Published in final edited form as:

Science. 2012 July 27; 337(6093): 421–422. doi:10.1126/science.1225787.

Use and Abuse of RNAi to Study Mammalian Gene Function:

Molecular Biology

William G. Kaelin Jr

Department of Adult Oncology, Office M457, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA. william_kaelin@dfci.harvard.edu

For decades, scientists studying mammalian cells could only marvel at the genetic tools available to scientists who studied organisms such as yeasts or drosophila. However, many genes linked to human diseases do not have orthologs in such model organisms, or they require an appropriate cellular context to display their disease-relevant phenotypes. The wish for a facile method to disrupt gene function in somatic mammalian cells appeared to be granted with the discovery of RNA interference (RNAi) and small interfering RNA (siRNA) (or short hairpin RNA, shRNA), which brought with them great promise—particularly for discovering novel drug targets through the use of genetic screens (1–3). However, the honeymoon is now over, and although some new discoveries have been made, the yield has fallen far short of expectations. Many drug targets identified by means of si/shRNA technology in academic laboratories are not robust when tested in industrial laboratories (4, 5). Avoiding this fate requires a more sophisticated interpretation of si/shRNA results, especially in the context of high-throughput screens.

As with any new technology, the initial euphoria is now being tempered by a growing awareness of the pitfalls. Perhaps the most damaging of these is the potential for any given si/shRNA to affect genes other than its intended target (off-target effects) (6). This concern is compounded when the phenotype being scored could also reflect a loss of cellular fitness, such as a decrease in cell proliferation or viability, which is regulated by many genes. More generally, inhibiting a complex biological phenotype (an “off” or “down” assay) is more susceptible to spurious results than the activation or restoration of a complex phenotype (an “on” or “up” assay). The common practice of focusing on the si/shRNAs for a given gene that have the strongest phenotypes in off/down screens regardless of their relative knockdown efficiencies runs the risk of enriching for off-target effects.

The minimum standard in the field is to show that the same phenotype is observed with two or more independent si/shRNAs targeting the same gene, ideally linked to a statement as to how many si/shRNAs were actually tested. In theory, the degree of target knockdown achieved with individual si/shRNAs would also correlate with the magnitude of their phenotypic effects. In practice, however, this is often not the case, presumably because the gene activity–phenotype relationship for many genes is not linear and/ or because si/shRNA off-target effects can suppress or enhance the on-target phenotype.

In principle, the gold standard for confirming that a si/shRNA phenotype is on-target is the demonstration that the phenotype can be reversed (“rescued”) by a si/shRNA-resistant mRNA for the gene of interest. However, this is again often difficult in practice. For example, it might be necessary to recapitulate the physiological expression of the targeted

Realizing the full potential of si/shRNA technology requires more sophisticated approaches to address the pitfalls.

Supplementary Materials

www.sciencemag.org/cgi/content/full/337/6093/421/DC1 Fig. S1

gene, in which case conventional expression vectors will usually be inadequate. The use of minigenes or bacterial artificial chromosomes might be helpful here (7). However, even when successful, rescue experiments do not exclude that the observed si/shRNA phenotype requires inhibition of the intended target and one or more off-target effects (that is, reflects synthetic interactions between the target gene and other genes).

It is therefore highly desirable to corroborate si/shRNA findings with alternative approaches, of which there are many available. For example, dominant-negative mutants and pharmacological agents are especially useful when the goal is to credential a drug target, because absence of a target (such as achieved with a si/shRNA) need not phenocopy an inactive protein (such as achieved with a drug). Another useful approach, especially for off/down si/shRNA phenotypes, is to show that increasing the activity of the si/shRNA target (such as through overexpression) induces changes that are the opposite of those seen with the si/shRNA. In some cases, known epistatic relationships can also be exploited to reverse si/shRNA phenotypes and to strengthen the case for an on-target relationship. Finally, other methods for inactivating gene function, such as adeno-associated virus-based homologous recombination and TAL (transcription activator-like) effector endonucleases (8, 9), might help but are currently low throughput. It may be feasible to conduct unbiased, relatively high-throughput, insertional mutagenesis screens in human haploid cancer cells (10).

