Published in final edited form as:

Comb Chem High Throughput Screen. 2014 February ; 17(2): 98–113.

Comparative analysis of RNAi screening technologies at genome-scale reveals an inherent processing inefficiency of the plasmid-based shRNA hairpin

Bhavneet Bhinder, David Shum, and Hakim Djaballah*

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

Abstract

RNAi screening in combination with the genome-sequencing projects would constitute the Holy Grail of modern genetics; enabling discovery and validation towards a better understanding of fundamental biology leading to novel targets to combat disease. Hit discordance at inter-screen level together with the lack of reproducibility is emerging as the technology's main pitfalls. To examine some of the underlining factors leading to such discrepancies, we reasoned that perhaps there is an inherent difference in knockdown efficiency of the various RNAi technologies. For this purpose, we utilized the two most popular ones, chemically synthesized siRNA duplex and plasmid-based shRNA hairpin, in order to perform a head to head comparison. Using a previously developed gain-of-function assay probing modulators of the miRNA biogenesis pathway, we first executed on a siRNA screen against the Silencer Select V4.0 library (AMB) nominating 1,273, followed by an shRNA screen against the TRC1 library (TRC1) nominating 497 gene candidates. We observed a poor overlap of only 29 hits given that there are 15,068 overlapping genes between the two libraries; with *DROSHA* as the only common hit out of the seven known core miRNA biogenesis genes. Distinct genes interacting with the same biogenesis regulators were observed in both screens, with a dismal cross-network overlap of only 3 genes (*DROSHA*, *TGFBRI*, and *DIS3*). Taken together, our study demonstrates differential knockdown activities between the two technologies, possibly due to the inefficient intracellular processing and potential cell-type specificity determinants in generating intended targeting sequences for the plasmid-based shRNA hairpins; and suggests this observed inefficiency as potential culprit in addressing the lack of reproducibility.

INTRODUCTION

RNAi technology has presented researchers with an opportunity to study and gain valuable insights into functional genomics through phenotypic perturbations since its initial discovery in 1998 by Fire and Mello in *Caenorhabditis elegans* [1]. Shortly afterwards, similar mechanisms of sequence-dependent cleavage of mRNA were discovered in plants and

*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.

mammalian cells [2]. A decade later, RNAi has become a widely used technique for target discovery, validation, and therapeutic development; and as a screening platform, it has enabled scientists to perform large-scale screens in areas ranging from cancer biology to host-pathogen interactions [3]. RNAi screening has evolved around two different technologies largely attributed to their delivery and processing inside cells. Small interfering RNA (siRNA) are chemically synthesized duplexes comprised of guide and passenger strands typically 20-25 nucleotides (nt) in length with overhangs on their 5' and 3' ends [4]. For gene silencing to occur, the duplexes must be delivered inside cells and a catalog of transfection reagents, ranging from liposomes to nanoparticles, are commercially available for this purpose. Once inside cells, an ATP-dependent helicase unwinds the siRNA duplexes and the guide strand is incorporated into a large multiprotein complex known as the RNA-induced silencing complex (RISC) while the passenger strand is released [5]. The complementary seed region of the guide strand binds and directs cleavage of the corresponding mRNA sequence resulting in translational repression or degradation [6].

Small hairpin RNA (shRNA), on the other hand, employs a series of steps involving delivery into host cell, integration into its genome, its transcription, and processing to ultimately achieve gene-silencing. shRNA utilizes a plasmid-based system to express a precursor insert of 57-58 nt in length [7]. Expression of the shRNA constructs can be achieved through plasmid delivery or through viral vectors including lentiviral delivery particles. Lentiviruses mediate integration of the shRNA insert into the host cellular genome and transcription usually occurs through RNA polymerase III in the nucleus; leading to precursor shRNA or pre-shRNA. The pre-shRNA are eventually transported into the cytoplasm and loaded into an RNase III complex containing Dicer where the hairpin loop is processed off into a mature RNA duplex. The Dicer complex then coordinates the unwinding and loading of the guide strand into RISC as described above for siRNA processing [8]. Eventually, the complementary guide strand binds and targets mRNA for cleavage. It is important to note that since shRNA hairpin undergoes intracellular processing into an active duplex, the site of hairpin cleavage can only be predicted during its design stages, and therefore the intended targeting sequence is merely theoretical; in sharp contrast to its siRNA counterpart, where it is known a priori as to the identity of the intended targeting sequence.

RNAi as a screening platform has enabled scientists to identify therapeutic targets; however a marginal reproducibility has consistently been observed when the screen data output for similar screens among different laboratories is compared [9-10]. As an example, four RNAi screens were published reporting on novel host factors required for infection by the human immunodeficiency virus (HIV); three were pooled siRNA duplex screens performed by Konig and co-workers [11], Brass and co-workers [12], and Zhou and co-workers [13]; and one was pooled shRNA hairpin screen reported by Yeung and co-workers [14]. On average, 300 gene candidates were reported as factors required for HIV infection in one or the other screen, but the most striking observation was the total lack of overlap across all the four screens, while only three genes (*RELA*, *MED6* and *MED7*) were found to be common to only the three siRNA screens [9]. These reports have stirred up the RNAi screening community leading to a couple of follow up comparative studies driven towards evaluation

of output discrepancies which the associated poor reproducibility to assay types, readouts, cell lines, RNAi technology, library coverage, screening platform, and hit selection methodology [9]; while highlighting the fact that most of these host-virus factors identified did share common functions and biological networks. The four HIV host-virus screens had shared the same scope and therefore should have presented an excellent opportunity to identify strong host factor for HIV infection, and the shRNA hairpin screen would have been expected to identify genes similar to the ones identified in the three siRNA duplex screens. Similarly, we have recently reported on a systematic re-analysis, using the BDA method, of published shRNA screening data sets for 121 cell lines seeking the identification of essential genes and performed using the same TRC1 shRNA library at the Broad Institute, either as arrayed or pooled format [15]. We could not find a single essential gene common to all the cell lines screened and, furthermore, none of the nominated genes showed consistent enrichment and overlap across all the data sets [15]. Moreover, there have been instances where pursuits towards validation and confirmation of targets identified by one screening group did not materialize as successful endeavors [9]. Therefore, these combined observations clearly highlight an inherent variability phenomenon within the RNAi screening data output; more so for the shRNA-based technology either as arrayed or pooled formats.

