RNA interference screens to uncover membrane protein biology

Anthony B. Mak and Jason Moffat

Advance Access publication date 22 June 2013

Abstract

In this review, we discuss the use of RNA interference screens to identify genes involved in the regulation and function of membrane proteins. Briefly, cells expressing the membrane protein of interest can be transduced with a pooled lentiviral short-hairpin RNA (shRNA) library containing tens of thousands of unique shRNAs. Transduced cells are then selected or fractionated based on specific criteria, such as membrane protein expression or function. shRNAs from selected cell populations are then deconvoluted and quantified using microarray analyses or high-throughput sequencing technologies. This allows individual shRNAs to be scored and cutoffs can be made to generate a list of shRNA hits. Bioinformatic analyses of gene targets of shRNA hits can be used to identify pathways and processes associated with membrane protein biology. To illustrate this functional genomics approach, we discuss pooled lentiviral shRNA screens that were performed to identify genes that regulate the transcription and cell-surface expression of the cancer stem cell marker CD133. This approach can be adapted to study other membrane proteins, as well as specific aspects of membrane proteins, such as their function or downstream signaling effects.

Keywords: RNA interference; short hairpin RNA; high-throughput screening; membrane proteins; CD133; AC133

INTRODUCTION

Membrane proteins play an essential role in both normal and cancer biology, but our knowledge of their regulation, structure and function is orders of magnitude less than cytosolic and nuclear proteins. Importantly, approximately one fifth of the human genome is predicted to encode for membrane proteins and mutations in these genes can result in cancer [1]. The study of membrane proteins has been challenging due to their inherent traits, such as increased hydrophobicity, instability and flexibility [2]. As membrane proteins are embedded in a lipid environment, their extraction requires binding of detergents to hydrophobic domains, which can result in protein inactivation, denaturation and dissociation of molecular interaction partners. Therefore, to obtain sufficient stability for downstream analyses, membrane protein solubilization may require laborious and non-biological optimization of various types and amounts of detergent, lipids and buffers. These conditions make it challenging to solve membrane protein structures, as well as to identify protein–protein interactions and protein post-translational modifications.

Additional challenges with studying membrane proteins may include their relatively low expression level in the cell. Experimental approaches that attempt to overcome issues of low abundance, such as heterologous overexpression, can result in issues such as cytotoxicity, protein misfolding and protein instability. Taken together there is a need of experimental approaches that can cooperate with the inherent properties of membrane proteins to study their regulation, structure and function.
EXPERIMENTAL APPROACH

A functional genomics approach to study membrane proteins in live mammalian cell culture is RNA interference (RNAi) screens. RNAi is a genetic tool used to reduce gene expression by post-transcriptional silencing of genes resulting in transcript degradation or translational inhibition. This phenomenon is mediated by double-stranded RNA (dsRNA) (e.g. small interfering RNA (siRNA), endoribonuclease-prepared siRNA (esiRNA) and short-hairpin RNA (shRNA)) encoded to target certain transcripts resulting in reduced expression of the gene product. siRNAs and esiRNAs can be introduced into cells using transfection to temporarily repress gene transcript targets. However, not all cell types can be efficiently transfected with dsRNAs. This limitation can be circumvented with the use of recombinant lentiviral technology to allow for long-term target gene transcript repression due to stable integration of shRNAs into host cell genomes. Importantly, recombinant lentiviral particles can transduce virtually all mammalian cell type without the need for cell division.

Due to the availability of genome sequences and strategies to design effective dsRNA molecules, lentiviral shRNA libraries have been generated to study gene fractions. For example, The RNAi consortium (TRC) has constructed a lentiviral shRNA library containing ~5 shRNAs designed to target each of the ~22,000 human and mouse genes [3]. The incorporation of multiple shRNAs per target gene in the library increases the likelihood that a target gene will be repressed. These shRNA libraries are encoded on the lentiviral vector pLKO.1, which allows for recombinant lentiviral production, host cell transduction and selection of stable shRNA integrations using puromycin. pLKO.1 is compatible with second generation lentiviral packaging systems resulting in the production of replication incompetent lentiviral particles, while maintaining sufficiently high viral titers [4].

