BMC
Genomics

METHODOLOGY ARTICLE   **Open Access**

Functional microRNA screening using a comprehensive lentiviral human microRNA expression library

Jos B Poell[1], Rick J van Haastert[2], Francesco Cerisoli[2], Anne S Bolijn[3], Lisette M Timmer[3], Begoña Diosdado-Calvo[3], Gerrit A Meijer[3], Andre AFL van Puijenbroek[2], Eugene Berezikov[1,2], Roel QJ Schaapveld[2] and Edwin Cuppen[1,2*]

Abstract

Background: MicroRNAs (miRNAs) are a class of small regulatory RNAs that target sequences in messenger RNAs (mRNAs) to inhibit their protein output. Dissecting the complexities of miRNA function continues to prove challenging as miRNAs are predicted to have thousands of targets, and mRNAs can be targeted by dozens of miRNAs.

Results: To systematically address biological function of miRNAs, we constructed and validated a lentiviral miRNA expression library containing 660 currently annotated and 422 candidate human miRNA precursors. The miRNAs are expressed from their native genomic backbone, ensuring physiological processing. The arrayed layout of the library renders it ideal for high-throughput screens, but also allows pooled screening and hit picking. We demonstrate its functionality in both short- and long-term assays, and are able to corroborate previously described results of well-studied miRNAs.

Conclusions: With the miRNA expression library we provide a versatile tool for the systematic elucidation of miRNA function.

Background

MicroRNAs (miRNAs) were discovered as a class of small regulatory molecules ten years ago [1-3]. These ~21 nucleotide (nt), small RNAs recognize partially complementary sequences on target mRNAs [4,5]. Following the initial discovery of miRNAs, substantial effort has gone into characterization of the canonical miRNA pathway [6,7] and into miRNA discovery; by identifying miRNAs in more species and by adding to the list of known miRNAs [8]. Although cDNA cloning and northern blotting techniques can be used to detect the most abundant miRNAs, the advent of massively parallel sequencing technologies has propelled the miRNA field, allowing for both discovery and quantification of all miRNAs in a given sample [9,10].

With the bulk of the miRNAs revealed in commonly studied species, the next challenge lies in elucidating the

biological processes in which miRNAs play a role. Current bioinformatic approaches rely on the identification of partially complementary sequences in mRNAs to predict miRNA targeting. Yet these approaches still come with one limitation; the exact parameters governing targeting remain unknown. Several prediction algorithms have been developed to overcome this difficulty by ascribing different weights to key parameters, such as binding energy between target and miRNA, conservation of the target site, quality of the "seed pairing", et cetera [11]. Still no single algorithm emerges as the best performer [12], and most algorithms predict thousands of targets for each miRNA [13]. Combining different target prediction algorithms generates shorter list of targets by creating more stringent cut-offs. This can provide some enrichment in true positives, but at the cost of more false negatives [14]. In addition to bioinformatics prediction, several approaches to genome-wide experimental miRNA target identification have been developed. These experiments utilize Argonaute pull-down assays (HITS-CLIP and PAR-CLIP) [15,16], changes in mRNA levels

* Correspondence: e.cuppen@hubrecht.eu
[1]Hubrecht Institute, Cancer Genomics Center, and University Medical Center Utrecht. Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
Full list of author information is available at the end of the article

[17], and protein expression after introduction or ablation of a specific miRNA [18-20]. These studies support that miRNAs indeed function by targeting hundreds of genes. Still, it is a daunting task to derive a function for a miRNA from these long lists of potential target genes. Despite progress in systematic approaches to find sets of related gene that are enriched within these long target lists [21], we are still far from satisfactory *in silico* prediction of miRNA function.

Alternatively, differential expression of a miRNA is commonly used to infer its function [22,23]. Identification of conditions where a specific miRNA is expressed versus an opposing condition where it is not, offers some clues as to the potential action of the miRNA. While this approach has been very successful in leading investigators to uncover miRNA functions, it still requires direct experimentation to prove effects due to miRNA activity beyond providing an only coincidental biomarker.

Another approach to determine the function of a miRNA is by knocking it down [24-26], or knocking it out, of the genome of a model organism [27-29]. Experimental knockdown of miRNAs may confirm or invalidate predicted functions, but it requires prior knowledge where a miRNA is expressed. Even with this knowledge, sufficient knockdown to demonstrate an observable effect is not guaranteed. Complete knockout delivers a clean result, but may not result in an obvious phenotype. Adding to this challenge is the possibility that many miRNAs may elicit only subtle changes or are redundant with other family members entirely. Indeed, only a fraction of all *C. elegans* miRNA families display pronounced abnormal phenotypes when deleted [30]. Given these challenges, knocking out a miRNA in mice or in a human cell line may often prove a fruitless endeavor.

In order to unravel the functions of specific miRNAs, while overcoming much of the challenges discussed above, we proposed to introduce or overexpress miRNAs in a system of interest. Moreover, we argue that it is more efficient to examine the effect of any miRNA for a predetermined phenotype, rather than blindly investigating one miRNA at a time. Such screens have been performed on different scales. Most are based on transfection of miRNA mimics [31]; synthetic RNAs that usually have a modified backbone. Although this approach ensures the presence of the mature miRNA in the target cells, a miRNA mimic is not processed via the canonical miRNA biogenesis pathway. This multi-cleavage process starts with the recognition of a hairpin in the primary transcript and ends with a mature miRNA produced from one or both of the arms of the hairpin [32]. Bypassing this physiological processing step has several implications. For instance, several variants of a

miRNA from the same primary transcript can arise due to variations in the processing pathway, such as arm switching, non-templated additions of adenines or uracils, and variations in the 3' and 5' cleavage sites [33,34]. Besides the loss of miRNA variants, the most-widely used miRNA mimics have modified backbones to increase stability within the cell. Consequently, they are not cleared naturally from the cells as endogenous miRNAs. Furthermore, it is possible that miRNA mimic transfection achieves cellular concentrations beyond physiological relevance. On the other hand, loading into the RISC complex, which is essential for biological activity, is not necessarily efficient for mimic transfection, as this process is known to be coupled to hairpin processing [35,36]. To express transcripts for endogenous processing into mature miRNAs, plasmid [37] or viral vectors [38,39] can be employed. While the retroviral library described by Agami and colleagues is a valuable tool for miRNA functional screens in pooled format, the viral supernatants are not available as individual isolates amenable to arrayed high-throughput screens [38].

Here, we describe the construction and application of a lentiviral human miRNA expression library. This library contains 660 annotated human miRNAs and 422 candidate miRNAs [10,40]. All are expressed from their genomic backbone, ensuring physiological processing of the miRNAs. The library is organized for high-throughput screening to provide a resource for the systematic elucidation of miRNA function. In addition, all lentiviral miRNA expression constructs can be applied individually to evaluate primary results. Lastly, we demonstrate the utility of this library in various types of screens to present the miRNA expression library as a versatile tool to study miRNA function.

