

In the name of reproducibility

Adhering to mouse nomenclature guidelines ensures that research is discoverable, replicable, and less wasteful. So why don't researchers do it?

Smriti Mallapaty

John Sundberg, a 67-year-old veterinary pathologist at The Jackson Laboratory in Bar Harbor, Maine, is an unyielding enforcer of mouse nomenclature. He points out discrepancies with accepted norms in almost every paper that he reviews and is wont to publicly rebuke speakers at conferences for referring to incorrect genes and strains.

That goes for his employer too. A few years ago, he called out a colleague for his laxity in citing 'Jax lab'. "I brought to his attention that the official name of this institution is 'The' Jackson Laboratory," says Sundberg.

Dropping the definite article, he told the man, risked confusing the biomedical research institute with an Indian pharmaceutical company or an American antibodies manufacturer, both called some variation of "Jackson lab".

Sundberg's pedantry is encouraged at The Jackson Laboratory, the global custodian of mouse-related terminology. As the institute's newest hire, Sundberg was himself the subject of such censure. In 1988, he had recently published a paper on the mouse papillomavirus in which he had failed to explicate the full strain name of a mouse model. It didn't take long for his colleagues to inform him that all papers were to be reviewed for naming accuracy prior to publication. "I got my knuckles snapped."

Sundberg now feels morally obliged to uphold the laboratory's high standards. "I am carrying the torch now; it is my turn to try to make people understand the significance of doing things correctly."

But the work of protocol purists like Sundberg is becoming ever more difficult as genetic research advances and more and more researchers adopt mouse models in their experiments. "The level of ignorance is the same, but the level of complexity of mammalian genetics has grown exponentially," he says.

Sundberg remembers receiving a complaint call regarding nude mice. These mice lack a type of white blood cell called T cells, which makes them ideal for tissue transplant studies. But the caller's specimens had all rejected the transplant. "Turns out the person who had ordered the mice had



The same, but different: They may all be the same species, but lab mice can vary genetically. Proper nomenclature is critical to account for those differences. (Credit: ocipalla / iStock / Getty Images Plus)

requested hairless mice, which are cheaper, but only have a mild T cell abnormality," says Sundberg.

The biomedical literature is rife with such lapses in clarity. At the pre-publication stage, Sundberg and other reviewers claim papers that get it right are the exception.

A 2013 study, led by biocurator and ontologist Nicole Vasilevsky at Oregon Health & Science University, attempted to quantify the scale of the nomenclature problem¹. In a review of 238 articles in 84 journals, Vasilevsky and her colleagues found that mouse models could be identified in only 67% of mentions. For the remaining 33%, a researcher unconnected to the study would not be able to obtain the resources to reproduce the experiments just from the information provided in the published paper.

"From reading papers every day, the use of standardized nomenclature is very spotty; it is certainly not consistent," says Caroline Zeiss, a veterinary pathologist at Yale School of Medicine. Yet, "of all the variables that affect reproducibility, the use

of standardized nomenclature is one of the easiest fixes."

So what are the rules?

Kinky and wobbly

When geneticist Muriel Davisson created a new mouse model for Down syndrome in 1990, she assigned the name and registered it herself. She was familiar with the protocols, having been responsible for approving new mutations and gene names at The Jackson Laboratory till her retirement in 2012.

Standard nomenclature for mouse models includes details on the strain of the mouse, the lab in which it originated and those it is maintained in, as well as the type and name of mutations it contains (Table 1). The Mouse Genome Informatics (MGI) project at The Jackson Laboratory is the ultimate resource for genetic, genomic, and biological information about the laboratory mouse. Also housed at The Jackson Laboratory is the International Mouse Strain Resource, which collates information about where to find mice that are available from commercial vendors or public mouse strain repositories.

