Published in final edited form as:

Comb Chem High Throughput Screen. 2013 December ; 16(10): 791–805.

Modulators of the microRNA biogenesis pathway via arrayed lentiviral enabled RNAi screening for drug and biomarker discovery

David Shum, Bhavneet Bhinder, and Hakim Djaballah*

HTS Core Facility, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA

Abstract

MicroRNAs (miRNAs) are small endogenous and conserved non-coding RNA molecules that regulate gene expression. Although the first miRNA was discovered well over sixteen years ago, little is known about their biogenesis and it is only recently that we have begun to understand their scope and diversity. For this purpose, we performed an RNAi screen aimed at identifying genes involved in their biogenesis pathway with a potential use as biomarkers. Using a previously developed miRNA 21 (miR-21) EGFP-based biosensor cell based assay monitoring green fluorescence enhancements, we performed an arrayed short hairpin RNA (shRNA) screen against a lentiviral particle ready TRC1 library covering 16,039 genes in 384-well plate format, and interrogating the genome one gene at a time building a panoramic view of endogenous miRNA activity. Using the BDA method for RNAi data analysis, we nominate 497 gene candidates the knockdown of which increased the EGFP fluorescence and yielding an initial hit rate of 3.09%; of which only 22, with reported validated clones, are deemed high-confidence gene candidates. An unexpected and surprising result was that only *DROSHA* was identified as a hit out of the seven core essential miRNA biogenesis genes; suggesting that perhaps intracellular shRNA processing into the correct duplex may be cell dependent and with differential outcome. Biological classification revealed several major control junctions among them genes involved in transport and vesicular trafficking. In summary, we report on 22 high confidence gene candidate regulators of miRNA biogenesis with potential use in drug and biomarker discovery.

Keywords

miRNA; biogenesis; shRNA; H score; BDA method; RNAi; HCS; biomarker; HCA; miRNA 21; *DROSHA*; biomarker; diagnostics

INTRODUCTION

miRNAs are small endogenous, non-coding RNA molecules that regulate gene expression of specific mRNA(s) at the post-transcriptional level. In the current biogenesis model, miRNAs are first transcribed inside the nucleus by RNA polymerase II or RNA polymerase III into primary miRNA (pri-miRNA) containing a stem-loop structure and up to several hundred nucleotides (nt) in length [1]. In the next processing step, pri-miRNA undergoes cleavage by the microprocessor complex [2] composed of *DROSHA* [3, 4] and *DGCR8* [5] resulting in a

*Corresponding Author: Hakim Djaballah, PhD, HTS Core Facility, MSKCC, New York, USA; djaballh@mskcc.org; Tel: (646) 888-2198.

DISCLOSURE STATEMENT The authors declare no competing financial interests.

60–110 nt long precursor-miRNA (pre-miRNA). Exportin-5 (*XPO5*) and Ran-GTP complex export the pre-miRNA out of the nucleus where subsequent processing takes place in the cytoplasm [6]. The RNA-induced silencing complex (RISC) [7] composed of *DICER*, tar-RNA binding protein (*TRBP*), and argonaute proteins cleave the stem-loop structure yielding a 20–22 nt long mature miRNA. The complementary seed region of the miRNA binds to mRNA and *EIF2C2* [8] cleaves the target transcript resulting in translational repression or degradation.

Since its initial discovery in 1993 by Lee and co-workers [9], miRNAs are now implicated in a number of biological functions including cellular development, differentiation, proliferation, and apoptosis. According to current database searches [10], the genome encodes for over 1,600 miRNAs and expression profiling shows a diverse patterns specific to tissues and organs [11]. For example, miR-293 and miR-294 were preferentially expressed embryonic stem (ES) and not in differentiated cells suggesting a role in maintaining pluripotent state [12]. miRNAs play an essential role in normal development and not surprisingly; aberrant functioning is strongly correlated with certain diseases such as diabetes, cancer, and hypertension [13]. Abnormal miRNA expression has been observed in certain tumor types especially in breast cancers where miR-10 and miR-21 are shown to be upregulated [14]. Furthermore, miRNA profiling of tumors and various states have identified expression signatures leading to differential prognoses [15]. Consequently, studies are now driven toward understanding the key regulatory genes and pathways that modulate their biogenesis for the potential development as therapeutic targets or biomarkers for miRNA levels.

RNA interference (RNAi) technology has become a widely used technique to study and gain valuable insights into functional genomics through phenotypic perturbations. RNAi has largely evolved around two different technologies attributed to their delivery and processing inside cells: short hairpin RNA (shRNA) versus small interfering RNA (siRNA). For shRNA, gene targeted silencing occurs through concerted effort of events involving integration, expression, and processing. First, a plasmid-based system or viral vector is used to express a precursor insert of 57–58 nt in length [16]. Viral vectors such as lentiviruses mediate stable integration of the shRNA insert into the host cellular genome and subsequent transcription by RNA polymerase III leads to expression of the precursor shRNA (pre-shRNA). Eventually, the pre-shRNA are transported into the cytoplasm through *XPO5* and loaded into an RNase III complex containing *DICER* where the hairpin loop is processed off into a mature RNA duplex. Processing of the hairpin loop by *DICER* is primarily determined by the 5' end and loop region; whereby precise cleavage is critical for the functioning of shRNA in targeted silencing [17]. Finally, RISC coordinates the unwinding and loading of the guide strand along with *EIF2C2* to target multiple mRNA transcripts for cleavage or repression [18].

Due to the inherent complexity of this technology, shRNA functioning is dependent on the precise processing of *DICER* for targeting specificity and gene knockdown. Moreover, these intracellular processing events may be differential across cell lines as recently shown [19]. Sigma-Aldrich in collaboration with the Broad Institute have in part addressed this issue by providing validation data for 30% of the TRC1 library. Data for gene knockdown of individual hairpins was performed by measuring mRNA expression levels using SYBR green-based real-time quantitative PCR (RT-qPCR) in any one of six cell lines: 293T/17, A3, A549, HeLa, MCF7, and MCH58. Nevertheless, the use of shRNA technology has provided valuable insights into the roles of genes and pathways by loss-of-function analysis. A recent genome-wide shRNA screen was performed using an arrayed approach and successfully identified essential gene candidates [20].