It could also be argued that the fact some of the published RNAi screens were performed at different geographical locations, and by different research laboratories using different sources of cell lines, etc., would systematically lead to the observed poor overlap of nominated hits. If this is the case, then RNAi as a screening tool is not as robust and reliable for such expensive endeavors of performing large-scale screens. For this reason, we postulated that the best comparative analysis, perhaps, would be the one where the same experimental logistics and personnel were to be used to perform controlled RNAi screens across the two popular technologies (siRNA and shRNA), and allows for the best possible unbiased scenario for addressing the hit discordance issues and concerns at inter-screen level; and ultimately shed some light and provide some guidelines as to the reported discrepancies. Since we have reported on the execution of an siRNA screen against the AMB library covering 21,565 genes to score for those which modulate miR-21 biogenesis, and leading to the nomination of 1,274 gene candidates [16]; we reasoned that by performing an additional genome-wide screen against the TRC1 library, an shRNA plasmid based RNAi technology, covering 16,039 genes would allow us in principle to compare and contrast the performances of the siRNA duplex and shRNA hairpin technologies in arrayed formats. As such, we have recently completed the execution of an arrayed genome-scale shRNA screen against the TRC1 library containing 80,598 hairpins and covering 16,039 genes with an average of 5 shRNA hairpins per gene [17]. We applied a high stringency hit nomination method encompassing criteria of at least 3 active shRNA hairpins per gene and filtered for potential OTEs [15], and leading to the identification of 497 gene candidates as modulators of the miR-21 biogenesis pathway; the knockdown of which resulted in enhancement of the EGFP fluorescence signal output [17]. A critical indicator of the overall performance of the screen was its ability to identify any or all of the seven well-known core genes of the miRNA biogenesis pathway; but only *RNASEN* (*DROSHA*) was identified as a

hit in this shRNA screen [17]; supporting our previous findings that plasmid-based shRNA hairpin screening is not as sensitive as initially perceived to be [7].

In this report, we present the first comprehensive comparative analysis between the two screens performed using either the siRNA or shRNA technologies in 384-well microtiter plate arrayed formats. Our analysis revealed an unexpected and surprising poor overlap between the two technologies with only 29 gene candidates considering a commonality of 15,068 genes, even though this process has been as a controlled environment for reproducibility as it can possibly be. Furthermore, we provide data representative of the differential phenotypic perturbations conferred by identical guide strand sequences obtained from AMB and TRC1; while also providing putative supporting evidence of potential off-target phenotypic perturbations via a simplistic walk-through sliding across the hairpin sequence by one nucleotide to mimic differential hairpin processing inside cells, therefore exploring beyond the theoretical sites of intracellular hairpin cleavage. It is our hope that this study helps us to gain deeper insights into RNAi screening hit discordance [9-10, 18-19], and provides cautionary notes for result interpretation from shRNA based screening for both arrayed and pooled formats.

MATERIALS AND METHODS

Comparative analysis between genome-scale siRNA and shRNA screens

For the purpose of our comparative analysis between the two screens, the gene names were standardized across the two libraries screened and overlaps were performed using custom scripts written in Perl. The network analysis was done in two parts. First, a network map was created in GeneGo's Metacore software using all the known modulators of the miRNA biogenesis machinery and the genes nominated as active in any of the two genome-wide screens were highlighted in the network map. Second, a network map was created using the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA), where 14 selected genes from the known miRNA biogenesis modulators were used as the seed nodes [20-22] and the path explorer function was used to map protein-protein interactions between the nominated hits and the seed nodes. The functional classes and the number of genes identified in each class were derived from the functional annotation clustering using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional analysis tool [23]. The functional annotations were also obtained using the PANTHER Classification System and enriched Gene Ontology (GO) categories were visualized in Cytoscape using BiNGO System [24, 25]. For the purpose of sequence mapping, the mRNA transcripts were obtained from Pubmed for the selected genes based on the RefSeq IDs and the duplex sequences used were as provided by the respective vendors. The pattern matching and match positions were determined using scripts written in Perl.

Comparative analysis between duplex guide sequences

Guide sequences corresponding to all siRNA duplexes in the AMB library were compiled. Likewise, guide sequences corresponding to all shRNA hairpins in the TRC1 library were collected based on two putative hairpin cleavage sites at nt position 32 and position 34; position 32 refers to the nt following the end of the CTCGAG loop on the hairpin and

position 34 was identified based on the 5' counting rule [26]. In the next step, guide sequences from the AMB library were cross compared to guide sequences from the TRC1 library and duplexes with identical guide sequences between the two libraries were selected. The list of identical sequences thus selected was further refined to contain only those sequences which were active in at least one of the two screens, thereby sequences inactive in both the screens were filtered out at this stage of the analysis. The sequences were then grouped based on their differential activity between the two screens.

Sequence walk-through for shRNA hairpins

To qualify for the sequence walk-through analysis, an shRNA hairpin must have the following: 1) An identical sequence match with the guide strand from AMB library targeting the same gene, and 2) Scores as active in the shRNA screen but inactive in the siRNA screen. For the purpose of walk through, each hairpin was sub-divided into all possible 19 nt long sub-sequences, referred to as cleavage variants, across the entire length of the hairpin shifting by one nucleotide each time a sub-sequence was generated, hence a step size of one. The CCGG flanking sequence on the 5' end of the hairpin was excluded and the walk-through was started from nt position 5 of the hairpin. This was repeated for all individual hairpins selected for this analysis. On an average, 37 theoretical cleavage variants per hairpin were obtained and the reverse-complements of these variants were used to determine the potential targeted sequences. These potential target sequences were compared against the human genome using Standard Nucleotide Basic Local Alignment Search Tool (BLASTn) restricting the output to utmost 10 matching transcripts [27].

The results from the BLASTn were manually curated to identify transcripts with at least 10 nt exact match without gaps or mismatches to the cleavage variants and inclusion of the seed heptamer in the matched portion was given preference. Seed heptamer used in this analysis refers to the 7-mer starting at nt position 2 from the 5' end of the guide sequence [15]. A list of all the shifted seed heptamers corresponding to each cleavage variant were also compiled and matched against the known human miRNA sequences to identify miRNAs with identical seed heptamers. The human miRNA sequences were obtained from miRBase release 18 [28] and the information relating to their experimentally validated targets was obtained from TarBase 6.0 [29]. The sequence analysis was performed using scripts written in Perl.

RESULTS

Comparative analysis of the genome-scale siRNA versus shRNA screens

We have reported on the execution of an siRNA screen to identify modulators of the miRNA biogenesis pathway and nominated 1,273 gene candidates with a putative role in the process [16]. Using a consistent approach mimicking a controlled experimental environment, we subsequently performed an additional shRNA screen against the TRC1 library and leading to the identification of 497 gene candidates [17]. The assays for both screens were developed and optimized according to the technology employed (Fig 1A). For siRNA screening, we used a reverse transfection protocol to introduce siRNA duplexes into the cells and measured EGFP signal, NUCL count, and Alamar Blue readout after seven days

[16]. Whereas for shRNA screening, we introduced shRNA hairpins through a transduction process using lentiviral particles and measured EGFP signal and NUCL count post cell fixing and imaging after ten days [7]; shRNA technology uses lentiviral particles for delivery and a puromycin selection step which inherently prolongs the length of the assay (Fig 1A). The overall commonality between the two screens provided us with an opportunity to compare the results so as to assess the performance between the siRNA and shRNA technologies. Frequency distribution analysis of EGFP fluorescence enhancement output revealed differences between the two screens, with the shRNA screening performance was skewed to the left resulting in fewer active hairpins (Fig 1B-C). Similarly, frequency distributions for NUCL counts and a reflection of cell killing, the shRNA screening performance, once more, was different to its siRNA counterpart, and with a much broader differential activity of individual hairpins (Fig 1D-E). We further analyzed and curated the gene coverage of the two libraries. The AMB library contains 64,755 chemically modified duplexes covering 21,565 genes with an average of 3 duplexes per gene target; meanwhile the TRC1 library consists of 80,598 hairpins covering 16,039 genes with an average of 5 hairpins per gene target [15]. Both libraries are in an arrayed format allowing for a one duplex to one gene systematic assessment of knockdown. An analysis of the two libraries revealed a total of 15,068 shared genes with overlap coverage of ~ 94% (Fig 2A).

