Spotlight on AI and Scientific Discovery

Authors: Rohita Biswas, Cinthya Souza Simas, Sara Tóth Martínez, María Belén Moyano, Felicitas S. Holzer, Gerhard G. Steinmann, Roland Mertelsmann

Submitted: 1. April 2026
Published: 8. April 2026
Volume: 13
Issue: 2

Affiliation: Journal of Science, Humanities and Arts (JOSHA), Freiburg im Breisgau, Germany

Languages: English
Keywords: Foundation Models; Interactome; Dynamics; DNA Repair; Chromatin.
Categories: News and Views
DOI: 10.17160/josha.13.2.1137

Abstract:

This edition explores how artificial intelligence is reshaping scientific discovery across molecular prediction, biological mapping, and model transparency. From structure-based foundation models that predict protein folds, ligand binding, energetics, and conformational change, to proteome-scale inference of human protein-protein interactions, these advances are turning computation into a powerful engine for decoding cellular function. Generative deep learning further expands this shift by emulating protein ensembles and free-energy landscapes at speeds far beyond conventional simulations. Alongside these AI-driven tools, the collection highlights the value of openness and scrutiny in model development, as illustrated by the analysis of DeepSeek R1, and the continued importance of high-resolution experimental resources such as REPAIRome and ExIGS. Together, these studies show that discovery increasingly emerges from the interplay between predictive models, scalable datasets, and experimentally grounded

JOSHA Journal of Science, Humanities and Arts

March 2026 Volume 13, Issue 2

Spotlight on AI and Scientific Discovery

Rohita Biswas, Cinthya Souza Simas, Sara Tóth Martínez, María Belén Moyano, Felicitas Holzer, Gerhard G. Steinmann, Roland Mertelsmann
editorial@josha-archive.org
Journal of Science, Humanities and Arts, Freiburg im Breisgau, Germany

Abstract

This edition explores how artificial intelligence is reshaping scientific discovery across molecular prediction, biological mapping, and model transparency. From structure-based foundation models that predict protein folds, ligand binding, energetics, and conformational change, to proteome-scale inference of human protein-protein interactions, these advances are turning computation into a powerful engine for decoding cellular function. Generative deep learning further expands this shift by emulating protein ensembles and free-energy landscapes at speeds far beyond conventional simulations. Alongside these AI-driven tools, the collection highlights the value of openness and scrutiny in model development, as illustrated by the analysis of DeepSeek R1, and the continued importance of high-resolution experimental resources such as REPAIRome and ExIGS. Together, these studies show that discovery increasingly emerges from the interplay between predictive models, scalable datasets, and experimentally grounded biological insight.

Keywords: Foundation Models; Interactome; Dynamics; DNA Repair; Chromatin.

1. AI to rewire life's interactome: Structural foundation models help to elucidate and reprogram molecular biology

By Zhuoran Qiao

Artificial intelligence (AI) is transforming molecular biology by creating a "computational microscope" that predicts biomolecular structures and interactions faster than traditional experimental methods. Tools like AlphaFold2 use evolutionary data to accurately model protein structures, while Qiao's team developed NeuralPLexer to predict protein-ligand interactions and their dynamic conformations. NeuralPLexer generates multiple conformational snapshots in one step, revealing cryptic binding pockets and ligand-induced structural changes. It also assigns confidence scores to distinguish strong binders from weak ones. Complementing this, OrbNet-Equi models molecular energetics, including protonation and electron states, with near-quantum accuracy at a fraction of the computational cost. These AI tools help elucidate enzyme mechanisms and protein activation, offering insights previously difficult to achieve. Together, these advances pave the way for rewiring cellular signaling, accelerating drug discovery, and enabling rational molecular design.

This article was previously published in *Science Volume 389, Issue 6757,* on July 17, 2025.

Read the full article here

2. Predicting protein-protein interactions in the human proteome

By Jing Zhang *et al*

By integrating deeper evolutionary information from omicMSAs built from unannotated eukaryotic genomic data with RoseTTAFold2-PPI, a deep-learning model trained on a substantially expanded domain-domain interaction dataset derived from 200 million predicted protein structures, the study establishes a scalable framework for proteome-wide prediction of human protein-protein interactions. Applied to 190 million possible human protein pairs, the approach achieved higher performance than ColabFold and AlphaFold3 while using substantially less computational power, producing more than 29,000 predicted

interactions with an estimated precision of 80% and a high-confidence subset of nearly 18,000 interactions, including about 3600 not previously reported. The predictions were enriched for challenging classes such as transmembrane proteins, enabled reconstruction of higher-order assemblies, identified potential new components of known complexes, and mapped disease-associated amino acid variants to interaction interfaces in 4950 predicted protein pairs, offering a useful resource for studying human biology, complex formation, and mechanisms of genetic disease at interactome scale.

