an easy system for reliable RNAi. It makes use of a simplified cloning process via PCR-free, BsmBI-mediated insertion of shRNA-coding oligonucleotides directly into the expression vectors. Thus, our system allows for fast and efficient cloning of single shRNAs or pooled cloning of custom shRNA libraries with minimum complexity and maximized potency. The simplicity and reliability of our system makes focused shRNA screens feasible for the broad scientific community. In line with previous improvements, our study improved the set of RNAi systems available and will alone or in combination with other tools, such as CRISPR-Cas9, facilitate reliable loss-of-function screens to invent novel biomaterials and to elucidate disease weaknesses and mechanisms.

Acknowledgements

We thank M. Morgan, D. Klatt and J.W. Schott for critically reading the manuscript, M. Wichmann for his assistance with next generation sequencing, M. Ballmaier and his team of the Cell Sorting Core Facility (Hannover Medical School), which is partially supported by Braukmann-Wittenberg-Herz-Stiftung and Deutsche Forschungsgemeinschaft, for cell sorting.

References

[1] Y. Dorsett, T. Tuschl, siRNAs: Applications in functional genomics and potential as therapeutics, Nat. Rev. Drug Discov. 3 (2004) 318–329. doi:10.1038/nrd1345.

[2] T.R. Brummelkamp, R. Bernards, R. Agami, A System for Stable Expression of Short

Interfering RNAs in Mammalian Cells, Science (80-.). 296 (2002) 550–554.

[3] P.J. Paddison, A.A. Caudy, G.J. Hannon, Stable suppression of gene expression by RNAi in mammalian cells., Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1443–8. doi:10.1073/pnas.032652399.

[4] D. Grimm, K.L. Streetz, C.L. Jopling, T.A. Storm, K. Pandey, C.R. Davis, P. Marion, F. Salazar, M.A. Kay, Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways, Nature. 441 (2006) 537–541. doi:nature04791 [pii]10.1038/nature04791 [doi].

[5] J. Moffat, D.A. Grueneberg, X. Yang, S.Y. Kim, A.M. Kloepfer, G. Hinkle, B. Piqani, T.M. Eisenhaure, B. Luo, J.K. Grenier, A.E. Carpenter, S.Y. Foo, S.A. Stewart, B.R. Stockwell, N. Hacohen, W.C. Hahn, E.S. Lander, D.M. Sabatini, D.E. Root, A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen, Cell. 124 (2006) 1283–1298. doi:10.1016/j.cell.2006.01.040.

[6] C. Watanabe, T.L. Cuellar, B. Haley, Quantitative evaluation of first, second, and third generation hairpin systems reveals the limit of mammalian vector-based RNAi., RNA Biol. 13 (2016) 25–33. doi:10.1080/15476286.2015.1128062.

[7] J.M. Silva, M.Z. Li, K. Chang, W. Ge, M.C. Golding, R.J. Rickles, D. Siolas, G. Hu, P.J. Paddison, M.R. Schlabach, N. Sheth, J. Bradshaw, J. Burchard, A. Kulkarni, G. Cavet, R. Sachidanandam, W.R. McCombie, M. a Cleary, S.J. Elledge, G.J. Hannon, Second-generation shRNA libraries covering the mouse and human genomes., Nat. Genet. 37 (2005) 1281–8. doi:10.1038/ng1650.

[8] R.A. Dickins, M.T. Hemann, J.T. Zilfou, D.R. Simpson, I. Ibarra, G.J. Hannon, S.W. Lowe, Probing tumor phenotypes using stable and regulated synthetic microRNA precursors., Nat. Genet. 37 (2005) 1289–1295. doi:10.1038/ng0306-389b.

[9] F. Stegmeier, G. Hu, R.J. Rickles, G.J. Hannon, S.J. Elledge, A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 13212–7. doi:10.1073/pnas.0506306102.

[10] T.T. Nielsen, I. van Marion, L. Hasholt, C. Lundberg, Neuron-specific RNA interference using lentiviral vectors, J. Gene Med. 11 (2009) 559–569. doi:10.1002/jgm.

[11] K.-H. Chung, C.C. Hart, S. Al-Bassam, A. Avery, J. Taylor, P.D. Patel, A.B. Vojtek, D.L. Turner, Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155, Nucleic Acids Res. 34 (2006) e53–e53. doi:10.1093/nar/gkl143.

[12] J. Zuber, K. McJunkin, C. Fellmann, L.E. Dow, M.J. Taylor, G.J. Hannon, S.W. Lowe, Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi., Nat. Biotechnol. 29 (2011) 79–83. doi:10.1038/nbt.1720.

[13] R.L. Boudreau, A.M. Monteys, B.L. Davidson, Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs., RNA. 14 (2008) 1834–44. doi:10.1261/rna.1062908.

[14] C. Fellmann, T. Hoffmann, V. Sridhar, B. Hopfgartner, M. Muhar, M. Roth, D.Y.Y. Lai, I.A.M.A.M. Barbosa, J.S.S. Kwon, Y. Guan, N. Sinha, J. Zuber, An Optimized microRNA Backbone for Effective Single-Copy RNAi, Cell Rep. 5 (2013) 1704–1713. doi:10.1016/j.celrep.2013.11.020.

[15] M. Kampmann, M.A. Horlbeck, Y. Chen, J.C. Tsai, M.C. Bassik, L.A. Gilbert, J.E. Villalta, S.C. Kwon, H. Chang, V.N. Kim, J.S. Weissman, Next-generation libraries for robust RNA interference-based genome-wide screens., Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3384-91. doi:10.1073/pnas.1508821112.

[16] C. Fellmann, J. Zuber, K. McJunkin, K. Chang, C.D. Malone, R.A. Dickins, Q. Xu, M.O. Hengartner, S.J. Elledge, G.J. Hannon, S.W. Lowe, Functional identification of optimized RNAi triggers using a massively parallel sensor assay., Mol. Cell. 41 (2011) 733–46. doi:10.1016/j.molcel.2011.02.008.

[17] S.R.V. Knott, A.R. Maceli, N. Erard, K. Chang, K. Marran, X. Zhou, A. Gordon, O. El Demerdash, E. Wagenblast, S. Kim, C. Fellmann, G.J. Hannon, A Computational Algorithm to Predict shRNA Potency, Mol. Cell. 56 (2014) 796–807. doi:10.1016/j.molcel.2014.10.025.

[18] D. Heckl, M.S. Kowalczyk, D. Yudovich, R. Belizaire, R. V Puram, M.E. McConkey, A. Thielke, J.C. Aster, A. Regev, B.L. Ebert, Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing, Nat. Biotechnol. 32 (2014) 941–946. doi:10.1038/nbt.2951.

[19] Thermo Fisher Scientific Inc., GeneOptimizer, (n.d.). https://www.thermofisher.com/de/de/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis/geneoptimizer.html (accessed April 17, 2017).

[20] D. Raab, M. Graf, F. Notka, T. Schödl, R. Wagner, The GeneOptimizer Algorithm: Using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization, Syst. Synth. Biol. 4 (2010) 215–225. doi:10.1007/s11693-010-9062-3.