A comparative analysis at gene-level reveals 29 overlapping gene candidates

As a first step towards the analysis, we compared the degree of overlap among the genes nominated in each screen. We observed a dismal overlap with only 29 gene candidates identified in common, among them was *DROSHA*, a well known core modulator/member of the miRNA biogenesis machinery (Fig 2B). The functional classes associated with the 29 overlapping genes extend over seven broad categories, which were transcription, signal transduction, immune response, metabolism, protein modification, transport, and cell cycle (Table 1). Among them were five regulators of transcription from RNA polymerase II including the transcription initiation factor *GTF2A2*. However, we were not able to identify the key functional classes of exosome and epigenetic control as previously identified in the siRNA screen [16], as an example only one member of the exosome complex *DIS3* was found in common.

Both screens were designed to identify the factors participating in the same biological process; at least, one would have expected to observe a significant overlap among the enriched functional classes. However, our findings indicate a differential enrichment of the nominated gene candidates in GO based functional categories derived from the DAVID functional annotation tool (Fig 2C). We found higher number of gene candidates from the siRNA duplex screen when compared to the shRNA hairpin screen prominently in the GO terms associated with the exonuclease activity (7 in siRNA versus 0 in shRNA), transcription factor activity (68 in siRNA versus 22 in shRNA), transcription from RNA polymerase II promoter (20 in siRNA versus 4 in shRNA), and vesicular localization and targeting (6 in siRNA versus 0 in shRNA). In addition, we have identified 10 members of TGF β /BMP pathway in the siRNA duplex screen while no such enrichment was observed in the shRNA hairpin screen.

Performance evaluation via reference set of known miRNA biogenesis modulators

Identification of known modulators of the miRNA biogenesis machinery in a random screen would serve as a means to benchmark its performance adding a level of confidence to the obtained results. To this end, we compared the performance of the two screens based on the number of known modulators identified by either one. We obtained about 96 genes in total from literature associated with miRNA biogenesis, among which were the core biogenesis genes namely: *DROSHA*, *DICER1*, *EIF2C2* (*AGO2*), *PRKRA* (*PACT*), *TARBP2* and *DGCR8*, and others with a putative regulatory role of the miRNA biogenesis pathway [20-22]. Specifically we looked at the performance difference between the siRNA duplex versus the shRNA hairpin in the identification of some of the core biogenesis genes (Table 2). We observed that the clones validated by the knockdown experiments performed by Sigma-Aldrich, especially the ones associated with the core biogenesis genes did not produce phenotype corresponding to a gain in EGFP signal as would otherwise have been expected. For example *DICER1*, a core biogenesis gene had 4 validated hairpins out of the 5 total targeting hairpins in the TRC1 library; yet only one hairpin scored as active in the shRNA screen. Similarly, 9 hairpins out of 11 hairpins in total were validated for another core biogenesis gene, *PACT*, but none of the hairpins displayed an active phenotype in the shRNA screen. It is important to note that these hairpins were validated in either of these two cell lines A549 or MCF7, and none of them were validated in HeLaS3 cells (Suppl Table 1).

Furthermore, while we identified a total of 8 known factors in the siRNA screen (*DROSHA*, *DICER1*, *EIF2C2*, *DGCR8*, *NHLH1* (*HEN1*), *PDCD6IP* (*ALIX*), *HDAC1*, and *CBP80*), we could only identify 3 known factors in the shRNA screen (*DROSHA*, *ADARI*, and *hnRNPR*). Importantly, four of the nominated gene candidates (*DROSHA*, *DICER1*, *EIF2C2*, and *DGCR8*) which constitute the core genes of the miRNA biogenesis machinery emerged as strong candidates in the siRNA screen while only *DROSHA* was nominated in the shRNA screen. In the next step, we used IPA's build pathway tool to create a network for 14 widely known miRNA biogenesis genes, namely *DHX9*, *DICER1*, *DROSHA*, *EIF2C2*, *HSP90*, *MOV10*, *PACT*, *RAN*, *SMAD2*, *SMAD3*, *TARBP2*, *TGFBR1*, *TNRC6B*, and *XPO5*. We identified and mapped genes which were nominated in the siRNA duplex screen and 67 gene candidates were found to be engaged in protein-protein interactions (PPI) with the selected genes (Fig 3A). Based on the PPI, we found a large number of 35 SMAD-binding proteins. To review the number of nominated genes from the shRNA hairpin screen, which would map to the interaction identified from the siRNA duplex screen, we overlaid the gene candidates from the shRNA hairpin screen onto this network map. Surprisingly, only 3 gene candidates (*DROSHA*, *TGFBR1*, and *DIS3*) overlapped, while all other genes were rendered inactive in the network (Fig 3A). All of these three genes have previously been implicated to play a plausible role in miRNA biogenesis. In the next step, we explored the PPI of the gene candidates nominated in the shRNA hairpin screen in a network with the 14 selected biogenesis genes and found a total of 50 gene candidates from the screen to participate in the network, majority of which 29 gene candidates were associated with the SMAD proteins (Fig 3B). We have found PPIs of the nominated candidates from both screen with the known biogenesis regulators; but unfortunately observed an overall poor gene overlap between the two connectivity maps (Fig 3).

Sequence level analysis suggests differential hairpin cleavage

Our comparative analysis efforts regarding the overlaps at the individual gene level, function, and network maps were indicative of a clear difference in performance between the two RNAi technologies. To explore this difference further and at the duplex's sequence level, we constructed a positional map illustrating the spread of individual duplex matches along the length of the target transcript: 8 representative genes with a putative role in miRNA biogenesis were selected for this analysis (Table 3). Interestingly, we found few examples (*EIF2C2*, *HDAC1*) where the siRNA duplexes and shRNA hairpins sharing the same positional space produced variable phenotypic response in the screen (Table 3). Taking this a step further, we extracted all sequences from the AMB and TRC1 libraries screened, so as to find the total number of overlapping sequences with an identical 19 nt guide sequence. For the purpose of matching to the guide sequence of the siRNA duplexes, the guide sequence from the shRNA hairpin was determined to be 19 nt long sequence starting from two potential cleavage sites at nt pos 32 and nt pos 34, resulting in 276 guide sequences corresponding to the former and 125 guide sequences corresponding to the later found to be common between the two libraries as well as active in at least one of the two screens, giving a total of 401 sequences (Fig 4A, Suppl Table 2). Out of the active sequences, 28 were found to be active in both the screens (7%), while 227 were active in AMB and inactive in TRC1 (Fig 4A). Besides, we also found that 146 sequences (~ 36%) out of the active sequences were exclusively active in TRC1, and therefore inactive in AMB (Suppl Table 2).

