DDT Vol. 6, No. 12 June 2001

Transgenic knockouts as part of high-throughput, evidence-based target selection and validation strategies

Stephen Harris

The worldwide genome sequencing projects are helping to define the size and complexity of the expressed genome and are thereby identifying an unprecedented number of genes of uncertain disease alignment and unknown function. It is widely recognized that, within the pharmaceutical industry, a significant commercial advantage will accrue to those companies that most effectively gather and integrate additional biological information into their therapeutic target selection and drug progression strategies. This article presents the rationale for including comparative phenotypic information obtained from transgenic gene knockouts as an integral part of any future therapeutic target selection strategy.

Stephen Harris
Technical Evaluation (Europe)
GlaxoSmithKline Research & Development
Medicines Research Centre
Gunnels Wood Road
Stevenage
UK SG1 2NY
tel: +44 1438 764964
fax: +44 1438 764810

▼ The completion of the first draft of the human genome sequence is, with the other ongoing genome sequencing projects, helping to define the size and complexity of the expressed mammalian genome, including an unprecedented number of genes of uncertain disease alignment and unknown function. As all gene sequences represent potential therapeutic targets, strategies must be developed to gain the additional biological insight that allows rate-limited R&D resources to be focused on those genes with the greatest therapeutic, and thereby commercial, potential. The emphasis within exploratory research is therefore shifting towards the evaluation and adoption of high-throughput technology platforms that can add additional value to the gene selection process, through functional studies or other measures of disease alignment, including genetics, differential gene expression, proteomics, tissue distribution, comparative species data and so on. The competition to achieve this will be particularly intense for those additional candidate gene family members that currently represent the chemically tractable or 'drugable' gene targets[1–4].

The potential of phenotypic information derived from gene knockouts to contribute to a high-throughput target selection/validation strategy has hitherto been limited by the resources required to generate and characterize a large number of gene knockout transgenics. Recently, several biotechnology companies have set out to address these issues and thereby to create an opportunity for the pharmaceutical industry to access this information in a timely fashion and on a previously unprecedented scale[5]. In this article, these opportunities are assessed with respect to the strategic business needs and changing organizational models being adopted within the pharmaceutical industry.

Therapeutic target selection: key considerations

The pharmaceutical industry is primarily focused on those key diseases for which there is a significant unmet clinical need and thereby future commercial value. To achieve this, the overall drug discovery process can be viewed as a series of key milestones or checkpoints at which go/no-go decisions are made about proceeding with a portfolio of projects. These decisions are based on the available evidence that individual projects have attained a minimum set of progression criteria. Until recently, maximizing diversity within HTS and the properties of compounds in preclinical and clinical settings has been a primary focus within the industry. This trend has been a result of, in part at least, the fact that the steady (rather than spectacular) rate of biological target identification and validation has meant that the portfolio of therapeutic targets has been relatively small. As such, the

1359-6446/01/$ – see front matter PII: S1359-6446(01)01794-9

physiochemical properties of a therapeutic product are particularly important to market success[6].

With the emergence of the human genome sequence, companies are increasingly beginning to focus on developing target selection strategies that will aggressively collect and integrate biological data into therapeutic target selection decisions, especially for those candidate genes considered to be most likely to deliver value to the business. To prosecute such a strategy, and thereby to minimize the overall R&D risk inherent in the initial choice of a therapeutic target from within a candidate gene pool, a pragmatic set of evidence-based selection criteria are needed upon which to base go/no-go decisions. Ideally, the outcome of such a shift in strategic emphasis should be an exploratory discovery organization capable of delivering a sustainable output of highly validated molecular targets into drug discovery[1].

In the short-to-medium term, therapeutic target selection decisions within exploratory research organizations are likely to become focused on two key properties of a candidate gene (Fig. 1). In the simplest terms, these can be portrayed in two dimensions as the relative positions of a portfolio of individual candidate genes along 'therapeutic' and 'chemical space' axes. In this model, the 'chemical space' axis represents the relative probability of obtaining viable chemical entities for progression within the drug discovery process after screening against the candidate gene, an estimate based on the historical precedent for the target class in question. The therapeutic axis represents the relative strength of the evidence for a disease association (or scientific rationale) between a given candidate gene and the desired therapeutic and/or mechanistic profiles of the disease of interest[6].

At a given time, there are likely to be exploratory projects, including multiple candidate genes, occupying points throughout this two-dimensional matrix, each seeking to gather evidence that results in a candidate gene achieving the relevant target selection threshold and thereby progressing into the drug discovery pipeline. As a natural consequence of using these criteria, most projects progressing from exploratory discovery into the drug discovery pipeline are likely to become focused on genes from within the chemically tractable families for which there is greatest evidence of a disease association (Zone A, Fig. 1). In addition, a limited number of pathway expansion projects focused on biological mechanisms with a strong scientific rationale (e.g. genetic, clinical or functional correlation with a disease phenotype) are likely to be pursued to identify a chemically tractable gene as a point of therapeutic intervention (Zone B, Fig. 1). These two axes are considered in more detail below with particular reference to the potential role of gene knockout phenotypes in helping to define the therapeutic alignment of individual members of the chemically tractable classes of candidate gene.