Previously, we have described and validated our multiplexed high-content strategy to identify modulators of miR-21 in both chemical libraries and arrayed genome-wide siRNA screening [21, 22]. Following a similar approach, we adapted our miR-21 EGFP-based biosensor cell based assay for shRNA screening to identify genes and pathways that modulate miR-21 biogenesis by monitoring enhanced green fluorescence protein (EGFP) signal output while simultaneously assessing cell death through imaged Hoechst-stained nuclei (NUCL). In this study, we describe assay development and report on the execution of our arrayed genome-wide shRNA screen against the TRC1 library containing 80,598 hairpins covering 16,039 genes with an average of 5 hairpins per gene to survey EGFP activity. We applied a high-stringency hit nomination method encompassing criteria of at least 3 active duplexes per gene and filtered for potential off-target effects (OTE), referred to as the BDA method [23], and leading to the identification of 497 modulator candidates of the miR-21 biogenesis pathway; the knockdown of which resulted in enhancement of the EGFP fluorescence signal output; 22 candidates are deemed high-confidence hits with their targeting shRNA hairpins reported as validated by Sigma-Aldrich. Biological classification revealed several control junctions that are involved miRNA biogenesis including vesicular trafficking. Altogether, we present the complete results of our screen, describe how these newly identified candidate regulators may act as potential modulators of the miRNA biogenesis pathway, and discuss their potential use as biomarkers in miRNA diagnostics.

MATERIALS AND METHODS

Cell culture and materials

The reporter cell line harboring EGFP under miRNA regulation was generated as previously reported [21, 22]. In brief, HeLaS3 cells were transfected with pcDNA/TO/EGFPmiR21 (Addgene, Cambridge, MA) using Lipofectamine 2000 transfection reagent and Zeocin-resistant cells were harvested for storage of cell stocks at -170°C . Cells were grown at 37°C and 5% CO_2 in complete growth media containing DMEM, high glucose with L-glutamine, D-glucose and sodium pyruvate supplemented with 10% heat inactivated FBS, and 200 $\mu\text{g}/\text{mL}$ of Zeocin. All cell culture supplies were from Life Technologies (Carlsbad, CA), and Sigma-Aldrich (St Louis, MO).

Liquid dispensing and automation

Several liquid dispensing devices were used throughout this study. shRNA lentiviral particles were transferred using a 384 stainless steel head with disposable low-volume polypropylene tips on a PP-384-M Personal Pipettor (Apricot Designs, Monrovia, CA). The addition of cell suspensions and growth media was performed using the Multidrop 384 (Thermo Fisher Scientific, Waltham, MA). Cell fixation and staining was performed using the ELx405 automated washer (BioTek, Winooski, VT). Assay plates were incubated in the Cytomat 6001H automated incubator (Thermo Fisher Scientific) under controlled humidity at 37°C and 5% CO_2 -95% air. The assay was performed on a fully automated linear track robotic platform (Thermo Fisher Scientific) using several integrated peripherals for plate handling, cell incubators, liquid dispensing, and detection systems.

Assay development with control lentiviral particles

Cell titration studies into 384-well plate format were previously described [21]. To optimize the reporter assay for the screen, we performed a matrix format-based design to assess polybrene tolerance, puromycin killing curve, control shRNA performances, and shRNA transduction efficiency assessment. For polybrene tolerance, cell suspensions were made at 500 and 1,000 cells per well in 50 μL of media plus polybrene at doubling serial dilutions between 12 ng/mL to 25 $\mu\text{g}/\text{mL}$. At 24 h, 48 h, 72 h and 96 h post-seeding, cells were fixed in 4% paraformaldehyde (w/v) for 20 min followed by nuclei staining in a solution

containing 1 μ M Hoechst and 0.05% Triton X-100 (v/v) for 15 min. Cells were washed twice with PBS and assay plates stored at 4 °C until imaging. For the puromycin killing curve, cell suspensions were made in 50 μ L of media plus 8 μ g/mL of polybrene. At 96 h, media was aspirated and 45 μ L of media was added containing puromycin at doubling serial dilutions between 12 ng/mL to 25 μ g/mL. After an additional 120 h, cells were fixed and stained as described.

Transduction and efficiency experiments were performed with the following control lentiviral particles: a pLKO.1-puro empty vector control (SHC001V, Sigma-Aldrich), a pLKO.1-puro non-targeting shRNA control (SHC002V, Sigma-Aldrich), and a pLKO.1-puro-CMV-TurboGFP expression vector (SHC003V, Sigma-Aldrich). Cell suspensions were made in 50 μ L of media plus 8 μ g/mL of polybrene and seeded at 1,000 cells per well. At 24 h post-seeding, lentiviral particles were added to the assay plates at a multiplicity of infection (MOI) of 4 followed by an 8 min spin on a bench top centrifuge at a speed of 1,300 rpm to facilitate viral particle integration into host cells. At 72 h post-transduction, media was aspirated and 45 μ L of media containing puromycin at 1 μ g/mL was added followed by incubation for 120 h to complete selection. At 120 h post-selection, cells were fixed and stained as described.

Arrayed lentiviral particle shRNA-based library

The TRC1 is a hairpin based library containing 80,598 shRNA hairpins covering 16,039 human genes with an average of 5 hairpins per gene target [24]. Each hairpin oligonucleotide has the following sequence design: a 4 nt overhang on the 5' end with CCGG sequence, 21 nt passenger strand, a 6 nt loop with a sequence motif of CTCGAG, and followed by a 21 nt guide strand with 4 nt overhang on the 3' end with TTTT sequence cloned into the pLKO.1 vector harboring a puromycin drug resistance marker [20]. High-throughput lentiviral production of each hairpin was outsourced to Sigma-Aldrich resulting in an overall lentivirus titer yields were at least 10^6 transducing units (TU) per mL. The lentiviral particles harboring shRNA hairpins were pre-arrayed as single clones per well into 295 intermediate 384-well microtiter polypropylene plates, leaving columns 13 and 14 empty for control additions, and with a built-in puromycin selection control wells in rows O_{wells:15–24} and P_{wells:15–24} and stored at –80°C until use. Individual shRNA hairpin validation data was provided by Sigma-Aldrich.

Genome-wide shRNA screen for modulators of miR-21 biogenesis pathway

For the genome-wide screen, cell suspensions, 1,000 cells per well, were dispensed into the assay plates in 45 μ L media. After overnight incubation, media was aspirated and replaced with media plus 8 μ g/mL of polybrene. The TRC1 was thawed from storage at room temperature, followed by a 1 min spin at a speed of 800 rpm on a bench top centrifuge, after which 4 μ L was transferred into the assay plates to achieve a final MOI of 4. The assay plates were then centrifuged for 8 min at a speed of 1,300 rpm. At 72 h post-transduction, media was aspirated and 45 μ L of media containing puromycin at 1 μ g/mL was added to the cells and further incubated to complete selection. For internal reference, controls in rows A-P_{wells:13} contained non-transduced cells with no puromycin as negative control and controls in rows A-P_{wells:14} contained TurboGFP transduced cells with puromycin added as positive control. At 120 h post-selection, cells were fixed and stained as previously described [21].

