OPINION

Common pitfalls in preclinical cancer target validation

William G. Kaelin Jr

Abstract | An alarming number of papers from laboratories nominating new cancer drug targets contain findings that cannot be reproduced by others or are simply not robust enough to justify drug discovery efforts. This problem probably has many causes, including an underappreciation of the danger of being misled by off-target effects when using pharmacological or genetic perturbants in complex biological assays. This danger is particularly acute when, as is often the case in cancer pharmacology, the biological phenotype being measured is a 'down' readout (such as decreased proliferation, decreased viability or decreased tumour growth) that could simply reflect a nonspecific loss of cellular fitness. These problems are compounded by multiple hypothesis testing, such as when candidate targets emerge from high-throughput screens that interrogate multiple targets in parallel, and by a publication and promotion system that preferentially rewards positive findings. In this Perspective, I outline some of the common pitfalls in preclinical cancer target identification and some potential approaches to mitigate them.

Pharmaceutical and biotechnology companies rely heavily on academic laboratories to identify protein targets, the functions of which, if modulated, would produce therapeutic effects in patients. The supposition that modulating a particular target will cause a given therapeutic effect is often based on the results of laboratory experiments in which the protein of interest was genetically or chemically manipulated in cells (*in vivo* or *ex vivo*) that were then scrutinized with respect to phenotypes believed to be predictive of a desired therapeutic outcome. This process of preclinical target identification and validation is at the foundation of modern drug discovery because it determines which candidate targets are likely to receive the substantial investments in time and money needed to make drugs against them. Unfortunately, both published and unpublished studies suggest that a disturbingly high percentage of the findings in papers published by academic investigators related to preclinical cancer target validation cannot be reproduced or are not sufficiently robust to justify moving forward with drug discovery efforts[1–5]. In this Perspective I try to highlight some of the likely contributors to this problem and offer some potential solutions.

Reproducibility versus robustness

It is worth first commenting on the distinction between lack of reproducibility and lack of robustness. The inability to reproduce the findings of a given published experiment under ostensibly the same conditions suggests several possibilities. One, hopefully rare, explanation is that the original results were fraudulent. Another possibility is that the original results were true but occurred either by chance (that is, they were noise rather than signal) or as a result of operator error.

Many times, however, the explanation relates to the practical impossibility of precisely replicating the conditions of the original set of experiments as that would, in principle, require knowing and controlling for a potentially infinite number of variables, including variables that might not have originally been suspected of influencing the experimental outcome. This problem was exacerbated by the trend that began more than 20 years ago of marginalizing and minimizing the descriptions of experimental methods in published papers. As a result, published experimental protocols frequently lack important experimental details. The ability to post detailed descriptions of experimental methods online, however, might help to mitigate this problem in the future. Similarly, increased sharing of data (whether in print or online) might lead to a better understanding of the true variability of experimental results and provide insights into reproducibility.

This then leads to a discussion of robustness. In colloquial terms, robustness is sometimes used to refer to size or strength, whereas robustness really refers to the ability to withstand perturbations. For example, commercial aeroplanes and the Internet are designed to be robust; they will continue to function even if a certain number of their subcomponents fail. Although it is true that large effects are more likely, a priori, to be robust than small effects, in the context of the laboratory, the robustness of a finding really refers to the degree to which that finding holds true over a range of experimental conditions. A result that is true but only true under an extremely narrow set of conditions is not robust irrespective of effect size. Robust results are more likely to make predictions that will be true under real-world conditions, such as those that occur in human clinical trials.

The classic way to minimize noise and to increase robustness is to ensure that results have been reproduced multiple times and, whenever possible, under a variety of conditions (for example, by studying more than one cell line or a range of buffer compositions). Ideally, particularly important results would also be independently replicated or 'beta tested' by another scientist or technician before publication. Investigators should be encouraged to provide a sense of the range of conditions they explored to obtain their findings and should, ideally, describe the heterogeneity they encountered in the course of their studies (for example, responses to a particular drug that are heterogeneous

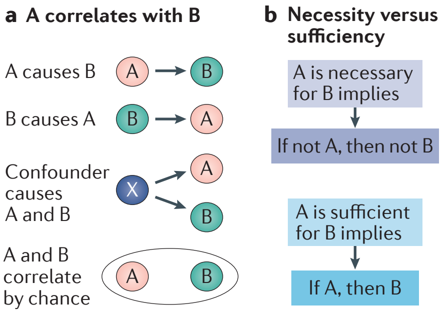
a A correlates with B
A causes B
A
B
B causes A
B
A
Confounder causes A and B
X
A
B
A and B correlate by chance
A
B
b Necessity versus sufficiency
A is necessary for B implies
If not A, then not B
A is sufficient for B implies
If A, then B

Figure 1 | **Distinguishing correlation from causation. a** | Possible reasons that two variables, A and B, correlate with one another. **b** | Definitions of necessity and sufficiency.

across cell lines or heterogeneous responses to different drugs that inhibit the same target). Heterogeneity in cancer biology is the rule rather than the exception, in both the laboratory and the clinic. Unfortunately, this can create conflicts with some editors and reviewers who prefer papers that are seemingly whole and internally consistent. Punishing authors who acknowledge the reality of cancer heterogeneity by including outlier results (for example, by including both sensitive and insensitive cell lines in the initial characterization of a new target or treatment, even if the differential sensitivity is not yet fully understood) exacerbates the problem of lack of robustness because it encourages investigators to 'cherry-pick' their findings. It also delays progress because it suppresses the publication of unexpected findings that, not infrequently, stimulate new discoveries and new ways of thinking.

Another factor that contributes to an apparent lack of reproducibility and robustness is the frequency with which investigators draw, and then publish, erroneous conclusions from their own experimental data. This is a problem because in most cases investigators do not attempt to literally repeat the experiments reported by others (which, as discussed above, is often impossible from a practical standpoint). Instead, they ask whether the predictions made by the conclusions of others (for example, that inhibiting a particular protein kills cancer cells) holds true in their models and systems. When those predictions prove to be false, the original work is assumed to have been flawed or at least not robust. It is likely that many of the published conclusions surrounding potential new cancer targets are incorrect or misleading because of illogical reasoning with respect to, for example, cause and effect relationships, and pitfalls associated with perturbing a complex system, such as a cancer cell, and then subjecting that system to a 'down' assay.

In this Perspective I attempt to highlight the importance of logical thinking in experimental design and data interpretation. Although my focus is on preclinical cancer target identification and validation, many of these concepts transcend target identification and validation and apply to experimental biology and clinical medicine more broadly.

Causation versus correlation

Several technological advances over the past few decades have made it much easier to generate cancer-relevant data, in both small-scale experiments carried out by individual investigators and large-scale experiments conducted by teams of investigators. Moreover, increasing amounts of these data are publicly available and mineable by third parties. To draw meaningful conclusions from such data, it is essential to distinguish between correlation and causation.

Two things (A and B) can correlate with one another for several reasons (FIG. 1a). For example, A might cause B, B might cause A, or perhaps both statements are true. However, there are many examples in biology where two things that correlate with one another do not have a causal relationship (that is, A does not cause B and B does not cause A). Sometimes this occurs because both A and B are under the control of a confounder, X, that causes both of them. Other times this is because A and B simply correlate with one another by chance. It is especially important to consider this latter possibility when the correlation of A and B emerges from testing for many such correlations simultaneously, a process referred to as multiple hypothesis testing. For example, it is well appreciated that retrospectively analysing disappointing clinical trial data in the hope of identifying subgroups that might nonetheless have benefited from a new therapy is fraught with danger unless those subgroups were prespecified and the study was powered to specifically detect differences in those subgroups. Otherwise, apparent differences can emerge by chance, especially if the number of subgroups analysed is large. Likewise a certain number of shRNAs will be randomly lost or gained in any pooled shRNA-based screen that targets, for example, thousands of genes with multiple shRNAs per gene, including instances in which more than one shRNA per gene covaried by chance. This noise needs to be corrected for when drawing conclusions based on shRNAs that are lost or gained under specific experimental conditions (for example, cell lines of one class versus cell lines of another class). When drawing inferences from analyses of large data sets investigators need to distinguish between hypotheses that they had formed before data analysis and hypotheses that they formed only after data analysis, as the latter require additional steps (for example, using separate training and validation gene expression data sets or testing shRNAs other than the shRNAs that scored in a pooled shRNA screen) to avoid drawing erroneous conclusions.