Results and Discussion
Approach and setup
We aimed for the construction of a miRNA expression library that fulfills four criteria: 1) the library contains all human miRNAs; 2) each miRNA is represented in a separate stock to allow for arrayed screening; 3) the library can be used over a wide range of different cell types; 4) the miRNAs are swiftly and stably overexpressed. The first two criteria will be discussed below. The third and fourth criteria were met by choosing a lentiviral expression system, employing the pCDH vector. The lentiviral particles express VSV-G, a glycoprotein that grants broad tropism [41]. The glycoprotein's receptor is a lipid component of the plasma membrane that occurs on most cell types over a wide range of species. Lentiviruses integrate into the host genome [42], but unlike other retroviruses, lentiviruses do not require cell division for genomic integration [43,44]. For instance, the lentivirus is able to transduce quiescent

stem cells and terminally differentiated neurons [43]. By integrating into the host genome, the construct is retained through cell divisions and can be stably expressed for an indefinite period of time. We chose a lentiviral backbone with a puromycin-resistance cassette to be able to select for successfully transduced cells [45].

To create a comprehensive library containing all human miRNAs, we included all miRNAs known at the time we started composing the library. We obtained genomic loci from the widely adopted miRBase miRNA repository (version 14) [46]. We also included a set of candidate miRNAs from previous experiments [10,40]. Although we did not update the library after viral particles for the entire library were produced, some of the candidate miRNAs have since entered the registry. See additional file 1, table S1 for a list of the current annotation of all miRNAs in the library.
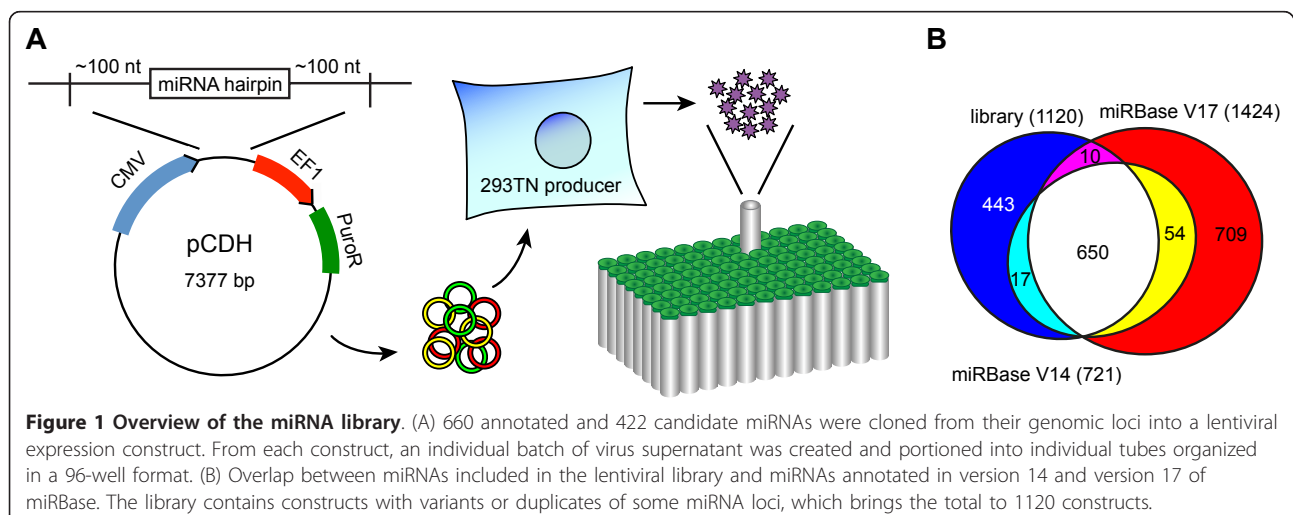
All miRNA loci, containing the full-length hairpin and ~100 nt flanking sequence on each side, were cloned into pCDH behind the CMV promoter. Care was taken to clone each miRNA separately, even when this miRNA resides in close proximity to another miRNA. After construction of the plasmid library, all constructs were packaged into lentiviral particles and portioned into individual tubes organized in ready-to-use 96-well format; the fact that tubes can be accessed separately adds versatility to the library. All virus supernatants were concentrated to titers of ~$5*10^8$ IFU/mL, in order to achieve efficient transduction with only small volumes of virus supernatant. This eliminates the need for infection under centrifugation or replacement of medium with virus-conditioned medium. While lentiviral particles were produced for all constructs individually, it should be mentioned the original plasmid library can also be applied in its own right. It can also be used for p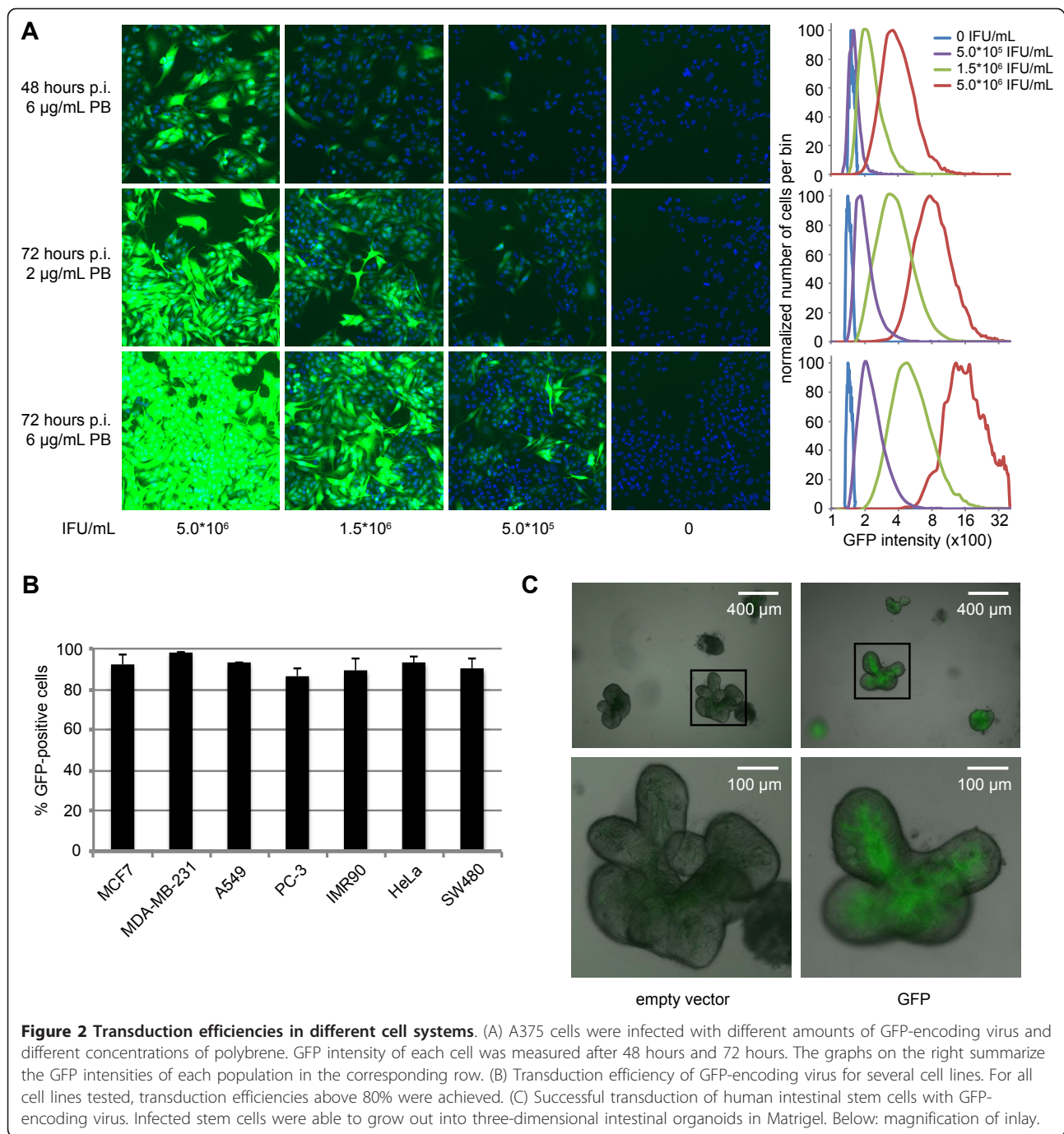lasmid-based screens, recloning purposes, or to create a pool of constructs containing all or a subset of miRNA constructs for generating mixed lentiviral stocks. Figure 1 is a schematic of the library construction and its most important features.

