ABBS

Review

A primer on using pooled shRNA libraries for functional genomic screens

Guang Hu[1]* and Ji Luo[2]*

[1]Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA
[2]Medical Oncology Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD 20892, USA
*Correspondence address. Tel: +1-919-541-4755; Fax: +1-919-541-0146; E-mail: hug4@niehs.nih.gov (G.H.)/Tel: +1-301-451-4725;
Fax: +1-301-480-8780; E-mail: ji.luo@nih.gov (J.L.)

The discovery of RNA interference (RNAi) has revolutionized genetic analysis in mammalian cells. Loss-of-function RNAi screens enable rapid, functional annotation of the genome. Of the various RNAi approaches, pooled shRNA libraries have received considerable attention because of their versatility. A number of genome-wide shRNA libraries have been constructed against the human and mouse genomes, and these libraries can be readily applied to a variety of screens to interrogate the function of human and mouse genes in an unbiased fashion. We provide an introduction to the technical aspects of using pooled shRNA libraries for genetic screens.

Introduction

Recent advances in RNA interference (RNAi) technologies have made it possible to interrogate the genetic dependencies of mammalian cells by loss-of-function screens on a genome-wide scale. RNAi screens can be carried out with either siRNA-based transient transfection or shRNA-based stable gene knockdown. Vector-based shRNA libraries have several unique advantages that make them particularly attractive: they can be screened in pools and this significantly reduces the cost of the screen; they afford long-term gene knockdown and thus can reveal slow phenotypic changes in the cell; they can be readily adapted for *in vivo* screens in mouse models. For these reasons we and others have developed resources and technologies for using pooled shRNA libraries in various screens. In this mini-review we provide a primer and a practical guide on how one could use pooled shRNA libraries to study cancer biology.

Library Construction

Several groups and companies have described the design and construction of genome-scale shRNA libraries (**Table 1**).

These libraries include the Netherlands Cancer Institute (NKI) libraries [1,15], the RNAi consortium (TRC) libraries [16,17], the Hannon–Elledge libraries [18–20], and the System Biosciences (SystemBio) libraries. They differ in size, coverage, shRNA sequence design, and, most importantly, they use different strategies to generate the miRNA-mimicking sequences for gene silencing. The NKI, TRC, and SystemBio libraries use RNA polymerase III (RNA Pol III) to express simple hairpin RNAs to mimic the pre-miRNAs for RNAi (for a review on miRNA biogenesis, please see [21]). The Hannon–Elledge libraries, on the other hand, use RNA polymerase II (RNA Pol II) to express hairpin RNAs in the context of a natural miRNA to mimic the pri-miRNA for RNAi. Owing to the differences, each of these libraries has unique characteristics that we discuss below.

The NKI shRNA library

The NKI shRNA libraries [1,15] use the RNA Pol III promoter H1 to express shRNAs that consist of a 19-nt double-stranded stem and a 9-nt loop. Once expressed in cells, the pre-miRNA-like shRNAs are processed into functional siRNAs that have 19-bp double-stranded RNA and 2-nt overhangs on each end. The libraries are generated in a mouse stem cell virus (MSCV)-based self-inactivating retroviral vector pRSC which also contains a puromycin selectable marker. The NKI shRNA library is the smallest among all with ~54,000 shRNAs in total. It targets ~8000 human genes with three shRNAs per gene and ~15,000 mouse genes with two shRNAs per gene.

The target-specific 19-nt sequences in the NKI libraries were selected using the following criteria: (i) target mRNA coding sequence; (ii) target all RefSeq transcript variants of the gene; (iii) no sequence overlap among shRNAs targeting the same gene; (iv) minimal sequence similarity to

Table 1 Properties of shRNA libraries discussed in this review

Library	NKI	TRC	SystemBio	Hannon−Elledge
Vector	pRSC	pLKO.1	pSIH-1-H1 pSIF-1-H1 pGreenPuro	pSMP pGIPZ pTRIPZ
shRNA promoter	H1 (Pol III)	U6 (Pol III)	H1 (Pol III)	MSCV-LTR, CMV, TRE2 (Pol II)
Mature siRNA size (bp)	19	21	27	22
Library coverage	54,000 shRNAs: 2−3 shRNAs/gene for 23,000 human and mouse genes	300,000 shRNAs: 5 shRNAs/ gene for 60,000 human and mouse genes	350,000 shRNAs: 4 shRNAs/gene for 90,000 human/mouse genes	160,000 shRNAs: 2−3 shRNAs/gene for 60,000 human and mouse genes
Sample pooled shRNA screens	[1−4]	[5,6]	[7−9]	[10−14]
Additional infomation	The NKI libraries is not commercially available, but the technical information is available at http://screeninc.nki.nl/	The TRC libraries are available from Thermo Scientific Openbiosystems: http://www.openbiosystems.com/ and from Sigma-Aldrich: http://www.sigmaaldrich.com/	The SystemBio libraries are available from System Biosciences: http://www.systembio.com/	The Hannon−Elledge libraries are available from Thermo Scientific Openbiosystems: http://www.openbiosystems.com/

other genes; (v) abide the thermodynamic asymmetry rules governing the incorporation of the correct RNA strand into the RNA-induced silencing complex (RISC); (vi) begin with a G or C after an AA dimer in the 5′-flanking sequence; (vii) end prior to a TT, TG, or GT doublet in the 3′-flanking sequence; (viii) no stretches of four or more T or A; (ix) 30%−70% G + C content; and (x) no *Eco*RI or *Xho*I sites. Because the NKI library was designed when knowledge of the miRNA biogenesis was relatively incomplete, these rules did not include all the known constraints for optimal design of shRNAs.