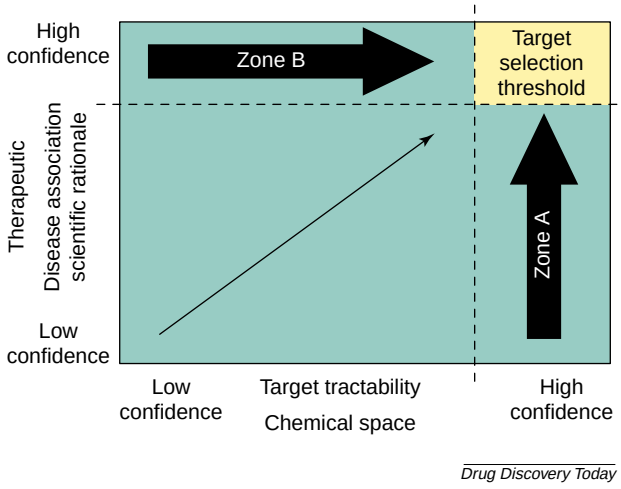

Figure 1. Key components of therapeutic target selection. In order to minimize the overall R&D risk inherent in the initial choice of a therapeutic target, such decisions often need to focus on two key properties of individual members of any pool of candidate genes. In the simplest terms, these can be described as (1) the 'therapeutic' axis, or the strength of the disease association (or scientific rationale) between a candidate gene and the disease of interest, and (2) the 'chemical space' axis, or the gene product's innate chemical tractability or 'drugability', an estimate based on historical precedent. Ideally, at any one point in time, the portfolio of therapeutic candidate genes moving from exploratory research into the drug discovery pipeline should be the subset of any pool of candidate genes attaining at least the minimum target selection thresholds established for each of these axes.

Bioinformatics: defining candidate gene targets

The draft human genomic sequence[7] is being intensively analyzed by bioinformaticians throughout the world to identify and catalogue all the genes within the expressed human genome. The uncertainty inherent in this task is illustrated by the broad range of values that had previously been estimated for the overall size of the expressed human genome, ranging from 28,000 up to >120,000 genes[8,9]. Although the final number is more likely to be in the range 30,000–40,000, this will no doubt continue to engender informed debate and good-humoured speculation (http://www.ensembl.org/) up until, and even beyond, the completion of the final draft of the human genome sequence, scheduled for 2003.

For the pharmaceutical industry, this represents a once-in-a-lifetime opportunity to participate in determining the size and complexity of the human candidate gene pool that

will contain all future therapeutic targets. More importantly, in the short term, this exercise will help define all the novel members of the gene families that are currently known to be chemically tractable or drugable, based on historical precedent (i.e. they are within Zone A, Fig. 1). Based on an extrapolation of the current therapeutic target classes, this subset of the genome is estimated at some 5000–10,000 genes, representing the G-protein-coupled receptors (GPCRs), ion channels, proteases, kinases and so on[1,6].

The potential value attributed to this DNA sequencing and data-mining activity is illustrated by the significant effort that has been expended in seeking a commercial advantage through development of a portfolio of gene-based intellectual property (IP)[10]. Although the future commercial value of the first phase of DNA-sequence-based IP remains uncertain, further biological insight around disease alignment and/or function might result in 'reach-through' claims that undermine the traditional commercial value of third-party chemical IP established around the biological target. In the short-to-medium term, therefore, companies that fail to establish biological IP (alone or in partnership) within the chemically tractable therapeutic gene-families might either be blocked from working on key molecular targets aligned to disease or obliged to pay significant royalties and/or milestone fees to third parties.

Despite significant efforts directed at understanding the biological significance of the expressed genome throughout the academic and commercial worlds, there is still an enormous number of genes of uncertain disease relevance and unknown function. Bioinformaticians can only infer biological function for the vast majority of genes, at least until additional biological annotation is deposited within public and proprietary databases[9]. The biological data that contribute to an increased level of confidence (validation) of any candidate gene on the therapeutic axis (Fig. 1) will come from a diverse set of activities, not all of which will be applicable to every disease state.

In the future, bioinformatics will, therefore, play an increasingly crucial role not only in continuing to identify and catalogue the chemically tractable genes within the genome but also in supporting the capture, integration and mining of gene-based biological annotation from a diverse set of experimental paradigms. Organizations that develop efficient knowledge management capabilities, as an integral part of any high-throughput biology effort, will benefit most from any shift in emphasis to a genomics-based target selection strategy.

Disease-based and gene-based approaches

In response to the opportunity and challenge represented by the scale of the expressed genome, the emphasis is beginning to shift from predominantly disease-to-target to gene-to-target strategies (Fig. 2). At this point, it is essential not to lose sight of the fact that both strategies are crucially dependent upon, and sensitive to, continued improvements in our background understanding of disease phenotypes and associated mechanisms and, therefore, need to be flexible enough to adapt to new knowledge. They are also not mutually exclusive strategies; rather they represent two extremes, both of which will already exist in a state of dynamic equilibrium within many pharmaceutical companies. There is, however, a widespread recognition of the need to adopt organizational models and processes aimed at streamlining the larger-scale collection and evaluation of biological information, including significant outsourcing options[2–4].

