

COMMENT

ANTHROPOLOGY Call for unity in the science of human beings **p.166**



GENETICS Reviewed: two primers on personal genomics **p.169**

POLICY Sanitation, not vaccination, is most important in Haiti **p.175**

OBITUARY Jack Oliver, key player in proof of plate tectonics, remembered **p.176**

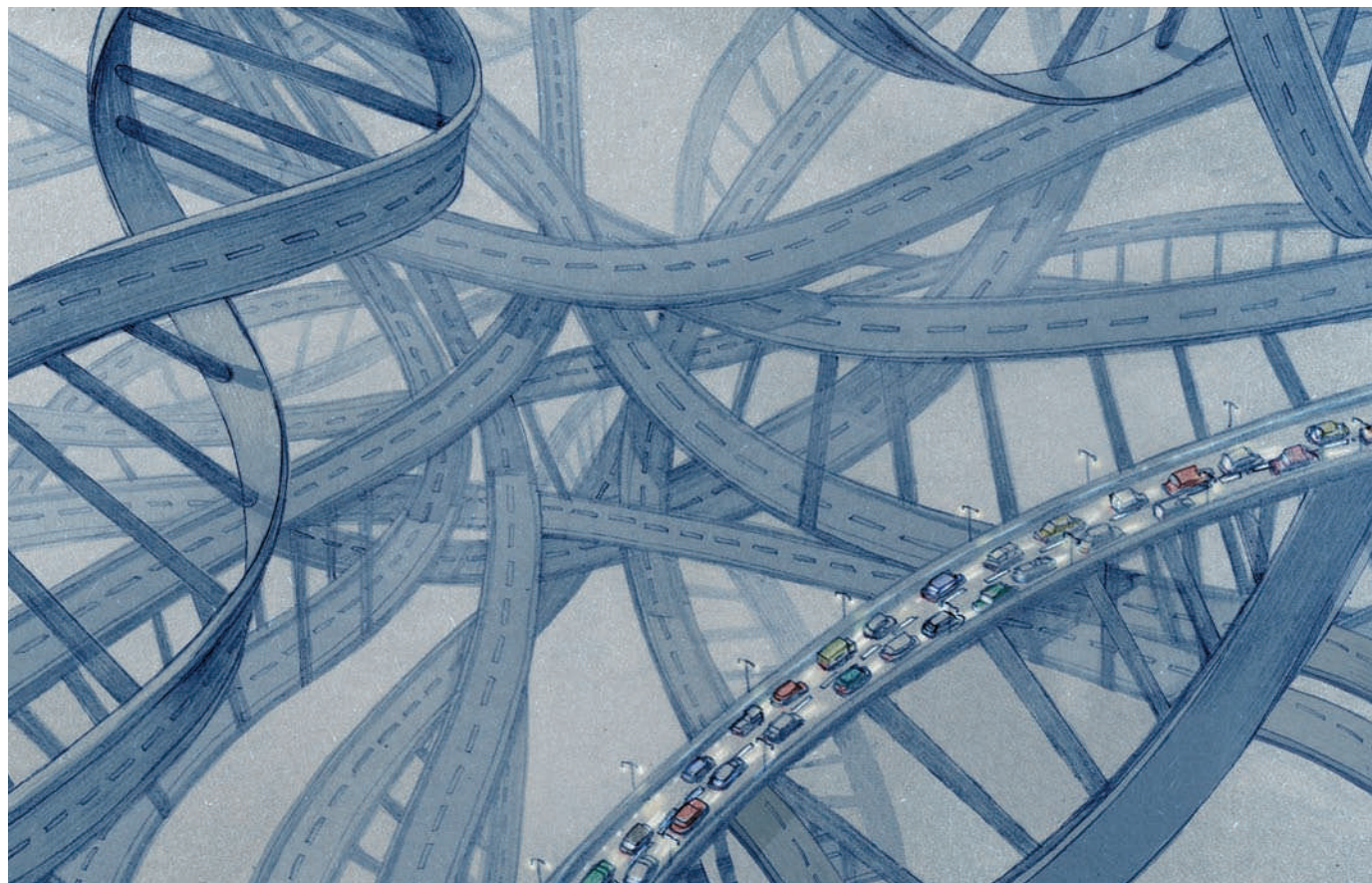


ILLUSTRATION BY JONATHAN BURTON

Too many roads not taken

Most protein research focuses on those known before the human genome was mapped. Work on the slew discovered since, urge **Aled M. Edwards** and his colleagues.

When a draft of the human genome was announced in 2000, funders, governments, industry and researchers made grand promises about how genome-based discoveries would revolutionize science. They promised that it would transform our understanding of human biology and disease, and provide new targets for drug discovery. Yet more than 75% of protein research still focuses on the 10% of proteins that were known before the genome was mapped — even though many more have been genetically linked to disease.

We performed a bibliometric analysis to assess how research activity has altered over time for three protein families that are central in disease and drug discovery: kinases, ion channels and nuclear receptors. For all three, we found very little change in the pattern of research activity — which proteins are associated with the highest number of publications — over the past 20 years¹.

