

# Knowledge about animal suffering is too rarely used

Researchers in different fields should pool information to minimize unnecessary harm.

*Sir* — We at the Humane Society of the United States strongly agree with your Opinion article “Rights, wrongs and ignorance” (*Nature* **416**, 351; 2002), that the science of animal suffering and cognition should be given higher priority. The public is often told that the pain, distress and suffering of research animals are already being minimized. But given the lack of scientific knowledge about these issues, much more can and should be done to develop knowledge to identify, assess, avoid and/or alleviate such suffering.

Our current knowledge of animal suffering and cognition is relatively isolated, and the principles too rarely applied in animal research. Scientists engaged in research on farm-animal welfare, for example, should be consulted on how to investigate animal suffering in laboratories. Widely used techniques, such as the use of carbon dioxide for rodent euthanasia and the creation of genetically modified animals, should be reviewed and discussed, and a consensus should be reached on the level of suffering caused.

We challenge all stakeholders, including the research community, supervising agencies and the animal-protection community to tackle these issues together in a coordinated way (see [www.hsus.org/ace/11350](http://www.hsus.org/ace/11350)), to minimize animal suffering and to reduce the uneasiness that we all experience (biomedical scientists included) when animals are harmed in the cause of scientific progress.

In your earlier Opinion article “Time to cut regulations that protect only regulators” (*Nature* **414**, 379; 2001), you rightly state that the primary purpose of the Institutional Animal Care and Use Committees (IACUC) system should be the protection of laboratory animals from pain, distress and harm. We are no more in favour of unnecessary paperwork than the research community. We have proposed a revision of the pain and distress classification system that would focus attention on the riskier projects and could help avoid unnecessary paperwork for projects that carry no risk of pain and distress — but the research community has protested

vociferously against our suggestions.

You are, however, incorrect to state that IACUC reviews used to be carried out after a grant application had passed an initial approval stage. This has never been the case, although the system may indeed change soon, in line with the recent changes by the Institutional Review Boards (IRBs) for human subjects to the ‘just in time’ system you outline, to cut down unnecessary paperwork. You are also incorrect to state that cross-talk between institutions is not permitted: various publications advise on procedures in this regard, though we agree that far too little is done.

Any fear among institutions is being driven by the IRB system, not by the IACUC system. Death or injury to a human being generates risks of adverse publicity and litigation that are orders of magnitude greater than when laboratory animals are used.

**Andrew N. Rowan**

*The Humane Society of the United States,  
2100 L St NW, Washington DC 20037, USA*

## Conceptual biology: a semantic issue and more

*Sir* — Mikhail V. Blagosklonny and Arthur B. Pardee in their Concepts essay “Unearthing the gems” (*Nature* **416**, 373; 2002) discuss the emergence of ‘conceptual biology’ — the iterative process of analysing existing facts and models available in published literature to generate new hypotheses — and its value in integrating existing information conceptually as an essential part of scientific research. This process has parallels with the process of drug discovery by iteration of existing biomedical literature.

Scientists have traditionally worked in discrete communities, creating discipline-specific language. The natural consequence is that today we are faced with an overwhelming array of nomenclature for genes, proteins, drugs and even diseases.

The  $\beta$ -amyloid converting enzyme BACE, for example, has synonyms ASP2, memapsin 2 and  $\beta$ -secretase, because of different research groups working in different research communities. Catatonic schizophrenia is the more common term for what used to be known as schizophrenic catalepsy, schizophrenic flexibilitas cerea, catatonic dementia

praecox or schizophrenic catatonia. The drug pindolol has more than 30 synonyms, which include LB46, Carvisken, Calvisken and Prindolol.

In some areas, such as the naming of chemicals, there are controls that prevent names from being used more than once. In others, the same name has been used for different entities. The term ‘hunk’ is not only an English word, but also a cell type (human natural killer) and a gene (hormonally upregulated Neu-associated kinase). A search for the term ‘hunk’ in PubMed gives references to all three contexts.

The problem for scientists trying to perform ‘conceptual’ searches precisely and comprehensively is self-evident. It may be reasonably straightforward (though time-consuming) to search in a ‘concept-driven manner’ within their area of expertise, in which they are likely to be familiar with the language. It is considerably more difficult to search outside the discipline, as they may not know where to look and will be less familiar with variable nomenclature. As soon as a name is changed for a compound, disease or target, data associated with that entity could easily be lost. Delving back into history to seek those potentially valuable resources of information can often be impossible.

