Biochemical and Biophysical Research Communications 430 (2013) 987–992

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

A rational design of completely random shRNA library

Guodong Yang [a,*], Lijun Yuan [b,1], Xiaozhao Lu [a,1], Zifan Lu [a,c,*], Libo Yao [a,*]

[a] The State Key Laboratory of Cancer Biology, Department of Biochemistry and Molecular Biology, The Fourth Military Medical University, Xi'an 710032, PR China
[b] Tangdu Hospital, The Fourth Military Medical University, Xi'an 710038, PR China
[c] The State Key Laboratory of Cancer Biology, Department of Pharmacogenomics, School of Pharmacy, The Fourth Military Medical University, Xi'an 710032, China

ARTICLE INFO

Article history:
Received 17 November 2012
Available online 19 December 2012

Keywords:
shRNA library
Random
Vector construction

ABSTRACT

It has been well established that the shRNA library has a significant advantage for screening the important genes involved in the interested biological pathways. Currently, the available libraries mainly target the known protein genes in human and mouse. With the expanding roles of lncRNA in biology, there is a great demand to design shRNAs targeting these non-coding RNAs. In this regard, a completely random shRNA library targeting all the genes with known or unknown sequences is of priority. Here we provide a practical workflow for construction of such a random shRNA library. In the novel shRNA library, there are about tens of different shRNAs targeting one gene, and thus significantly avoids the off-target effects.

1. Introduction

Loss-of-function of the interested genes is essential for exploring their biological functions and the underlying mechanisms [1]. The discovery of RNA interference (RNAi) makes the loss-of-function practical and easily realized. RNA interference (RNAi) is a phenomenon in which small dsRNA induces sequence-specific gene-silencing of homologous mRNA. The dsRNA could be either synthesized chemically or transcribed from the shRNA (short hairpin RNA) expressing vectors. The siRNAs enter the RNA-induced silencing complex (RISC), and the antisense strand guides the RISC complex to its target for mRNA degradation [2–4].

Gene loss-of-function mediated by shRNA has helped us understand the roles of interested genes comprehensively. Besides, large-scale shRNA library targeting the entire genome is also undergoing development for large scale screen of the genes involved in certain biological processes. Currently, there are some versions of both synthetic RNAi duplex- and vector-based libraries, and several groups have reported great success in using these tools in genetic screens [5–10]. Most importantly, the RNAi consortium (TRC) has produced the human and mouse libraries, which currently contain >135,000 lentiviral clones targeting 27,000 genes. Initial screening efforts have demonstrated that these libraries and methods are practical and powerful tools for high-throughput lentivirus RNAi screens [11]. Although such screens have provided rich and unappreciated insights into many biological processes, there are still some challenges. For example, the currently used lentivirus based shRNA libraries mainly target the known protein coding genes in mammalian cells [11,12], limiting their implication in screening the important non-coding RNAs.

In the past decades, multiple lncRNAs have been identified, such as Xist [13] and H19 [14], opening an era of lncRNA biology. With the advent of advanced sequencing technologies and findings from large-scale consortia focused on characterizing functional genomic elements, such as ENCODE (encyclopedia of DNA elements), more and more lncRNA are being identified and awaited for functional validation. According to the recent data by ENCODE Project Consortium, there are about 9640 long non-coding RNA (lncRNA) loci in human genome [15,16], shedding light on the promising future of lncRNA study. The expanding roles of lncRNA in biological processes await a larger scale of RNAi library. Theoretically, a completely random shRNA library is of priority for such purposes. However, there are still no such stable random shRNA libraries available. There were multiple excellent efforts trying to construct random shRNA libraries [17–20]. However, these strategies are of some limitations, such as insertion of additional sequence, competition of the two promoters or not compatible to the pLKO.1, the mostly used lentiviral vectors. To this end, developing a strategy to construct completely random shRNA library is still needed.

Here we provide a strategy for lentivirus based random shRNA library construction. This kind of shRNA library can be easily constructed, providing another choice for whole-genome screening of genes.