[21] Q. Geissmann, C. Costa, S. Yang, C. Mello, W. dos Santos, M. Rodrigues, A. Candeias, C.

Gusmo, A. Severini, L. Borrs, A. Cirilo, M. Forero, J. Pennack, A. Hidalgo, A. Yati, S. Dey, H. Mansberg, D. Mukherjee, A. Pal, S. Sarma, D. Majumder, M. Putman, R. Burton, M. Nahm, Z. Cai, N. Chattopadhyay, W. Liu, C. Chan, J. Pignol, J. Bewes, N. Suchowerska, D. McKenzie, M. Niyazi, I. Niyazi, C. Belka, M. Clarke, R. Burton, A. Hill, M. Litorja, M. Nahm, S. Sieuwerts, F. De Bok, E. Mols, W. De Vos, J.V.H. Vlieg, S. Brugger, C. Baumberger, M. Jost, W. Jenni, U. Brugger, J. Marotz, C. Lbbert, W. Eisenbei, OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects, PLoS One. 8 (2013) e54072. doi:10.1371/journal.pone.0054072.

[22] D. Sims, A.M. Mendes-Pereira, J. Frankum, D. Burgess, M.-A. Cerone, C. Lombardelli, C. Mitsopoulos, J. Hakas, N. Murugaesu, C.M. Isacke, K. Fenwick, I. Assiotis, I. Kozarewa, M. Zvelebil, A. Ashworth, C.J. Lord, High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing., Genome Biol. 12 (2011) 1–13. doi:10.1186/gb-2011-12-10-r104.

[23] S. Andrews, P. Lindenbaum, B. Howard, P. Ewels, FastQC: A quality control tool for high throughput sequence data, (2010). http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed December 2, 2015).

[24] E. Warlich, J. Kuehle, T. Cantz, M.H. Brugman, T. Maetzig, M. Galla, A.A. Filipczyk, S. Halle, H. Klump, H.R. Schöler, C. Baum, T. Schroeder, A. Schambach, Lentiviral Vector Design and Imaging Approaches to Visualize the Early Stages of Cellular Reprogramming, Mol. Ther. 19 (2011) 782–789. doi:10.1038/mt.2010.314.

[25] RStudio, Easy web applications in R, (2013). http://www.rstudio.com/shiny/.

[26] J.-P. Vert, N. Foveau, C. Lajaunie, Y. Vandenbrouck, An accurate and interpretable model for siRNA efficacy prediction., BMC Bioinformatics. 7 (2006) 520. doi:10.1186/1471-2105-7-520.

[27] M.C. Bassik, R.J. Lebbink, L.S. Churchman, N.T. Ingolia, W. Patena, E.M. LeProust, M. Schuldiner, J.S. Weissman, M.T. McManus, Rapid creation and quantitative monitoring of high coverage shRNA libraries., Nat. Methods. 6 (2009) 443–5. doi:10.1038/nmeth.1330.

[28] L. Osório, R. Gijsbers, M. Oliveras-Salvá, A. Michiels, Z. Debyser, C. Van den Haute, V. Baekelandt, Viral vectors expressing a single microRNA-based short-hairpin RNA result in potent gene silencing in vitro and in vivo, J. Biotechnol. 169 (2014) 71–81. doi:10.1016/j.jbiotec.2013.11.004.

[29] S.E. Mohr, N. Perrimon, RNAi screening: New approaches, understandings, and organisms,

Wiley Interdiscip. Rev. RNA. 3 (2012) 145–158. doi:10.1002/wrna.110.

[30] S.E. Mohr, J. a Smith, C.E. Shamu, R. a Neumüller, RNAi screening comes of age: improved techniques and complementary approaches, Nat. Publ. Gr. 15 (2014) 591–600. doi:10.1038/nrm3860.

[31] V.C. Auyeung, I. Ulitsky, S.E. McGeary, D.P. Bartel, Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing, Cell. 152 (2013) 844–858. doi:10.1016/j.cell.2013.01.031.

[32] Q. Du, H. Thonberg, J. Wang, C. Wahlestedt, Z. Liang, A systematic analysis of the silencing effects of an active siRNA at all single-nucleotide mismatched target sites, Nucleic Acids Res. 33 (2005) 1671–1677. doi:10.1093/nar/gki312.

[33] J.G. Doench, P.A. Sharp, Specificity of microRNA target selection in translational repression., Genes Dev. 18 (2004) 504–11. doi:10.1101/gad.1184404.

[34] L.M. Alemán, J. Doench, P.A. Sharp, Comparison of siRNA-induced off-target RNA and protein effects., RNA. 13 (2007) 385–95. doi:10.1261/rna.352507.

[35] C. Dahlgren, H.-Y. Zhang, Q. Du, M. Grahn, G. Norstedt, C. Wahlestedt, Z. Liang, Analysis of siRNA specificity on targets with double-nucleotide mismatches, Nucleic Acids Res. 36 (2008) e53–e53. doi:10.1093/nar/gkn190.

[36] A. Schwarzer, H. Holtmann, M. Brugman, J. Meyer, C. Schauerte, J. Zuber, D. Steinemann, B. Schlegelberger, Z. Li, C. Baum, Hyperactivation of mTORC1 and mTORC2 by multiple oncogenic events causes addiction to eIF4E-dependent mRNA translation in T-cell leukemia., Oncogene. 34 (2015) 3593–3604. doi:10.1038/onc.2014.290.

[37] R. Loew, N. Heinz, M. Hampf, H. Bujard, M. Gossen, Improved Tet-responsive promoters with minimized background expression, BMC Biotechnol. 10 (2010) 81. doi:10.1186/1472-6750-10-81.

[38] M. Stahlhut, A. Schwarzer, M. Eder, M. Yang, Z. Li, M. Morgan, A. Schambach, O.S. Kustikova, Lentiviral vector system for coordinated constitutive and drug controlled tetracycline-regulated gene co-expression., Biomaterials. 63 (2015) 189–201. doi:10.1016/j.biomaterials.2015.06.022.

[39] N. Heinz, A. Schambach, M. Galla, T. Maetzig, C. Baum, R. Loew, B. Schiedlmeier, Retroviral and Transposon-Based Tet-Regulated All-In-One Vectors with Reduced Background Expression and Improved Dynamic Range, Hum. Gene Ther. 22 (2011) 166–176.

doi:10.1089/hum.2010.099.

[40] X. Zhou, M. Vink, B. Klaver, B. Berkhout, A.T. Das, Optimization of the Tet-On system for regulated gene expression through viral evolution, Gene Ther. 13 (2006) 1382–1390. doi:10.1038/sj.gt.3302780.

[41] K. Ohyashiki, J.H. Ohyashiki, A.A. Sandberg, Cytogenetic Characterization of Putative Human Myeloblastic Leukemia Cell Lines (ML-1, -2, and -3): Origin of the Cells, Cancer Res. 46 (1986).

