

# Oncology studies using siRNA libraries: the dawn of RNAi-based genomics

Christoph Sachse<sup>1</sup> and Christophe J Echeverri<sup>\*1</sup>

<sup>1</sup>Cenix BioScience GmbH, Tatzberg 47, 01307 Dresden, Germany

**High-throughput, human cell-based applications of RNA-mediated interference (RNAi) have emerged in recent years as perhaps the most powerful of a ‘second wave’ of functional genomics technologies. The available reagents and methodologies for RNAi screening studies now enable a wide range of different scopes and scales of investigation, from single-parameter assays applied to focused subsets of genes, to comprehensive genome-wide surveys based on rich, multiparameter readouts. As such, RNAi-based screens are offering important new avenues for the discovery and validation of novel therapeutic targets for several disease areas, including oncology. By enabling a ‘clean’ determination of gene function, that is the creation of direct causal links between gene and phenotype in human cells, RNAi investigations promise levels of pathophysiological relevance, efficiency, and range of applicability never before possible on this scale. The field of oncology, with its many assays using readily transfectable cell lines, has offered particularly fertile ground for showcasing the potential of RNAi-based genomics. However, like any other technology before it, RNAi is not without its own challenges, limitations, and caveats. Many of these issues stem directly from the choice of silencing reagent to be used in such studies, and the design of the overall screening strategy. Here, we discuss the basic design issues, potential advantages, and technical challenges of large-scale RNAi screens based on the use of chemically synthesized siRNA libraries.**

*Oncogene* (2004) 23, 8384–8391. doi:10.1038/sj.onc.1208072

**Keywords:** RNA interference; gene silencing; siRNA; functional genomics; screening

## The advent of RNAi-based genomics

The completion, in recent years, of key animal genome sequencing projects, including the human, mouse, and rat genomes, has ushered in the ‘postgenomic’ era, with its major challenge of developing novel, efficient strategies to best exploit this unprecedented wealth of raw biological data. The so-called ‘functional genomic’ and ‘proteomic’ technologies, including microarray-based transcriptional profiling, a wide

variety of bio-informatics mining tools (i.e. ‘*in silico* biology’), and large-scale efforts to map protein interaction networks, represent the first successes in meeting this challenge. However, while these technologies have indeed yielded countless new insights into a wide array of biological processes, their impact on the discovery and development of new therapeutic drugs, both in oncology and other disease areas, has failed to meet expectations.

Most major drug developers have embraced these advances at significant cost in recent years, hoping for a bounty of ‘smarter’ targets that would significantly reduce attrition rates during the drug development process. Instead, although the functional insights afforded by this first wave of technologies have been compelling enough to foster countless academic follow-ups, they have all-too-often fallen well short of the pathophysiological relevance needed to warrant their further pursuit in drug development programs. For example, the genes identified by microarray analyses as being over- or underexpressed in cancerous biopsies may offer excellent potential as diagnostic biomarkers, but the inherent inability of this technology to determine whether these differences are causal or consequential to the pathological state leaves their therapeutic potential too far from certain. The result, then, has been a huge accumulation of ‘candidate drug targets’ emerging from academic as well as industry labs, most of which are now stalled in industry pipelines, requiring significant further validation work before they can be followed-up on. Meanwhile, attrition rates have not substantively improved, suggesting that the new, first wave of so-called ‘genomics targets’ may still not be ‘smart’ enough.

The inability of first-wave functional genomics technologies to achieve the quality and detail of analyses normally afforded by conventional laboratory investigations has been the main hurdle here. As a result, some disillusioned drug developers are now pulling away from genomics-based discovery endeavours, a particularly sad outcome, since this happens just at the time when the tools they had been hoping for are finally emerging. The best, most direct way of analysing gene function and, at the same time, predicting the action of an eventual drug targeting this gene is to inhibit either the activity of the targeted protein or the expression of the underlying gene, and then analyse the resulting loss-of-function phenotypes. Until recently however, high-throughput (HT) gene silencing at the genome scale

\*Correspondence: CJ Echeverri;  
E-mail: echeverri@cenix-bioscience.com

was not feasible in mammalian systems (despite some heroic, although ultimately failed, efforts from the biotech sector), restricting such direct tests of gene function to the domain of model organisms. In view of this, the excitement generated by the discovery of RNA-mediated interference (RNAi), and particularly by its HT, genome-scale applications in human cells, should come as no surprise. The dizzyingly fast progression of this technology, from intriguing observations in plants (Mueller *et al.*, 1995; English *et al.*, 1996), to powerful experimental applications in *Caenorhabditis elegans* (Fire *et al.*, 1998), and now to genome-scale screening in human cells, has bred sometimes-exuberant optimism among academic and industry researchers alike. What emerges clearly, nonetheless, is that RNAi, the experimental ‘hijacking’ of a powerful, highly conserved gene silencing pathway functioning primarily through the targeted destruction of individual mRNAs, now represents arguably the most powerful second-generation functional genomics technology available to date (Carpenter and Sabatini, 2004).

