

Weekly Big Bioinformatics Papers Report

Coverage window: 2026-03-27 to 2026-04-03

Journal scope: Nature; Science; Cell; Nature Biotechnology; Nature Methods; Nature Genetics; Nature Medicine; Nature Machine Intelligence; Nature Computational Science; Cell Systems; Cell Genomics

Executive Summary

This report summarizes 30 recent papers selected from the configured journal scope and ranked for likely bioinformatics relevance, user-interest relevance, and potential field impact.

Impact Score Method

- Impact score is a heuristic weekly triage score, not a citation-based impact metric.
- Score = journal prestige score + generic bioinformatics relevance score × 1.00 + user-interest keyword score × 2.00 + small citation-count bonus + abstract-availability bonus + small bonus for benchmark/resource/atlas-style papers.
- Only papers whose strict publication date falls inside the requested date window are retained. Publication date is taken in this priority order: published-print, then published-online, then published.
- Journal prestige score is manually assigned by journal title.
- Generic bioinformatics relevance is estimated from broad field-related keywords in title, abstract, and subject metadata.
- User-interest keyword relevance is computed from matched user-supplied keywords. Keywords used in this run: None provided.
- If user-interest keywords are supplied, only papers matching at least one user-interest keyword are retained in the report.
- When user-interest keywords are supplied, the report displays only matched user-interest keywords. Built-in generic keywords are displayed only when no user-interest keywords are provided.
- Each paper includes a matched-keywords column so the ranking can be interpreted transparently.
- Fast mode was used: the script first ranked papers using Crossref metadata and only enriched the top candidate set with PubMed abstracts.
- If PubMed enrichment was used, only the top 20 candidates from the first-pass ranking were enriched with PubMed abstracts.
- This score is intended to prioritize likely high-interest papers for manual review.

Top Papers This Week

Rank	Title	Journal	Date	Matched keywords	Impact score
1	CREsted: modeling genomic and synthetic cell-type-specific enhancers across tissues and species	Nature Methods	2026-04-02	foundation model; atlas; single-cell; software; deep learning	24.70
2	Scaling and quantization of large-scale foundation model enables resource-efficient predictions in network biology	Nature Computational Science	2026-03-27	foundation model; resource; single-cell; computational	21.60

Rank	Title	Journal	Date	Matched keywords	Impact score
3	Compressing the collective knowledge of ESM into a single protein language model	Nature Methods	2026-03-30	language model; benchmark; methods; protein structure	21.30
4	DefensePredictor: A machine learning model to discover prokaryotic immune systems	Science	2026-04-02	crispr; language model; machine learning	21.30
5	Targeting modulated vascular smooth muscle cells in atherosclerosis via FAP-directed immunotherapy	Science	2026-04-02	atlas; single-cell; spatial; transcriptomics	20.66
6	General scales unlock AI evaluation with explanatory and predictive power	Nature	2026-04-02	language model; llm; benchmark	20.50
7	Benchmarking alignment methods for spatial transcriptomics data	Nature Computational Science	2026-04-03	benchmark; spatial; methods; transcriptomics	18.90
8	Predicting new research directions in materials science using large language models and concept graphs	Nature Machine Intelligence	2026-04-01	language model; methods; machine learning	17.33
9	Protein and genomic language models uncover the unexplored diversity of bacterial immunity	Science	2026-04-02	language model; machine learning	17.30
10	Unpaired data as a first-order challenge in single-cell and spatial proteomics	Nature Biotechnology	2026-03-27	single-cell; spatial; proteomics	16.70
11	The MicrobeAtlas database: Global trends and insights into Earth's microbial ecosystems	Cell	2026-04-01	atlas; database	15.93
12	Scalable single-cell total RNA sequencing unifies coding and noncoding transcriptomics	Nature Biotechnology	2026-03-31	single-cell; transcriptomics	14.60
13	Mapping cis-regulatory mutations at scale in sorghum enables modulation of gene expression	Nature Biotechnology	2026-03-27	crispr	14.50
14	Functional RNA splitting drove the evolutionary emergence of type V CRISPR-Cas systems from transposons	Cell	2026-04-01	crispr	14.00
15	Atomic-resolution imaging of gold species at organic liquid-solid interfaces	Science	2026-04-02	deep learning	13.80
16	AlphaFold as a prior: experimental structure determination conditioned on a pretrained neural network	Nature Methods	2026-04-01	alphafold	13.50
17	Senescence in cancer: Hallmarks, paradoxes, and therapeutic promise	Cell	2026-04-01	llm	13.00
18	Assessing uncertainty of sequence representations generated by protein language models	Nature Methods	2026-04-01	language model	13.00
19	Bridging the gap between hybrid and sequence-only protein language models	Nature Methods	2026-03-30	language model	13.00
20	Exome sequencing and analysis of 44,028 British South Asians enriched for high autozygosity	Nature Genetics	2026-03-27	resource	12.40