The disease-orientated strategy has historically favoured an exploratory research organization built around disease-focused multidisciplinary teams seeking to identify and validate the role of 'novel' candidate genes within the current understanding of the disease, and in clinical practice, cellular systems and comparative species (Fig. 2a). In this model, disease knowledge shapes the source of candidate genes and the crucial path of a portfolio of exploratory research projects moving outward from the most validated point(s) in biological space. This strategy classically involves repeated cycles of gene identification and functional validation (often partial), followed by HTS and drug development, with each cycle tailored to the hypothesized properties of the most current 'nearest-neighbour' set of candidate genes. This approach explains, at least in part, the consolidation of pharmaceutical interest around a similar set of clinically precedented or biologically well-validated therapeutic targets and/or pathways. As part of such a disease-orientated model, scarce resources to create and characterize gene knockouts are generally, but not exclusively, focused on supporting target validation of a prioritized shortlist of disease-aligned candidate genes, rather than driving target selection decisions *per se*.

By contrast, the gene-orientated strategy assumes the same background disease knowledge as above but favours an exploratory research organization built around a suite of technology platforms focused on high-throughput delivery of the key comparative biological data required to select individual targets from any pool of candidate genes (Fig. 2b). In this model, the prioritization of therapeutic targets occurs after systematically generating the minimum set of comparative biological properties that any pool of candidate genes must exhibit to meet the target selection threshold defined for the disease of interest, and thereby progress to drug discovery. The option to incorporate comparative phenotypic information derived from gene