The feasibility of applying RNAi as a systematic genome-scale screening method was first demonstrated early on in *C. elegans* (Gönczy *et al.*, 2000; Fraser *et al.*, 2000; also further reviewed in this issue), and soon thereafter in *Drosophila* cells (Clemens *et al.*, 2000; Kiger *et al.*, 2003; Lum *et al.*, 2003; also further reviewed in this issue). At that time, the hope of applying RNAi in vertebrate systems remained doubtful in view of the well-known interferon-based stress response elicited there by the presence of double-stranded RNAs (dsRNAs), the only triggering molecule then known for RNAi. Thankfully, the first solution to this problem was demonstrated in 2001 by Tuschl and colleagues (Elbashir *et al.*, 2001a). Their success came from the use of chemically synthesized dsRNA molecules designed to mimic the small size and specific structure of the so-called short interfering RNAs (siRNAs), initially identified in plants as apparent intermediates in the RNAi pathway (Hamilton and Baulcombe, 1999). Others have since extrapolated on this theme, most notably demonstrating comparable success using vector-based expression of short hairpin RNAs (shRNAs) to trigger RNAi (reviewed elsewhere in this issue). The result of Tuschl’s breakthrough has been nothing short of a revolution. Most researchers, from diverse fields including but certainly not limited to oncology, have been quick to experience success with small-scale applications of RNAi, focusing on handfuls of their ‘favourite’ genes. Those aiming for larger, even genome-scale applications, while also seeing the huge new potential, quickly recognized the new challenges inherent to this pursuit, starting with the crucial selection of RNAi silencing reagents, the optimization of screening protocols, and the refinement of the readout strategies. The present article discusses the exciting potential, but also those new challenges, associated with RNAi-based screening in human cells using chemically synthesized siRNA libraries, and how these are being addressed in the study of cancer.

## Designing siRNAs/shRNAs to target an entire genome

Whether using an siRNA or shRNA-based approach, the need to avoid the interferon response in mammalian cells has required the development of new criteria and new bio-informatics tools to select optimal targeting sequences. In the vast majority of cases to date, these have been designed as 21-mers, integrating a 19-mer duplex core and two nucleotide overhangs at both 3’ ends (for siRNAs; for shRNAs, see elsewhere in this issue). Applying the first wave of publicly available design parameters, approximately 50–60% of siRNAs were found to be capable of reducing target mRNA levels by at least 70% after 48 h, a success rate that was more than satisfactory for most small-scale applications. However, the building of genome-scale siRNA libraries demanded a significantly higher success rate, therefore spawning multiple efforts to define superior design rules. The relative complexity of the problem led several groups, including our own, to generate large experimental data sets of siRNA-derived silencing efficacies measured under tightly standardized conditions, culminating in the empirical development of several siRNA design algorithms.

These algorithms typically comprise a wide range of different criteria to maximize silencing efficacy, which, in our case and others (Reynolds *et al.*, 2004), have included specific base compositions at key positions along the 19 core siRNA base pairs, thermodynamic base-pairing profiles defining ‘regional base compositions’ (GC content in particular), base composition of 3’ overhangs, positions along the targeted mRNA, and lack of variability of the targeted mRNA over the relevant site (avoiding known SNPs, etc.). One of the most potent design features implemented in our own algorithm as well as others is the creation of a strong differential between the two ends of the siRNA duplex region in base-pairing thermodynamics, insuring a significantly weaker pairing of the 5’ end of the antisense strand. As recently documented by Zamore and co-workers (Schwarz *et al.*, 2003), such asymmetry strongly favours the ‘loading’ of the antisense strand over that of the sense strand into available RISC complexes, thereby insuring recognition and subsequent destruction of the correct target mRNA by the largest possible proportion of the available RISC pool.

Ultimately, no matter how good the *in silico* predictions look, the silencing performance of the algorithm-designed siRNAs must be demonstrated experimentally in cultured human cells under strictly standardized conditions. In our experience, such tests should ideally measure silencing of endogenously expressed genes, applying a suitable quantitative analysis of target mRNA levels (e.g. real-time RT-PCR or branched DNA-based methods), and integrating large enough numbers of genes so as to generate a statistically relevant sample. In our own case (others have reported generally comparable results), the resulting success rates have indicated that ~82% of these first-generation algorithm-designed siRNAs yield more than 70%

silencing after 48 h, enough to build up a first generation of genome-scale siRNA libraries.