Each 19-nt sequence was synthesized as two complementary 60-nt oligos that contain the 19-nt stem region, a nine-base loop, a termination sequence, and *Bgl*II and *Hin*dIII sites. For library construction, pairs of complementary oligos were annealed and cloned into *Hin*dIII/ *Bgl*II-digested pRSC in 96-well plate format. The shRNA expression cassette containing the H1 promoter and the hairpin construct is flanked by *Eoc*RI and *Xho*I sites and can be shuttled into different vectors.

The TRC shRNA library
The TRC shRNA libraries [16,17] use the RNA Pol III promoter U6 to express shRNAs that contain a 21-nt double-stranded stem and a 6-nt loop. The pre-miRNA-like shRNAs are then processed into functional siRNAs in cells. The libraries were cloned in the lentiviral vector pLKO.1, which contains a puromycin selectable marker. The TRC shRNA library is the biggest among all, consisting of ∼300,000 shRNAs targeting ∼60,000 human and mouse genes with 5 shRNAs per gene.

The 21-nt sequences in the TRC libraries are selected with the following procedure. For each RefSeq transcript, all 21-nt sequences were generated *in silico* from 25 bp after the coding sequence start site to 150 bp before the end. Each sequence was scored using a set of bioinformatics criteria for knockdown potential and ease of synthesis. The top 100 sequences were BLASTed against NCBI RefSeq and Unigene databases for specificity such that each shRNA must have >3 mismatches to all other RefSeqs with at least two of the mismatches between positions 3 and 19. The top four 21-mers from the CDS and the best one from the 3′ UTR were selected for library construction.

For library construction, complementary 65-nt oligos were synthesized that contain the 21-nt shRNA stem region, a six-base loop with a *Xho*I site, a TTTTT termination sequence, and overhangs compatible with *Eco*RI and *Age*I sites. Annealed oligos were ligated into the *Eco*RI and *Age*I digested pLKO.1 vector in 96-well format and sequence verified.

The Hannon−Elledge shRNA libraries

The first generation of the Hannon−Elledge shRNA library used a similar design as the NKI and TRC libraries [19]. With the advance in the understanding of the miRNA biogenesis, this was replaced by second-generation libraries that adopt completely different design compared with other shRNA libraries [18,20]. The second-generation libraries use RNA Pol II or Pol III to express shRNAs in the context of a natural miRNA, miR-30 (125 bases 5′ and 3′ of the pri-miR-30 sequence). This strategy was shown to be up to 12 fold more efficient than pre-miRNA hairpin designs in producing mature siRNA, and was therefore more effective in gene silencing. More importantly, the use of RNA Pol II promoter greatly increases the flexibility of this library. For examples, promoters of different strengths can be used to achieve optimal RNAi in different cell types or tissues; inducible promoters can be used to control the timing and severity of RNAi; and reporter genes can be expressed concomitantly with the shRNA to label cells that received RNAi.

The second-generation Hannon−Elledge libraries contain ∼87,000 shRNAs against all human genes and ∼76,000 shRNAs against all mouse genes with ∼2−3 shRNAs per gene. These libraries exist several different vectors, including the retroviral vector pSM2, the MSCV-based retroviral vector pSMP (also named MSCV-PM), the lentiviral vector pGIPZ, and the inducible lentiviral vector pTRIPZ (discussed below). The 22-nt shRNA sequences were designed by Rosetta Inpharmatics (Kirkland, USA) using proprietary algorithms developed based on empirical testing of thousands of siRNAs. The algorithm also introduced additional positional biases and thermodynamic rules suggested by analysis of siRNA and endogenous miRNA incorporation into the RISC complex.

To synthesize the library, 97-nt shRNAmir templates containing the 22-nt shRNA stem region, a 15-nt miR-30 loop, and flanking 5′ and 3′ miR-30 sequences were synthesized on printed microarrays, cleaved off, and polymerase chain reaction (PCR)-amplified as a pool using universal primers. Pools of 22,000 shRNAmirs were inserted between XhoI and EcoRI sites in the pSM2 vector. The pSM2 shRNA plasmids were then entered into a high-throughput sequencing pipeline and clones with unique and correct shRNA sequences were retained individually in 96-well plates. For each pool, the sequencing progress was monitored dynamically and stopped when accumulations of new clones slowed down. At that point, new arrays were synthesized to produce additional shRNAmir pools. When the number of shRNA clones for a given gene exceeded three, no additional shRNAs for this gene were made. With this iterative synthesis, 87,283 verified human shRNA clones and 76,896 verified mouse shRNA clones have been produced so far. The libraries were re-cloned into different vectors mentioned above. These different vectors allowed the Hannon−Elledge libraries to be used for almost any kind of assays and thus made them the most flexible and versatile shRNA libraries currently available.

The SystemBio shRNA libraries

Similar to the NKI library, the SystemBio shRNA libraries also use the H1 RNA Pol III promoter to express pre-miRNA-like shRNAs. The shRNAs consist of a 27-nt double-stranded stem and a 12-nt loop. They were designed with a proprietary algorithm and cloned into the System Biosciences pSIH-1-H1, pSIF-1-H1, or pGreenPuro vectors. The SystemBio libraries only exist in pooled format: (i) human genome shRNA library: ∼200,000 shRNAs targeting ∼50,000 human genes; (ii) mouse genome library: ∼150,000 shRNAs targeting ∼40,000 mouse genes; (iii) human apoptosis library: 6876 shRNAs targeting 597 human apoptosis genes; (iv) human kinase library: 10,453 shRNAs targeting 897 human kinase genes; (v) human phosphatase library: 2719 shRNAs targeting 244 human phosphatase genes. These libraries have not been sequence validated.