Necessity versus sufficiency

Clear thinking about causation versus correlation is aided by using words that have precise meanings rather than vague terms, such as 'linked to' and 'associated with', which often create ambiguity (intentionally or not) about whether two things are causally related to one another. Two words that are particularly useful in describing causal relationships are necessity and sufficiency (FIG. 1b). The statement 'A is necessary for B' means that if A is not true, then B cannot be true. The statement 'A is sufficient for B' implies that if A is true, then B will be true.

Failure to distinguish between necessity and sufficiency can lead to illogical conclusions. For example, when BRAF mutations were first detected in malignant melanoma, I heard it argued by some participants at scientific advisory board meetings that mutant BRAF would not be a good drug target because BRAF mutations are also present in benign naevi. However, the latter observation indicated only that BRAF mutations are not sufficient to cause malignant melanoma. The more important question from a therapeutic perspective is whether BRAF activity is necessary for the maintenance of BRAF-mutant melanomas, which has now been answered affirmatively[6].

Likewise, some researchers have intimated that successfully treating tumours with multiple driver mutations will require multidrug combinations aimed at targeting each mutation. This again reflects confusion about necessity and sufficiency. The fact that multiple driver mutations were necessary to cause a particular cancer would imply, in a perfect world, that it was sufficient to target any one of those mutations to have a therapeutic effect (potential exceptions here would include mutations that act purely to compromise the fidelity of DNA replication or act in a 'hit and run' manner). Indeed, in the laboratory this often appears to be the case. For example, restoring p53 function in *p53*$^{-/-}$ tumours often induces apoptosis despite the presence of additional driver mutations[7]. By way of analogy, it is necessary for all of the tumblers to be in place for a combination lock to open, and it is therefore sufficient for any one of the tumblers to be out of place to prevent the lock from opening. Combination therapy will usually be required to effectively control cancer for other reasons, such as the need to minimize the emergence of resistance, but not because a tumour with *n* drivers necessarily requires *n* drugs.

Misinterpretation of Kaplan–Meier data

Correlative clinical data are often misinterpreted when it comes to new cancer drug targets. For example, it is now common practice for cancer target validation papers to include data linking expression of the target in question to poor outcomes in patients, with the implicit message that the former somehow caused the latter. Factoring in such correlations when choosing drug targets is, however, fraught with danger. To see why this is the case, first consider the hypothetical survival curves of patients with chronic lung disease who either have or have never required a ventilator in FIG. 2a. It would come as no surprise that patients who have required a ventilator do worse than those who have not, but no one would interpret these data to mean that ventilators caused the bad outcomes. Instead, requiring a ventilator is a biomarker of more advanced lung disease.

There are many examples in cancer biology of molecular changes that correlate with increased aggressiveness of cancer without necessarily causing the aggressive behaviour. For example, intratumoural hypoxia and the resulting upregulation of the transcription factor hypoxia-inducible factor (HIF) in tumours almost invariably correlates with poor outcomes in patients[8] (FIG. 2b). This could signify that hypoxia and HIF cause some tumours to become more aggressive. Alternatively, it could simply reflect the fact that aggressive tumours outgrow their blood supplies, become hypoxic and therefore upregulate HIF. In this latter scenario, there is a causal relationship between hypoxia, HIF and poor outcomes, but the causality direction has been reversed (FIG. 2c). Other cause-and-effect scenarios could also explain the correlation. For example, there could be tumours in which hypoxia promotes aggressive growth but HIF does not, with HIF serving only as a marker of hypoxia. Similar considerations apply to other proteins that could reflect, without necessarily causing, other hallmarks of cancer, such as certain proteins that increase in abundance when cells enter the cell cycle or accumulate mutations.

Next, consider the hypothetical survival curves of patients with breast cancers that do or do not express the oestrogen receptor (ER) (FIG. 2d). Patients with tumours that are ER$^+$ have better outcomes than patients with ER$^-$ tumours, not because ER inhibits tumour growth but because tumours driven by ER have a more indolent biology than do other breast cancer subtypes. In fact, ER is one of the best-validated targets in cancer. In summary, being associated with a bad prognosis is neither necessary nor sufficient to be a good therapeutic target in cancer.

Correlation + plausibility ≠ causation

Establishing that a correlation between two things reflects a causal relationship usually involves, when feasible, experiments in which one of the two things (A) is perturbed and the effect on the second thing (B) is measured. For example, if disrupting the function of protein A using, for example, genetic or chemical tools, invariably leads to loss of phenotype B, then one can conclude that A is necessary for B. If activation of protein A always produces phenotype B, then one concludes that A is sufficient for B. Of course, these experiments do not distinguish between direct effects and indirect effects. Ideally they would be complemented with experiments, such as biochemical and structural studies, that begin to address the mechanisms by which A causes B.

Unfortunately, a very common shortcut in the literature is to infer causation based on biological plausibility, as in 'A correlates with B, it is plausible that A causes B, therefore A causes B'. For example, it is certainly plausible that the correlation of high HIF levels with poor outcomes in cancer reflects the fact that HIF upregulates genes that encode proteins, such as vascular endothelial growth factor A (VEGFA) and matrix metalloproteinases (MMPs), that promote angiogenesis and tumour invasion[8]. However, this argument ignores the fact that HIF also upregulates the expression of genes encoding proteins that

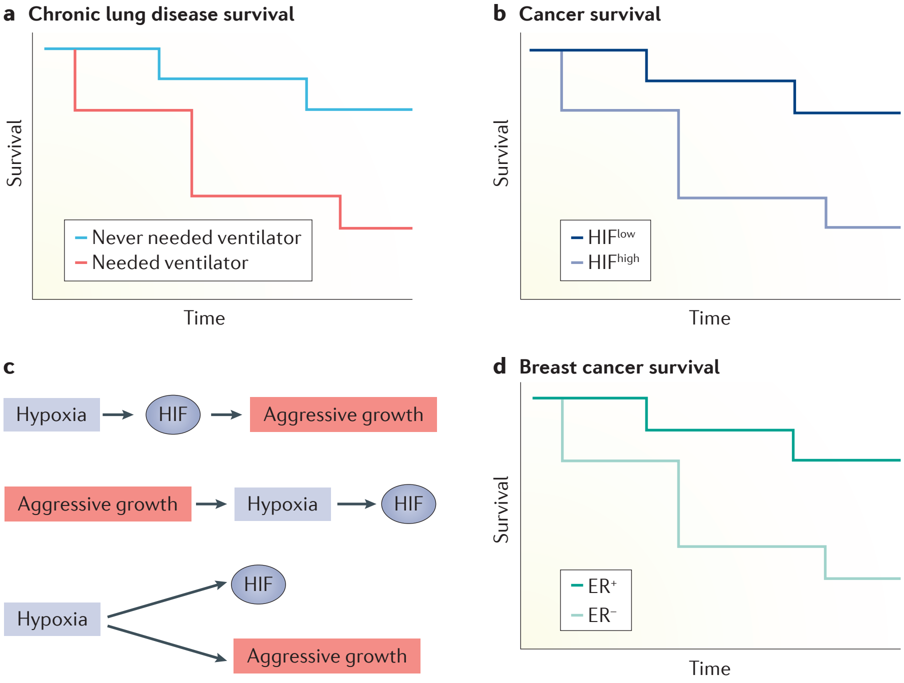

Figure 2 | **Drawing erroneous causal inferences from Kaplan–Meier curves. a** | Hypothetical Kaplan–Meier curve for patients with chronic lung disease. **b** | Hypothetical Kaplan–Meier curve for patients with cancer. **c** | The association between hypoxia and hypoxia-inducible factor (HIF) and poor survival depicted in panel **b** could, in a non-mutually exclusive manner, reflect the ability of hypoxia and HIF to cause aggressive tumour growth or the fact that aggressive tumours are likely to outgrow their blood supplies, become hypoxic and induce HIF. It is also possible that hypoxia causes aggressive tumour growth, but HIF does not. **d** | Hypothetical Kaplan–Meier curve for patients with breast cancer. ER, oestrogen receptor.