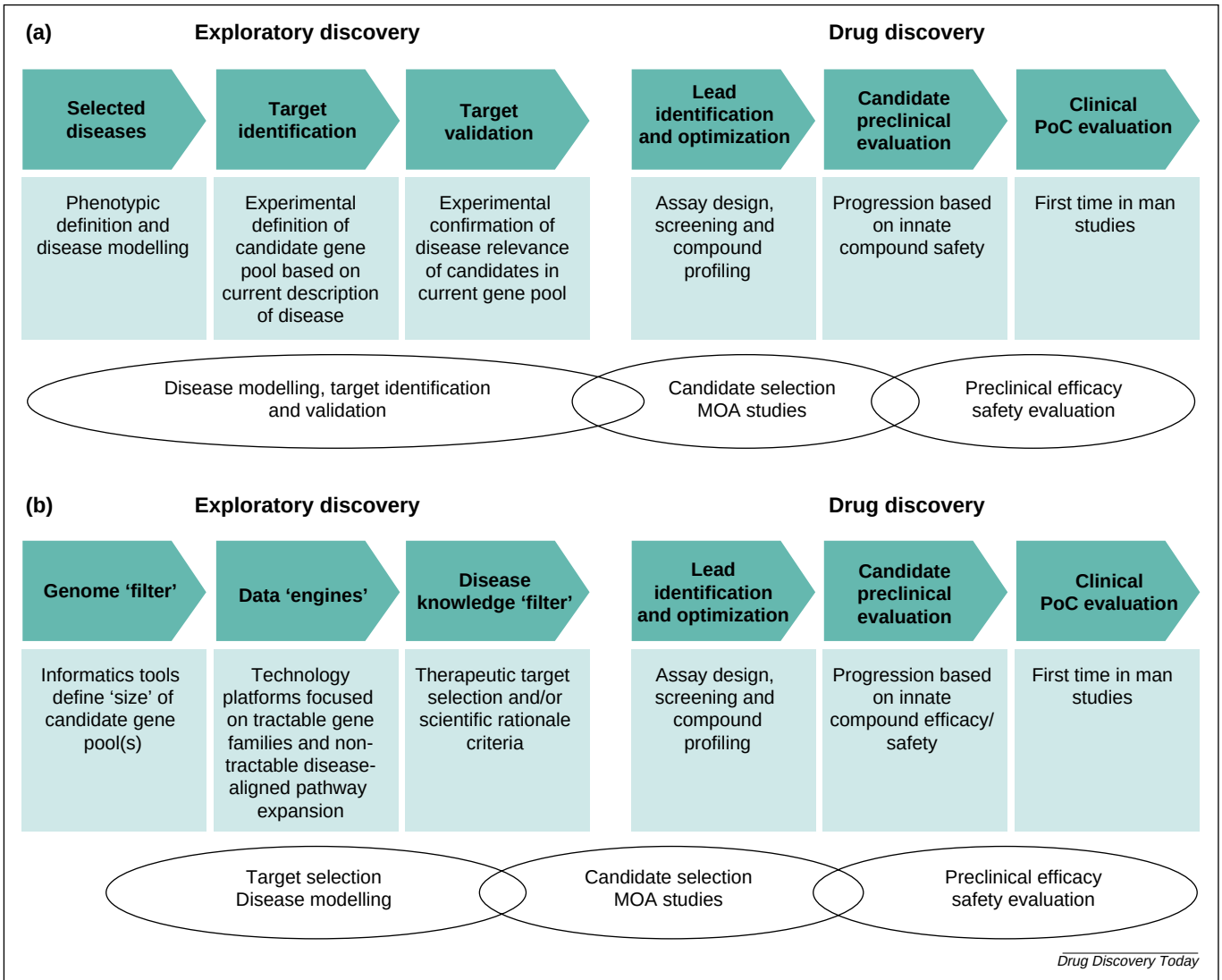

Figure 2. Therapeutic target selection: disease-based and gene-based approaches. Information obtained from transgenic gene knockouts can be used in decision making at various points within the overall pharmaceutical discovery process, several stages of which are illustrated here. The greatest effect of the timely delivery of comparative phenotypic information from gene knockouts will be within exploratory discovery, where the selection and validation of therapeutic targets from within the portfolio of chemically tractable candidate genes is becoming increasingly important. The gene-knockout animals, tissues or cells derived from them might also prove useful at several other stages of the drug discovery process. Abbreviations: MOA, mode of action; PoC, proof of concept.

knockouts into this strategy has, until recently at least, been limited by the resources required to systematically generate and characterize a large number of gene knockout transgenics in a timely manner[5] (discussed also later).

Attributing a relative value to biological data

Faced with a portfolio of 5000–10,000 chemically tractable candidate genes, there are few, if any, technology platforms that can provide definitive biological evidence that a particular candidate gene, or subset of candidate genes, are therapeutically relevant in a systematic or timely fashion[2,6]. The industry is, therefore, in a transition phase in which it is currently obliged to make a series of pragmatic, intermediate target-selection decisions based on incomplete biological information, while continuing to seek access to the comparative biological information it needs to make better informed decisions in the future. Based on historical precedent, these biological data filters fall into three categories, or combinations thereof, that broadly reflect their relative value to the overall therapeutic target-selection threshold depicted in Fig. 1: (1) expression studies, (2) *in vitro* and *ex vivo* functional studies and (3) molecular genetics and *in vivo* studies. Moreover, biological insight gained from the human population, or material sourced from the human

population, has the highest value but data from one or more comparative species are often used as a surrogate in exploratory biology[2,5,6].

Gene expression studies

Gene expression profiling has a relatively high throughput and a low to medium value, and is often used as a 'first-pass' means of defining a subset of genes with an appropriate tissue and/or cellular distribution pattern considered relevant to the pathophysiology of the disease of interest. Array-based transcriptional profiling[11] and proteomics[12] are complementary approaches that are radically changing the way in which such expression studies are conducted. Both generate significant amounts of information about steady-state expression levels and/or changes in expression profile under differing conditions (e.g. diseased and undiseased tissue, treated and untreated cells). Using these approaches, not only can changes in expression profile be observed but also some degree of functional validation can be inferred, given appropriate experimental design and a suitable framework of prior art. The highly parallel nature of these platforms can, however, result in large volumes of data, which in itself can be a challenge when considering which candidate genes to take forward for further validation.

In vitro *and* ex vivo *functional studies*

There are several *in vitro* and *ex vivo* approaches for performing functional studies with low to medium throughput and medium to high value. They are based on well-established techniques such as the yeast two-hybrid system, cloning by complementation and expression cloning and so on. They all provide a means of identifying and/or validating a candidate therapeutic target, usually within a defined cellular or biochemical context. The challenge will be designing and scaling these mechanism-based approaches and combining them with gene-modulation tools such as antisense oligonucleotides[13] to directly assess the functional contribution of candidate genes from within the chemically tractable target classes. In the long term, these types of studies will contribute to intracellular 'pathway maps', especially when combined with expression profiling techniques. These maps will allow researchers rapidly to make biological connections between disease-aligned intractable candidate genes to those chemically tractable genes that are most likely to be developed into products with a therapeutic effect.

Molecular genetics and in vivo *studies*

Molecular genetics and *in vivo* studies have a low throughput and a medium to high value. Population genetics in both humans and comparative species has a proven track record of unambiguously identifying genes responsible for particular disease phenotypes[14–16]. Although this success has, in the main, been restricted to monogenic disorders, recent advances in this area are beginning to extend the utility of this approach to the more common polygenic disease states. Rodents, especially the mouse, are expected to play an increasingly significant role in determining the functional significance that specific genes play in complex diseases states[5].

Although molecular genetics provides a high degree of confidence that a particular gene is responsible for a specific disease state, there is a relatively low probability (~0.1) that the gene will be a member of a chemically tractable gene family. Furthermore, it is proving to be a significant challenge to translate a genetically defined but chemically intractable gene of limited or unknown function into a chemically tractable therapeutic target that is amenable to drug discovery (i.e. to cross Zone B, Fig. 1). In many organizations, the resources available to gain the additional biological understanding required to progress with a novel genetically defined candidate gene or pathway within exploratory research are limited, relative to the study of chemically tractable candidate genes.