Image acquisition, data analysis, and screening data management

Images were acquired on the IN Cell Analyzer 3000 (INCA3000), an automated laser confocal microscope using the following wavelengths: 364 nm excitation / 450 nm emission in the blue channel for Hoechst-stained nuclei and 488 nm excitation / 535 nm emission in the green channel for TurboGFP and EGFP signal with an exposure time of 1.5 msec. For

assay development and screening, four images per well were collected using a 40x magnifying objective covering 40% of the well and required 10 sec per well, with a total imaging time of 60 min for a complete 384-well microtiter plate. Images were analyzed using the Raven 1.0 software's built-in object intensity analysis module to assess green fluorescence intensity per well and count number of Hoechst-stained nuclei. Analysis of TurboGFP or EGFP signal and NUCL count required approximately 20 min for a complete 384-well microtiter plate. Screening data files were loaded into Oncology Research Informatics System, a custom built suite of modules for RNAi registration, plating, and data management.

Data analysis using the BDA method

For the data analysis of the genome-wide screen, we employed the step-by-step hit nomination workflow using the previously described BDA method [23]. We used the control based method for selection of active duplexes which conferred an increase in EGFP signal gain and the threshold was defined at greater than mean (μ) plus 2 standard deviations (σ) of the negative control wells. For active duplex identification for genes whose knockdown conferred a cytotoxic phenotype, the threshold was determined at a reduction of NUCL count by 80% relative to the median of the NUCL count of the shRNA hairpins in the TRC1 library. The active duplexes were assessed for clear breakpoint in EGFP signal gain and active genes were identified based on the H score of ≥ 60 or the p-value of < 0.05 . The genes that scored as active for EGFP signal gain as well as cytotoxic were classified as essential gene candidates while the ones that were active but non-cytotoxic were classified as non-essential gene candidates. This was followed by OTE filtering and re-scoring to nominate high-confidence gene candidates. Enriched gene ontology (GO) functional categories and InterPro (IPR) protein domains were found using DAVID functional annotation tool. A functional activity map was created after removing redundant GO and IPR terms for the biological functions, molecular functions, and protein domains in non-essential and essential genes. The statistical data analysis was performed using Perl and Sigmplot (SYSTAT, San Jose, CA).

RESULTS

Assay optimization for screening the shRNA-based library

Using a matrix format-based design for assay optimization, we adapted a previously described and validated cell-based system in which the HeLaS3 line contains a plasmid that encodes an EGFP fused to a sequence with perfect complementarity to miR-21 in its 3'-UTR region [21, 22]. Accordingly, endogenous miR-21 processing specifically destabilizes the EGFP mRNA resulting in low level protein expression hence low fluorescence output. Knockdown of genes involved in the miR-21 biogenesis pathway will ultimately result in reduced miR-21 levels; thereby allowing for the stabilization and expression of the EGFP mRNA resulting in high fluorescence output.

As a first step, the reporter cell line was seeded at densities of 500 and 1,000 cells per well and treated with polybrene; a cationic polymer used to increase efficiency of infection by neutralizing charge repulsion between lentiviral particles and the cell surface [25]. The effects of polybrene on NUCL count were assessed at concentrations ranging from 12 ng/mL to 25 μ g/mL for up to 96 h. Typically, polybrene remains with cells through the transduction period and is removed in place of selection media; therefore does not exceed 96 h incubation. Polybrene was very well tolerated at concentrations between 12 ng/mL to 12.5 μ g/mL and slightly cytotoxic at the highest concentration of 25 μ g/mL (Fig. 1A). Therefore, the concentration of polybrene was selected at 8 μ g/mL for maximum efficiency of infection while maintaining low cytotoxicity. Next, the reporter cell line was seeded at densities of

500 and 1,000 cells per well plus 8 $\mu\text{g/mL}$ of polybrene for 96 h; followed by treatment with puromycin at concentrations ranging from 12 ng/mL to 25 $\mu\text{g/mL}$ for 120 h to select for only those cells that have been successfully transduced. Puromycin was cytotoxic as expected allowing us to generate a kill curve on NUCL count. The calculated IC_{50} and IC_{95} values for the puromycin kill curve at 500 cells per well seeding was 0.18 $\mu\text{g/mL}$ and 0.45 $\mu\text{g/mL}$, respectively; whereas the IC_{50} values was 0.28 $\mu\text{g/mL}$ and IC_{95} values was 1 $\mu\text{g/mL}$ at 1,000 cells per well seeding (Fig. 1B). Based on our optimization efforts and analysis, the following assay conditions were used for subsequent studies: a cell seeding density of 1,000 cells per well in growth media supplemented with 8 $\mu\text{g/mL}$ polybrene and a puromycin selection concentration of 1 $\mu\text{g/mL}$ was selected.

To further validate our optimized assay parameters for shRNA-based screening, we performed transduction experiments using control lentiviral particles: empty vector, non-targeting shRNA, TurboGFP scoring for transduction on NUCL count and green fluorescence signal. The reporter cell line was transduced at a MOI of 4 for 72 h followed by puromycin selection for 120 h. Control lentiviral particles for empty vector, non-targeting shRNA, and TurboGFP exhibited puromycin resistance at a concentration of 1 $\mu\text{g/mL}$ indicating successful transduction. As expected, non-transduced cells with addition of 1 $\mu\text{g/mL}$ of puromycin resulted in cytotoxicity (Fig. 1C). Furthermore, a subset of wells contained no puromycin and we observed no significant adverse effects of lentiviral particles on the reporter cell. Using TurboGFP as a phenotypic control for transduction, we observe an increase green fluorescence signal at 6-fold greater in comparison to background with no puromycin selection translating into a signal gain of 39% and 8-fold greater with puromycin selection translating into signal gain of 71% (Fig. 1D & Fig. 1E). Taken together, the selected assay conditions for performing the shRNA screen were: a cell seeding density of 1,000 cells per well in growth media supplemented with 8 $\mu\text{g/mL}$ polybrene for 24 h, transduce the cells with the lentiviral particles at an MOI of 4 for 72 h, perform puromycin selection at a concentration of 1 $\mu\text{g/mL}$ for 120 h, followed by cell fixation and Hoechst-staining of the nuclei for automated imaging microscopy (Table 1).

Genome-wide shRNA screen for modulators of miR-21 biogenesis pathway

Using the optimized conditions and established workflow (Table 1), we have successfully executed on an arrayed genome-wide shRNA screen against the TRC1 library containing 80,598 hairpins covering 16,039 genes with approximately 5 hairpins per gene target to identify genes and pathways that modulate miR-21 biogenesis. To monitor the assay's performance throughout the screen, non-transduced cells in column 13 acted as negative control and TurboGFP lentiviral particle was added to column 14 as positive control. The controls worked as expected with box plot analysis of green signal gain showing good separation of controls from 1% to 100% (Fig. 2A). Z' value was less than 0 due to large standard deviation given the heterogeneous nature of lentiviral transduction and furthermore did not affect hit identification as the assay utilizes a gain-of-function readout. To assess the performance of the library, a frequency distribution plot of EGFP signal shows a narrow distribution reflective of a small assay window (Fig. 2B). For NUCL count, a frequency distribution plot shows a wide dynamic assay window likely due to cell growth heterogeneity (Fig. 2C).