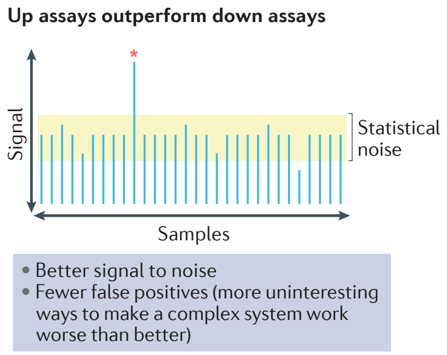
Up assays outperform down assays
*
Signal
Statistical noise
Samples
• Better signal to noise
• Fewer false positives (more uninteresting ways to make a complex system work worse than better)

Figure 3 | **Up assays outperform down assays.** Up assays typically outperform down assays for two reasons: a better signal to noise ratio and fewer false positives. The former relates to the ability to see a positive or increased signal (asterisk) in a field of negative or decreased signals, respectively. The latter stems from the fact that there are more uninteresting or trivial ways to make a complex system perform worse than there are ways to make it perform better. Many pharmacodynamic assays in cancer pharmacology are based on loss of an analyte (for example, a phosphoepitope) and hence are down assays. An analyte that is rapidly induced upon target engagement could form the basis for a pharmacodynamic up assay. Likewise, most chemical and genetic screens in cancer biology are based on down assays, but one can envision screens based on up assays. For example, one could screen for inhibitors of a genetically validated oncogenic pathway in reporter cells in which that pathway actually suppresses cell fitness, owing to naturally occurring or engineered differences between those reporter cells and cells in which that pathway normally promotes tumorigenesis. The former approach might exploit the fact that some oncogenic pathways actually suppress cell growth in certain cellular contexts and the latter might entail cells that have been molecularly engineered to die in response to specific oncogenic signals.

suppress protein synthesis and promote autophagy, processes that can be tumour suppressive[9–11]. Indeed, it appears that the effects of HIF on tumour growth, when tested experimentally, are highly context dependent, with HIF promoting tumour growth in some models and suppressing tumour growth in others[12]. Similarly, a drug against a target known to have a role in cancer survival might engage its target, as determined by pharmacodynamic assays, and might also kill cells. This does not, however, formally prove that the drug kills the cells by inhibiting its intended target. In other words, target engagement and cell killing by the drug might be 'true, true and unrelated'. For example, maternal embryonic leucine zipper kinase (MELK) was widely believed to have an important role in many cancers and small-molecule MELK kinase inhibitors advanced to clinical trials based on their antitumour activity in preclinical models[13]. However, Sheltzer and colleagues[13] recently showed that cell lines used in these preclinical models tolerate CRISPR–Cas9-mediated removal of *MELK* and that such *MELK*$^{-/-}$ cells remain sensitive (with the half-maximal inhibitory concentration (IC$_{50}$) in *MELK*$^{-/-}$ cells being comparable to that in isogenic *MELK*$^{+/+}$ cells) to at least one MELK inhibitor, strongly arguing that the drug was not killing the cells by inhibiting MELK.

Up assays versus down assays

The concepts of necessity and sufficiency are also very helpful when analysing phenotypic assays that are commonly used in cancer biology and cancer pharmacology. Phenotypic assays can be categorized as being either 'up' assays or 'down' assays. In an up assay one looks for an increase in a readout that reflects the performance of the system being measured (for example, increased activity of an enzyme or increased fitness of a cell) whereas in a down assay one looks for a decrease in such a readout (for example, decreased expression of a reporter or decreased fitness of a cell). Up assays typically outperform down assays with respect to delivering biological insights for several reasons. First, they usually have superior signal-to-noise characteristics because it is typically easier to see a positive signal against a background of negatives than the other way around (FIG. 3). In fact, the noise created by random fluctuations in signal can make it impossible to distinguish decreased signals from noise in down assays. More importantly, up assays are less likely than down assays to produce signals that have relatively trivial or uninteresting explanations. This is because the more things that are necessary for a given phenotype, the easier it is to compromise that phenotype (because it should be sufficient to interfere with any one of those necessary things). Put another way, there are many more ways to decrease the performance of a complex system than there are ways to enhance its performance. For example, randomly removing parts from an internal combustion engine or hitting a television set with a hammer will almost always decrease the performance of the engine and the television, respectively, instead of making them work better. Likewise, most athletic injuries result in decreased, rather than increased, athletic performance. It is for these reasons that biologists who design and conduct high-throughput chemical and genetic screens usually prefer up assays.

Unfortunately, the phenotypic assays that are commonly used by cancer biologists and cancer scientists are almost always configured as down assays (FIG. 4). For example, investigators might perturb the function of a particular protein in cancer cells and report that this decreases the viability, proliferation, invasiveness or tumorigenicity of those cells. For the reasons outlined above, such assays are prone to producing results that, although real, are biologically uninteresting and therapeutically intractable. Any perturbation that has the end result of disrupting cellular homeostasis and decreasing cellular fitness might score in such an assay. This latter characteristic of down assays also makes them particularly susceptible to being confounded by off-target effects.

On-target versus off-target effects

Cancer scientists routinely use reagents, such as chemicals, siRNAs, shRNAs and single guide RNAs (sgRNAs), to perturb the functions of specific targets in cancer cells. The phenotypes caused by these perturbants can be due to effects on their intended targets (on-target activity), unintended targets (off-target activity) or some combination of the two (FIG. 5). In practice the total number of off-target activities of such perturbants when applied to cells and animals is unknowable.

When a perturbant causes a 'down' phenotype in a cellular assay it should immediately raise the question of whether that phenotype was due to an unappreciated off-target activity that impaired cellular fitness (FIG. 6). This is especially true when the perturbant has been dosed until it achieved the desired phenotype in a way that was agnostic to target engagement, such as when a drug is used at a concentration well above the concentration needed to inhibit its intended target. A perturbant might achieve the expected pharmacodynamic effect on its intended target, and it might be biologically plausible that perturbing the target would produce the observed phenotype, but that does not prove that the perturbant is causing the phenotype by engaging its intended target (that is, that the phenotype is on-target) (FIG. 6a). Despite this caveat, this is another setting in which investigators frequently implicitly or explicitly invoke arguments based on correlation and plausibility to interpret their findings instead of performing the additional controls necessary to prove causation.