When considering how best to address the challenge of gathering sufficient biological evidence to allow comparative, evidence-based therapeutic target selection decisions, the industry is, therefore, faced with some difficult, and often expensive, choices. The scientific rationale and business assessment (principally the risk:return ratio) leading to any portfolio of investments in a high-throughput biology capability has, ultimately, to be a matter of individual organizational judgement based on several factors:

- The perceived value that an individual biological observation makes to the comparative therapeutic target selection process.
- An assessment of the likelihood that a sufficient proportion of all such observations will be of similar value, thereby enabling candidate genes to be prioritized as part of a comparative, evidence-based therapeutic target selection strategy.
- The perceived returns justify the overall cost of gaining access to the information or tools.

Functional genomics and transgenic gene knockouts

The selection of biological targets for the development of potential new medicines relies, in part, on the quality of the *in vivo* biological data that relates a particular molecular target with the underlying pathophysiology of a disease. Since the late 1970s, several techniques have been developed that allow the production of transgenic animals with defined genome modifications, and so the mouse has increasingly become the species of choice for mammalian

gene function studies (Box 1). Within the pharmaceutical industry, transgenic animals, especially gene knockouts, are proving to be invaluable sources of functional information and tools that can be used in studies at various other stages of the drug discovery process (Fig. 2). For example, in preclinical candidate drug selection, information obtained from gene knockout and/or gene addition transgenics is increasingly being accepted as a viable cost-effective alternative for mutagenicity and carcinogenicity testing[17,18]. In addition, where relatively imprecise pharmacological reagents are available, gene knockouts can be used to define the biological mode of action by helping to discriminate between the *in vivo* gene function(s) of closely related members of a gene family[19–22]. These few examples illustrate the potential power of comparative studies when the appropriate gene knockout reagents are available.

The most significant impact of transgenics is currently in the exploratory phase, where gene knockouts are predominantly, but not exclusively, created to support target validation as part of a disease-to-target strategy (Fig. 2a). The manipulation of gene function *in vivo* can provide a high degree of confidence that the gene of interest is a crucial component of the biology under investigation and thereby help to focus scarce resources on progressing candidates that exhibit the phenotype of greatest clinical relevance (Box 2). The potential of gene knockout transgenics to contribute to high-throughput target selection and/or validation has hitherto been limited by their availability, and/or the resources required to generate and characterize a large number of gene knockouts.

Box 1. Generating and characterizing transgenic animals

Transgenic animals are commonly generated either by pronuclear DNA microinjection[a] or by gene targeting via homologous recombination in embryonic stem cells[b]. The recent demonstration that gene targeting can be performed in sheep means that targeted gene modification might become routinely available in other species[c–e]. Although there are several theoretical and practical caveats and limitations associated with using gene addition and gene knockout transgenics in functional analysis, including copy number, site of integration effects, embryonic lethality and genetic background effects, the continued development and adoption of conditional knockout and knock-in approaches, along with other techniques, will probably provide the opportunity to overcome some of these limitations and thereby obtain further mechanistic insights into *in vivo* gene function[f–l].

In the context of exploratory drug discovery, the success of this technology platform for target selection ultimately depends on the phenotypic description of the gene knockout used to make a gene-to-disease correlation. A serious 'phenotyping gap' is emerging that is, in part at least, a result of the practical considerations inherent in establishing the breadth and depth of first-pass analysis currently used in the high-throughput phenotypic screening of both chemically induced mutants and gene knockouts[m,n]. For some diseases, it will be essential, albeit challenging, to invest in the secondary and tertiary phenotypic screens, including aging studies, that will be required to build confidence in comparative target selection in certain therapeutic areas.

References

a Gordon, J.W. and Ruddle, F.H. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 21, 1244–1246
b Capecchi, M.R. (1994) Targeted gene replacement. *Sci. Am.* 270, 34–41
c Capecchi, M.R. (2000) Choose your target. *Nat. Genet.* 26, 159–161
d McCreath, K.J. *et al.* (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405, 1066–1069
e Suraokar, M. and Bradley, A. (2000) Genetics. Targeting sheep. *Nature* 405, 1004–1005
f Baudoin, C. *et al.* (1998) Knockout and knockin of the beta-1 exon D define distinct roles for integrin splice variants in heart function and embryonic development. *Gene Dev.* 12, 1202–1216
g Lobe, C.G. and Nagy, A. (1998) Conditional genome alteration in mice. *Bioessays* 20, 200–208
h Brusa, R. (1999) Genetically modified mice in neuropharmacology. *Pharmacol. Res.* 39, 405–419
i Geng, Y. *et al.* (1999) Rescue of cyclin D1 deficiency by knockin cyclin E. *Cell* 97, 767–777
j Leissring, M.A. *et al.* (2000) Capacitative calcium entry deficits and elevated luminal calcium content in mutant presenilin-1 knockin mice. *J. Cell Biol.* 149, 793–797
k Ohtani, T. *et al.* (2000) Dissection of signaling cascades through gp130 *in vivo*: reciprocal roles for STAT3- and SHP2-mediated signals in immune responses. *Immunity* 12, 95–105
l Van Dijk, K.W. *et al.* (2000) Use of transgenic mice to study the role of apolipoprotein E in lipid metabolism and atherosclerosis. *Int. J. Tissue React.* 22, 49–58
m Williams, R.S. and Wagner, P.D. (2000) Transgenics animals in integrative biology: approaches and interpretations of outcome. *J. Appl. Physiol.* 88, 1119–1126
n Nadeau, J.H. (2000) Muta-genetics or muta-genomics: the feasibility of large-scale mutagenesis and phenotyping programs. *Mamm. Genome* 11, 603–607