Hit nomination using the BDA method

To identify modulators of the miR-21 biogenesis pathway, we used the BDA method encompassing a stringent and systematic approach for hit nomination. As a first step of active duplex identification for EGFP signal readout, individual shRNA hairpins were scored using a threshold based on 2σ from the mean of the negative control that translates to an EGFP signal gain of 25% and 7,114 duplexes were identified as active (Fig. 2D & Suppl

Table 1). In parallel for active duplex identification of NUCL count readout, individual shRNA hairpins were scored using a median based method at 80% and 2,225 active duplexes were cytotoxic (Fig. 2E). As the second step, we set an H score of 60 to nominate 523 active genes as modulators of miR-21 biogenesis and 110 active genes as cytotoxic. To nominate the final list of genes involved in miR-21 biogenesis, the remaining active sequences were subjected to OTE filtering. For the purpose of seed heptamer selection, we used the guide strand starting at nt position 32 based on the ideal location on the oligonucleotide as well as nt position 36 empirically found using the ESS method [23]. After merging the OTE results from the two seed selection methods, 51 active duplexes qualified as high confidence-OTE and were removed from subsequent analysis to nominated 497 gene candidates in total; out of which 481 were scored as non-essential and 16 as essential gene candidates for their potential roles as modulators of the miRNA biogenesis pathway (Fig. 3 & Suppl Table 2). Furthermore, we consider 22 of them as high-confidence candidates (Table 2) since each gene had an H score of 100 with at least three active duplexes being independently validated by Sigma-Aldrich in one of six different cell lines (Fig. 4).

497 nominated gene candidates versus known modulators of miRNA biogenesis machinery

We surveyed the miRNA biogenesis literature and collected approximately 92 genes with a known or putative regulatory role in miRNA biogenesis; out of which seven genes constituted the core set of essential to the biogenesis machinery, *DGCR8*, *DICER1*, *DROSHA*, *EIF2C2* (*AGO2*), *PRKRA* (*PACT*) [26], *TARBP2*, and *XPO5*. As a first step, we studied the degree of overlap between the known modulators and the 497 gene candidate nominated in the genome-wide screen reported here. The overlap analysis revealed surprising results as only a single core biogenesis gene *DROSHA* was identified (Table 3). Although *DGCR8*, *DICER1* and *EIF2C2*, the three key players in the biogenesis machinery had utmost 2 hairpins scored as active but failed to qualify at the stringent gene activity threshold determined by H score of 60. For the remaining two core biogenesis genes, *PRKRA* and *TARBP2*, none of the hairpins scored active above the qualifying threshold for gain in EGFP signal readout. Of note, the nuclear transporter *XPO5* was not present in the TRC1 library. Besides *DROSHA*, two other regulatory genes were identified *ADAR* [27] and *hRNPR* [28] which is involved in miRNA editing and processing, respectively.

Having identified only three genes with previously implicated roles in miRNA biogenesis, we next studied protein-protein interactions of the 497 nominated genes in this screen with all known modulators (Fig. 5A). In this analysis, we identified 52 candidates to interact with known biogenesis regulators such as *TGFBRI* [29], *MAPK1* [30], *CDK1* [31] and *RAFI* [32]. Consistent with our overlap analysis, we found sparse interactions of nominated gene candidates with core biogenesis genes. None of our gene candidates were found to have interactions with *DGCR8* and *DROSHA*. Only one gene, *UPF1* [33] a component of exon junction complex was found to interact with core biogenesis components *DICER1* and *EIF2C2*.

Biological classification of 497 nominated gene candidates

The 497 nominated gene candidates were subjected to a gene clustering analysis based on the similarity in their biological functions or domains and found 12 distinct gene groups (Table 4). Of prominence were four gene groups associated with protein kinases involved in adenylyl ribonucleotide binding, small GTPase that mediated signaling, glycosylation via Golgi apparatus and G-protein coupled receptors. In the next step of the analysis, we looked at the biological classification and identified 17 distinct categories of prominence in metabolic processes (285 genes in GO:0008152), cellular processes (200 genes in GO:

0009987), cell communication (155 genes in GO:0007154), development (88 genes in GO:0032502), and transporters (81 genes in GO:0006810) (Fig. 5B). Amongst the metabolic processes, we identified components involved in metabolism of carbohydrates (12 genes), proteins (112 genes), hormones (10 genes), and carboxylic acid (28 genes). In addition, we found 11 genes each associated with cellular localization and protein complex assembly. A detailed breakup of the 497 nominated gene candidates is provided (Suppl Table 3). An evaluation of the protein-protein interactions among the 497 nominated gene candidates revealed top five gene clusters within the master network: Cluster 1 comprised of nine genes associated with eukaryotic translation process, Cluster 2 comprised of six genes involved in mRNA processing, Cluster 3 comprised of five genes involved in GTPase signaling, Cluster 4 comprised of five genes involved in mitochondrial electron transport, Cluster 5 comprised of seven genes involved in signal transduction, carbohydrate metabolism, and actin cytoskeleton (Fig. 5C).

Furthermore, we grouped the nominated gene candidates into specific protein classes using the PANTHER classification system and were able to identify 26 distinct protein categories. Prominent protein classes (PC) include transferases (94 genes in PC00220), followed by 76 nucleic acid binding proteins (76 genes in PC00171), hydrolases (72 genes in PC00121), 57 receptors (57 genes in PC00197), and enzyme modulators (52 genes in PC00095) (Fig. 5D). Moreover, we analyzed the canonical pathway associations using Metacore software and the top 15 pathways were grouped into six broad categories including transcription (2 pathways), immune response (1 pathway), G-protein coupled signaling (3 pathways), development (5 pathways), cytoskeleton remodeling (3 pathways), and the cell cycle (1 pathway) (Fig. 5E).

DISCUSSION

miRNAs modulate gene expression at the post-transcriptional level and influence a large number of biological functions including development, differentiation, proliferation, and apoptosis. Expression profiling has revealed tissue specific miRNA patterns in both normal and disease states including cancer. Not surprisingly, miRNA patterns in tumors have identified differential expression signatures that can lead to a variety of prognoses and provide an excellent rationale in understanding their regulation to develop as biomarkers for miRNA levels. For this purpose, we performed the first arrayed genome-wide shRNA screen against the lentiviral particle ready TRC1 library aimed at identifying genes that modulate the miR-21 biogenesis pathway. Previously, we have developed the EGFP-based biosensor [21] as a surrogate for miR-21 activity and validated this assay strategy in an arrayed genome-wide siRNA screen using the Ambion Silencer Select V4.0 library [22]. From the TRC1 screen of 80,598 hairpins covering 16,039 genes, we identified 7,114 duplexes as active based on EGFP signal threshold and also scored 2,225 active duplexes as cytotoxic with NUCL count readout. We implemented the high-stringent BDA method to nominate 481 non-essential and 16 essential gene candidates as modulators of miRNA biogenesis of which 22 are of high-confidence with an H score of 100 and at least three of their targeting shRNA has been deemed validated (Table 2).