NATURE.COM
Protein mapping gains a human focus:
go.nature.com/vbqctf

Even those proteins that have been directly associated with disease

remain 'hidden in plain sight', with scientists proving very reluctant to study them.

Where there has been a shift in research activity, it was often spurred by the emergence of tools to study a particular protein, not by a change in the protein's perceived importance. We believe that ensuring high-quality tools are developed for all the proteins discovered may be all that is needed to drive research into the unstudied parts of the human genome — even within funding and peer-review systems that are inherently conservative.

We searched for mention of every human ▶

► kinase, ion channel and nuclear receptor in either the title, abstract, keywords or 'MeSH' terms (used for indexing articles in Medline and PubMed) in the almost 20 million papers published between 1950 and 2009. We discovered that for all three classes of protein, the same small fraction of family members have remained 'the favourites' for nearly 20 years (see 'Fondling our problems').

For instance, the human genome encodes more than 500 protein kinases, of which hundreds have been shown to have genetic links with human diseases. Yet around 65% of the 20,000 kinase papers published in 2009 focused on the 50 proteins that were the 'hottest' in the early 1990s. Similarly, 75% of the research activity on nuclear hormone receptors in 2009 focused on the 6 receptors — out of the 48 encoded in the genome — that were most studied in the mid 1990s (ref. 1).

BIASED APPROACH

Although academics may be surprised by the magnitude of this research bias, they generally acknowledge its existence. It was first identified in kinase research² in 2008 and last year its effects were demonstrated in kinase drug discovery³. But a common assumption is that previous research efforts have preferentially identified the most important proteins. The evidence doesn't support this.

Patterns of gene expression and links between DNA sequences and breast cancer suggest, for instance, that 11 protein kinases are key nodes in the signalling pathways underlying the disease. Yet in the 2009 literature, one of these kinases, CDC2, received more attention than seven of the others combined, and three received just one mention. Likewise, various genetic approaches, including genome-wide association studies, have directly linked 37 of the 48 human nuclear receptor genes to disease. Among these, more than half of the total research

activity in 2009 was focused on just three. These three were also 'top of the nuclear receptor charts' in the 1990s.

Why the reluctance to work on the unknown? As the Nobel-prizewinning biochemist Roger Kornberg put it, scientists are wont to "fondle their problems": they have a natural tendency to dig deeper into their areas of expertise. Plus, funding and peer-review systems are risk-averse; funders and reviewers alike are less willing to support research on unstudied proteins, for which it is often harder to explain the rationale and significance. Moreover, the time frames associated with academic promotion and training encourage researchers to focus on systems that are likely to generate results rapidly, and for which research infrastructure and methods are already available.

Some funders are developing strategies to address the conservative nature of peer review. The Wellcome Trust, the largest non-governmental funder of biomedical research in the United Kingdom, for instance, is withdrawing its project grants in favour of providing longer-term support to outstanding investigators. And many universities are examining the pitfalls of their current reward systems. Unfortunately, institutional systems are ponderously slow to change. So what else can be done?

To establish a protein's function, and especially the details of how it works and its suitability for drug discovery, molecular biologists draw on an arsenal of tools. For instance, antibodies can help them identify where in the body the protein is being expressed; chemical inhibitors can be used to block a protein's activity in human cells and in animal models. These antibodies and small molecules also provide a launch pad for the development of new medicines by the biotechnology and pharmaceutical industries. Yet because of the cost and time

required to generate and characterize such tools, they are currently available for only a handful of well-studied proteins.

Our analysis of publication patterns for the human nuclear hormone receptor family suggests that making such tools readily available for all proteins could dramatically shift the balance in biomedical research.

WAKE-UP CALL

Nuclear receptors are transcription factors that bind small signalling molecules, such as steroids and hormones. Genetic data now suggest that all the receptors are directly or indirectly linked to human disease. About 30 of the family members were discovered at around the same time in the 1990s, allowing us to compare publication trends for numerous related proteins over time. We know exactly when the receptors were cloned, when genetic links with diseases were established and when research tools (in this case, chemical probes)^{4,5} became available.

When the 'novel' nuclear receptors were identified in the 1990s, all the family members were thought to have therapeutic potential.

"Making protein-based research tools readily available must be a major objective in the decade to come."

Interest developed most rapidly in those that were found to have genetic links to disease⁶⁻⁸ or that had interesting knockout phenotypes⁹, such as infertility. However, over the next 15 years, research activity refocused on a subset of 8 of these receptors.

From a genomics point of view, these 8 are no more interesting than any of the other 29 with known links to disease.