Genome wide vs. focused libraries

Often for practical reasons it is desirable to screen only a subset of genes in the genome such as kinases, phosphatases, transcription factors, or epigenetic enzymes. The TRC, Hannon−Elledge, and SystemBio libraries are all available as 'gene family' or 'pathway' shRNA libraries. These 'focused' libraries are less complex in size and therefore allow more thorough and cost-effective screening. However, 'focused' libraries are inherently limited in their discovery potentials as only well-annotated genes are included in these libraries. Given a significant fraction of human genes are poorly annotated in terms of their biological functions, the main appeal for using genome-wide libraries is the discovery of novel functions for new genes.

Vector Technology

Over the years, shRNA vectors have evolved from the initial simple designs that can only accomplish gene silencing into more sophisticated versions that are tailored for advanced applications. Features such as easy delivery, inducibility, and trackable shRNA expression have been added to enable different assays. Here, we review several of these features found in current generation of the shRNA vectors, a summary of the properties of these vectors is provided in **Table 2**.

Viral shRNA vectors

Viral vectors, including retroviral, lentiviral, and adenoviral vectors, are now widely used as vehicles to deliver shRNAs to target cells because of their high delivery

Table 2 Properties of shRNA vectors discussed in this review

shRNA vectors	Promoter driving shRNA expression	Constitutive or inducible	Trackable shRNA expression	Virus type
pRSC	H1 (RNA Pol III)	Constitutive	No	Retrovirus
pLKO.1	U6 (RNA Pol III)	Constitutive	No	Lentivirus
pSMP	MSCV-LTR (RNA Pol II)	Constitutive	No	Retrovirus
pGIPZ	CMV (RNA Pol II)	Constitutive	Yes (GFP)	Lentivirus
pSuperior.retro	Tet-inducible H1 (RNA Pol III)	Inducible	No	Retrovirus
Tet-pLKO-Puro	Tet-inducible U6 (RNA Pol III)	Inducible (one-vector design)	No	Lentivirus
pLKO-puro-IPTG-1 × LacO/3 × LacO	IPTG-inducible-U6 (RNA Pol III)	Inducible (one-vector design)	No	Lentivirus
pPRIME	CMV, TRE, TREX (RNA Pol II)	Consititutive or Inducible	Yes	Lentivirus
pSLIK	Tet-inducible TRE2 (RNA-Pol II)	Inducible (one-vector design)	Yes (GFP)	Lentivirus
pTRIPZ	Tet-inducible TRE2 (RNA-Pol II)	Inducible (one-vector design)	Yes (turboRFP)	Lentivirus
pInducer	Tet-inducible TRE2 (RNA-Pol II)	Inducible (one-vector design; faster and easier isolation of cells with inducible knockdown)	Yes (turboRFP)	Lentivirus

efficiency. Retrovirus- and lentivirus-based vectors are most commonly used in basic research, because they can provide long-term gene silencing due to the integration of the shRNA expression unit into the host cell genome [22]. Adenovirus vector is more suited for *in vivo* experiments and therapeutic applications, because adenoviral DNA does not integrate into the genome and is not replicated during cell division [23].

All the above shRNA libraries use virus-based vectors. The NKI libraries (pRSC) and some versions of the Hannon–Elledge libraries (pSM2 and pSMP) used retroviral vectors. Since retroviruses only infect dividing cells but not non-dividing cells, these vectors raise less biosafety concerns. In addition, they can be used to selectively infect and label proliferating cells in a population, an important advantage for certain cancer and stem cell research. The TRC (pLKO.1), SystemBio, and other versions of the Hannon–Elledge libraries (pGIPZ and pTRIPZ) used lentiviral vectors. Lentiviral shRNA vectors have high transduction efficiency. More importantly, they transduce non-dividing cells such as neurons and other hard-to-transduce cells such as primary cells, thus greatly expanding the possibility of RNAi and RNAi screens [24].

Trackable shRNA expression
For many applications, such as *in vivo* animal studies, genetic mosaic studies, and lineage tracing studies, it is necessary to identify the subset of cells that underwent gene

knockdown within a complex population. Trackable shRNA expression is thus a feature particularly valuable for these applications. Because the Hannon–Elledge shRNA libraries use RNA Pol II to drive the shRNA expression, a reporter gene such as a drug-selection or a fluorescence marker can be directly placed in front of the Mir30-shRNA cassette. Both the reporter gene and the shRNA are expressed as a chimera RNA from the same promoter, which is cleaved to generate the reporter mRNA and the shRNA. The reporter can thus faithfully track shRNA expression quantitatively. This feature enables one to select for cells with the desired level of knockdown. The constitutive pPRIME vectors were the first to adopt this strategy [25]. They use the RNA Pol II promoter CMV (cytomegalovirus) to express green fluorescent protein (GFP) or a drug-selection maker followed by miR30-shRNA. The pGIPZ vector (Open Biosystems, Huntsville, USA) was modeled after the pPRIME vectors and uses CMV to drive the expression cassette consisting of turbo GFP followed by internal ribosome entry site (IRES) followed by puromycin-resistance gene cassette and the miR30-shRNA.

The same strategy cannot be applied to RNA Pol III-based shRNA systems. As an approximation, RNA Pol III-based shRNA vectors often express drug-resistance markers or fluorescence reporters from a separate Pol II promoter. Since the shRNA and the reporter are expressed independently, cells can only be selected based

on their transduction status and not on the expression level of shRNA.

Inducible vectors

Inducible shRNA vectors have several advantages over constitutive shRNA vectors in many experimental settings. For example, they provide temporal and reversible control of gene expression; they allow for the study of essential genes where stable knockdown can lead to lethality; they minimize experimental variations due to clonal heterogeneity within an experiment. Different inducible systems have been developed to accommodate different shRNA library designs.