[42] J. Zuber, A.R. Rappaport, W. Luo, E. Wang, C. Chen, A. V Vaseva, J. Shi, S. Weissmueller, C. Fellmann, C. Fellman, M.J. Taylor, M. Weissenboeck, T.G. Graeber, S.C. Kogan, C.R. Vakoc, S.W. Lowe, An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance., Genes Dev. 25 (2011) 1628–40. doi:10.1101/gad.17269211.

[43] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences., Nucleic Acids Res. 18 (1990) 6097–100. http://www.ncbi.nlm.nih.gov/pubmed/2172928 (accessed February 13, 2017).

[44] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, LEAST ANGLE REGRESSION, Ann. Stat. 32 (2004) 407–499.

[45] R Core Team, R: A Language and Environment for Statistical Computing, (2016). https://www.r-project.org/.

[46] D.W. Morgens, R.M. Deans, A. Li, M.C. Bassik, Systematic comparison of CRISPR / Cas9 and RNAi screens for essential genes, Nat. Biotechnol. 34 (2016) 1–4. doi:10.1038/nbt.3567.

[47] S. Gu, L. Jin, Y. Zhang, Y. Huang, F. Zhang, P.N. Valdmanis, M.A. Kay, The Loop Position of shRNAs and Pre-miRNAs Is Critical for the Accuracy of Dicer Processing In Vivo, Cell. 151 (2012) 900–911. doi:10.1016/j.cell.2012.09.042.

[48] T. Maetzig, J. Kuehle, A. Schwarzer, S. Turan, M. Rothe, A. Chaturvedi, M. Morgan, T.C. Ha, M. Heuser, W. Hammerschmidt, C. Baum, A. Schambach, All-in-One inducible lentiviral vector systems based on drug controlled FLP recombinase, Biomaterials. 35 (2014) 4345–4356. doi:10.1016/j.biomaterials.2014.01.057.

[49] L.E. Dow, P.K. Premsrirut, J. Zuber, C. Fellmann, K. McJunkin, C. Miething, Y. Park, R.A. Dickins, G.J. Hannon, S.W. Lowe, A pipeline for the generation of shRNA transgenic mice, Nat. Protoc. 7 (2012) 374–393. doi:10.1038/nprot.2011.446.

Figures

Figure Captions and Tables

Figure 1: miR-N as a potent miR backbone for efficient cloning of shRNAmiRs

(A) Schematic of differences between four miR-30 backbones. Cloning sites are indicated by arrow heads. Inserted DNA fragment between the cloning sites is indicated by darker gray. 5' com, complement. 3' cons, 3' conserved region (ACNNC). **(B)** Schematic presentation of the miR-N cloning process. Steps indicated by arrows: Target site selection by Sensor criteria, phosphorylation by polynucleotide kinase (PNK) and annealing of oligonucleotides. BsmBI digestion of miR-N backbone integrated into a vector lacking BsmBI sites. Ligation of oligonucleotides into digested backbone. **(C)** KD reporter assay for validation of shRNAmiR potency. Steps indicated by arrows: Stable transduction and sorting of 32D or RagMEF ($rtTA^+$ MEF) with the reporter vector that harbors shRNAmiR target sites coupled to BFP. Transduction of reporter cells with the shRNAmiR containing vector. Measurement of BFP fluorescence to assess of shRNAmiR potency five to six days after transduction. **(D)** Comparison of the knockdown potencies of different miR backbones for four different shRNAmiRs. **(E)** Quantitative comparison of miR-(E) and miR-(N) knockdown potencies. Data represents mean + standard deviation (sd) from five shRNAs, each with minimum three values from independent experiments (W-test, P = 0.87). **(F)** Comparison of miR-E and miR-N potency by western blotting. LTR, long terminal repeat. ψ, packaging signal. gag, group-specific antigen. RRE, Rev response element. PPT, polypurine tract. SFFV, spleen focus-forming virus promoter. EGFP, enhanced green fluorescent protein. miR-N, micro RNA (miR) backbone for shRNA expression. wPRE, Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element. BFP, blue fluorescent protein. shRNAmiR, miR embedded short hairpin RNA. rtTA, reverse tetracycline-controlled transactivator.

Figure 2: Sanger and deep sequencing validate parallel and pooled miR-N cloning

(A) Left panel: Analysis of n = 379 clones by Sanger sequencing for mutated shRNAmiRs. Percentages apply to successful sequencing runs. Right panel: Frequency of different shRNA mutations detected by Sanger sequencing after cloning (n = 36). del, deletion. Misc, Miscellaneous. ps, passenger strand. gs, guide strand. **(B)** Overview of pooled cloning in two different pool sizes (about 3x350 or 1050 oligonucleotides) into pLenti.c and subsequent transfer to pGamma.TTMPVIV-(N) via restriction digest (RE). **(C)** Negative correlation of bacterial colonies with number of pooled oligonucleotides (shRNAmiRs). Data was acquired with OpenCFU (Geissmann 2013) from >3 experiments. **(D,E)** Uniform raw read counts of shRNAmiRs in all libraries from pooled cloning assessed by deep sequencing of PCR amplified guide strands and compared to a manually assembled small library from individual DNA preparations (n > 500, except pGamma.c_Manual with n = 96). Raw reads **(D)** and density distribution **(E)**. **(F,G)** Scatter plots for detailed pairwise comparison of read counts from different libraries (n > 500).

Figure 3: A new lentiviral vector with improved properties

(A) Leakiness of pGamma.TTMPVIR-(E).shRptor.2 as demonstrated by knockdown of Rptor in the absence of doxycycline. **(B,C,D)** Significantly reduced leakiness of pLenti.TTMPVIV-(N) compared to pGamma.TTMPVIV-(N) by ~60 % measured by western blotting **(B)** and quantitatively evaluated by KD reporter assays **(C,D)** without doxycycline treatment. **(B)** 32D cells were transduced with VSVg pseudotyped pLenti.TTMPVIV-(N) or pGamma.TTMPVIV-(N) harboring the respective shRNAs with a transduction rate of less than 20 %. To ensure a high percentage of functionally transduced cells for western blotting, transduced cells were sorted for $Venus^{+}dTomato^{+}$ after short doxycycline incubation. Western blotting was performed after several days without doxycycline as described in materials and methods. L, pLenti.TTMPVIV-(N). G, pGamma.TTMPVIV-(N). **(C)** Representative flow cytometry plots with non-treated (nt) cells in the upper panel and doxycycline (Dox) treated cells in the lower panels. Black lines indicate the center of the reporter populations. For non-treated (nt) cells in the upper flow cytometry plots, red and blue lines indicate the difference of transduced cells to reporter population (leakiness). For doxycycline treated cells in the lower panel, arrows indicate a population expressing dTomato but not the shRNA. **(D,E)** Data from 3 different shRNAs (shLuc, shPten.1, shRptor.1) measured with the KD reporter assay without (t-test, P = 0.0005) **(D)** or with (W-test, P = 1) **(E)**