The systematic application of these algorithms over the entire genome, or more precisely, the entire transcriptome sequence of human, mouse, and rat, is, in itself, not without its own complexities. One key pitfall of this strategy has been its dependence on the science of gene structure predictions, which remains very much in rapid evolution today. Indeed, many gene predictions are still changing significantly, as more experimental data surface to correct the *in silico* predictions, thus requiring regular updating of RNAi libraries to keep up with the new annotations. The alternative, namely the use of cDNA libraries as a target source, avoids this issue by insuring that all silencing reagents are necessarily targeting expressed gene sequences. Although arguably of 'genome-scale', this precludes the achievement of anything close to genome-wide coverage, and biases such libraries heavily towards those highly expressed genes that are best represented in the cDNA library. Of course, this bias can be used to significant advantage in those cases where the scope of the screen is intentionally focused on a subset of genes that may be preferentially enriched in the said cDNA library or otherwise preselected subset. For example, siRNA libraries can be designed to focus specifically on those large collections of genes identified by transcriptional profiling or other methods as being up- or downregulated in cancerous tissues. This represents a particularly powerful application of RNAi screening to follow-up and thereby help in fully realizing the promise of this and other 'first-generation' functional genomics technologies.

#### **Silencing reagents: siRNAs if you can, shRNAs if you must**

The second challenge faced by those defining RNAi screening strategies in human cells has typically been the choice of silencing reagent to be used. Chemically synthesized siRNAs and vector-encoded shRNAs are currently the best-characterized reagents for all scales of RNAi experimentation in human cells. Using viral vector-mediated expression of shRNAs (Arts *et al.*, 2003; Rubinson *et al.*, 2003) offers the primary advantages of facilitating studies in cell types that are otherwise difficult to transfect, and generating RNAi-based silencing in a more sustained manner, beyond the 5- to 7-day transient effect typically afforded by siRNAs (Holen *et al.*, 2002). In our experience, however, most oncology-relevant processes are clearly measurable within that transient assay window, and fortunately, most transformed human cells commonly used in oncology studies can be transfected readily through lipid-mediated delivery (further discussed below). One notable exception outside the realm of large-scale screening but that is of particular interest in oncology is the use of RNAi in cancerous cells to be used for xenograft studies, as these inevitably require silencing to

be sustained over weeks, that is, clearly beyond the scope of siRNAs.

The ability to closely control the concentration of siRNAs delivered to the cells also offers an important advantage over the shRNA vector approach, which inherently lacks this level of control, instead driving the highest possible expression of shRNAs in the infected cells. As further discussed below, this is particularly problematic in view of recent reports that have clearly demonstrated an increased risk of off-target and nonspecific effects when using excessive concentrations of RNAi silencing reagents (Sledz *et al.*, 2003; Persengiev *et al.*, 2004). In addition, many researchers have noted significantly higher variability in silencing efficacies exhibited by shRNAs *vs* siRNAs, leading to the common observation of weaker or less detectable phenotypes. Finally, in those cases where the high transduction efficiencies afforded by viral approaches cannot be realized, the use of plasmid-based shRNA strategies becomes particularly limiting, as their typically low and variable transfection efficiencies all but preclude biochemical readouts of any sort. While these issues clearly do not preclude the overall use of vector-based shRNA libraries for RNAi screening (as reviewed elsewhere in this issue and already successfully demonstrated: Arts *et al.*, 2003; Berns *et al.*, 2004; Paddison *et al.*, 2004), they do limit their breadth of applicability, making them best adapted for groups who are unconcerned with the issue of being comprehensive, and who accept that some phenotypes may well be missed.

Those groups, however, who do aim for maximally comprehensive genome coverage, have focused on the use of chemically synthesized siRNAs. As the best-characterized reagents in the field to date, these 'gold standards' can be rapidly designed and synthesized from available genome sequence data to target literally any predicted gene (pre-designed and custom-designed siRNAs are now available from several vendors to target virtually any human, mouse, or rat gene). When optimally designed and applied, siRNAs offer the highest, most robust silencing performance currently achievable. It must be noted however that their cost, although posing no significant barrier for low- and medium-scale applications, is currently limiting the widespread adoption of siRNAs for genome-scale work by academic groups who must invariably apply for grant funding to buy their own libraries, or to subcontract the screening work to specialists.

