Pharm Res (2012) 29:1717–1721
DOI 10.1007/s11095-012-0712-x

PERSPECTIVE

Bottlenecks Caused by Software Gaps in miRNA and RNAi Research

Sean Ekins • Ron Shigeta • Barry A. Bunin

Received: 2 January 2012 / Accepted: 15 February 2012 / Published online: 24 February 2012

ABSTRACT Understanding the regulation of gene expression is critical to many areas of biology while control via RNAs has found considerable interest as a tool for scientific discovery and potential therapeutic applications. For example whole genome RNA interference (RNAi) screens and whole proteome scans provide views of how the entire transcriptome or proteome responds to biological, chemical or environmental perturbations of a gene's activity. Small RNA (sRNA) or MicroRNA (miRNA) are known to regulate pathways and bind mRNA, while the function of miRNAs discovered in experimental studies is often unknown. In both cases, RNAi and miRNA require labor intensive studies to tease out their functions within gene networks. Available software to analyze relationships is currently an ad hoc and often a manual process that can take up to several hours to analyze a single candidate RNAi or miRNA. With experiments frequently highlighting tens to hundreds of candidates this represents a considerable bottleneck. We suggest there is a gap in miRNA and RNAi research caused by inadequate current software that could be improved. For example a new software application could be created that provides interactive, comprehensive target analysis that leverages past datasets to lead to statistically stronger analyses.

KEY WORDS informatics · microRNA · screening · siRNA · software

MIND THE GAP

Small RNAs have found an increasing role as biological tools and potential therapeutic modulators (1). RNA interference (RNAi) silencing has introduced a faster, less expensive approach than genetic screening by random mutagenesis. At the same time it has become the most widely used technique for analyzing loss of function phenotypes of individual genes and dissecting complex regulatory pathways (2) with genome-scale screens in nematode, *Caenorhabditis elegans*, and assays in fly, mouse, and human cells (3–8). This technology is an invaluable tool in identifying novel therapeutic targets, developing therapeutic agents and in analysis of mechanisms of action of small molecules (9). MicroRNA (miRNA) can regulate genes and represent a large percentage of the human genome, yet the functional relevance of the majority is unknown (10). Recent studies have identified miRNAs involved in neurogenesis (11), type 2 diabetes mellitus (12), and many other diseases. The availability of miRNA array technologies (Applied Biosystems, Affymetrix and Capital Bio MicroRNA product lines as well as academic consortia efforts such as ORB Sanger15 Multispecies MIcroRNA Microarray) in the past 2 years creates an expectation of more data appearing in the immediate future.

S. Ekins (✉) • B. A. Bunin
Collaborative Drug Discovery
1633 Bayshore Highway, Suite 342
Burlingame, California 94010, USA
e-mail: ekinssean@yahoo.com

S. Ekins
Collaborations in Chemistry
5616 Hilltop Needmore Road
Fuquay-Varina, North Carolina 27526, USA

S. Ekins
Department of Pharmaceutical Sciences, University of Maryland
College Park, Maryland 21201, USA

R. Shigeta
Affymetrix, Inc.
6550 Vallejo Street, Suite 100
Emeryville, California 94608, USA

siRNA share common mechanisms of action, but the effectiveness of a given siRNA is difficult to predict. They vary in terms of their uptake by cell types and experimental conditions, different sites on the target mRNA will yield different qualities of results. While there are several methods to design RNAi reagents (13), RNAi reagents need experimental confirmation to show effective behavior. RNAi high throughput (HT) screens can be improved substantially with a system that tracks the specific miRNA probe sets or RNAi molecule in a way like other cheminformatics software products already available for small molecules or reagents. RNA panels typically have thousands of reagents, so tracking and annotation of their individual designs would be a desirable improvement to data analysis.

HT screens tend to yield hundreds of statistically significant results. Yet a typical analysis of an HT RNAi screen focuses on a limited number of genes with the greatest knockdown effects, a method that is dictated more by the limitations of low replicate count in the experiment and minimal screening for biological function. For each gene, available functional and structural annotations across many databases (Table I) are examined in the search for clues to understanding the functional response (2). The researcher can also consult available annotations on homologous genes in other organisms. This *ad hoc* process is laborious and error-prone, usually requiring several hours of work per gene, repetitive mouse clicks, and copying/pasting among multiple websites. Some specialized software exists but may be too sophisticated and expensive for most researchers (Table II). Researchers typically persevere by using other alternatives, either software packages for small molecule chemical screens or MS Excel. These applications are adequate for storing, retrieving and sorting RNAi data, but are not optimized to facilitate efficient and thorough analysis and statistical methods used for small molecules screens may not be as useful for RNAi with its lower Z scores, increased experimental variability and decreased signal to noise (14).

