nature biotechnology

Modeling credentials

The time has come to systematically assess the pros and cons of different mouse models as predictors of patient responses to cancer therapies.

There are numerous reasons why cancer therapies can stumble in clinical testing. Trials may be underpowered. Treatments may be tested in patients with advanced-stage disease who have failed all other treatment options. Heterogeneity of a cancer type may lead to heterogeneity of patient response. Trial endpoints (such as tumor response rates) may fail to predict overall survival. But all these reasons are moot if the experimental therapy itself lacks therapeutic efficacy. And all too often, mouse models of cancer predict therapeutic efficacy that goes missing in human beings. That is why the cancer community needs to devote greater effort to systematically identifying features of preclinical models that are important for successful predictions of clinical success against different tumor types.

Between 2003 and 2011, the likelihood that an oncology drug entering a phase 1 trial would ultimately be approved by the US Food and Drug Administration was a measly 7%—the lowest among all disease areas. The consultants Cutting Edge Information have estimated the cost of a phase 3 oncology trial as high as $100,000 per patient. The large number of costly failures in cancer highlights the special challenges of drug development in this area.

In most cases, the decision to commence clinical evaluation is based on positive experimental data from at least one, but in many cases several, mouse model(s). These models come in many different flavors, ranging from simple cancer cell line xenografts to highly engineered genetic mouse models. For each model, one can draw up a list of potential advantages or disadvantages. How easy is the model to work with? Does it model the genetic diversity observed in patients? Does it contain all relevant driver mutations? Does it model the complexities of the genetic background of the patient population and the mutational load and heterogeneity observed in human tumors? Does it model the tumor microenvironment? Is a functional immune system present? Is the model stable over time? Does it model metastasis? But which of these features matter most for predicting therapeutic success in the real world is not always obvious.

Of course, many fields of biology have had to face issues of validity and reproducibility of their methods and protocols. A good way of finding optimal assays is to test competing methods on the same gold-standard data set to make a fair comparison of the results possible. The question is, Why haven't oncology researchers made an effort to use similar approaches to validate the utility of preclinical cancer models?

In the 2000s, several studies retrospectively evaluated correlations between the activity of (mainly cytotoxic) compounds in preclinical xenograft models and phase 2 clinical trials. The results were mixed, highlighting the differences among various cancer types and the importance of studying drug response in multiple models. More recently, genetically engineered mouse models have been evaluated both in dedicated mouse trials and in so-called co-clinical trials in which mice and human patients are treated side by side. However, few studies have undertaken a real comparative effort to evaluate the merits of different modeling strategies.

Such studies are not without challenges. One major issue is that different types of treatment may demand different types of mouse models. With the anticancer armamentarium now including radio- and chemotherapies, DNA repair inhibitors, kinase inhibitors, histone deacetylase and demethylase inhibitors and immune-checkpoint inhibitors—not to mention adoptive T cell therapies and oncolytic viruses—it is clear that models must be fit for purpose: they must mimic the aspects of the human tumor that are important in the response to the therapy (for example, carry a kinase mutation targeted by a therapeutic or have the relevant humoral and cellular immune processes to interact with a checkpoint inhibitor). Another challenge is to make models recapitulate as closely as possible the human malignancy they represent—a stiff task, given the different selective pressures on tumors in animals and patients and the gross differences between rodent and primate biology.

One option would be to perform a retrospective meta-analysis of mouse data and human data from already published reports. But the numbers of mice in published studies often are too small, making it difficult to perform statistically robust comparisons; the experimental conditions (such as relative tumor size at the start of treatment or metastatic status) are not comparable to those in the patient population; and in many cases, the drug doses used and tolerated in mice are simply too toxic for use in people.

Thus, there is a need to design *ab initio* experimental mouse trials that compare the relative merits for efficacy prediction of different models. What cancers and what range of drugs should be tested? What are the most widely used models for the given cancer, and are they compatible with all the selected drugs? What doses of the drugs should be used in the mice? What endpoints should be studied (for example, progression free survival, overall survival, response rates or something else)? And should assays from *in vitro* models (such as two-dimensional cultures or organoids) be included for comparison?

Just last year, the US National Cancer Institute (NCI) recognized the validation of mouse models as a problem, calling for proposals for "projects devoted to ensuring that mice and mouse models used for translational research questions are used appropriately and that the models provide reliable information for patient benefit." Other funders and industry should consider either joining forces with the NCI or setting up similar funding lines.

There is no doubt that these systematic studies will be expensive with no guarantee that they will reveal any surprises in terms of which models are the best predictors of response. But until we carry out such studies, we will never truly know whether the community's favored models really are the best. And avoiding just one needless phase 3 oncology trial would likely pay for several mouse credential studies many times over. ⬡