Library validation

The first step in the characterization of the library is the determination of transduction efficiency and optimal titration of the lentiviral particles. To this end, we used a variant of pCDH containing a GFP cassette in place of the puromycin-resistance cassette. We infected A375 melanoma cells with different virus concentrations, in combination with several concentrations of polybrene. While transduction efficiency reached over 90% under several conditions, increased amounts of virus led to brighter GFP expression within the population of cells (Figure 2a). This indicates multiple copies are integrated per cell when using higher titers. GFP expression was higher at 72 hours after infection than after 48 hours. This implies 72 hours post infection is a more sensitive time point for determination of transduction efficiency, rather than earlier timepoints. We further optimized conditions for virus infection in a panel of seven cell lines (Figure 2b). Notably, this panel includes the diploid fetal lung fibroblast line IMR-90; a non-transformed cell line [47]. Optimal transduction efficiencies were reached between 10^6 and 10^7 IFU/mL.

As mentioned, lentivirus is a favorable vector for its ability to transduce cells that are hard to transfect, such as stem cells. We have successfully transduced primary human intestinal cells using a GFP-encoding vector. These cells were subsequently used to create intestinal organoids *in vitro* [48]. After prolonged culture, entire organoid bodies turned fluorescent, indicating the stem cells from which the organoids developed were also successfully transduced (Figure 2c).

Figure 1 Overview of the miRNA library. (A) 660 annotated and 422 candidate miRNAs were cloned from their genomic loci into a lentiviral expression construct. From each construct, an individual batch of virus supernatant was created and portioned into individual tubes organized in a 96-well format. (B) Overlap between miRNAs included in the lentiviral library and miRNAs annotated in version 14 and version 17 of miRBase. The library contains constructs with variants or duplicates of some miRNA loci, which brings the total to 1120 constructs.

Figure 2 Transduction efficiencies in different cell systems. (A) A375 cells were infected with different amounts of GFP-encoding virus and different concentrations of polybrene. GFP intensity of each cell was measured after 48 hours and 72 hours. The graphs on the right summarize the GFP intensities of each population in the corresponding row. (B) Transduction efficiency of GFP-encoding virus for several cell lines. For all cell lines tested, transduction efficiencies above 80% were achieved. (C) Successful transduction of human intestinal stem cells with GFP-encoding virus. Infected stem cells were able to grow out into three-dimensional intestinal organoids in Matrigel. Below: magnification of inlay.

While ectopic expression of GFP is useful to determine transduction efficiency, it does not address whether miRNAs are also successfully overexpressed. We used a panel of ten test miRNAs to assess miRNA overexpression in five human cell lines: PC-3 (prostate cancer), MDA-MB-231 (breast cancer), A549 (lung cancer), IMR-90 (fetal lung fibroblast), and MCF-7 (breast cancer). Cells were infected with $1.0*10^7$ IFU/mL and subjected to puromycin selection from 24 hours after

infection. RNA was collected 48 hours (MCF-7), 72 hours (PC-3, MDA-MB-231 and A549) or 96 hours after infection (IMR-90). Relative mature miRNA expression was measured by qPCR. Values were standardized against the corresponding cell line infected with an empty vector control. We were able to overexpress all miRNAs in all cell lines, with the exception of miR-221 in IMR-90 cells (table 1). The fact that miR-221 is very effectively overexpressed in MCF7 indicates the

Table 1 Overexpression of miRNAs in various cell lines after lentiviral transduction

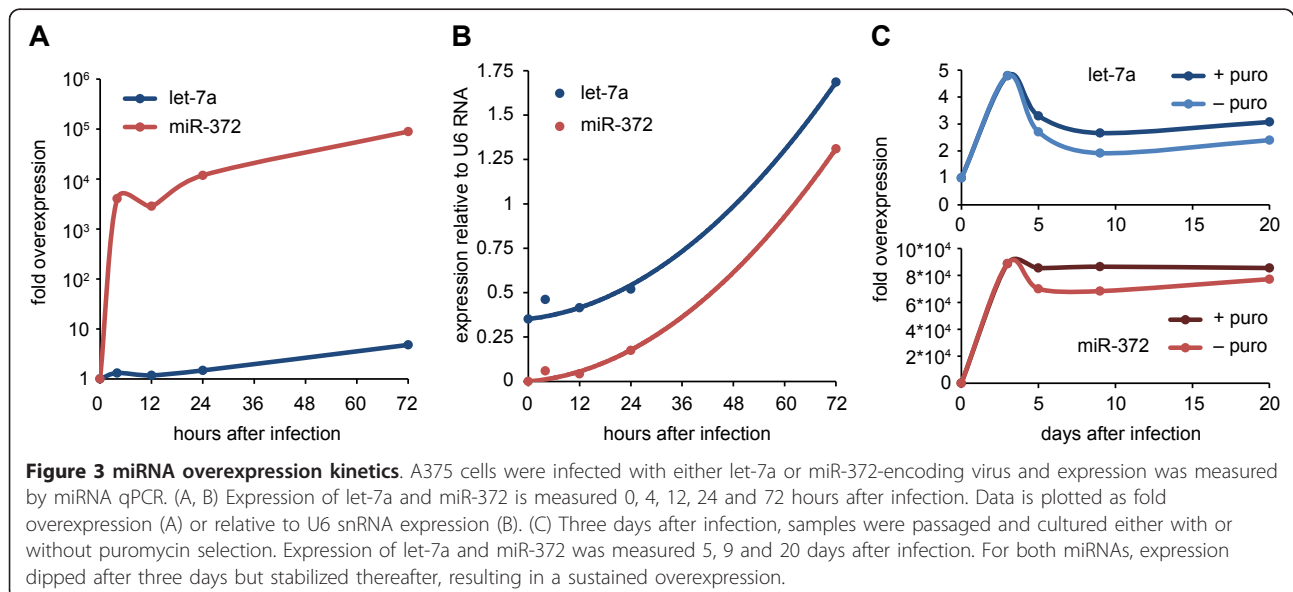| microRNA	cell line				
	PC-3	MDA-MB-231	A549	IMR-90	MCF7
miR-221	5.0	3.1	8.7	0.7	527
miR-203	> 1, 000	> 1, 000	> 1, 000	> 1, 000	45.6
miR-21	4.7	2.2	3.1	3.0	3.0
let-7a	11.8	114	18.9	17	38.3
miR-372	> 1, 000	> 1, 000	> 1, 000	> 1, 000	> 1, 000
miR-34a	> 1, 000	> 1, 000	26.0	17.6	170
miR-126	> 1, 000	> 1, 000	220	> 1, 000	293
miR-10b	> 1, 000	> 1, 000	321	406	> 1, 000
miR-335	> 1, 000	> 1, 000	> 1, 000	> 1, 000	> 1, 000
miR-205	> 1, 000	> 1, 000	14.1	> 1, 000	159

expression construct is functional. Besides, we were able to overexpress all other miRNAs in IMR-90. The results with miR-221 can at least partially be ascribed to very high endogenous miR-221 levels in this cell line (additional file 2, table S2) and also suggest that overexpression of miRNAs using our standard conditions remains within physiological levels.