Inducible vectors compatible with the NKL shRNA library pSuperior.retro. The tetracycline-inducible vector pSuperior.retro (Oligoengine, Seattle, USA) is fully compatible with the NKL shRNA library. In the pSuperior.retro vector, a tetracycline operator 2 (TetO2) site was inserted at the end of the H1 promoter and the RNA hairpin transcription start site. The TetO2 sequence serves as the binding site for the Tet repressor (TetR). In the absence of tetracycline, TetR binds to the TetO2 sequence and represses transcription of the shRNA. Doxycycline binding releases TetR from the Tet operator and the transcription of the shRNA can proceed. As pSuperior.retro itself does not encode TetR, it requires cell lines that already express TetR.

Inducible vectors compatible with the TRC shRNA library Tet-pLKO-Puro. The tetracycline-inducible vector Tet-pLKO-Puro is derived from pLKO.1 and is therefore fully compatible with the TRC shRNA library [26]. It uses a similar strategy for inducible shRNA expression as the pSuperior.retro vector by replacing the constitutive U6 polymerase III promoter with a Tet-inducible H1 promoter. Importantly, it also expresses TetR and the puromycin-resistance gene from a constitutive promoter PGK. Thus, Tet-pLKO-Puro allows inducible expression of shRNAs from a single vector and obviates the need to separately generate TetR-expressing cell lines.

pLKO-puro-IPTG-1xLacO and pLKO-puro-IPTG-3xLacO. The IPTG-inducible vector pLKO-puro-IPTG-1xLacO and pLKO-puro-IPTG-3xLacO (Sigma-Aldrich, St Louis, USA) are derived from pLKO.1 by inserting the Lac operator (LacO) at the end of the U6 promoter and including the Lac repressor (LacI) in the PGK-Puro cassette. In the absence of isopropyl-β-D-thio-galactoside (IPTG, an analog of lactose), LacI binds to LacO and prevent expression of the shRNA. In the presence of IPTG, LacI is released from lacO and the expression of the shRNA is allowed. Similar to Tet-pLKO-Puro, pLKO-puro-IPTG-LacO allows inducible gene silencing from a single vector.

Inducible vectors compatible with Hannon–Elledge shRNA library pPRIME and pGIPZ. pPRIME is a series of constitutive and inducible lentiviral vectors for high penetrance and trackable gene silencing in the Hannon–Elledge shRNA system [25]. The inducible pPRIME vectors use tetracycline-inducible promoters TRE (Clontech, Mountain View, USA) or TREX (Invigrogen, Carlsbad, USA) to express the GFP fluorescence reporter gene and the miR30-shRNA on a single transcript. They allow for tetracycline- or doxycycline-regulated RNAi and the simultaneous tracking of shRNA expression with GFP. Furthermore, these vectors provide high-penetrance gene silencing at single copy, ensuring the use of bar-coding strategies to deconvolve large pools of shRNAs.

pSLIK and pTRIPZ. pSLIK has a similar design to the tet-inducible pPRIME vectors [27]. Like pPRIME, it uses TRE to express GFP and miR30-shRNA. In addition, it uses a separate, constitutive Ubc promoter to express the reverse tetracycline transactivator (rtTA) linked to a drug resistance marker by IRES. With these elements, pSLIK can achieve trackable and inducible gene silencing in a single vector. Furthermore, pSLIK also allows for simultaneous silencing of multiple genes if multiple miR30-shRNA cassettes are concatenated. The pTRIPZ vector is modeled after pSLIK. It uses TRE promoter to drive the expression of turbo RFP (tRFP) followed by the miR30-shRNA, and it uses the Ubc promoter to drive the expression of an rtTA-IRES-Puro cassette. The pTRIPZ vector allows for tightly regulated shRNA expression in a variety of cells and cell lines.

pInducer. the pInducer vectors enables the tracking of shRNA induction in mammalian cells both *in vitro* and *in vivo* [28]. They use the TRE promoter to express tRFP or luciferase followed by the miR30-shRNA. They use a separate, constitutive promoter (Ubc or EF1a) to drive the expression of rtTA together with a drug resistance marker, GFP, or luciferase (depending on the application). By fluorescence-activated cell sort, high rtTA expressing cells can be isolated, which allows uniform temporal, dose-dependent, and reversible control of gene expression without lengthy drug selection. With these features, the pInducer vectors are especially suited for *in vivo* screens in animal models.

Assay Design

Screening format

Whereas siRNA libraries can only be screened in a well-by-well format, shRNA libraries can be screened either in a well-by-well format or in a pooled format. Well-by-well screen has the advantage of being more thorough as each shRNA is interrogated individually. A unique

advantage of well-by-well screen is that it is amenable to high-content image-based screenings such as those studying morphological changes, protein localization, or protein phosphorylation [16,29–31]. A major disadvantage of well-by-well screen is the high cost of robotics, consumables, and technical support required to carry out such screens. On the contrary, pooled shRNA screens are low cost, flexible, and can be carried out under standard laboratory settings without robotics. With the rapid democratization of low-cost next-gen sequencing technologies, the only 'high-tech' part of pooled shRNA screen—library deconvolution—is no longer a hurdle. Below we provide considerations on assay design for pooled shRNA screen.

Types of assay

Like all high-throughput screens, designing a robust assay is the key to a successful shRNA screen. Several assay designs have been implemented with pooled shRNA libraries. These include cell viability assay, reporter assay, and morphological assay (**Fig. 1**).

Cell viability is among most straightforward assays for pooled shRNA library screen [10,32]. The unique advantage of shRNA is that it affords stable gene knockdown, and thus even small effect on cell growth can be detected by extending the assay duration. For example, if the knockdown of a gene decreases the growth rate of cells by 10% per cell cycle in a cell line that doubles daily, a 5-day assay would yield a ~40% decrease in cell viability, whereas a 10-day assay yield a ~65% decrease, and a 15-day assay ~80% decrease. Analogous principle applies to shRNAs that increase cell proliferation. This advantage is of practical value because current shRNA designs rarely afford complete gene knockdown, thus one is often screening for hypomorphic phenotypes. Extending assay duration would allow one to compensate for modest effect on cell viability due to partial gene knockdown. In other words, extending assay duration increases the detection threshold of the screen.