To address the question as to why the duplexes with identical guide sequences exhibited differential activities in the shRNA screen, we focused our analysis on the 146 common guide sequences from the shRNA screen that were deemed inactive in the siRNA screen (Fig 4A). For this purpose, we collected the oligonucleotide sequences corresponding to the 146 sequence and disregarded the first four flanking nts, we generated all possible 19 nt sub-sequence along the length of each of those oligonucleotide with an increment of 1 nt (Fig 4B). Depending on the length of the oligonucleotide, on average 37 to 38 different cleavage variants per oligonucleotide were obtained at this stage of the analysis, each of them was subsequently subjected to a sequence match search against the human genome using BLASTn [27]. Top ten hits from the BLASTn search were selected and reviewed for targets other than the intended transcript and found transcripts with either the full-length match or a perfect match around the seed heptamer (Suppl Table 3). A closer look at the BLASTn output revealed multiple occurrences of known or putative miRNA biogenesis modulators as off-targeted transcripts like *DGCR8*, *DHX9* and *EIF2C2* to name a few (Table 4, Suppl Table 4); at which point it can be reasoned that the heterogeneous hairpin splicing perhaps triggered silencing of a critical biogenesis gene, therefore yielding the observed and unintended phenotype. In addition, we extracted the seed heptamer from all the cleavage variants and searched the miRBase to identify exact seed-to-seed heptamer matches with 350 known human miRNAs (Suppl Table 5). Among them, the most enriched were two of the following miRNAs, *miR-516a* and *miR-4444*; which had the exact seed heptamer match with 50 cleavage variants each. Of note we could not identify any experimentally validated targets for *miR-4444*, however some of their predicted targets included *AKT3*, and *E2F1*, *KLK10*, all three of which were identified as active in the shRNA screen [28]. For

miR-516a, *KLK10* was identified as a validated target that was one among the nominated gene candidates in the shRNA screen [30].

DISCUSSION

RNAi as a screening platform will soon be celebrating 15 years of trials and tribulations; this quinquennial occasion will be marked by an impressive achievement of up to ~ 600 publications reporting on its use and identifying gene targets through its random process. Like many new technologies before it, RNAi is plagued by controversy surrounding the ever-growing evidence of different hit lists from these random screens, dismal gene correlation at inter-screen level, and the lack of reproducibility of published results [9, 8]. We reasoned that to gain deeper insights into such variability, we could start by investigating the hit reproducibility and likely performance difference between the two leading RNAi technologies, siRNA duplex and shRNA hairpin. This would be the first comparative analysis between these two most popular flavors of RNAi, with the hope and expectation that the overall performance is on par for both and results in similar hit list content. For the purpose of a fair comparison, we use same equipments to execute the screens, same scientists performing the screens, same cell based assay, and same stringent data analysis method; making this endeavor the best controlled experimental execution at this scale.

We opted to use a gain-of-function assay for this type of comparison because, in our experience; loss-of-function assays tend to be more heterogeneous and noisier, for example lethality screens. We implemented a cell based assay system which measured gain in EGFP expression, under the control of miR-21, upon knockdown of critical components of miRNA biogenesis pathway, as assessed by automated microscopy [31]. We first executed on a siRNA genome-wide screen against the Ambion Silencer Select V4.0 library covering 21,565 genes and leading to the nomination of a hit list of 1,273 candidates [16]; followed by an shRNA genome-wide screen against the TRC1 library covering 16,039 genes and leading to the nomination of a hit list of 497 candidates [17]. At the first stage of gene-level comparison, we observed a marginal overlap of only 29 gene candidates between the two hit lists (Table 1); a surprising result considering the fact that 15,068 genes are commonly covered in both AMB and TRC1. In terms of identification and overlap of known biogenesis factors, we could identify only one core component, *DROSHA*, out of the seven well known miRNA biogenesis modulators, in common to both screens (Table 2). Furthermore, a PPI network level analysis using selected miRNA biogenesis modulators as seed nodes revealed distinct gene candidates from the two screens interacting with the same biogenesis pathway; 67 genes candidates from AMB and 50 from TRC1 were found to interact with the selected seed nodes. However, an overlay of the network maps from AMB upon TRC1 and vice versa, exhibited a dismal outcome as most of the PPIs were exclusive to either one of the screens with only 3 gene candidates (*DROSHA*, *TGFBRI*, and *DIS3*) found in common (Fig 3). This observation is consistent with previous comparative analysis reports where disparate gene lists were harmonized with regards to their participation in similar pathways and functional classes [9]. However, the question remains whether such a participation of different hits in common pathways is sufficient to assess the success and reproducibility of

an RNAi screen, if so then what would be the value proposition of the individual gene hits identified in independent RNAi screens.

The ability of shRNA hairpins to confer a robust on-target knockdown relies on four major transactions within the host cell: 1) Cargo delivery via the viral particle harboring an shRNA construct, 2) random integration into the host genome, 3) efficient transcription of the genomic insert into pre-shRNA, and 4) intracellular processing of pre-shRNA into a mature siRNA duplex [7, 32]. The first three transactions would produce heterogeneous cell populations considering that the transduction efficiencies are never 100% and with a higher probability of multiple integrations per host cell genome. The fourth transaction, however, includes processing of the transcribed hairpin into a functional duplex, accuracy of which heavily relies on the site of cleavage by the *DICER* complex. Models to determine exact sites of hairpin cleavage have been previously proposed, but those by far are theoretical estimators [26]. The only experimental evidence towards actual cleavage products generated intracellularly was provided by Gu and co-workers, who published data demonstrating the presence of multiple sequence variants inside HEK293 cells shortly after transduction with the miR30-based plasmids [33]. Besides Gu elegant demonstration, unfortunately, there are no other reports strongly supporting the following: 1) demonstration of a precise intracellular hairpin cleavage, 2) demonstration of an exclusive silencing of the intended target gene, and finally providing comparable silencing thresholds across multiple cell lines. With in mind, we decided to look for possible reasons as to the differential performance between siRNA and shRNA based screens at the sequence level.

We hypothesized that duplexes with identical sequences would most likely result in comparable phenotypic perturbations; at this stage, we assembled data on those duplexes which increased EGFP expression above the threshold in either AMB or TRC1; their sequences were cross-compared and the resulting common targeting sequences were selected. Of note, AMB was a collection of predetermined guide strands while TRC1 was a collection of the theoretical guide strands. Surprisingly, the results were striking in terms of discrepant phenotypic outcomes among identical sequences from the two libraries; 56% of the sequences were active exclusively in AMB, and 37% were active exclusively in TRC1, while a very small proportion of 7% exhibited an inter-screen correlation (Fig 4A).

It can be argued that perhaps the TRC1 shRNA hairpin counterparts of the identical sequences solely active in AMB underwent an inaccurate intracellular processing and therefore failed to exhibit the desired phenotype. However, a more perplexing observation was the identification of 146 active identical sequences in TRC1 with no observed activity in AMB. To investigate this further, we performed a sequence walk-through for the active shRNA hairpins to survey all theoretical byproducts in an event of improper hairpin processing. Surprisingly, we observed a strong enrichment in random genes upon performing a BLASTn search for the cleavage variants against the human genome; up to 123 variants corresponded to a complete 19 nt sequence coverage in ~ 15 different genes. Interestingly, 217 of these off-targeted transcripts were previously nominated as gene candidates in the siRNA screen and 121 of these off-targeted transcripts were nominated as gene candidates in the shRNA screen.