#### **RNAi specificity and off-target effects: a moot point with the right controls**

Some groups have detected complex sequence-dependent 'off-target' effects in RNAi experiments, at the mRNA level (Jackson *et al.*, 2003; Persengiev *et al.*, 2004; Scacheri *et al.*, 2004; Snove and Holen, 2004). These results come in contrast to earlier studies, which indicated that as few as two mismatches could completely disrupt the silencing efficacy of an siRNA

(Elbashir *et al.*, 2001b). Thus, while the general understanding of these observations remains incomplete, it does suggest that many siRNAs in use today may not only recognize the intended perfectly matched target mRNA, but may also direct the destruction of other, imperfectly matched ‘secondary target’ mRNAs, although almost always to a significantly lesser extent. These sequence-dependent off-target effects are proving difficult to predict, as the underlying stretches of base-pair complementarity have not exhibited clear thresholds in size, composition, or other obvious patterns. Furthermore, the top BLAST ‘hits’ for a given siRNA do not necessarily represent the most likely secondary targets. Nonetheless, the sequence dependence of this type of off-target effect can be used to completely and easily neutralize this issue. Since each individual siRNA’s off-target ‘footprint’ is defined by its sequence, one can quickly ascertain the gene specificity of any observed phenotype simply by using multiple distinct siRNAs targeting the same gene of interest. Despite the continuing efforts to further optimize siRNA targeting specificities, it is expected that this simple control will always be advisable, and most likely always be required for publication.

A second type of off-target effect that has been noted in RNAi experiments consists in the concentration-dependent modulation of nontargeted stress response genes, including factors of the interferon response pathway (Sledz *et al.*, 2003; Persengiev *et al.*, 2004). While still poorly understood also, these effects are currently thought to be dependent on such factors as cell type and delivery method, and, unlike the sequence-dependent off-target effects noted above, their risk is known to be significantly increased by the use of excessive concentrations of silencing reagents (Jackson *et al.*, 2003; Persengiev *et al.*, 2004). Thus, titration of the siRNA concentration (e.g. from 100 nM down to about 10 pM) has the potential of attenuating or even eliminating these effects, rare though they may be. Also, one or several ‘scrambled’ or ‘unspecific’ siRNAs, that is, siRNAs that do not exhibit significant matches to any gene of the targeted genome, should be included in all RNAi experiments to control for this issue. In fact, in essentially all RNAi screens, the vast majority of tested genes yield negative results, and therefore provide a huge abundance of baseline control values for this issue.

#### **Of plates and barcodes: making siRNA-based assays ‘screenable’**

In view of the typically high costs and intense efforts needed to conduct systematic, genome-scale siRNA-based screens, considerable work is usually invested first and foremost in making the basic biological test as robust, as sensitive, and as meaningful as possible. This starts with a demonstration of the assay principle at normal laboratory scale using siRNAs targeting known positive control genes for the process of interest. A wide range of known regulators of cell cycle progression, proliferation, and apoptosis are commonly used in

oncology studies. Next, the same principle must be adapted and optimized for implementation in a highly parallelized format (e.g. 96- or 384-well plates are most often used to get statistically relevant numbers of cells within each well), to achieve a throughput that is suitable for the intended screening scope. The assay optimization efforts usually focus on the siRNA delivery method, the kinetics of the assay’s readout (i.e. defining the time point(s) for the readout(s) after siRNA delivery), the cell culture conditions, and importantly, the overall logistics including data flow management.

Oncology screens in transformed human cell lines are particularly well served by the siRNA-based approach, as delivery of these reagents using a wide variety of commercially available cationic lipid formulations typically yields very high transfection efficiencies (>95%) with low toxicity. What has emerged clearly from the field so far is the value of methodically testing several of these ‘lipofection’ reagents with each new cell line, to identify the best choice for that line. Primary cells have proven more difficult to deliver into using lipofection methods, leading many to explore viral-based shRNA methods, and stimulating renewed intense efforts to solve this problem. Although electroporation-based protocols have shown some success, particularly in cells of haematopoietic lineage, parallelized, HT applications of this technology remain problematic, being, at best, in their infancy.

The kinetics of target loss of function in RNAi experiments are inevitably quite slow when compared to compound-based assays, relying first on the depletion of the targeted mRNA pool, and then on the natural – uncontrolled – turnover of the encoded protein. As a result, the optimization of RNAi screening assays almost always benefits greatly from detailed time-course experiments, tracking the progression of observed phenotypes from the subtly hypomorphic to near-null, a process that often yields invaluable functional insights of its own. Ideally, most screening assays should integrate kinetic data in this way, as demonstrated by our own group and others using a time-lapse microscopy analysis of cell division processes in *C. elegans* (Fraser *et al.*, 2000; Gönczy *et al.*, 2000). However, in many cases, this is technically unfeasible, therefore requiring the selection of a single time point as an optimal compromise, at least for the first pass of a screen. With this limitation in mind, it is often advantageous to select a later time point for the first pass, thereby minimizing the risk of missing hits by insuring that most phenotypes have had more time to develop and become detectable. Although, in such cases, the initial phenotypes observed in the first screening pass may be more complex and difficult to interpret – that is, consisting of both primary and secondary biological consequences of the silenced target – these can be characterized in more detail during subsequent screening passes, which deal with much smaller numbers of genes. A quite effective screening strategy therefore emerges as a multipass approach, the first of which is focused primarily on reducing the scale of the problem in a maximally inclusive way (accepting relatively high

false-positive rates to be as comprehensive as possible), followed by 'clean-up' or 'confirmation' rounds to refine the phenotypic analyses and establish target specificities. Ultimately, a true positive hit is considered one that is confirmed with at least two distinct siRNAs, and where the observed phenotype can be directly correlated to an observed reduction in target expression levels in at least two independent experiments.

