Cellular phenotyping by RNAi

Florian Fuchs and Michael Boutros

Advance Access publication date 23 February 2006

Abstract

A systematic characterization of genes with unknown function is a key challenge after the sequencing of the human genome and the genomes of many model organisms. High-throughput RNA-interference (RNAi) screenings have become a widely used approach in invertebrate model organisms and also promise to revolutionize cell biology in mammals. Genome-wide RNAi screens in *Caenorhabditis elegans* and *Drosophila*, and in a smaller scale in mammalian cells have proven to be a valuable and successful method for the dissection of diverse biological processes. A number of RNAi libraries have become available that rely on different technologies, such as long double-stranded (ds) RNAs, *in vitro* diced short-interfering (si) RNAs, synthetic siRNAs and short-hairpin (sh) RNAs, which all have specific advantages and disadvantages. In addition, progress in screening technologies and data analysis allows the adaptation of screening methods to analyse more complex cellular processes. This review will summarize strategies in combining genome-scale RNAi libraries, high-throughput screening technologies, integrated high-content data analysis and will discuss future challenges.

Keywords: RNAi screening; functional genomics; high-throughput cell-based assays

INTRODUCTION

Gene silencing by the conserved RNAi-pathway as first described in *Caenorhabditis elegans* is initiated by the introduction of double-stranded RNAs (dsRNA), resulting in sequence-specific degradation of homologous endogenous mRNA [1]. Later, RNAi has been shown to function in a similar manner in every metazoan and has been applied to study a wide variety of phenotypes *in vivo* and in cells. Experiments in *C. elegans* and *Drosophila* primarily make use of long dsRNAs, which are intracellularly diced into functional 21mer short-interfering RNAs (siRNAs). However, long dsRNAs are not effective in most mammalian cells due to the induction of antiviral pathways that lead to host cell shutdown. This can be circumvented by transfecting synthetic or plasmid-encoded siRNA that in most cases do not elicit an interferon or other host cell response [2]. These approaches have been effectively used to study many different biological pathways in loss-of-function analysis, and several large-scale efforts have recently generated libraries that target every predicted gene in major model organisms and humans.

RNAi libraries in model organisms and mammalian cells

To construct large-scale RNAi libraries, several strategies have been pursued that rely on the use of long dsRNA, chemically synthesized siRNAs or expression of short-hairpin RNAs (shRNAs). RNAi libraries in *C. elegans* and *Drosophila* are mainly based on long dsRNAs, which are either synthesized *in vitro* or for experiments in *C. elegans* expressed in *Escherichia coli* that are fed to worms and elicit efficient gene silencing. Several libraries that target the majority of genes have become available, and are in part distributed through public resource centres. In mammals, *in vitro* processing of long dsRNA by recombinant RNaseIII or Dicer can be used to generate siRNA pools that provide efficient and specific silencing effects without induction of an interferon response [3]. Chemically synthesized siRNAs have high-transfection efficiencies, but are relatively expensive and can result in non-specific off-target effects [4]. In addition, vector-based expression of shRNAs has been used to silence gene expression from transiently or stably transfected vectors [5]. This approach can also be adapted to

Corresponding author. Michael Boutros, Boveri-Group Signaling and Functional Genomics, German Cancer Research Center, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany. Tel: +49 6221 42-1951; Fax: +49 6221 42-1959; E-mail: m.boutros@dkfz.de

Florian Fuchs and Michael Boutros are members of the Boveri-Group on Signaling and Functional Genomics at the German Cancer Research Center.
lenti- or retroviral vectors to silence gene expression in cells that are otherwise difficult to transfect [6]. Both siRNA and shRNA libraries have been successfully used to screen human cells for various phenotypes [7, 8]. Several computational tools have been developed for the assessment of efficiency and specificity of RNAi probes [9, 10], which provide means to evaluate or to rationally design long dsRNAs and siRNAs. Currently available libraries for large-scale RNAi studies in model organisms and mammalian cell culture are summarized in Table 1. As algorithms to predict efficient and specific siRNA sequences will be further improved, it is likely that currently available libraries will be modified and updated also to include changes in genome annotations.

**Simple, complex and more complex assays**

In classical genetic approaches, mutants are generated by chemical or transposon-mediated mutagenesis, selected for a certain phenotype and analysed by mapping the gene mutation contributing to this phenotype. These screens have been successfully employed in particular in invertebrate model organisms to dissect many conserved cellular pathways [11]. With the sequencing of many genomes and development of RNAi-based approaches that silence gene expression in a sequence specific manner, reverse genetics that start with a particular gene and determine the effect following its disruption, are now more feasible for a wide variety of organisms that lack classical genetic approaches. A typical workflow of a genome-wide RNAi screen is depicted in Figure 1, which requires a comprehensive RNAi library, a suitable cell-based assay system monitoring phenotypic changes, data analysis procedures to identify ‘hits’, re-testing and integration of screening data with meta-data from public databases.

