DRUG DISCOVERY TODAY TARGETS

High-throughput target discovery using cell-based genetics

P. David Jackson and John J. Harrington

High-throughput target discovery requires robust disease models and the ability to rapidly survey the genome for function. In the post-genomics era, there has been a strong emphasis placed upon 'gene-to-function' approaches that take advantage of the large amount of gene sequence information now available. Here, we advocate a return to 'function-to-gene' approaches as a first step in target discovery (and validation), followed by hypothesis-driven research to validate new targets identified by their activity in cell-based disease models.

In the past decades, a fundamental paradigm shift in drug development occurred in the pharmaceutical industry. Previously, drug discovery was carried out in cellular and animal models to identify molecules that affect human disease relevant phenotypes. Compound libraries were screened and active compounds were subsequently optimized, often with little or no knowledge of the target or mechanism of action. While this approach has yielded many drugs, it suffers from several limitations. Without knowledge of the target, it is difficult to develop structure–activity relationships (SARs) to guide lead optimization; it is not possible to counterscreen compounds against related targets to enhance selectivity, nor is it possible to carry out structure-guided drug design; and the design of clinical trials suffers from a poor understanding of drug action. Thus, there are significant advantages to identifying specific drug targets *a priori* and then screening for compounds that modulate their activity. Today, most new drug candidates are created using a target-centric approach to drug screening.

The need for improved target discovery is emphasized by the backlog of unmet medical needs in treating increasingly prevalent diseases such as obesity, diabetes, cardiovascular disease, immune disorders and cancer. A rational selection of drug targets benefits from a thorough understanding of disease processes, including the identification and characterization of genes playing causative roles in human disease. We advocate the use of high-throughput genetics to perturb cellular disease models as a starting point for target discovery. This approach allows the entire genome to be surveyed for biological function and results in the identification of a small number of genes whose altered function has an impact on a disease-relevant cellular phenotype. The relevance of these genes can be further investigated employing lower-throughput organismal disease models, such as transgenic and gene-knockout animals. Ultimately, the identification of rigorously validated, disease-relevant targets will increase the efficiency and therapeutic success of drug discovery. Here, we review emerging techniques for high-throughput target identification and advocate a return to the genetic discovery of disease determinants.

Current target discovery

Drug targets are currently discovered using combinations of three basic strategies (Figure 1). First, expression profiling and proteomic analyses have been used to identify genes and proteins whose expression or post-translational modification is altered

P. David Jackson
John J. Harrington*
Athersys Inc.
3201 Carnegie Ave,
Cleveland,
OH, 44115, USA
*email:
jharrington@athersys.com

1359-6446/04/$ – see front matter PII: S1359-6446(04)03303-3

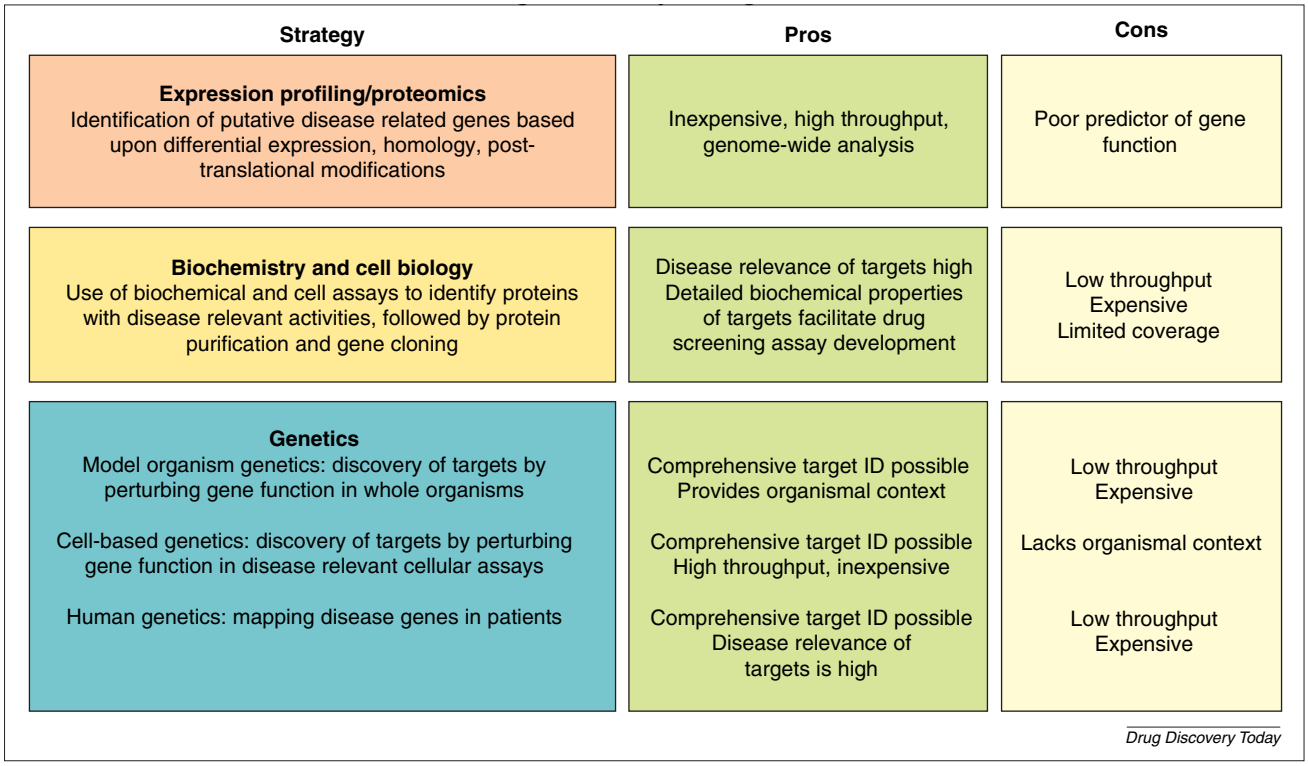

FIGURE 1

Three general strategies for target discovery.

