

COMMENT

NATURAL HISTORY Derring-do and drawer-delving at Chicago's Field Museum **p.618**

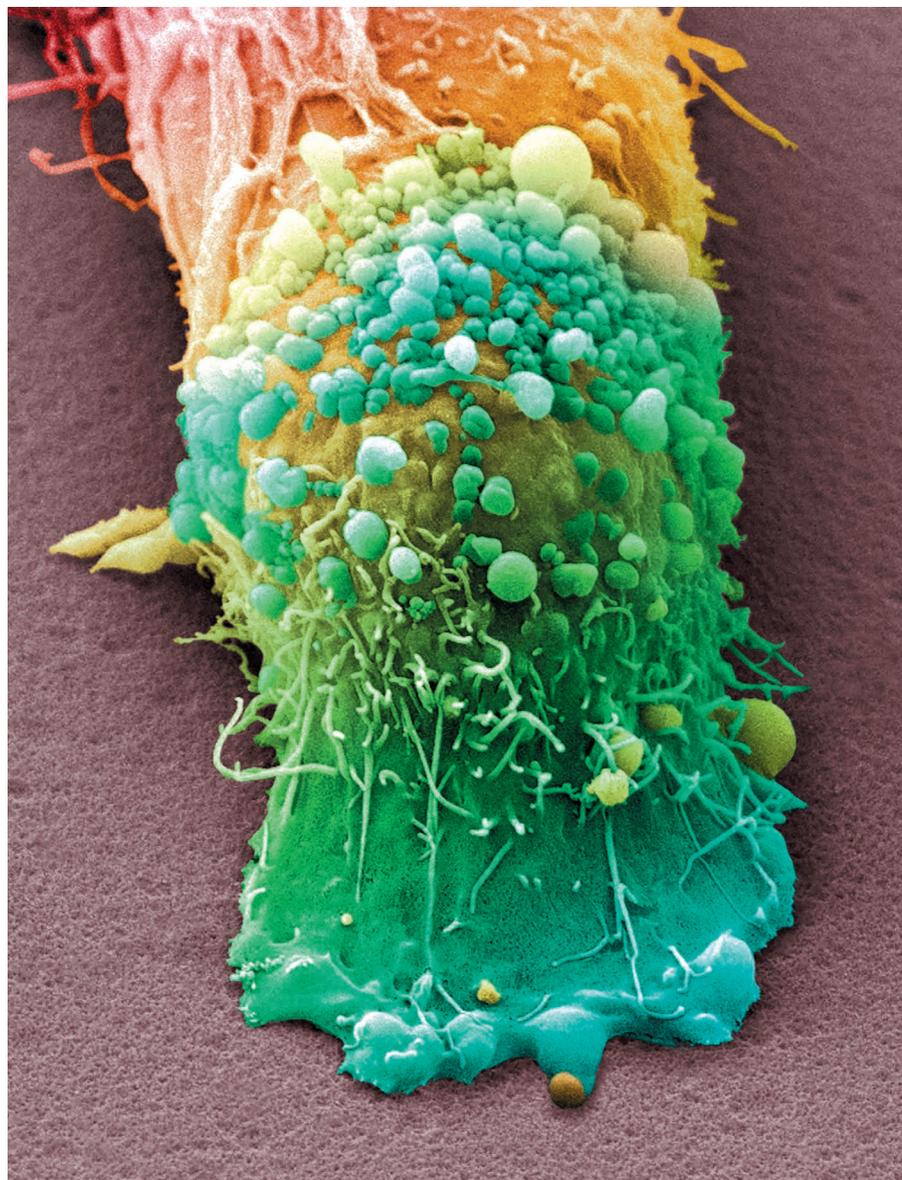
EVOLUTION Did pointing and waving kickstart language? **p.620**



ELITES Get out of the ivory tower and take a road trip **p.623**

OBITUARY Kenneth J. Arrow, Nobel-winning welfare economist, remembered **p.624**

STEVE GSCHMEISSNER/SPL



A cell from a melanoma shows the blebs and microvilli characteristic of a cancer cell.

Remember why we work on cancer

Levi Garraway reflects on the three things that keep his compass true when the going gets tough.

I first realized I'd been bitten by the science bug in the summer of 1987. I was walking home from the laboratory, mulling over an organic chemistry reaction that I had been attempting — and mostly failing — to execute. Suddenly, a notion coalesced in my 19-year-old brain: all human biology and disease must ultimately come down to reactions that either proceed properly or go awry. As I savoured the evening breeze, I knew that I wanted to dedicate my career to understanding these mechanisms and thereby to hasten new treatments.