Rank	Title	Journal	Date	Matched keywords	Impact score
21	Systematic analysis of snRNA genes reveals frequent RNU2-2 variants in dominant and recessive developmental and epileptic encephalopathies	Nature Genetics	2026-03-30	transcriptomics	12.30
22	An atlas to navigate environmental factors and health	Nature Medicine	2026-03-30	atlas	12.20
23	Large-scale proteomics across neurological disorders uncovers biomarker panel and targets in multiple sclerosis	Cell	2026-04-01	proteomics	12.10
24	HorusEye: a self-supervised foundation model for generalizable X-ray tomography restoration	Nature Computational Science	2026-03-27	foundation model	11.90
25	A deep joint-learning proteomics model for diagnosis of six conditions associated with dementia	Nature Medicine	2026-03-31	proteomics	11.90
26	Quemliclustat and chemotherapy with or without zimberelimab in metastatic pancreatic adenocarcinoma: a randomized phase 1 trial	Nature Medicine	2026-03-30	spatial	11.90
27	3d-OT: a deep geometry-aware framework for heterogeneous slices alignment of spatial multi-omics	Nature Methods	2026-03-27	spatial	11.60
28	Identifying maximally informative signal-aware representations of single-cell data using the information bottleneck	Cell Systems	2026-04-01	single-cell	11.40
29	Single-cell morphodynamical trajectories enable prediction of gene expression accompanying cell state change	Cell Systems	2026-04-01	single-cell	11.40
30	The potential of wheat spatial omics	Nature Genetics	2026-04-01	spatial	11.30

1. CREsted: modeling genomic and synthetic cell-type-specific enhancers across tissues and species

Journal: Nature Methods **Date:** 2026-04-02 **DOI:** 10.1038/s41592-026-03057-2

Authors: Niklas Kempynck; Seppe De Winter; Casper H. Blaauw; Vasileios Konstantakos; Eren Can Ekici; Sam Dieltiens; Darina Abaffyová; Valérie Bercier; Ibrahim I. Taskiran; Gert Hulselmans; Katina Spanier; Valerie Christiaens

Matched generic keywords: foundation model; atlas; single-cell; software; deep learning

Impact score: 24.70 (journal=9.50, generic=14.20, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Sequence-based deep learning models have become the state of the art for analyzing the genomic regulatory code.

Major discoveries: Particularly for enhancers, these models excel at deciphering sequence grammar that underlies their activity. To enable end-to-end enhancer modeling and design, we developed a software package called CREsted (cis-regulatory element sequence training, explanation and design).

Likely impact: Matched keywords: foundation model; atlas; single-cell; software; deep learning. Likely relevant to single-cell analysis workflows. Likely important for AI-driven bioinformatics. Builds an atlas-scale reference or dataset. Likely offers methodological or software impact.

Abstract excerpt: Sequence-based deep learning models have become the state of the art for analyzing the genomic regulatory code. Particularly for enhancers, these models excel at deciphering sequence grammar that underlies their activity. To enable end-to-end enhancer modeling and design, we developed a software package called CREsted (cis-regulatory element sequence training, explanation and design). It combines preprocessing and analysis of single-cell assay for transposase-accessible chromatin using sequencing data, modeling chromatin accessibility from sequence, sequence design and downstream analysis to decipher enhancer grammar.

2. Scaling and quantization of large-scale foundation model enables resource-efficient predictions in network biology

Journal: Nature Computational Science **Date:** 2026-03-27 **DOI:** 10.1038/s43588-026-00972-4

Authors: Han Chen; Madhavan S. Venkatesh; Javier Gómez Ortega; Siddharth V. Mahesh; Tarak N. Nandi; Ravi K. Madduri; Karin Pelka; Christina V. Theodoris

Matched generic keywords: foundation model; resource; single-cell; computational

Impact score: 21.60 (journal=8.40, generic=11.20, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=1.00)

Why it matters: Foundation models for network biology are pretrained on large-scale biological data to enable context-aware predictions in a diverse array of downstream tasks through transfer learning.

