

COMMENT

AVIAN INFLUENZA Shift expertise to track mutations where they emerge **p.534**

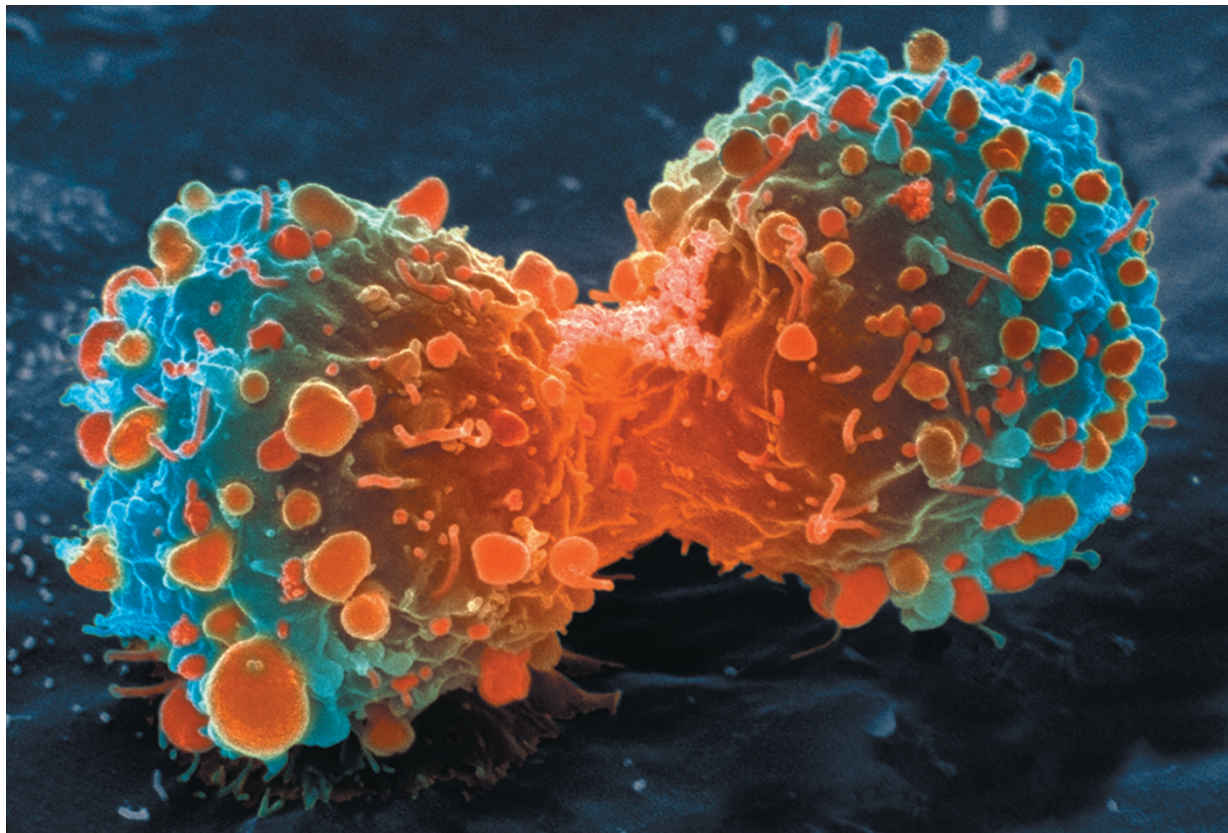
EARTH SYSTEMS Past climates give valuable clues to future warming **p.537**

HISTORY OF SCIENCE Descartes' lost letter tracked using Google **p.540**

OBITUARY Wylie Vale and an elusive stress hormone **p.542**



S. GSCHMEISSNER/SPL



Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low¹. Sadly, clinical

trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and

investigators must reassess their approach to translating discovery research into greater clinical success and impact.

Many factors are responsible for the high failure rate, notwithstanding the inherently difficult nature of this disease. Certainly, the limitations of preclinical tools such as inadequate cancer-cell-line and mouse models² make it difficult for even ▶

► the best scientists working in optimal conditions to make a discovery that will ultimately have an impact in the clinic. Issues related to clinical-trial design — such as uncontrolled phase II studies, a reliance on standard criteria for evaluating tumour response and the challenges of selecting patients prospectively — also play a significant part in the dismal success rate³.

Unquestionably, a significant contributor to failure in oncology trials is the quality of published preclinical data. Drug development relies heavily on the literature, especially with regards to new targets and biology. Moreover, clinical endpoints in cancer are defined mainly in terms of patient survival, rather than by the intermediate endpoints seen in other disciplines (for example, cholesterol levels for statins). Thus, it takes many years before the clinical applicability of initial preclinical observations is known. The results of preclinical studies must therefore be very robust to withstand the rigours and challenges of clinical trials, stemming from the heterogeneity of both tumours and patients.