Nearly every scientist remembers moments like these. I am saddened, therefore, by the cynical view that has become increasingly common in both academia and industry: that much biomedical science, even — or perhaps especially — that which appears in 'high-profile' journals, is bogus.

I am one of many scientists who have seen their past research subjected to unexpected scrutiny as a result. An attempt to replicate work from my team was among the first described by the Reproducibility Project: Cancer Biology, an initiative that independently repeated experiments from high-impact papers. In this case, as an editorial that surveyed the first replications explained, differences between how control cells behaved in the two sets of experiments made comparisons uninformative¹. The replicators' carefully conducted experiment showed just how tough it can be to reproduce results².

This experience — along with ongoing discussions — has set me reflecting on how we scientists can be confident of our work. Has the larger scientific enterprise overestimated its collective ability to elicit how nature works? What does it mean when one scientist is unable to reproduce the work of another?

Medical residents have a saying: "If you've never had a complication, you've not done enough medical procedures." By analogy, I reckon that if you have seldom wrestled with whether your results or conclusions are correct, you may not be as good a scientist as you think.

A path out of the headache and heartache becomes clearer when one considers the three 'Rs' that comprise high-quality scientific research — rigour, reproducibility ▶

► and robustness. These remind us of the reason why we became scientists in the first place.

RIGOUR

The need for an appropriate level of circumspection when interpreting scientific results seems axiomatic. However, it can prove surprisingly challenging to apply. Sometimes, lack of rigour stems from inadequate training. At least as often, such lapses stem from more elemental vagaries.

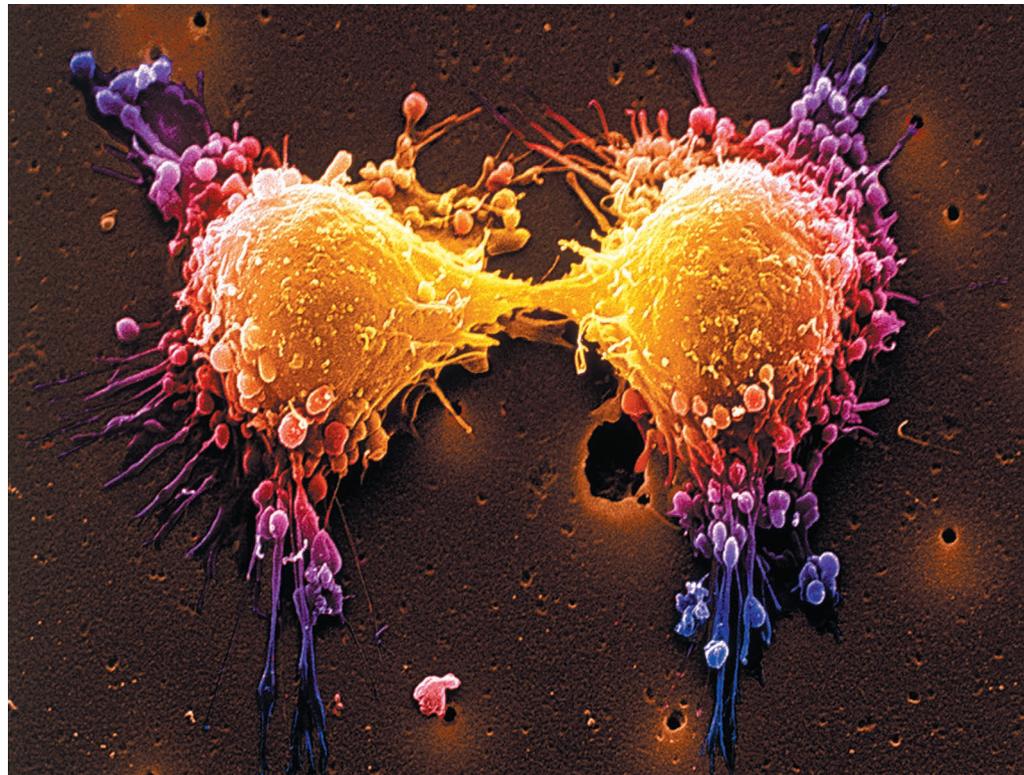
A harrowing experience in my group at Harvard Medical School in Boston, Massachusetts, several years ago, drove this point home. We were using RNA interference ‘knockdown’ screens to discover genes that were essential to cancer cells only if those cells already carried another ‘driver’ mutation (for example, a mutation essential for malignant properties). One big winner emerged, which I will call ‘MFG’ (My Favourite Gene). We looked for the same association in an independent RNAi data set and the finding seemed to hold. Next we performed what we considered very clever experiments to work out the mechanism.