Major discoveries: However, increasing model sizes with the expansion of available pretraining data also increases the computational resources required for fine-tuning and inference in downstream applications. Here we first assemble a corpus comprising ~104 million human single-cell transcriptomes from a broad range of tissues and diseases and pretrain successively larger models, defining the scaling laws for transcriptional masked learning.

Likely impact: Matched keywords: foundation model; resource; single-cell; computational. Introduces a reusable community resource. Likely relevant to single-cell analysis workflows. Likely important for AI-driven bioinformatics. Likely offers methodological or software impact.

Abstract excerpt: Foundation models for network biology are pretrained on large-scale biological data to enable context-aware predictions in a diverse array of downstream tasks through transfer learning. However, increasing model sizes with the expansion of available pretraining data also increases the computational resources required for fine-tuning and inference in downstream applications. Here we first assemble a corpus comprising ~104 million human single-cell transcriptomes from a broad range of tissues and diseases and pretrain successively larger models, defining the scaling laws for transcriptional masked learning. We then demonstrate that model quantization preserves the contextual gene and cell embedding space of the full-precision model, matching performance in zero-shot and fine-tuning applications while requiring

only 15% of the time and 34% of the memory as the full model for fine-tuning with the same batch size.

3. Compressing the collective knowledge of ESM into a single protein language model

Journal: Nature Methods **Date:** 2026-03-30 **DOI:** 10.1038/s41592-026-03050-9

Authors: Tuan Dinh; Seon-Kyeong Jang; Noah Zaitlen; Vasilis Ntranos

Matched generic keywords: language model; benchmark; methods; protein structure

Impact score: 21.30 (journal=9.50, generic=10.80, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Protein language models (PLMs) have recently emerged as a promising approach for next-generation variant-effect prediction (VEP).

Major discoveries: Most high-performing VEP methods currently utilize PLMs combined with additional information, such as homology, protein structure and population genetics data to improve prediction accuracy. This performance gain, however, comes with added complexity or limited applicability compared to pure PLMs trained only on raw, unaligned sequences, such as evolutionary scale modeling (ESM).

Likely impact: Matched keywords: language model; benchmark; methods; protein structure. Provides a benchmark or evaluation framework. Likely important for AI-driven bioinformatics. Likely offers methodological or software impact. Potentially important for structural bioinformatics.

Abstract excerpt: Protein language models (PLMs) have recently emerged as a promising approach for next-generation variant-effect prediction (VEP). Most high-performing VEP methods currently utilize PLMs combined with additional information, such as homology, protein structure and population genetics data to improve prediction accuracy. This performance gain, however, comes with added complexity or limited applicability compared to pure PLMs trained only on raw, unaligned sequences, such as evolutionary scale modeling (ESM). Here we challenge the prevailing view that sequence-only PLMs are intrinsically limited and present an efficient co-distillation approach to adapt them for high-accuracy VEP without requiring additional information beyond evolutionary signals captured during pretraining.

4. DefensePredictor: A machine learning model to discover prokaryotic immune systems

Journal: Science **Date:** 2026-04-02 **DOI:** 10.1126/science.adv7924

Authors: Peter C. DeWeirdt; Emily M. Mahoney; Michael T. Laub

Matched generic keywords: crispr; language model; machine learning

Impact score: 21.30 (journal=10.00, generic=10.30, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Antiphage defense systems protect bacteria from viral infection and have inspired important biotechnologies such as CRISPR-Cas9 while also revealing the evolutionary roots of eukaryotic innate immunity.

Major discoveries: Many systems have been discovered by genomic colocalization, but this approach cannot identify systems outside of defense islands. We present DefensePredictor, a machine learning model that uses protein language model embeddings to classify proteins as defensive.

Likely impact: Matched keywords: crispr; language model; machine learning. Likely important for AI-driven bioinformatics.

Abstract excerpt: Antiphage defense systems protect bacteria from viral infection and have inspired important biotechnologies such as CRISPR-Cas9 while also revealing the evolutionary roots of eukaryotic innate immunity. Many systems have been discovered by genomic colocalization, but this approach cannot identify systems outside of defense islands. We present DefensePredictor, a machine learning model that uses protein language model embeddings to classify proteins as defensive. Applying DefensePredictor to 69 diverse *Escherichia coli* strains, we predicted hundreds of previously unknown systems and experimentally validated 42 of them.