in disease. Such studies permit cost-effective, genome-wide measurement of changes using microarrays, serial analysis of gene expression (SAGE), or proteomic measurement tools [1–4]. Unfortunately, transcriptional profiling does not always accurately predict protein levels, and, more importantly, altered gene expression by itself has been a poor predictor of gene function [5–8]. Thus, these correlation studies offer speed, cost-effectiveness and genome coverage but rarely establish causal relationships. Second, biochemical and cell-biological assays have been extensively utilized to identify genes and proteins directly participating in disease pathways. For example, biochemical characterization of the cholesterol biosynthetic pathway led to the identification of key enzymes, such as hydroxymethylglutaryl-coenzyme A (HMG-CoA), that served as targets for the development of cholesterol lowering drugs [9]. Similarly, cell differentiation assays in combination with protein fractionation methods have been used to identify and purify numerous therapeutically relevant growth factors, including erythropoietin, granulocyte colony stimulating factor and human growth hormone [10–12]. Such efforts are arduous, but they have also clearly been successful in identifying therapeutic targets, especially where the cause of the disease is understood, as was the case in the examples cited above. Thus, biochemical and cell biological approaches have the ability to establish disease relevance and identify drug targets, but are low-throughput, expensive and lack genome coverage. Finally, genetics allows one to identify relevant genes on the basis of the phenotype that results when their function is disturbed; this method has been used to identify genes and pathways regulating cell cycle, apoptosis and development [13–16]. Genetic screens have begun in mammals; however, achieving genome-wide coverage has been hindered by the time and expense required for comprehensive genetic screens in animals [17–19]. Thus, mammalian genetics can identify causal relationships and disease relevance but suffers from low throughput and high cost.

Each of the above target discovery approaches has utility and, indeed, plays an important role in target identification and validation. Not one of the approaches will provide all of the information required to move a target into drug screening with complete confidence. The key consideration is how best to use these approaches together to cost-effectively and efficiently identify and validate new targets. We argue that cell-based genetics offers the best balance between speed, cost, genome coverage and biological relevance, and is therefore the most effective method for initiating the target discovery process.

The case for cell-based genetics

Cell-based genetics refers to all molecular approaches for modifying the expression or activity of all or nearly all genes in a cell type of interest for the purpose of assessing the phenotypic effects upon the cell. By using highly scaleable cell culture assays, it is possible to cost-effectively analyze the entire genome for biological function and disease relevance. Mutagenesis can generate complete loss or gain of gene function and a continuum of gene activity between these extremes. This broad range of gene activities

encompasses the set of perturbations most likely to cause disease and lead to identification of lists of candidate drug targets. Once relevant genes are identified, lower-throughput approaches in animals, such as transgenics or gene knockouts, can be applied in combination with hypothesis-driven experimentation to further validate a target of interest. Validated targets can then be screened to identify compounds that modulate their activity, and serve as starting points for lead optimization and development of new drug candidates (Figure 2).

Requirements for cell-based target screens

In order for cell-based genetics to be useful for target discovery, cellular assays must meet several requirements. First and foremost, the assay must accurately model a critical aspect of the disease of interest. Faithful modeling of human disease requires the use of cell types crucial to disease pathology, an ability to modify gene function, and a method for detecting changes in a disease-relevant phenotype. Numerous cell-based assays have been developed to identify genes involved in cellular transformation/oncogenesis, differentiation, activation, migration and invasion, metabolism, viral infection, signaling, apoptosis, neo-vascularization and other biological processes relevant to human disease [20–24]. However, not all human diseases and biological processes, such as cognition and memory, are easily modeled in a cell system. The inability to model some biological and disease processes in cell culture is a result of several factors, including insufficient understanding of the role of specific cell types in the process and an inability to culture specialized cell lineages. As we address these issues, increasingly sophisticated cell-based models incorporating new cell lineages, multiple cell types, disease triggers and reporter systems may overcome this limitation to enable target discovery for most human diseases.

Second, the cellular assay must possess many of the attributes required for any high-throughput screen. Specifically, the assay must have a suitable signal-to-noise ratio permitting the identification of rare genetic events over a background of a large number of negatives. The assay format must be scaleable, permitting analysis of a large number of genes and alleles, often 100 000 to 1 000 000 cellular clones or more to achieve genome coverage. Depending on the phenotype of interest and the genetic methodology used for perturbation, the cells used in the assay must be able to proliferate in culture to allow phenotype detection.

Third, the screening assay must allow isolation of cells displaying the desired phenotype. Separating the cells of interest from the general population can be accomplished by a variety of approaches, including drug selection for cells activated for a transgenic selectable marker/reporter gene or physical separation of cells displaying a new surface antigen associated with the phenotype of interest using magnetic beads or fluorescence activated cell sorting (FACS).