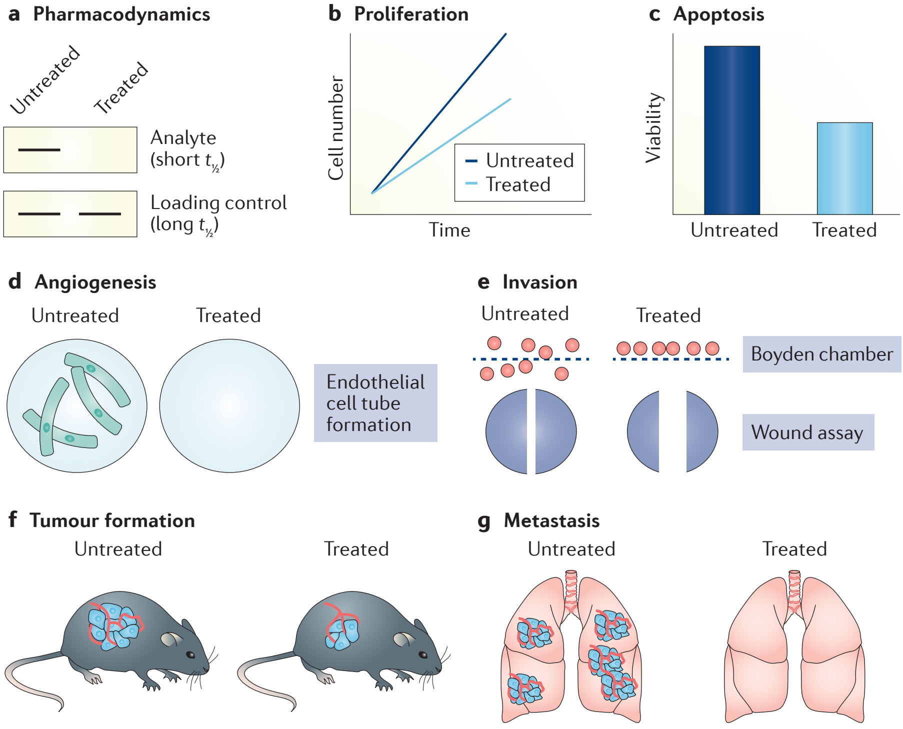

Figure 4 | **Commonly used down assays in cancer biology and cancer pharmacology.** **a** | Pharmacodynamic assay based on loss of an analyte (for example, a protein or phosphoepitope). Note that if the analyte has a shorter half-life than the normalization control (as is often the case), then the analyte will seem to specifically disappear in response to agents that are nonspecifically toxic and cause a global decrease in transcription or translation. **b** | Proliferation assays. **c** | Cell viability assays. **d** | *Ex vivo* angiogenesis assays. **e** | *Ex vivo* invasion assays. **f** | *In vivo* tumour assays such as subcutaneous xenograft assays. **g** | *In vivo* metastasis assays, such as those carried out after tail vein injection of tumour cells.

Rescue experiments

The classic way to show that a phenotype caused by a perturbant is on-target is to show that it can be reversed (or rescued) by a perturbant-resistant version of the target (FIG. 6b). For example, when imatinib mesylate was developed as an inhibitor of ABL kinase activity, it was biologically plausible that it killed chronic myelogenous leukaemia (CML) cells by inhibiting the pathogenic BCR–ABL fusion protein, but the formal proof that the effect of the drug was on-target came with the demonstration that BCR–ABL variants that are insensitive to imatinib mesylate render CML cells resistant to the drug[14,15].

Introducing an expression vector containing a target cDNA that is resistant to a given drug can theoretically be used to assess the extent of off-target toxicity of that drug (FIG. 7). There are several methods for generating such drug-resistant variants of targets of interest. In one approach, a mammalian expression vector for the target of interest is propagated in error-prone *Escherichia coli*, resulting in the accumulation of missense mutations (FIG. 8a). The randomly mutagenized pool of expression vectors is then introduced into a mammalian cell line that is sensitive to the drug. After drug treatment, drug-resistant clones (for example, surviving cells, in the case of drugs that induce cell death) are expanded so that their corresponding target expression vectors can be recovered and sequenced. The target variants identified in this way can then be tested to see whether they are indeed resistant to the biochemical effects of the drug and whether they confer phenotypic drug resistance when introduced into naive cells[15–17].

A complementary method of identifying drug-resistant variants that is target agnostic (that is, does not require prior knowledge of the relevant drug target) has been developed by Tarun Kapoor and co-workers and further optimized by Deepak Nijhawan and co-workers[18–20] (FIG. 8b). In this method, sensitive cells that are naturally mismatch repair (MMR) deficient, and hence riddled with missense mutations, are treated with the drug in question. Resistant cells are clonally expanded and subjected to RNA-seq and whole-exome sequencing to identify candidate drug-resistant alleles, which are then validated, as described above. This approach has the added advantage of potentially identifying alternative resistance mechanisms. In principle, CRISPR–Cas9 can be used to generate MMR-defective cells (by inactivating one or more MMR genes) in situations in which naturally occurring MMR-deficient cell lines are not inherently sensitive to the drug being studied or are otherwise unsuitable because of their biological properties. In a similar vein, Stephen Jackson and co-workers[21] recently described the use of chemically mutagenized haploid mammalian cells to identify drug-resistant target alleles.

In the case of genetic perturbants such as siRNAs, shRNAs and sgRNAs, rescues can be performed by co-introducing an expression vector containing a target cDNA that is insensitive to the perturbant (sgRNAs) or that directs the expression of an mRNA that is insensitive to the perturbant (siRNAs or shRNAs) (FIG. 7). In the case of shRNAs, the rescue cDNA might lack the shRNA-binding site entirely (as can be done when the shRNA targets an untranslated region) or might contain multiple translationally silent mutations (exploiting the redundancy in the genetic code) in the shRNA recognition sequence. Similarly for sgRNAs, the cDNA might lack the sgRNA-binding site entirely (for example, if the sgRNA targets an intron–exon boundary) or might contain translationally silent mutations that destroy the sgRNA-binding site (for example, by eliminating the protospacer adjacent motif (PAM) sequence).

It can be useful in such rescue experiments to compare the behaviour of a cDNA encoding a wild-type version of the target of interest with that of cDNAs encoding mutant versions of the target in which specific functional domains have been disrupted. For example, one could compare the ability of a wild-type kinase to rescue the phenotype with that of a missense mutant of the kinase that is no longer catalytically active. This approach can help to distinguish between phenotypes that, although on-target, are the result of total loss of the target rather than loss of a specific biochemical activity that might be phenocopied by a small-molecule functional antagonist.

Another, non-mutually exclusive, way to reverse an on-target phenotype is to pharmacologically or genetically restore an effector function immediately downstream of the target of interest (FIG. 6c). For example,

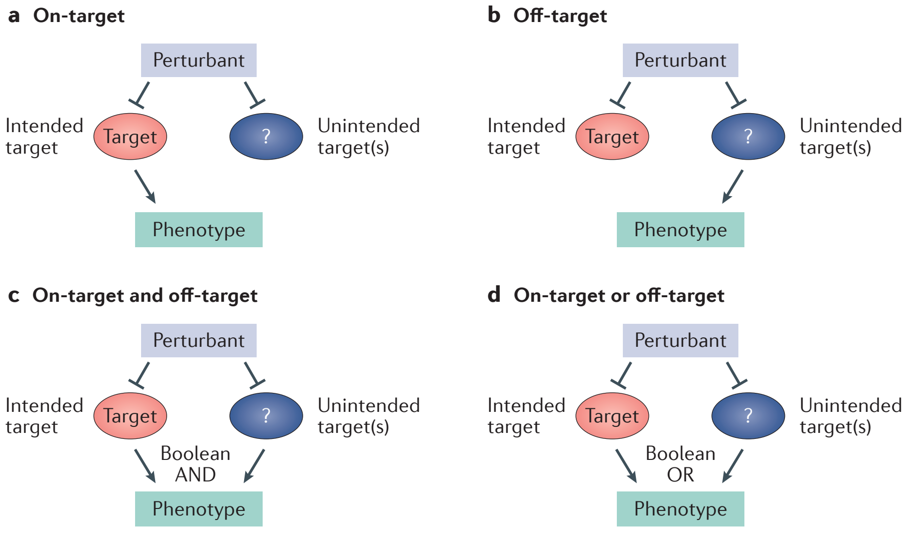

Figure 5 | **On-target versus off-target effects. a** | Perturbant (for example, drug, siRNA, shRNA or single guide RNA (sgRNA))-induced phenotype is on-target when it is caused by engagement of its intended target. **b** | Perturbant-induced phenotype is off-target when it is caused by engagement of an unintended target. **c** | Sometimes the phenotype induced by a perturbant reflects and requires engagement of its intended target and at least one unintended target. Note that in this scenario the phenotype can be reversed (rescued) by mitigating the effect of the perturbant on its unintended target. **d** | Sometimes a phenotype induced by a perturbant reflects and requires engagement of its intended target or at least one unintended target. Note that in this scenario the phenotype cannot necessarily be reversed (rescued) by mitigating the effect of the perturbant on its unintended target.

