COMMENTARY Pharmacology

www.the-innovation.org/the-innovation-drugdiscovery

Reshaping the drug discovery ecosystem with open science and collaborative innovation

Chaoyang Xie,[1,3] Xiaorui Wang,[4] Yawen Dong,[5] Xiaojun Yao,[3,*] Gefei Hao,[2,*] and Yuquan Li[1,2,*]

[1]State Key Laboratory of Public Big Data, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China

[2]State Key Laboratory of Green Pesticide, Center for Research and Development of Fine Chemicals, Guizhou University, Guiyang 550025, China

[3]Centre for Artificial Intelligence Driven Drug Discovery, Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR 999078, China

[4]College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

[5]School of Pharmaceutical Sciences, Guizhou University, Guiyang 550025, China

*Correspondence: xjyao@mpu.edu.mo (X.Y.); gfhao@gzu.edu.cn (G.H.); yvquan.li@gzu.edu.cn (Y.L.)

Received: March 29, 2026; Accepted: April 8, 2026; Published Online: April 13, 2026; https://doi.org/10.59717/j.xinn-drugdisc.2026.100016

Citation: Xie C., Wang X., Dong Y., et al. (2026). Reshaping the drug discovery ecosystem with open science and collaborative innovation. *The Innovation Drug Discovery* **1**:100016.

The pharmaceutical industry faces an unprecedented productivity crisis. The average cost of developing a new drug has surpassed 2.6 billion, with development timelines stretching beyond a decade, yet clinical success rates remain stubbornly below 10%.[1] This productivity crisis—where escalating costs fail to translate into proportional therapeutic advances—signals a fundamental limitation in the traditional closed-innovation model. Proprietary research behind corporate walls creates data silos, redundant efforts waste billions on repeated failures, and the "file drawer problem" of unpublished negative results perpetuates costly mistakes. These barriers are increasingly untenable. Modern drug discovery generates unprecedented data volumes through artificial intelligence and multi-omics technologies—volumes that exceed any single organization's capacity to exploit. A transformative shift is underway: open science—the movement toward transparency, data sharing, and collaborative research practices—and collaborative innovation are breaking down these barriers and reshaping how drugs are discovered, offering a path toward greater efficiency, reduced costs, and accelerated therapeutic breakthroughs.

FROM SILOS TO SYNERGY: BREAKING THE DATA BARRIER

Traditional drug discovery operates on a proprietary model where pharmaceutical companies protect intellectual property at every stage. While this approach has delivered important medicines, it erects formidable barriers to progress. The "file drawer problem"—where unsuccessful experiments remain unreported—leads to wasteful repetition of failed approaches across different organizations. Furthermore, modern drug discovery increasingly exceeds any single organization's capacity. Understanding disease mechanisms now requires integrating multi-omics data, structural biology, clinical phenotypes, and real-world evidence across diverse populations. No institution possesses all necessary expertise, technologies, and patient access, yet institutional walls prevent the data integration essential for therapeutic innovation.

Open science breaks these barriers by enabling researchers worldwide to build upon each other's work and collectively tackle challenges too large for any single entity. Large-scale genomic initiatives exemplify how dismantling data silos accelerates drug discovery: the UK Biobank's whole-genome sequencing of nearly half a million participants, released as an open resource, has enabled thousands of researchers globally to identify disease-associated variants and validate therapeutic targets that would have remained hidden behind institutional barriers.[2] Similarly, open chemical databases like ChEMBL and PubChem now contain millions of compounds with associated bioactivity data, eliminating redundant screening campaigns and enabling researchers globally to identify drug discovery starting points without duplicating expensive efforts.

OPEN DATA ECOSYSTEMS: THE LIGAND-AI PARADIGM

The convergence of artificial intelligence and open science is catalyzing a new era in drug discovery. Open data ecosystems—the technical infrastructure, platforms, and collaborative frameworks that operationalize open science principles—are central to this transformation. Launched in early 2026, the LIGAND-AI initiative represents a landmark collaboration in pharmaceutical research, uniting industry leaders such as Pfizer with academic consortia including the Structural Genomics Consortium. This multinational effort, spanning nine countries with 18 institutional partners and funded at over €60 million, embodies a strategic shift toward collective data generation. The initiative's core objective is creating comprehensive, openly accessible datasets mapping molecular interactions between proteins and potential drug compounds across thousands of human targets. This addresses a fundamental bottleneck: artificial intelligence algorithms require massive, high-quality training data that no single organization can feasibly produce alone.