* Corresponding authors. Address: No. 17 Changlexi Road, The Fourth Military Medical University, 710032 Xi'an, PR China. Fax: +86 29 84773947.
E-mail addresses: guodong.yang029@gmail.com (G. Yang), luzfliuq@fmmu.edu.cn (Z. Lu), bioyao@fmmu.edu.cn (L. Yao).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.11.126

2. Materials and methods

2.1. Oligonucleotide design and synthesis

The oligonucleotide sequences are designed as follows: "Partial-hairpin" structured oligonucleotide, 5′-NNNNNNNNNNNNNNNNNNNNCTCGAGATCACCATTATAAGATCTCGAG-3′; Linker 1, 5′-TTTTTGAATTCCTCAGCAGAGAACTCA-3′; Linker 2, 5′-ACCGGTCCACCAACTATCCAGAC-3′; Primer1, 5′-TGAGTTCTCTGCTGAGGAATTC-3′; and Primer 2, 5′-GTCTGGATAGTTGGTGGACCG-3′. All of these oligos were synthesized and purified with HPLC in Sangon, Shanghai China.

2.2. Fill-in of the "Partial-hairpin" structured oligonucleotides

The synthesized "Partial-hairpin" structured oligos were resolved in ddH_2O to a concentration of 20 μM, and then were added with DNA polymerase I Klenow buffer before annealing. The annealing reaction was done as follows. The dissolved oligonucleotides were incubated for 5 min at 95 °C in a beaker of boiling water and then slowly cooled to room temperature over the period of about 1 h. Then DNA polymerase I Klenow fragment and dNTP were added to the annealed oligonucleotides, and the reaction was run at 25 °C for 2 h.

2.3. 5′-phos modification of the linkers and the PCR primers

The synthesized linkers 1 and 2, primers 1 and 2 used in PCR reaction were added 5′ phos with the T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's instructions. Briefly, the reaction was done in a 50 μl volume aliquot and incubated at 37 °C for 45 min. The reaction contained up to 300 pmol of 5′ termini, 1X T4 Polynucleotide Kinase Buffer, 1 mM ATP and 10 units of T4 Polynucleotide Kinase.

2.4. Ligation of the single strand DNA with the linker

Either the filled-in product or its complementary strand was purified with phenol/chloroform (Sigma–Aldrich) and precipitated with ethanol before resolved in ddH_2O. The 3'-hydroxyl group end of the linker was blocked with ddATP using terminal deoxynucleotidyltransferase (TdT) before ligation. The linker and the purified single strand DNA were mixed at a molar ratio of 1:1 for proper ligation. Ligation can be completed by T4 RNA ligase or its homologue. The ligated DNA was further purified as described above using phenol/chloroform/ethanol method.

2.5. PCR amplification

The ligation product of linker 2 and complementary DNA was used as template, and amplified with the previously 5′ phos modified primers (just for easy cloning). The PCR reaction was done as follows: 94 °C for 5 min, 94 °C 40 s, 68 °C 1 min, for 10 cycles. The PCR products were run on PAGE gel and the amplicons with the exact size were purified.

2.6. Cloning and subcloning

The PCR products were inserted into the pMD-18 *T* simple vector, which has no EcoR1, Xho1, Bbs1, Age1 enzyme sites. To remove the excess sequence in the loop, the pMD-18 *T* simple vector was further digested with Xho1 and self-ligated using T4 DNA ligase. Then, the constructed *T* vector with only one Xho1 site was further digested with Age1 and EcoR1 and the cleaved fragment was further subcloned into the pLKO.1 vector.

3. Results and discussion

3.1. Generation of the random shRNA library

It is technically impossible to synthesize oligonucleotides for a random shRNA library directly. And most of the current available shRNA libraries are a collection of custom designed shRNAs [11]. Previously, multiple efforts have tried to construct the random shRNA library. However, none of these strategies were based on pLKO.1. What's more, these strategies are either inefficient or including some additional sequences, rather than the exact sequence for a typical shRNA expression [17–20]. To this end, constructing a random shRNA library based on pLKO.1 vector and expressing the exact the shRNA with no additional nucleotides, is preferred.