activating MEK1, which acts downstream of BRAF, renders BRAF-mutant melanoma cells resistant to BRAF inhibitors[22,23]. Similarly, a membrane-permeable version of the oncometabolite *R*-2-hydroxyglutarate (*R*-2HG) reverses the effect of blocking *R*-2HG production by mutant isocitrate dehydrogenase (IDH) in IDH-mutant leukaemia cells[24]. However, exploiting such epistatic relationships in rescue experiments requires prior knowledge of the downstream effector function or functions that are most crucial for the phenotype observed after manipulating the target.

One way to further strengthen the case that a loss-of-function (or gain-of-function) genetic perturbant is acting on-target would be to show that the corresponding gain of function (or loss of function) causes the opposite phenotype. For example, if target knockdown suppresses tumour growth in an on-target manner one would predict that experimentally overexpressing that target would increase tumour growth. However, there are several caveats to this assumption that are important to keep in mind. First, if a tumour cell has already maximally activated a particular oncogenic target, further overexpression of that target may have no effect. Alternatively, overexpressing a particular oncogenic target can, paradoxically, cause a loss-of-function phenotype (for example, by sequestering associated proteins that normally participate in multiprotein complexes with precise stoichiometries) or, in some cases, cause cell death by exceeding some maximally tolerated signal. More generally, the phenotypic effects of some targets may not be linearly related to their expression levels or activities. In other words, some phenotypes caused by subphysiological target levels are not the converse of the phenotypes observed when those targets are present at supraphysiological levels[25].

Boolean logic and rescue phenotypes

The ability to rescue a perturbant phenotype using the approaches described above increases the likelihood that the perturbant is acting on-target, whereas failure to do so should raise suspicion that the perturbant is not acting on-target. However, a caveat with respect to the latter situation is that reversing some phenotypes might require that exogenous expression of the target faithfully mirrors the abundance and physiological regulation (for example, cell cycle-specific modulation) of the endogenous target. One approach to address this issue, at least with respect to protein abundance, is to place the exogenous target under the control of a regulatable promoter and to then eliminate the endogenous target using CRISPR–Cas9. Another approach is to perform rescue experiments with constructs such as bacterial artificial chromosomes (BACs) that harbour some of the target's *cis*-acting regulatory elements[26]. Finally, it should be possible, in principle, to use CRISPR–Cas9 to eliminate siRNA- or shRNA-binding sites or to introduce drug-resistance mutations into endogenous genes.

Another, perhaps more nuanced, concern regarding the interpretation of rescue experiments stems from Boolean logic. If the phenotype induced by a perturbant is caused by an on-target effect and an off-target effect (that is, both are necessary to induce the phenotype) (FIG. 5c), then the phenotype will be rescued when just the on-target effect is prevented. This outcome would be misleading with respect to the ability of a second, independent, perturbant directed against the same target to replicate the phenotype. This is one of the reasons why it is important to probe the consequences of perturbing a target with more than one agent (for example, by examining multiple, independent shRNAs or by complementing genetic approaches with pharmacological tools).

It is also formally possible, especially with down assays, that a perturbant could cause a phenotype because of either its on-target effect or one or more of its off-target effects (that is, either is sufficient to induce the phenotype) (FIG. 5d). In this scenario, a perturbant-resistant target will not rescue the perturbant-associated phenotype, which could lead to a potentially important on-target effect being mistakenly dismissed as being off-target.

Positive and negative controls

The power of any experiment lies in the appropriate use of controls. Positive controls enable interpretation of negative results and negative controls enable interpretation of positive results. For example, a positive control might provide evidence that the absence of a signal for a particular analyte is, in fact, due to the absence of that analyte, rather than to a technical failure of the assay. Conversely, a negative control might provide evidence that a signal for a particular analyte would not be present if the analyte were actually absent. Some negative controls can function as specificity controls to provide information about the permissiveness of an assay. For example, negative controls can be designed to assess how many other analytes could have produced a similar signal in an assay that was designed to interrogate a specific analyte of interest.

Negative controls are particularly important in down assays. At minimum, negative controls when using a chemical perturbant should include the vehicle in which the perturbant is delivered. A better control, when possible, is to use a closely related chemical, such as an alternative enantiomer or a similar chemical based on the same scaffold, that lacks the ability to engage the target. The finding that closely related chemicals that lack activity against the target fail to induce the putative on-target phenotype can significantly strengthen the case that the chemical perturbant is acting on-target. In practice, however, this latter control can be problematic for several reasons. First, academic investigators often do not have access to, or the ability to generate, such chemicals, even if such chemicals were generated as part of a structure–activity relationship study by a company. Second, it is impossible to know, a priori, whether a given chemical modification that enfeebled the on-target activity of a chemical perturbant also lessened its off-target activity. Fortunately, as outlined above, the ability of a perturbant-resistant version of an intended target to reverse the effects of a chemical perturbant provides gold-standard evidence that the effect of the perturbant is at least partially on-target.

A classic way to generate negative controls for peptidic and nucleic acid perturbants is to randomize or 'scramble' their sequences. This controls for effects that are not sequence specific but are instead due to the physiochemical properties (for example, the charge) of the molecules. For example, scrambled controls that are CG rich have been shown to have nonspecific antitumour effects *in vivo* by inducing an interferon response[27]. Unfortunately, the sequences for the negative control siRNA and shRNA sequences used in cancer biology and pharmacology papers are often not provided, sometimes because they are considered proprietary by commercial vendors. Disclosure of such sequences should be a requirement for publication, just as many journals now require disclosure of chemical structures. Moreover, the term scrambled is often used to mean randomly chosen or unrelated, rather than scrambled in the sense described above. This is readily apparent when an experiment includes a single scrambled negative siRNA or shRNA control that is compared with multiple other siRNAs or shRNAs. Finally, and perhaps most concerningly, many negative control perturbants are selected, consciously or unconsciously, because they have been empirically determined to have minimal measurable effects on cellular viability or fitness. For example, a negative control siRNA or shRNA might be chosen because it has been empirically found to have little or no effect on the transcriptome of a cell or be recommended by a colleague because it has been repeatedly found not to score in down assays. This introduces a bias when such negative controls are compared with randomly chosen targeting siRNAs or shRNAs that have not been similarly filtered for potential nonspecific adverse effects.