The transformation from traditional closed models to open collaborative ecosystems represents a fundamental shift in pharmaceutical innovation (Figure 1). This model addresses a critical challenge: AI models for drug discovery require vast training datasets that exceed what any single organization can generate. AlphaFold, DeepMind's protein structure prediction system, demonstrated this principle by training on publicly available data and releasing predicted structures for hundreds of millions of proteins, catalyzing thousands of research projects.[3] Yet effectively deploying AI in open data environments requires addressing model robustness when integrating heterogeneous datasets with varying quality standards and formats. Emerging AI paradigms—particularly foundation models and generative systems—are tackling these challenges by integrating diverse data types and enabling better generalization across different experimental platforms and data sources. The evolution from AlphaFold 2's fully open-source model to AlphaFold 3's more restricted access illustrates the ongoing negotiation between openness and commercial viability. Successful consortia navigate this tension by establishing clear boundaries: precompetitive data generation remains open and collaborative, while competitive applications stay proprietary and protected. Initiatives like Open Targets demonstrate this balance, enabling companies to collaborate on target identification while competing on downstream drug development.[4]

BALANCING OPENNESS WITH INCENTIVES

The transition to open science requires rethinking intellectual property frameworks that have governed pharmaceutical innovation for decades. The current patent system creates inherent tensions with openness, yet these need not be mutually exclusive. Several models are emerging to balance collaboration with appropriate incentives while protecting core commercial interests.

Successful consortia address the "free rider" problem through membership structures requiring meaningful contributions—financial, data, or expertise. They create value beyond immediate research outputs, including priority access to shared resources, influence over research directions, and enhanced reputation. The Open Targets platform, integrating genetic, genomic, and chemical data to prioritize therapeutic targets, demonstrates how companies can collaborate on target identification while competing on drug development. This model shows that when stakeholders—academia, industry, government, and patients—align their goals, innovation efficiency can increase substantially.

INFRASTRUCTURE AND CULTURAL TRANSFORMATION

Realizing open science's full potential requires robust infrastructure for data sharing and standardization. Current efforts remain fragmented, with data scattered across incompatible repositories using inconsistent formats. A critical challenge is ensuring data quality: open datasets often suffer from noise, systematic biases, and inconsistencies that can severely compromise AI model performance. Addressing this requires rigorous data curation pipelines, standardized data formats and quality metrics, and community-