To address this all-too-familiar problem, many groups are applying

standards to biomedical nomenclature; expert committees are reaching consensus on the naming of gene families, diseases and compounds; there are consortia such as Gene Ontology ([www.geneontology.org](http://www.geneontology.org)), which is defining a controlled vocabulary for the association of genes across different species. From the opposite perspective, there are resources such as the Metathesaurus of the Unified Medical Language System ([www.nlm.nih.gov/research/umls](http://www.nlm.nih.gov/research/umls)), which brings together numerous classification systems so that all alternative names and meanings defined by different groups are in one repository.

These are all extremely valuable resources, but what the biomedical community needs is the ability to use such resources to search the vast sources of information more effectively, to extract more meaning.

The annotation of Medline with medical subject headings (MeSH) is probably the best attempt so far ([www.nlm.nih.gov/mesh](http://www.nlm.nih.gov/mesh)), as it at least helps users to link their search-term to abstracts containing different terms with the same meaning. What is still needed is a way to control the context of the search, so that terms having different meaning in different contexts can be retrieved appropriately. We also need ways to enable scientists to cross disciplines and search in areas outside their expertise, so that they

can extract information critical for new discoveries. Knowledge-based systems will no doubt provide the best opportunity in this regard.

**Julie C. Barnes**

*BioWisdom Limited, Babraham Hall, Babraham, Cambridge, CB2 4AT, UK*

## Uncovering the complex mysteries of mosaicism

*Sir*—I read with great interest your News feature “Dual identities”, bringing together intriguing results that are starting to emerge in a relatively unrecognized area of human genetics, chimaerism and mosaicism<sup>1</sup>. It did not, however, cover single-gene mutations leading to somatic mosaicism, or germline mosaicism. I would like to draw readers’ attention to these developments because the clinical significance of these observations is probably more far-reaching than previously thought.

One of the initial lines of evidence implicating germline mosaicism in single-gene disorders was haplotype-analysis-based detection of the transmission of a muscular dystrophy gene by an unaffected male<sup>2</sup>. Today, we know that a significant number of boys born with Duchenne muscular dystrophy are members of families in which one parent carries the disease-causing mutation only in his or her germ cells, an important issue in genetic counselling. It is increasingly being realized that mosaicism in germ cells can be an important underlying cause of disease in a growing number of genetic disorders<sup>3</sup>.

With respect to somatic mosaicism, we are now beginning to understand the biological mechanism by which some boys with X-linked dominant diseases such as Rett syndrome<sup>4</sup> and incontinentia pigmenti<sup>5</sup> survive.

Although it is not clear whether single-gene mutations mainly occur as a postzygotic event or before fertilization at the half-chromatid stage<sup>6</sup>, the presence of an X-linked dominant disease in a male is an important diagnostic sign of somatic mosaicism.

**Tayfun Özçelik**

*Ayhan Sahenk Foundation and Department of Molecular Biology and Genetics, Bilkent University, Bilkent, Ankara 06533, Turkey*

1. Pearson, H. *Nature* **417**, 10–11 (2002).
2. Darras, B. T. & Francke, U. *Nature* **329**, 556–558 (1987).
3. Zlotogora, J. *Hum. Genet.* **102**, 381–386 (1998).
4. Topçu, M. *et al. Eur. J. Hum. Genet.* **10**, 77–81 (2002).
5. The International Incontinentia Pigmenti Consortium. *Am. J. Hum. Genet.* **69**, 1210–1217 (2001).
6. Gartler, S. M. & Francke, U. *Am. J. Hum. Genet.* **27**, 218–223 (1975).

## Bioinformatics code must enforce citation

*Sir*—Despite repeated calls for the development of open, interoperable databases and software systems in bioinformatics (for example refs 1–3), Lincoln Stein in his Commentary “Creating a bioinformatics nation”, with some justification compares the state of bioinformatics to the mediaeval city-states of Italy, and proposes a unifying code of conduct<sup>4</sup>. In considering his proposal, we must ask why such a chaotic situation arose, and why it has been so persistent.