doxycycline. Data were normalized to pGamma.TTMPVIV-(N) to compensate for different shRNA potencies. **(F)** Representative western blot confirming tightness (no visible leakiness) and potency of pLenti.TTMPVIV-(N) in 32D cells. **(G,H)** Significantly increased doxycycline sensitivity of rtTA-V10 versus rtTA3 assessed by percentage of dTomato^{+} cells (W-test, $p < 2E-5$) **(G)** or dTomato MFI (W-test, $p < 2.6E-4$) **(H)**, both normalized to lowest and highest doxycycline concentrations. Depicted is a local regression (LOESS) over two independent experiments and 4-5 different shRNAmiRs, respectively. Significance was calculated by estimation of area under curve. **(I)** Schematic setting of the inducible pLenti.TTMPVIV-(N). Changes to original vector and expression cassette are indicated by arrows. Abbreviations of vector elements in addition to Fig. 1. dTom, dTomato. PGK, phosphoglycerate kinase promoter (constitutive). IRES, internal ribosome entry site. rtTA, reverse tetracycline-controlled transactivator. cPPT, central polypurine tract. SV40, simian virus 40. CMV IE, cytomegalovirus immediate-early. **(I)** Titer determined on HT1080 for VSVg pseudotyped pGamma.TTMPVIV-(N) and pLenti.TTMPVIV-(N) (t-test, $P = 0.032$). Data for each vector represents the titer for 4 different shRNAmiRs calculated from 6 dilutions. Each dilution was measured in triplicate. Only values below 35 % transduction rate were included.

Figure 4: Functional validation of our optimized lentiviral vector

(A) Representative western blot illustrating knockdown of PTEN by pLenti.TTMPVIV-(N) in Venus^{+}dTomato^{+}-sorted human K-562 cells after 10 or 15 days with or without doxycycline. **(B)** Long-term KD reporter assay on RagMEF cells with two different MOIs and shRNAmiRs over more than 20 days. Doxycycline was added on day 0. **(C,D,E)** Functional validation of pLenti.TTMPVIV-(N) by long-term expression of control (shLuc) or validated depleting shRNAmiRs (shMYB and shDOT1L) over more than 30 days. **(C)** Schematic illustration of a functional depletion experiment. **(D)** Doxycycline (Dox) treated group showing shRNAmiR expressing cells. **(E)** Non-treated group shows stably maintained transduced cells. **(F)** Flow cytometry plot for shLuc and shMYB of doxycycline treated cells at day 31. shRNA+, population expressing the shRNAmiR. shRNA^{-}, population not expressing the shRNAmiR. **(G,H)** Competitive assay of depleting shMYB.1 in comparison to shLuc expressed from pLenti.TTMPVIV-(N) in ML-2 cells within one cell culture dish. shMYB.1 was coupled to dTomato **(G)** or E2-Crimson **(H)**, shLuc to the corresponding color, respectively.

Figure 5: shRNA design and sequence features for potent RNAi triggers

(A) Potency of shRNAs prefiltered for the Sensor criteria. Percentage of shRNAmiRs with the respective knockdown potency. Dotted straight lines indicate that 85 % of the shRNAmiRs reached a knockdown of 70 %. n = 277 shRNAs individually measured with KD reporter assays. **(B)** Sequence logos for the most potent 20 % (upper panel) and weakest 20 % (lower panel) of shRNAs. **(C)** Diagram indicating the process of feature extraction, model building, prediction and RMSE (root-mean-square error) calculation on 90 % of the Fellmann dataset (n = 18720). **(D)** Discrimination of the most predictive features by RMSE with a threshold of RMSE at half maximum density (black line). **(E)** Fellmann dataset prefiltered for Sensor criteria (n = 2648) and used for model building on all possible feature combinations that were detected before in addition to the Sensor criteria. **(F)** Decreasing RMSE with model complexity. Black line links the best model for each complexity (number of combined features). **(G)** Feature weights for the best model with complexity 9. **(H,I)** Grouped boxplot showing the significant correlation between prediction with our model and **(H)** the Fellmann Score for the remaining 10 % of the Fellmann dataset (W-test, P = 0.006) or **(I)** our completely independent, individually measured dataset (W-test, P = 1.1E-4). Low, shRNAs within lower third of score. High, shRNAs within higher third of score.

Table 1: Comparison of shRNA expression vectors

Overview of major differences between shRNA expression vectors and their respective properties. pGamma.TTMPVIR-(E), T3.dTomato.miRE.PGK.Venus.IRES.rtTA3. pLenti.TTMPVIV-(N), pL40C.T3.dTomato.miRN.PGK.Venus.IRES.rtTA-V10.WPRE. SFFV, spleen focus-forming virus promoter. SV40, simian virus 40. CMV IE, cytomegalovirus immediate-early. rtTA, reverse tet-transactivator. 4-OHT, 4-Hydroxytamoxifen.

Table 2: Sequence features for identification of potent RNAi triggers

Column “Sensor criteria” lists the sequence features that were defined previously [16]. Reanalysis of the dataset resulted in additional features, indicated in the column “Fellmann data”. Our model is based on the difference between the additional features from our reanalysis and the original Sensor criteria. Sensor criteria, as listed in [49]: Sensor1, A/T at position 1. Sensor2, A/T content 40 – 80 %. Sensor3, A/T content in positions 1 – 14 >50 %. Sensor4, ratio (A/T% 1 – 14) / (A/T% 15 – 22) > 1.

Sensor5, no A at position 20. Sensor6, A/T at position 13 or T at position 14. Sensor7, no “AAAAAA”, “TTTTT”, “CCCC” or “GGGG” pattern. AT114, A/T content in positions 1-14.

Supplementary Data

Supplement A

6 supplementary figures and 3 supplementary tables.

Supplement B

Supplementary data set containing the results of all shRNA KD reporter assays. Given is the shRNA name, forward oligonucleotide sequence and the respective knockdown measured with KD reporter assays.