The development of these lentiviral shRNA libraries allows for high-throughput RNAi screens. Depending on the downstream screening approach, lentiviral particle production can present in an arrayed or pooled format. Although the arrayed screening format has its advantages, such as knowledge of the shRNA target identity and the ability to systemically screen for phenotypes by imaging, it is limited by cost and labor. These limitations can be partially avoided by using a pooled screening approach that allows the simultaneous interrogation of tens of thousands shRNAs.

In the pooled screening approach, a library of lentiviral shRNAs is used to transduce cells at a low multiplicity of infection (MOI), usually less than 1.0, to avoid synthetic genetic interactions (Figure 1A). Furthermore, a large enough number of cells are transduced to ensure sufficient representation of the shRNA library. Details about the optimization of transducing cells with a pooled lentiviral library have been previously reviewed in detail [5]. Transduced cells are then selected using puromycin to generate a population of cells with stable shRNA knockdown (Figure 1A). After obtaining a transduced mammalian cell population stably harboring the pooled lentiviral shRNA library, they can be screened for certain cellular or molecular properties (Figure 1B). For example, cells can be fractionated based on membrane protein expression using cell sorting. However, one must exercise caution when cell dissociation reagents are used, which may undesirably cleave integral and peripheral membrane proteins of interest and lead to false results. The type and duration of dissociation reagent use should be carefully assessed and optimized in the experimental design prior to performing large-scale cell fractionation.

As shRNAs are designed to target a unique DNA sequence, they inherently act as unique barcodes that can be identified and quantified by microarray analysis or high-throughput sequencing. This is achieved by using primers to PCR amplify shRNAs from isolated genomic DNA (gDNA) of shRNA transduced cells, followed by appropriate processing for downstream quantitation analyses (Figure 1C) [5]. Regardless of the screening approach taken, individual shRNA abundances are compared to an appropriate reference population (e.g. untreated or control treated cells) to measure a relative fold change (Figure 1C). To determine significant relative fold changes of shRNA abundance between treated and reference samples, each shRNA can be scored using a variety of approaches. For example, individual shRNAs can be scored using a shRNA Activity Rank Profile (shARP) scoring approach [6]. Another perspective that can be used is to score shRNAs on the level of their gene targets, such as using the Gene Activity Ranking Profile (GARP) [6]. Other similar scoring approaches include the RNAi gene enrichment ranking (RIGER) [7] and the redundant siRNA activity (RSA) [8]. These scoring approaches may require adaptions to best serve the experimental design of the shRNA screen. Using
these or other scoring approaches, a cutoff can be determined for shRNAs considered as hits (Figure 1D). The genes corresponding to these shRNAs can be analysed using bioinformatics to determine enrichment or suppression of certain biological processes or pathways (Figure 1D).

Genes corresponding to shRNA hits require validation. Individual shRNA hits can be isolated from the library and independently validated in the experimental conditions used in the screen. This may allow for the simultaneous assessment of target gene knockdown efficacy of individual shRNA hits. However, this would not address possible off-target effects of individual shRNAs, which can be difficult to assess. To avoid possible off-target effects, gene hits can be validated using orthogonal RNAi technologies, such as siRNA or esiRNA. The identification and validation of gene hits should be performed prior to further investigation and downstream experiments.

IDENTIFYING GENES THAT REGULATE THE CANCER STEM CELL MARKER AC133 EPITOPE

The availability of RNAi libraries allows for the interrogation of large gene sets for their involvement in membrane protein biology. To illustrate the utility of this approach, we will discuss pooled lentiviral shRNA screening strategies that our research group has taken to identify genes involved in the regulation of CD133 on two levels: post-translational regulation and transcription. These strategies could be applied or adapted to study the regulation and function of other membrane proteins.

CD133 was first identified as the protein bearing the AC133 epitope [9], a cell-surface marker that has been used to select for stem cells and cancer stem cells from a broad range of tissue types [10]. In a number of cases, it has also been used to characterize certain cellular properties, such as chemoresistance. Given this, it may not be surprising that cell-surface AC133 and CD133 detection can be associated with a poor prognosis. Although CD133 detection may have clinical applications and its AC133 epitope has allowed characterization of subpopulations of cancer cells, little is known regarding its regulation and function.