Here we provided an easy protocol for constructing such a library, as seen in Fig. 1. The steps are briefly described as follows:

(1) The oligonucleotides composed of a randomized 21-nt sequence, a loop and a short hairpin structure, were chemically synthesized. After annealing, the synthesized oligonucleotides form a secondary structure as shown in Fig. 1 (up panel). The partially short hairpin structure is further extended by DNA polymerase I Klenow fragment.
(2) After the first step, the 3′ ends of resultant stem-loop DNA were purified then ligated with the linker1, in which 5′ phos group was added by PNK.
(3) The ligated oligonucleotides the served as a template for synthesis of the complementary strand, which was completed by the PCR mix.
(4) The 3′ ends of the newly synthesized strands were further ligated with linker 2.
(5) The products of the step 4 were further amplified by PCR reaction for 10 cycles, using primer 1 and 2; the products were further added additional "A" in the 3′end if the PCR enzyme with high 3′ exonuclease activity.
(6) The above products with expected size were then purified by gel electrophoresis before insertion into the *T* vectors, and the ligated vector was further transformed into the high efficiency competent cells.
(7) The clones were mixed together and cultured in the LB medium for 16 h before harvested for plasmid purification, and the purified plasmids were further digested with Xho1 and then self-ligated and transformed into competent cells again; the newly formed vectors, which harbored the exact sequence for shRNA expression and processing, were further amplified and purified.
(8) The shRNA expression insertions were then cut off and subcloned into the pLKO.1 vector, forming a new library for further shRNA lentiviral construction.

Usually, we could get about 10^3 clones when we ligate 10 ng of the PCR products in step 5 with 50 ng of *T* vectors, and more than 95% of the clones are positive with insertions, suggesting that scaling up the reaction is practical to cover all the random shRNAs.

3.2. Further modification and implication

It is essential to produce enough high-titer viral samples for genome-scale loss-of-function screens. Currently, the third-generation self-inactivating lentiviral vector pLKO.1 vector is widely used for shRNA expression and commonly used to maximize viral titers, in which a human U6 promoter is used to drive the expression of shRNAs, and a PGK promoter driving expression of the puromycin resistance gene to allow selection of transduced