Funding

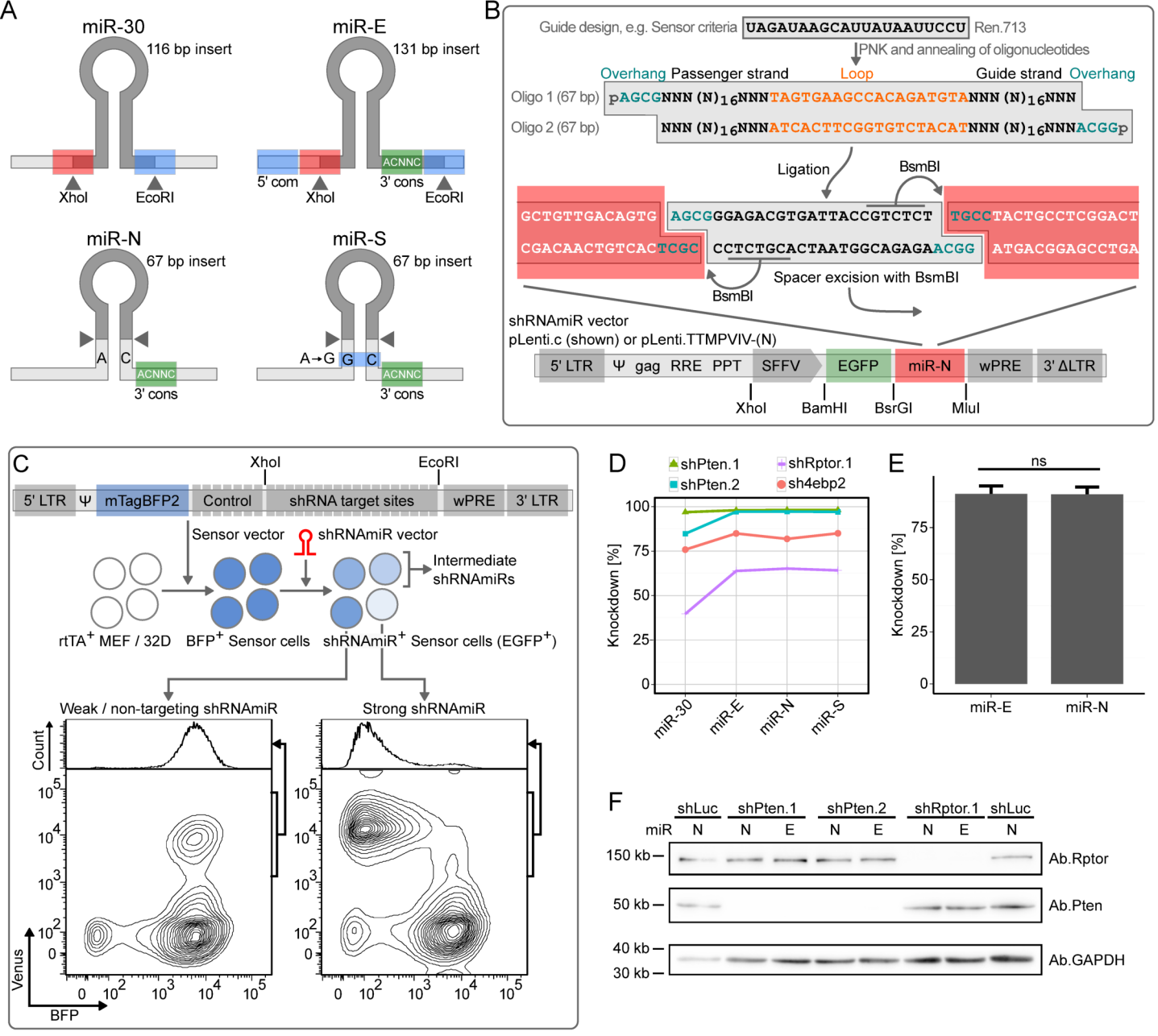
This work was supported by DFG [SFB738, REBIRTH Cluster of Excellence EXC62/1, HE 5240/5-1, HE 5240/6-1], EU [PERSIST, CELL-PID, SCIDNET], Hannover Medical School [HiLF and Young Academy-Grants to AdS]; and Deutsche Krebshilfe [110284, 110287, 110292 and 111267, 111743].