CD133 is a single copy gene on chromosome 4 and its transcription is regulated by five alternative promoters [11]. These promoters have been demonstrated to be hypomethylated in certain CD133 expressing cancer cells and methylated at CpG islands in cells with low or undetectable CD133 levels.
CD133 promoter methylation is regulated, at least in part by the activation of TGFβ signaling, which results in the downregulation of the DNA methyltransferases (DNMT1 and DNMT3β) [15]. The RAS/ERK signaling pathway has also been implicated in the regulation of CD133 promoter, as its ETS2 and ELK1 are thought to bind to two ETS sites in the CD133 P5 promoter [16]. However, treatment of CD133-expressing cells with the MEK/ERK inhibitor U0126 resulted in CD133 downregulation in only certain cancer cell lines, suggesting that other factors that regulate CD133 expression remain unidentified.

Concerns about the specificity of CD133 as a marker of stem and progenitor cells were raised due to the detection of CD133 gene and protein expression in differentiated cell types [17–19]. This was partially addressed by the report that cell-surface recognition of the AC133 epitope was specific to stem and progenitor cells and becomes masked during cell differentiation while total CD133 protein levels could be maintained [20]. As a change in CD133 molecular weight was detected during cell differentiation, it was proposed that cell-surface AC133 expression may be regulated by post-translational modifications.

To identify genes that regulate CD133 on the post-translational level, we performed a pooled lentiviral shRNA screen in cells exogenously expressing CD133 from a constitutive promoter. We used a lentiviral-based system to stably introduce CD133 into human embryonic kidney 293 (HEK293) cells [21, 22], which do not endogenously express CD133. Importantly, we were able to isolate a subpopulation of cells that stably and homogenously expressed cell-surface AC133 (HEK293/AC133). The cell-surface detection of AC133 demonstrated proper expression and post-translational regulation of the CD133 membrane protein in HEK293 and served as a readout for our screen. This engineered cell line allowed us to reduce noise or background that would come from a cell line expressed CD133 from its endogenous promoter, e.g. genes that regulated CD133 transcription.

HEK293/AC133 cells were transduced with a pooled lentiviral shRNA library containing ~54,000 (54 K) unique shRNAs, designed to target a total of 11,248 human genes (~5 unique shRNAs designed to target each gene) [7, 22]. A sufficient number of cells were transduced with lentiviral shRNAs to achieve a 185-fold representation of the 54 K library. After allowing gene knockdown to occur, cells were stained with AC133 antibody conjugated to a fluorophore and analysed using fluorescent-activated cell sorting (FACS). FACS allowed us to isolate AC133low and AC133high fractions. The AC133low fraction can be defined as a complete or near complete loss of cell-surface AC133 staining (<2% of the population), whereas the AC133high fraction represented the remaining ~98% of the population that had maintained cell-surface AC133 expression. Similar to the transduction of cells with the lentiviral shRNA library, a sufficient number of cells were fractionated to achieve a 185-fold representation of the 54 K library. shRNA identities from both low and high AC133 fractions were deconvoluted using microarray analyses. A greater than or equal to 5-fold enrichment in signal intensity in the AC133low fraction versus the AC133high fraction constituted as a shRNA hit. We performed a secondary screen in an arrayed format using an orthogonal RNAi approach (esiRNAs) on the gene targets of randomly selected shRNA hits and found a statistically significant reduction of cell-surface AC133 expression in ~70% of randomly selected gene hits, demonstrating that pooled lentiviral shRNA screening is a powerful approach to identify genes involved in membrane protein regulation.

Bioinformatic analysis of the gene targets of shRNA hits found that the N-glycosylation pathway was significantly enriched. This finding was consistent with the hypothesis that cell-surface AC133 expression is regulated by post-translational modifications [20]. Indeed, the generation of N-glycan mutants, in vitro glycosidase experiments and inhibiting various N-glycosylation steps with small molecules demonstrated that N-glycan processing contributes to cell-surface AC133 expression [22]. It has been proposed that specific CD133 N-glycan processing allows for a certain CD133 conformation or tertiary structure that exposes AC133 at the cell-surface [20]. Taken together, a pooled lentiviral shRNA screening approach allowed us to identify genes and a biological process/pathway that regulates cell-surface AC133 expression, which help addressed a longstanding question in the cancer stem cell biology field.