5. Targeting modulated vascular smooth muscle cells in atherosclerosis via FAP-directed immunotherapy

Journal: Science **Date:** 2026-04-02 **DOI:** 10.1126/science.adx1736

Authors: Junedh M. Amrute; In-Hyuk Jung; Tracy Yamawaki; Wen-Ling Lin; Andrea Bredemeyer; Johanna Diekmann; Sikander Hayat; Xianglong Zhang; Devin L. Wakefield; Xin Luo; Sidrah Maryam; Gyu Seong Heo

Matched generic keywords: atlas; single-cell; spatial; transcriptomics

Impact score: 20.66 (journal=10.00, generic=9.60, interest=0.00, citation bonus=0.06, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Vascular smooth muscle cell (VSMC) diversification drives atherosclerotic coronary artery disease (CAD), but the mechanisms governing these cell state transitions remain unclear.

Major discoveries: We applied multiomic single-cell profiling, epitope mapping, and spatial transcriptomics across 27 human coronary arteries, identifying fibroblast activation protein (FAP) as a marker of modulated VSMCs. Lineage tracing in mice indicated that FAP + cells originate from Myh11 + VSMCs, and FAP positron emission tomography imaging in CAD patients showed plaque uptake.

Likely impact: Matched keywords: atlas; single-cell; spatial; transcriptomics. Likely relevant to single-cell analysis workflows. Builds an atlas-scale reference or dataset. Potentially important for spatial omics analysis.

Abstract excerpt: Vascular smooth muscle cell (VSMC) diversification drives atherosclerotic coronary artery disease (CAD), but the mechanisms governing these cell state transitions remain unclear. We applied multiomic single-cell profiling, epitope mapping, and spatial transcriptomics across 27 human coronary arteries, identifying fibroblast activation protein (FAP) as a marker of modulated VSMCs. Lineage tracing in mice indicated that FAP + cells originate from Myh11 + VSMCs, and FAP positron emission tomography imaging in CAD patients showed plaque uptake. FAP + cell states resided in the macrophage-rich neo-intima.

6. General scales unlock AI evaluation with explanatory and predictive power

Journal: Nature **Date:** 2026-04-02 **DOI:** 10.1038/s41586-026-10303-2

Authors: Lexin Zhou; Lorenzo Pacchiardi; Fernando Martínez-Plumed; Katherine M. Collins; Yael Moros-Daval; Seraphina Zhang; Qinlin Zhao; Yitian Huang; Luning Sun; Jonathan E. Prunty; Zongqian Li; Pablo Sánchez-García

Matched generic keywords: language model; llm; benchmark

Impact score: 20.50 (journal=10.00, generic=9.50, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Ensuring safe and effective use of artificial intelligence (AI) requires understanding and anticipating its performance on new tasks, from advanced scientific challenges to transformed workplace activities 1–3 .

Major discoveries: So far, benchmarking has guided progress in AI but has offered limited explanatory and predictive power for general-purpose AI systems 4–8 , attributed to limited transferability across specific tasks 9–11 . Here we introduce general scales for AI evaluation that elicit demand profiles explaining what capabilities common AI benchmarks truly measure, extract ability profiles quantifying the general strengths and limits of AI systems and robustly predict AI performance for new task instances.

Likely impact: Matched keywords: language model; llm; benchmark. Provides a benchmark or evaluation framework. Likely important for AI-driven bioinformatics. Likely offers methodological or software impact.

Abstract excerpt: Abstract Ensuring safe and effective use of artificial intelligence (AI) requires understanding and anticipating its performance on new tasks, from advanced scientific challenges to transformed workplace activities 1–3 . So far, benchmarking has guided progress in AI but has offered limited explanatory and predictive power for general-purpose AI systems 4–8 , attributed to limited transferability across specific tasks 9–11 . Here we introduce general scales for AI evaluation that elicit demand profiles explaining what capabilities common AI benchmarks truly measure, extract ability profiles quantifying the general strengths and limits of AI systems and robustly predict AI performance for new task instances. Our fully automated methodology builds on 18 rubrics, capturing a broad range of cognitive and intellectual demands, which

place different task instances on the same general scales, illustrated on 15 large language models (LLMs) and 63 tasks.

7. Benchmarking alignment methods for spatial transcriptomics data

Journal: Nature Computational Science **Date:** 2026-04-03 **DOI:** 10.1038/s43588-026-00977-z

Authors: Yunzhi Yan; Tianyi Gu; Chengcheng Sun; Yinghao Zhang; Yan Cui; Senlin Lin; Qi Zou; Yixuan Du; Chuangyi Han; Kairan Kang; Sijie Li; Yihan Zhao

Matched generic keywords: benchmark; spatial; methods; transcriptomics

Impact score: 18.90 (journal=8.40, generic=9.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=1.50)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Computational Science.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: benchmark; spatial; methods; transcriptomics. Provides a benchmark or evaluation framework. Likely offers methodological or software impact. Potentially important for spatial omics analysis.