There is also a need to reduce false positives and false negatives. A typical screen yielding 400 hits often requires many hours over weeks or months to analyze manually, leading many gene results to simply be ignored. One example of a typical user was described (15). After controlling for false positives, 225 hits were found; after reducing cutoffs by 50–75%, 30 hits remained. This experiment indicates the amount of effort currently required, resulting in publishing the top 30 hits that are conveniently manually searchable. Another example is a case in which gene interaction databases are combined with RNAi results to give strong clustering of results (16). Custom software may sometimes expedite this process. Differences between custom solutions across institutions also result in the data produced not being easily reproduced or shared with others. Gene-by-gene analysis has been shown to be lacking for other types of genome-scale experiments and is subject to serious flaws (17). A critical gap therefore exists in available software to facilitate effective extraction of statistically significant information from RNAi screens. The ability to compare RNAi and coding RNA microarray data would also provide a broader window on this data. In our minds this is comparable to the development of microarray expression analysis software where Partek (www.partek.com) and GeneSpring (Agilent) produced software suites providing comprehensive and integrated analysis tools and therefore found a route to commercial success.

While there are several small molecule screening analysis systems (Accelrys Isentris, IDBS ActivityBase and CambridgeSoft (now Perkin Elmer) Enterprise Solutions), none to our knowledge currently supports RNAi screening data, but they can be used to archive RNAi plate configurations and assay data. These systems can perform standard analyses of readouts but cannot integrate readout information with the biological context of genes by linking to the many available databases. Doing so still requires time-consuming, manual collation. Freely available commercial software for analysis of differential gene expression in microarray experiments (e.g. FatiScan (18), GoMiner (19), GSEA (17)) can be used to analyze RNAi screening hits to find effects correlated with groups of genes sharing a common annotation. The major shortfall of these tools is that they are not integrated with other sources of information needed to analyze an RNAi screen. Databases such as Ingenuity Systems' Pathway Knowledgebase and Thomson Reuters/GeneGo's MetaCore application

Table I Databases that Contain Functional and Structural Annotations that Could be Used in Understanding RNAi and miRNA Targets

	Website
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/
UniProt	http://www.uniprot.org/
OMIM	http://www.ncbi.nlm.nih.gov/omim
MeSH	http://www.ncbi.nlm.nih.gov/mesh
PUBMED	http://www.ncbi.nlm.nih.gov/pubmed/
Bio-grid	http://thebiogrid.org/
STRING	http://string-db.org/

Table II Tools for Whole Genome Screen Data Analysis

	Website	References
Spotfire	http://spotfire.tibco.com/	(29,30)
cellHTS	http://www.bioconductor.org/packages/bioc/1.8/html/cellHTS.html	(20)
Ingenuity Pathways Analysis	http://www.ingenuity.com/	(31,32)
MetaCore	http://www.genego.com/	(33–35)

(Table II) provide high quality, integrated information for downstream RNAi screen analyses. Another major problem that plagues RNAi screens is that there is no standard platform for primary screening, assay instrumentation and/or data analyses, compared to what exists for sequencing projects such as microarrays. It is widely accepted that two RNAi screens using exactly the same readout for the same pathway often result in two distinct hit lists, with less than 10% overlap. Software that integrates datasets from a variety of RNAi-HTS must consider the quality of the data generated. Other software tools specifically developed for RNAi screen analysis include: cellHTS (20) written for the R statistical computing environment (www.r-project.org). This software and the R statistical platform are powerful, but require computational and statistical expertise to use correctly, prohibiting widespread adoption. Integrating this software with the researcher's data and other information requires additional software development. There are several web-based tools for siRNA probe design and RNAi screening databases (Table III) but none that address the gap we have identified above.