Next, we examined miRNA expression kinetics after infection. To determine the optimal conditions for a screen, such as assay length and virus titer, it is important to know when and to what level the miRNA will be expressed. We have tested this extensively for let-7a and miR-372 in A375 cells. While let-7a is abundant in these cells, miR-372 is barely detectable (Figure 3). Both expression profiles show a quick surge of miRNA expression within 4 hours after infection. The early source of these miRNAs is unclear, but they may have

been present in the virus supernatant, encapsidated within the virus, or processed intracellularly from the viral RNA genome. miRNA levels exhibit a slight drop, but are quickly overtaken by expression from integrated constructs 16 to 20 hours after infection. Our data demonstrate that already after 24 hours, ectopic expression of miRNAs reaches comparable levels of endogenously predominant miRNAs, as evident from the expression relative to endogenous let-7a and U6 RNA (Figure 3b).

For long-term experiments, it is necessary for miRNA expression to be sustained. 72 hours after infection with let-7a or miR-372, we passaged A375 cells and cultured them either with or without puromycin selection. We examined miRNA expression by miRNA qPCR at 5, 9 and 20 days after infection (Figure 3c). A decline in miRNA expression was observed with both miRNAs between day 3 and day 5 after infection. This may be caused either by passaging or changed culture conditions, but it is also possible that it is a secondary effect of highly elevated miRNA processing. This effect was markedly stronger in the let-7a-infected cells, which may indicate that there is a selection disadvantage for cells with high let-7a expression. Despite the small drop after three days, overexpression of both miRNAs stabilized and persisted for weeks. Even after prolonged culture without puromycin selection, miRNA overexpression was sustained. We conclude there is little or no genetic loss or epigenetic silencing of the integrated construct.

As the goal of the library is to identify biologically relevant functions of miRNAs, we performed a proof-of-principle experiment by testing the well-documented effects of let-7a [49-51] as an inhibitor of cell

Figure 3 miRNA overexpression kinetics. A375 cells were infected with either let-7a or miR-372-encoding virus and expression was measured by miRNA qPCR. (A, B) Expression of let-7a and miR-372 is measured 0, 4, 12, 24 and 72 hours after infection. Data is plotted as fold overexpression (A) or relative to U6 snRNA expression (B). (C) Three days after infection, samples were passaged and cultured either with or without puromycin selection. Expression of let-7a and miR-372 was measured 5, 9 and 20 days after infection. For both miRNAs, expression dipped after three days but stabilized thereafter, resulting in a sustained overexpression.
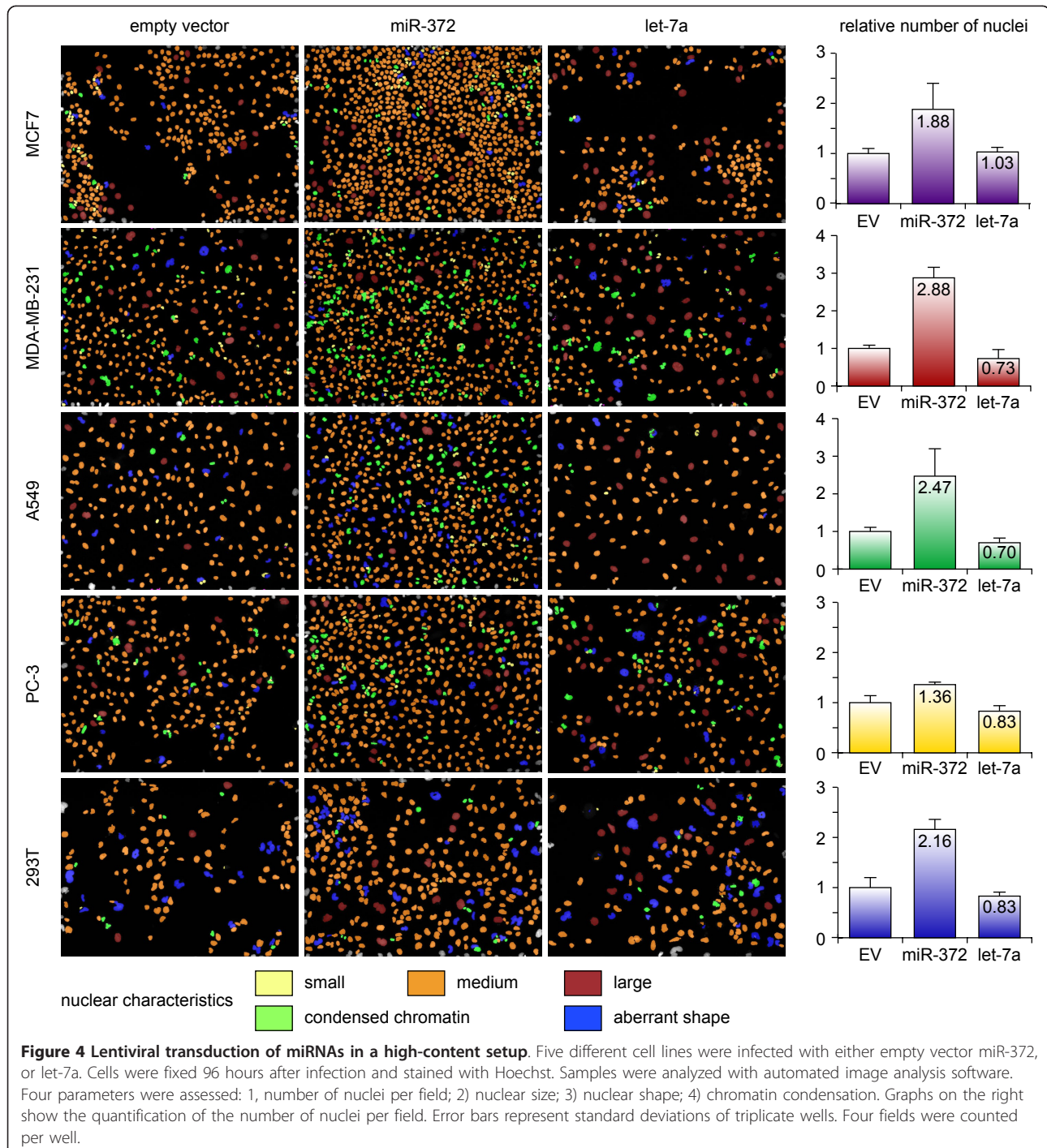
proliferation and miR-372 [38] as a stimulator in five cell lines. The experiment was set up in the fashion of a high-content screen. Cells were seeded in 384-well plates and infected after 8 hours. 72 hours after infection, cells were fixed and nuclei were stained with Hoechst. The automated image analysis software not only counts the number of nuclei, but also detects aberrations in nuclear shape and condensed chromatin

(Figure 4). Although we did not find any novel effects of the miRNAs on nuclear shape, we were able to confirm the expected effects of let-7a and miR-372 on proliferation in a high-content setup.

