INNOVAPATH™
The Premier Journal Bridging Discovery & Innovation

Editorial
https://doi.org/10.63501/qzsx8v40

The Wrong Animal: Genetic Mislabeling in Laboratory Mouse Research and the Promise of AI-Assisted Quality Control

Sohail Rao, MD, MA, DPhil[1]

[1]Editor-in-Chief, INNOVAPATH™ Journal, 6918 Camp Bullis Road, San Antonio, TX 78256.

[1]Corresponding Author Email: editor@innovapath.us

It is a habit of mine, and one I commend to younger colleagues, to set aside a portion of each week for a deliberate reading of the news pages of Nature. The peer-reviewed literature tells one where the field has been; the news pages, read attentively, tell one where the field is going, and where it is quietly going wrong. This week, I came across a piece in Nature News, published on May 15th, 2026, that summarizes a Science study by Fernando Pardo-Manuel de Villena and colleagues at the University of North Carolina at Chapel Hill, published the same day. The finding is sobering and deserves the attention of every investigator, sponsor, and reviewer whose work touches on preclinical biology.[1,2]

Roughly half of the laboratory mouse strains the team genotyped, drawn from a network of repositories established by the U.S. National Institutes of Health and distributed globally for animal research, do not match the genetic identity under which they are cataloged and used. Of 611 samples spanning 341 strains, 47% were inconsistent with their descriptions. For those of us who have spent decades navigating the long road from bench to first-in-human studies, this is not a curiosity confined to a specialist subfield. It is a quiet structural problem sitting beneath an enormous edifice of published biomedical claims and beneath a great many regulatory submissions that have moved, on the strength of those claims, into the clinic.[1]

The Mechanism of the Failure

The mechanism is mundane, and that is precisely what makes it dangerous. Introducing a single genetic alteration, a knockout, a conditional allele, or a reporter into an inbred background is supposed to require ten to twenty generations of careful backcrossing, with documentation to match. When that discipline is shortcut, donor-strain genetic variation is retained in the recipient, and the supposed clean comparison between mutant and wild-type animals quietly becomes something else: a comparison between two genetically heterogeneous populations whose differences are no longer confined to the gene of interest. Time and money are the usual culprits. Twenty generations, as Pardo-Manuel de Villena has observed, is a long and costly road.[9]

The consequences are not hypothetical. Daniel Rawle and colleagues showed in 2022 that genotype discrepancies in granzyme A-knockout mice had produced a false mechanistic conclusion about protection against chikungunya-induced arthritis. One mislabeled colony, one wrong inference, one chain of downstream studies built on sand. How many such chains exist across oncology, immunology, neuroscience, and metabolic disease? We genuinely do not know, and that uncertainty is itself the indictment.[3,4]

https://doi.org/10.63501/qzsx8v40

Three Implications Worth Airing

First, the reproducibility conversation has been dominated for more than a decade by statistics, blinding, registration of analytic pipelines, and the discipline of pre-specified hypotheses. This study is a useful corrective. It reminds us that the [5,6]*biological substrate itself* can be the failure point. No amount of methodological rigor downstream rescues a misidentified starting material upstream. A perfectly pre-registered, perfectly blinded, perfectly analyzed experiment performed on the wrong animal is not a contribution to knowledge; it is a contribution to confusion.

Second, journals, funders, and regulators should treat strain-genotype verification with the same seriousness as cell-line authentication, which emerged after the HeLa contamination scandals and which the NIH eventually formalized as a condition of funding. Nominal strain identity is not identity. A cage card is not a genotype. The minimum acceptable standard for a methods section ought to be documented, contemporaneous genotype verification, not the name a colony has been called for the past fifteen years.[9][7,8]

Third, those of us who chair scientific review and institutional review committees must ask sharper questions of the preclinical packages submitted in support of first-in-human studies. *What strain did you use?* is no longer sufficient. *How do you know?* must follow. So, must: When was the strain last genotyped? Against what reference? By whom? These are not pedantic questions. They are the difference between a mechanism that exists and a mechanism that was confected by genetic drift

A Broader Reflection

There is a wider lesson here, one that resonates with concerns I have raised in other contexts about the cultures of speed and credentialism that have come to dominate parts of biomedical research. The pressure to publish, to patent, to translate, and to move to the clinic has compressed the time horizons of careful experimental husbandry to a degree that the founding generation of mouse geneticists would have found alarming. Twenty generations of backcrossing is not bureaucratic obstruction; it is the price of knowing what one is looking at. When that price is deemed too high, the bill comes due later in failed Phase II trials, in irreproducible mechanistic literature, and in the slow erosion of public trust in the basic claims of biomedical science.

The mouse has carried biomedical science for the better part of a century. It has served as the proximate witness to nearly every major therapeutic advance of the modern era. We owe it, and the patients who stand downstream of its work, the elementary courtesy of knowing who it actually is. And we owe ourselves the weekly discipline of looking up from our own benches long enough to notice when a foundation we have long taken for granted has begun, quietly, to shift.

Can Artificial Intelligence Help?