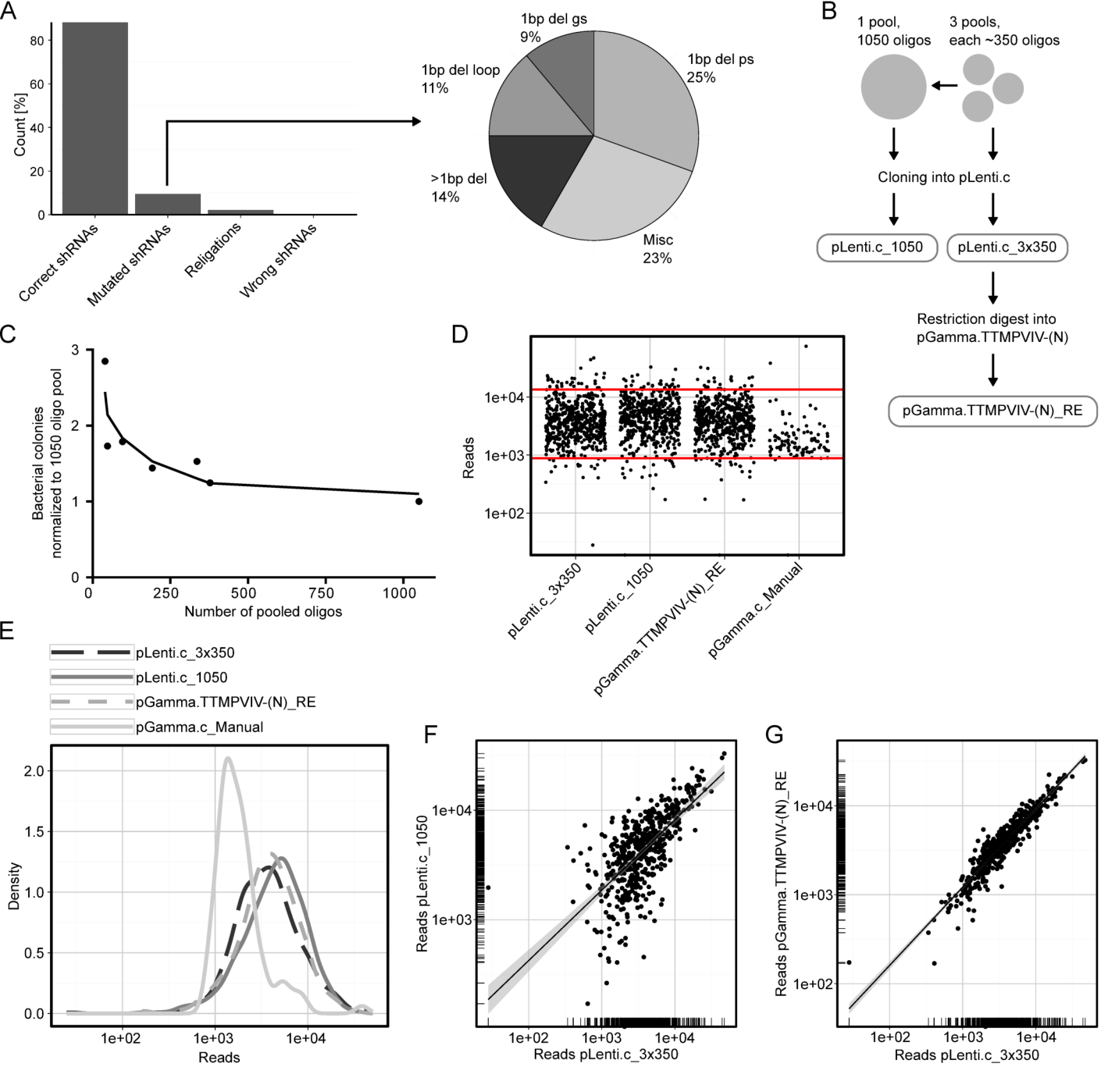
Author Contributions

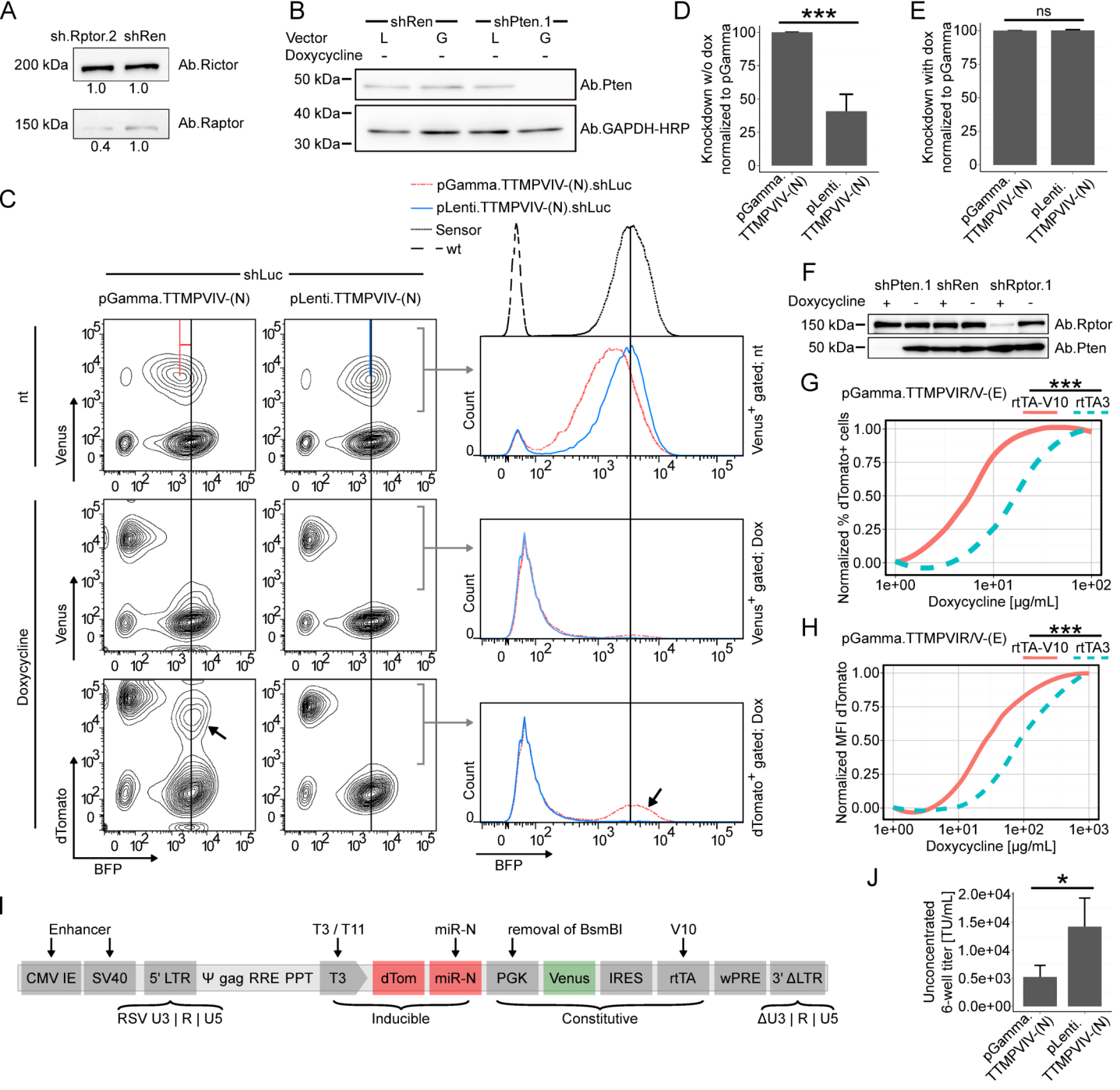
FFA, DH, AxS and AdS designed experiments. FFA and AdS performed experiments and analyzed data. JZ and TH provided unpublished reagents, protocols and technical advice. FFA, AdS and ST performed statistical analyses. FFA wrote the online tool. MH and AK designed and conducted experiments with human primary leukemia cells. FFA and FT performed and analyzed NGS-sequencing. FFA, AxS and AdS wrote the manuscript with input from all co-authors.