FILLING THE GAP

In summary, users of RNAi HT screens and miRNA arrays need access to tools that will give them the best options available to screen their data. Currently available tools provide valuable functionality, but none provides a biologist-friendly, open platform for integrated analysis of RNAi screening data. This may be accomplished by leveraging the experimental performance of the RNA in question as well as integrating as much of the biological functional data possible, tracking the performance of specific probe or siRNA panel designs, decreasing false positives and clustering results into interpretable groups. Users of RNAi continue to expend extraordinary time and effort using manual means and abstracting from existing programs to solve problems. Perhaps what is needed is a software tool that combines the available open algorithms in an integrated system, allowing users to select statistical models, annotations, sources and annotation evidence codes to include in their analysis. Briefly this would proceed by; (1) reviewing and screening hit data with Bayes priors calculated from experiments selected from users own or shared data; (2) query screening data using arbitrary nucleotide or amino acid sequences to detect relationships between sequence and function not already recorded in existing databases; (3) interactively reviewing and selecting from a variety of cluster hit data analyses using information from other RNAi experiments or existing gene annotation databases and (4) record and share customized annotations representing the investigator's conclusions (labeling all genes that are 'hits' in one screen for use in querying genetic screens with other biological readouts). Although the use of Bayesian estimates has shown significant improvements of hit estimates (21), Bayesian estimates of RNAi performance across several data sets is not available in any of the available RNAi analysis packages to date (14).

Such an approach will also allow researchers to search not just for effects on sets of genes sharing a common

Table III RNAi Related Databases and Software Resources

	Website	References
GenomeRNAi	http://genomernai.de/GenomeRNAi/	(13,36)
FLIGHT	www.flight.licr.org	(37)
RNAiDB	http://research.imb.uq.edu.au/rnadb/	(38)
flyRNAi	http://www.flyrnai.org/	(39)
E-RNAi	http://www.dkfz.de/signaling/e-rnai3/	(40)
siDirect	http://sidirect2.rnai.jp/	(41)
DEQOR	http://bioinformatics.age.mpg.de/bioinformatics/DEQOR.html	(42)
siRNA Selection Program at Whitehead	http://jura.wi.mit.edu/bioc/siRNAext/home.php	(43)
BLOCK-iT™ RNAi Designer	https://rnaidesigner.invitrogen.com/sirna/	
The RNAi Web	http://www.rnaiweb.com/RNAi/siRNA_Design/	
MIARE	http://miare.sourceforge.net/HomePage	(44)
Broad RNAi Genome Browser	http://www.broadinstitute.org/scientific-community/software?criteria=RNAi	
The RNAi Global Consortia	http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/ProjTRC.shtml	
The Trans NIH RNAi Initiative	http://rnai.nhgri.nih.gov/wordpress/?page_id=16	
miRBase	http://www.mirbase.org/	(45)
RNAi Global	http://www.rnaiglobal.org/Home/	

annotation, but also for effects on genes sharing possible novel, investigator-specified sequence motifs or structural domains. Such software could accelerate RNAi-based research focused on elucidating pathways for new therapeutic strategies in many disease areas (22–26). While research using miRNAs is relatively nascent we believe that this too could benefit from a proposed tool in the same way as RNAi. The utility of a database that could pull multiple datasets together could lead to new insights in the same way that a recent study has integrated many public RNA datasets to identify serum biomarkers for organ transplant rejection (27).

We believe researchers using miRNA and RNAi would also benefit from software approaches that foster intergroup collaboration. Scientists need to manage and analyze their RNA data more effectively while optionally being able to share their data securely. From our own experience of developing a cheminformatics platform targeted at neglected disease and academic researchers that enables toggling between and simultaneously mining across private, shared, and public data sets (28), developing a miRNA and RNAi user network could increase the efficiency of scientific research.

In summary, RNA researchers urgently need new ways to gain an overview of HT arrays and panels that contain tens of thousands of data points and elucidate the biological mechanisms underlying their observations. Manual processes used for most data analyses take considerable time to draw on the many databases available. In the majority of cases these processes are only performed for a fraction of the data that shows a significant effect. Leveraging and clustering the data, will allow users to evaluate genes which are working in concert but may be overlooked by manual processes. Also, data obtained in different laboratories are often stored in incompatible formats and inaccessible databases, preventing effective data mining across laboratories. This is especially true of sharing negative data, which rarely occurs. Two potential benefits of a proposed new miRNA and RNAi software platform over existing methods are; (1) potential for increased accuracy and reproducibility, and (2) achieving more comprehensive analyses with additional significant time-savings. Such an approach described in this perspective could have an important impact on the miRNA and RNAi field and remove a bottleneck that is impeding rapid progress for pharmaceutical research.