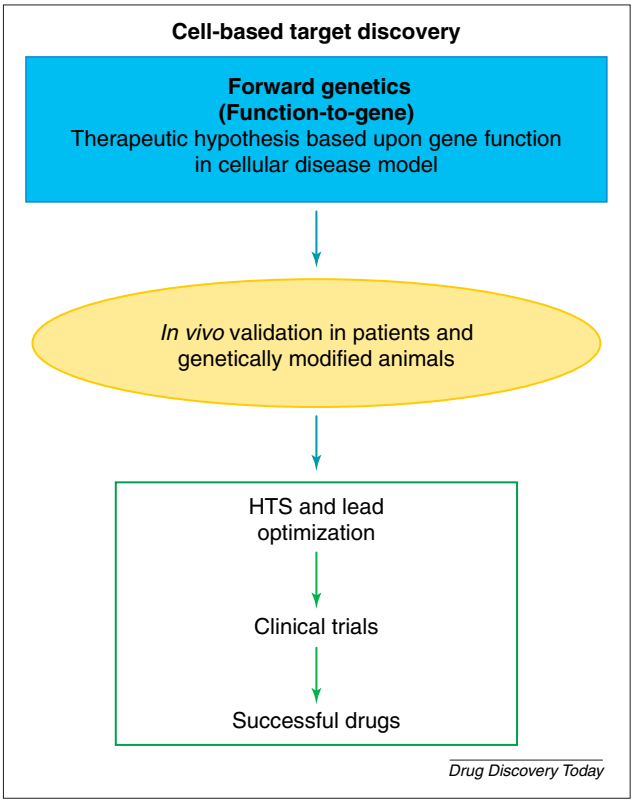

FIGURE 2

Cell-based target discovery and validation. Cell-based target discovery uses disease-relevant cellular assays to survey the genome for gene function and disease relevance. Use of cell-based assays makes genome-wide target discovery feasible as a consequence of the speed and cost-efficiency of cellular assays relative to *in vivo* discovery. Since genes are identified on the basis of function, there is a higher probability of success in the lower-throughput, more expensive *in vivo* validation phase of testing. As a result, fewer genes need to be tested *in vivo,* and the overall cost of discovering new targets is lower. The ability of genetics to survey the genome comprehensively also results in fewer missed genes. Following validation, targets may be taken into high-throughput screening (HTS) to identify compounds that modulate their activity.

Finally, the methodology used to perturb the cells must allow efficient identification of genes responsible for an observed phenotype. This is best achieved by making a distinguishing molecular link between the perturbing agent and the affected gene, since finding untagged genes in the expanse of a mammalian genome is otherwise a nearly impossible task. In practice, known nucleotide sequences present on the perturbation vector (e.g. cDNA expression vector or insertional mutagen) provide the means for identifying genes present either on the vector (in the case of expression vectors) or adjacent to the integrated vector (in the case of insertional mutagens).

Emerging technologies for high-throughput target screens in mammalian cells

Based upon the requirements cited above, two complementary strategies for manipulating gene expression/activity in cells have been developed. Loss-of-function (LOF) strategies focus on decreasing or eliminating gene

expression/activity, while gain-of-function (GOF) strategies focus on increasing gene expression/activity in the cell. From a target discovery perspective, these two approaches mimic drug action in different ways. For example, decreasing the activity of a target gene can be used to mimic the effect of an antagonist or enzyme inhibitor, whereas increasing the activity of a target gene can be used to mimic the effect of an agonist. In both strategies, targets are identified when gene activity changes detectably alter a disease-relevant cellular phenotype. It is important to recognize, however, that up-regulation of some genes may have no effect on cellular phenotype, whereas down-regulation of these same genes may have a profound effect on the cell. Conversely, up-regulation of other genes may have a profound effect on phenotype, whereas down-regulation may have no effect. Thus, the most comprehensive yield of drug targets results from applying both LOF and GOF strategies, as these complementary approaches often identify different target sets.

Loss-of-function genetics

Much ingenuity has gone into developing mutagenesis techniques that disrupt gene activity and permit the rapid identification of biologically active genes. The emerging LOF techniques fall into three major classes on the basis of the mechanism of gene inactivation at the level of, first, the gene, second, the transcript or, third, the protein (Figure 3).