Cell-based readout systems of genetic screens already performed using RNAi libraries can be grouped to focused reporters [12–16] and high-content, such as microscope imaging readout [3, 17–20] assays. Specificity and sensitivity of any reporter system that is used to analyse individual genes for their function in the pathway of interest, needs to be evaluated using appropriate negative and

<table>
<thead>
<tr>
<th>Organism</th>
<th>Description</th>
<th>References</th>
<th>Web resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>Long dsRNA generated from the DGC EST collection</td>
<td>Lum et al. [14]</td>
<td></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>Long dsRNA generated by <em>in vitro</em> transcription</td>
<td>Foley and O’Farrell [26]</td>
<td>See also <a href="http://rnai.dkfz.de/nextrnai">http://rnai.dkfz.de/nextrnai</a></td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>RNAi library of ~16 000 dsRNA designed against BDGP4 and additional gene-models</td>
<td>–</td>
<td><a href="http://rnia.dkfz.de/nextrnai">http://rnia.dkfz.de/nextrnai</a></td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Collection of ~18 000 dsRNAs</td>
<td>–</td>
<td><a href="http://www.ambion.com">http://www.ambion.com</a></td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Library of retroviral vectors encoding shRNAs targeting 28 500 human genes</td>
<td>Paddison et al. [8]; Da Silva et al. [27]</td>
<td><a href="http://codex.cshl.edu">http://codex.cshl.edu</a>; <a href="http://www.rragnal.org">http://www.rragnal.org</a></td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Genome-wide siRNA library covering 21 125 human genes (Dharmacon)</td>
<td>–</td>
<td><a href="http://www.ambion.com">http://www.ambion.com</a></td>
</tr>
<tr>
<td>H. sapiens</td>
<td>Genome-wide siRNA library</td>
<td>–</td>
<td><a href="http://www.ambion.com">http://www.ambion.com</a></td>
</tr>
</tbody>
</table>
positive controls. The quality of focused reporter assays, such as fluorescence or luminescence readouts, is determined by their variability and the dynamic range of the signal. A comparison reflecting advances and limitations of single-readout systems and high-content assays is shown in Table 2.

**Figure 1**: Typical workflow in genome-wide RNAi screening and data analysis.

**Table 2**: Comparison of single-readout experiments and high-content assays

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-throughput screens using luminescence or fluorescence readouts</td>
<td>High-content RNAi analysis using automated microscopy</td>
</tr>
<tr>
<td>Simple automation</td>
<td>Simple data management</td>
</tr>
<tr>
<td>Speed of screening is high</td>
<td>Potentially higher false positives</td>
</tr>
<tr>
<td>Easily adaptable to specific read-outs</td>
<td>Sensitivity depends on reporter construct</td>
</tr>
<tr>
<td>Simple data management</td>
<td></td>
</tr>
</tbody>
</table>

Advantages: Simple automation, speed of screening is high, easily adaptable to specific read-outs.

Disadvantages: Potentially higher false positives, sensitivity depends on reporter construct.

High-throughput screens using luminescence or fluorescence readouts:
- Simple automation
- Speed of screening is high
- Easily adaptable to specific read-outs
- Simple data management

High-content RNAi analysis using automated microscopy:
- High sensitivity
- Multiple cellular descriptors can be analysed (e.g., DNA content, number of nuclei, cell morphology)
- Spatial and temporal resolution
- Low false positives
- Analysis on single cell and subcellular level
- High costs (e.g., specific antibodies)
- Assay development is rather complex
- Throughput is often limited
- Automated image analysis needs to be adapted to every assay
- Large amounts of data

**High-content assays using RNAi libraries**

In contrast to focused reporter assays, high-content screening approaches by microscopy generate a very large amount of data that pose significant challenges for data storage and analysis. Image analysis software can be applied to allow determination of parameters such as the number of cells or nuclei, nuclear shape or DNA content per cell in single-cell analysis, but in many cases needs to be further enhanced to perform tasks in an automated fashion.

Given that appropriate probes are available, fluorescence microscopy is highly sensitive and many phenotypic changes can be detected that remain unnoticed in focused reporter assays. A combination

...
of compound or genome-wide RNAi libraries and automated microscopy is well suited to monitor phenotypic changes of several cellular descriptors in space and time. Recently, a parallel screening strategy aiming at the identification of both, small molecules that inhibit cytokinesis and genes that are required for this biological process, was applied to identify targets and their potential inhibitors [18]. These experiments identified 214 genes required for cell division, many of which were previously uncharacterized.

Another approach has provided a way for adaptation of high-content assay technology to high-content screening [3]. In these experiments, a genome-scale endoribonuclease-prepared siRNA screen in human cells identified genes essential for cell division using a two-step screening strategy. In the first step, a high-throughput cell viability test identified 275 candidate genes. These candidates were further analysed with a second time-lapse microscopy-based high-content approach. Thirty-seven out of 5305 genes analysed, were identified to be essential for cell division. These studies very impressively showed the feasibility of profiling complex cellular phenotypes using automated microscopy.

**Future perspectives**

Currently available RNAi libraries, in particular for mammalian cells, are still limited by potential off-target effects, limited gene silencing efficiency and often incomplete genome coverage. It is likely that a more detailed knowledge about the mechanism of RNA interference will improve currently available prediction algorithms with regards to silencing efficiency and target specificity.

The application of genome-wide RNAi libraries in high-content assays is still often limited by technical challenges in the acquisition, managing and analysis of data. Image acquisition often offers only a limited throughput, whereas higher-throughput scanners only offer low optical resolutions. Similarly, improvements in image analysis software are foreseeable that allow the automated analysis of more cellular descriptors than currently feasible. Furthermore, tools for organizing and visualizing multidimensional data will be required to make both raw and processed image data publicly available.

Systematic phenotyping by RNAi and other approaches will provide new perspectives on a gene’s function in the context of the genome on a gene-by-gene level. The integration of phenotypic information from other functional genomic datasets will allow to dissect many important cellular processes with an unprecedented spatial and temporal resolution.

---

**Key Points**

- RNA interference is a powerful tool for gene discovery and functional characterization.
- Genome-wide RNAi screening is now feasible for many organisms.
- Large-scale phenotypic screens require new approaches for computational analysis.

---

**References**


