[Cell Cycle 6:8, 898-901, 15 April 2007]; ©2007 Landes Bioscience

Extra View

High-Throughput Screening by RNA Interference

Control of Two Distinct Types of Variance

David J. Stone[1,*]
Shane Marine[2]
John Majercak[3]
William J. Ray[3]
Amy Espeseth[4]
Adam Simon[3]
Marc Ferrer[2]

Departments of [1]Molecular Profiling; [2]Automated Biotechnology; [3]Alzheimer's Research; and [4]Molecular and Cellular Technology; Merck & Co., Inc.; West Point, Pennsylvania USA

*Correspondence to: David J. Stone; Merck & Co., Inc.; 770 Sumneytown Pike; West Point, Pennsylvania 19486 USA; Tel.: 215.652.3843; Fax: 215.993.1835; Email: david_stone4@merck.com

Original manuscript submitted: 03/07/07
Manuscript accepted: 03/22/07

Previously published online as a *Cell Cycle* E-publication:
http://www.landesbioscience.com/journals/cc/abstract.php?id=4184

KEY WORDS

siRNA, functional genomics, high-throughput screening

ABSTRACT

The availability of genome-wide RNAi libraries has enabled researchers to rapidly assess the functions of thousands of genes; however the fact that these screens are run in living biological systems add complications above and beyond that normally seen in high-throughput screening (HTS). Specifically, error due to variance in both measurement and biology are large in such screens, leading to the conclusion that the majority of "hits" are expected to be false positives. Here, we outline basic guidelines for screen development that will help the researcher to control these forms of variance. By running a large number of positive and negative control genes, error of measurement can be accurately estimated and false negatives reduced. Likewise, by using a complex readout for the screen, which is not easily mimicked by other biological pathways and phenomena, false positives, can be minimized. By controlling variance in these ways, the researcher can maximize the utility of genome-wide RNAi screening.

Genome-wide siRNA libraries have opened up the era of true functional genomics. Although there have been many claims of "functional genomics" since the completion of the human genome project at the turn of the millennium, many of these have in fact been expression genomics studies. In many of these studies, function was actually inferred or deduced on the basis of information taken largely from the literature. Therefore, in many cases the derived function is no better than a guess. Although true functional genomics studies have been done using complex experimental readouts from known physiological states or positive controls and pattern matching the resulting readouts,[1-6] RNAi allows for the first time the direct measurement of gene function in pathways of interest on a genome-wide scale. This approach has been quickly embraced in both academic and industrial research.[7-12]

High-throughput siRNA screens combine the issues particular to both the analysis of HTS and to the study of complex biological systems, and as a result, the researcher has two forms of variance to contend with: measurement and biological. Most experimental scientists are faced with biological variance on a daily basis, but are less concerned with error of measurement as this is often small relative to that found between the organisms in question. On the other hand, HTS laboratories frequently face issues with error of measurement, but are not always forced to contend with the variance introduced when working on living systems. Because both of these forms of variance are present in high-throughput RNAi screens, consideration of both is paramount in the design and analysis of a genome wide siRNA screen.

ERROR OF MEASUREMENT

In large siRNA screens testing ten to twenty thousand genes, it would appear that the researcher has an excellent estimate of measurement error because of the large number of data points. While this is to some extent true, the fact that in most libraries the majority of genes are represented only one time and those that are represented multiple times are often done so under different siRNA designs makes the analysis of the error of measurement more complex. With an n = 1 per gene, it is impossible to determine the error of measurement, and then in turn to determine how much of the total variance is due to differences in the effects of genes on the pathway being studied. In a recently published paper,[12] individual siRNAs were found to have a coefficient of variance of roughly twenty (i.e., standard deviation roughly 20% of the mean levels for all readouts tested) calculated from multiple runs of positive controls and non-silencing negative controls. This fact, combined with the large number of genes tested, leads one to the conclusion that in a genome-wide