Clonal screening
As a first use of the complete library we screened miR-NAs for their effect on A375 melanoma cell

Figure 4 Lentiviral transduction of miRNAs in a high-content setup. Five different cell lines were infected with either empty vector miR-372, or let-7a. Cells were fixed 96 hours after infection and stained with Hoechst. Samples were analyzed with automated image analysis software. Four parameters were assessed: 1, number of nuclei per field; 2) nuclear size; 3) nuclear shape; 4) chromatin condensation. Graphs on the right show the quantification of the number of nuclei per field. Error bars represent standard deviations of triplicate wells. Four fields were counted per well.

proliferation. We tested whether we could detect proliferation effects in A375 cells in a 96-well setup, again using let-7a and miR-372 as test miRNAs. The effects were highly similar to those seen in Figure 4 (Figure 5a). Every single well infected with either let-7a or miR-372 proved a statistical outlier compared to empty-vector transduced wells (Figure 5a, right panel).

For the entire library, we performed an arrayed screen by MTS assay, which measures the viability of cells in a well [52]. This measure is reported to correlate strongly with cell number. All samples were tested in duplicate, and for each replicate a B-score was calculated [53]. This method calculates how a miRNA scores relative to the other miRNAs on the same plate. It is a robust method as it is not sensitive to outliers. B-scores of replicates are shown in Figure 5b. Also shown are the measures for all let-7a constructs and miR-372. The strong correlation between replicates indicates a high reproducibility of the assay. The effects of let-7a and miR-372 are similar as in the cell count assay. Importantly, all let-7a constructs cluster together. Since these constructs have different backbones but produce the same mature miRNA, we surmise the observed effects are likely to be caused by the mature miRNA. A confounding factor in short-term arrayed screens can be the range of virus titers of the different constructs. However, we saw no correlation between virus titer and cell viability (additional file 3, figure S1). Thus, with these experimental conditions, virus titers were not a significant factor contributing to toxicity.

Pooled screening

Some assays do not lend themselves to screening in 96-well format. In such cases the lentiviral library can be used for a pooled screen. There are two approaches: pool virus supernatant of several or all constructs together, or infect cells with the individual constructs and subsequently pool cells. The latter is generally favorable, because it precludes the possibility that two or more virus particles containing different miRNAs infect a single cell. A puromycin selection step can be added to make certain each construct is represented in transduced cells and to reduce background caused by non-transduced cells. We applied this approach to the poorly invasive HCT15 colorectal cell line. HCT15 cells were individually infected, puromycin-selected, and pooled in groups of 40 constructs. A fraction of the pool was used for genomic DNA isolation and the rest was subjected to two rounds of an invasion assay. Cells that successfully transmigrated through extracellular matrix were subcultured and genomic DNA was isolated. To identify the miRNA constructs that were enriched in the transmigrated fraction, we took advantage of the fact that the constructs are integrated into the host genome. We PCR-amplified the constructs from the genomic DNA using universal primers that anneal to the vector backbone

Figure 5 Arrayed screening using the lentiviral library. (A) A375 cells were seeded in a 96-well plate and infected with virus containing an empty vector, let-7a, or miR-372. Five days after infection, cells expressing let-7a were significantly less numerous than cells expressing empty vector, while cells expressing miR-372 were significantly more numerous (left panel, * indicates p < 10⁻⁴ in a two-tailed t-test). Values are average number of cells per well ± standard deviation, n = 6. Each measurement of both let-7a and miR-372 samples is separated from the empty vector samples by more than three standard deviations (Z-score, right panel). (B) Results of an arrayed screen on cell viability of A375 cells. Cells in individual wells were infected with single miRNA constructs and subjected to an MTS assay 5 days after infection. The screen was performed in duplicate. For each measurement a B-score was calculated. Correlation between duplicate B-scores is shown. The green dots represent let-7a-1, -2, and -3, while the red dot represents miR-372.

(Figure 6a). The PCR products were subjected to high-throughput sequencing. Relative abundance of the various constructs was derived from the sequence reads mapping to them. Hits were selected on the basis of a weighted score calculated from the enrichment and the number of sequence reads in the invasive fraction. We found evidence for 45 miRNAs that may enhance the invasive capacity of HCT15 cells (Figure 6b).

In a pooled invasion screen, as performed above, the possibility exists that a portion of the hits can be
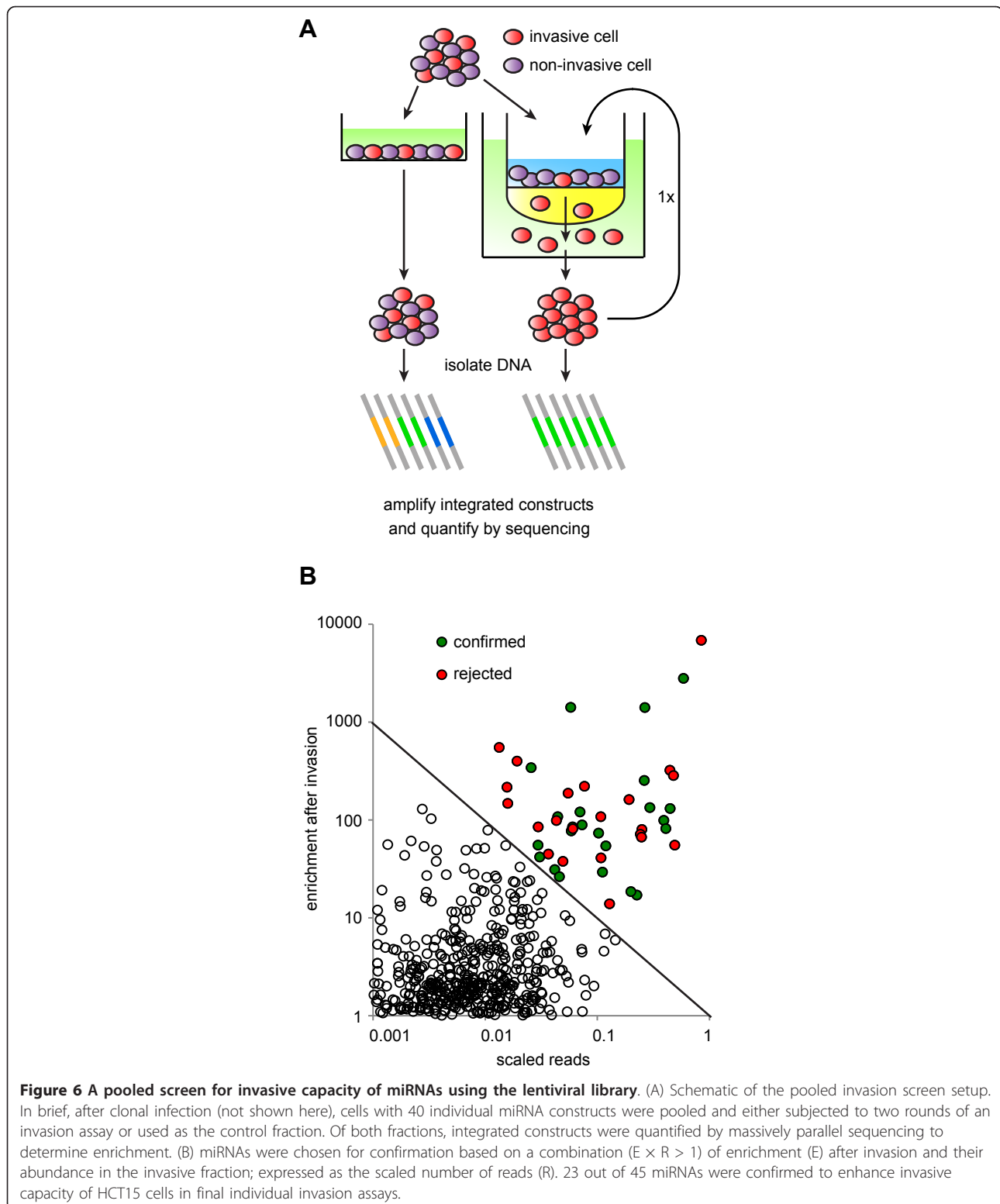
Figure 6 A pooled screen for invasive capacity of miRNAs using the lentiviral library. (A) Schematic of the pooled invasion screen setup. In brief, after clonal infection (not shown here), cells with 40 individual miRNA constructs were pooled and either subjected to two rounds of an invasion assay or used as the control fraction. Of both fractions, integrated constructs were quantified by massively parallel sequencing to determine enrichment. (B) miRNAs were chosen for confirmation based on a combination (E × R > 1) of enrichment (E) after invasion and their abundance in the invasive fraction; expressed as the scaled number of reads (R). 23 out of 45 miRNAs were confirmed to enhance invasive capacity of HCT15 cells in final individual invasion assays.