Importance of corroboration

Any experiment, even if well controlled, can be misleading if viewed in isolation. Corroborating lines of experimental evidence in support of a conclusion increase the likelihood that that conclusion will be correct and robust. For example, a conclusion from an experiment based on a genetic loss-of-function perturbant against a given target is more likely to be true if a similar conclusion emerges from complementary experiments using a chemical perturbant against the same target, or from analysis of the consequences of enhancing the function of the target in gain-of-function experiments. Obtaining similar results with different types of perturbant can decrease the need to carry out rescue experiments with each one, especially when the perturbant is already known to be highly potent and specific (for example, an antibody or small molecule that binds to its intended target with picomolar affinity and that has already been effectively counterscreened against other targets). Conversely, obtaining the same result with multiple perturbants of the same type (for example, multiple shRNAs against a given target) does not fully obviate the need to do rescue experiments. For example, sgRNAs targeting amplified genes decrease cellular proliferation and fitness irrespective of whether those genes are expressed[28,29], and the phenotypes induced by shRNAs across cell lines, even when against the same gene, can be confounded by differences in the levels of argonaute 2 (AGO2) and DICER, which are needed for shRNA processing (W. Hahn, personal communication).

Similarly, the conclusions drawn from cell culture experiments are strengthened if they are congruent with conclusions drawn from animal experiments, and conclusions from animal experiments are strengthened if they are congruent across multiple models. In this regard, all of the commonly used animal models in cancer pharmacology (for example, cell line subcutaneous xenografts, patient-derived xenografts and genetically engineered mouse models) have inherent strengths and weaknesses. For example, xenograft experiments can be informative in cases of cell-intrinsic targets that are linked to cell survival, but misleading in cases of targets that rely on the host immune response. It is also important to note that subcutaneous xenograft models and orthotopic models can

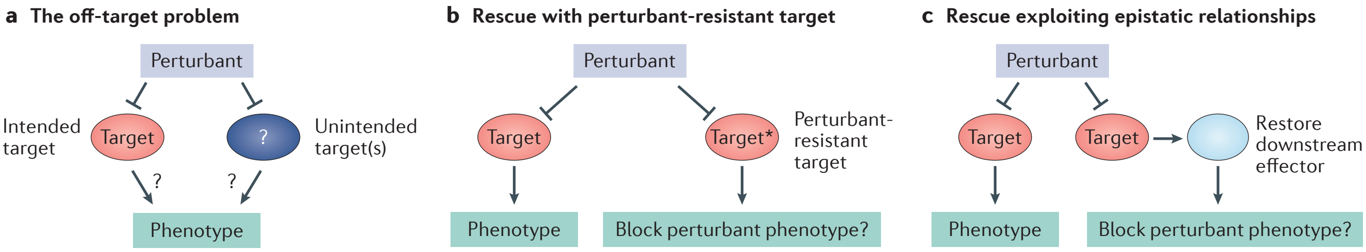
a The off-target problem
Perturbant
Intended target
Target
?
Unintended target(s)
?
?
Phenotype
b Rescue with perturbant-resistant target
Perturbant
Target
Target*
Perturbant-resistant target
Phenotype
Block perturbant phenotype?
c Rescue exploiting epistatic relationships
Perturbant
Target
Target
Restore downstream effector
Phenotype
Block perturbant phenotype?

Figure 6 | **Distinguishing on-target versus off-target effects with rescue experiments. a** | The challenge with a phenotype caused by a perturbant (for example, drug, siRNA, shRNA or single guide RNA (sgRNA)) is to determine whether that phenotype is on-target, off-target or both. This is especially important when the phenotype in question could simply reflect a loss of fitness stemming from a nonspecific toxic effect. **b** | The classic way to address this is to ask whether the perturbant still induces that phenotype in the face of a target that is resistant to that perturbant. **c** | An alternative approach, which requires prior knowledge of the target's function and downstream activities, is to ask whether the perturbant still induces the phenotype if the downstream effector function of the target is maintained.

sometimes lead to divergent conclusions, presumably owing to microenvironmental differences that affect interactions between tumour cells and the host[30]. For example, subcutaneous growth appears to place a premium on the ability of tumour cells to induce the formation of new blood vessels[31,32].

There has been a trend, especially in papers in high-profile journals, towards making far-reaching claims in an attempt to paint a seemingly complete picture that incorporates both new mechanistic insights and the physiological or clinical relevance of a given set of findings. The field would be better served if papers claimed less, but provided more lines of corroborating evidence in support of their conclusions. Describing a properly controlled and complementary set of target validation experiments can easily constitute an entire manuscript. It should not be an afterthought relegated to the last figure of a manuscript, in a gratuitous attempt to achieve *in vivo* or clinical relevance. Nor is it reasonable to require that an initial cancer target validation paper should also provide deep mechanistic insights into how that target alters cancer cell fitness and, ideally, does so in a cancer cell-selective manner. These questions are often best left to future studies and, in the case of most approved cancer drugs, still remain unanswered today.

Modelling therapy of established disease

A tumour as small as one cubic centimetre might already contain as many as 10^9 cells, and most patients with recurrent or metastatic disease harbour 10^{10}–10^{12} cancer cells. The challenge of cancer target discovery is ultimately to identify drugs that can halt tumour growth and, ideally, cause tumour regression, at clinically tolerable doses. Manipulating a target in cancer cells *ex vivo* before cell implantation can overestimate the importance of the target in tumour maintenance and the likelihood that inhibiting that target will cause tumour regression rather than simply slowing progression. For example, 10^6 tumour cells injected into an immunocompromised mouse must undergo at least ten doublings to achieve a size of 1 cm^3, at which point most tumour xenograft experiments have to be terminated. An intervention that causes only a very modest slowing of cell proliferation, such that nine doublings occur during the same time period, would score as causing a 50% reduction in tumour growth in this scenario. Similar considerations apply when a pharmacological intervention is applied at the time of tumour cell implantation or very soon thereafter. Ideally, interventions should be applied in models of established tumours including, whenever possible, models of metastatic disease because most cancer patients die of metastatic disease rather than organ-confined disease. However, as with any experiment, there are several caveats that must be considered when assessing the benefits and risks of inhibiting a particular target in such animal models. For example, a drug that causes only tumour stasis in an immunocompromised mouse might cause tumour regression in the presence of an intact immune system.

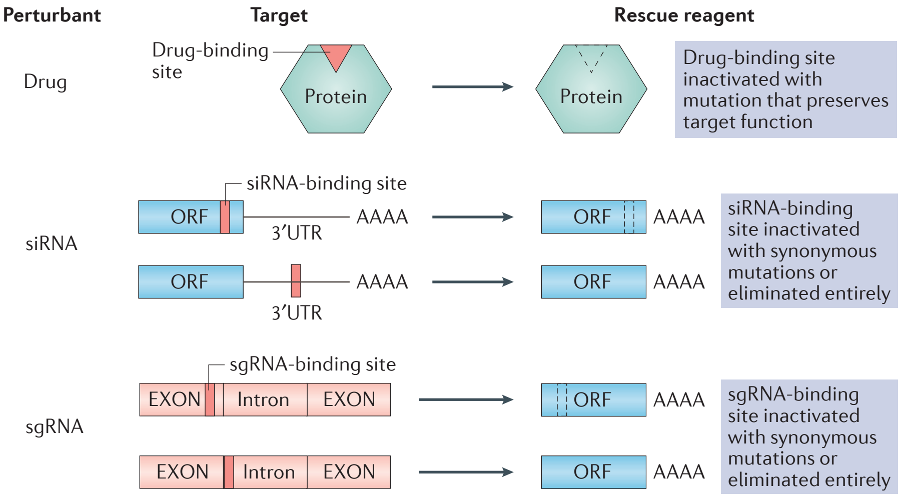
Perturbant
Target
Rescue reagent
Drug
Drug-binding site
Protein
Protein
Drug-binding site inactivated with mutation that preserves target function
siRNA
siRNA-binding site
ORF
3′UTR
AAAA
ORF
AAAA
ORF
3′UTR
AAAA
ORF
AAAA
siRNA-binding site inactivated with synonymous mutations or eliminated entirely
sgRNA
sgRNA-binding site
EXON
Intron
EXON
ORF
AAAA
EXON
Intron
EXON
ORF
AAAA
sgRNA-binding site inactivated with synonymous mutations or eliminated entirely