**IDENTIFYING REGULATORS OF CD133 TRANSCRIPTION**

To identify genes that regulate CD133 transcription from its native promoter we wanted to perform a
pooled lentiviral shRNA screen in cells that endogenously express CD133. Like many other membrane proteins, the study of CD133 is challenging as its expression is generally limited to rare cell populations (e.g. certain stem and progenitor cells). However, there are several cancer cell lines that are nearly entirely CD133\(^+\) (e.g. the colon adenocarcinoma cell line Caco-2 and the acute lymphoblastic leukemia (ALL) cell line SEM-K2). Therefore, pooled shRNA screens on cells that endogenously express CD133 may be limited to specific cell lines.

We performed a screen in Caco-2 cells in parallel with the HEK293/AC133 screen [23]. Although cell-surface AC133 could be detected in nearly the entire Caco-2 population, it has been previously demonstrated that this AC133 levels are dynamic and are regulated by the cell cycle [24]. This may lead to increased noise and background as cells in certain parts of the cell cycle that are expressing low cell-surface AC133 levels may contaminate the AC133\(^{low}\) fraction, resulting in false positives. Therefore, in an attempt to reduce background noise and to identify genes that regulate CD133 transcription, we performed a comparative analysis between the HEK293/AC133 and the Caco-2 screens and focused on shRNA hits that were specific to the Caco-2 screen [23].

Bioinformatic analyses identified a statistically significant enrichment of genes involved in transcription. We validated all genes with known or predicted functions in transcription that corresponded to at least two shRNAs. Among the gene hits from the pooled lentiviral shRNA screen and that could be validated was the transcriptional activator AF4. Further experiments demonstrated that AF4 directly regulated CD133 transcription elongation across a number of cancer cell lines [23] and was most likely involved in a higher order protein complex involving the positive elongation factor b (PTEF-b) [25].

AF4 is the most commonly identified fusion with the mixed lineage leukemia (MLL) gene in pediatric ALL and is associated with a particularly dismal prognosis [26]. Others have demonstrated that MLL-AF4 directly regulates CD133 transcription in ALL [27, 26]. As the MLL-AF4 fusion alone can result in ALL, we had hypothesized that part of its transforming ability is through CD133. Interestingly, similar to MLL-AF4 knockdown [26], CD133 knockdown resulted in a dramatic decrease in colony forming units and apoptosis of ALL cells.

Our study demonstrated that the pooled lentiviral shRNA screening approach can identify genes that regulate membrane proteins. There are still numerous gene hits from both the HEK293/AC133 and Caco-2 screens that have been validated [22, 23] but require further investigation to appreciate their role in regulating CD133 expression and cancer biology. Furthermore, our screening strategies focused on genes that positively regulate cell-surface AC133 expression or CD133 transcription. One adaptation of our approach could be to enrich for cells that have increased membrane protein expression. This may allow for the identification of genes that act to negatively regulate or inhibit expression of certain membrane proteins.

**DISCUSSION**

We demonstrated that pooled lentiviral shRNA screening can be a powerful approach to identify genes involved in membrane protein regulation. However, a number of noteworthy limitations of this approach should be considered. Many shRNAs do not efficiently knockdown their gene target, while others result in off-target effects. Together, this can possibly result in false negatives and positives, respectively. Another consideration is that a number of genes that may regulate membrane proteins may also be essential for cell-survival.

shRNAs can display discrepancy in their efficiency to repress gene target depending on factors such as timing, MOI, integration location, cell type or culture condition, which can alter the shRNA level and the target gene expression level in the cell. Therefore, one should not completely ignore negative results and performing pooled lentiviral shRNA screens in multiple cell types and contexts that work toward the same objective may provide a more comprehensive answer to the biological question being asked. Similarly, validation of shRNA and gene hits should be performed in more than one cell line or cell type, if multiple ones were screens.

Representation of individual shRNAs within the pool are variable, some of which are at such low abundance they are within background levels when analysed by microarrays [28]. This may be rectified by using a more sensitive approach to quantify shRNAs, such as high-throughput sequencing. Another circumstance that may lead to false negatives is the approach of scoring shRNA hits. In fact, biologically significant hits may be below...
cut-offs and thresholds. For example, we recently demonstrated that histone deacetylase 6 (HDAC6) plays a crucial role in positively regulating CD133 protein at the plasma membrane [29]. HDAC6 was a gene hit that fell below our cutoff criteria in both the HEK293/AC133 and the Caco-2 screens. Although low abundance of HDAC6 shRNAs in the pool was not the case and could not necessarily explain the exclusion of HDAC6 as a gene hit, it is possible that a single lentiviral shRNA integration (resulting from a MOI ≤ 1.0, as described in the experimental approach) in a transduced cell results is insufficient HDAC6 knockdown to observe CD133 downregulation. Therefore, like HDAC6, there are likely shRNAs and genes that are reported as false negatives using this screening approach.