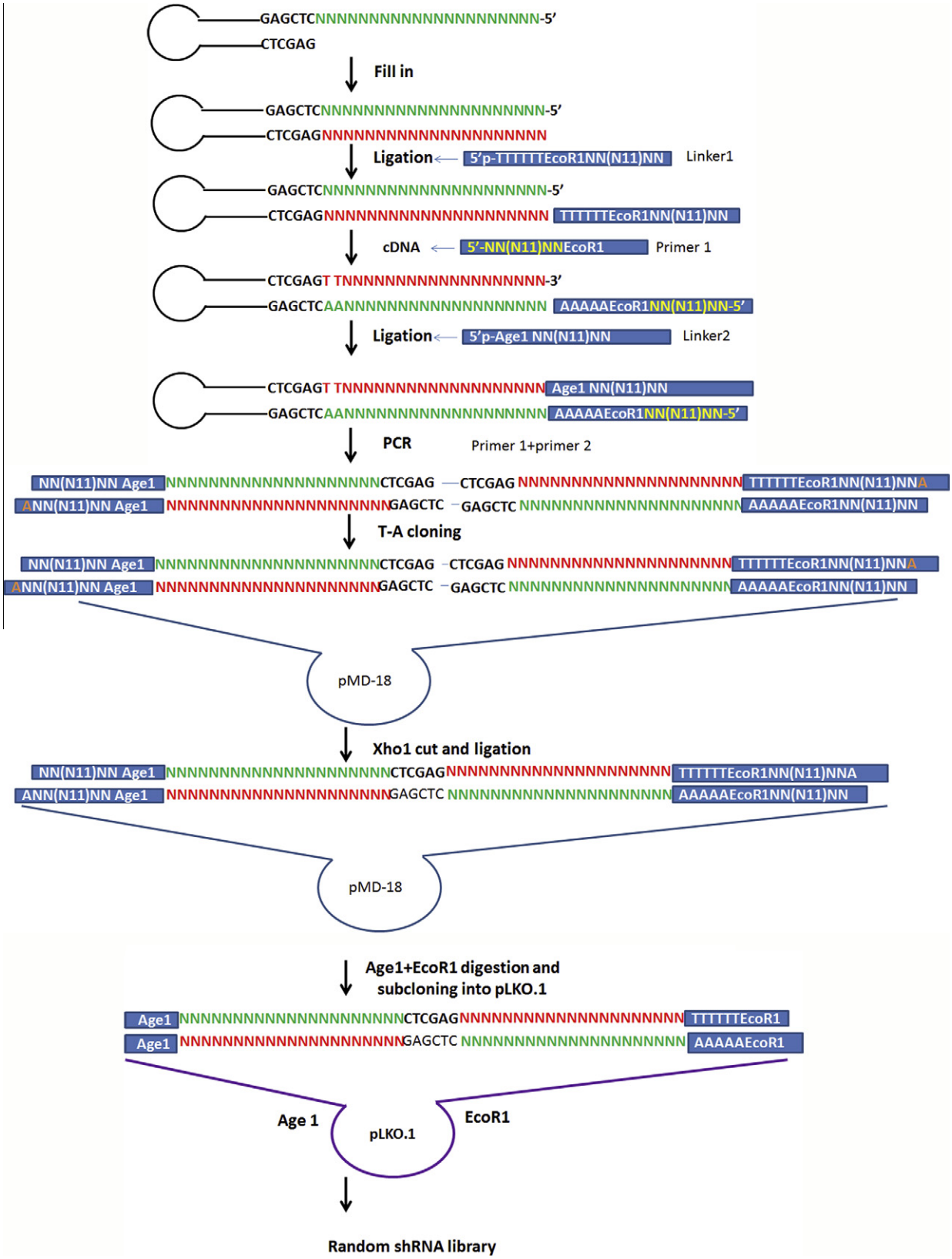

Fig. 1. Workflow of random shRNA vector construction. The procedure for generating randomized shRNA expressing pLKO.1 library. The synthesized oligonucleotides with 21 nt random 5′ overhangs were annealed and filled in. The resultant stem-loop DNA was ligated with the 5′ phosphorylated linker and then the reverse strand was synthesized by PCR reaction with primer 1. The resultant reverse strand was further ligated with linker 2. The sequence was then amplified with primer 1 and 2 for 10 cycles and cloned into pMD-18 *T* simple vector. Additional nucleotides in the loop were removed by Xho1 digestion. Then the insertion was subcloned into pLKO.1.