Our finding seemed particularly exciting because MFG encoded a protein that had the potential to be uniquely ‘druggable’. In other words, we felt confident we could design an inhibitor of the protein that would have few undesirable side effects. We received a grant to look for small-molecule inhibitors and study further. This was what I had dreamed of since I was 19! Needless to say, we submitted our work to a high-profile journal.

While awaiting the reviewer critiques, a brilliant computational biologist in my group asked to meet with me urgently. He had taken an early look at new RNAi screening data that included many additional cancer cell lines, and saw no differences in cell behaviour when MFG was silenced in mutant cells compared with other cell lines. Put another way, our exciting finding could not be validated in this new, much larger, data set. I was dumbfounded. How could this be, when we had performed our own validation analysis and even amassed much mechanistic data?

One of my mentors, fellow cancer researcher William Kaelin, often says: “The most dangerous result is the one you were hoping for.” Such emotionally captivating results can cause a researcher to overlook more trivial explanations. In our case, careful reassessment found that although the result of each individual experiment seemed valid, our aggregate interpretation was not. Confounding factors included a

“Believing in exciting interpretations too early can cause a cascade of miscues.”



Unregulated cell division is a prerequisite for cancer to develop.

potential discovery bias from small numbers of mutant cell lines alongside an outsized inhibitory effect by certain RNAi constructs in our follow-up experiments.

I immediately withdrew the manuscript. I also contacted the agency that had awarded our grant, described the situation and was given permission to pivot to a different project. I still lie awake at night on occasion wondering what might have transpired if we had not discovered the MFG flaws until after publication. Would I have accepted the new analysis so quickly? I certainly hope so.

Believing in exciting interpretations too early can cause a cascade of miscues: too-favourable views of modest effects; a failure to include key controls; and a tendency to prioritize experimental models that yield positive results while downplaying others that do not. The essence of rigour is maintaining scepticism when reviewing experimental results. It always pays dividends.

REPRODUCIBILITY

In medicine, the quest to distinguish between distinct conditions with similar symptoms is called ‘differential diagnosis’. By analogy, three questions can guide a diagnosis of apparent irreproducibility: (1) Was the replication attempt backed by the requisite expertise? (2) Does a systematic comparison reveal a basis for discordant results? (3) Could the original result be wrong?

To illustrate the first question, consider the following example: extensive evidence exists that, in his prime, Tiger Woods could consistently hit a golf ball more than 260 metres straight down the fairway using his driver (the largest club in the bag). I too play golf; I have achieved credible results with my own driver (well, some of the time) and I am roughly the same height and weight as Tiger. I would very much like to be able to reproduce his results with my own hands.

Over the years, the golf industry has made untold sums of money from many golfers (myself included), who aspire to reproduce Tiger Woods’ results. However, only a small percentage of us can pull this off (and I am not one of them). Does this mean that Tiger Woods’ results were ‘wrong’, or that the remarkable physics he seemingly exemplified does not truly stand up to independent scrutiny? Of course not; it means that reproducing such results consistently requires a level of mastery that the typical golfer does not possess.

I do not believe that experimental skill is as elusive as that necessary to win the green jacket. I just want to underscore that we cannot assume that any given scientist — even a very good one — is properly equipped to reproduce an experiment if it involves new reagents, systems or biological context. Before even attempting to reproduce an ‘index’ experiment, a lab that lacks the needed experience often sends a trainee to another lab to work with a scientist who

regularly performs the experiment. This is so, even if both labs are recognized experts.

If the question of reproducibility persists once the requisite expertise is established, the answers often reside in subtle differences in methods, cell-line properties or reagents that become apparent only after scrutiny. Two acclaimed cancer researchers required more than a year to harmonize techniques to get similar measures for experiments; success depended on cross-country visits and on reconciling minor differences in how cells were prepared³.

If these steps do not work, we must consider whether the original result was really correct. And we must be prepared to accept the brutal facts, make the required corrections and move on. Great scientists are always willing to embrace the truth with humility and grace — even when it hurts.

ROBUSTNESS

Sometimes we get hung up on particularities of reagents and experiments, but what really matters is what we can credibly infer about how biology works. Moving from experiment to scientific finding requires the third ‘R’ of good science: robustness.

Robust findings become established over time as multiple lines of evidence emerge. Achieving robustness takes rigour and reproducibility, plus patience and judicious attention to the big picture.

As with the other Rs, my own impressions about robustness have been forged through some testing experiences. One involved a large collaborative project that I led

for nearly a decade — the Cancer Cell Line Encyclopedia. Reported inconsistencies with a related project, called the Genomics of Drug Sensitivity in Cancer, prompted intense, clarifying discussions on precisely how measurements are taken and on what sorts of data give the best insights for developing therapies^{4–15}.