Aberrant miRNA expression has been strongly correlated to certain diseases and key signatures are different among pancreatic, leukemic, and lung cancers. For instance, abnormally high levels of miR-155 and low let-7a-2 were found in lung adenocarcinoma suggesting a substantial role in carcinogenesis [34]. In cystic fibrosis patients, elevated levels of miR-145 and miR-494 in nasal epithelial tissues indicating a role in pathogenesis [35]. As such, recognizing the key regulators that modulate their biogenesis and hence their expression may be exploited as targets for miRNA dysregulation or predicting miRNA levels. For example, a study identified *AKT2* dependent pathway in which hypoxia promotes

tumor resistance through induction of miR-21 [36]. Moreover, a genomic analysis in a drug resistant MCF7 cell line found that *AKT3* correlates inversely with miR-505 expression and modulates drug sensitivity [37]. As such, we hope that our RNAi-based approach will elucidate the underlying mechanisms of pathology for aberrant miRNA expression and provide better therapeutic drugs for diseases.

To develop biomarkers for miRNA levels, a biological classification of our nominated genes revealed several major functional groups that potentially control the miR-21 biogenesis pathway: transport via Golgi apparatus, vesicular trafficking, cytoskeletal remodeling, and TGF β /BMP signaling. Studies have suggested the involvement of the Golgi apparatus in premiRNA processing or transfer of miRNA onto EIF2C2 [38]. Our results show an enrichment of genes involved in glycosylation within the Golgi network including *B4GALNT4*, *CHST5*, *FUT11*, *FUT6*, *GALNT10*, *GALNT3*, *GALNT6*, *GALNT7*, *GALNTL4*, *ST3GAL3*, and *ST6GAL2*. In particular, *GALNT7* has been reported to play a role melanoma metastasis with upregulated miR-30b/30d clustering [39]. For vesicular trafficking, we identified 11 Rab genes that belong to the Ras super family including *RAB3B*, *RAB3C*, *RAB6C*, *RAB13*, *RAB17*, *RAB21* [40], *RAB23*, *RAB28* [41] of which known biological functions include intracellular targeting/transport and therefore plausible that these genes participate in miRNA localization. Additionally, our results show a cluster of five genes of the Rho family that may be involved in mobilization of miRNA through cytoskeletal remodeling and actin dynamics [42]. Although, we did not identify any enrichment of genes directly involved in the TGF β /BMP pathway that is known to play a role in miRNA biogenesis; however, we found TGF β -dependent induction of epithelial to mesenchymal transition (EMT) that has quite recently been shown to play a regulatory role on miRNA expression [43]. Of the gene candidates mentioned *GALNT7*, *RAB21*, and *RAB28* may be starting points for development as biomarkers for miRNA levels as were previously shown to interact with the biogenesis pathway.

To identify additional genes as potential biomarkers, we performed a KEGG's database search to determine if any of our candidate genes were previously associated with diseases. A KEGG's database search of human diseases identified a total of 26 genes related to different tumor types including melanoma, pancreatic cancer, colorectal cancer, and leukemia [44]. Our nominated candidates included the following 22 genes *ABL1*, *AKT3*, *CDK6*, *E2F1*, *FGF13*, *FGF16*, *FGF2*, *FGF3*, *FGFR1*, *FOS*, *GLI2*, *IL6*, *ITGB1*, *MAPK1*, *PIK3R1*, *RALGDS*, *RXRA*, *SHH*, *SOS2*, *TGFB2*, *TGFBR1*, and *WNT3*; with 4 high-confidence gene candidates *GSTP1*, *JAK1*, *PTK2*, and *RAF1*. Amongst these genes *AKT3*, *JAK1*, and *WNT3* are promising candidates for development as biomarkers for miRNA levels. As mentioned earlier, *AKT3* correlated with miR-505 expression in a drug-resistant MCF7 cell line [37]. *JAK1* was a previously suggested biomarker for drug sensitivity in lung cancer [45] and *WNT3* plays an important role in miRNA signaling for development in human osteoblasts [46]. Altogether, our nominated gene candidates suggest a plausible link between miRNA biogenesis and subsequent aberrant expression whose dysregulation may potentially be useful as predictive indicators for miRNA levels in diseases.

Lastly, among the core essential genes involved in miRNA biogenesis we only identified *DROSHA* as a hit out of the TRC1; *XPO5* was not covered in the TRC1 library. Furthermore, only two genes *ADAR* and *hmrNPR* out of the known miRNA regulators were identified; however, a network level analysis further revealed 54 gene candidates to interact with the several known core and regulators of the miRNA machinery. The lack of activity amongst known core genes and their regulators in the screen supports our previous notion that shRNA technology may not be as robust as initially perceived since intracellular processing remains a major determinant of target specificity and knockdown [20]. Similarly, Boettcher and co-workers recently observed that shRNA produced differential knockdowns

in OVCAR8 and MCF7 revealing that cell line type plays a role in knockdown efficiencies of hairpins [19] offering a plausible explanation for our observations. Of the 7,114 active hairpins identified in the screen, 2,304 were deemed validated by Sigma-Aldrich among six different cell lines 293T/17, A3, A549, HeLa, MCF7, and MCH58 (Fig. 4). A majority of the validated hairpins was performed in the A549 line as opposed to our screen which was completed in HeLaS3 line and must be considered during the confirmation process.

In summary, we have reported on the successful execution of the first arrayed genome-scale shRNA screen to identify modulators of miRNA biogenesis and nominated 481 non-essential and 16 essential gene candidates. Our gene analysis by means of biological classification reveals several major control junctions including Golgi transport, vesicular trafficking, and cytoskeletal remodeling. Finally, we provide a complete list of the 497 gene candidates with 22 deemed high-confidence. Moreover, we nominate six candidate genes, *AKT3*, *GALNT7*, *JAK1*, *RAB21*, *RAB28*, and *WNT3*, as potential biomarkers for intracellular miRNAs having previously been associated with biogenesis and diseases. We anticipate these newly identified genes will provide additional insights into the miRNA biogenesis pathway leading to the development of new therapeutics and biomarkers for miRNA diagnostics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank members of the HTS Core Facility especially Christina N. Ramirez, Constantin Radu, and Paul Calder for their help during the course of this study. The HTS Core Facility is partially supported by Mr. William H. Goodwin and Mrs. Alice Goodwin and the Commonwealth Foundation for Cancer Research, the Experimental Therapeutics Center of MSKCC, the William Randolph Hearst Fund in Experimental Therapeutics, the Lillian S Wells Foundation and by an NIH/NCI Cancer Center Support Grant 5 P30 CA008748-44.