"passenger hits" that do not impart invasive capacity, but are present in cells that trail the truly invasive cells as they make their way through the extracellular matrix. To distinguish these two types, we subjected each hit individually to an invasion assay. We were able to confirm 23 miRNAs that increase invasiveness of HCT15 cells (Figure 6b).

Conclusions

We present a lentiviral miRNA expression library that is optimized for use in arrayed screens. The library allows miRNAs to be individually assessed in any gain-of-function screen. We envision the library being used in a wide range of functional screens. To name some possibilities: reporter-based assay screens, screens to study cell cycle, senescence, and epithelial-to-mesenchymal transition. The broad tropism of the lentivirus makes it applicable to infect cells of different species. This may prove particularly valuable, since lentivirus can be directly used *in vivo* [43]. Thus, the same lentivirus can be used to screen for a miRNA function *in vitro* and verify this functionality *in vivo*.

We have demonstrated the transduction of stem cells that were subsequently used for *in vitro* organ culture. The ability to stably transduce stem cells opens up avenues for studying miRNAs required for differentiation and stemness.

An arrayed library comes with several advantages. Arrayed screening offers more sensitivity and faster results than pooled screening, without the need of data deconvolution. It also allows for the assessment of several more complex parameters, as is commonplace in high-content screens. Assays that require long-term culture may not be amenable to arrayed screening and require pooling. Still, infecting cells individually before pooling is likely to decrease false-negative results in pooled screens, suggesting that even for pooled screens an arrayed library may be preferred.

Our library currently contains the majority of all human miRNAs. However, the most recent update (version 17) of miRBase includes a large number of newly identified miRNAs. While we do not wish to question their validity or understate their potential, we argue that most miRNA functions will be covered by the most highly expressed or broadly conserved miRNAs, which were already present in earlier versions of miRBase and thus represented in our library. By the same token, the merit of candidate miRNAs in our library can be questioned. In anticipation of this, candidate miRNAs were positioned on separate plates in the library, presenting researchers the choice to screen the entire library, only the annotated miRNAs, or only the candidate miRNAs.

Expression of the miRNAs from their native genomic background ensures the physiological processing of the miRNA. Not only the pre-miRNA hairpin, but also the flanking sequence contributes to proper processing [54]. Transcribing the miRNA from an integrated construct enables expression of both arms and all isomiRs that would naturally derive from the primary transcript. Therefore, the library could be employed for further characterization and validation of the included miRNAs. Such experiments have been extensively done for mouse miRNAs [55], but not for human miRNAs.

We have shown that ectopic expression of miRNAs using a lentiviral vector can be used to screen for biologically relevant effects. While the library is widely applicable and can be used to study various aspects of biology, our primary focus is on cancer-related processes. A first screen assessing miRNA-induced effects on melanoma cell growth demonstrated the value of the library in an arrayed screen. The next step is to perform such screens over a panel of cell lines to determine which miRNAs may have a growth effect on specific cancer indications, and which miRNAs have a general growth inhibitory or stimulatory function. These results will aid in finding miRNAs suited for tumor-specific treatment. Indeed, such therapeutic options have been successfully explored in a murine model of hepatocellular carcinoma [56]. With the miRNA expression library, we offer a platform that facilitates the identification of miRNAs with therapeutically relevant functions.

Methods

Construction of the lentiviral library and other constructs

Backbone for all constructs in the library is the lentiviral expression construct pCDH (cat. no. CD510B-1, System Biosciences). Individual loci containing a single miRNA hairpin were PCR amplified from human genomic DNA and cloned into the multiple cloning site of the plasmid. Loci were either obtained from miRBase http://mirbase.org/ or from mapping of candidate miRNAs found in previous experiments [10,40]. Primers were designed using Primer3. PCR was performed using Pfu polymerase (Agilent). The cloned fragments contain the full-length miRNA hairpin and approximately 100 flanking base pairs on both sides. An expression construct with EGFP was cloned by excision and ligation of the EGFP sequence from pEGFP-N1 (Clontech) into the multiple cloning site of pCDH. All constructs were packaged into lentiviral particles commercially by System Biosciences using the pPACKH1 HIV-based lentiviral packaging kit (cat. no. LV500A-1, System Biosciences). Viral particles were recieved in concentrated form with a median titer of $5.9*10^8$ IFU/mL in a 96-well format. Sequence of all inserts was confirmed from both the plasmid and the virus supernatant (we were able to obtain high-quality specific sequences from the virus supernatant using universal primers, most likely due to trace amounts of

plasmid). A list of all included miRNAs in their current annotation (miRBase17) is available in additional file 1, table S1. Lentiviral particles for pCDH with copGFP instead of puroR (cat. no. CD511B-1) were ordered separately from System Biosciences.

Cell culture and viral infections

293T, A375, MDA-MB-231, A549, MCF-7, IMR-90, PC-3, and HCT15 cells were maintained on 10% FCS complete medium: DMEM Glutamax (GIBCO) with 10% FCS (Sigma) supplemented with non-essential amino acids (GIBCO) and penicillin/streptomycin (GIBCO). All viral infections followed by RNA isolation were done 8 hours after seeding 10, 000 cells in a 6-well plate in 2 mL 10% FCS complete medium. All transduction efficiency experiments were performed by infection with pCDH-MCS-EF1-copGFP, except for the organoid culture, for which we used pCDH-EGFP-EF1-PuroR. Infection mix contained 2 μL virus supernatant, 12 μL 1 mg/mL polybrene (Sigma), and 86 μL PBS0, unless stated otherwise. Intestinal organoid bodies were cultured as described before [48]. Because the organoids are grown in Matrigel, cells were infected before seeding. Cells were infected in 250 μL Wnt-3a-conditioned medium containing 2 μL virus supernatant. Infection took place during a centrifugation step at 150 rcf for 1 hour at room temperature. Unless indicated otherwise, infections were performed in 96-well plates using the following set-up: 1000 cells were seeded in 100 μL 5% FCS complete medium per well of a 96-well plate and infected after 8 hours with 10 μL infection mix. 10 μL infection mix contained 0.6 μL 1 mg/mL polybrene, 0.5 μL virus supernatant, and 8.9 μL PBS0. In the arrayed screen, medium was replaced with 150 μL fresh medium 24 hours after infection.