The walk-through analysis revealed yet another astounding finding that among the numerous off-targeted transcripts were the miRNA biogenesis regulators; importantly, *DGCR8*, *EIF2C2*, *DICER* and *DROSHA* were identified as potential off-targeted transcripts. Since some of the off-targets transcripts reported here are in fact core modulators essential for miRNA biogenesis, and whose down regulation would undoubtedly result in a positive phenotypic response, it becomes difficult to assess the role of TRC1 nominated gene candidates with regards to miRNA biogenesis as factual versus fallacious. In the majority of the cases, match length between the guide strand and off-target was between 10 to 14 nts, and more than often it included a seed heptamer match; considering that even a partial transcript match enables a guide strand to down regulate the mRNA expression level [34], can we *de facto* correlate the knockdown produced by an shRNA hairpin to its intended target vis-à-vis an off-targeted transcript(s)? We are not assuming that all these transcripts are being active inside the host cell, however, we demonstrate that a plethora of off-targets become vulnerable to unintended silencing by inefficient intracellular processing of the shRNA hairpin (Fig 5).

There is an emergence of clone validation among the shRNA hairpin libraries to survey the knockdown efficiencies of a given shRNA hairpin; 24,368 clones in TRC1 are documented to confer a stable knockdown in a battery of cell lines. In the shRNA screen, we found that 32% of the active hairpins in the HeLaS3 host cell line screened were regarded as independently validated [17]. However, we also found that 1,270 TRC1 shRNA hairpins corresponding to nominated gene candidates in AMB were validated but found inactive in the shRNA screen. Since both screens were a gain of function, the inactivity of validated hairpins targeting nominated candidates from AMB demonstrates hit discordance not just at a gene level but also at the sequence level with regards to differential duplex performance.

Furthermore, we surveyed the validation status of the hairpins targeting the core miRNA biogenesis genes, more so since we were able to nominate only *DROSHA* and not the other six essential biogenesis regulators as active gene candidates in TRC1. Surprisingly, we found 27 validated shRNA hairpins out of the 58 inactive hairpins targeting the 14 representative biogenesis genes and exhibited an average knockdown efficiency up to 96%; a rather bafflingly observation since one would have expected validated shRNA hairpins to produce a stable knockdown of the core miRNA biogenesis genes at least, given that it is a gain-of-function screen. Upon a closer scrutiny of the validation data set, we found that they were performed by measuring reduction of mRNA expression levels by individual hairpins in any one of these 6 different cell lines: 293T/17, A3, A549, HeLa, MCF7, or MCH58. In a recent genome-scale arrayed shRNA hairpin screen to identify factors essential for cellular survival, we made striking observations with regards to inactivate phenotypic response rendered by validated hairpins targeting known essential genes; importantly, *PLK1*, a routine negative control, had 20 validated hairpins out of which only 4 hairpins resulted in reduction of residual NUCL [7]. These 20 hairpins targeting *PLK1* were independently validated in the MCF7 cell line; an observation that led us to postulate that the hairpins might have undergone a differential intracellular processing inside the HeLa cell line instead, which was used in the screen, versus the MCF7 cell line, which was used in the validation. These preliminary observations along with a similar inference reported in a

recent RNAi review, are all indicative of a possible association between the levels of knockdown efficiencies and the type of cell line used, and perhaps such associations should be explored in depth [35]. Furthermore, there is growing evidence of differential processing of shRNA hairpins across different cell lines resulting in off-targeted transcripts, and dampening the theoretical notion that shRNA hairpins confer knockdown specificity. Therefore, it may be more prudent and meaningful to catalog the knockdown efficiencies of individual shRNA hairpins in a multitude of cell lines to substantiate and confirm their stable and specific knockdown.

An era of skepticism regarding RNAi data output has emerged due to growing evidence of disparate hit lists originating from a multitude of RNAi screens with up to ~ 600 publications reporting on genes identified through this random process. Bushman and co-workers had published a pair-wise gene-level overlap for host-virus factors required for HIV infection obtained from 9 screens, 3 of which were pooled siRNA duplex screens performed by different groups, in different experimental set-up and implementing different methods of hit nomination; surprisingly none of the genes was identified in common across the board [36]. Recently, we reported on the re-analysis of data outputs from two published lethality screens; a pooled shRNA hairpin screen performed by Cheung and co-workers in 102 cell lines, and an arrayed shRNA hairpin screen performed by Barbie and co-workers in 19 cell lines [15]. A consistent methodology for hit nomination was utilized by applying the BDA method but once again, none of the genes emerged as common among the 121 cell lines even though all screens were catered towards measuring cell death [15].

In summary, we have reported on a head-to-head comparative analysis between siRNA duplex and shRNA hairpin screens at a genome-scale. A plasmid-based shRNA hairpin genome-wide screen revealed a darker side in a gain-of-function assay in comparison to the illuminating siRNA duplex genome-wide screen, brighter and populous in hit candidates (Fig 6). Both screens shared the same scope and performed using similar experimental set up with only 29 gene candidates in common to both TRC1 and AMB; including only one core miRNA biogenesis gene (*DROSHA*) while the six other core biogenesis modulators were completely missed in the shRNA screen, even though they are independently validated but failed to confer a phenotypic perturbation in the HeLaS3 cell line. Though a dismal overlap at the inter-screen level was noted, the obtained hits shared similarity at the biological network level of analysis. In addition, we provide, for the first time, a simple and systematic walk-through of the shRNA hairpin sequence, surveying off-targeted transcripts, and based on the plethora of potential off-targets obtained which included the seven core biogenesis modulators, we postulate that differential intracellular processing of the plasmid-based, and even for the miR30-based shRNA, within different host cell lines may well be the sole culprit for the observed discrepancies to date. Experiments are ongoing to further test our hypothesis of differential hairpin processing originating from a single hairpin resulting in unintended down regulation of other mRNA transcripts within the cell; and perhaps validate the notion that poor hit reproducibility observed among published RNAi screen hit lists lies within the chosen RNAi technology.

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 A. Calder for their help during the course of this study. We also thank Wenjing Wu, Medical Graphics, MSKCC for her help with the artwork depicted in Figure 6. 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
siRNA	small interfering RNA
shRNA	short hairpin interfering RNA
BDA	Bhinder-Djaballah Analysis method
H score	hit rate per gene score
OTE	off-target effect
miRNA	microRNA

REFERENCES

1. Fire A, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific gene interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998; 391(6669):806–811. [PubMed: 9486653]
2. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. 1999; 286(5441):950–952. [PubMed: 10542148]
3. Aagaard L, Rossi JJ. RNAi therapeutics: principles, prospects and challenges. *Adv. Drug Deliv. Rev.* 2007; 59(2-3):75–86. [PubMed: 17449137]
4. Gavrillov K, Saltzman WM. Therapeutic siRNA: principles, challenges, and strategies. *Yale J. Biol. Med.* 2012; 85(2):187–200. [PubMed: 22737048]
5. 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]
6. Parker JS, Roe SM, Bardford D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*. 2005; 434(7033):663–666. [PubMed: 15800628]
7. Bhinder B, Antczak C, Ramirez CN, Shum D, Liu-Sullivan N, Radu C, Frattini MG, Djaballah H. An Arrayed Genome Scale Lentiviral Enabled shRNA Screen Identifies Lethal & Rescuer Gene Candidates. *Assay Drug Dev. Technol.* 2013; 11(3):173–190. [PubMed: 23198867]
8. 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]
9. Bhinder B, Djaballah H. A Decade of RNAi Screening: Too Much Hay & Very Few Needles. *Drug Disc. World.* 2013; 14:31–41.
10. Bhinder B, Djaballah H. Systematic analysis of RNAi reports identifies dismal commonality at gene-level & reveals an unprecedented enrichment in pooled shRNA screens. *Comb. Chem. High Throughput Screen.* 2013 (In press).

11. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, Chiang CY, Tu BP, DeJesus PD, Lilley CE, Seidel S, Opaluch AM, Caldwell JS, Weitzman MD, Kuhlen KL, Bandyopadhyay S, Ideker T, Orth AP, Miraglia LJ, Bushman FD, Young JA, Chanda SK. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*. 2008; 135(1):49–60. [PubMed: 18854154]
12. Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, Lieberman J, Elledge SJ. Identification of host proteins required for HIV infection through a functional genomic screen. *Science*. 2008; 319(5865):921–926. [PubMed: 18187620]
13. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, Stec E, Ferrer M, Strulovici B, Hazuda DJ, Espeseth AS. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe*. 2008; 4(5):495–504. [PubMed: 18976975]
14. Yeung ML, Houzet L, Yedavalli VS, Jeang KT. A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J. Biol. Chem*. 2009; 284(29):19463–19473. [PubMed: 19460752]
15. Bhinder B, Djaballah H. A simple method for analyzing actives in random RNAi screens: Introducing the “H score” for gene nomination and prioritization. *Comb. Chem. High Throughput Screen*. 2012; 15(9):686–704. [PubMed: 22934950]
16. Shum D, Bhinder B, Ramirez CN, Radu C, Calder PA, Beauchamp L, Farazi T, Landthaler M, Tuschli 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]
17. Shum D, Bhinder B, Djaballah H. Modulators of the microRNA biogenesis pathway via arrayed lentiviral enabled RNAi screening for drug and biomarker discovery. *Comb. Chem. High Throughput Screen*. 2013 (In press).
18. Kaelin WG Jr. Use and Abuse of RNAi to Study Mammalian Gene Function. *Science*. 2012; 337(6093):421–422. [PubMed: 22837515]
19. Djaballah H. Random RNAi Screening Data Analysis: A Call for Standardization. *Comb. Chem. High Throughput Screen*. 2012; 15(9):685. [PubMed: 23017031]
20. Saj A, Lai E.C. Control of microRNA biogenesis and transcription by cell signaling pathways. *Curr. Opin. Genet. Dev*. 2011; 21(4):504–510. [PubMed: 21592778]
21. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet*. 2010; 11(9):597–610. [PubMed: 20661255]
22. Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol*. 2009; 11(3):228–234. [PubMed: 19255566]
23. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc*. 2009; 4(1):44–57. [PubMed: 19131956]
24. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, Ladunga I, Ulitsky-Lazareva B, Muruganujan A, Rabkin S, Vandergriff JA, Doremiex O. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res*. 2003; 31(1):334–341. [PubMed: 12520017]
25. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005; 21(16):3448–3449. [PubMed: 15972284]
26. Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, Patel DJ, Kim VN. Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*. 2011; 475(7355):201–205. [PubMed: 21753850]
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol*. 1990; 215(3):403–410. [PubMed: 2231712]
28. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 2011; 39:D152–D157. [PubMed: 21037258]
29. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, Gerangelos S, Koziris N, Dalamagas T, Hatzigeorgiou AG. Tarbase 6.0: Capturing the Exponential Growth of

- miRNA Targets with Experimental Support. *Nucleic Acids Res.* 2012; 40:D222–229. [PubMed: 22135297]
30. White NM, Chow TF, Mejia-Guerrero S, Diamandis M, Rofael Y, Faragalla H, Mankaruous M, Gabril M, Girgis A, Yousef GM. Three dysregulated miRNAs control kallikrein 10 expression and cell proliferation in ovarian cancer. *Br. J. Cancer.* 2010; 102(8):1244–1253. [PubMed: 20354523]
31. 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(9):529–541. [PubMed: 22540737]
32. Rao DD, Vorhies JS, Senzer N, Nemunaitis J. siRNA vs. shRNA: similarities and differences. *Adv. Drug Deliv. Rev.* 2009; 61(9):746–759. [PubMed: 19389436]
33. Gu S, Jin L, Zhang Y, Huang Y, Zhang F, Valdmanis PN, Kay MA. The Loop Position of shRNAs and Pre-miRNAs Is Critical for the Accuracy of Dicer Processing In Vivo. *Cell.* 2012; 151(4): 900–911. [PubMed: 23141545]
34. Jackson AL, Burchard J, Schelter J, Chau BN, Cleary M, Lim L, Linsley PS. Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA.* 2006; 12(7):1179–1187. [PubMed: 16682560]
35. Boettcher M, Hoheisel JD. Pooled RNAi Screens - Technical and Biological Aspects. *Curr. Genomics.* 2010; 11(3):162–167. [PubMed: 21037854]
36. Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, Zhou H, Hazuda DJ, Espeseth AS, König R, Bandyopadhyay S, Ideker T, Goff SP, Krogan NJ, Frankel AD, Young JA, Chanda SK. Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* 2009; 5(5):e1000437. [PubMed: 19478882]

A)

Figure 1. Genome-scale siRNA and shRNA assay methodology and screening performance

A) Workflow comparison of the optimized assays for the siRNA and shRNA screens; for siRNA, cells (500 cells per well) were screened against the AMB library at 50 nM using reverse transfection and evaluated at 6 days post-seeding. For shRNA, cells (1,000 cells per well) were screened against the TRC1 library at a multiplicity of infection (MOI) of 4 and evaluated 9 days post-transduction. **B)** Histogram frequency distribution plot of individual siRNA duplex performances as assessed by EGFP signal enhancement for the AMB library. **C)** Histogram frequency distribution plot of individual shRNA hairpin performances as assessed by EGFP signal enhancement for the TRC1 library. **D)** Histogram frequency distribution plot of individual siRNA duplex performances on cellular viability as assessed by residual nuclear counts for the AMB library. **E)** Histogram frequency distribution plot of individual siRNA duplex performances on cellular viability as assessed by residual nuclear counts for the TRC1 library.

A) **B)**

C)

Figure 2. Comparative analysis of genome-scale siRNA and shRNA libraries and resulting nominated gene candidates
A) Overlap analysis for common genes between the AMB and the TRC1 library; 15,068 genes were found to be common to both. **B)** Overlap analysis of the nominated gene candidates from the AMB and the TRC1 library; 29 hits were found to be common to both screens. **C)** GO functional analysis of nominated genes from the AMB and the TRC1 libraries showing differential outcome from the screens.

A)

B)

siRNA ← shRNA

Figure 3. Network analysis of nominated genes from siRNA and shRNA screening

A) Evaluation of nominated gene candidates from siRNA to shRNA screening using IPA. **B)** Evaluation of nominated gene candidates from shRNA to siRNA screening using IPA.

