Title of proposed program: Publicly Available siRNA Libraries for the Unrestricted Release of Screening Data and Corresponding Sequence Information

Submitting Source: NIH

What is the major obstacle/challenge/opportunity that the Common Fund should address? What would the goals of the program be? siRNA screening has become a powerful tool to interrogate gene function. However, RNAi screening suffers from severe limitations that leave the promise of this technology woefully unfulfilled. For example, off-target effects dominate RNAi screens and make the data difficult to interpret. This is evidenced by the observation that there is little correlation between different siRNAs designed to target the same gene in any given assay. Recently, a number of studies have shown that off-target effects can be identified, and even interpreted, through a global understanding of how siRNA sequences influence an assay. Unfortunately, sequence data for commercial libraries cannot be released, rendering subsequent analysis of these data by the community of bioinformatics researchers impossible, and thus severely limiting the contribution of genome-wide siRNA studies to the understanding of genome function in health and disease. To overcome these obstacles, we propose the creation of publicly available siRNA libraries along with open access to corresponding sequence information. The goals of such a program would be to:

1. Design and obtain siRNA libraries (beginning with human and mouse) built on a state-of-the-art understanding of RNAi biology and empirical data from actual large-scale screens. Custom libraries will allow for the unrestricted release of sequence information.
2. Via NIH or third-party distributors, make the libraries available to the research community.
3. Release and archive all screen data and corresponding sequences obtained with these libraries.

Why is a trans-NIH strategy needed to achieve these goals? What initiatives might form the strategic plan for this topic? This type of program would benefit all areas of research and advance our understanding of gene function in the context of disease and fundamental biology. A broad, trans-NIH strategy is needed to gather the resources and expertise necessary to realize the benefits of such a program. This plan would require a consortium to help design and produce the libraries, and would nucleate a community to utilize and for the first time publically release the data generated.

If a Common Fund program on this topic achieved its objectives, what would be the impact? It is sobering to realize that, 10 years after completion of the Human Genome Project and 6 years after the Nobel Prize was awarded for RNAi, there is no publically available compendium of genome-wide siRNA data, since the siRNA sequence data are proprietary. This program would solve that problem and thus catalyze a new field of bioinformatics based on these data, enabling secondary and meta-analyses of siRNA screens and thus delivering unprecedented insights in systems biology and an ever-improving understanding of how siRNA/miRNAs influence cellular processes relevant to physiology and disease.