ABBREVIATIONS

RNAi	RNA interference
shRNA	short hairpin interfering RNA
BDA	Bhinder-Djaballah Analysis method
H score	hit rate per gene score
miRNA	microRNA
INCA3000	IN Cell Analyzer 3000

REFERENCES

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116(2):281–297. [PubMed: 14744438]
2. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. The Microprocessor complex mediates the genesis of microRNAs. *Nature*. 2004; 432(7014):235–240. [PubMed: 15531877]
3. Lee Y, Ahn C, Han J, Choi H, Kim 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. [PubMed: 14508493]
4. Lee Y, Han J, Yeom KH, Jin H, Kim VN. Drosha in primary microRNA processing. *Cold Spring Harb Symp Quant Biol*. 2006; 71:51–57. [PubMed: 17381280]

5. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* 2004; 18(24):3016–3027. [PubMed: 15574589]
6. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* 2003; 17(24):3011–3016. [PubMed: 14681208]
7. 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. [PubMed: 15973356]
8. Okamura K, Ishizuka A, Siomi H, Siomi MC. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 2004; 18(14):1655–1666. [PubMed: 15231716]
9. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell.* 1993; 75(5):843–854. [PubMed: 8252621]
10. [(Accessed March 28, 2013)] miRBase: the microRNA database. <http://mirbase.org>
11. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* 2004; 5(3):R13. [PubMed: 15003116]
12. Houbaviy HB, Murray MF, Sharp PA. Embryonic stem cell-specific microRNAs. *Dev Cell.* 2003; 5(2):351–358. [PubMed: 12919684]
13. Kloosterman WP, Plasterk RH. The diverse functions of microRNAs in animal development and disease. *Dev Cell.* 2006; 11(4):441–451. [PubMed: 17011485]
14. Singh R, Mo YY. Role of microRNAs in breast cancer. *Cancer Biol Ther.* 2013; 14(3):201–212. [PubMed: 23291983]
15. Iorio MV, Croce CM. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. *EMBO Mol Med.* 2012; 4(3):143–159. [PubMed: 22351564]
16. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat Methods.* 2006; 3(9):715–719. [PubMed: 16929317]
17. Maczuga P, Lubelski J, van Logtenstein R, Borel F, Blits B, Fakkert E, Costessi A, Butler D, van Deventer S, Petry H, Koornneef A, Konstantinova P. Embedding siRNA sequences targeting apolipoprotein B100 in shRNA and miRNA scaffolds results in differential processing and in vivo efficacy. *Mol Ther.* 2013; 21(1):217–227. [PubMed: 23089734]
18. Manjunath N, Wu H, Subramanya S, Shankar P. Lentiviral delivery of short hairpin RNAs. *Adv Drug Deliv Rev.* 2009; 61(9):732–745. [PubMed: 19341774]
19. Boettcher M, Hoheisel JD. Pooled RNAi screens - technical and biological aspects. *Curr Genomics.* 2010; 11(3):162–167. [PubMed: 21037854]
20. Bhinder B, Antczak C, Ramirez CN, Shum D, Liu-Sullivan N, Radu C, Frattini MG, Djaballah H. An arrayed genome-scale lentiviral-enabled short hairpin RNA screen identifies lethal and rescuer gene candidates. *Assay Drug Dev Technol.* 2013; 11(3):173–190. [PubMed: 23198867]
21. Shum D, Bhinder B, Radu C, Calder P, Ramirez CN, Djaballah H. An image-based biosensor assay strategy to screen for modulators of the microRNA 21 biogenesis pathway. *Comb Chem High Throughput Screen.* 2012; 15(7):529–541. [PubMed: 22540737]
22. Shum D, Bhinder B, Ramirez CN, Radu C, Calder PA, Beauchamp L, Farazi T, Landthaler M, Tuschi T, Magdaleno S, Djaballah H. An arrayed RNA interference genome-wide screen identifies candidate genes involved in the microRNA 21 biogenesis pathway. *Assay Drug Dev Technol.* 2013; 11(3):191–205. [PubMed: 23153064]
23. Bhinder B, Djaballah H. A simple method for analyzing actives in random RNAi screens: Introducing the “H Score” for hit nomination and gene prioritization. *Comb Chem High Throughput Screen.* 2012; 15(9):686–704. [PubMed: 22934950]
24. [(Accessed April 17, 2013)] Sigma-Aldrich: MISSION® Whole-Genome shRNA Library. <http://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/shrna/library-information/current-library.html>
25. Davis HE, Rosinski M, Morgan JR, Yarmush ML. Charged polymers modulate retrovirus transduction via membrane charged neutralization and virus aggregation. *Biophys J.* 2004; 86(2):1234–1242. [PubMed: 14747357]

26. Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN. The role of PACT in the RNA silencing pathway. *EMBO J.* 2006; 25(3):522–532. [PubMed: 16424907]
27. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol.* 2006; 13(1):13–21. [PubMed: 16369484]
28. Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem.* 1993; 62:289–321. [PubMed: 8352591]
29. Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, Calin GA, Rossi S, Fernandez AF, Carneiro F, Oliveira C, Ferreira B, Liu CG, Villanueva A, Capella G, Schwartz S Jr, Shiekhattar R, Esteller M. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet.* 2009; 41(3):365–370. [PubMed: 19219043]
30. Paroo Z, Ye X, Chen S, Liu Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell.* 2009; 139(1):112–122. [PubMed: 19804757]
31. Chivukula RR, Mendell JT. Circular reasoning: microRNAs and cell-cycle control. *Trends Biochem Sci.* 2008; 33(10):474–481. [PubMed: 18774719]
32. Zhang X, Mao H, Chen JY, Wen S, Li D, Ye M, Lv Z. Increased expression of microRNA-221 inhibits PAK1 in endothelial progenitor cells and impairs its function via c-Raf/MEK/ERK pathway. *Biochem Biophys Res Commun.* 2013; 431(3):404–408. [PubMed: 23333386]
33. Jin H, Suh MR, Han J, Yoem KH, Lee Y, Heo I, Ha M, Hyun S, Kim VN. Human UPF1 participates in small RNA-induced mRNA downregulation. *Mol Cell Biol.* 2009; 29(21):5789–5799. [PubMed: 19704008]
34. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, Harris CC. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell.* 2006; 9(3):189–198. [PubMed: 16530703]
35. Megiorni F, Cialfi S, Cimino G, De Biase RV, Dominici C, Quattrucci S, Pizzuti A. Elevated levels of miR-145 correlate with SMAD3 down-regulation in Cystic Fibrosis patients. *J Cyst Fibros.* doi:pii: S1569-1993(13)00043-X. 10.1016/j.jcf.2013.03.007.
36. Polytarchou C, Iliopoulos D, Hatziaepostolou M, Kottakis F, Maroulakou I, Struhl K, Tschlis PN. Akt2 regulates Akt isoforms and promotes resistance to hypoxia through induction of miR-21 upon oxygen deprivation. *Cancer Res.* 2011; 71(13):4720–4731. [PubMed: 21555366]
37. Yamamoto Y, Yoshioka Y, Minoura K, Takahashi RU, Takeshita F, Taya T, Horii R, Fukuoka Y, Kato T, Kosaka N, Ochiya T. An integrative genomic analysis revealed the relevance of microRNA and gene expression for drug-resistance in human breast cancer cells. *Mol Cancer.* 2011; 10:135. [PubMed: 22051041]
38. Gibbins D, Voinnet O. Control of RNA silencing and localization by endolysosomes. *Trends Cell Biol.* 2010; 20(8):491–501. [PubMed: 20630759]
39. Gaziel-Sovran A, Hernando E. miRNA-mediated GALNT modulation of invasion and immune suppression: A sweet deal for metastatic cells. *Oncoimmunology.* 2012; 1(5):746–748. [PubMed: 22934269]
40. Pellinen T, Arjonen A, Vouriluoto K, Kallio K, Fransén JA, Ivaska J. Small GTPase Rab21 regulates cell adhesion and controls endosomal traffic of beta1-integrins. *J Cell Biol.* 2006; 173(5):767–780. [PubMed: 16754960]
41. Lee SH, Baek K, Dominguez R. Large nucleotide-dependent conformational change in Rab28. *FEBS Lett.* 2008; 582(29):4107–4111. [PubMed: 19026641]
42. Ridley AJ. Rho GTPases and actin dynamics in membrane protrusions and vesicle trafficking. *Trends Cell Biol.* 2006; 16(10):522–529. [PubMed: 16949823]
43. Cai ZG, Zhang SM, Zhang H, Zhou YY, Wu HB, Xu XP. Aberrant expression of microRNAs involved in epithelial-mesenchymal transition of HT-29 cell line. *Cell Biol Int.* 2013; 37(7):669–674. [PubMed: 23483606]
44. [(Accessed April 17, 2013)] KEGG PATHWAY Database. <http://www.genome.jp/kegg/pathway/hsa/hsa05200.html>
45. Shimokawa T, Seike M, Soeno C, Uesaka H, Miyanaga A, Mizutani H, Kitamura K, Minegishi Y, Noro R, Okano T, Yoshimura A, Gemma A. Enzastaurin has anti-tumour effects in lung cancers

- with overexpressed Jak pathway molecules. *Br J Cancer*. 2012; 106(5):867–875. [PubMed: 22333600]
46. Kapinas K, Kessler C, Ricks T, Gronowicz G, Delany AM. miR-29 modulates Wnt signaling in human osteoblasts through a positive feedback loop. *J Biol Chem*. 2010; 285(33):25221–25231. [PubMed: 20551325]

Figure 1.

Assay optimization for shRNA screening

Assay was optimized for shRNA screening using a matrix format-based design. **A)** Assessment of cell tolerance to polybrene of reporter cells at two seeding densities of 500 and 1,000 cells per well in dose response studies for 96 h with NUCL count readout. **B)** Puromycin kill curve following treatment of reporter cells at two seeding densities of 500 and 1,000 cells per well in 8 $\mu\text{g/mL}$ of polybrene for 96 h followed by dose response studies in puromycin for 120 h with NUCL count readout. **C)** Transduction of control lentiviral particles of cells at a seeding density of 1,000 cells per well in 8 $\mu\text{g/mL}$ of polybrene for 72 h followed by puromycin selection of 1 $\mu\text{g/mL}$ for 120 h with NUCL count readout, **D)** TurboGFP signal, and **E)** images acquired using the IN Cell Analyzer 3000 (INCA3000) with green channel for TurboGFP signal and blue channel for Hoechst-stained nuclei.

Figure 2.

Assessment of assay performance during shRNA screening

Cells (1,000 cells per well) were screened against TRC1 at a MOI of 4 covering 16,039 genes. **A)** Box plot and analysis of controls during screening. **B)** Distribution plot of individual shRNA hairpin activity during screening for EGFP signal. **C)** Distribution plot of individual shRNA hairpin activity during screening for NUCL count. **D)** Plot of active shRNA hairpins per gene for EGFP signal. **E)** Plot of active shRNA hairpins per gene for NUCL count.

Figure 3.

BDA analysis method workflow

High-stringency analysis nominates 481 non-essential and 16 essential candidate modulators of miRNA biogenesis. HC_OTE; High confidence off-target effects, LC_OTE; Low confidence off-target effects, No_OTE; No off-target effects.

Figure 4.

Validation data for the shRNA hairpins scored as active in the first step of active duplex identification by the BDA method.

Validation data was provided by Sigma-Aldrich.

A)

B)

Figure 5.

Biological classification of 497 nominated gene candidates

A) Network mapping using Ingenuity Pathway Analysis shows protein-protein interactions among the nominated gene candidates (blue) and known core components (bold red) or regulators of miRNA biogenesis (grey). **B)** Categorization of nominated gene candidates into functional classes based on the enrichment of Gene Ontology (GO) terms. The number of genes participating in each category were obtained from PANTHER classification system. **C)** Five gene clusters found within nominated candidates generated using the MiMI and MCODE analysis. **D)** Classification of nominated gene candidate into protein classes (PC) and enrichment determined using PANTHER classification system. **E)** Prominent pathways among nominated gene candidates determined at a p-value of < 0.05 and false discovery rate of 10% using the Metacore pathway analysis software.

Table 1

Assay workflow using optimized conditions for the execution of the genome-scale shRNA screen in the reporter cell line

Step	Parameter	Value	Description
1	Cell plating	45 μ L	1,000 cells in growth media
2	Incubation time	24 h	37°C, 5% CO ₂
3	Polybrene addition	45 μ L	Growth media containing 8 μ g/mL polybrene
4	Transduction – arrayed format	4 μ L	Addition of shRNA lentiviral particles; 4,000 viral particles
5	Assay plate spin	8 min	Centrifuged at 1,300 rpm
6	Incubation time	72 h	37°C, 5% CO ₂
7	Puromycin selection	45 μ L	Growth media containing 1 μ g/mL puromycin
8	Incubation time	120 h	37°C, 5% CO ₂
9	Fix	45 μ L	4% PFA (w/v) in 1X PBS for 20 min
10	Nuclear staining	45 μ L	1 μ M Hoechst solution with 0.05% Triton X-100 (v/v) in 1X PBS
11	Assay readout	364nm/450nm & 488nm/535nm (ex/em)	INCA 3000 automated microscope
12	Image analysis	--	Multiparametric analysis using Raven 1.0 software