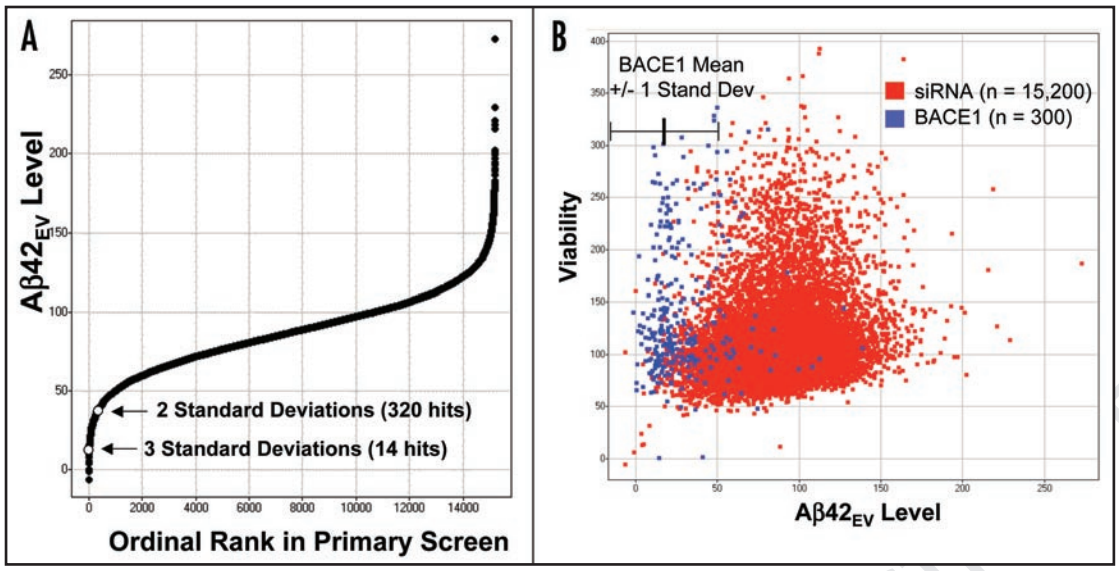

Figure 1. Hit Selection in an APP processing siRNA screen. When hits are selected based on the biological variability (A) a small number of genes are selected, and it is unclear what the power is to detect positive control genes which are in the library (and thus true positives as well). A large number of runs on positive control genes, such as BACE1 in this example (B) can be used to determine the error of measurement associated with a given siRNA or siRNA pool. The researcher can then use this information to determine the cutoff necessary for the desired power to detect the control gene. An APP transgene with a β-secretase cleavage site mutated to NFEV was used in this screen to increase APP protein metabolite levels. Metabolites are therefore designated Aβ42EV, Aβ40EV, sAPPβNF, and sAPPαNFEV, to reflect this fact12.

siRNA screen for effects on a given biological process, the majority of hits are expected to be false positives. Consider a screen with twenty thousand genes tested, searching for a reduction in the production of protein X. Assume an error of measurement as seen in Majercak et al.,[12] and 0.1% of genes actually decreasing protein production by 40% or more. If 5% of the remaining 19,980 genes fall two standard deviations from the mean (as expected by chance), we should see almost 500 genes showing a reduction in protein X levels of 40% or more due to chance alone, resulting in roughly 95% false positives at this level. Indeed, if we were to run a screen using only non-silencing, negative control siRNAs, the data would most likely be normally distributed, with a number of genes 2 and 3 standard deviations from the mean for any readout tested. Again, this brings us to the conclusion that for most screens, and especially for those where the readout of interest turns out to be normally distributed, we must assume that the majority of hits are false positives, especially when a large number of genes are tested.

Testing each siRNA multiple times would provide an estimate of the error of measurement and thus reduce the number of false positives; however due to the expense of genome-wide screens this is not practical in many circumstances. Most researchers therefore prefer the obvious alternative of retesting a subset of siRNA from the primary screen (i.e., the top scoring hits). However, it is critical that the cutoff criteria used for hit selection be determined based upon the error of measurement rather than biological variance. For example, consider the study cited above.[12] This screen was designed to detect genes which affect the processing of the amyloid-precursor protein (APP), especially production of the Aβ42 metabolite which has been implicated in the pathogenesis of Alzheimer's disease. Hit selection cutoffs based strictly upon biological variance are shown in Figure 1A. The mean and standard deviation for the entire data set (n = 15,200 genes knocked down by siRNA, each individual siRNA considered once), were calculated and marked as potential cutoffs for secondary screening. Choosing a cutoff of three standard deviations below the mean, the researcher would only consider those siRNAs lowering amyloid to 12% of the non-silencing control (fourteen hits total); likewise choosing a less stringent cut of two standard deviations would included siRNAs lowering amyloid levels to 37% of control (320 hits total). Although at first pass these seem reasonable, Figure 1B shows the behavior of a single, positive control siRNA (*BACE1*) over the course of 312 runs. BACE1 is the first enzyme to cleave APP in the production of Aβ, and has been shown to be essential for this process.[13] The average amyloid reduction by *BACE1* knock-down was roughly 70% from that of the non-silencing control, and the standard deviation (for *BACE1* knock down only) was roughly 22%. Note that here we are only considering the knock-down of a single gene to determine an estimate of the error of measurement, rather than biological variance.