We used a 54 K pooled lentiviral shRNA library in the example studies in this review, which is designed to target 11 248 genes. Although this is a large collection of shRNAs, there remains a large number of human gene targets that would not be interrogated. Recently, the use of a pooled lentiviral shRNA library containing 78 432 shRNAs (80 K) designed to target 16 056 human genes has been employed by our laboratory to profile essential genes in breast, pancreatic and ovarian cancer cell lines [6]. It is noteworthy that although a larger lentiviral shRNA library may allow the interrogation of more genes, it does increase cost and labor. For example, tens of millions of cells are required for infection to maintain a certain fold representation of the library. Furthermore, this large number of cells needs to be maintained during the selection process, which in the illustrative examples in this review resulted in tens of hours of FACS. Nevertheless, pooled lentiviral shRNA screening of such a large number of shRNAs remains less laborious than an arrayed approach. However, to answer certain hypothesis driven or very specific questions relating to membrane protein biology, a customized pool of lentiviral shRNAs or selected individual shRNAs arranged in an array format may be a more efficient screening approach. This, of course, introduces bias to the screen.

The screening approach in our examples employed a fluorophore-conjugated antibody to detect cell-surface AC133 and properly localized CD133 protein. However, many membrane proteins do not have available antibodies that recognize their cell-surface epitopes. One option is to generate and characterize antibodies for a specific membrane protein prior to performing a screening approach, which can be developed in animals or by using synthetic antibody libraries. Another option is to generate a fusion protein between the membrane protein and a fluorescent marker. This would allow monitoring of protein expression during FACS analyses. However, this may create issues with specificity of shRNA hits toward the membrane protein versus the fluorescent marker or the fusion itself. Furthermore, the generation of a membrane protein fusion may result in abnormal membrane protein regulation and function. Therefore, we encourage performing validation of gene hits with the endogenous membrane protein too.

Membrane proteins have functions in a wide range of biological processes and pathways in the cell. Knowledge of these functions can allow selection of an appropriate readout. Membrane proteins that function as a receptor in a signaling pathway could rely on reporter genes, allowing cells to be fractionated or selected based on reporter gene expression (e.g. a TCF/LEF fluorescent reporter as a readout for Wnt/β-catenin signaling activity). Genes involved in the regulation of function of membrane proteins that act as an efflux pumps of certain drug(s) can be determined by performing the screen in the presence and absence of that drug(s) and looking for shRNA hits that result in chemosensitivity. For example, a genome-wide pooled shRNA screen in non-small cell lung cancer expressing an activating mutation of the epidermal growth factor receptor (EGFR) was performed to identify genes and pathways that contributed to the resistance of EGFR inhibitors [30]. Another example, cells transduced with a pooled lentiviral shRNA library were treated with or without FAS ligand to induce apoptosis to identify genes that would inhibit death receptor-mediated apoptosis [31]. The utility of pooled shRNA screens can be viewed from a perspective other than using membrane protein expression or function as a readout. Gene hits that encode a membrane protein from screens focused on studying cellular processes can also associate functions for those membrane proteins. For example, the ephrin receptor A7 was identified as a tumor suppressor from a pooled shRNA screen in follicular lymphoma [32]. These examples further illustrate the utility of RNAi screens to uncover membrane protein biology in areas that conventional experimental approaches have faced difficulties.
Key points
- siRNA screens can be used to study membrane protein regulation, structure and function.
- This screening approach relies on membrane protein expression or function as a readout to report the effects of gene knockdown by siRNA.
- We illustrate the utility of this approach by discussing studies that identified genes that regulate CD133 transcription and post-translational regulation.

Acknowledgements
We thank Dr T. Ketela and D. Kasmer for establishing and optimizing the pooled lentiviral shRNA screening approach. We also thank P.-A. Penttila for her consultation on cell sorting.

FUNDING
This work was supported by funds from the Stem Cell Network (to J.M.), the Canadian Institutes of Health Research (to J.M.) and the Ontario Institute for Cancer Research (to J.M.).

References