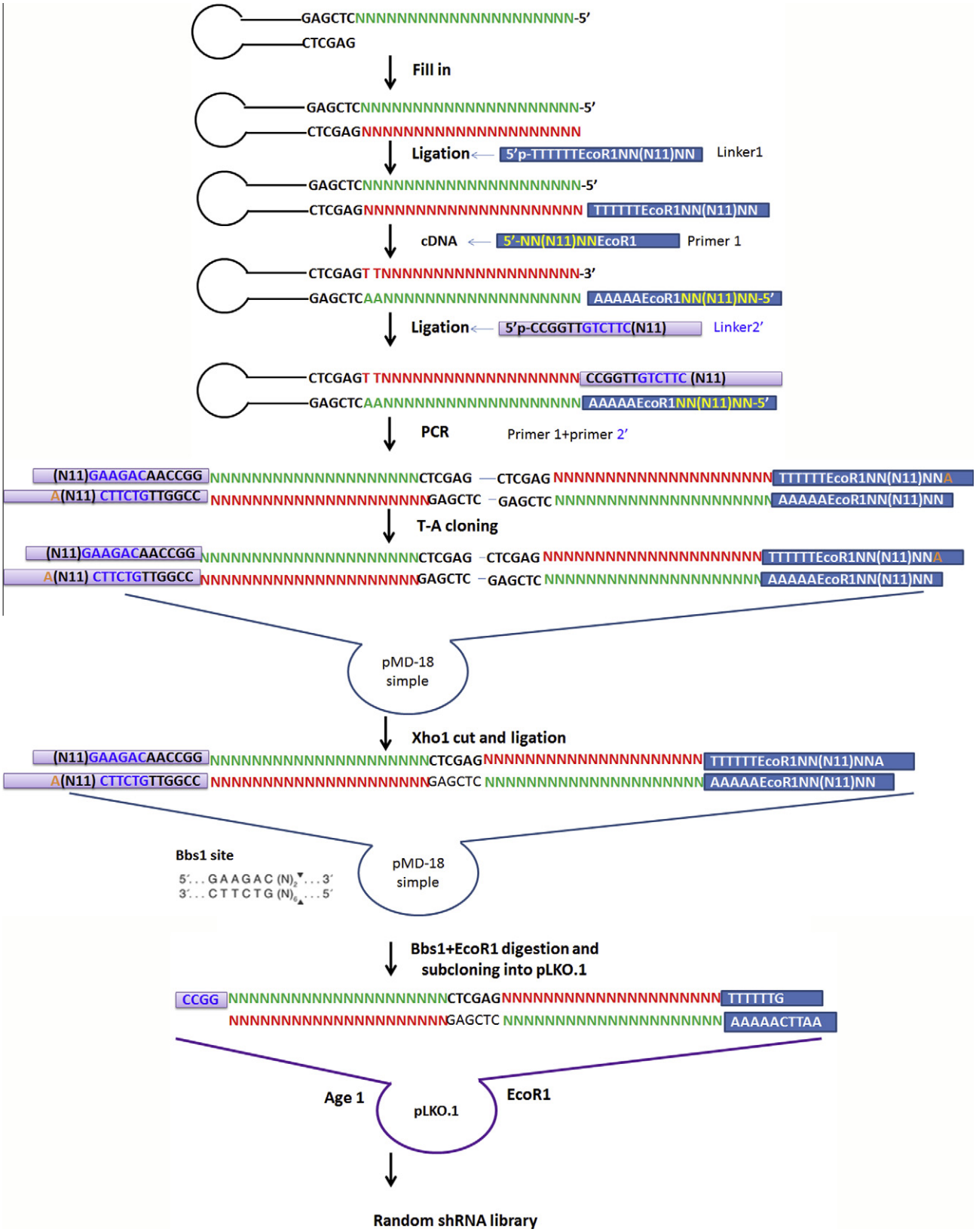

Fig. 2. Workflow of random shRNA vector with Age1 site mutated. The whole process was similar as that in Fig. 1B, with the Linker 2 and Primer 2 changed. The linker 2 harbors a Bbs1 site. After Bbs1 digestion, the sequence would produce the same overhang as the Age1 while without *T* behind, making it the same as the desired sequence as recommended by the pLKO.1 protocol.

cells. Usually, we need to synthesize two reversely complementary DNA sequences and clone the annealed sequence into the pLKO.1 vector. From the above protocol, we can expect a really random shRNA library is around the corner. However, there are still some modifications needed to be done to make the protocol really cover all the effective random shRNAs.

Table 1
The 21-nucleotide distribution of the random shRNA library based on the rational design criterion.

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Nucleotide	A	A	A	A	A	A	A	A	A	A	A	U	A	A	A	A	A	A	A	A	A
			U	U		U	U	U	U	U	U		U	U	U	U	U	U	U	U	U
			C	C		C	C	C	C	C	C		C	C	C	C	C	C	C	C	
			G	G		G	G	G	G	G	G		G	G		G	G	G	G	G	

3.2.1. The transcription start site and transcription efficiency

As we see, the Age1 site is still kept in the above shRNA library, while most of the previous constructions using the synthesized oligonucleotides directly have mutated the enzyme site, with no *T* after the flanking end. Notably, the remaining "*T*" in our protocol will not be complementary to the antisense strand of the hairpin siRNA, and there are lots of effective shRNAs starts with "*T*". This extra nucleotide in our protocol appears to have no effect on the activity of the hairpin siRNA as it only appears in the sense strand.