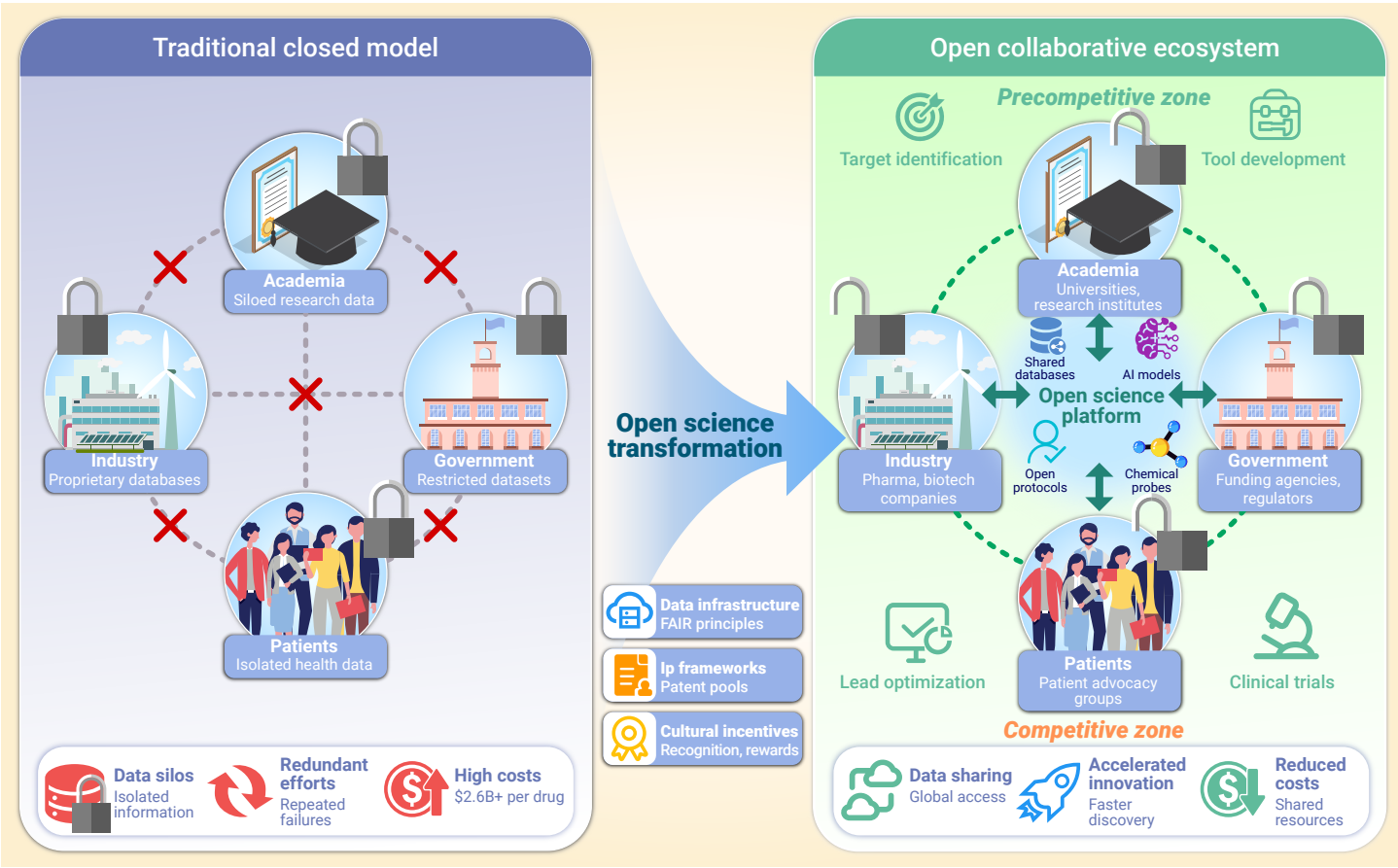

Figure 1. Transformation from closed to open drug discovery ecosystems

driven benchmarking initiatives to establish quality standards. Establishing FAIR (Findable, Accessible, Interoperable, Reusable) data principles as the pharmaceutical research standard is essential. This requires investment in unified data standards, ontologies, and platforms enabling seamless integration of diverse data types—a technical bottleneck that can introduce bias into AI model training if not properly addressed.

Preprint servers and open-access publishing accelerate dissemination, but cultural change is needed to ensure sharing negative results becomes standard practice. The current academic evaluation system still emphasizes traditional journal publications and must be reformed to reward data sharing and collaborative contributions. Funding agencies and institutions must recognize and reward open science contributions in hiring, promotion, and grant decisions. Training the next generation in open science practices—including technical skills and understanding of ethical, legal, and social implications—is equally critical. The evidence increasingly supports openness: genetic evidence supporting drug targets, when openly shared, more than doubles the probability of clinical success.[5]

THE PATH FORWARD

The future of drug discovery lies not in choosing between open and closed models, but in thoughtfully integrating both. Precompetitive research—target identification, tool development, and foundational biology—benefits enormously from openness. Later-stage development, where differentiation and commercial return are critical, may retain traditional approaches. The key is identifying the optimal boundary and creating mechanisms enabling smooth transitions between collaborative and competitive phases, as illustrated in Figure 1.

Policymakers must create regulatory frameworks facilitating data sharing while protecting patient privacy and legitimate commercial interests. Priority actions include: (1) establishing global unified data standards and sharing protocols for drug discovery; (2) promoting "open innovation" contracts that clarify data usage rights and benefit distribution; (3) developing secure data-sharing platforms with robust provenance tracking and access controls. Most importantly, the research community must recognize that in an era of unprecedented complexity, collaboration is not just ethically desirable—it is scientifically essential.

The barriers to drug discovery are formidable, but not insurmountable. By breaking down walls between organizations, disciplines, and sectors, open science and collaborative innovation offer a path toward more efficient, equitable, and effective drug development. The LIGAND-AI initiative and similar efforts demonstrate that this transformation is already underway, though significant challenges remain in balancing openness with commercial incentives, standardizing data infrastructure, and reforming institutional cultures. Success will require sustained commitment from all stakeholders—academia, industry, government, and patients—to prioritize collective progress over individual advantage in precompetitive research while maintaining appropriate incentives for competitive development. The urgency of unmet medical needs demands nothing less.

REFERENCES

1. Zhang, K., Yang, X., Wang, Y., et al. (2025). Artificial intelligence in drug development. *Nat. Med.* **31**:45–59. DOI:10.1038/s41591-024-03434-4
2. The UK Biobank Whole-Genome Sequencing Consortium. (2025). Whole-genome sequencing of 490,640 UK Biobank participants. *Nature,* **645**:692–701. DOI:10.1038/s41586-025-09272-9
3. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature,* **596**:583–589. DOI:10.1038/s41586-021-03819-2
4. Hulcoop, D. G., Trynka, G., & McDonagh, E. M. (2025). Open Targets: 10 years of partnership in target discovery. *Nat. Rev. Drug Discov.* **24**:153–154. DOI:10.1038/d41573-024-00204-2
5. Minikel, E. V., Painter, J. L., Dong, C. C., et al. (2024). Refining the impact of genetic evidence on clinical success. *Nature,* **629**:624–629. DOI:10.1038/s41586-024-07316-0

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.