The more consequential near-term intervention may therefore be a policy one: making AI-assisted genotype verification a mandatory condition of NIH funding for mouse-based research, in precisely the way cell-line authentication was eventually formalized as a condition of award. The technology is ready. The question, as it so often is in these matters, is whether the institutions are.[7,8]

https://doi.org/10.63501/qzsx8v40

The honest limits of this optimism deserve equal space. Artificial intelligence does not resolve the underlying problem, which is cultural and economic in nature. Twenty generations of backcrossing remain expensive and slow regardless of how sophisticated the verification technology becomes. The incentive structures that led researchers to shortcut that process, the pressure to publish, the undervaluation of quality control work in grant review, the invisibility of animal husbandry in budgets and timelines are not amenable to algorithmic correction. What AI can do is lower costs and increase verification speed to the point where impracticability is no longer a credible excuse, and make the consequences of non-compliance visible early enough to be corrected before they are published.[5,6]

On remediation, AI's role is more limited but real. It cannot correct a mislabeled colony retroactively. It can, however, help prioritize which published findings are most likely to be compromised, guide the sequencing of replication efforts, and assist journals and funders in triaging the reanalysis workload. Causal inference models trained on the phenotypic literature could, in principle, estimate how much of an observed effect is attributable to background genetic contamination rather than the experimental manipulation of interest, partitioning variance in a way that allows at least partial salvage of conclusions from otherwise compromised datasets.

Upon detection in the published literature, natural language processing models can now scan methods sections at scale and flag papers that report the use of a strain subsequently identified as mislabeled in repository records, describe backcrossing procedures inconsistent with the generation numbers claimed, or cite a colony source that has been implicated in documented discrepancies. This approach is already being applied in the cell-line authentication context; text-mining tools, combined with the ICLAC register, have identified thousands of papers that use known misidentified lines. The same logic transfers directly to the mouse strain problem, and at a scale no team of human reviewers could replicate.[7,12]

On prevention, AI's contribution is most immediate and most tractable. The MiniMUGA array developed by Pardo-Manuel de Villena's group is already a high-throughput genotyping platform; coupling it with machine-learning models trained on verified reference-strain profiles enables rapid, low-cost authentication of mouse colonies before experiments begin. AI-assisted genotype calling can flag discrepancies between a colony's declared identity and its actual genomic signature in hours rather than weeks. Colony management systems incorporating AI-driven anomaly detection can monitor allele frequencies in real time, identifying genetic drift inconsistent with the claimed inbred background long before it propagates into experimental cohorts and the literature.[11][9,11]

The question that follows naturally from this account is whether the tools now reshaping so much of biomedical research can be turned on the problem itself. The answer is yes, and in several distinct ways that are worth separating carefully, because the role of artificial intelligence differs depending on whether we are addressing prevention, detection, or remediation.

Keywords: *mouse strain authentication, genetic mislabeling, preclinical reproducibility, genotype verification, laboratory mouse, artificial intelligence, research integrity, MiniMUGA, cell-line authentication, NIH quality control*

REFERENCES

1. Pardo-Manuel de Villena F, Amos-Landgraf JM, Bell TA, et al. Improve genetic quality control to increase rigor and reproducibility of mouse research. Science. 2026;392:698–700. https://doi.org/10.1126/science.aec3177
2. Maxmen A. Genetic survey exposes flaws in widely used mouse models. Nature. 2026 May 15. https://www.nature.com/articles/d41586-026-01534-4
3. Rawle DJ, Le TT, Dumenil T, Bishop C, Yan K, Nakayama E, Bird PI, Suhrbier A. Widespread discrepancy in Nnt genotypes and genetic backgrounds complicates granzyme A and other knockout mouse studies. eLife. 2022;11:e70207. https://doi.org/10.7554/eLife.70207
4. Wilson JAC, Prow NA, Schroder WA, et al. RNA-Seq analysis of chikungunya virus infection and identification of granzyme A as a major promoter of arthritic inflammation. PLOS Pathog. 2017;13(2):e1006155. https://doi.org/10.1371/journal.ppat.1006155
5. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483:531–533. https://doi.org/10.1038/483531a
6. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011;10:712. https://doi.org/10.1038/nrd3439-c1
7. International Cell Line Authentication Committee (ICLAC). Register of Misidentified Cell Lines. Version 12. 2023. https://iclac.org/databases/cross-contaminations/
8. NIH National Center for Advancing Translational Sciences. Cell Line Authentication. https://ncats.nih.gov/research/research-activities/cell-line-authentication
9. Blanchard MW, Bell TA, Buis JM, et al. The updated Mouse Universal Genotyping Array bioinformatic pipeline improves genetic QC in laboratory mice. G3 (Bethesda). 2023;13(4):jkad036. https://doi.org/10.1093/g3journal/jkad036
10. Pardo-Manuel de Villena F, Replication data for: Improve genetic quality control standards to increase rigor and reproducibility of mouse research. UNC Dataverse. 2025. https://doi.org/10.15139/S3/MLJJBK
11. VanDenBerg KR, Oravecz-Wilson K, Krolikowski L, Hill V, Reddy P, Freeman ZT. Impact of automated genotyping and increased breeding oversight on overall mouse breeding colony productivity. Front Physiol. 2022;13:925784. https://doi.org/10.3389/fphys.2022.925784
12. Hatherley R, Rankin N. Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. PLoS ONE. 2019;14(2):e0205418. https://doi.org/10.1371/journal.pone.0205418