Target selection using large-scale knockout phenotyping

The value (impact) of a gene knockout phenotype on the target selection and candidate drug progression process can be assessed by examining how the phenotypic information derived from gene knockouts has been historically used in decision making within the drug discovery process. Although this exercise can be performed for any pool of candidate genes, the example used here summarizes the conclusions for one class of chemically tractable gene, the gene family encoding the GPCRs (Ref. 5). An internal survey of discovery projects revealed that, if gene knockout information were readily available, up to ~25% of all non-olfactory *GPCR* gene knockout phenotypes could contribute significantly to the scientific rationale leading to the selection of a *GPCR* family member as a therapeutic target for drug discovery. A further ~60% of gene knockout phenotypes could potentially help to discriminate between candidate genes within a gene family whereas 10–15% of

Box 2. Case study: obesity phenotype, gene knockouts and therapeutic target selection

Obesity is becoming an increasingly important health problem around the world, not least because it causes or exacerbates other disease states. Remarkable progress has been made recently in understanding body-weight regulation through studies of this phenotype in the mouse. These mechanistic insights into the complex pathophysiology of this disease have been gained using both molecular genetic approaches and targeted gene knockouts, and are, in some cases at least, being complemented by studies in humans[a].

More specifically, the phenotypes of knockout mice lacking genes encoding components of the melanocortin system have highlighted the fact that specific members of the G-protein-coupled melanocortin receptor (MCR) family are involved in regulating body weight through distinct and complementary mechanisms[b]. Moreover, within this five-member gene family, the available gene knockouts have helped to distinguish distinct functional roles for the *MCR-3/4* and *MCR-5* genes in spontaneous obesity and exocrine gland dysfunction, respectively[c–f].

This example illustrates the effect that phenotypes exhibited by gene knockouts might have on therapeutic target selection. That is, the primary output of a comparatively high-throughput knockout phenotyping capability can help to focus rate-limited R&D resources on candidates with the greatest therapeutic, and thereby commercial, potential. Moreover, it is specifically proposed that, for the pool of chemically tractable genes of uncertain or unknown function within the human genome, accessing comparative *in vivo* phenotypic information from gene knockouts will provide a competitive advantage if incorporated within future therapeutic target selection and validation strategies within the pharmaceutical industry.

References

a Barsh, G.S. *et al*. (2000) Genetics of body-weight regulation. *Nature* 404, 644–651
b Cummings, D.E. and Schwartz, M.W. (2000) Melanocortins and body weight: a tale of two receptors. *Nat. Genet.* 26, 8–9
c Chen, W. *et al*. (1997) Exocrine gland dysfunction in MC5-R-deficient mice: evidence for coordinated regulation of exocrine gland function by melanocortin peptides. *Cell* 91, 789–798
d Huszar, D. *et al*. (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 88, 131–141
e Butler, A.A. *et al*. (2000) A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* 141, 3518–3521
f Chen, A.S. *et al.* (2000) Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nat. Genet.* 26: 97–102

gene knockout phenotypes would initially be misleading in the absence of additional biological information.

This sample set is small and inherently biased towards those gene family members with significant prior art, including pharmacological reagents and/or natural ligands. The analysis was, therefore, extended to include an assessment of the gene knockout phenotypes observed for a total of 63 *GPCR* genes, as described in 111 publications. Approximately, 50% of those gene knockouts examined gave rise to a readily observed phenotype. Of these, ~25% were either overt phenotypes (e.g. obesity, infertility, aggressiveness, embryonic lethality) or phenotypes revealed by a relatively simple experimental measure (e.g. altered pain threshold, haematology, clinical chemistry, behaviour). Approximately 30% of gene knockouts required a more sophisticated experimental challenge before a clinically relevant pathophysiological phenotype was observed (e.g. impaired cellular recruitment in inflammation, altered cognition and memory, glucose intolerance).

These conclusions are dependent upon the level of phenotypic characterization within the literature; for example, a single gene knockout might have exhibited an overt, a simple and/or a complex phenotype and, therefore, contributed more than once to this analysis. Furthermore, based on the experimental hypotheses tested within the literature, ~50% of gene knockouts had no discernible phenotype and, therefore, could not be used to derive a meaningful disease alignment for the gene in question, at least in the absence of a more broadly based or sophisticated battery of phenotypic tests.