CONFIRMING RESEARCH FINDINGS

The scientific community assumes that the claims in a preclinical study can be taken at face value — that although there might be some errors in detail, the main message of the paper can be relied on and the data will, for the most part, stand the test of time. Unfortunately, this is not always the case. Although the issue of irreproducible data has been discussed between scientists for decades, it has recently received greater attention (see go.nature.com/q7i2up) as the costs of drug development have increased along with the number of late-stage clinical-trial failures and the demand for more effective therapies.

Over the past decade, before pursuing a particular line of research, scientists (including C.G.B.) in the haematology and oncology department at the biotechnology firm Amgen in Thousand Oaks, California, tried to confirm published findings related to that work. Fifty-three papers were deemed 'landmark' studies (see 'Reproducibility of research findings'). It was acknowledged from the outset that some of the data might not hold up, because papers were deliberately selected that described something completely new, such as fresh approaches to targeting cancers or alternative clinical uses for existing therapeutics. Nevertheless, scientific findings were confirmed in only 6 (11%) cases. Even knowing the limitations of preclinical research, this was a shocking result.

Of course, the validation attempts may have failed because of technical differences or difficulties, despite efforts to ensure that this was not the case. Additional models were also used in the validation, because

to drive a drug-development programme it is essential that findings are sufficiently robust and applicable beyond the one narrow experimental model that may have been enough for publication. To address these concerns, when findings could not be reproduced, an attempt was made to contact

“The scientific process demands the highest standards of quality, ethics and rigour.”

the original authors, discuss the discrepant findings, exchange reagents and repeat experiments under the authors' direction, occasionally even in the laboratory of the original investigator.

These investigators were all competent, well-meaning scientists who truly wanted to make advances in cancer research.

In studies for which findings could be reproduced, authors had paid close attention to controls, reagents, investigator bias and describing the complete data set. For results that could not be reproduced, however, data were not routinely analysed by investigators blinded to the experimental versus control groups. Investigators frequently presented the results of one experiment, such as a single Western-blot analysis. They sometimes said they presented specific experiments that supported their underlying hypothesis, but that were not reflective of the entire data set. There are no guidelines that require all data sets to be reported in a paper; often, original data are removed during the peer review and publication process.

Unfortunately, Amgen's findings are consistent with those of others in industry. A team at Bayer HealthCare in Germany last year reported⁴ that only about 25% of published preclinical studies could be validated to the point at which projects could continue. Notably, published cancer research represented 70% of the studies analysed in that report, some of which might overlap with the 53 papers examined at Amgen.

Some non-reproducible preclinical papers had spawned an entire field, with hundreds of secondary publications that expanded on elements of the original observation, but did not actually seek to confirm or falsify its fundamental basis. More troubling, some of the research has triggered a series of clinical studies — suggesting that many patients had

subjected themselves to a trial of a regimen or agent that probably wouldn't work.

These results, although disturbing, do not mean that the entire system is flawed. There are many examples of outstanding research that has been rapidly and reliably translated into clinical benefit. In 2011, several new cancer drugs were approved, built on robust preclinical data. However, the inability of industry and clinical trials to validate results from the majority of publications on potential therapeutic targets suggests a general, systemic problem. On speaking with many investigators in academia and industry, we found widespread recognition of this issue.

IMPROVING THE PRECLINICAL ENVIRONMENT

How can the robustness of published preclinical cancer research be increased? Clearly there are fundamental problems in both academia and industry in the way such research is conducted and reported. Addressing these systemic issues will require tremendous commitment and a desire to change the prevalent culture. Perhaps the most crucial element for change is to acknowledge that the bar for reproducibility in performing and presenting preclinical studies must be raised.

An enduring challenge in cancer-drug development lies in the erroneous use and misinterpretation of preclinical data from cell lines and animal models. The limitations of preclinical cancer models have been widely reviewed and are largely acknowledged by the field. They include the use of small numbers of poorly characterized tumour cell lines that inadequately recapitulate human disease, an inability to capture the human tumour environment, a poor appreciation of pharmacokinetics and pharmacodynamics, and the use of problematic endpoints and testing strategies. In addition, preclinical testing rarely includes predictive biomarkers that, when advanced to clinical trials, will help to distinguish those patients who are likely to benefit from a drug.

Wide recognition of the limitations in preclinical cancer studies means that business as usual is no longer an option. Cancer researchers must be more rigorous in their approach to preclinical studies. Given the inherent difficulties of mimicking the human micro-environment in preclinical research, reviewers and editors should demand greater thoroughness.

REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

*Source of citations: Google Scholar, May 2011.

RECOMMENDATIONS

Improving the reliability of preclinical cancer studies

We recommend the following steps to change the culture of oncology research and improve the relevance of translational studies:

- There must be more opportunities to present negative data. It should be the expectation that negative preclinical data will be presented at conferences and in publications. Preclinical investigators should be required to report all findings, regardless of the outcome. To facilitate this, funding agencies, reviewers and journal editors must agree that negative data can be just as informative as positive data.
- Journal editors must play an active part in initiating a cultural change. There must be mechanisms to report negative data that are accessible through PubMed or other search engines. There should be links to journal articles in which investigators have reported alternative findings to those in an initial (sometimes considered landmark) publication. One suggestion is to include 'tags' that report whether the key findings of a seminal paper were confirmed.
- There should be transparent opportunities for trainees, technicians and colleagues to discuss and report troubling or unethical behaviours without fearing adverse consequences.