MTS assay

Five days after infection, 30 μL MTS One Solution (Promega) was added to all samples. After each hour, plates were gently tapped to disperse the coloration of the MTS, and absorbance was measured at 492 nm. Last measurements were taken 4 hours after start of the assay. The time point with measurements showing the highest dynamic range without saturation of signal was used for data analysis.

Nuclear staining, GFP quantification and high-content applications

Five days after infection, 100 μL 8% PFA (Sigma) was added to all samples. Cells were fixed for 15 minutes. Samples were washed once with PBS0 followed by 10 minutes staining in 100 μL PBS0 containing 0.5 μg/mL Hoechst 33342 (Sigma). Cells were washed twice with PBS0 and kept on PBS0 at 4°C. Cells were quantified on

a Cellomics ArrayScan VTI using the accompanying software by counting nuclei in 4 fields per well under 10× magnification. Nuclei were identified as shapes with a contiguous Hoechst stain. Nuclear demarcations were used to quantify GFP intensity. All data acquisition was done using adaptions of the TargetActivation program of the ArrayScan software. Data and images displayed in table 1 and Figure 4 were generated by Cenix BioScience, GmbH.

RNA isolation and miRNA qPCRs

RNA was isolated using TRIzol (Invitrogen) isolation following the manufacturer's protocol. Small RNAs qPCR reactions were performed using the TaqMan MicroRNA reverse transcription kit (ABI), Taqman MicroRNA qPCR assays (ABI) and TaqMan Universal PCR Master Mix, No AmpErase UNG (ABI) using 10 ng total RNA input. qPCR reactions were set up using the suggested reaction conditions on a Bio-Rad MyiQ thermal cycler. U6 was used as a housekeeping control RNA in the experiments concerning Figure 3, RNU6B was used in the experiments concerning table 1. Relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. miRNA qPCR data in table 1 were generated by Cenix BioScience GmbH.

Invasion assay

HCT15 cells were seeded at 2500 cells per well in a 96-well plate in 10% complete medium. Cells were infected with individual virus supernatants (0.5 μL per well) after 24 hours. 72 hours after infection cells were selected with puromycin-containing medium and subsequently grown to 100% confluency. Pools were made of 40 samples per pool and grown for another 3 days on puromycin-containing medium. Half of the pooled culture was used for genomic DNA isolation, the other half was subjected to two rounds of invasion assay. For each pool, $2*10^5$ cells were applied to the upper compartment of a Boyden chamber (BD Fluoroblok 24-Multiwell, 8 μm pores) coated with extracellular matrix (ECM, Sigma) and containing serum-free medium, and allowed to invade the lower compartment containing 10% FCS medium. Cells were collected from the bottom compartment and expanded for 2-3 weeks on complete medium. $2*10^5$ cells of this subculture were applied to a second round of invasion as described above. Cells in the resulting bottom compartment were subcultured to be used for genomic DNA isolation.

Genomic DNA isolation and massively parallel sequencing

Genomic DNA was isolated from cells using a Qiagen Genomic Tip kit following the manufacterer's instructions. For both the invasive fraction and the control fraction, 10 ng DNA of each pool was pooled together

and subjected to PCR-amplification (20 cycles) of integrated constructs using Platinum PCR Supermix (Invitrogen) and primers flanking the inserts. In a subsequent, secondary (3 cycles) and tertiary (3 cycles) PCR the adapters and barcodes used for massively parallel sequencing were added to the flanks of the products. Samples were sequenced on the ABI SOLiD platform. The invasive and control fraction yielded $1.0*10^6$ and $0.9*10^6$ reads that mapped to integrated constructs respectively.

Library availability

Both the plasmid and virus library are publicly available when requested through InteRNA Technologies. Contact information is available at http://www.interna-technologies.com.

Additional material

Additional file 1: Table S1. Contains a list of the miRNAs included in the library and their cognate genomic sequences that were cloned into the vector, and associated virus titers.

Additional file 2: Table S2. Contains the endogenous levels of the set of 10 test miRNAs in the five tested cell lines. Expression calculated as $2^{-\Delta Ct}$ relative to RNU6B.

Additional file 3: Figure S1. Demonstrates the lack of correlation between virus titer and toxicity.

Author details

[1]Hubrecht Institute, Cancer Genomics Center, and University Medical Center Utrecht. Uppsalalaan 8, 3584 CT Utrecht, The Netherlands. [2]InteRNA Technologies BV, Padualaan 8, 3585 CH Utrecht, The Netherlands. [3]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands.

Authors' contributions

JBP constructed the plasmid library, designed and performed most experiments, analyzed all data and drafted the manuscript. RJvH constructed the plasmid library and designed and performed the arrayed screen. FC, ASB, LMT, BD and GAM designed and/or performed the pooled screen and its confirmation follow-up. AAFLvP was involved in the design of the library validation experiments and the arrayed screen. EB, RQJS and EC conceived and designed the lentiviral library. EC supervised all studies and critically revised the manuscript. All authors read and approved the final manuscript.

References

1. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of novel genes coding for small expressed RNAs.** *Science* 2001, **294(5543)**:853-858.
2. Lau NC, Lim LP, Weinstein EG, Bartel DP: **An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans.** *Science* 2001, **294(5543)**:858-862.
3. Lee RC, Ambros V: **An extensive class of small RNAs in Caenorhabditis elegans.** *Science* 2001, **294(5543)**:862-864.
4. Lai EC: **Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation.** *Nat Genet* 2002, **30(4)**:363-364.
5. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115(7)**:787-798.
6. Hutvagner G, Zamore PD: **A microRNA in a multiple-turnover RNAi enzyme complex.** *Science* 2002, **297(5589)**:2056-2060.
7. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425(6956)**:415-419.
8. Berezikov E, Cuppen E, Plasterk RH: **Approaches to microRNA discovery.** *Nat Genet* 2006, **38(Suppl)**:S2-7.
9. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ: **Elucidation of the small RNA component of the transcriptome.** *Science* 2005, **309(5740)**:1567-1569.
10. Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RH: **Diversity of microRNAs in human and chimpanzee brain.** *Nat Genet* 2006, **38(12)**:1375-1377.
11. Rajewsky N: **microRNA target predictions in animals.** *Nat Genet* 2006, **38(Suppl)**:S8-13.
12. Sethupathy P, Megraw M, Hatzigeorgiou AG: **A guide through present computational approaches for the identification of mammalian microRNA targets.** *Nat Methods* 2006, **3(11)**:881-886.
13. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.** *Cell* 2005, **120(1)**:15-20.
14. Alexiou P, Maragkakis M, Papadopoulos GL, Reczko M, Hatzigeorgiou AG: **Lost in translation: an assessment and perspective for computational microRNA target identification.** *Bioinformatics* 2009, **25(23)**:3049-3055.
15. Chi SW, Zang JB, Mele A, Darnell RB: **Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.** *Nature* 2009, **460(7254)**:479-486.
16. Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T: **Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP.** *Cell* 2010, **141(1)**:129-141.
17. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.** *Nature* 2005, **433(7027)**:769-773.
18. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N: **Widespread changes in protein synthesis induced by microRNAs.** *Nature* 2008, **455(7209)**:58-63.
19. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, **455(7209)**:64-71.
20. Jovanovic M, Reiter L, Picotti P, Lange V, Bogan E, Hurschler BA, Blenkiron C, Lehrbach NJ, Ding XC, Weiss M, Schrimpf SP, Miska EA, Grosshans H, Aebersold R, Hengartner MO: **A quantitative targeted proteomics approach to validate predicted microRNA targets in C. elegans.** *Nat Methods* 2010, **7(10)**:837-842.
21. Tsang JS, Ebert MS, van Oudenaarden A: **Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures.** *Mol Cell* 2010, **38(1)**:140-153.
22. Yi R, Poy MN, Stoffel M, Fuchs E: **A skin microRNA promotes differentiation by repressing 'stemness'.** *Nature* 2008, **452(7184)**:225-229.
23. Tavazoie SF, Alarcon C, Oskarsson T, Padua D, Wang Q, Bos PD, Gerald WL, Massague J: **Endogenous human microRNAs that suppress breast cancer metastasis.** *Nature* 2008, **451(7175)**:147-152.
24. Meister G, Landthaler M, Dorsett Y, Tuschl T: **Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing.** *RNA* 2004, **10(3)**:544-550.
25. Hutvagner G, Simard MJ, Mello CC, Zamore PD: **Sequence-specific inhibition of small RNA function.** *PLoS Biol* 2004, **2(4)**:E98.
26. Ebert MS, Sharp PA: **MicroRNA sponges: progress and possibilities.** *RNA* 2010, **16(11)**:2043-2050.
27. van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN: **Control of stress-dependent cardiac growth and gene expression by a microRNA.** *Science* 2007, **316(5824)**:575-579.
28. Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, Xue Y, Murphy A, Frendewey D, Valenzuela D, Kutok JL, Schmidt-Supprian M, Rajewsky N,