Viability screens have gained particular popularity in identifying the genetic vulnerabilities of cancer cells. We and others have conducted pooled shRNA screens in panels of cancer and normal cell lines to identify genes that are more essential in cancer cells than in normal cells [5,6,10,33]. We and others have also applied this approach to study synthetic lethal interactions with oncogene such as KRAS [11,34–36]. In addition, pooled shRNA library screen can be combined with drug treatment to identify synthetic lethal interactions with small-molecule inhibitors [12,37,38]. Conversely, viability assay can be used for positive selection. For examples, enrichment assays can be set up to identify shRNAs that either increase cell proliferation [1], prevent cellular senescence [2,13], or confer resistance to a small-molecule inhibitor [3,4].

Pooled shRNA library screen is amenable to reporter assays based on the use of GFP and other trackable reporters. For example, GFP can be used either alone as a transcription reporter or as a fusion protein to report protein stability. To identify shRNAs that either turn the reporter 'on' or 'off', cells are sorted into GFP-high or GFP-low bins, and the composition of shRNAs in each bin is deconvoled separately to identify those shRNAs that have become enriched or depleted in each bin (**Fig. 1**).

Pooled shRNA library is also amenable for assays involving changes in cell behavior. For example, genes that control anchorage-independent growth of cells can be identified by introducing an shRNA library into immortalized epithelial cells that cannot grow in soft agarose. Those shRNAs that enable cells to form colonies in soft agarose are enriched in the screen. This approach has successfully indentified several tumor-suppressor genes [14]. shRNAs that regulate cell motility can be identified by subject shRNA library-infected cells to migration and invasion assay in order to enrich for shRNAs that either promote or impede these processes [39] (**Fig. 1**).

It is important to note that many of the aforementioned assays can be carried out *in vivo*. This is a unique advantage of shRNA over siRNAs. *In vivo* enrichment and drop-out screens have been successfully used to identify both candidate tumor-suppressor genes and cancer lethal genes [7,40–42]. *In vivo* screens tend to be noisier as the behavior of cells (often in the context of a xenograft tumor) is more variable. Thus smaller pool size and higher representation might be employed to increase the signal-to-noise ratio (see below).

Assay optimization

A critical consideration in assay design is its robustness, which is determined by both the signal-to-noise ratio and the dynamic range of the assay. This is often quantified in z-scores that measure how far the positive controls deviate from the background mean [43]. Achieving high signal-to-noise ratio and large dynamic range are critical for the success of a screen, and thus it is always beneficial to optimize assay conditions. Furthermore, for pooled shRNA screens it is essential that individual cells behave *uniformly* in the assay such that clonal variation in the starting population will not confound the effect of the shRNA library. To illustrate how heterogeneity in population behavior could drastically affect the outcome of a screen, consider a screen design where one wishes to identify shRNAs that cause resistance to a small-molecule inhibitor. It was determined by standard viability assay that the inhibitor should be applied at its IC_{90} concentration to allow the selection of resistant shRNAs on a relatively clean background. This assumption is true if at the IC_{90} all cells' viability is reduced by 90% uniformly. However, if a small fraction of cells in the
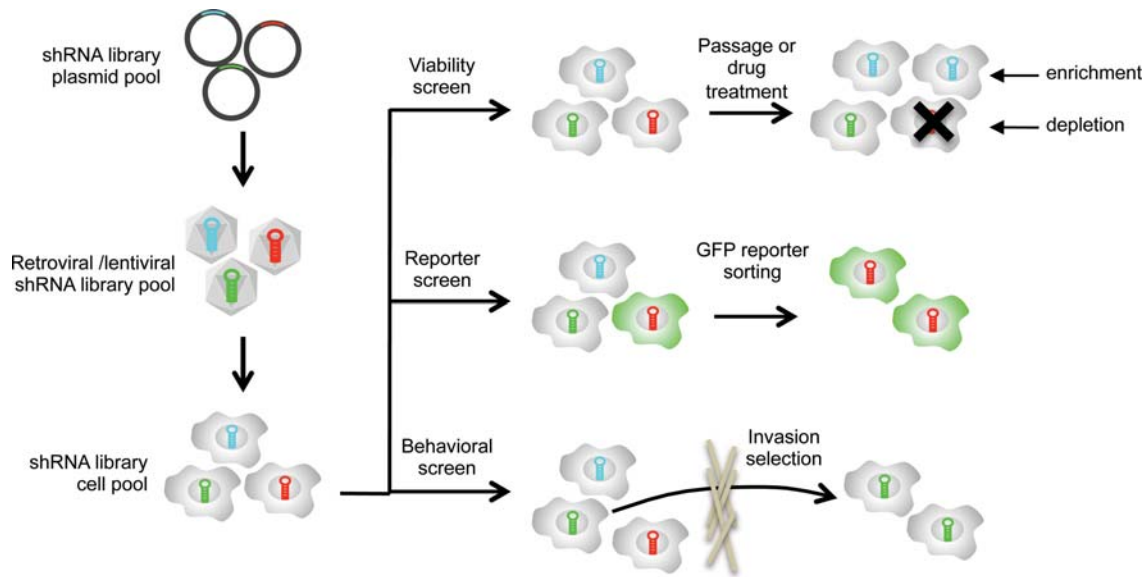
Figure 1 Examples of pooled shRNA screen design   Any assay that enables the selection of a sub-population of cells from the pool is amenable to pooled shRNA screen. Those shRNAs that altered the cell's behavior will be enriched or depleted from the pool. They can be identified by comparing the compositions of the pool before and after the selection.

population are intrinsically more resistant to the inhibitor, these cells will contribute disproportionally to the 10% survivors. Thus, any shRNAs that are delivered into the resistant sub-population will be selected as false positives. This scenario is in fact not uncommon: in a cancer cell line the dose–response curves of many drugs often does not reach zero viability even at high concentrations. The inherent genomic instability in cancer cells often account for such behavior.