These issues are potentially compounded by problems related to multiple hypothesis testing when large si/shRNA libraries are used in high-throughput screens. Many of the first-generation screens used libraries containing three to five si/shRNAs per target gene where the knockdown efficiencies and off-target effects of the individual si/shRNAs were unknown. In such large screens, and especially those approaching genome-wide scale, the probability that two or more si/shRNAs for a given gene will have similar phenotypes (especially in off/down screens) because of random off-target effects increases, leading to false discoveries. Although these screens clearly do generate valuable data, the extent to which false-positives contaminate the results is currently unclear.

These concerns can be partially mitigated through the use of algorithms that weigh the behavior of each of the si/shRNAs for a given gene to generate a gene score that can then be compared to the scores obtained for the other genes in the library. At three to five si/shRNAs per gene, however, different algorithms that are reasonable a priori can give very different scores. For example, one algorithm might place greater weight on the “top scoring” one, two, or three si/shRNAs out of a set, and another might place greater weight on the absence of outlier si/shRNAs within a set.

Ranking genes relative to one another becomes even more challenging because the rank for a given gene will potentially reflect a number of variables that are unknown (or unknowable), including the knockdown efficiencies of the si/shRNAs targeting that gene, their off-target effects, and the dose-response curve linking the activity of that gene to the phenotype being scored (see the figure and fig. S1). It is therefore not surprising that screens designed to achieve the same end (for example, to kill cells harboring a particular pathogen or cancer-relevant mutation) but using different libraries and different ranking algorithms can yield very different results (11–15).

So, what’s the best way forward? In the short term, one approach would be to conduct more focused screens, perhaps involving fewer than 100 genes, to allow deeper interrogation of primary screen hits in lower-throughput secondary screens. For example, one might focus on the genes within a particular disease-associated amplicon or that encode a particular family of enzymes. In the long term, the performance of genomescale screens will improve with further library enhancements (including increasing the number of si/shRNAs per gene and

eliminating si/shRNA empirically found to produce false-positives across multiple screens), the use of algorithms that take into account si/shRNA knockdown efficiencies, and incorporation of orthogonal data sets (for example, data from genomic studies or chemical screens). In the meantime, greater prudence is needed when analyzing si/shRNA data if we are to remain in love beyond the honeymoon.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Elbashir SM, et al. *Nature*. 2001; 411:494. [PubMed: 11373684]
2. Brummelkamp TR, et al. *Science*. 2002; 296:550. [PubMed: 11910072]
3. Sui G, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:5515. [PubMed: 11960009]
4. Prinz F, Schlange T, Asadullah K. *Nat. Rev. Drug Discov.* 2011; 10:712. [PubMed: 21892149]
5. Begley CG, Ellis LM. *Nature*. 2012; 483:531. [PubMed: 22460880]
6. Echeverri CJ, et al. *Nat. Methods*. 2006; 3:777. [PubMed: 16990807]
7. Kittler R, et al. *Proc. Natl. Acad. Sci. U.S.A.* 2005; 102:2396. [PubMed: 15695330]
8. Khan IF, et al. *Nat. Protoc.* 2011; 6:482. [PubMed: 21455185]
9. Bogdanove AJ, Voytas DF. *Science*. 2011; 333:1843. [PubMed: 21960622]
10. Carette JE, et al. *Science*. 2009; 326:1231. [PubMed: 19965467]
11. Barrows NJ, Le Sommer C, Garcia-Blanco MA, Pearson JL. *J. Biomol. Screen.* 2010; 15:735. [PubMed: 20625183]
12. Goff SP. *Cell*. 2008; 135:417. [PubMed: 18984154]
13. Luo J, et al. *Cell*. 2009; 137:835. [PubMed: 19490893]
14. Scholl C, et al. *Cell*. 2009; 137:821. [PubMed: 19490892]
15. Barbie DA, et al. *Nature*. 2009; 462:108. [PubMed: 19847166]

Figure. si/shRNA screen variables

For a phenotypic si/shRNA screen involving n genes, the ranking of any one gene will be influenced by the knockdown efficiencies of the si/shRNAs (A to D in the example shown) for that gene, their off-target effects (for simplicity, on-target effects not shown), and the dose-response curves linking the activity of that gene and the phenotype being measured (for example, loss of viability).