This article was previously published in *Science* on October 23, 2025.

Read the full article here

3. Scalable emulation of protein equilibrium ensembles with generative deep learning

By Sarah Lewis *et al*

The article introduces BioEmu, a generative deep learning model designed to rapidly emulate protein equilibrium ensembles and functionally relevant motions with accuracy comparable to long-timescale molecular dynamics simulations, but at vastly reduced computational cost. BioEmu takes sequence–structure representations (derived from AlphaFold) as input to a diffusion-based architecture and is trained in stages on large structural databases, extensive molecular dynamics trajectories, and hundreds of thousands of experimental stability measurements, enabling it to jointly model conformational distributions and thermodynamic properties. It shows that BioEmu can recover diverse conformational changes such as folding/unfolding transitions, large domain motions, and cryptic pocket formation, reproducing equilibrium free-energy landscapes with sub–1 kcal/mol errors while achieving speedups of 4–5 orders of magnitude compared with conventional simulations. Additionally, it demonstrates that the model predicts protein stability and mutational effects with good quantitative accuracy, suggesting it can serve both as a tool for mechanistic interpretation of experiments and as a high-throughput engine for exploring how sequence variation shapes structure, dynamics, and function across many proteins.

This article was previously published in *Science Volume 389, Issue 6761,* on July 10, 2025.

Read the full article here

4. Secrets of DeepSeek AI model revealed in landmark paper

By Elizabeth Gibney

The Chinese company DeepSeek has revealed interesting details about its new AI model, R1. By sharing this information, other scientists can evaluate the system, which researchers consider highly useful. Experts also suggest that U.S. companies should follow this example of transparency. According to the report, this new AI model is much cheaper than its North American counterparts. Training R1 cost only about $294,000, plus roughly $6 million for the base model. R1 was created using reinforcement learning. During training, the model was rewarded for producing correct answers. To further improve performance, the system uses a technique called group relative policy optimization (GRPO), in which the model scores its own attempts using estimates. The model has had a significant impact on the market, inspiring similar initiatives.

This article was previously published in *Nature*, on September 17, 2025.

Read the full article here

5. A comprehensive genetic catalog of human double-strand break repair

By Ernesto López de Alba *et al*

The article introduces REPAIRome, a comprehensive genome-wide catalog generated by combining CRISPR knockout screening with massively parallel indel profiling at Cas9-induced DNA double-strand breaks (DSBs) in human cell lines, revealing how nearly 20,000 genes influence DSB repair outcomes and mutation patterns. It uncovers previously uncharacterized mechanisms, including opposing roles of XLF and PAXX paralogs in controlling DNA end processing (favoring insertions versus deletions), sequential multinucleotide insertions after Cas9 recutting, HLTF-mediated removal of post-cleavage Cas9 impacting repair, SAGA chromatin-remodeling complex involvement in microhomology-mediated end joining (MMEJ), and insertional NHEJ as the basis for the ID11 cancer mutational signature

linked to VHL deficiency and hypoxia. The resource, accessible via a public webtool (https://repairome.bioinfo.cnio.es/), provides insights into nonhomologous end joining (NHEJ) and MMEJ pathways, enhances understanding of CRISPR editing precision, explains cancer genome instability, and serves as a broadly applicable tool for investigating gene-specific effects on DSB repair with implications for therapeutics and genome engineering.

This article was previously published in *Science*, *Volume 390, Issue 6768,* on October 2, 2025.

Read the full article here

6. Expansion in situ genome sequencing links nuclear abnormalities to aberrant chromatin regulation

By Ajay S. Labade *et al*

By combining expansion microscopy with in situ genome sequencing, the study introduces expansion in situ genome sequencing (ExIGS), a platform that enables simultaneous nanoscale imaging of nuclear proteins and three-dimensional mapping of genomic loci within physically expanded nuclei. Validation in human fibroblasts showed that the method improves measurement of DNA-protein interactions at high spatial resolution, and its application to fibroblasts from an individual with Hutchinson-Gilford progeria syndrome revealed that abnormal lamin A/C structures, including invaginations extending into the nuclear interior, accumulate across passages and are associated with localized disruptions in chromatin compartmentalization rather than broad genome-wide disorganization. Paired imaging of lamin A/C and RNA polymerase II further indicated that lamin-associated structures are generally linked to transcriptional repression and vary not only in progeria but also across normal tissue and aging contexts, positioning ExIGS as a broadly useful approach for connecting nuclear morphology with genome regulation across disease and physiology.

This article was previously published in *Science* on July 24, 2025.

Read the full article here

Acknowledgements

GPT-5.2 version of Chat-GPT (Open AI) was used during the writing process as part of JOSHA's policy of experimentation with AI tools. However, JOSHA takes full responsibility for its content.