These data allow us to ask the question "What is the likelihood that a researcher using the described cutoffs would detect a gene with an effect equal to *BACE1* in this screen?". *BACE1* is necessary for amyloid production, and is thus the gold standard for this assay. Surprisingly, at a cutoff of three standard deviations, the researcher would have only a 16% chance of detecting this gene if represented only once in the library. Even when the cutoff is reduced to two standard deviations, the researcher still has a 25% chance of not detecting a gene with an effect equal to *BACE1*. Even if a "gold standard" positive control gene is not available/ not known, one can still calculate the power to detect genes with a given effect size if some genes are run in all 384-well plates and the error of measurement is calculated. If a researcher fails to do this they will be unaware if the screen is underpowered; an unacceptable risk considering the time and expense involved in whole genome siRNA HTS.

Examination of the error of measurement has led to two conclusions that must be considered when designing/running a high-throughput RNAi screen: the majority of the hits are false positives, and one must adjust the cutoff level for genes to be selected for secondary screening downward to accommodate the error of

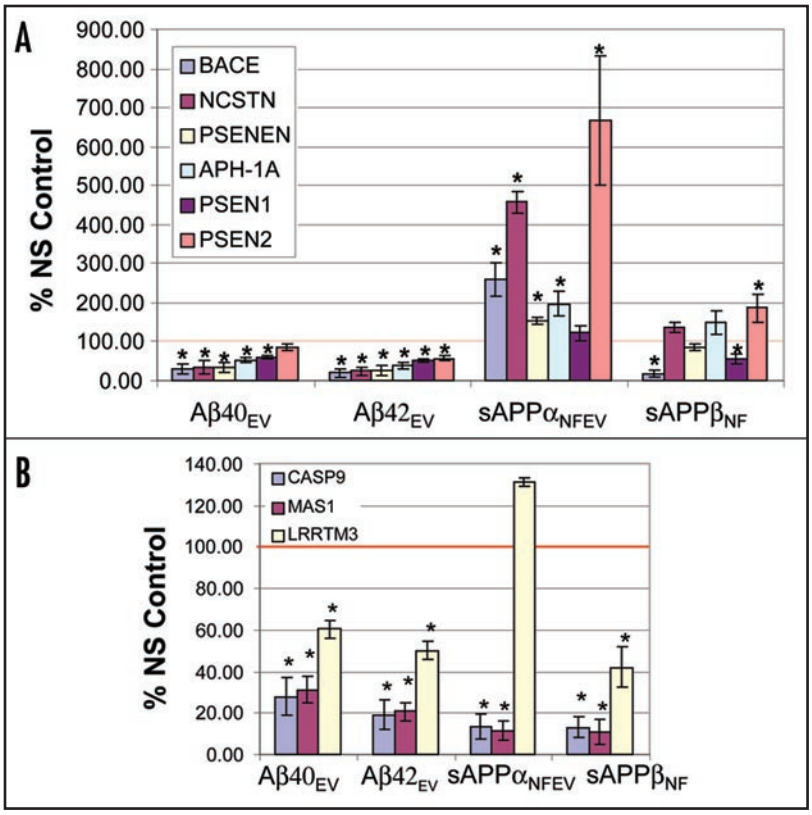

Figure 2. Multiple readouts separate true from false positives. Known secretase genes were used as positive controls (A) and none showed a profile reducing all metabolites measured, but rather shifting ratios between the various products. Many hits from the screen (B) had profiles similar to *CASP9* and *MAS1*, reducing all metabolites. These were shown to be reducing APP through altered expression of the transgene. True hits, such as *LRRTM3*, shifted the ratios between the various APP metabolites as was seen in positive controls. All values are shown relative to non-silencing controls for that metabolite (fixed at 100%). * = p< 0.05 vs. non-silencing control siRNA treated cells.

measurement specific to the screen in question. The combination of these two factors suggests that the only way to ensure a successful high-throughput siRNA screen is to run a large number of siRNAs a number of times. One way to partially circumvent this is to address the biological variance inherent in genome-wide analysis through measurement of a complex, multi-dimensional phenotype/readout.