A)

A Active duplex
I Inactive duplex
— AMB Guide strand
— TRC1 Guide strand

B)

Figure 4. Sequence analysis for duplexes with identical 19-nucleotide guide strand match

A) Identification of sequences from the TRC1 and the AMB libraries with a common guide sequence and phenotypic activity in respective screens leading to a selection of 147 sequences exclusively active in the TRC1 screen. **B)** Workflow describing breakup of active TRC1 hairpin sequences into all possible 19 nucleotide cleavage variants; with BLASTn analysis to identify potential off-targets and miRNA mimics.

Figure 5. Illustration of differential shRNA hairpin processing

A representative shRNA hairpin randomly selected from the list of active duplexes in TRC1 is shown; the hairpin has been independently validated. A precise cleavage of the hairpin at its theoretical site would yield an accurate guide strand with a potential to silence the target gene. Any deviation from the theoretical site of hairpin cleavage yields numerous distinct guide strands with shifted seed locations, resulting in off-targeted transcripts and misleading phenotypic outcomes.

Figure 6. Revealing the dark side of shRNA hairpin processing in a gain-of-function assay

A siRNA duplex has a pre-defined guide strand, which unwinds inside the cell and targets a specific gene, resulting in the desired phenotypic perturbation, as shown by the bright glow of EGFP signal. On the other side, precise theoretical cleavage of a shRNA hairpin inside a cell remains uncertain, the final (un)known duplex produced would determine the phenotypic outcome, and perhaps a misleading one, as shown by the dull dark side.

Table 1

List of the 29 overlapping genes from the siRNA and the shRNA screens and their GO functional analysis

Gene	Refseq ID	Description	Molecular function	Biological function	Protein class
ABHD7	NM_173567	Abhydrolase domain-containing protein 7	serine-type peptidase activity	immune system process	serine protease
ARHGEF19	NM_153213	Rho guanine nucleotide exchange factor 19	Guanine nucleotide exchange factor	regulation of signal transduction (Rho-GTPase)	-
ATF4	NM_001675	Cyclic AMP-dependent transcription factor ATF-4	transcription factor activity	regulation of transcription from RNA polymerase II promoter	CREB transcription factor
BRD4	NM_058243	Bromodomain-containing protein 4	nucleic acid binding	transcription from RNA polymerase II promoter, establishment/maintenance of chromatin architecture	chromatin/chromatin-binding protein
C17ORF51	XM_378661	Chromosome 17 open reading frame 51	-	-	-
CDC2	NM_001786	Cell division control protein 2 homolog	protein kinase activity	mitosis	non-receptor serine/threonine protein kinase
CKS2	NM_001827	Cyclin-dependent kinases regulatory subunit 2	protein binding, kinase regulator activity	cell cycle	kinase modulator
CLEC4M	NM_014257	C-type lectin domain family 4 member M	receptor activity/cell adhesion	macrophage activation	defense/immunity protein, receptor
CPT1C	NM_152359	Carnitine O-palmitoyltransferase 1, brain isoform	acyltransferase activity	cellular amino acid metabolic process, fatty acid metabolic process	acyltransferase
DIS3	NM_014953	Exosome complex exonuclease RRP44	exoribonuclease activity	rRNA processing	exoribonuclease
DROSHA	NM_013235	Ribonuclease 3	endoribonuclease activity	rRNA metabolic process	endoribonuclease
FOSL2	NM_005253	Fos-related antigen 2	transcription factor activity	regulation of transcription from RNA polymerase II promoter	transcription factor
GABRG2	NM_000816	Gamma-aminobutyric acid receptor subunit gamma-2	GABA receptor activity	signal transduction	GABA receptor
GATM	NM_001482	Glycine amidinotransferase, mitochondrial	transferase activity	cellular amino acid metabolic process	transferase
GGPS1	NM_004837	Farnesyltransferase	acyltransferase activity	metabolic process	acyltransferase
GKAP1	NM_025211	G kinase-anchoring protein 1	structural molecule activity	-	structural protein
GTF2A2	NM_004492	Transcription initiation factor IIA subunit 2	transcription factor activity	transcription initiation from RNA polymerase II promoter	transcription factor
HAND2	NM_021973	Heart- and neural crest derivatives-expressed protein 2	transcription factor activity	regulation of transcription from RNA polymerase II promoter	basic helix-loop-helix transcription factor
HSPA9	NM_004134	Stress-70 protein, mitochondrial	-	immune system process, protein folding, protein complex assembly	Hsp70 family chaperone

Gene	Refseq ID	Description	Molecular function	Biological function	Protein class
PLXNC1	NM_005761	Plexin-C1	transmembrane receptor protein kinase activity	transmembrane receptor protein tyrosine kinase signaling pathway	tyrosine protein kinase receptor
PRMT7	NM_019023	Protein arginine N-methyltransferase 7	methyltransferase activity	protein amino acid methylation	methyltransferase
RAB3C	NM_138453	Ras-related protein Rab-3C	GTPase activity	intracellular protein transport/signaling cascade, receptor-mediated endocytosis	small GTPase
RAP1A	NM_002884	Ras-related protein Rap-1A	GTPase activity	MAPKKK cascade, cell adhesion	small GTPase
RFTN1	NM_015150	Raftlin	-	-	-
RPN2	NM_002951	Dolichyl-diphosphooligosaccharide-protein glycosyltransferase subunit 2	transferase activity	protein modification process	glycosyltransferase
SLC30A3	NM_003459	Zinc transporter 3	transmembrane transporter activity	cation transport	transporter
TGFBR1	NM_004612	TGF-beta receptor type-1	transmembrane receptor protein kinase activity	cytokine-mediated signaling pathway, transforming growth factor beta receptor activity	serine/threonine protein kinase receptor
TUBA1C	NM_032704	Tubulin alpha-1C chain	structural constituent of cytoskeleton	intracellular protein transport, tubulin complex	tubulin
XRCC3	NM_005432	DNA repair protein XRCC3	hydrolase activity	immune system process, DNA recombination	DNA strand-pairing protein

Table 2
Comparative analysis of the observed activities of the core miRNA biogenesis genes in the siRNA and shRNA screens

miRNA biogenesis machinery	Genes	Refseq ID	Primary screens			
			AMB		TRC1	
			Activity (EGFP signal)	H score		
Core components						
Microprocessor complex	DGCR8	NM_022720	3/3 siRNA	100	2/4 shRNA	50
	DROSHA	NM_013235	2/3 siRNA	67	3/5 shRNA	60
RISC loading complex	EIF2C2	NM_012154	3/3 siRNA	100	2/5 shRNA	40
	DICER1	NM_030621	3/3 siRNA	100	1/5 shRNA	20
	PRKRA	NM_003690	0/3 siRNA	0	0/11 shRNA	0
	TARBP2	NM_004178	1/3 siRNA	33	0/4 shRNA	0
Nuclear export	XPO5	NM_020750	0/3 siRNA	0	Not found in TRC1	NA
Regulators						
RISC formation & activity	TNRC6B	NM_015088	1/3 siRNA	33	1/5 shRNA	20
	DHX9	NM_001357	0/3 siRNA	0	2/5 shRNA	40
	MOV10	NM_020963	0/3 siRNA	0	0/5 shRNA	0
	HDAC1	NM_004964	2/3 siRNA	67	0/10 shRNA	0
miRNA processing	hmrNPR	NM_005826	0/3 siRNA	0	4/5 shRNA	80
	ADAR	NM_001111	1/3 siRNA	33	4/5 shRNA	80
miRNA stability	NHLH1	NM_005598	3/3 siRNA	100	1/5 shRNA	20
	PDCD6IP	NM_013374	3/3 siRNA	67	2/5 shRNA	40