However, RNA pol III is considered to prefer to initiate transcription with a purine, and thus mutation of Age1 is supposed to facilitate efficient transcription (*Instruction Manual of psilencer3.1*). To this end, we need to mutate the Age 1 site for a more efficient library. In this regard, we provide the following modifications (Fig 2). The most of the steps are the same, with linker 2 and primer 2 changed. In this modified protocol, the Age1 was replaced by Bbs1 site and an additional 6 nucleotide in the end. Digestion with the Bbs1 site would produce the exact Age1 flanking end, while without the "*T*".

3.2.2. The GC content for efficient knockdown

Our strategy here includes a step of PCR. The advantage is that the resultant products are double strand and is efficient for ligation. The disadvantage is that PCR would produce preference for those sequences with lower complexity of secondary structures. To this, these preferred products might be exactly what we want. As we know, efficient shRNA sequences often have a GC content less than 50% [21]. To achieve a theoretically random shRNA library, the DNA polymerase, such as the sequencing enzyme (sequenase) and Phusion PCR enzyme might be desired, which are highly processive, and are not impeded by secondary structures [22] in the fill-in process and the PCR process.

3.2.3. To reduce the clonal numbers rationally

For a random shRNA with 19 + 2 nt structure, there would be 274,877,906,944 individual shRNAs, this large number challenges the capacity for clone construction and virus packaging. And what is more, this large random shRNA library would also challenge the effort for screening the interested genes. To this end, minimize the shRNA library rationally would be important. In fact, the efficient shRNAs share some common features, and these features would largely reduce the shRNA number. According to the rational design principle proposed by Reynolds [21], the nucleotide in some positions can be fixed (Table 1), and then the shRNA number can be minimized to 6,442,450,944, making the shRNA library much smaller and more practical.

In summary, to our knowledge this is the first protocol for constructing complete random shRNA library with exact the sequence expressing typical shRNAs. One of the most appealing advantages is the shRNA library could target one gene by tens to hundreds different shRNA, and thus easy for screen and avoiding the off-target effects. Notably, the proposed strategy for constructing the shRNA is still with some shortcomings. For example, the ligation efficiency of the single strand DNAs is very low, which might compromise the random library scale. In this regard, efforts improving the ligation efficiency are needed. Alternatively a stem-loop constructed linker might facilitate the ligation by T4 DNA ligase and thus might be preferred [17]. More efforts involved to make this strategy really feasible and practical are still needed.

Acknowledgments

This study was funded by National Science Foundation of China, NSFC 81030046, NSFC 31100979, NSFC81170149, and NSFC 81101050.