Table 1 | What's in a name? Some questions to consider:

What's the strain?	This is the specific type of mouse. The inbred C57BL/6 is the most popular lab mouse strain; others include inbred BALB/c and DBA mice, or outbred CD-1 mice. Hybrids will need abbreviations for their parents' strains.
Where'd it come from?	When mice are bred in separate facilities, their genetic backgrounds can diverge over time, even if they come from the same original strain—the substrain is indicated by a number or a laboratory code for where it is maintained. For example, C57BL/6J mice are from The Jackson Laboratory, while C57BL/6NTac are from Taconic, by way of the NIH. There are about 50 different C57BL/6 substrains out there, with subtle differences in genetics that can have consequences for scientific outcomes.
What gene is being studied?	Abbreviated gene names come after (sub)strain information. For mice, genetic information is italicized and the first letter is capitalized; particular alleles are italicized and superscripted above the gene.
Was the gene deliberately edited?	Shorthand exists to indicate the method. For example, 'tm' is used for targeted mutations, 't' for translocations, and 'em' indicates an endonuclease-mediated mutation. Transgenic mouse names should include the shorthand 'Tg' and the promoter-gene pair used to make them. As gene editing advances, expect more abbreviations to come.
Who provides it now?	At the end of the name, include a '/' and the laboratory code where the mouse line lives.

Laboratory codes are designated by the International Laboratory Code Registry, maintained by the United States National Academies of Sciences, Engineering, and Medicine. MGI coordinates with these and other collections to maintain its database.

Davisson's Down syndrome model was named Ts(17¹⁶)65Dn. Ts for trisomic, because it contains three copies of parts of chromosomes 16 and 17. It is the 65th chromosomal aberration discovered in Davisson's lab, registered with the code Dn. Other investigators have crossed it into other genetic backgrounds, or combined it with other mutations, so there are a number of different strains that carry this chromosomal aberration registered in MGI. Two repository strains that carry this chromosomal rearrangement include B6EiC3Sn.BLiA-Ts(17¹⁶)65Dn/DnJ and B6EiC3Sn a/A-Ts(17¹⁶)65Dn/J.

That name is the culmination of decades of work. Standards for referring to mouse genetic information were first published in September 1939 by a newly convened Committee on Mouse Genetics Nomenclature, now known as the International Committee on Standardized Genetic Nomenclature for Mice. The committee decided the original rules by ballot, with 23 out of 43 members voting to use symbols based on the first initial of the gene of interest, such as *dw* for the *dwarf* gene, and expressed in italics. The guidelines included details on how to distinguish between recessive and dominant genes, as well as alleles.

Initially, researchers kept abreast of new genetic discoveries via an informal biannual (and later quarterly) bulletin launched in 1949, known as Mouse News Letter. "It was quaint," says bioinformatician Janan Eppig, who managed the MGI database from its inception till her retirement last year. The

newsletters published updates from labs all over the world—the first edition described mutants named *kinky*, *wobbly*, *trembler*, and *paralytic*. "It would include everything from, 'I found this new gene that makes purple spots' to 'so-and-so has moved from my lab over to so-and-so's lab.'" Details about nomenclature and new gene discovery ensured, says Eppig, "that each gene that was discovered had its own unique identity."

Harwell (now known as MRC Harwell Institute) published the newsletter for 48 years, by which time The Jackson Laboratory had set up its own Encyclopedia of the Mouse Genome, a precursor to MGI². Initially released in 1990, before the large-scale adoption of the World Wide Web, the data was distributed via floppy disk to more than 300 investigators and contained details on some 800 known mouse genes, explains Eppig.

MGI today contains data on hundreds of thousands of genes, mutations, alleles and strains. These are integrated with additional information such as gene function, expression, and tumor data. The site tallied 7.3 million page-views in 2016, and served some 100,000 regular visitors.

Researchers use the database to look up existing names, and publications associated

with each biological resource. They can also connect directly with the MGI team for help, or to assign a name to their newly discovered entity.

Sonic, radical fringe

These days, a younger team responds to queries on the MGI nomenclature hotline. Among them is first-responder David Shaw, who answers calls and emails, or redirects requests to specialized groups. Monica McAndrews is practiced in the ways of genetic nomenclature, assigning names to new entries. For alleles and strains, requests are handled by Cynthia Smith's staff.

But the team can't always play by a rulebook. "People often give memorable symbols and names to these elements," says McAndrews. Asking them to give up a clever nickname or an ego-rubbing eponym for an impersonal series of letters can trigger a rebuff. "Sometimes we just have to back off and wait, or make an exception if things are entrenched," she says. But they prefer not to, rejoins Smith.