These two historical analyses strongly suggested that both the frequency and the value (impact) of comparative gene knockout phenotypes observed for members of the chemically tractable gene families could make a significant contribution to therapeutic target selection, if available at scale and in a timely manner. These general conclusions are highly dependent on the breadth, depth and quality of the comparative phenotypic analysis performed and, in this case, published for each gene knockout. They are also sensitive to the type of phenotypic selection criteria and relative target selection thresholds used in different organizations. Furthermore, for such information to lead to meaningful discrimination between a portfolio of candidate genes, it is important that any decision-making process takes account of any 'negative' phenotype (i.e. does not fit the prevailing scientific rationale) and 'no discernible' phenotype results when prioritizing therapeutic targets.

Accessing a high-throughput knockout phenotyping capability

The key to making comparative evidence-based target selection decisions is the timely delivery of phenotypic information for decision making, rather than the resources to generate the gene knockouts and performing the phenotypic

screens[23]. Many pharmaceutical companies have some in-house transgenic capability or outsource their needs, either via collaboration or to those commercial organizations that generate transgenics on a fee-for-service basis. There is little evidence to suggest that any individual pharmaceutical company is willing to commit their own resources to generate, breed and, most importantly, characterize a large number of gene knockouts in a high-throughput, systematic and timely manner in the near future.

The option of outsourcing phenotypic information from gene knockouts and/or accessing tools, albeit at a reasonable cost, has recently become a reality as biotechnology companies have set out to address the challenges involved in developing a high-throughput gene knockout production and/or phenotyping capability, including Deltagen, Lexicon Genetics and Paradigm Therapeutics. These companies each provide a mechanism by which interested parties can gain access to many candidate gene knockouts and/or the phenotypic information derived from them upon which to make comparative evidence-based target selection decisions. In addition, they offer some type of fee-for-service arrangement whereby potential partners can access specific tools (gene knockout, conditional knockout and knock-in transgenics) for in-house or collaborative phenotypic studies. In this regard, they are similar to several other companies offering transgenic fee-for-service arrangements, albeit seeking to deliver reagents in an accelerated timeframe.

Among other activities, Deltagen (Menlo Park, CA, USA; http://www.deltagen.com/) offers potential partners the option to take out a non-exclusive subscription to Deltabase™. This is a proprietary database that will contain phenotypic information derived from a portfolio of gene knockouts selected from the pool of drugable gene family members. Deltagen has selected >1000 mammalian genes thought to be relevant to small-molecule drug discovery for inclusion in Deltabase. The focus on the 'up-front' delivery of primary comparative phenotypic information for a significant portfolio of gene knockouts of potential interest to the pharmaceutical industry is the distinctive feature of Deltabase. A subscription to Deltabase also provides access to the knockout mice for additional phenotypic studies, as required. The immediate value (impact) of the primary comparative phenotypic information within Deltabase for target selection will depend on how well the breadth and depth of the phenotypic analysis performed by Deltagen aligns with the therapeutic selection criteria of a potential subscriber.

Lexicon Genetics (Woodlands, TX, USA; http://www.lexicon-genetics.com/) is developing Omnibank™, a proprietary sequence database linked to a physical bank of pretargeted mouse embryonic stem cell clones generated by random insertional mutagenesis[24]. This strategy overcomes the need to perform the often time-consuming molecular and cellular biology that is involved in targeting large numbers of candidate genes on a gene-by-gene basis. Once knockout animals have been generated, any phenotypes exhibited by the gene-trap event can be explored either via in-house analysis or using the range of phenotypic screening options Lexicon offers to partners. The Omnibank concept is particularly suited to those interested in gaining access to gene knockout information for potentially any of the expressed genes within the genome. This includes those seeking to validate the functional significance of a portfolio of candidate genes implicated in a priority therapeutic area, or that reside within a defined region of the genome as determined by a genetic approach. More recently, Lexicon has introduced its LexVision™ programme, which closely resembles the Deltabase concept offered by Deltagen. As previously, the impact of the information within the LexVision data set on target selection will depend upon the portfolio of genes analyzed and how well the phenotypic information gathered by Lexicon aligns with the therapeutic selection criteria of a potential subscriber.

Paradigm Therapeutics (Addenbrooke's Hospital site, Cambridge, UK) offers partners access to several transgenic technology platforms. Their aim is to use optimized molecular, cellular and husbandry techniques to generate and discern the phenotypes of gene knockouts for a portfolio of candidate genes from within the drugable gene families, in this case with an emphasis on CNS and metabolic diseases.

Future prospects

The impending fruition of the various genome sequencing projects is helping to define the pool of chemically tractable candidate gene targets and is thereby shifting the emphasis from target identification and validation *per se* to target selection. In the short term, gene-to-target candidate gene selection strategies are likely to have an increasingly significant impact on therapeutic target selection decisions because, by definition, they can be directed towards the chemically tractable classes of gene family. The selection of biological targets for the development of potential new medicines relies, in part, on the quality of the *in vivo* biological data that correlates a particular molecular target with the underlying pathophysiology of a disease. In recent years, transgenic techniques, especially gene targeting, have revolutionized our ability to infer the biological function(s) of genes within an *in vivo* mammalian context.