Finally, logistics often become a key challenge. Insuring a maximally streamlined process not only reduces costs but also minimizes the likelihood of mistakes, and perhaps most importantly, maximizes the standardization of the overall screen, thus insuring that all data, whether emerging over days, weeks, or months, are created, acquired, and analysed in a uniform manner throughout the study. The use of liquid handling robots, bar-coded labelling systems, a thorough system for information management and process tracking (whether electronic or paper-based), and most importantly, a well-structured database system represent the most impactful (but also quite challenging) elements of the typical infrastructure needed for these studies. In striving for adequate level of experimental reproducibility and performance, the monitoring of the 'screening window coefficient', as defined by Zhang *et al.* (1999), during thorough analyses of intraplate, interplate, interoperator, and interexperimental levels of variability has become a widely used and accepted guide.

### The readouts: moving beyond plate readers

Until recently, the vast majority of HT cell-based assays were based on single-parameter readouts, using colorimetric, luminescent, or fluorescent reporters for a given pathway or process, so as to be measured quantitatively in a plate reader. This approach affords very high-screening throughputs and is best adapted for answering very tightly focused questions on single pathways. When trying to gain broader insights into more complex cellular processes, however, researchers have long sought richer, more contextual, multiparametric readouts. In particular, the plate reader approach cannot exploit the huge wealth of information that is accessible through *in situ* analyses. Microscopy-based screening has come of age in recent years, largely through the development and commercialization of automated microscopy instruments by several reputable vendors. As a result, the concept of 'high-content' readouts is gaining increasing support, whereby multiple parameters, including at least four distinct fluorescent markers (antibodies, dyes, or other ligands) can be analysed simultaneously to document their relative intensities as well as their distributions at the subcellular and cell population levels. Although acquisition technologies have progressed very well, yielding good image quality at impressive throughputs, the key bottleneck, as aforementioned, remains the development of adequate automated image analysis algorithms to at least approximate the classification powers of the human brain. It should be noted that the manual analysis and

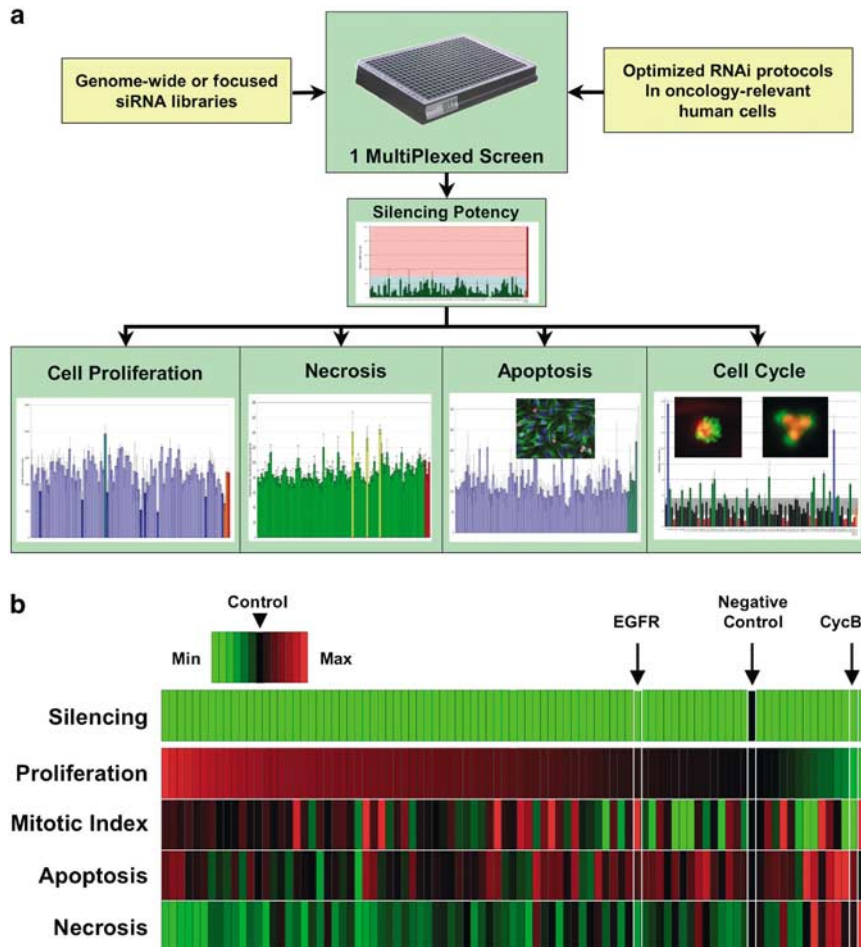
annotation of very large microscopy data sets is not impossible: our own group has indeed completed such a task over approximately 40 000 time lapse microscopy recordings for a genome-wide screen in *C. elegans* embryos (Sönnichsen *et al.*, manuscript). However, if an adequate classification can be realized through an automated approach, this, of course, becomes a highly preferred course, as it will greatly reduce very long timelines otherwise needed and the high risk of introducing human biases that may thwart the standardization of the overall analysis.