References

[1] K. Chang, S.J. Elledge, G.J. Hannon, Lessons from nature: microRNA-based shRNA libraries, Nat. Methods 3 (2006) 707–714.
[2] P.D. Zamore, T. Tuschl, P.A. Sharp, D.P. Bartel, RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals, Cell 101 (2000) 25–33.
[3] A. Fjose, S. Ellingsen, A. Wargelius, H.C. Seo, RNA interference. mechanisms and applications, Biotechnol. Annu. Rev. 7 (2001) 31–57.
[4] S.M. Hammond, S. Boettcher, A.A. Caudy, R. Kobayashi, G.J. Hannon, Argonaute2, a link between genetic and biochemical analyses of RNAi, Science 293 (2001) 1146–1150.
[5] K. Berns, E.M. Hijmans, J. Mullenders, T.R. Brummelkamp, A. Velds, M. Heimerikx, R.M. Kerkhoven, M. Madiredjo, W. Nijkamp, B. Weigelt, R. Agami, W. Ge, G. Cavet, P.S. Linsley, R.L. Beijersbergen, R. Bernards, A large-scale RNAi screen in human cells identifies new components of the p53 pathway, Nature 428 (2004) 431–437.
[6] J. Mullenders, A.W. Fabius, M. Madiredjo, R. Bernards, R.L. Beijersbergen, A large scale shRNA barcode screen identifies the circadian clock component ARNTL as putative regulator of the p53 tumor suppressor pathway, PLoS ONE 4 (2009) e4798.
[7] T.R. Brummelkamp, A.W. Fabius, J. Mullenders, M. Madiredjo, A. Velds, R.M. Kerkhoven, R. Bernards, R.L. Beijersbergen, An shRNA barcode screen provides insight into cancer cell vulnerability to MDM2 inhibitors, Nat. Chem. Biol. 2 (2006) 202–206.
[8] I.G. Kolfschoten, B. van Leeuwen, K. Berns, J. Mullenders, R.L. Beijersbergen, R. Bernards, P.M. Voorhoeve, R. Agami, A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity, Cell 121 (2005) 849–858.
[9] V.N. Ngo, R.E. Davis, L. Lamy, X. Yu, H. Zhao, G. Lenz, L.T. Lam, S. Dave, L. Yang, J. Powell, L.M. Staudt, A loss-of-function RNA interference screen for molecular targets in cancer, Nature 441 (2006) 106–110.
[10] G. Sethi, H.B. Pathak, H. Zhang, Y. Zhou, M.B. Einarson, V. Vathipadiekal, S. Gunewardena, M.J. Birrer, A.K. Godwin, An RNA interference lethality screen of the human druggable genome to identify molecular vulnerabilities in epithelial ovarian cancer, PLoS ONE 7 (2012) e47086.
[11] D.E. Root, N. Hacohen, W.C. Hahn, E.S. Lander, D.M. Sabatini, Genome-scale loss-of-function screening with a lentiviral RNAi library, Nat. Methods 3 (2006) 715–719.
[12] C.J. Echeverri, N. Perrimon, High-throughput RNAi screening in cultured cells: a user's guide, Nat. Rev. Genet. 7 (2006) 373–384.
[13] C.J. Brown, B.D. Hendrich, J.L. Rupert, R.G. Lafreniere, Y. Xing, J. Lawrence, H.F. Willard, The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus, Cell 71 (1992) 527–542.
[14] R. Feil, J. Walter, N.D. Allen, W. Reik, Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes, Development 120 (1994) 2933–2943.
[15] B.E. Bernstein, E. Birney, I. Dunham, E.D. Green, C. Gunter, M. Snyder, An integrated encyclopedia of DNA elements in the human genome, Nature 489 (2012) 57–74.
[16] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C.A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, R. Guigo, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Res. 22 (2012) 1775–1789.
[17] Y. Nishikawa, T. Sugiyama, A shRNA library constructed through the generation of loop-stem-loop DNA, J. Gene Med. 12 (2010) 927–933.
[18] F. Buchholz, R. Kittler, M. Slabicki, M. Theis, Enzymatically prepared RNAi libraries, Nat. Methods 3 (2006) 696–700.

[19] M. Chen, L. Zhang, H.Y. Zhang, X. Xiong, B. Wang, Q. Du, B. Lu, C. Wahlestedt, Z. Liang, A universal plasmid library encoding all permutations of small interfering RNA, Proc. Natl. Acad. Sci. USA 102 (2005) 2356–2361.
[20] Q. Guo, Y. Kong, L. Fu, T. Yu, J. Xu, W. Chen, A randomized lentivirus shRNA library construction, Biochem. Biophys. Res. Commun. 358 (2007) 272–276.
[21] A. Reynolds, D. Leake, Q. Boese, S. Scaringe, W.S. Marshall, A. Khvorova, Rational siRNA design for RNA interference, Nat. Biotechnol. 22 (2004) 326–330.
[22] S. Tabor, C.C. Richardson, Selective inactivation of the exonuclease activity of bacteriophage T7 DNA polymerase by *in vitro* mutagenesis, J. Biol. Chem. 264 (1989) 6447–6458.