

International, cross-sector collaboration commits nearly \$2 billion to build foundational data for AI models to predict and treat disease

REDWOOD CITY, CA, October 7, 2026 — Biohub, the U.S. Department of Energy, the National Institutes of Health, and new funding partners today announced a major expansion of an international effort to generate and make accessible the data enabling predictive AI models of biology. Together, the organizations are investing \$1.8 billion in funding, data, computation, and new measurement technology, the largest coordinated commitment to generating AI-ready biological data to date. The result will be an open resource for the research community that provides the foundation for greater understanding and ultimately treatment of human diseases.

As part of this announcement, Biohub has partnered with the Department of Energy (DOE) Office of Science and the National Institutes of Health (NIH) to advance the frontier of artificial intelligence in biology. DOE will invest more than \$500 million over five years in lab measurement, modeling and computation toward the international effort to build an AI-ready open data resource. NIH will coordinate the contribution of relevant datasets, repositories, and knowledge bases developed through more than \$500 million in prior federal investment aligned to this initiative. Biohub will work with NIH to standardize these datasets for AI model training.

In addition, Google DeepMind, Isomorphic Labs, and Meta are collectively investing \$300 million in the [Virtual Biology Initiative](#) to create the technologies and multi-modal datasets needed to build predictive models of life.

These datasets will enable the global scientific community to collectively build and use AI models that allow researchers to ask, predict, and answer biological questions digitally, accelerating the path to new ways of preventing and treating diseases. This initiative will deliver the foundational measurements to train these models, expanding cell response data to interventions across far more cell types and conditions than have yet been studied, and building and validating technologies for studying cells and cellular interactions at greater scale, speed, and accuracy.

“An accurate predictive model of biology could dramatically accelerate scientific discovery by enabling scientists to perform experiments digitally. The insights that come from this could unlock a far greater understanding of disease and open up completely new paths for cures,” said Biohub Head of Science Alex Rives. “Because of this potential, the creation of a virtual cell is one of the most important challenges for the next era of science. It will require coordinated data generation efforts at a national and international scale, which is why these partners are coming together. We invite the worldwide scientific community to join us in this project.”

The Virtual Biology Initiative, announced in April 2026, will coordinate data generation across institutions and disciplines to build AI-ready open datasets to enable predictive models of life. Biohub’s founding \$500 million commitment anchors that work: \$400 million supports new technologies that expand what biologists can measure: cryo-electron tomography, which resolves near-atomic detail inside the cell; microscopy that can image millions to billions of cells in living tissue; and engineering tools to build and perturb biology at molecular, cellular, tissue, and whole-organism levels. A further \$100 million funds research outside Biohub.

“Generating the data to solve predictive systems biology requires scaling past the limits of what any single organization can produce today,” said Max Jaderberg, President of Isomorphic Labs. “By joining the Virtual Biology Initiative as a founding member, Isomorphic Labs is helping build a massive, multimodal data foundation. This initiative will generate the data needed to push the industry closer to the next significant breakthrough for biology.”

Through the Genesis Mission, a cross-agency initiative led by the Department of Energy, DOE will contribute more than \$500 million over five years in fundamental cell research — data collection, AI analytics, measurement and imaging, modeling, and computation — drawing on exascale supercomputing, X-ray and neutron scattering, cryo-electron microscopy and tomography, and autonomous laboratories across the National Laboratory system.

“This partnership represents a critical step forward in leveraging artificial intelligence for public benefit,” said Darío Gil, DOE’s Under Secretary for Science. “By combining DOE’s exascale computing, experimental measurement, and modeling assets, including premier user facilities at the Joint Genome Institute, the Environmental Molecular Sciences Laboratory, and advanced structural beamlines with the unique AI models, tool development, and biological data capabilities of Biohub, we are setting a new standard for open science that will accelerate discoveries in both medicine and biotechnology.”

Through its [Bio Genesis Mission](#), NIH will bring together existing biomedical datasets, national data infrastructure, and research programs to help build AI-ready resources for the broader scientific community. NIH’s extensive investments in biomedical research provide a foundation for this work; resources include national biomedical repositories catalogued by NIH’s National Library of Medicine (NLM) and the [National Center for Biotechnology Information](#), as well as NIH Common Fund [programs](#) that are already developing coordinated biological atlases, shared data standards, and AI-ready biomedical datasets.

“By combining resources and expertise, we can accelerate the development of universal cell models with sufficient biological complexity to predict how any cell responds to an intervention,” said Nicole Kleinstreuer, Ph.D., NIH Deputy Director Program Coordination, Planning, and Strategic Initiatives (DPCPSI). “The return from these models could be broad and profound, resulting in substantially faster timelines for medical breakthroughs as compared with attempting to attain the same results through laboratory experiments alone.”

In addition, leading scientific institutions and consortia with experience in organizing transformative international collaborations, from the Human Genome Project onwards, have come together to help nucleate the scientific community across academia and industry around developing effective scientific strategies to maximize the impact of virtual biology. The groups include the Allen Institute, Broad Institute, Gladstone Institutes, the Human Cell Atlas, the Human Protein Atlas, and the Wellcome Sanger Institute. These groups are committed to working together as part of the Virtual Biology Initiative as well as through independent efforts toward this shared goal. NVIDIA will support the initiative to leverage accelerated computing infrastructure, domain-specific software, and technical expertise. Renaissance Philanthropy is helping to expand funding for data generation.

As the initiative takes shape, Biohub is bringing together partners across disciplines and industries to build the layer that lets their datasets work in a unified fashion — shared standards, common identifiers, and a single point of access. Equally important is building the scientific community around these resources — convening researchers across institutions and disciplines, connecting complementary expertise and capabilities, and creating opportunities to define and pursue ambitious scientific questions together. Over the past decade, Biohub has expanded the reach and impact of measurement technologies and open datasets, leading projects such as [Tabula Sapiens](#), [OpenCell](#), and [Zebrahub](#). It has also built and maintained community data infrastructure, including [CELLxGENE](#) and the [CryoET Data Portal](#). The Virtual Biology Initiative builds on these experiences to enable coordinated efforts at a scale that no single institution could achieve alone.

“The quest to build a virtual cell is one of the great collective scientific challenges and key to understanding the mechanisms of life. We will not solve this challenge without open, experimental biological data at an unprecedented scale, showing how living cells behave and respond to changes,” said Pushmeet Kohli, VP, AI for Science at Google DeepMind and Google Cloud's Chief Scientist. “This investment in biological data generation will help create an open, standardized data commons, which will lay the foundations researchers around the world need to better model biology.”

###

About Biohub

Biohub is a 501(c)(3) nonprofit research institute combining frontier AI and biology to accelerate science. With its compute capacity, AI research and engineering, and state-of-the-art technology for measuring, imaging, and programming biology, Biohub is enabling scientists worldwide to use AI-powered biology to study how cells operate and organize as systems — with the ultimate mission to cure or prevent all disease. Learn more at biohub.org.

Press Contact

Biohub
press@biohub.org