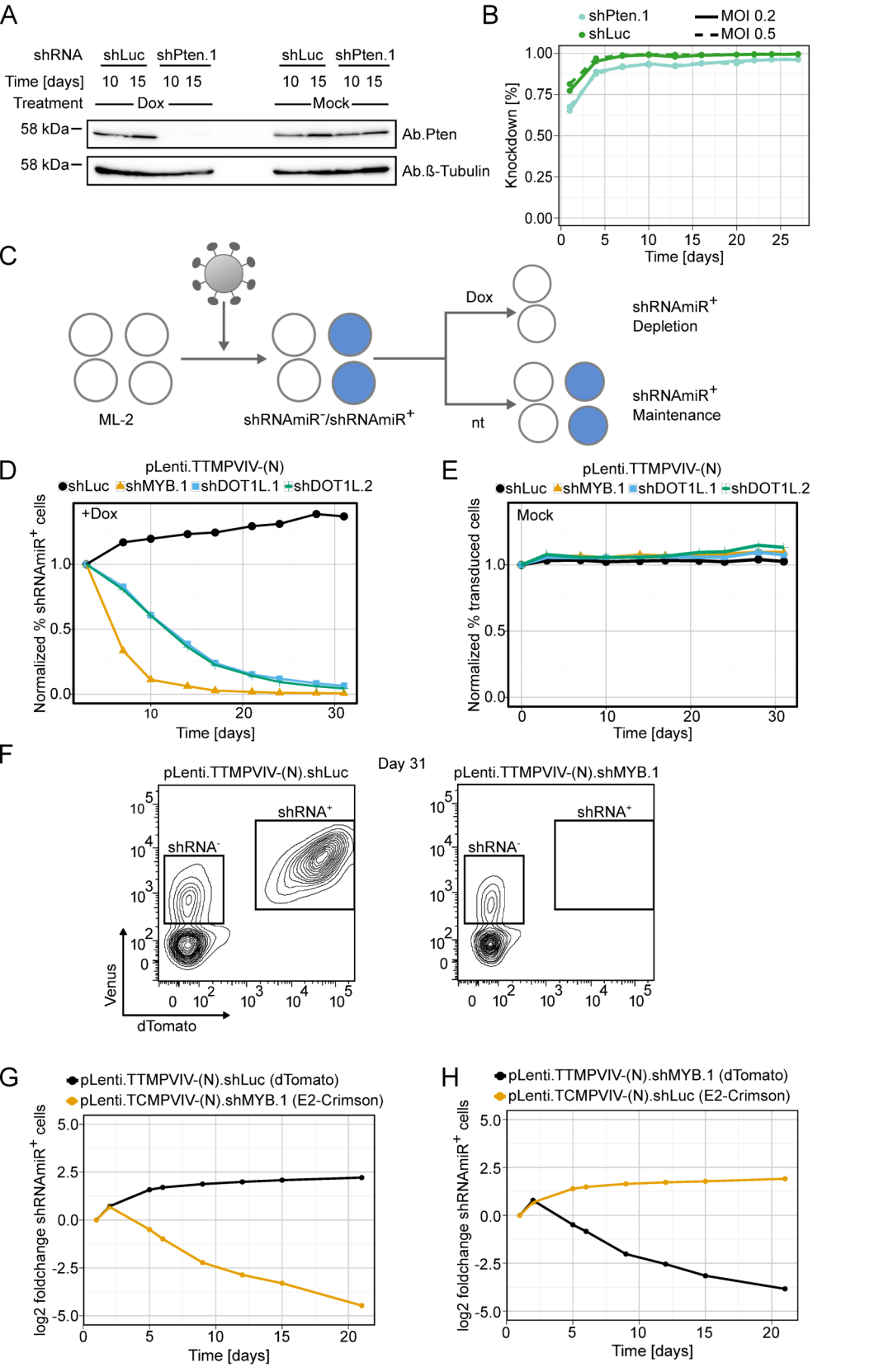
Conflict of interest disclosure

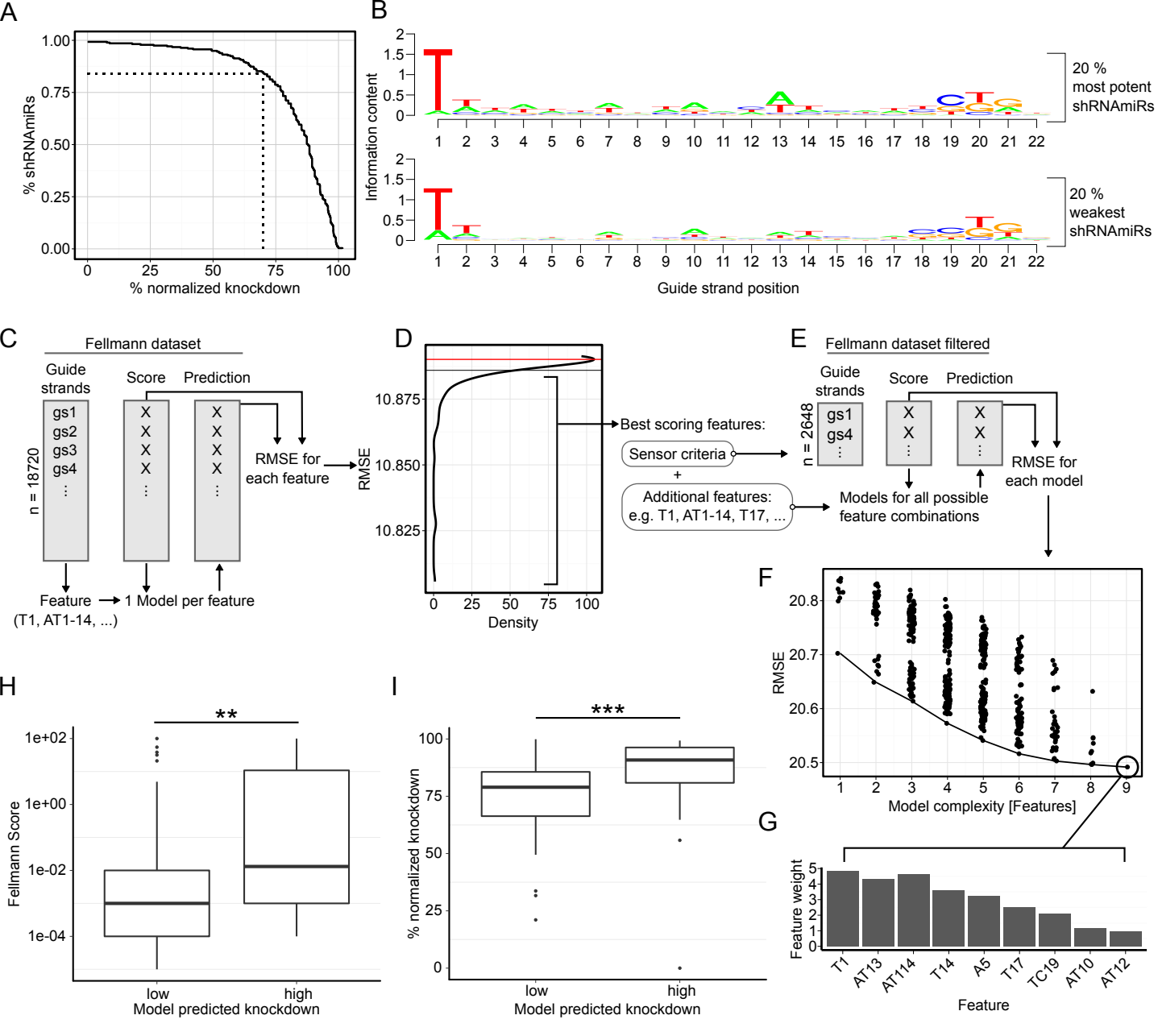
The authors confirm that there are no personal or financial conflicts of interest to declare. All authors have approved the final article.


A
miR-30
116 bp insert
XhoI
EcoRI
miR-E
131 bp insert
5' com
XhoI
ACNNC
3' cons
EcoRI
miR-N
67 bp insert
A
C
ACNNC
3' cons
miR-S
67 bp insert
A→G
G
C
ACNNC
3' cons
B
Guide design, e.g. Sensor criteria
UAGAUAAGCAUUAUAAUUCCU
Ren.713
PNK and annealing of oligonucleotides
Overhang
Passenger strand
Loop
Guide strand
Overhang
Oligo 1 (67 bp)
pAGCGNNN(N)16NNNTAGTGAAGCCACAGATGTANNN(N)16NNN
Oligo 2 (67 bp)
NNN(N)16NNNATCACTTCGGTGTCTACATNNN(N)16NNNACGGp
Ligation
BsmBI
GCTGTTGACAGTG
AGCGGGAGACGTGATTACCGTCTCT
TGCCTACTGCCTCGGACT
CGACAACTGTCACTCGC
CCTCTGCACTAATGGCAGAGAACGG
ATGACGGAGCCTGA
BsmBI
Spacer excision with BsmBI
shRNAmiR vector
pLenti.c (shown) or pLenti.TTMPVIV-(N)
5' LTR
Ψ
gag
RRE
PPT
SFFV
EGFP
miR-N
wPRE
3' ΔLTR
XhoI
BamHI
BsrGI
MluI
C
XhoI
EcoRI
5' LTR
Ψ
mTagBFP2
Control
shRNA target sites
wPRE
3' LTR
Sensor vector
shRNAmiR vector
Intermediate shRNAmiRs
rtTA+ MEF / 32D
BFP+ Sensor cells
shRNAmiR+ Sensor cells (EGFP+)
Weak / non-targeting shRNAmiR
Strong shRNAmiR
Count
Venus
BFP
D
shPten.1
shPten.2
shRptor.1
sh4ebp2
Knockdown [%]
miR-30
miR-E
miR-N
miR-S
E
ns
Knockdown [%]
miR-E
miR-N
F
shLuc
shPten.1
shPten.2
shRptor.1
shLuc
miR
N
N
E
N
E
N
E
N
150 kb
Ab.Rptor
50 kb
Ab.Pten
40 kb
30 kb
Ab.GAPDH