Step	Notes
1	Dispensing into assay plate with Multidrop 384; 30 sec per 384-well microtiter plate
2	Assay plates stored in the Cytomat 6001 H; an automated incubator
3	Aspirating on the ELx405 automated washer and dispensing with Multidrop 384; 1 min per 384-well microtiter plate
4	Dispensing on the PP-384-M Personal Pipettor using a custom 384 head; 30 sec per 384-well microtiter plate
5	Spin assay plates on bench top centrifuge
6	Assay plates stored in the Cytomat 6001 H; an automated incubator
7	Aspirating on the ELx405 automated washer and dispensing with Multidrop 384; 1 min per 384-well microtiter plate
8	Assay plates stored in the Cytomat 6001 H; an automated incubator
9-10	Aspirating on the ELx405 automated washer and dispensing with Multidrop 384; 1 min per 384-well microtiter plate
11	For assay development and screening, four images per well for 10 sec per well with total imaging time of 60 min per 384-well microtiter plate
12	Analysis of EGFP signal output and Hoechst-stained nuclei, 20 min per 384-well microtiter plate

Table 2

List of 22 high-confidence gene candidates

Gene Symbol	RefSeq ID	Gene Description	Active duplexes	Total duplexes	H score	Validated duplexes	Category
ACADL	NM_001608	Acyl-coenzyme A dehydrogenase, long-chain	5	5	100	4	Non essential
ALDH1A2	NM_170697	Aldehyde dehydrogenase 1 family, member A2	5	5	100	5	Non essential
AOI1	XM_173173	Amine oxidase, flavin containing 1	5	5	100	3	Non essential
ARIH2	NM_006321	Ariadne homolog 2 (Drosophila)	5	5	100	4	Non essential
DIS3	NM_014953	DIS3 mitotic control homolog	5	5	100	4	Non essential
DYRK3	NM_003582	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 3	10	10	100	6	Non essential
FAM134C	NM_178126	Family with sequence similarity 134, member C	5	5	100	3	Non essential
GATM	NM_001482	Glycine amidinotransferase	5	5	100	5	Non essential
GSTP1	NM_000852	Glutathione S-transferase, pi 1	5	5	100	5	Non essential
GTF2A2	NM_004492	General transcription factor IIA, 2, 12kDa	4	4	100	4	Non essential
JAK1	NM_002227	Janus kinase 1	19	19	100	18	Non essential
KPNB1	NM_002265	Karyopherin (importin) beta 1	5	5	100	5	Essential
MAN2A1	NM_002372	Mannosidase 2, alpha 1	5	5	100	4	Non essential
MEITL3	NM_019852	Methyltransferase like 3	4	4	100	3	Non essential
OXAIL	NM_005015	Oxidase assembly 1-like	5	5	100	4	Non essential
PLK4	NM_014264	Polo-like kinase 4	19	19	100	14	Non essential
PTK2	NM_005607	PTK2 protein tyrosine kinase 2	23	23	100	20	Non essential
RAF1	NM_002880	v-Raf-1 murine leukemia viral oncogene homolog 1	9	9	100	7	Non essential
RPS27A	NM_002954	Ribosomal protein S27a	5	5	100	3	Essential
SFRS2	NM_003016	Splicing factor, arginine/serine-rich 2	32	32	100	3	Non essential
TIE1	NM_005424	Tyrosine kinase receptor 1	19	19	100	3	Non essential
TRAPPC3	NM_014408	Trafficking protein particle complex 3	5	5	100	5	Non essential

Performance of shRNA hairpins from screen against the seven core components and eight representative regulator genes of miRNA biogenesis machinery

Table 3

miRNA biogenesis machinery	Genes	RefSeq ID	Activity (EGFP signal)	H score	BDA
Core components					
Microprocessor complex	DGCR8	NM_022720	2/4 shRNA	50	Inactive
	DROSHA (RNASEN)	NM_013235	3/5 shRNA	60	Active
	EIF2C2 (AGO2)	NM_012154	2/5 shRNA	40	Inactive
RISC loading complex	DICER1	NM_030621	1/5 shRNA	20	Inactive
	PRKRA (PACT)	NM_003690	0/11 shRNA	0	Inactive
	TARBP2	NM_004178	0/4 shRNA	0	Inactive
Nuclear export	XPO5	NM_020750	Not present in TRC1	NA	NA
Regulators[†]					
RISC formation & activity	TNRC6B	NM_015088	1/5 shRNA	20	Inactive
	DHX9 (RHA)	NM_001357	2/5 shRNA	40	Inactive
	MOV10	NM_020963	0/5 shRNA	0	Inactive
miRNA processing	HDAC1	NM_004964	0/10 shRNA	0	Inactive
	hnRNPR	NM_005826	4/5 shRNA	80	Inactive
	ADAR	NM_001111	4/5 shRNA	80	Inactive
miRNA stability	NHLH1 (HEN1)	NM_005598	1/5 shRNA	20	Inactive
	PDCD6IP (ALIX)	NM_013374	2/5 shRNA	40	Inactive

[†] Few examples shown

Associated functions of 12 gene groups identified using DAVID functional annotation tool

Table 4

Gene group	Group function	# of genes	Nominated genes in group
1	Protein kinases (adenyl ribonucleotide binding)	18	AKT3, CDKL3, CSNK1D, DYRK3, MAP2K4, MAP2K5, MAP4K2, MARK3, NEK3, PLK3, PLK4, SRPK1, STK32C, TAOK3, TSSK4, TXK, VRK1, YES1
2	Small GTPase mediated signaling	18	ARF4, ARF5, ARL4C, ARL5B, RAB13, RAB17, RAB21, RAB23, RAB28, RAB30, RAB38, RAB39B, RAB3B, RAB3C, RAB6C, RAPIA, RHOC, RHOG
3	GTPase regulatory activity	4	KNDC1, RALGDS, RAPGEF1, RGL1
4	Transmembrane protein kinase receptor	4	EPHB1, EGFR1, MSTIR, TIE1
5	Golgi apparatus, glycosylation	11	B4GALNT4, CHST5, FUT11, FUT6, GALNT10, GALNT3, GALNT6, GALNT7, GALNTL4, ST3GAL3, ST6GAL2
6	Ribosomal protein	4	RPS13, RPS3A, RPS5, RPS7
7	Electron transport chain	4	NDUFA11, NDUFA2, NDUFAF1, NDUFB7
8	Transcription regulation	5	GTF2A2, HAND2, MAML3, TAF8, TEAD3
9	Immunoglobulin-like	5	BTNL9, CADM3, FGFR1, NCR2, SIRPB1
10	G-protein coupled receptors	13	FFAR2, GPR111, GPR112, GPR148, GPR151, GPR62, MAS1, OR1A1, OR52E8, OR52M1, OR5T1, OR8B8, TAS2R5
11	Ion gated channel	4	KCNG3, KCNH2, SCN9A, TRPM2
12	Zinc finger	8	MKRN2, RCHY1, RFWD3, RNFT1, ZNF333, ZNF436, ZNF549, ZNF768