BIOLOGICAL VARIANCE AND COMPLEX READOUTS

Biological variance is present in all research conducted in living systems. Researchers in the life sciences are generally aware of how to control for this; however when examining ten to twenty thousand genes, the total variance on the readout of interest can be massive. Because of the connectedness of biological pathways, genes affecting the screen readout of interest are expected to fall into one of three basic categories: direct effects on process, indirect effect on process, and effects on screen readout through unrelated processes. Genes exhibiting a direct effect on the process under study are frequently the smallest class, though they are of the highest interest. Genes having an indirect effect on the process of study (through mechanisms such as control of mRNA levels, intracellular protein transport, protein phosphorylation, etc. which in turn affect the genes directly involved) can also be of interest in determining the overall physiology involved. However the third class of genes, those not involved in the pathway of interest but affecting the screen readout through unrelated mechanisms will be the focus of this section. These genes, which are false positives, present a challenge different from those due to error of measurement. While those false positive that result from error of measurement can be weeded out in a secondary/confirmation screen, false positives resulting from true biological effects on unrelated pathways affecting the screen readout cannot. These will be highly reproducible and difficult to detect, thus they pose the greatest threat for wasting time and resources in follow-up work.

In the siRNA screen conducted by Majercak and others,[12] an HEK cell line expressing an *APP* transgene driven by a CMV promoter was used to detect the effects of siRNA gene knock down on APP processing. In the primary HTS, several hundred genes were found to reduce amyloid levels with potency equal to or greater than the known secretases upon siRNA knock down. However, when the majority of these hits were compared to the known secretase positive controls, an important difference was noted. The known secretases did not drive all APP metabolites down simultaneously; specifically sAPPα was found to be unaffected or increased (Fig. 2A). When all metabolites of APP were examined for the hits in the primary screen, the majority (>90%) showed a different profile; one in which all metabolites were decreased (Fig. 2B). This suggested some form of experimental artifact, which in this case was traced back to the CMV promoter used to produce transgenic over-expression within the HEK cells. CMV-driven transgenes are frequently used in research such as this, to increase the levels of the protein of interest to easily readable levels. In doing so, however, the researcher has introduced a non-native aspect to the system; one that can be manipulated via siRNA. The result is that genes affecting the promoter can be easily confused with those affecting APP processing. Upon inspection, hits which reduced all metabolites (or pan-reducers) were shown to decrease *APP* expression, but importantly only *APP* expression driven by the CMV promoter (not endogenous APP expression).[12] These genes represent the most insidious type of false positives; they have a robust effect on the primary readout, which is readily reproducible in follow-up screens and assays in the same cell line. Likewise, these false positives do not necessarily all come from the same artifact; siRNAs which globally effect protein secretion could be expected to have a similar readout even when they do not affect the process of interest (APP metabolism) either directly or indirectly.

There is no set way to deal with false positives resulting from this form of biological variance (i.e., variance in the data resulting from siRNAs affecting unrelated pathways altering the readout only while not affecting the process being studied, as opposed to variance in the size of the effect on the pathway of interest). The tests necessary to detect these will vary from screen to screen, and will depend on the specifics of the cell type being used. One general principle, however, can be taken from recent biomarker and microarray profiling work. In general, a readout that is more complex and multi-dimensional will be more difficult to mimic via an alternate biological pathway than one consisting of a single measurement. In the Majercak et al study,[12] this was accomplished by measuring four separate APP metabolites; true positives altering processing changed the ratio of these readouts, and false positive pan-reducers were easily removed.

In screens measuring levels of proteins of interest, any gene affecting overall mRNA transcription or translation, cellular viability, cell cycle, trafficking, etc can show up as a hit. This concept is not new; in clinical medicine a diagnosis is rarely made on the basis of a single readout. Multiple tests are performed to rule out various other conditions until the final diagnosis is reached. In high throughput RNAi screen this could be done serially, with a number of follow-up screens to determine the "true" hits; however as a general rule when more information can be taken from the initial assay it should be used to better classify the resulting data. In screens with up to 20,000 genes being examined, there are likely to be hundreds if not thousands of potential hits, most of which are expected to be false positives.