A
Count [%]
80
60
40
20
0
Correct shRNAs
Mutated shRNAs
Religations
Wrong shRNAs
1bp del gs
9%
1bp del ps
25%
1bp del loop
11%
>1bp del
14%
Misc
23%
B
1 pool,
1050 oligos
3 pools,
each ~350 oligos
Cloning into pLenti.c
pLenti.c_1050
pLenti.c_3x350
Restriction digest into
pGamma.TTMPVIV-(N)
pGamma.TTMPVIV-(N)_RE
C
Bacterial colonies
normalized to 1050 oligo pool
3
2
1
0
0
250
500
750
1000
Number of pooled oligos
D
Reads
1e+04
1e+03
1e+02
pLenti.c_3x350
pLenti.c_1050
pGamma.TTMPVIV-(N)_RE
pGamma.c_Manual
E
pLenti.c_3x350
pLenti.c_1050
pGamma.TTMPVIV-(N)_RE
pGamma.c_Manual
Density
2.0
1.5
1.0
0.5
0.0
1e+02
1e+03
1e+04
Reads
F
Reads pLenti.c_1050
1e+04
1e+03
1e+02
1e+03
1e+04
Reads pLenti.c_3x350
G
Reads pGamma.TTMPVIV-(N)_RE
1e+04
1e+03
1e+02
1e+02
1e+03
1e+04
Reads pLenti.c_3x350


A
sh.Rptor.2 shRen
200 kDa
Ab.Rictor
1.0 1.0
150 kDa
Ab.Raptor
0.4 1.0
B
shRen shPten.1
Vector L G L G
Doxycycline - - - -
50 kDa
Ab.Pten
40 kDa
30 kDa
Ab.GAPDH-HRP
D
***
Knockdown w/o dox normalized to pGamma
pGamma. TTMPVIV-(N)
pLenti. TTMPVIV-(N)
E
ns
Knockdown with dox normalized to pGamma
C
pGamma.TTMPVIV-(N).shLuc
pLenti.TTMPVIV-(N).shLuc
Sensor
wt
shLuc
pGamma.TTMPVIV-(N)
pLenti.TTMPVIV-(N)
nt
Doxycycline
Venus
dTomato
BFP
Count
Venus+ gated; nt
Venus+ gated; Dox
dTomato+ gated; Dox
F
shPten.1 shRen shRptor.1
Doxycycline + - + - + -
150 kDa
Ab.Rptor
50 kDa
Ab.Pten
G
pGamma.TTMPVIR/V-(E)
rtTA-V10 rtTA3
***
Normalized % dTomato+ cells
Doxycycline [µg/mL]
H
pGamma.TTMPVIR/V-(E)
rtTA-V10 rtTA3
***
Normalized MFI dTomato
Doxycycline [µg/mL]
I
Enhancer
T3 / T11
miR-N
removal of BsmBI
V10
CMV IE
SV40
5' LTR
Ψ gag RRE PPT
T3
dTom
miR-N
PGK
Venus
IRES
rtTA
wPRE
3' ΔLTR
RSV U3 | R | U5
Inducible
Constitutive
ΔU3 | R | U5
J
*
Unconcentrated 6-well titer [TU/mL]
pGamma. TTMPVIV-(N)
pLenti. TTMPVIV-(N)


A
shRNA
shLuc
shPten.1
shLuc
shPten.1
Time [days]
10 15 10 15
10 15 10 15
Treatment
Dox
Mock
58 kDa
Ab.Pten
58 kDa
Ab.ß-Tubulin
B
shPten.1
shLuc
MOI 0.2
MOI 0.5
Knockdown [%]
Time [days]
C
ML-2
shRNAmiR⁻/shRNAmiR⁺
Dox
nt
shRNAmiR⁺
Depletion
shRNAmiR⁺
Maintenance
D
pLenti.TTMPVIV-(N)
shLuc
shMYB.1
shDOT1L.1
shDOT1L.2
+Dox
Normalized % shRNAmiR⁺ cells
Time [days]
E
pLenti.TTMPVIV-(N)
shLuc
shMYB.1
shDOT1L.1
shDOT1L.2
Mock
Normalized % transduced cells
Time [days]
F
Day 31
pLenti.TTMPVIV-(N).shLuc
pLenti.TTMPVIV-(N).shMYB.1
shRNA⁻
shRNA⁺
Venus
dTomato
G
pLenti.TTMPVIV-(N).shLuc (dTomato)
pLenti.TCMPVIV-(N).shMYB.1 (E2-Crimson)
log2 foldchange shRNAmiR⁺ cells
Time [days]
H
pLenti.TTMPVIV-(N).shMYB.1 (dTomato)
pLenti.TCMPVIV-(N).shLuc (E2-Crimson)
log2 foldchange shRNAmiR⁺ cells
Time [days]


A
% shRNAmiRs
% normalized knockdown
B
Information content
20 % most potent shRNAmiRs
20 % weakest shRNAmiRs
Guide strand position
C
Fellmann dataset
Guide strands
Score
Prediction
n = 18720
gs1
gs2
gs3
gs4
RMSE for each feature
Feature (T1, AT1-14, ...)
1 Model per feature
D
RMSE
Density
Best scoring features:
Sensor criteria
+
Additional features: e.g. T1, AT1-14, T17, ...
E
Fellmann dataset filtered
n = 2648
gs1
gs4
Models for all possible feature combinations
RMSE for each model
F
RMSE
Model complexity [Features]
G
Feature weight
T1
AT13
AT114
T14
A5
T17
TC19
AT10
AT12
Feature
H
**
Fellmann Score
low
high
Model predicted knockdown
I
***
% normalized knockdown
low
high
Model predicted knockdown

	Vector	pGamma.TTMPVIR-(E)	pLenti.c	pLenti.TTMPVIV-(N)	FKBP12-HA.FLPs-ERT2
	Reference	[12,14]	[18]		[48]
	Genus	Gammaretrovirus	Lentivirus	Lentivirus	Lentivirus
Components	Enhancer			SV40, CMV IE	
	shRNA promoter	T3	SFFV	T3	SFFV
	Transactivator	rtTA3		rtTA-V10	
Properties	miR backbone	miR-30E	miR-30N	miR-30N	miR-30
	miR potency	High	High	High	Low
	shRNA cloning	Several steps	1-step	1-step	Several steps
	shRNA expression	Constitutive	Constitutive	Dox inducible	Shield-1 + 4-OHT
	Reversibility	Yes	Yes	Yes	No
	Leakiness	High		Low	None
	Dosage control	Yes		Yes	No
	Titer	Low	High	Medium	Low

Base Position	Sensor criteria	Fellmann data	Model
1	AT1	T1	T1
2			
3			
4			
5		A5	A5
6			
7			
8			
9			
10		TA10	TA10
11			
12		TA12	TA12
13	AT13 or T14	AT13	AT13
14		T14	T14
15			
16			
17		T17	T17
18			
19		TC19	TC19
20	no A20	no A20	
21			
22			
na	Sensor2	Sensor2	
na		AT114	AT114
na	Sensor3	Sensor3	
na	Sensor4	Sensor4	
na	Sensor6	Sensor6	
na	Sensor7	no Sensor7	