

## Science &amp; Society

## Towards the ethical development of functional genomics research

Ian Stevens <sup>1,\*</sup>, Danielle Pacia<sup>1</sup>, Jan N. Hansen<sup>2</sup>, Jillian Parker<sup>3</sup>, Pamela Payne-Foster<sup>4</sup>, Trey Ideker<sup>3</sup>, Jean-Christophe Bélisle-Pipon<sup>5</sup>, and Vardit Ravitsky<sup>1</sup>

**Contemporary functional genomics researchers have sought to couple high-throughput experimental technologies with computational methodologies. However, the ethical issues associated with bench science at the early stages of the artificial intelligence (AI) lifecycle have received less attention and merit further examination, as addressing issues early on can mitigate downstream challenges in healthcare.**

### Introduction to functional genomics and its context

Functional genomics aims to explore and understand how genes and intergenic regions of the genome influence various biological processes. These components encompass different cellular systems, each studied through distinct 'omics' fields, including genomics, transcriptomics, proteomics, and metabolomics, to name a few. When taken together, a multi-omics approach facilitates a broader understanding of the molecular connections that lead to phenotypic characteristics resulting from an underlying genotype [1]. By leveraging these insights, functional genomics has the potential to identify novel biomarkers and predict individual patient responses

to interventions, thereby addressing challenges in understanding rare diseases and population-level variations.

In the mid-2000s, the ethics of functional genomics focused on the increasingly blurred distinction between human and animal research and the potential for this issue to undermine the justification for prioritizing animal experimentation. This distinction became even less clear due to scientific advancements that enabled the transfer of uniquely human genetic sequences into animals, coupled with the growing recognition that humans share varying degrees of genomic sequences with animals [2]. While these ethical and scientific discussions have continued, advancements in -omics profiling and high-throughput experimental techniques to study human samples and the concomitant increase in the use of AI/machine learning (ML) have brought novel ethical tensions to light.

### The role of ethics in basic research

There are several examples of successful collaborations between ethicists and laboratory scientists, such as determining whether a product of stem cell research should be classified as an embryo and how to best describe nervous system organoids of varying morphologies [3]. Such collaborations strive to address ethical issues as they emerge in basic science, to respect ethical values, and to improve practices that may have downstream implications in terms of how results are implemented. Ethics plays a critical role not as a restrictive force, but as an enabler of responsible innovation by embedding moral considerations early in the research lifecycle. Neglecting ethics can result in the unintentional incorporation of biases into AI models, further exacerbating inequalities and reducing public trust. Successful collaboration between scientists and ethicists needs to occur over time to foster mutual trust through co-reasoning, where each person openly participates in

shared decision making [3,4]. Such collaboration can help integrate ethics into AI research in ways that facilitate responsible innovation, rather than exacerbating the perception that ethics 'polices' scientific research [5].

### Structural impacts on decision making

Multiple structural factors implicitly or explicitly influence how functional genomics research is conducted. These factors need to be considered when evaluating the moral integrity of scientific practices, especially as the application of AI may increase the risk of exacerbating historically biased practices. Scientists' high level of expertise and specialization in a specific topic may guide their choice of which conditions to investigate, which methods to use, and which patient populations to prioritize. The priorities of the funding agencies, which can be based on societal values, may impact which research areas benefit from investment and which ones are neglected. In the lab, scientists may prefer a particular cell line because it is more accessible, less expensive, or more conducive to experimental manipulation. Furthermore, scientists may choose a cell line because it has been used by many others in the past and hence it is associated with large amounts of previous data useful for integration and cross-comparison. Such factors can result in an ethically problematic under-representation of cell lines from minority populations or rare diseases, since the majority of widely available cell lines are derived from individuals of European descent. This parallels other commercial and open genomic databases, such as the Personal Genome Project [6], and sets the stage for training AI models on historically biased datasets.

### (Some) ethical points to consider

#### Implications for generalizability

When AI-aided functional genomic research examines the multi-omics characteristics of a small number of cell lines,

the generalizability of the project has a defined scope. This aspect follows the trend set by the Human Genome Project's legacy of focusing on discovery-driven science, while also implicitly contributing to the contemporary 'precision medicine' movement that has been growing since the early 2010s [7]. In general, precision medicine strives to acquire large amounts of health information from patients with the goal of finding each individual's optimal treatment. The mantra that 'knowledge is power' may result in patients being considered increasingly responsible for their own health, as increasing amounts of data are collected from them regardless of causative, external sociologic factors that may be out of their control [8]. In the research context, precision medicine is also challenging traditional clinical trial methodology – which relies on large

cohort studies to evaluate a treatment's safety and efficacy – as cohorts are stratified into smaller and smaller sample sizes based on the unique physiologic presentation of patients (Table 1) [9].