While high-throughput RNAi screens are a powerful tool and will undoubtedly elucidate the specific genes in poorly understood molecular pathways involved in many unmet medical conditions, they also pose numerous pitfalls which could cause users to spend considerable resources unnecessarily. A basic understanding (and control/reduction when possible) of the variance resulting from both measurement and the biology of the system specifically employed allows the researcher to fully utilize these libraries. We suggest 2 rules of thumb when designing a genome wide screen: (1) Determine the error of measurement in the system, and select cutoffs based on this level. By using the positive and negative controls employed on all 96- or 384-well plates, researchers can have several hundred runs of multiple genes. Assuming this level of measurement error for "true" hits, the researcher can then determine the power to detect hits of a given magnitude at any given level, and cutoffs for secondary screening are chosen accordingly. (2) When possible, use a complex readout to control for biological variance resulting from genes involved in pathways which affect parts of the screen readout but not the process of interest. Screens with a single readout are more likely to have a large percentage of false positives due to alternative pathways affecting the readout. When multiple variables are employed, it is less likely that a given pathway will alter all of them is the same direction as the pathway of interest, resulting in fewer "repeatable" false positives.

Addressing the variance in this manner will allow for the maximization the number of biologically relevant siRNA hits while reducing false positives and false negatives to an acceptable level.

References

1. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburtty K, Simon J, Bard M, Friend SH. Functional discovery via a compendium of expression profiles. Cell 2000; 102:109-26.
2. Gunther EC, Stone DJ, Rothberg JM, Gerwien RW. A quantitative genomic expression analysis platform for multiplexed in vitro prediction of drug action. Pharmacogenomics J 2005; 5:126-34.
3. Gunther EC, Stone DJ, Gerwien RW, Bento P, Heyes MP. Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro. Proc Natl Acad Sci USA 2003; 100:9608-13.
4. Slatter JG, Templeton IE, Castle JC, Kulkarni A, Rushmore TH, Richards K, He Y, Dai X, Cheng OJ, Caguyong M, Ulrich RG. Compendium of gene expression profiles comprising a baseline model of the human liver drug metabolism transcriptome. Xenobiotica 2006; 36:938-62.
5. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 2006; 313:1929-35.
6. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of *escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol 2007; 5:e8.
7. Gonczy P, Echeverri C, Oegema K, Coulson A, Jones SJ, Copley RR, Duperon J, Oegema J, Brehm M, Cassin E, Hannak E, Kirkham M, Pichler S, Flohrs K, Goessen A, Leidel S, Alleaume AM, Martin C, Ozlu N, Bork P, Hyman AA. Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. Nature 2000; 408:331-6.
8. Lee SS, Lee RY, Fraser AG, Kamath RS, Ahringer J, Ruvkun G. A systematic RNAi screen identifies a critical role for mitochondria in *C. elegans* longevity. Nat Genet 2003; 33:40-8.
9. Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, Kamath RS, Fraser AG, Ahringer J, Plasterk RH. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol 2003; 1:e12.
10. Sonnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, Brehm M, Alleaume AM, Artelt J, Bettencourt P, Cassin E, Hewitson M, Holz C, Khan M, Lazik S, Martin C, Nitzsche B, Ruer M, Stamford J, Winzi M, Heinkel R, Roder M, Finell J, Hantsch H, Jones SJ, Jones M, Piano F, Gunsalus KC, Oegema K, Gonczy P, Coulson A, Hyman AA, Echeverri CJ. Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. Nature 2005; 434:462-9.
11. Friedman A, Perrimon N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. Nature 2006; 444:230-4.
12. Majercak J, Ray WJ, Espeseth A, Simon A, Shi XP, Wolffe C, Getty K, Marine S, Stec E, Ferrer M, Strulovici B, Bartz S, Gates A, Xu M, Huang Q, Ma L, Shughrue P, Burchard J, Colussi D, Pietrak B, Kahana J, Beher D, Rosahl T, Shearman M, Hazuda D, Sachs AB, Koblan KS, Seabrook GR, Stone DJ. LRRTM3 promotes processing of amyloid-precursor protein by *BACE1* and is a positional candidate gene for late-onset Alzheimer's disease. Proc Natl Acad Sci USA 2006; 103:17967-72.
13. Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, Teplow DB, Ross S, Amarante P, Loeloff R, Luo Y, Fisher S, Fuller J, Edenson S, Lile J, Jarosinski MA, Biere AL, Curran E, Burgess T, Louis JC, Collins F, Treanor J, Rogers G, Citron M. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. Science 1999; 286:735-41.