Within the pharmaceutical industry, the opportunity to use comparative phenotypic information derived from gene knockouts as part of any high-throughput, evidence-based target selection and validation strategy has been limited by availability, either through the literature or by a resource-constrained internal capacity. The emergence of several biotechnology companies specifically focusing on the high-throughput generation and phenotyping of gene knockouts means that the industry now has the opportunity to access this highly informative source of phenotypic information and/or tools in a timely fashion on a previously unprecedented scale. The portfolio of gene knockouts and the breadth and depth of the phenotypic analysis on offer will both be key factors that will ultimately dictate the choice, scale and effect of future pharmaceutical–biotechnology company partnerships in this area.

It is both an exciting and a challenging time to be involved in establishing functional genomics strategy within the pharmaceutical industry. Recent advances in comparative molecular genetics and other techniques have heightened interest in the mouse as a means of identifying the genes underlying both monogenic and polygenic disease states[5]. Foremost among these is the prospect of gaining access to additional phenotypes from gene knockout mice on a previously unprecedented scale, thereby extending the portfolio of biological information used within a comparative, evidence-based target selection strategy. The challenge for individual corporate bodies will be balancing their strategic investments in mammalian gene knockouts against the other types of biological insight offered by the alternative functional genomic technology platforms.

Acknowledgements

I would like to thank the many colleagues within the former GlaxoWellcome and elsewhere whose enthusiasm for basic science and drug discovery helped to shape the views expressed within this article. In particular, Steven Foord for sharing his extensive knowledge of the GPCR literature and Glenda Watson for her comments on the manuscript.

References

1 Drews, J. (1996) Genomic sciences and the medicines of tomorrow. *Nat. Biotechnol.* 14, 1516–1518
2 Spence, P. (1998) Obtaining value from the human genome: a challenge for the pharmaceutical industry. *Drug Discov. Today* 3, 179–188
3 Spence, P. (1999) Genomics: the race is on. *Drug Discov. Today* 4, 103–104
4 Dyer, M.R. *et al.* (1999) Functional genomics: from genes to new therapies. *Drug Discov. Today* 4, 109–114
5 Harris, S. and Foord, S. (2000) Transgenic gene knock-outs: functional genomics and therapeutic target selection. *Pharmacogenomics* 1, 433–443
6 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
7 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
8 Aparicio, S.A.J.R. (2000) How to count…human genes. *Nat. Genet.* 25, 129–130
9 Sanseau, P. (2001) Impact of the human genome sequence for *in silico* target discovery. *Drug Discov. Today* 6, 316–323
10 Thomas, S.M. (1999) Genomics and intellectual property rights. *Drug Discov. Today* 4, 134–138
11 Young, R.A. (2000) Biomedical discovery with DNA arrays. *Cell* 102, 9–15
12 Thomas, J.D. and VanBogelen, R.A. (2000) Experimentalism in Proteomics. In *Proteomics: A Trends Guide* (Blackstock, W. and Mann, M., eds), pp. 7–11, Elsevier Science
13 Taylor, M.F. *et al.* (1999) Antisense oligonucleotides: a systematic high-throughput approach to target validation and gene function determination. *Drug Discov. Today* 4, 562–567
14 Antonarakis, S.E. and McKusick, V.A. (2000) OMIM passes the 1,000-disease-gene mark. *Nature* 25, 11
15 Bedell, M.A. *et al.* (1997) Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. *Genes Dev.* 11, 1–10
16 Bedell, M.A. *et al.* (1997) Mouse models of human disease. Part II: recent progress and future directions. *Genes Dev.* 11, 11–43
17 Dean, S.W. *et al.* (1999) Transgenic mouse mutation assay systems can play an important role in regulatory mutagenicity testing *in vivo* for the detection of site-of-contact mutagens. *Mutagenesis* 14, 141–151
18 Storer, R.D. (2000) Current status and use of short/medium term models for carcinogenicity testing of pharmaceuticals – scientific perspective. *Toxicol. Lett.* 112, 557–566
19 Kieffer, B.L. (1999) Opioids: first lessons from knockout mice. *Trends Pharmacol. Sci.* 20, 19–26
20 Kable, J.W. *et al.* (2000) *In vivo* gene modification elucidates subtype-specific functions of alpha(2)-adrenergic receptors. *J. Pharmacol. Exp. Ther.* 293, 1–7
21 Nyce, J.W. (1999) Insight into adenosine receptor function using antisense and gene-knockout approaches. *Trends Pharmacol. Sci.* 20, 79–83
22 Murphy, D.L. *et al.* (1999) Molecular manipulations as tools for enhancing our understanding of 5-HT neurotransmission. *Trends Pharmacol. Sci.* 20, 246–252
23 Capecchi, M.R. (2000) Choose your target. *Nat. Genet.* 26, 159–161
24 Zambrowicz, B.P. *et al.* (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* 392, 608–611