In addition to using positive controls for assay optimization, assay robustness can be empirically tested using a negative control shRNA pool. Such pool typically consists of several hundred shRNAs targeting assay-irrelevant genes such as luciferase or GFP. By putting this pool through the screen, background variation introduced by the selection process can be quantified and the assay conditions can be adjusted accordingly.

As mentioned above, shRNAs have the unique advantage of affording stable gene knockdown to allow a longer assay period. However, the duration of the assay should be carefully optimized to avoid over-selection. In a viability assay for growth-enhancing shRNAs, too long a selection could result in the strongest shRNAs over-taking the pool and thus 'pushing out' less strong hits, leading to increased false-negative rate. Conversely, in a viability assay for growth inhibitory shRNAs, once *bona fide* toxic shRNAs have dropped out from the pool, extending the assay any longer will not provide additional dynamic range but rather introduce additional noise.

Library representation

In pooled shRNA screen, the effect of a given shRNA is assayed through the aggregate behavior of all cells carrying the shRNA. In a typical screen, each individual cell will, on average, receive a single shRNA viral particle and thus express one shRNA only. This is achieved through controlling the multiplicity of infection of the cells. Widely variable infection efficiencies are especially common among primary cells, thus each cell line should be tested empirically for its infection efficiency by the library vector.

The expression of an shRNA in a given cell is influenced by its site of integration and the nearby host genetic elements. This problem can be mitigated by generating a sufficiently high number of independent infections (i.e. the representation rate) for each shRNA in the library. A high representation rate for each shRNA is also necessary to prevent its spurious loss from the pool. For these two reasons, we recommend maintaining the representation of an shRNA at 500–2000 [10]. As the dynamic ranges for drop-out assays tend to be lower than enrichment assays, we recommend higher representations (1000–2000) for drop-out assays. Such representation must be carefully maintained throughout all steps in the screen from shRNA delivery to shRNA recovery. Any single step in the screen that significantly reduces the representation of the library would bottle-neck the pool and introduces additional, random fluctuations in library composition and thus contributes to the noise in the screen.

Library Deconvolution and Hits Validation

To identify shRNAs that confer the desired phenotype in the library, the composition of shRNAs in the pool at the end of the assay must be compared with the composition of the pool at the beginning of the assay. For viability

screen, this could simply be comparison of pool composition immediately after infection and at a later time point. For GFP reporter assays, for example, this would be comparison of pool composition between the total population and the desired GFP+ or GFP− sub-populations.

To deconvolve the shRNA library, all shRNA integrants are PCR-recovered as a single mixture using vector-backbone directed universal primers from genomic DNA extracted from the pool of cells [10]. The abundance of each shRNA in the mixture can be measured either by custom barcode microarrays or by next-gen sequencing. For microarray deconvolution, the principle is similar to two-color gene-expression microarray: two samples to be compared are labeled with different colors and hybridize to a 'barcode' microarray containing probes that are specific to each shRNA in the library. Such probes can be designed in two ways: 'half-hairpin' barcode probes are complementary sequences to one-half of the shRNA sequence. As shRNA sequences are primarily selected by their knockdown efficiency, half-hairpin barcode probes have non-uniform hybridization properties (due to variable GC content) and have a higher failure rate. A second method involves the use of a dedicated 'barcode' for each shRNA vector that are specifically optimized for hybridization properties [44]. Recently, barcode microarrays have been replaced by next-gen sequencing as the cost of these two platforms became comparable. Thus, after PCR recovery, the shRNA composition of the library can be directly sequenced and counted. The main advantages of sequencing over barcode microarray are that it eliminates the concerns of probe-cross-hybridization and it does not suffer from dynamic range compression at extreme levels of shRNAs. When using next-gen sequencing to deconvolve the library, we recommend the coverage to be 500–2000 fold.

Standard statistical analyses for high-throughput screens can be readily adapted for analyzing pooled shRNA screening data [43]. One major issue with existing RNAi libraries (shRNAs and siRNAs) is their incomplete penetrance, i.e. many shRNAs/siRNAs in the library fail to achieve sufficient knockdown of the target gene to give a phenotype. This is a particularly salient problem for proteins with enzymatic activities such as kinases and phosphatases. Consequently, current shRNA/siRNA screens are far from saturation and one cannot interpret negative results in an RNAi screen. As a result, similar screens carried out with different shRNA libraries in different cell lines sometimes yield different hits. Library penetrance can be somewhat improved by using a deeper libraries with more shRNA per gene to improve the chance of having at least one potent shRNA for each gene. Specialized analysis methods such as RIGER [5] have been developed to account for varying shRNA knockdown efficiency in the library. Another practical problem with low library penetrance is that often the majority of hit genes have only a single

shRNA scoring in the screen, making it difficult to judge the extent of off-target effects. A solution to the problem is to look for enrichment of shRNAs targeting genes in a common molecular pathway. A more fundamental solution to the penetrance problem is to generate a knockdown-validated library containing only shRNAs with known knockdown efficiency. Traditionally the construction of such library would be both costly and slow, as many shRNAs must be tested for each gene using quantitative PCR. Recently, an shRNA 'Sensor' method has been developed to rapidly interrogate hundreds of shRNAs for each gene to identify the most potent ones [45]. Thus, we expect that the next generation of shRNA libraries would both be deep (>10 shRNAs/gene) and knockdown validated (>70% depletion of mRNA). This will significantly improve the penetrance of shRNA screens to allow better saturation.