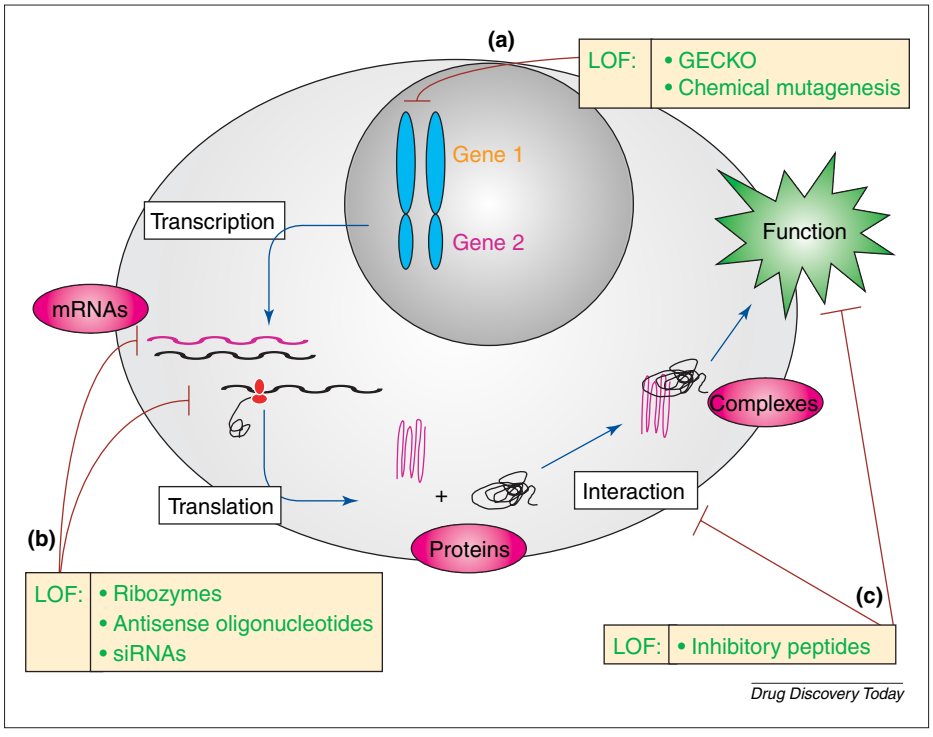

FIGURE 3

Molecular targets of cell-based loss-of-function (LOF) technologies. Loss of gene function can be achieved in the cell by **(a)** mutating genes to change either their protein coding potential or level of transcription, **(b)** targeting transcripts to decrease their stability or translatability, or **(c)** providing interfering peptides to disrupt protein structure or interactions necessary for function.

LOF by gene mutation

LOF at the gene level has the broadest potential for target discovery as it is possible to create a large range of mutations at every gene to collectively provide the best chance of perturbing gene function to produce a disease-relevant phenotype. Two LOF methods are aimed at the gene level. The first approach utilizes extensive chemical mutagenesis to mutate a large number of genes in each cell. By saturating the genome with mutations within a population of cells, a library is produced in which different clones contain different combinations of mutations. In libraries of sufficient size, every gene will be mutated at both alleles in at least one clone to produce a complete loss of gene function. Following screening, mutated cells displaying the phenotype of interest can then be isolated and the gene responsible for the phenotype identified by cDNA rescue, an approach that relies on genetic complementation of the mutant phenotype with cDNA expression libraries [25–27]. To identify the mutated genes using cDNA rescue, cells displaying the phenotype of interest are transfected with a cDNA expression library produced from the parental non-mutagenized cells. Transfected clones displaying a corrected phenotype are isolated and the cDNA responsible for the complementation is identified. While this approach has been used effectively to create mutations, identification of the gene of interest using cDNA rescue is laborious and prone to failure because of the general limitations of cDNA expression cloning discussed below. In one recent screen, only mutations in known regulators of Fas-induced apoptosis were found, emphasizing the difficulty of target discovery using this approach [28]. The lack of novel genes identified in this screen likely reflects a combination of incomplete mutagenesis and the biased and incomplete cDNA expression libraries used to identify mutated genes.

A more recent gene-level LOF technique is GECKO (Genome-wide Cell-based Knock-Out) [29]. As described in Figure 4, GECKO produces genome-wide LOF mutations by combining chemical mutagenesis with insertional mutagenesis. The combination of these approaches produces the high mutation efficiency associated with chemical mutagenesis and the gene identification capability associated with insertional mutagenesis. Thus, GECKO solves the major deficiencies of chemical mutagenesis (i.e. difficulty of identification of mutated genes) and insertional mutagenesis (i.e. inefficiency of creating homozygous mutations). Accordingly, GECKO produces a library containing a small number of clones in which each gene has been knocked out