- Greater dialogue should be encouraged between physicians, scientists, patient advocates and patients. Scientists benefit from learning about clinical reality. Physicians need better knowledge of the challenges and limitations of preclinical studies. Both groups benefit from improved understanding of patients' concerns.
- Institutions and committees should give more credit for teaching and mentoring: relying solely on publications in top-tier journals as the benchmark for promotion or grant funding can be misleading, and does not recognize the valuable contributions of great mentors, educators and administrators.
- Funding organizations must recognize and embrace the need for new cancer-research tools and assist in their development, and in providing greater community access to those tools. Examples include support for establishing large cancer cell-line collections with easy investigator access (a simple, universal material-transfer agreement); capabilities for genetic characterization of newly derived tumour cell lines and xenografts; identification of patient selection biomarkers; and generation of more robust, predictive tumour models. **C.G.B. and L.M.E.**

lines may allow elucidation of markers of sensitivity or resistance. Journals and grant reviewers must allow for the presentation of imperfect stories, and recognize and reward reproducible results, so that scientists feel less pressure to tell an impossibly perfect story to advance their careers.

Although reviewers, editors and grant-committee members share some responsibility for flaws in the system, investigators must be accountable for the data they generate, analyse and submit. We in the field must remain focused on the purpose of cancer research: to improve the lives of patients. Success in our own careers should be a consequence of outstanding research that has an impact on patients.

The lack of rigour that currently exists around generation and analysis of preclinical data is reminiscent of the situation in clinical research about 50 years ago. The changes that have taken place in clinical-trials processes over that time indicate that changes in prevailing attitudes and philosophies can occur (see 'Improving the reliability of preclinical cancer studies').

Improving preclinical cancer research to the point at which it is reproducible and translatable to clinical-trial success will be an extraordinarily difficult challenge. However, it is important to remember that patients are at the centre of all these efforts. If we in the field forget this, it is easy to lose our sense of focus, transparency and urgency. Cancer researchers are funded by community taxes and by the hard work and philanthropic donations of advocates. More importantly, patients rely on us to embrace innovation, make advances and deliver new therapies that will improve their lives. Although hundreds of thousands of research papers are published annually, too few clinical successes have been produced given the public investment of significant financial resources. We need a system that will facilitate a transparent discovery process that frequently and consistently leads to significant patient benefit. ■ **SEE EDITORIAL P.509**

C. Glenn Begley is a consultant and former vice-president and global head of Hematology and Oncology Research at Amgen, Thousand Oaks, California 91359, USA. **Lee M. Ellis** is at the University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. e-mail: lellis@mdanderson.org

1. Hutchinson, L. & Kirk, R. *Nature Rev. Clin. Oncol.* **8**, 189–190 (2011).
2. Francia, G. & Kerbel, R. S. *Nature Biotechnol.* **28**, 561–562 (2010).
3. Rubin, E. H. & Gilliland, D. G. *Nature Rev. Clin. Oncol.* <http://dx.doi.org/10.1038/nrclinonc.2012.22> (2012).
4. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10**, 712 (2011).
5. Fanelli, D. *PLoS ONE* **5**, e10271 (2010).

As with clinical studies, preclinical investigators should be blinded to the control and treatment arms, and use only rigorously validated reagents. All experiments should include and show appropriate positive and negative controls. Critical experiments should be repeated, preferably by different investigators in the same lab, and the entire data set must be represented in the final publication. For example, showing data from tumour models in which a drug is inactive, and may not completely fit an original hypothesis, is just as important as showing models in which the hypothesis was confirmed.

Studies should not be published using a single cell line or model, but should include a number of well-characterized cancer cell lines that are representative of the intended patient population. Cancer researchers must commit to making the difficult, time-consuming and costly transition towards new research tools, as well as adopting more robust, predictive tumour models and improved validation strategies. Similarly, efforts to identify patient-selection biomarkers should be mandatory at the outset of drug development.

Ultimately, however, the responsibility

for design, analysis and presentation of data rests with investigators, the laboratory and the host institution. All are accountable for poor experimental design, a lack of robust supportive data or selective data presentation. The scientific process demands the highest standards of quality, ethics and rigour.

BUILDING A STRONGER SYSTEM

What reasons underlie the publication of erroneous, selective or irreproducible data? The academic system and peer-review process tolerates and perhaps even inadvertently encourages such conduct⁵. To obtain funding, a job, promotion or tenure, researchers need a strong publication record, often including a first-authored high-impact publication. Journal editors, reviewers and grant-review committees often look for a scientific finding that is simple, clear and complete — a 'perfect' story. It is therefore tempting for investigators to submit selected data sets for publication, or even to massage data to fit the underlying hypothesis.

But there are no perfect stories in biology. In fact, gaps in stories can provide opportunities for further research — for example, a treatment that may work in only some cell