Genome-wide shRNA screens often yield hundreds of hits even after careful bioinformatics filtering. Secondary assays therefore must be devised to further validate and prioritize candidate genes for functional characterization. Careful re-testing of hit shRNAs using the primary screen assay is a necessary first step. The design of secondary assays should fulfill the following goals:

(1) *Improve on-target probability of candidate gene*: additional shRNA or siRNA sequence should be included in the secondary screen. This can be done either with a well-by-well format or with a new pooled library. In this latter approach, a new shRNA library can be synthesized to target only the candidate genes—as the gene list is now relatively small, the library could be made substantially deeper (e.g. 10 fold more shRNAs per gene). A major advantage of this approach is the rapidity with which hundreds of candidate genes can be tested.

(2) *Rule out cell line-specific effects*: it is often cost-prohibitive to conduct genome-wide screen in multiple cell lines. Thus, the same assay should be repeated with the primary screen hits in other relevant cell lines to ensure that the phenotype is generally reproducible.

(3) *Broaden phenotypic readout*: As with all screens, it is important to test candidate genes using a counter screen at this stage to rule out primary screen artifacts. Furthermore, in most cases the question to be addressed in the primary screen can in fact be studied using multiple assays. With fewer shRNAs to test and lower throughput, diverse secondary assays are often amenable at this stage to allow functional categorization of hits. For example, hits from a cell viability screen can be interrogated using cell cycle, apoptosis or senescence assay in the secondary screen to elucidate their biological functions.

Having multiple, independent shRNAs scoring for a gene is insufficient to warrant an on-target hit. Nevertheless, a strong correlation between the severity of phenotype and the degree of knockdown across multiple shRNAs indicates that the phenotype is unlikely to be an off-target effect. Ultimately, rescue experiments using shRNA-resistant cDNA are necessary to rule out off-target effects. If an shRNA targets the 3′-UTR of the gene, simply expressing a cDNA without the 3′-UTR should rescue the shRNA's phenotype. If an shRNA targets the coding region of a gene, expression of a cDNA-containing synonymous mutations in the target sequence should rescue. Importantly, rescue experiments should be performed to validate every phenotype of the shRNA to ensure all observed effects are on-target. In practice, rescue experiments could prove difficult for a gene whose dosage is critical for normal cellular function, thus both loss of expression and over-expression could be toxic to the cell. In this scenario using an inducible vector should allow tunable expression of the cDNA to match the endogenous protein expression level.

Concluding Remarks

RNAi is a powerful approach to provide functionally annotate of mammalian genome. Like many other recent 'omics' approaches such as gene expression, whole-genome sequencing, ChIP-seq, epigenomics, interactome, proteomics, global protein stability and localization analysis, and metabolomics, the major leap forward with RNAi screen is the ability to parallel a biological assay to a degree that all genes can be interrogated in an unbiased fashion. However, whereas the aforementioned 'omics' approaches provide information on the state of a gene or protein (mutated or not mutated, expressed or not expressed, bound to certain other proteins, phosphorylated on certain sites. etc.), RNAi screens yield functional information for a gene. With continuous advancement in technology, more and more functional assays should become amenable to RNAi screen, thus making it possible to richly annotate the many mammalian genes that are poorly studied.

Acknowledgement

Funding

References

1 Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M and Kerkhoven RM, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature 2004, 428: 431–437.

2 Nicke B, Bastien J, Khanna SJ, Warne PH, Cowling V, Cook SJ and Peters G, et al. Involvement of MINK, a Ste20 family kinase, in Ras oncogene-induced growth arrest in human ovarian surface epithelial cells. Mol Cell 2005, 20: 673–685.

3 Berns K, Horlings HM, Hennessy BT, Madiredjo M, Hijmans EM, Beelen K and Linn SC, et al. A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. Cancer Cell 2007, 12: 395–402.

4 Brummelkamp TR, Fabius AW, Mullenders J, Madiredjo M, Velds A, Kerkhoven RM and Bernards R, et al. An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors. Nat Chem Biol 2006, 2: 202–206.

5 Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X and Hinkle G, et al. Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci USA 2008, 105: 20380–20385.

6 Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA and East A, et al. Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. Proc Natl Acad Sci USA 2011, 108: 12372–12377.

7 Gumireddy K, Li A, Gimotty PA, Klein-Szanto AJ, Showe LC, Katsaros D and Coukos G, et al. KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. Nat Cell Biol 2009, 11: 1297–1304.

8 Yeung ML, Houzet L, Yedavalli VS and 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: 19463–19473.

9 Hattori H, Zhang X, Jia Y, Subramanian KK, Jo H, Loison F and Newburger PE, et al. RNAi screen identifies UBE2D3 as a mediator of all-trans retinoic acid-induced cell growth arrest in human acute promyelocytic NB4 cells. Blood 2007, 110: 640–650.

10 Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ and Zhao Z, et al. Cancer proliferation gene discovery through functional genomics. Science 2008, 319: 620–624.

11 Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF and Wong KK, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 2009, 137: 835–848.

12 Smogorzewska A, Desetty R, Saito TT, Schlabach M, Lach FP, Sowa ME and Clark AB, et al. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. Mol Cell 2010, 39: 36–47.

13 Burrows AE, Smogorzewska A and Elledge SJ. Polybromo-associated BRG1-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. Proc Natl Acad Sci USA 2011, 107: 14280–14285.

14 Westbrook TF, Martin ES, Schlabach MR, Leng Y, Liang AC, Feng B and Zhao JJ, et al. A genetic screen for candidate tumor suppressors identifies REST. Cell 2005, 121: 837–848.