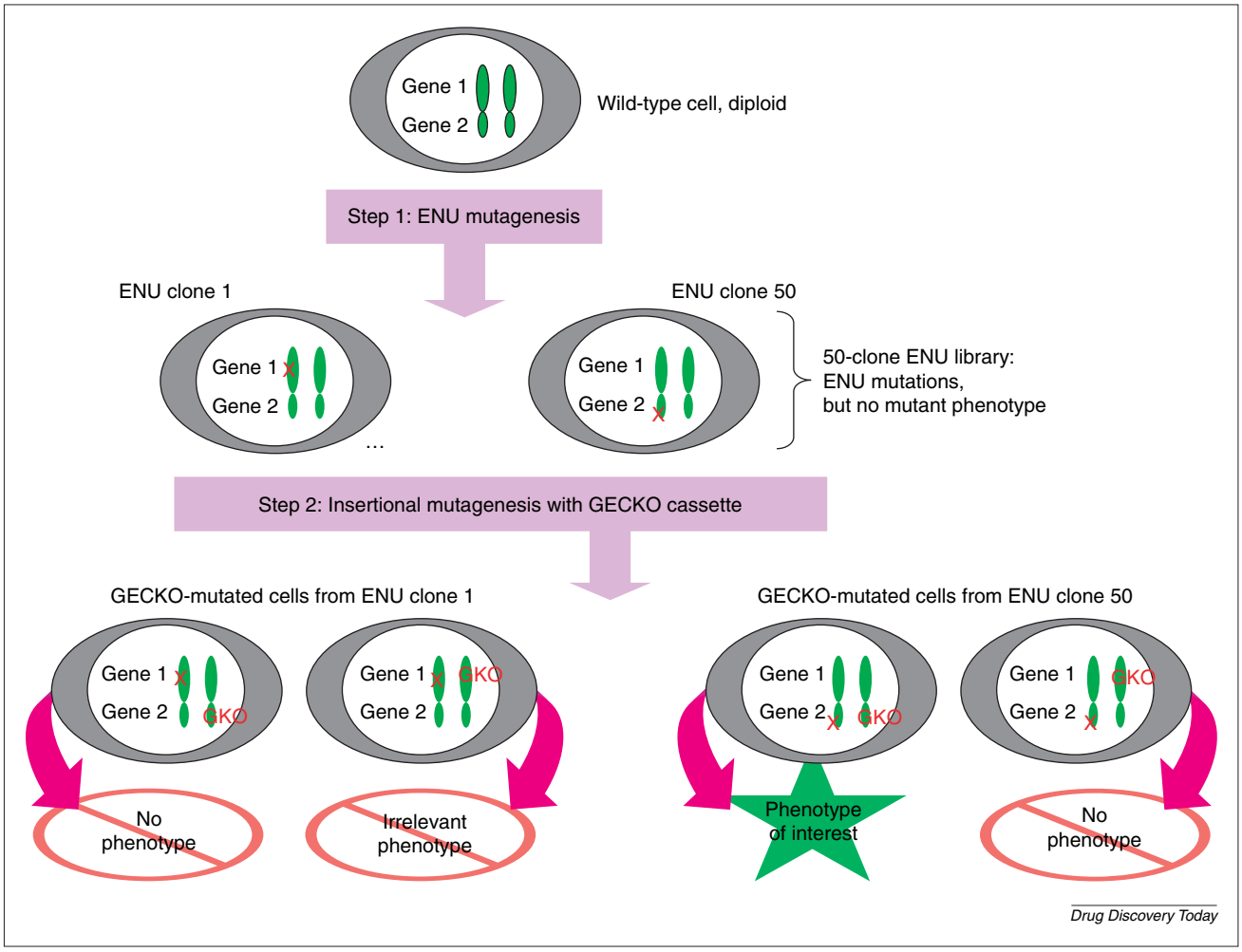

FIGURE 4

Use of GECKO to create and screen genome-wide LOF libraries. GECKO mutagenesis is accomplished in two stages. In step 1, wild type cells are mutagenized with a chemical mutagen, such as N-ethyl N-nitrosourea (ENU), to create mutations in a large number of alleles in each cell. In this example, 100 LOF mutations in each of 50 ENU clones is sufficient to inactivate one allele of each of the ~5000 expressed genes once, on average. Under these conditions, phenotypes are rarely created in diploid cells at this stage, but each mutated clone is examined to ensure that none have yet acquired the phenotype of interest. In step 2, chemically mutated cells are expanded and treated with a gene trap vector. Each cell will contain a unique combination of mutations. In some cells, one allele of a given gene is mutated by the chemical mutagen and the other allele mutated by the gene trap, resulting in loss of function at both alleles. Following construction, the library is screened for a disease relevant phenotype and clones are isolated. Since the gene trap is inserted at the locus of the gene that it inactivates, 5′ RACE using gene tap sequences can be used to rapidly and unambiguously identify the gene responsible for the observed phenotype.

in at least one clone within the library, and at least one allele of the knocked-out gene is tagged by an insertional mutagen to permit its rapid and unambiguous identification. Importantly, GECKO does not require foreknowledge of gene sequence to create genome-wide LOF libraries, an advantage shared with genetic screens using a chemical or insertional mutagen in model organisms. In a recent screen for activation-defective T cells, GECKO successfully identified many of the well-established determinants of lymphocyte activation, as well as several novel determinants (manuscript in preparation). The independent confirmation of the novel genes using siRNA not only verified their role in T cell activation, but, perhaps more importantly, illustrates how multiple LOF techniques can be used together to increase the level of validation around newly discovered targets.

LOF by RNA modulation

LOF, by decreasing RNA transcript level, provides an alternative approach to reducing gene expression in the cell. Three distinct RNA-targeting approaches – ribozymes, antisense (AS) oligonucleotides, and small interfering RNA (siRNA) – have been developed. All of these RNA-targeting techniques use base-pairing between the targeting agent and the target mRNA transcript to confer specificity and all require knowledge of the gene sequence. The oldest mRNA-level LOF approach, AS oligonucleotide inhibition, decreases gene expression by introducing AS oligonucleotides into the cell, which then anneal to the target mRNA resulting in Rnase-H-dependent degradation of the mRNA transcript and/or in the physical blockade of protein translation. While this approach has been used effectively to decrease expression of individual and small groups of