**Incorporating diversity into functional genomics research**

Given the biases that racial categories entail, it is tempting to simply try and remove the socially constructed notion of race from science altogether. However, some structural practices of science are built to perpetuate it (<https://www.congress.gov/bill/103rd-congress/senate-bill/1>). Despite this challenge, an honest pursuit of genetic diversity in research studies, such as using race and/or ethnic categories as imperfect guides to increasing inclusivity, helps balance the pursuit of objective scientific findings in an inherently biased world. For

example, in the case of the increased mortality from triple-negative breast cancer among African American women, research is ongoing to parse environmental causes of genomic instability from underlying inequalities in healthcare access [10]. Going beyond the need for genomic diversity, it remains unclear where in the multi-omics framework the ethical imperative to strive for inclusive samples becomes outweighed by the perceived objective nature of the research being conducted. Given the insights that different populations can contribute to scientific knowledge, it makes sense to strive for genomic diversity; however, it may be less clear why there should be such a requirement for scientists studying the basic, fundamental science of protein-protein interactions. Thus, which omics level(s) ought to incorporate diversity remains a topic open for further examination (Table 1).

Table 1. Ethical points to consider in functional genomics research using AI

Ethical point	Description	Key implications
Implications for generalizability	Functional genomics research often relies on small, specific datasets, raising concerns about the generalizability of results. The rise of precision medicine emphasizes personalized data but may unfairly shift responsibility for health outcomes to individuals while ignoring broader sociological factors	Risk of bias in precision medicine Challenges to traditional clinical trial models due to smaller, stratified cohorts
Incorporating diversity	While genomic diversity is essential for understanding population-specific traits and addressing health disparities, the need for diversity at other 'omics' levels (e.g., proteomics) is less clear. Structural scientific practices perpetuate racial biases, making it vital to approach diversity as an ethical imperative	Ethical tension in balancing inclusivity with scientific objectivity Need to ensure that research benefits all populations, especially under-represented groups
Value of interpretability	AI research's 'black box' nature can undermine trust and transparency. Efforts to improve interpretability, such as using VNNs, are promising but require further empirical research to determine where transparency matters most in omics studies	Trustworthiness of AI-powered research depends on transparency Engagement with stakeholders (e.g., clinicians, patients) is needed to improve trust and acceptance
Consent and legacy cell lines	Many legacy cell lines lack clear consent documentation. As AI advances increase re-identification risks, reliance on de-identified specimens raises ethical concerns. Researchers are encouraged to use cell lines with documented consent while balancing resource availability and inclusion of under-represented groups	Growing ethical challenges in using legacy cell lines Need for evolving consent practices in response to AI's data re-identification capabilities

**The value of interpretability**

The absence of explainability or interpretability in AI-based research may reduce the trustworthiness of the science, which represents an ethical challenge. Recently, visible neural networks (VNNs) that use experimentally derived maps of cellular architecture as the foundation for ML approaches have been developed and successfully applied to mitigate some of the 'black box' nature of AI-driven predictions [11]. However, it remains unclear where else interpretability should be prioritized in any given project. For example, which component of the omics data ought to be easily interrogatable and why? Additional empirical social science research (e.g., clinical surveys and interviews) is necessary to explore whether transparency (i.e., explainability or interpretability) has an impact on trust in and trustworthiness of interventions that are based on AI-powered research. Moreover, empirical research is needed to explore methods of engagement with users and characterize how these methods may impact trust. This is especially important in the context of a healthcare system serving marginalized

communities familiar with the racist and classist history of medicine that continues to influence the present (Table 1) [12].

### Impact of legacy cell lines on consent practices

The informed consent documentation associated with many legacy cell lines is vague and often non-existent. Even though the Common Rule does not currently require that de-identified specimens be obtained with informed consent [13], such a standard is becoming harder to rely on as AI improvements continue to increase the capacity for patient re-identification through their genomic information [14]. Thus, it is strongly recommended that, when possible, researchers use cell lines with clear, documented consent. Such a principle is not absolute, as it is also understood that such ethical considerations must be balanced against availability of resources, reliability of the cell line, and prioritization of under-represented demographics and diseases. Researchers should continue to consider all these variables carefully, weighing them against one another, while aiming to select cell lines that have clear, documented consent. However, as the medical landscape continues to adjust to the integration of AI technologies, simply relying on informed consent could prove insufficient in a few decades (Box 1). Thus, in the spirit of transparency that consent fosters, scientists are encouraged to interact with the relevant communities regarding their research endeavors (e.g., tissue

sample donors or their descendants, patient populations, disease specific advocacy organizations, etc.) (Table 1).

### Concluding remarks

The ethical tensions associated with functional genomics research that employs AI deserve more dedicated scholarship. Prioritizing this multidisciplinary research helps tease apart the inherent structural factors that influence value-laden decisions in the scientific process, the commitments that come with research relying on small sample sizes, the ambiguities surrounding diversity and interpretability requirements within multi-omics research, and the challenges of utilizing legacy cell lines and de-identified specimens moving forward. Integrating ethical considerations into these areas – such as ensuring transparency in AI models, incorporating diverse datasets for the sake of generalizability, and balancing consent practices – helps mitigate potential downstream challenges in resulting medical applications. Proper integration of ethics into the research ecosystem, particularly early on in the lifecycle, ensures that research practices benefit from insights before ethical challenges are imprinted into the resulting products of such labor-intensive work.