In 1993, molecular biologist Robert Riddle discovered a gene that regulates vertebrate growth in chick embryos and named it *sonic hedgehog*, or *Shh* for short. When the same gene was isolated in mice, the researchers retained the name. "To change it would have caused lots of confusion," says Smith. The human nomenclature community, however, was not pleased, out of concern for children that might be affected by the mutation. *Sonic hedgehog* was among several genes, including *lunatic fringe* and *radical fringe*, that the human genome committee tried to eradicate more than a decade ago. The latter two gene names were eventually changed in human, rat and mouse, but their original symbols remain the same.



Janan Eppig managed Mouse Genome Informatics as it grew from floppy discs to a major online data resource for all things lab mouse.
(The Jackson Laboratory)



Need nomenclature help? Monica McAndrews and her colleagues on the Mouse Genome Informatics team are standing by. (The Jackson Laboratory)

Synchronizing standards with those in other communities is important for MGI. “I communicate with them [the human and rat nomenclature committees] almost daily,” says McAndrews. Names act as links in the long chain of discovery. By certifying their accuracy and consistency within and across species, the MGI team ensures that researchers build on the work of their predecessors, and don’t waste resources duplicating experiments. Researchers who shun the rules also make their own work harder to discover. Unfortunately, many do.

Querying the MGI database is revealing. Take for example the non-existent gene *TAP*, which has been used in articles to refer to five different genes. “Someone reading one of those papers has to first figure out which of those five genes the researcher is talking about,” says Shaw.

The problem is compounded when trying to identify specific alleles. Knocking out the tumor suppressor gene *Brca2*, for example, can increase the risk of breast cancer, but only when targeting the correct allele. “If you don’t tell me which exact mutation you are working on, or which exact genetic alteration you have made, it makes reproducibility a lot more difficult,” says Smith.

That goes for broader genetic backgrounds too. In 2015, Zeiss uncovered a mutation with major implications for research into Alzheimer’s disease. She analyzed some 500 intervention studies of Alzheimer’s disease and found that more than half used mouse models carrying mutations that make them blind³. “Many of the tests to assess memory depend on a mouse being able to see,” says Zeiss. “Failure to recognize the confounding influence of background genetics undermines research results. It implies that we may be using a lot of mice in studies that ultimately aren’t all that impactful.”

Up-to-date

In the 1920s and 1930s, The Jackson Laboratory generated the C57BL/6 strain, which has since become the most popular

laboratory mouse. Over generations of breeding, however, descendants of the original ancestor have gained random mutations, some of which can have physiological effects, from liver toxicity to immunodeficiency and their preference for alcohol⁴.

These substrains are distinguished by symbols for the lab code of the investigator or institution where they were developed or discovered. C57BL/6, for example, was appended with J (for The Jackson Laboratory) a few generations after its initial characterization to distinguish it from C57BL/6 substrains originating elsewhere.

Every so often though, a researcher will identify mutations in substrains no-one previously knew existed, prompting a review and potential revision of strain nomenclature. In 2001, researchers in London found that a subpopulation of C57BL/6J distributed by the mouse supplier Harlan, now Envigo, contained mutations in the alpha-synuclein gene, which has been associated with rare inherited forms of Parkinson’s disease. The researchers proposed renaming the strain C57BL/6S, but since all animals in the substrain carried the mutation, Envigo maintained the original name: C57BL/6J^{OlaHsd} (both *Ola* and *Hsd* are lab codes for Envigo). “In these cases, vendors typically consider the mutation a characteristic of the model,” says Paul Surdez, a vice president at the company.

More recently, in 2016, a team led by Shiv Pillai at Harvard Medical School discovered a population of C57BL/6N^{Hsd} that had spontaneously generated mutations in the *Dock2* immune system gene. The mutations were found in animals bred in 6 of Envigo’s 19 facilities. Envigo decided not to continue selling the subpopulation of mice that contained the mutation, and therefore did not change the nomenclature.

Just as strain names must be revisited over time, the International Committee on Standardized Genetic Nomenclature for Mice is also currently undergoing a retrospective related to genes. The Wellcome Sanger Institute in the United Kingdom recently released complete genomes for 16 of the most common inbred mouse models used in biological studies. Comparing these with the current mouse reference genome, which is C57BL/6J, reveals lots of variations.