genes, screening with genome-wide AS libraries has not yet been practical because of difficulties constructing comprehensive antisense oligonucleotide libraries corresponding to all genes and because of the problem of off-target effects [30–33]. In addition, the apparent need for relatively high levels of AS per cell has hindered quantitative gene knock-down, particularly of abundant transcripts. As an alternative approach to antisense mediated LOF, engineered ribozymes have been developed to target and catalyze the hydrolysis of specific mRNA transcripts. While ribozymes have been shown to reduce expression of specific genes, experience has shown that extensive optimization of the ribozyme is required to reliably achieve RNA knock-down. This optimization generally focuses on identification of accessible regions of the target mRNA, improving the specificity of target recognition, increasing catalytic turnover by modifying sequences throughout the ribozyme molecule, and, in the case of synthetic ribozymes, increasing half-life by incorporating DNA or non-natural nucleotide analogs into the ribozyme molecule [34–36]. Because of the empirical nature of ribozyme optimization, genome-wide screens using gene sequence databases to design libraries of ribozymes have not been possible. Random sequence ribozyme libraries, however, have been used to identify both known and novel inhibitors of TNF-α and FasL-induced apoptosis [37–39]. Thus, random ribozyme approaches appear to have promise in cell-based assays. The third and perhaps most promising mRNA-level LOF approach, siRNA, decreases gene expression by the introduction of short, sequence-specific double stranded RNA that guides the endogenous RNA-induced silencing complex (RISC) machinery to degrade specific mRNA transcripts. SiRNA-mediated transcript degradation is catalytic, thereby reducing the amount of siRNA that must be delivered to the cell to effectively decrease gene expression. Unlike AS oligonucleotides and ribozymes, design algorithms for siRNA more reliably achieve at least 50–80% knockdown of gene activity, a reduction sufficient to see phenotypes associated with some but not all genes. siRNA libraries targeting about one-third of the known human genes have been constructed and used to screen for genes involved in P53, NFkB and proteosome pathways. The identification of several known and novel genes validates this approach [40–43]. The authors noted, however, that many known pathway components were missed despite representation in the siRNA library, suggesting that more extensive siRNA redundancy (i.e. larger libraries) might be necessary to identify some genes. Furthermore, in other studies, off-target effects have been noted with some siRNAs, and this has the potential of producing 'false positives' in cell-based screens [44,45]. As new siRNA mechanistic studies are carried out and the method refined, it is likely that knock-down efficiency and target specificity will be further improved, enhancing the overall effectiveness of this approach in target discovery and validation.

LOF by protein inhibition

The third approach to decreasing target activity produces LOF at the protein level. In this approach, libraries of overexpressed random peptides are used to interfere with protein function and create cellular phenotypes. Loss-of-function is typically achieved in this approach when a peptide binds to a target protein and inhibits its enzymatic activity or its ability to interact with other proteins necessary for its function. Following perturbation, targets of bioactive peptides are then identified by cDNA rescue (as described above for chemical mutagenesis), affinity target purification of proteins that interact with the peptide perturbagen, or two-hybrid protein interaction screens between a cDNA library and the peptide perturbagen. Random peptide libraries have been screened to identify genes involved in T cell activation and resistance to chemotherapeutic agents. In these screens, multiple LOF peptides were identified, although endogenous targets were rarely identified, suggesting that this approach may be more useful for producing cellular phenotypes as a tool to study a biological process rather than as an approach to identifying new target genes [46–49]. Nonetheless, the random peptide library approach provides an important means of influencing cell phenotypes by interacting with pathways at the protein level and offers the theoretical possibility of using the inhibitory peptide sequences and/or structures identified in the screen to guide small molecule SAR for the creation of lead compounds.

Overall, LOF strategies have greatly improved in recent years, leading to the identification of many regulators of disease processes such as cancer and immune disorders [50–51]. Furthermore, it is now possible to routinely, rapidly and cost-effectively screen the entire genome for function in cellular models of human disease. As these approaches are refined and new approaches developed, LOF screens will increasingly contribute to drug target discovery in the coming years.

Gain-of-function genetics

The ability to perturb cells by gene overexpression provides an important complement to the LOF approaches described above. Since many diseases, such as cancer and inflammation, can be caused by overexpression of specific genes, GOF approaches can be used to identify genes capable of conferring a disease phenotype on a cell type of interest. Furthermore, other diseases and disorders result from loss of gene function and can be corrected by overexpressing a specific protein. Thus, GOF approaches can be used to identify genes that cause, as well as genes that correct, human disease.

There are two primary strategies for creating GOF libraries in cells. The first approach uses cDNA expression libraries to overexpress a large number of genes in a cell type of interest. Cells displaying a phenotype of interest are isolated and the cDNA cloned, typically by recovering the expression vector and gene of interest from the cell or

by amplifying the gene of interest using polymerase chain reaction. This approach has been used successfully to identify novel regulators of processes relevant to disease, such as immune cell function, osteogenesis, metastasis and angiogenesis [52–55]. While cDNA expression approaches can be used to identify new disease genes, DNA libraries are difficult to normalize, often biased toward smaller, abundantly expressed genes as the result of large variations in gene expression in the cells used to produce the library and difficulties in cloning long transcripts. As a consequence, rare transcripts and splice variants capable of producing a cellular phenotype may be missed as a result of their poor representation in existing cDNA libraries [56–57]. These issues are being addressed by systematically cloning full-length cDNAs corresponding to known and predicted genes. This emerging collection of full-length expression clones has allowed the testing of large numbers of genes of unknown function in cell-based assays [58]. Unfortunately, at the present time, only a minority of genes and splice variants are available for testing.