### Acknowledgments

This article was supported through the Bridge2AI program, National Institutes of Health Grant Number: 1OT2OD032742.

### Declaration of interests

T.I. is a co-founder of Data4Cure, Inc., is on the scientific advisory board, and has an equity interest. T.I. is on the scientific advisory board of Ideaya Biosciences, Inc. and has an equity interest. T.I. is on the advisory board of the Cell Systems journal. The remaining authors have no interests to declare.

<sup>1</sup>The Hastings Center, Harrison, NY, USA

<sup>2</sup>Stanford University, Stanford, CA, USA

<sup>3</sup>University of California San Diego, School of Medicine, San Diego, CA, USA

<sup>4</sup>The University of Alabama School of Medicine, Tuscaloosa, AL, USA

<sup>5</sup>Faculty of Health Science, Simon Fraser University Burnaby, BC, Canada

\*Correspondence:  
ims28@nau.edu (I. Stevens).

<https://doi.org/10.1016/j.tibtech.2025.10.015>

© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

### References

1. Ali, A.M. and Mohammed, M.A. (2024) A comprehensive review of artificial intelligence approaches in omics data processing: evaluating progress and challenges. *Int. J. Math. Stat. Comput. Sci.* 2, 114–167
2. Hoeyer, K. and Koch, L. (2006) The ethics of functional genomics: same, same, but different? *Trends Biotechnol.* 24, 387–389
3. Lunshof, J.E. and Rijssenbeek, J. (2024) Collaborative ethics: innovating collaboration between ethicists and life scientists. *Nat. Methods* 21, 1571–1574
4. Pacia, D.M. *et al.* (2024) Early AI lifecycle co-reasoning: ethics through integrated and diverse team science. *Am. J. Bioeth.* 24, 86–88
5. Anon (2024) Research ethics matter. *Nat. Methods* 21, 1569
6. Norori, N. *et al.* (2021) Addressing bias in big data and AI for health care: a call for open science. *Patterns* 2, 100347
7. Juengst, E. *et al.* (2016) From 'personalized' to 'precision' medicine: the ethical and social implications of rhetorical reform in genomic medicine. *Hast. Cent. Rep.* 46, 21–33
8. Collins, B.X. and Wilkins, C.H. (2024) Overcoming barriers to health equity in precision medicine research. *Am. J. Bioeth.* 24, 86–88
9. Vogt, H. and Hofmann, B. (2022) How precision medicine changes medical epistemology: a formative case from Norway. *J. Eval. Clin. Pract.* 28, 1205–1212
10. Dietze, E.C. *et al.* (2015) Triple-negative breast cancer in African-American women: disparities versus biology. *Nat. Rev. Cancer* 15, 248–254
11. Park, S. *et al.* (2024) A deep learning model of tumor cell architecture elucidates response and resistance to CDK4/6 inhibitors. *Nat. Can.* 5, 996–1009
12. Jangoan, S. *et al.* (2024) Demystifying explainable AI: understanding, transparency, and trust. *Int. J. Multidiscip. Res.* 6, 1–13
13. Spector-Bagdady, K. *et al.* (2019) Biospecimens, research consent, and distinguishing cell line research. *JAMA Oncol.* 5, 406–5410
14. Kudiabor, H. (2024) 'Anonymous' genetic databases vulnerable to privacy leaks. *Nature* 634, 764–765
15. Calleau, R. *et al.* (1978) Long-term human breast carcinoma cell lines of metastatic origin: preliminary characterization. *In Vitro* 14, 911–915

### Box 1. An example from the front lines

As an amalgamation of researchers, we too are coming at this topic from the lens of pre-clinical studies given we are involved with the Cell Maps for AI (CM4AI) consortium, which is itself affiliated with the broader Bridge2AI Project (<https://cm4ai.org/>). This functional genomics research endeavor is generating data in the triple-negative breast cancer cell line MDA-MB-468 and the induced pluripotent stem cell (iPSC) line KOLF2.1J under various treatment conditions and differentiation lineages. As it relates to this article, it is important to highlight that the MDA-MB-468 cell line was originally obtained from a 51-year-old Black woman during a procedure to aspirate a plural effusion at what is now known as MD Anderson Cancer Center in November in 1977 [15]. It is unclear to what degree her consent was obtained, or if she was aware that cells originating from her biological specimen had the potential for use in contemporary research protocols nearly 50 years later. As such, our team has been considering how these ambiguities surrounding the original consent could be disclosed to readers of scientific publications so that they can weigh this information with the findings of the study when determining its validity and ethical appropriateness for themselves. We hope this example helps demonstrate the complexity of achieving transparency due to scientific practices established in the past that are now becoming the foundation of AI models.