Table 3

Overlap analysis of seed sequences on mRNA transcripts for siRNA duplexes and shRNA hairpins

Gene	RefSeq ID	RNAi type	Supplier ID	Activity	Guide/Oligonucleotide Sequence (5' → 3') [†]
HDAC1 [*]	NM_004964	siRNA	s73	A	<u>TTTTCGGTAGAGCCATAGTT</u>
		shRNA	4818	I	CCGGGCTGCTCAACTATGGTCTCTACTCGAGTAGAGCCATAGTTGAGCAGCTTTTT
HDAC1 [*]	NM_004964	siRNA	s74	A	<u>TTACTTTGGACATGACCGGCT</u>
		shRNA	4816	I	CCGGGCCGGTCAATGTCCTCAAGAGTAATCTCGAGATTACTTTTGGACATGACCGGGCTTTT
EIF2C2 [*]	NM_012154	siRNA	s25932	A	<u>TTTGCTAAATCTCTCTCTGCGG</u>
		shRNA	7864	I	CCGGGGCAAGAAAGAGATTAGCAAACTCGAGTTTGGCTAACTCTCTCTGCGGTTTTT
RNASEN [*]	NM_013235	siRNA	s26492	A	<u>TTCCGAACAAAGTTAAAGTTC</u>
		shRNA	22250	I	CCGGCGAAGCTCTTTGGTGAATATCTCGAGATTATTCACCAAAAGAGCTTCGTTTTT
APOL3	NM_014349	siRNA	s37435	A	<u>TACCTTCAATCGGTCAATGCT</u>
		shRNA	160295	I	CCGGCAATTGACCGATTGAAGGTAATCTCGAGAAATACCTTCAATCGGTCAATGTTTTTTG
RGMB	NM_001012761	siRNA	s50017	A	<u>ATCCTTGAAAAGTTCTGAGGTG</u>
		shRNA	133928	I	CCGGCCTCAGAACTTTCAAGGATAACTCGAGTTATCCTTGAAAAGTTCTGAGGTTTTTTG
ALAS2	NM_000032	siRNA	s1225	A	<u>ATCAATTACTACACAGACGGA</u>
		shRNA	45760	I	CCGGCGTCTGGTGTAGTAATGATTACTCGAGTAATCAATTACTACACAGACGTTTTTG
FAM5C	NM_199051	siRNA	s50542	A	<u>TCTTGTGCTAAATCCCTGCCG</u>
		shRNA	153297	I	CCGGGCAAGGATTAGCACAAAGATACTCGAGTAATCTTGTGCTAAATCCCTGCTTTTTTG
MAST1	NM_014975	siRNA	s22771	A	<u>TAATAAGCAACTTCTTCACCA</u>
		shRNA	196385	I	CCGGGTGAAGAAAGTTGCTTATTATCTCGAGGATAATAAGCAACTTCTTCACTTTTTG
AURKB	NM_004217	siRNA	s17611	A	<u>ATAGTTGTAGAGACGCAGGAT</u>
		shRNA	777	A	CCGGCTGCGTCTCTACAACTATTCTCGAGAAATAGTTGTAGAGACGCAGGTTTTT
SMAD3 [*]	NM_005902	siRNA	s8400	I	<u>TTGAGTTTCTTGACCCAGGCTC</u>
		shRNA	20013	I	CCGGGAGCCTGGTCAAGAAACTCAACTCGAGTTTGTGTTTCTTGACCAAGGCTCTTTTT

The predicted loop region is highlighted in green, A; Duplex is scored active in screen, I; Duplex is scored inactive in screen

[†]The overlapping regions are highlighted in red text^{*} Known miRNA biogenesis gene

Table 4

Representative cleavage variants with off-target matches in known miRNA biogenesis modulators

Input Gene	Supplier ID	Cleavage variant Start	Guide Match Start	Guide Match End	Seed match	Oligonucleotide sequence [†]	OTE predicted	Screen performance	
								AMB	TRC1
ATP10D	51540	21	2	12	Yes	CCGGGCGCTATGTGAACAATCGAATACTCGAGTATTCGATTGTTCACATAGCGTTTTTG	DGCR8	A	I
OR2T4	61184	21	2	12	Yes	CCGGCCAAAACATCCAAATGGCCAATACTCGAGTATTGGCCATTGGATGTTTGGGTTTTTG	DGCR8	A	I
TAS2R49	63323	27	8	19	No	CCGGCTGGGCAGTAACCAATCATTTCTCGAGAAATGATTGGTTACTIGCCCAGGTTTTTG	DICER1	A	I
IL6	59207	27	1	14	Yes	CCGGCAGAACGAAATTGACAAACAAACTCGAGTTTTGTGTCAAATTCGTTCTGTTTTTG	EIF2C2	A	I
IL6	59207	17	1	14	Yes	CCGGCAGAACGAAATTGACAAACAAACTCGAGTTTTGTGTCAAATTCGTTCTGTTTTTG	EIF2C2	A	I
SLC26A1	43544	38	6	19	No	CCGGCAGCCTCTATACGTCCTTCTTCGAGAAAGAGGACGTATAGAGGCTGTTTTTG	MOV10L1	A	I
CTTN	40273	30	1	10	Yes	CCGGCGGCAAAATACGGGTATCGAACAACTCGAGTTGTCGATACCGTATTTGCCGTTTTTG	DROSHA	A	A
CLPTM1	82947	20	10	19	No	CCGGCTCCATCTACATCCACGTTTACTCGAGTAAACCGTGGAATGTAGATGGAGTTTTTG	TGFBF1	A	A
FOXF1	13954	33	1	14	Yes	CCGGCGCCTCTTATATCAAGCAGCACTCGAGTCTGCTTGCTTGATATAAGAGGCGTTTTT	DHX9	I	I
PHOX2B	13312	28	1	12	Yes	CCGGCGTCTCTATCTTCGCTCCAAACTCGAGTTTGGAGCGAAGATAGGACGCTTTTT	LIN28B	I	I
MEF2B	15738	23	1	13	Yes	CCGGCGGCGACTTTCCTAAGACCCTTCTCGAGAAAGGCTTAGGAAAGTCGCCGTTTTT	TARBP2	I	I
PAWR	20306	19	1	10	Yes	CCGGTAGATATTCTCGAACAGATACTCGAGTATCTGTTCGAGAAATATCTACTTTTT	TNRC6B	I	I
PTGFR	14169	33	1	12	Yes	CCGGCGGTGTATTGGAGTCAAACTCGAGTTTGTGACTCCAATACACCGCTTTTT	ADAR	I	A
TECTA	73541	33	1	16	Yes	CCGGCCGCACTGTCTATGTCAATAAACTCGAGTTATTGACATAGACAGTCCGGTTTTTG	EIF4E	I	A

A; Duplex is scored active in screen, I; Duplex is scored inactive in screen

[†] Blue; selected guide sequence, Red; matching sequence from BLASTn search, Gray; remaining oligonucleotide sequence.