8. Predicting new research directions in materials science using large language models and concept graphs

Journal: Nature Machine Intelligence **Date:** 2026-04-01 **DOI:** 10.1038/s42256-026-01206-y

Authors: Thomas Marwitz; Alexander Colsmann; Ben Breitung; Christoph Brabec; Christoph Kirchlechner; Eva Blasco; Gabriel Cadilha Marques; Horst Hahn; Michael Hirtz; Pavel A. Levkin; Yolita M. Eggeler; Tobias Schlöder

Matched generic keywords: language model; methods; machine learning

Impact score: 17.33 (journal=8.20, generic=8.10, interest=0.00, citation bonus=0.03, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Due to an exponential increase in published research articles, it is impossible for individual scientists to read all publications, even within their own research field.

Major discoveries: Here we investigate the use of large language models to extract the main concepts and semantic information from scientific abstracts in the domain of materials science to identify links that were not noticed by humans and to suggest inspiring near and/or mid-term future research directions. We show that large language models can extract concepts more efficiently than automated keyword extraction methods to build a concept graph as an abstraction of the scientific literature.

Likely impact: Matched keywords: language model; methods; machine learning. Likely important for AI-driven bioinformatics. Likely offers methodological or software impact.

Abstract excerpt: Abstract Due to an exponential increase in published research articles, it is impossible for individual scientists to read all publications, even within their own research field. Here we investigate the use of large language models to extract the main concepts and semantic information from scientific abstracts in the domain of materials science to identify links that were not noticed by humans and to suggest inspiring near and/or mid-term future research directions. We show that large language models can extract concepts more efficiently than automated keyword extraction methods to build a concept graph as an abstraction of the scientific literature. A machine learning model is trained to predict emerging combinations of concepts, that is, new research ideas, based on historical data.

9. Protein and genomic language models uncover the unexplored diversity of bacterial immunity

Journal: Science **Date:** 2026-04-02 **DOI:** 10.1126/science.adv8275

Authors: Ernest Mordret; Alexandre Hervé; Florian Tesson; Hugo Vaysset; Tyler Clabby; Arthur Loubat; Helena Shomar; Remi Planel; Rachel Lavenir; Jean Cury; Aude Bernheim

Matched generic keywords: language model; machine learning

Impact score: 17.30 (journal=10.00, generic=6.30, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: The bacterial pangenome contains a vast diversity of antiphage systems, whose overall extent is still unknown.

Major discoveries: In this study, we developed complementary machine learning approaches to systematically predict antiphage function from genomic context, protein sequence, or their combination, achieving up to 99% precision and 92% recall. We validated these models experimentally in *Escherichia* and *Streptomyces* with the discovery of 12 antiphage systems.

Likely impact: Matched keywords: language model; machine learning. Likely important for AI-driven bioinformatics.

Abstract excerpt: The bacterial pangenome contains a vast diversity of antiphage systems, whose overall extent is still unknown. In this study, we developed complementary machine learning approaches to systematically predict antiphage function from genomic context, protein sequence, or their combination, achieving up to 99% precision and 92% recall. We validated these models experimentally in *Escherichia* and *Streptomyces* with the discovery of 12 antiphage systems. Applied to over 32,000 bacterial genomes, these models expand the predicted antiphage repertoire, with ~1.5% of bacterial genomes devoted to defense and more than 85% of predicted protein families remaining uncharacterized.

10. Unpaired data as a first-order challenge in single-cell and spatial proteomics

Journal: Nature Biotechnology **Date:** 2026-03-27 **DOI:** 10.1038/s41587-026-03074-8

Authors: Mikaela Koutrouli; Francesca Finotello; Erwin M. Schoof; Panos Roussos; Wouter-Michiël Vierdag; Lucas Diedrich; Matthias Meyer-Bender; Malte Kuehl; Jose Nimo; Harald Vöhringer; Theodoros Visvikis; Vincenth Brennstainer

Matched generic keywords: single-cell; spatial; proteomics

Impact score: 16.70 (journal=9.50, generic=7.20, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Biotechnology.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: single-cell; spatial; proteomics. Likely relevant to single-cell analysis workflows. Potentially important for spatial omics analysis.