Yancopoulos G, Rao A, Rajewsky K: **Regulation of the germinal center response by microRNA-155.** *Science* 2007, **316(5824)**:604-608.

29. Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, Soond DR, van Dongen S, Grocock RJ, Das PP, Miska EA, Vetrie D, Okkenhaug K, Enright AJ, Dougan G, Turner M, Bradley A: **Requirement of bic/microRNA-155 for normal immune function.** *Science* 2007, **316(5824)**:608-611.

30. Miska EA, Alvarez-Saavedra E, Abbott AL, Lau NC, Hellman AB, McGonagle SM, Bartel DP, Ambros VR, Horvitz HR: **Most Caenorhabditis elegans microRNAs are individually not essential for development or viability.** *PLoS Genet* 2007, **3(12)**:e215.

31. Wang Y, Baskerville S, Shenoy A, Babiarz JE, Baehner L, Blelloch R: **Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation.** *Nat Genet* 2008, **40(12)**:1478-1483.

32. Kim VN: **MicroRNA biogenesis: coordinated cropping and dicing.** *Nat Rev Mol Cell Biol* 2005, **6(5)**:376-385.

33. Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K: **Complexity of the microRNA repertoire revealed by next-generation sequencing.** *RNA* 2010, **16(11)**:2170-2180.

34. Fernandez-Valverde SL, Taft RJ, Mattick JS: **Dynamic isomiR regulation in Drosophila development.** *RNA* 2010, **16(10)**:1881-1888.

35. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R: **TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing.** *Nature* 2005, **436(7051)**:740-744.

36. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R: **Human RISC couples microRNA biogenesis and posttranscriptional gene silencing.** *Cell* 2005, **123(4)**:631-640.

37. Silver SJ, Hagen JW, Okamura K, Perrimon N, Lai EC: **Functional screening identifies miR-315 as a potent activator of Wingless signaling.** *Proc Natl Acad Sci USA* 2007, **104(46)**:18151-18156.

38. Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, Liu YP, van Duijse J, Drost J, Griekspoor A, Zlotorynski E, Yabuta N, De Vita G, Nojima H, Looijenga LH, Agami R: **A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors.** *Cell* 2006, **124(6)**:1169-1181.

39. le Sage C, Nagel R, Egan DA, Schrier M, Mesman E, Mangiola A, Anile C, Maira G, Mercatelli N, Ciafre SA, Farace MG, Agami R: **Regulation of the p27(Kip1) tumor suppressor by miR-221 and miR-222 promotes cancer cell proliferation.** *EMBO J* 2007, **26(15)**:3699-3708.

40. Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, Verloop R, van de Wetering M, Guryev V, Takada S, van Zonneveld AJ, Mano H, Plasterk R, Cuppen E: **Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis.** *Genome Res* 2006, **16(10)**:1289-1298.

41. Seganti L, Superti F, Girmenia C, Melucci L, Orsi N: **Study of receptors for vesicular stomatitis virus in vertebrate and invertebrate cells.** *Microbiologica* 1986, **9(3)**:259-267.

42. Farnet CM, Haseltine WA: **Integration of human immunodeficiency virus type 1 DNA in vitro.** *Proc Natl Acad Sci USA* 1990, **87(11)**:4164-4168.

43. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, Gage FH, Verma IM, Trono D: **In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector.** *Science* 1996, **272(5259)**:263-267.

44. Cockrell AS, Kafri T: **Gene delivery by lentivirus vectors.** *Mol Biotechnol* 2007, **36(3)**:184-204.

45. Vara JA, Portela A, Ortin J, Jimenez A: **Expression in mammalian cells of a gene from Streptomyces alboniger conferring puromycin resistance.** *Nucleic Acids Res* 1986, **14(11)**:4617-4624.

46. Griffiths-Jones S: **The microRNA Registry.** *Nucleic Acids Res* 2004, , **32** Database: D109-111.

47. Nichols WW, Murphy DG, Cristofalo VJ, Toji LH, Greene AE, Dwight SA: **Characterization of a new human diploid cell strain, IMR-90.** *Science* 1977, **196(4285)**:60-63.

48. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, Clevers H: **Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.** *Nature* 2009, **459(7244)**:262-265.

49. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ: **RAS is regulated by the let-7 microRNA family.** *Cell* 2005, **120(5)**:635-647.

50. Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D, Wilson M, Wang X, Shelton J, Shingara J, Chin L, Brown D,

Slack FJ: **The let-7 microRNA represses cell proliferation pathways in human cells.** *Cancer Res* 2007, **67(16)**:7713-7722.

51. Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman J, Song E: **let-7 regulates self renewal and tumorigenicity of breast cancer cells.** *Cell* 2007, **131(6)**:1109-1123.

52. Cory AH, Owen TC, Barltrop JA, Cory JG: **Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture.** *Cancer Commun* 1991, **3(7)**:207-212.

53. Brideau C, Gunter B, Pikounis B, Liaw A: **Improved statistical methods for hit selection in high-throughput screening.** *J Biomol Screen* 2003, **8(6)**:634-647.

54. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: **Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex.** *Cell* 2006, **125(5)**:887-901.

55. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP: **Mammalian microRNAs: experimental evaluation of novel and previously annotated genes.** *Genes Dev* 2010, **24(10)**:992-1009.

56. Kota J, Chivukula RR, O'Donnell KA, Wentzel EA, Montgomery CL, Hwang HW, Chang TC, Vivekanandan P, Torbenson M, Clark KR, Mendell JR, Mendell JT: **Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model.** *Cell* 2009, **137(6)**:1005-1017.