The committee now has to decide whether some of the genes found in certain strains are sufficiently different to those present in the reference genome to be called by another name, explains Smith, who is the current committee chair. Duplicated genes, for example, might be given separate, sequential names, such as *Ren1* and *Ren2*.

Other genes in the same genetic location but with entirely different sequences of code might be renamed with alphabetical suffixes, such as *Tmem1a* or *Tmem1e*, depending on how substantially different they are.

“We are going to have to go case by case,” says Smith. “We just haven’t had enough time yet to go through all of the strains or find enough examples to start setting any rules.”

Into the journal

Granted, naming conventions are complex, but the support offered is steadfast. Shaw, McAndrews and Smith are happy to be flooded with email queries. “Contact us,” says Smith, for questions about the process, and to verify names, suggest revisions, or even dispute a decision. “We are always interested in feedback,” says McAndrews.

Scientists and their lexicologists aren’t the only ones responsible for ensuring quality, however. The entire scientific community should maintain a degree of oversight, says Cory Brayton, a pathologist at Johns Hopkins Medicine, from the funders to the peer reviewers, and, eventually, the journal editors.

“Journals represent a bottleneck that everybody has to go through to get their studies published,” says Zeiss. If they insisted on using standardized nomenclature, she says, it would ensure more accurate reporting. “Within the lab animal community this is a very old story,” she says. The challenge is getting the messages out to the larger scientific community, she adds.

Several initiatives have attempted to raise awareness about the issue, and its broader implications for reproducibility. In 2010, the United Kingdom’s National Centre for the Replacement, Refinement & Reduction of Animals in Research published a set of guidelines to improve the reporting of research that uses animals. The ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) guidelines have been translated into



“Of all the variables that affect reproducibility, the use of standardized nomenclature is one of the easiest fixes,” says Caroline Zeiss from Yale University. (Yale University)

seven languages and come with a 20-point checklist. Item 8b covers the need to include relevant details on the source, strain and genetic modification of animal models. In 2011, the Institute for Laboratory Animal Research (ILAR) at the US national academies published guidance for describing animal research. It outlined information related to an animal's source and genetic nomenclature, and more—age, sex, weight, life stage and breeding environment. And the US National Institutes of Health have also added a section to their grant applications, in which researchers have to describe the methods they will use to ensure the identity of key biological resources.

The idea has been making its way to publishers too. In 2013, *Nature* created a

checklist to ensure that authors in the life sciences are consistent and transparent about reporting relevant information for research reproducibility. The checklist refers to the ARRIVE guidelines and prompts researchers to note the species, strain and sex of laboratory animals.

A checklist isn't enough, says Brayton. Researchers would greatly benefit from online tutorials for the ARRIVE guidelines and the ILAR guidance, she says. "That would be a really powerful tool"

Despite these initiatives, many journals are rather lenient with their execution. Several chief editors who have been applauded by the mouse nomenclature committee for enforcing approved rules agree standardized reporting is important, and rely on a mixture of editors, reviewers,

and subeditors to check for errors. But none have strict procedures in place.

"You have to stop it at the point of distribution," says Sundberg who refuses to publish in journals (even "high-end" ones) that don't meet his standards. His work is far from done. □

Smriti Mallapaty

Freelance writer, Sydney, Australia.

e-mail: smriti.mallapaty@gmail.com

Published online: 25 June 2018

<https://doi.org/10.1038/s41684-018-0095-7>

References

1. Vasilevsky, N. A. et al. *PeerJ* **1**, e148 (2013).
2. Eppig, J. T. *ILAR J.* **58**, 17–41 (2017).
3. Zeiss, C. J. *Drug Discov. Today* **20**, 475–482 (2015).
4. Reardon, S. *Nature* **551**, 281 (2017).

PLAS LABS INC.®

CONCEPTION ■ INTERPRETATION ■ PRODUCTION ■ SATISFACTION

Producing equipment, helping
YOU produce results!
Since 1967!



PLAS LABS, INC.®

www.PLAS-LABS.com
PLI@PLAS-LABS.com
Lansing, Michigan USA