### RNAi screening strategies for oncology

The simplest applications of siRNA-based screens for oncology are designed to identify genes whose direct loss-of-function phenotypes result in the slowed proliferation or death of cancerous cells. A number of other oncology-relevant cell-based assays that have been developed for HT work, and that are now being converted to RNAi screens include cell migration and invasion assays to model metastatic processes, as well as colony formation and tube formation assays for angiogenesis studies, to name but a few.

Running the same screen in parallel in cancerous cells and in normal cells is expected to reveal particularly promising target candidates, and is therefore the basis of many ongoing efforts in this field. Nonetheless, these are still early days for large-scale siRNA-based screens in human cells, mainly due to the fact that high-quality siRNA libraries have only recently become available, and most groups are only now in the process of carrying out their first exploratory studies. Consequently, few large-scale siRNA-based screens have yet to reach publication so far, and the few that have are typically focused on gene family-wide or pathway-wide, rather than genome-wide, scopes (Hsieh *et al.*, 2004; Krönke *et al.*, 2004; Xin *et al.*, 2004). More than just cutting costs, such efforts also enrich the therapeutic relevance of their data sets by focusing the screens on 'high-priority genes', known from expression profiling data, for example, to be expressed in commonly occurring tumours, or genes linked to key pathways through *in silico* or proteomics data, or simply, genes predicted to be 'druggable' (Hopkins and Groom, 2002). Another promising application of more focused RNAi screening is the analysis, using appropriately designed cell-based assays, of all predicted ORFs at or near newly discovered disease loci, to efficiently identify the disease gene in question.

Whether the screening scope is comprehensive or more focused, many groups are now striving for more depth of analysis even at HT, mostly by implementing higher content readouts through multiplexed assays (as discussed above). Our own group, for example, has enhanced the ubiquitous proliferation assay by 'multiplexing' it with markers to simultaneously reveal the underlying causes of any observed proliferation effects, that is, necrosis, apoptosis, or deregulation of cell cycle progression (see Figure 1; Sachse *et al.*, 2004). The result



**Figure 1** RNAi-based phenotypic profiling: combining HT screening using siRNA libraries, with high-content readouts. (a) The concept of multiplexing functional readouts in an RNAi screen applying siRNA libraries: by assaying multiple parameters in parallel, in this case, measurements of cell proliferation, mitotic index, apoptosis, and necrosis, a wide range of oncology-relevant questions can be addressed simultaneously, which enables not only the screening application but also the efficient acquisition of precise functional insights into targeted genes. The inclusion of a quantitative analysis of silencing efficacy is crucial to creating a direct link between silenced gene and the observed phenotypes. (b) Sample data set produced by a single experimental run of the above screening paradigm: screen of 88 human kinases (modified from Sachse *et al.*, 2004) using prevalidated siRNAs (Ambion Inc., Austin, TX, USA) in HeLa cells. Validated siRNAs targeting cyclin B1 and EGFR served as positive controls, and an ‘unspecific siRNA’ (‘Negative 1’ from Ambion) served as negative control. Data are shown as a ‘heat map’, illustrating the relative changes compared to the negative control (unspecific siRNA), which was set to 100% (black). Red indicates increasing values and green decreasing values. Data are ordered by the proliferation data

is an extremely efficient screening paradigm, which can be applied over the entire genome or any subset thereof, creating cell line-specific data sets that offer considerable target discovery and validation potential, with minimal time and effort.

Other promising applications of siRNA-based screens include the so-called modifier or enhancer/suppressor screens, whereby existing anticancer compounds can be applied in the context of an RNAi screen, to enable several different strategies. For one, this can help identify new components of important cancer-relevant pathways that are already targeted by the screening compound. The resulting new targets can then be used to develop new stand-alone drugs, or to work towards two-drug sensitization or synthetic lethal strategies to ‘revive’ existing drugs. The same modifier screens can

also generate important new insights into the mechanism of action of the test compounds, an application that is only now beginning to be explored, but that bears very strong potential nonetheless.