15 Bernards R, Brummelkamp TR and Beijersbergen RL. shRNA libraries and their use in cancer genetics. Nat Methods 2006, 3: 701–706.

16 Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G and Piqani B, et al. A lentiviral RNAi library for human and mouse genes

applied to an arrayed viral high-content screen. Cell 2006, 124: 1283–1298.

17 Root DE, Hacohen N, Hahn WC, Lander ES and Sabatini DM. Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat Methods 2006, 3: 715–719.

18 Chang K, Elledge SJ and Hannon GJ. Lessons from nature: microRNA-based shRNA libraries. Nat Methods 2006, 3: 707–714.

19 Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S and Balija V, et al. A resource for large-scale RNA-interference-based screens in mammals. Nature 2004, 428: 427–431.

20 Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ and Siolas D, et al. Second-generation shRNA libraries covering the mouse and human genomes. Nat Genet 2005, 37: 1281–1288.

21 Ghildiyal M and Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet 2009, 10: 94–108.

22 Cullen BR. Induction of stable RNA interference in mammalian cells. Gene Ther 2006, 13: 503–508.

23 Grimm D, Pandey K and Kay MA. Adeno-associated virus vectors for short hairpin RNA expression. Methods Enzymol 2005, 392: 381–405.

24 Singer O and Verma IM. Applications of lentiviral vectors for shRNA delivery and transgenesis. Curr Gene Ther 2008, 8: 483–488.

25 Stegmeier F, Hu G, Rickles RJ, Hannon GJ and Elledge SJ. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. Proc Natl Acad Sci USA 2005, 102: 13212–13217.

26 Wiederschain D, Wee S, Chen L, Loo A, Yang G, Huang A and Chen Y, et al. Single-vector inducible lentiviral RNAi system for oncology target validation. Cell Cycle 2009, 8: 498–504.

27 Shin KJ, Wall EA, Zavzavadjian JR, Santat LA, Liu J, Hwang JI and Rebres R, et al. A single lentiviral vector platform for microRNA-based conditional RNA interference and coordinated transgene expression. Proc Natl Acad Sci USA 2006, 103: 13759–13764.

28 Meerbrey KL, Hu G, Kessler JD, Roarty K, Li MZ, Fang JE and Herschkowitz JI, et al. The pINDUCER lentiviral toolkit for inducible RNA interference in vitro and in vivo. Proc Natl Acad Sci USA 2011, 108: 3665–3670.

29 Gwack Y, Sharma S, Nardone J, Tanasa B, Iuga A, Srikanth S and Okamura H, et al. A genome-wide Drosophila RNAi screen identifies DYRK-family kinases as regulators of NFAT. Nature 2006, 441: 646–650.

30 Schmitz MH, Held M, Janssens V, Hutchins JR, Hudecz O, Ivanova E and Goris J, et al. Live-cell imaging RNAi screen identifies PP2A-B55alpha and importin-beta1 as key mitotic exit regulators in human cells. Nat Cell Biol 2010, 12: 886–893.

31 Friedman A and Perrimon N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature 2006, 444: 230–234.

32 Ngo VN, Davis RE, Lamy L, Yu X, Zhao H, Lenz G and Lam LT, et al. A loss-of-function RNA interference screen for molecular targets in cancer. Nature 2006, 441: 106–110.

33 Shaffer AL, Emre NC, Lamy L, Ngo VN, Wright G, Xiao W and Powell J, et al. IRF4 addiction in multiple myeloma. Nature 2008, 454: 226–231.

34 Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF and Schinzel AC, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009, 462: 108–112.

35 Scholl C, Frohling S, Dunn IF, Schinzel AC, Barbie DA, Kim SY and Silver SJ, et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. Cell 2009, 137: 821–834.

36 Vicent S, Chen R, Sayles LC, Lin C, Walker RG, Gillespie AK and Subramanian A, et al. Wilms tumor 1 (WT1) regulates KRAS-driven oncogenesis and senescence in mouse and human models. J Clin Invest 2010, 120: 3940–3952.

37 O'Connell BC, Adamson B, Lydeard JR, Sowa ME, Ciccia A, Bredemeyer AL and Schlabach M, et al. A genome-wide camptothecin sensitivity screen identifies a mammalian MMS22L–NFKBIL2 complex required for genomic stability. Mol Cell 2010, 40: 645–657.

38 Bivona TG, Hieronymus H, Parker J, Chang K, Taron M, Rosell R and Moonsamy P, et al. FAS and NF-kappaB signalling modulate dependence of lung cancers on mutant EGFR. Nature 2011, 471: 523–526.

39 Smolen GA, Zhang J, Zubrowski MJ, Edelman EJ, Luo B, Yu M and Ng LW, et al. A genome-wide RNAi screen identifies multiple RSK-dependent regulators of cell migration. Genes Dev 2010, 24: 2654–2665.

40 Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K and Sethumadhavan S, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature 2011, 476: 346–350.

41 Zender L, Xue W, Zuber J, Semighini CP, Krasnitz A, Ma B and Zender P, et al. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. Cell 2008, 135: 852–864.

42 Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison EA and Magoon D, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. Nature 2011, 478: 524–528.

43 Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E and Santoyo-Lopez J, et al. Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods 2009, 6: 569–575.

44 Xu Q, Schlabach MR, Hannon GJ and Elledge SJ. Design of 240,000 orthogonal 25mer DNA barcode probes. Proc Natl Acad Sci USA 2009, 106: 2289–2294.

45 Fellmann C, Zuber J, McJunkin K, Chang K, Malone CD, Dickins RA and Xu Q, et al. Functional identification of optimized RNAi triggers using a massively parallel sensor assay. Mol Cell 2011, 41: 733–746.