A second GOF approach, RAGE (random activation of gene expression), addresses many of the deficiencies of cDNA-based methodologies. RAGE uses insertional mutagenesis to overexpress endogenous genes in cells *in vitro* [59]. Briefly, RAGE utilizes specialized plasmid or retroviral vectors containing a promoter linked to a generic exon and splice donor site. Integration of the vector into the genome of a cell (by transfection or viral infection) results in the transcriptional activation of the first endogenous gene downstream of the insertion site. Following transcription, splicing occurs between the vector-encoded splice donor site and the first downstream splice acceptor site within an endogenous gene. The resulting chimeric transcript is translated into large amounts of the protein of interest. Since RAGE libraries activate genes directly in the genome in a relatively unbiased fashion, these libraries are inherently normalized and can be expanded to overexpress nearly all possible gene products. In a recent screen, RAGE successfully identified both known and novel genes in the NFkB signaling pathway, demonstrating the validity of this approach [60]. Importantly, RAGE can be applied to cells derived from any tissue and any species, and therefore can be used in a large range of disease-relevant cellular models.

Like LOF approaches, GOF methodologies play an important role in target discovery and validation. GOF approaches have proven to be highly scalable, efficient and cost-effective. With a demonstrated ability to identify therapeutically important genes, GOF screens will continue to discover new targets missed by other approaches.

Conclusions and future directions

The human genome project has led to the identification of over 30 000 human genes; however, the majority of these genes currently have no known biological function. Bioinformatic analysis, expression profiling and other reverse-genetics approaches are now being used in an effort to link these genes to biological pathways and disease. Unfortunately, these 'guilt-by-association' approaches are poor predictors of gene function. As a result, a large amount of time and resources are currently spent surveying the genome for associations and developing hypotheses that ultimately are not correct. Recent advances in cell-based genetics now enable a more efficient approach to establishing gene function and disease relevance. By beginning with demonstrated function, established in cell-based assays, a logical and high-probability set of experiments can be undertaken to further validate newly discovered genes.

In the future, we believe there will be a convergence between correlation approaches and cellular/organismal genetic approaches to produce a fully capable systems biology approach to target discovery and pathway characterization [61]. As the field moves in this direction, genome-wide cell-based genetic screens will play an important role in the development of *in silico* prediction models. Ultimately, target discovery and validation will likely occur simultaneously, with multiple experimental approaches being used in parallel. Until then, cell-based genetic approaches to target discovery offer the best balance between speed, cost-effectiveness, genome coverage and biological relevance.

References

1 Schena, M. *et al.* (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* 16, 301–306
2 Zhang, L. *et al.* (1997) Gene expression profiles in normal and cancer cells. *Science* 276, 1268–1272
3 Velculescu, V.E. *et al.* (1995) Serial analysis of gene expression. *Science* 270, 484–487
4 Gromov, P. *et al.* (2002) Human proteomic databases: a powerful resource for functional genomics in health and disease. *Prog. Biophys. Mol. Biol.* 80, 3–22
5 Birrell, G.W. *et al.* (2002) Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8778–8783
6 Gygi, S.P. *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730
7 Giaever, G. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391
8 Chen, G. *et al.* (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteomics* 1, 304–313
9 Tobert, J.A. (2003) Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat. Rev. Drug Discov.* 2, 517–526
10 Lee-Huang, S. (1984) Cloning and expression of human erythropoietin cDNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 81, 2708–2712
11 Souza, L.M. *et al.* (1986) Recombinant human granulocyte colony-stimulating factor: effects on normal and leukemic myeloid cells. *Science* 232, 61–65
12 Martial, J.A. *et al.* (1979) Human growth hormone: complementary DNA cloning and expression in bacteria. *Science* 205, 602–607
13 Horvitz, H.R. *et al.* (1994) The genetics of programmed cell death in the nematode *Caenorhabditis elegans*. *Cold Spring Harb. Symp. Quant. Biol.* 59, 377–385
14 Hartwell, L.H. (1991) Twenty-five years of cell cycle genetics. *Genetics* 129, 975–980

15 Murakami, H. and Nurse, P. (2000) DNA replication and damage checkpoints and meiotic cell cycle controls in the fission and budding yeasts. *Biochem. J.* 349, 1–12

16 St. Johnston, D. and Nusslein-Volhard, C. (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68, 201–219

17 Rathkolb, B. *et al.* (2000) Large-scale N-ethyl-N-nitrosourea mutagenesis of mice–from phenotypes to genes. *Exp. Physiol.* 85, 635–644

18 Justice, M.J. *et al.* (1997) Using targeted large deletions and high-efficiency N-ethyl-N-nitrosourea mutagenesis for functional analyses of the mammalian genome. *Methods* 13, 423–436

19 Zambrowicz, B.P. *et al.* (2003) Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14109–14114

20 Randow, F. and Seed, B. (2001) Endoplasmic reticulum chaperone gp96 is required for innate immunity but not cell viability. *Nat. Cell Biol.* 3, 891–896

21 He, Y.W. and Bevan, M.J. (1999) High level expression of CD43 inhibits T cell receptor/CD3-mediated apoptosis. *J. Exp. Med.* 190, 1903–1908

22 Lorincz, M.T. *et al.* (2004) Embryonic stem cells expressing expanded CAG repeats undergo aberrant neuronal differentiation and have persistent Oct-4 and REST/NRSF expression. *Mol. Cell. Neurosci.* 26, 135–143