Figure 7 | **Designing perturbant-resistant targets for rescue experiments.** For a drug, a version of its target (usually a protein) is discovered or designed that retains its biochemical activity (such as its enzymatic activity in the case of an enzyme) but is drug resistant because of a mutation that decreases drug binding. For siRNAs and shRNAs an expression vector for the target is introduced that encodes an mRNA that lacks the siRNA- or shRNA-binding site entirely, as can be done for siRNA- or shRNA-binding sites located in untranslated regions (UTRs), or the siRNA- or shRNA-binding site is crippled by the introduction of three or more synonymous mutations. For single guide RNAs (sgRNAs) one introduces an expression vector for the target that encodes an mRNA that partially or completely lacks the sgRNA-binding site (including the protospacer adjacent motif (PAM) site), as can be done for sgRNA-binding sites located at intron–exon boundaries, or by crippling the sgRNA binding, such as by introducing a synonymous mutation into the PAM site. ORF, open reading frame.

Predicting toxicity

One of my colleagues is fond of saying: "If you are smart enough, you can always think of a reason why something will fail". Many new prospective cancer targets are immediately attacked, and are sometimes even dismissed altogether, by reviewers, the pharmaceutical industry or both, because of a priori arguments that inhibiting those targets will lead to unacceptable toxicity. These arguments are often based on the roles those targets have during embryological development, as revealed by germline knockout studies in mice, their known functions in adult animals, or both.

These naysayer arguments overlook several important considerations. First, the role of a target during development does not necessarily predict its requirement in adults. For example, germline knockout of the imatinib mesylate target *Abl* is embryonic lethal in mice (this is also true for some of the off-targets of imatinib mesylate)[33], but imatinib mesylate is well tolerated in adults. Although more sophisticated tools now exist to inactivate genes in adult tissues, these approaches typically lead to complete loss of the gene of interest in a mosaic pattern. By contrast, most drugs inactivate specific functions of their targets. Extrapolating the effects of genetic knockout of a gene to chemical inhibition of its protein product can therefore be misleading.

A second important difference between genetic and chemical perturbation of a target is that most successful drugs are successful precisely because they can be titrated to achieve a level of target inhibition, with respect to both depth and duration of inhibition, that produces a therapeutic outcome without causing toxicity. For example, the proteasome inhibitor bortezomib has anti-myeloma activity at a concentration that causes a 50–80% decrease in proteasome activity, but is toxic at higher concentrations that more completely block proteasome activity[34]. This last observation underscores the fact that the issue in cancer therapeutics is not whether a target is

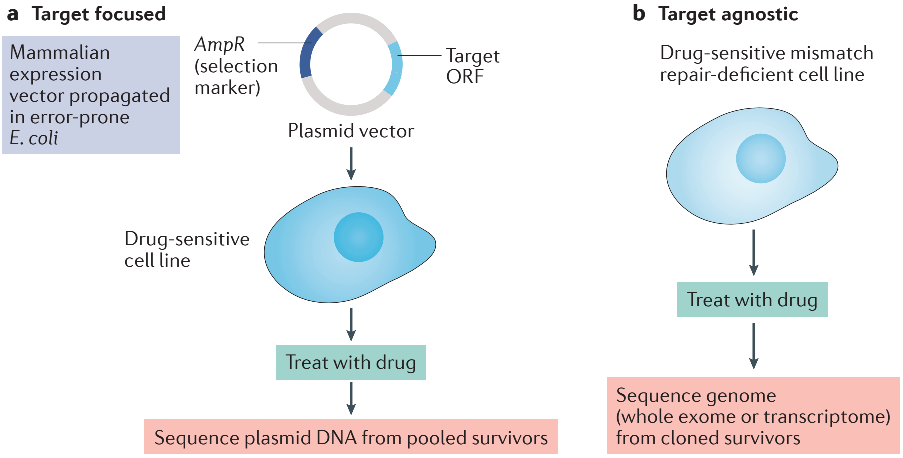
a Target focused
Mammalian expression vector propagated in error-prone *E. coli*
AmpR (selection marker)
Target ORF
Plasmid vector
Drug-sensitive cell line
Treat with drug
Sequence plasmid DNA from pooled survivors
b Target agnostic
Drug-sensitive mismatch repair-deficient cell line
Treat with drug
Sequence genome (whole exome or transcriptome) from cloned survivors

Figure 8 | **Generating drug-resistant protein targets. a** | A mammalian expression vector encoding the protein target of interest is first propagated in error-prone *Escherichia coli*, such as the strain XL1-red, and then introduced into a cell line that is sensitive to the drug. After drug exposure, drug-resistant clones are pooled, genomic DNA is isolated and the cDNA is amplified by PCR and sequenced using next-generation sequencing methods. Putative drug-resistance mutations are reintroduced into a wild-type target cDNA and tested for their ability to confer drug resistance to the target in biochemical assays and to confer drug resistance to cells expressing the mutant target. **b** | A mismatch repair-deficient cell line that is sensitive to the drug of interest is treated with that drug. Surviving cells are cloned, which are then analysed by RNA-seq and whole-exome sequencing for somatic mutations that recur among multiple independent clones. Performing multiple biological replicates or introducing a DNA bar code library before drug exposure can assist in identifying mutations that recur because of founder effects in the original population. *AmpR*, ampicillin resistance gene; ORF, open reading frame.

important or essential in normal cells, but whether the target is more important in cancer cells than it is in normal cells[35,36]. The degree to which there is a differential requirement for the target in cancer cells relative to normal cells is the biological determinant of the therapeutic window for inhibition of the target. Even in hindsight, it is not clear why most approved cancer drugs, including the above-mentioned imatinib mesylate and bortezomib, have therapeutic windows. This question is even more perplexing for many classic cytotoxic agents, including DNA-alkylating agents and microtubule poisons. Presumably, these therapeutic windows are driven by several factors, including cancer-specific mutations, lineage-specific epigenetic differences and microenvironmental factors, that quantitatively or qualitatively increase the requirements for specific targets in cancer cells compared with normal host cells[35,36].

A last point is that there are many examples of drugs that cause toxicities in animal models that are not observed in humans, and vice versa. For example, the development of imatinib mesylate was delayed because of a liver toxicity signal in an animal model that was not observed in humans. This underscores the fact that a number of variables including, for example, species-specific differences in drug metabolism, can influence the likelihood and degree to which a drug against a given target will cause toxicity.

Concluding remarks

Several sociological and technological changes related to the way we carry out cancer research has created a 'perfect storm', which is fuelling the reproducibility and robustness problems we increasingly face. Scientists, and particularly young trainees, feel increasing pressure from both funding agencies and high-profile journals, the latter driven by both editors and reviewers, to do work that is perceived to be clinically relevant. The availability of siRNA and shRNA technology means that target validation studies in human cells are now readily feasible, even for researchers who lack the ability to generate chemical perturbants and lack the biological insights necessary to, for example, design dominant-negative mutants or inhibitory antibodies. The tsunami of genomic data, including mutational and copy number alteration data, that is emerging from large-scale cancer sequencing projects means that there is suddenly no shortage of potential therapeutic targets to study. The widespread use of perturbants in phenotypic down assays that are being used as surrogates for therapeutic outcomes is leading to findings that are frequently confounded by off-target effects. The adoption of high-throughput screening technologies in academic laboratories is increasing the likelihood of false positives linked to multiple hypothesis testing. Finally, our system of publication and promotion creates a strong bias in favour of seemingly positive results. This makes it impossible to know how often findings occurred by chance, as the true denominator with respect to how often the hypothesis under study was interrogated and rejected by the scientific community is unknown, and also potentially incentivizes bad practices, such as consciously or unconsciously cherry-picking results.