ACKNOWLEDGMENTS & DISCLOSURES

The authors kindly acknowledge Dr. Moses Hohman and Dr. Alan Berson.

SE consults for Collaborative Drug Discovery, RS is an employee of Affymetrix, BB is a Board Member and CEO of Collaborative Drug Discovery Inc.

REFERENCES

1. Herrera R, Tien E. Regulation of gene expression by small, non-coding RNAs: practical applications. In: Ekins S, Xu JJ, editors. Wiley series on technologies for the pharmaceutical Industry. Hoboken: Wiley and Sons; 2009. p. 327–47.
2. Perrimon N, Mathey-Prevot B. Applications of high-throughput RNA interference screens to problems in cell and developmental biology. Genetics. 2007;175:7–16.
3. Eggert US, Kiger AA, Richter C, Perlman ZE, Perrimon N, Mitchison TJ, *et al.* Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. PLoS Biol. 2004;2:e379.
4. Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature. 2004;428:431–7.
5. Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, *et al.* A resource for large-scale RNA-interference-based screens in mammals. Nature. 2004;428:427–31.
6. Muller P, Boutros M, Zeidler MP. Identification of JAK/STAT pathway regulators–insights from RNAi screens. Semin Cell Dev Biol. 2008;19:360–9.
7. Muller P, Kuttenkeuler D, Gesellchen V, Zeidler MP, Boutros M. Identification of JAK/STAT signalling components by genome-wide RNA interference. Nature. 2005;436:871–5.
8. DasGupta R, Kaykas A, Moon RT, Perrimon N. Functional genomic analysis of the Wnt-wingless signaling pathway. Science New York, NY. 2005;308:826–33.
9. Kramer R, Cohen D. Functional genomics to new drug targets. Nat Rev Drug Discov. 2004;3:965–72.
10. Su WL, Kleinhanz RR, Schadt EE. Characterizing the role of miRNAs within gene regulatory networks using integrative genomics techniques. Mol Syst Biol. 2011;7:490.
11. Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, *et al.* MicroRNA-mediated conversion of human fibroblasts to neurons. Nature. 2011;476:228–31.
12. Karolina DS, Armugam A, Tavintharan S, Wong MT, Lim SC, Sum CF, *et al.* MicroRNA 144 impairs insulin signaling by inhibiting the expression of insulin receptor substrate 1 in type 2 diabetes mellitus. PLoS One. 2011;6:e22839.
13. Horn T, Arziman Z, Berger J, Boutros M. GenomeRNAi: a database for cell-based RNAi phenotypes. Nucleic Acids Res. 2007;35:D492–7.
14. Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, *et al.* Statistical methods for analysis of high-throughput RNA interference screens. Nat Meth. 2009;6:569–75.
15. Dekanty A, Romero NM, Bertolin AP, Thomas MG, Leishman CC, Perez-Perri JI, *et al.* Drosophila genome-wide RNAi screen identifies multiple regulators of HIF-dependent transcription in hypoxia. PLoS Genet. 2010;6:e1000994.
16. Ho H, Milenkovic T, Memisevic V, Aruri J, Przulj N, Ganesan AK. Protein interaction network topology uncovers melanogenesis regulatory network components within functional genomics datasets. BMC Syst Biol. 2010;4:84.
17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
18. Al-Shahrour F, Minguez P, Tarraga J, Montaner D, Alloza E, Vaquerizas JM, *et al.* BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. Nucleic Acids Res. 2006;34:W472–6.
19. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 2003;4:R28.