23 Aiken, C.T. *et al.* (2004) A cell-based screen for drugs to treat Huntington's disease. *Neurobiol. Dis.* 16, 546–555

24 Peeper, D.S. *et al.* (2002) A functional screen identifies hDRIL1 as an oncogene that rescues RAS-induced senescence. *Nat. Cell Biol.* 4, 148–153

25 Thompson, L.H. *et al.* (1970) Isolation of temperature-sensitive mutants of L-cells. *Proc. Natl. Acad. Sci. U. S. A.* 66, 377–384

26 Wang, R.J. (1974) Temperature-sensitive mammalian cell line blocked in mitosis. *Nature* 248, 76–78

27 Goldsmith, M.A. and Weiss, A. (1987) Isolation and characterization of a T-lymphocyte somatic mutant with altered signal transduction by the antigen receptor. *Proc. Natl. Acad. Sci. U. S. A.* 84, 6879–6883

28 Juo, P. *et al.* (1998) Essential requirement for caspase-8/FLICE in the initiation of the Fas-induced apoptotic cascade. *Curr. Biol.* 8, 1001–1008

29 Harrington, J.J. *et al.* (2004) Compositions and methods for making mutations in cell lines and animals. United States Patent Application 20040018624.

30 Marcusson, E.G. *et al.* (2004) Case study: use of a library of antisense inhibitors for gene functionalization and drug target validation. *Drug Discov. Today* 3, 117–122

31 Nasevicius, A. and Ekker, S.C. (2000) Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* 26, 216–220

32 Heasman, J. (2002) Morpholino oligos: making sense of antisense? *Dev. Biol.* 243, 209–214

33 Lebedeva, I. and Stein, C.A. (2001) Antisense oligonucleotides: promise and reality. *Annu. Rev. Pharmacol. Toxicol.* 41, 403–419

34 Romero-Lopez, C. (2004) Design and optimization of sequnce-specific hairpin ribozymes. *Methods Mol. Biol.* 252, 327–338

35 Joseph, S. and Burke, J.M. (1993) Optimization of an anti-HIV hairpin ribozyme by *in vitro* selection. *J. Biol. Chem.* 268, 24515–24518

36 Hovig, E. *et al.* (2001) Optimization of hammerhead ribozymes for the cleavage of S100A4 (CAPL) mRNA. *Antisense Nucleic Acid Drug Dev.* 11, 67–75

37 Welch, P.J. *et al.* (2000) Identification and validation of a gene involved in anchorage independent cell growth control using a library of randomized hairpin ribozymes. *Genomics* 66, 274–283

38 Kawasaki, H. and Taira, K. (2002) A functional gene discovery in the Fas-mediated pathway to apoptosis by analysis of transiently expressed randomized hybrid-ribozyme libraries. *Nucleic Acids Res.* 30, 3609–3614

39 Kawasaki, H. *et al.* (2002) Identification of genes that function in the TNF-α-mediated apoptotic pathway using randomized hybrid ribozyme libraries. *Nat. Biotechnol.* 20, 376–380

40 Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498

41 Zheng, L. *et al.* (2004) An approach to genome-wide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 101, 135–140

42 Paddison, P.J. *et al.* (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature* 428, 427–431

43 Berns, K. *et al.* (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 428, 431–437

44 Jackson, A.L. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* 21, 635–637

45 Jackson, A.L. and Linsley, P.S. (2004) Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* 2, 521-524.

46 Xu, X. *et al.* (2001) Dominant effector genetics in mammalian cells. *Nat. Genet.* 27, 23–29

47 Chu, P. *et al.* (2003) Systematic identification of regulatory proteins critical for T-cell activation. *J. Biol.* 2, 21

48 Kinsella, T.M. *et al.* (2002) Retrovirally delivered random cyclic peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J. Biol. Chem.* 277, 37512–37518

49 Geyer, C.R. *et al.* (1999) Mutagenesis by peptide aptamers identifies genetic network members and pathway connections. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8567–8572

50 Chu, P.C. *et al.* (2004) A novel role for p21-activated pro tein kinase 2 in T cell activation. *J. Immunol.* 172, 7324–7334

51 Berns, K. *et al.* (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 428, 431–437

52 Michiels, F. *et al.* (2002) Arrayed adenoviral expression libraries for functional screening. *Nat. Biotechnol.* 20, 1154–1157

53 Felts, K.A. *et al.* (2002) Functional cloning using pFB retroviral cDNA expression libraries. *Mol. Biotechnol.* 22, 25–32

54 Hitoshi, Y. *et al.* (1998) Toso, a cell surface specific regulator of Fas-induced apoptosis in T cells. *Immunity* 8, 461–471

55 Perez, O.D. *et al.* (2002) Activation of the PKB/AKT pathway by ICAM-2. *Immunity* 16, 51–65

56 Sun, M. *et al.* (2004) SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BMC Genomics* 5, 1–4

57 Chen, J. *et al.* (2002) Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12257–12262

58 Chanda, S.K. (2003) Genome-scale functional profiling of the mammalian AP-1 signal pathway. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12153–12158

59 Harrington, J.J. *et al.* (2001) Creation of genome-wide protein expression libraries using random activation of gene expression. *Nat. Biotechnol.* 19, 440–445

60 Jin, G.A. *et al.* (2004) Identification of a human NF-κB activating protein, TAB3. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2028–2033

61 Ideker, T. *et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929–934