The reproducibility and robustness problem in preclinical research has already led to new guidelines for conducting such research (for example, NIH Principles and Guidelines for Reporting Preclinical Research; see Further information) as well as initiatives such as The Reproducibility Project: Cancer Biology (see Further information)[4,5] to better understand the magnitude of the problem. Development of the cancer drugs of tomorrow will rely on the target identification and validation studies being carried out today. We in the academic community need to set a higher standard with respect to such studies, and we need to remind trainees, as well as ourselves, that publishing papers is a means to an end and not an end in itself. The first question that should be asked when assessing the importance of any paper is whether its findings are likely to withstand the test of time; not which journal it appeared in. Publishing findings that are false, misleading or simply not robust actually impedes progress and consumes precious resources. We should lower the bar with respect to the number of claims required for publication and raise the bar in terms of the burden of proof required to support those claims. Although raising our standards of rigor will be challenging, it is useful to remember that it is a virtual certainty, from a statistical point of view, that cancer will one day affect someone we love, such as a child, a grandchild or a spouse. Our patients today, as well as the patients of tomorrow, are desperately counting on us to do better and move the field forward.

William G. Kaelin Jr is at the Howard Hughes Medical Institute, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA.
william_kaelin@dfci.harvard.edu

doi:10.1038/nrc.2017.32
Published online 19 May 2017

1. Prinz, F., Schlange, T. & Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* **10**, 712 (2011).
2. Begley, C. G. & Ellis, L. M. Drug development: raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
3. Frye, S. V. *et al.* Tackling reproducibility in academic preclinical drug discovery. *Nat. Rev. Drug Discov.* **14**, 733–734 (2015).
4. Nosek, B. A. & Errington, T. M. Making sense of replications. *eLife* **6**, e23383 (2017).
5. Errington, T. M. *et al.* An open investigation of the reproducibility of cancer biology research. *eLife* **3**, e04333 (2014).
6. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363**, 809–819 (2010).
7. Harris, C. C. p53 tumor suppressor gene: from the basis research laboratory to the clinic — an abridged historical perspective. *Carcinogenesis* **17**, 1187–1198 (1996).
8. Semenza, G. L. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. *Oncogene* **29**, 625–634 (2010).
9. Mellor, H. R. & Harris, A. L. The role of the hypoxia-inducible BH3-only proteins BNIP3 and BNIP3L in cancer. *Cancer Metastasis Rev.* **26**, 553–566 (2007).
10. Reiling, J. H. & Hafen, E. The hypoxia-induced paralogs Scylla and Charybdis inhibit growth by down-regulating S6K activity upstream of TSC in *Drosophila. Genes Dev.* **18**, 2879–2892 (2004).
11. Brugarolas, J. *et al.* Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. *Genes Dev.* **18**, 2893–2904 (2004).
12. Keith, B., Johnson, R. S. & Simon, M. C. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat. Rev. Cancer* **12**, 9–22 (2012).
13. Lin, A., Giuliano, C. J., Sayles, N. M. & Sheltzer, J. M. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. *eLife* **6**, e24179 (2017).
14. Gorre, M. *et al.* Clinical resistance to STI-571 cancer therapy caused by BCR–ABL gene mutation or amplification. *Science* **293**, 876–880 (2001).
15. Azam, M., Latek, R. R. & Daley, G. Q. Mechanisms of autoinhibition and STI-571/imatinib resistance revealed by mutagenesis of BCR–ABL. *Cell* **112**, 831–843 (2003).
16. Burgess, M. R., Skaggs, B. J., Shah, N. P., Lee, F. Y. & Sawyers, C. L. Comparative analysis of two clinically active BCR–ABL kinase inhibitors reveals the role of conformation-specific binding in resistance. *Proc. Natl Acad. Sci. USA* **102**, 3395–3400 (2005).
17. Balbas, M. D. *et al.* Overcoming mutation-based resistance to antiandrogens with rational drug design. *eLife* **2**, e00499 (2013).
18. Wacker, S. A., Houghtaling, B. R., Elemento, O. & Kapoor, T. M. Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat. Chem. Biol.* **8**, 235–237 (2012).
19. Kasap, C., Elemento, O. & Kapoor, T. M. DrugTargetSeqR: a genomics- and CRISPR–Cas9-based method to analyze drug targets. *Nat. Chem. Biol.* **10**, 626–628 (2014).
20. Han, T. *et al.* The antitumor toxin CD437 is a direct inhibitor of DNA polymerase α. *Nat. Chem. Biol.* **12**, 511–515 (2016).
21. Forment, J. V. *et al.* Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells. *Nat. Chem. Biol.* **13**, 12–14 (2017).
22. Emery, C. M. *et al.* MEK1 mutations confer resistance to MEK and B-RAF inhibition. *Proc. Natl Acad. Sci. USA* **106**, 20411–20416 (2009).
23. Nazarian, R. *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973–977 (2010).
24. Losman, J. A. *et al.* (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* **339**, 1621–1625 (2013).
25. Kaelin, W. G. Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science* **337**, 421–422 (2012).
26. Kittler, R. *et al.* RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proc. Natl Acad. Sci. USA* **102**, 2396–2401 (2005).
27. Brignole, C. *et al.* Immune cell-mediated antitumor activities of GD2-targeted liposomal c-myb antisense oligonucleotides containing CpG motifs. *J. Natl Cancer Inst.* **96**, 1171–1180 (2004).
28. Munoz, D. M. *et al.* CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* **6**, 900–913 (2016).
29. Aguirre, A. J. *et al.* Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* **6**, 914–929 (2016).
30. Blouw, B. *et al.* The hypoxic response of tumors is dependent on their microenvironment. *Cancer Cell* **4**, 133–146 (2003).
31. Bridgeman, V. L. *et al.* Vessel co-option is common in human lung metastases and mediates resistance to anti-angiogenic therapy in preclinical lung metastasis models. *J. Pathol.* **241**, 362–374 (2017).
32. Donnem, T. *et al.* Vessel co-option in primary human tumors and metastases: an obstacle to effective anti-angiogenic treatment? *Cancer Med.* **2**, 427–436 (2013).
33. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the c-abl proto-oncogene. *Cell* **65**, 1153–1163 (1991).
34. Hamilton, A. L. *et al.* Proteasome inhibition with bortezomib (PS-341): a phase I study with pharmacodynamic end points using a day 1 and day 4 schedule in a 14-day cycle. *J. Clin. Oncol.* **23**, 6107–6116 (2005).
35. Kaelin, W. J. Choosing anticancer drug targets in the postgenomic era. *J. Clin. Invest.* **104**, 1503–1506 (1999).
36. Kaelin, W. G. Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nat. Rev. Cancer* **5**, 689–698 (2005).

Acknowledgements

The author thanks J. Losman, D. Nijhawan, W. Sellers and members of the Kaelin Laboratory for careful reading of this manuscript and useful suggestions. He dedicates this article to his late wife, Dr Carolyn M. Kaelin, who survived a breast cancer in 2003 but died of a glioblastoma in 2015.

Competing interests statement

The author declares no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

FURTHER INFORMATION

Principles and Guidelines for Reporting Preclinical Research: http://www.nih.gov/about/reporting-preclinical-research.htm

The Reproducibility Project: Cancer Biology: https://cos.io/our-services/research/rpcb-overview/

ALL LINKS ARE ACTIVE IN THE ONLINE PDF