There are many reasons for the existing chaos. Bioinformatics is a rapidly evolving field. Stable interfaces take time to design, implement and maintain. Algorithms and tools evolve and incorporate feedback from users, and the interfaces must necessarily evolve as well. But standards have been developed and widely accepted in other fields undergoing rapid technological change, such as the Internet.

The difference is that academic scientists are responsible for most of the software and data in bioinformatics. Academic careers are advanced by publications that establish priority and citations that validate the impact of the work. Being the first to develop a new approach forms the basis for a peer-reviewed publication, which is not the case for developing and maintaining a standard interface to an old tool or data set. Academic scientists cannot be expected to sacrifice their careers in the interest of community standards. Significant responsibility for standards development and implementation must fall to service organizations such as database providers. These organizations need the support of the academic community in standards development, whereas academic scientists need to benefit from the time and effort they contribute to the process.

Modern bioinformatics software systems are complex. A genome-annotation system, for example, draws on dozens of software components, involving teams of dozens or even hundreds of developers. We need ways to recognize the often-critical contributions of these individuals to the overall result. Stein’s code of conduct would facilitate the development of seamlessly interoperable systems in a way that hides the underlying complexity of a calculation from the user. From an academic scientist’s perspective, this goal is in direct conflict with the need for recognition and citation and will do nothing for the career of the developer. An academic scientist, therefore, has a strong career imperative to force users to deal

directly with their tool or website, and little incentive to make the technology accessible through interoperable systems.

BLAST<sup>5</sup> and FASTA<sup>6</sup> are “citation classics”, but they are also at the top of the list of “failure to be cited classics”. Projects like NCBI<sup>7</sup> and Ensembl<sup>8</sup> have made useful software tools and large volumes of data widely available, but do not give users the information necessary to cite appropriately the algorithms and software needed to access the system. Ensembl, for example, provides users with alignments performed by BLAST and SSAHA<sup>9</sup> using EST sequences<sup>10,11</sup> aligned to the human genome sequence<sup>12</sup> and gene models created by GenScan<sup>13</sup>. And yet Ensembl lists a citation only to itself on its home page, and the NCBI genome resources pages provide no citation information for the underlying bioinformatics.

If bioinformatics is to emerge as a strong “nation state”, Stein’s code of conduct needs to address the career imperatives of computational biologists. First and foremost, it must require people to cite their sources. Interfaces and data sets should include explicit citation information, so that systems assembled from components can recursively retrieve citation data from their components and present the user with information for all the modules used in a task.

Graphical user interfaces provide ready mechanisms to display the properties of an object. A user clicking on a gene model should be able to retrieve citation information quickly and automatically for the software and data used to assemble that model. This object- and task-specific citation approach would also provide a mechanism for recognizing the specific contributions of developers in large teams. The use of algorithms, software or data without attribution is plagiarism. Manuscripts that fail to cite bioinformatics sources properly are not acceptable for publication in peer-reviewed journals, and software systems that fail to cite their component sources are not appropriate for use by the scientific community.

**David J. States**

*University of Michigan School of Medicine, Ann Arbor, Michigan 48109, USA*

1. Boguski, M. S. *Curr. Opin. Genet. Dev.* **4**, 383–388 (1994).
2. Fasman, K. H. *J. Comput. Biol.* **1**, 165–171 (1994).
3. Karp, P. D. *Trends Biotechnol.* **14**, 273–279 (1996).
4. Stein, L. *Nature* **417**, 119–120 (2002).
5. Altschul, S. F. *et al. J. Mol. Biol.* **215**, 403–410 (1990).
6. Pearson, W. R. & Lipman, D. J. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
7. Wheeler, D. L. *et al. Nucleic Acids Res.* **30**, 13–16 (2002).
8. Hubbard, T. *et al. Nucleic Acids Res.* **30**, 38–41 (2002).
9. Ning, Z., Cox, A. J. & Mullikin, J. C. *Genome Res.* **11**, 1725–1729 (2001).
10. Hillier, L. D. *et al. Genome Res.* **6**, 807–828 (1996).
11. Adams, M. D. *et al. Nature Genet.* **4**, 373–380 (1993).
12. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
13. Burge, C. & Karlin, S. *J. Mol. Biol.* **268**, 78–94 (1997).