20. Boutros M, Bras LP, Huber W. Analysis of cell-based RNAi screens. Genome Biol. 2006;7:R66.
21. Zhang C, Wu C, Blanzieri E, Zhou Y, Wang Y, Du W, Liang Y. Methods for labeling error detection in microarrays based on the effect of data perturbation on the regression model. Bioinform (Oxford, England). 2009;25:2708–14.
22. Ashihara E, Kawata E, Maekawa T. Future prospect of RNA interference for cancer therapies. Curr Drug Targets. 2010;11:345–60.
23. Li L, Zhu J, Tian J, Liu X, Feng C. A role for Gcn5 in cardiomyocyte differentiation of rat mesenchymal stem cells. Mol Cell Biochem. 2010;345:309–16.
24. Lovejoy CA, Xu X, Bansbach CE, Glick GG, Zhao R, Ye F, *et al.* Functional genomic screens identify CINP as a genome maintenance protein. Proc Natl Acad Sci U S A. 2009;106:19304–9.
25. Mohr S, Bakal C, Perrimon N. Genomic screening with RNAi: results and challenges. Annu Rev Biochem. 2010;79:37–64.
26. Jung M, Peterson H, Chavez L, Kahlem P, Lehrach H, Vilo J, *et al.* A data integration approach to mapping OCT4 gene regulatory networks operative in embryonic stem cells and embryonal carcinoma cells. PLoS One. 2010;5:e10709.
27. Chen R, Sigdel TK, Li L, Kambham N, Dudley JT, Hsieh SC, *et al.* Differentially expressed RNA from public microarray data identifies serum protein biomarkers for cross-organ transplant rejection and other conditions. PLoS Comput Biol. 2010;6:e1000940.
28. Hohman M, Gregory K, Chibale K, Smith PJ, Ekins S, Bunin B. Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. Drug Disc Today. 2009;14:261–70.
29. Ahlberg C. Visual exploration of HTS databases: bridging the gap between chemistry and biology. Drug Discov Today. 1999;4:370–6.
30. Stoner CL, Gifford E, Stankovic C, Lepsy CS, Brodfuehrer J, Prasad JV, *et al.* Implementation of an ADME enabling selection and visualization tool for drug discovery. J Pharm Sci. 2004;93:1131–41.
31. Hummel R, Wang T, Watson DI, Michael MZ, Van der Hoek M, Haier J, *et al.* Chemotherapy-induced modification of microRNA expression in esophageal cancer. Oncol Rep. 2011;26:1011–7.
32. Suzuki S, Takeshita K, Asamoto M, Takahashi S, Kandori H, Tsujimura K, *et al.* High mobility group box associated with cell proliferation appears to play an important role in hepatocellular carcinogenesis in rats and humans. Toxicology. 2009;255:160–70.
33. Ekins S, Nikolsky Y, Bugrim A, Kirillov E, Nikolskaya T. Pathway mapping tools for analysis of high content data. Methods Mol Biol. 2007;356:319–50.
34. Luna C, Li G, Qiu J, Epstein DL, Gonzalez P. Role of miR-29b on the regulation of the extracellular matrix in human trabecular meshwork cells under chronic oxidative stress. Mol Vis. 2009;15:2488–97.
35. Kang CJ, Chen YJ, Liao CT, Wang HM, Chang JT, Lin CY, *et al.* Transcriptome profiling and network pathway analysis of genes associated with invasive phenotype in oral cancer. Cancer Lett. 2009;284:131–40.
36. Gilsdorf M, Horn T, Arziman Z, Pelz O, Kiner E, Boutros M. GenomeRNAi: a database for cell-based RNAi phenotypes. 2009 update. Nucleic Acids Res. 2010;38:D448–52.
37. Sims D, Bursteinas B, Gao Q, Zvelebil M, Baum B. FLIGHT: database and tools for the integration and cross-correlation of large-scale RNAi phenotypic datasets. Nucleic Acids Res. 2006;34:D479–83.
38. Gunsalus KC, Yueh WC, MacMenamin P, Piano F. RNAiDB and PhenoBlast: web tools for genome-wide phenotypic mapping projects. Nucleic Acids Res. 2004;32:D406–10.
39. Yan H, Venkatesan K, Beaver JE, Klitgord N, Yildirim MA, Hao T, *et al.* A genome-wide gene function prediction resource for Drosophila melanogaster. PLoS One. 2010;5:e12139.
40. Horn T, Boutros M. E-RNAi: a web application for the multi-species design of RNAi reagents–2010 update. Nucleic Acids Res. 2010;38:W332–9.
41. Naito Y, Yoshimura J, Morishita S, Ui-Tei K. siDirect 2.0: updated software for designing functional siRNA with reduced seed-dependent off-target effect. BMC Bioinform. 2009;10:392.
42. Henschel A, Buchholz F, Habermann B. DEQOR: a web-based tool for the design and quality control of siRNAs. Nucleic Acids Res. 2004;32:W113–20.
43. Yuan B, Latek R, Hossbach M, Tuschl T, Lewitter F. siRNA Selection Server: an automated siRNA oligonucleotide prediction server. Nucleic Acids Res. 2004;32:W130–4.
44. Haney SA. Increasing the robustness and validity of RNAi screens. Pharmacogenomics. 2007;8:1037–49.
45. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res. 2008;36:D154–8.