#### Outlook: the era of RNAi-based phenotypic profiling begins

One key lesson learned from the first wave of functional genomics technologies was that genomics-based discovery efforts must not only provide more targets but must also come with much clearer data on function and therapeutic relevance. Answering that call, siRNA-based screening now enables smarter discovery screens with a level of comprehensiveness never available

before. Despite some basic caveats, the development of this technology has been amazingly rapid and successful, as it is now possible to create direct, readily interpretable causal links between genes and therapeutically relevant phenotypes at HT, on a genomic scale. In addition to smarter target discovery, the same methods also offer the ideal solution for efficiently extracting maximal value out of existing 'piles' of poorly validated target candidates, piles that are often too large for useful follow-up with most other technologies available today.

Oncology research will also benefit from the development of more cell-based assays to model accurately those cellular processes found to underlie key intervention points for the pathology. The richer the assays, the better: multiparametric or 'high-content' readouts are compatible with high-screening throughputs, and offer a wider window into the biological consequences of experimental perturbations such as gene silencing or drug treatments, alike. The depth of functional insights afforded by a single screen is thereby increased, and the detection of undesirable side effects facilitated early on, when the investment is still relatively modest. Thus the concept of RNAi-based phenotypic profiling emerges: combining HT RNAi with high-content readouts to generate efficiently gene-specific 'signatures' of loss-of-function phenotypes under a wide range of screening conditions.

The fast growing interest in this type of experimentation is propelling several aspects of the underlying technologies forward. One elegant technique was actually pioneered by Sabatini and colleagues before RNAi screening took off (Carpenter and Sabatini, 2004) and is known as retro-transfection or solid-phase optimized transfection (SPOT). This has now been adapted to allow the creation of SPOT-RNAi arrays of siRNAs or even viruses on growth supports (slides or plates) onto which cells are seeded, offering significantly higher throughputs than conventional methods for those assays that prove compatible. The range of compatibility of screening assays with the SPOT-RNAi approach will depend on the method's experimental robustness, particularly in accommodating different cell

types, as well as the number of cells required within each sample (i.e. over each siRNA spot) to yield statistically significant results from the desired assay. Other up-and-coming developments of note may include the emergence of new RNAi reagents, stemming from such recent developments as so-called endonuclease-derived siRNAs, or esiRNAs, by Yang *et al.* (2002). These pools of siRNA-like molecules are generated through the controlled digestion of long dsRNA molecules using either Dicer or bacterial RNase III, which shows superior activity *in vitro*. The resulting pool of heterogeneous sequences, all targeting the same transcript, is thought to increase the potency of the reagent, while 'diluting out' the off-target effects of each individual molecule. The detailed characterization of esiRNA's experimental performance, especially in screening applications, is keenly awaited, especially in view of the method's inherently low production costs, which promise cheap libraries.

In the meantime, it remains important to strive for maximal detection sensitivity and accuracy in these screens, making the right choices to minimize the risk of unnecessarily missing much of what is reasonably detectable. Indeed, in view of the relatively high costs of these studies, it is not likely that the same screen will be funded twice, thus actually increasing the chance that missed targets may remain undiscovered for years to come. There is huge discovery space to be explored by carrying out high-content RNAi screens in different cell types, in combination with different compounds and treated backgrounds, all of which will inevitably add much greater depth to the functional understanding of targets being considered for therapeutic applications in oncology and other fields.

#### Acknowledgements

We thank Andrea Krönke, Anne Grabner, Michael Hannus, Andrew Walsh, Liisa Koski, and Birte Sönnichsen for their scientific input, and Mary-Ann Grosse, Sindy Kluge, and Moddassar Khan for their skilled technical assistance. The work presented in this article has partially been supported by Grant no. 0314002 of the German Federal Ministry of Education and Science (BMBF).

#### References

- Arts GJ, Langemeijer E, Tissingh R, Ma L, Pavliska H, Dokic K, Dooijes R, Mesic E, Clasen R, Michiels F, van der Schueren J, Lambrecht M, Herman S, Brys R, Thys K, Hoffmann M, Tomme P and Van Es H. (2003). *Genome Res.*, **13**, 2325–2332.
- Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL and Bernards R.. (2004). *Nature*, **428**, 431–437.
- Carpenter AE and Sabatini DM. (2004). *Nat. Rev. Genet.*, **5**, 11–22.
- Clemens JC, Worby CA, Simonson Leff N, Muda M, Maehama T, Hemmings BA and Dixon JE. (2000). *Proc. Natl. Acad. Sci., USA*, **97**, 6499–6503.
- Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K and Tuschl T. (2001a). *Nature*, **411**, 494–498.
- Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W and Tuschl T. (2001b). *EMBO J.*, **20**, 6877–6888.
- English JJ, Mueller E and Baulcombe DC. (1996). *Plant Cell*, **8**, 179–188.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE and Mello CC. (1998). *Nature*, **391**, 806–811.

- Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M and Ahringer J. (2000). *Nature*, **408**, 325–330.
- Gönczy 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 and Hyman AA. (2000). *Nature*, **408**, 331–336.
- Hamilton AJ and Baulcombe DC. (1999). *Science*, **286**, 950–952.
- Holen T, Amarzguioui M, Wiiger MT, Babaie E and Prydz H. (2002). *Nucleic Acids Res.*, **30**, 1757–1766.
- Hopkins AL and Groom CR. (2002). *Nat. Rev. Drug Discov.*, **1**, 727–730.
- Hsieh AC, Bo R, Manola J, Vazquez F, Bare O, Khvorova A, Scaringe S and Sellers WR. (2004). *Nucleic Acids Res.*, **32**, 893–901.
- Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G and Linsley PS. (2003). *Nat. Biotechnol.*, **21**, 635–637.
- Kiger A, Baum B, Jones S, Jones M, Coulson A, Echeverri C and Perrimon NA. (2003). *J. Biol.*, **2**, 27.
- Krönke A, Grabner A, Hannus M, Sachse C, Dorris D and Echeverri C. (2004). *Drug Discov. World*, **5**, 53–62.
- Lum L, Yao S, Mozer B, Rovescalli A, Von Kessler D, Nirenberg M and Beachy PA. (2003). *Science*, **299**, 2039–2045.
- Mueller E, Gilbert JE, Davenport G, Brigneti G and Baulcombe DC. (1995). *Plant J.*, **7**, 1001–1013.
- Paddison PJ, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Baliya V, O’Shaughnessy A, Gnoj L, Scobie K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ and Hannon GJ. (2004). *Nature*, **428**, 427–431.
- Persengiev SP, Zhu X and Green MR. (2004). *RNA*, **10**, 12–18.
- Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS and Khvorova A. (2004). *Nat. Biotechnol.*, **22**, 326–330.
- Rubinson DA, Dillon CP, Kwiatkowski AV, Sievers C, Yang L, Kopinja J, Rooney DL, Ihrig MM, McManus MT, Gertler FB, Scott ML and Van Parijs L. (2003). *Nat. Genet.*, **33**, 401–406.
- Sachse C, Krausz E, Krönke A, Hannus M, Walsh A, Grabner A, Ovcharenko D, Dorris D, Trudel C, Sönnichsen B and Echeverri C. (2004). *Methods Enzymol.*, **392**, (in press).
- Scacheri PC, Rozenblatt-Rosen O, Caplen NJ, Wolfsberg TG, Umayam L, Lee JC, Hughes CM, Shanmugam KS, Bhattacharjee A, Meyerson M and Collins FS. (2004). *Proc. Natl. Acad. Sci. USA*, **101**, 1892–1897.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N and Zamore PD. (2003). *Cell*, **115**, 199–208.
- Sledz CA, Holko M, de Veer MJ, Silverman RH and Williams BR. (2003). *Nat. Cell. Biol.*, **5**, 834–839.
- Snoe Jr O and Holen T. (2004). *Biochem. Biophys. Res. Commun.*, **319**, 256–263.
- Sönnichsen B, Koski L, Walsh A, Marshall P, Neumann F, Brehm M, Alleaume A-M, Artelt J, Bettencourt P, Cassin E, Hewitson M, Holz C, Khan M, Lazik S, Martin C, Nitschke B, Ruer M, Stanford J, Winzi M, Heinkel R, Röder M, Finell J, Häntsch H, Jones S, Jones M, Coulson A, Oegema K, Gönczy P, Hyman AA and Echeverri CJ. Genome-wide screening by RNA interference identifies 663 genes required for the first cell division of *C elegans*. manuscript.
- Xin H, Bernal A, Amato FA, Pinhasov A, Kauffman J, Breneman DE, Derian CK, Andrade-Gordon P, Plata-Salaman CR and Ilyin SE. (2004). *J. Biomol. Screen.*, **9**, 286–293.
- Yang D, Buchholz F, Huang Z, Goga A, Chen CY, Brodsky FM and Bishop JM. (2002). *Proc. Natl. Acad. Sci. USA*, **99**, 9942–9947.
- Zhang JH, Chung TDY and Oldenburg KR. (1999). *J. Biomol. Screen.*, **4**, 67–73.