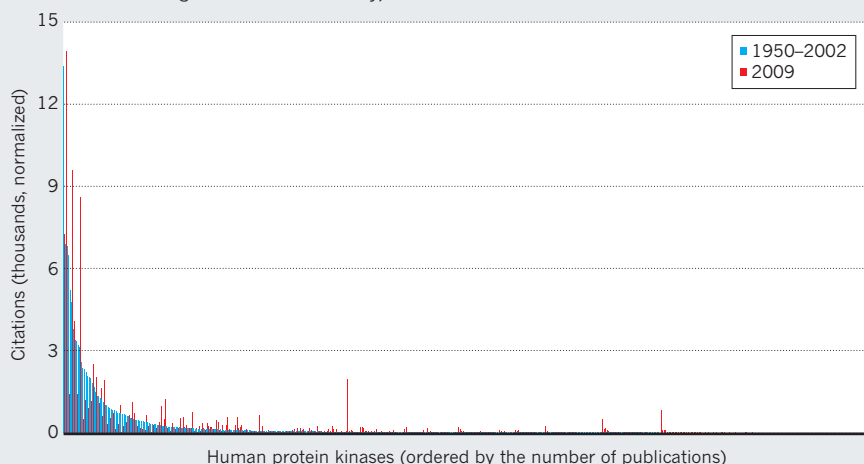
To our knowledge, the only connection among these 8 receptors is that for each there is a widely available, high-quality chemical probe that either enhances the receptor's activity or dampens it. In short, where high-quality tools are available (often commercially), there is research activity; where there are no tools, there is none (see 'Tools are telling').

Several other observations are consistent with the ideas that the availability of chemical probes for a given receptor dictates the level of research interest in it, and that the development of these tools is not driven by the importance of the protein. For instance, large and sustained increases in the rate of publications mentioning a nuclear receptor usually followed, not preceded, the release of a chemical probe.

Our findings should serve as a wake-up call to the biomedical and pharmaceutical research communities. Granting systems must be more daring, institutions must foster and reward risk, and the entire biomedical community must play down the legacy of the literature and let new evidence guide

FONDLING OUR PROBLEMS

Researchers' 'favourite kinases' have remained the same for decades with a few exceptions (kinases linked to diseases of great interest to industry).



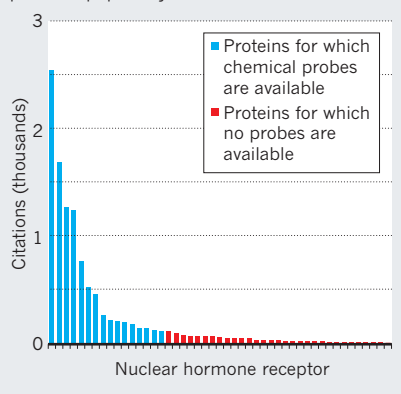
research. Genome-wide tools such as the DNA microarrays used in association studies have allowed geneticists to ignore preconceived ideas about disease mechanisms and pursue a remarkably successful broad-brush approach; this approach should be embraced more generally.

Our data also indicate that high-quality, readily available research tools can dramatically facilitate exploratory biomedical research. Funders such as the Wellcome Trust and the US National Institutes of Health have allocated some funding to tool-generating projects, but perhaps not enough. Part of the problem is that, unlike the high-energy physics community, which endorses the creation of large resources, the biomedical community often views projects focused on tool creation with some disdain, for lacking the elegance of 'real' science.

The budgets required also incite a visceral reaction. For example, the level of funding needed to develop even one chemical probe is enormous. Although it is only a fraction of the US\$100 billion spent on biomedical research each year — about several million dollars — it is huge compared with the amount customarily allocated to an individual scientist. Finally, the risk is significant. Large-scale efforts are not guaranteed to succeed; they require expertise in science

TOOLS ARE TELLING

The availability of research tools influences a protein's popularity.



and management, as well as collaboration between disciplines, between public and private sectors and — to avoid duplication of effort — even between countries.

Much of the work that has emerged from exploring the human genome over the past ten years lies fallow. Challenges notwithstanding, making protein-based research tools readily available must be a major objective in the decade to come. ■

Aled M. Edwards is *Banbury professor at the*

University of Toronto, Toronto, Ontario M5G 1L7, Canada. Ruth Isserlin and Gary D. Bader are at the Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada. Stephen V. Frye is at the Center for Integrative Chemical Biology and Drug Discovery, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. Timothy M. Willson is at GlaxoSmithKline, Research Triangle Park, North Carolina 27709, USA. Frank H. Yu is in the Program in Neurobiology and Dental Research Institute, School of Dentistry, Seoul National University, Seoul, 110-749, South Korea.

e-mail: aled.edwards@utoronto.ca

The authors declare competing financial interests: details accompany this article online at go.nature.com/kgw3oz.

1. Isserlin, R. *et al.* Preprint at <http://arxiv.org/abs/1102.0448> (2011).
2. Grueneberg, D. A. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 16472–16477 (2008).
3. Fedorov, O., Müller, S. & Knapp, S. *Nature Chem. Biol.* **6**, 166–169 (2010).
4. Frye, S. V. *Nature Chem. Biol.* **6**, 159–161 (2010).
5. Willson, T. M. & Moore, J. T. *Mol. Endocrinol.* **16**, 1135–1144 (2002).
6. Muscatelli, F. *et al.* *Nature* **372**, 672–676 (1994).
7. Yamagata, K. *et al.* *Nature* **384**, 458–460 (1996).
8. Achermann, J. C., Ito, M., Ito, M., Hindmarsh, P. C. & Jameson, J. L. *Nature Genet.* **22**, 125–126 (1999).
9. Lee, S. L. *et al.* *Science*. **273**, 1219–1221 (1996).