A satisfying example of robustness occurred when our team published an unexpected finding, one that points to a vulnerability in certain malignancies: cancer cells harbouring a frequent genetic deletion become more reliant on a protein whose activity is regulated by the very metabolite that becomes elevated when the deletion is present. Two other large groups published similar findings at around the same time, each using distinct approaches^{16–18}. Diverse data that converge on the same observation in aggregate provide robustness, even though any single approach or model system has limitations.

WELL-LIT PATH

Recalling the teenage ‘aha moment’ that kindled my own investigative career decades ago still makes me smile. I did not pursue cancer research because it would offer copious opportunities to coax specific experiments to work reproducibly (such as that vexing chemical reaction). I chose this path because I believe that there are answers out there for

“What really matters is what we can credibly infer about how biology works.”

human disease, and that science holds the key to their discovery and application.

We scientists search tenaciously for information about how nature works through reason and experimentation. Who can deny the magnitude of knowledge we have gleaned, its acceleration over time, and its expanding positive impact on society? Of course, some data and models are fragile, and our understanding remains punctuated by false premises. Holding fast to the three Rs ensures that the path — although tortuous and treacherous at times — remains well lit. ■

Levi Garraway is senior vice-president of Global Oncology at Eli Lilly in Indianapolis, Indiana, USA.

e-mail: garraway_levi@lilly.com

1. *eLife* **6**, e23693 (2017).
2. Horrigan, S. K., Courville, P., Sampey, D., Zhou, F. & Cai, S. *eLife* **6**, e21634 (2017).
3. Hines, W. C. *et al. Cell Rep.* **6**, 779–781 (2014).
4. Safikhani, Z. *et al. F1000Research* **5**, 825 (2016).
5. Havery, P. M. *et al. Nature* **533**, 333–337 (2016).
6. Barretina, J. *et al. Nature* **483**, 603–607 (2012).
7. Garnett, M. J. *et al. Nature* **483**, 570–575 (2012).
8. Haibe-Kains, B. *et al. Nature* **504**, 389–393 (2013).
9. Cancer Cell Line Encyclopedia Consortium, Genomics of Drug Sensitivity in Cancer Consortium. *Nature* **528**, 84–87 (2015).
10. Geleher, P. *et al. Nature* **540**, E1–E2 (2016).
11. Safikhani, Z. *et al. Nature* **540**, E2–E4 (2016).
12. Mpindi, J. P. *et al. Nature* **540**, E5–E6 (2016).
13. Safikhani, Z. *et al. Nature* **540**, E6–E8 (2016).
14. Mehdi Bouhaddou, M. *et al. Nature* **540**, E9–E10 (2016).
15. Safikhani, Z. *et al. Nature* **540**, E11–E12 (2016).
16. Kryukov, G. V. *et al. Science* **351**, 1214–1218 (2016).
17. Mavrakis, K. J. *et al. Science* **351**, 1208–1213 (2016).
18. Marion, K. *et al. Cell Rep.* **15**, 574–587 (2016).

Five ways consortia can catalyse open science

An analysis of more than 50 collaborations shows the secrets of success, write **Joel Cutcher-Gershenfeld** and colleagues from the Stakeholder Alignment Collaborative.

“I am going to my grave with my disk drive in my cold dead hands.” So a senior scientist told a junior researcher, who related the tale at a 2013 US National Science Foundation (NSF) workshop on the reuse of physical samples in the geosciences. Sharing — of data sets, metadata, models, software and other resources — promises to speed discoveries, improve reproducibility and expand economic development. But it requires people to change.

Overcoming personal reluctance is doubly difficult because many aspects of the scientific enterprise undermine sharing.

Right now, most departments, funders and journals presume that data are proprietary from collection to publication. Even when individual scientists and institutional leaders want to do things differently, they face reviewers, colleagues and competitors clinging to conventional models.

As philosopher of science Thomas Kuhn documented more than 50 years ago, the scientific community resists challenges to its orthodoxy. The open sharing of data and other resources is a prime example¹. Conservatism reigns because, as the ‘iron law of oligarchy’ devised by sociologist

Robert Michels in the 1910s predicts, institutions established to achieve a certain goal will prioritize continued existence over the stated objective². And we’ve all experienced ‘path dependency’: that practices are hard to change once established³. For example, despite widespread support, few academic departments have restructured tenure processes to accommodate interdisciplinary work.

Over the past four years, we have studied more than a dozen scientific consortia involved in data sharing, and we’ve mapped the landscape of these ▶