11. The MicrobeAtlas database: Global trends and insights into Earth's microbial ecosystems

Journal: Cell **Date:** 2026-04-01 **DOI:** 10.1016/j.cell.2026.01.021

Authors: João Frederico Matias Rodrigues; Janko Tackmann; Lukas Malfertheiner; David Patsch; Eugenio Perez-Molphe-Montoya; Nicolas Näpflin; Daniela Gaio; Gregor Rot; Mihai Danaila; Matteo Eustachio Peluso; Marija Dmitrijeva; Thomas Sebastian Benedikt Schmidt

Matched generic keywords: atlas; database

Impact score: 15.93 (journal=10.00, generic=4.90, interest=0.00, citation bonus=0.03, abstract bonus=0.00, extra bonus=1.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: atlas; database. Introduces a reusable community resource. Builds an atlas-scale reference or dataset.

12. Scalable single-cell total RNA sequencing unifies coding and noncoding transcriptomics

Journal: Nature Biotechnology **Date:** 2026-03-31 **DOI:** 10.1038/s41587-026-03068-6

Authors: Alina Isakova; Daniel Dan Liu; Ivana Cvijović; Rahul Sinha; Anna E. Eastman; Sirle Saul; Angela M. Detweiler; Norma Neff; Shirit Einav; Irving L. Weissman; Stephen R. Quake

Matched generic keywords: single-cell; transcriptomics

Impact score: 14.60 (journal=9.50, generic=5.10, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Biotechnology.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: single-cell; transcriptomics. Likely relevant to single-cell analysis workflows.

13. Mapping cis-regulatory mutations at scale in sorghum enables modulation of gene expression

Journal: Nature Biotechnology **Date:** 2026-03-27 **DOI:** 10.1038/s41587-026-03046-y

Authors: Evan D. Groover; David Ding; Flora Z. Wang; Gonzalo Benegas; Joseph Rivera; Shahar Schwartz; Stephen Chen; Michael F. Moubarak; Viktoriya Georgieva; Peggy G. Lemaux; Brian J. Staskawicz; Krishna K. Niyogi

Matched generic keywords: crispr

Impact score: 14.50 (journal=9.50, generic=4.00, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Precise modulation of gene expression through cis -regulatory editing holds promise for nontransgenic crop improvement.

Major discoveries: However, the sequence-to-function relationships that govern plant promoter activity remain poorly understood. Here we develop a massively parallel reporter assay in Sorghum bicolor to systematically measure the effects of >30,000 mutations spanning deletions, substitutions and motif insertions accessible through CRISPR editing across entire native promoters and 5' untranslated regions of 3 photosynthesis genes: PsbS , Raf1 and SBPase .

Likely impact: Matched keywords: crispr.

Abstract excerpt: Abstract Precise modulation of gene expression through cis -regulatory editing holds promise for nontransgenic crop improvement. However, the sequence-to-function relationships that govern plant promoter activity remain poorly understood. Here we develop a massively parallel reporter assay in Sorghum bicolor to systematically measure the effects of >30,000 mutations spanning deletions, substitutions and motif insertions accessible through CRISPR editing across entire native promoters and 5' untranslated regions of 3 photosynthesis genes: PsbS , Raf1 and SBPase . We find that gene expression is most tunable within a ~500-bp core promoter region.

14. Functional RNA splitting drove the evolutionary emergence of type V CRISPR-Cas systems from transposons

Journal: Cell **Date:** 2026-04-01 **DOI:** 10.1016/j.cell.2026.03.014

Authors: Shuai Jin; Zixu Zhu; Yunjia Li; Shouyue Zhang; Yijing Liu; Danyuan Li; Yuanqing Li; Yingfeng Luo; Zhiheng Cheng; Kevin Tianmeng Zhao; Qiang Gao; Guanglei Yang

Matched generic keywords: crispr

Impact score: 14.00 (journal=10.00, generic=4.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: crispr.

15. Atomic-resolution imaging of gold species at organic liquid-solid interfaces

Journal: Science **Date:** 2026-04-02 **DOI:** 10.1126/science.adw2469

Authors: Sam Sullivan-Allsop; Nick Clark; Wendong Wang; Rongsheng Cai; William Thornley; David G. Hopkinson; James G. McHugh; Ben Davies; Samuel Pattisson; Nicholas F. Dummer; Rui Zhang; Matthew Lindley

Matched generic keywords: deep learning

Impact score: 13.80 (journal=10.00, generic=2.80, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: The structure and dynamics of adsorbed atoms (adatoms) at solid-liquid interfaces determine the performance of advanced catalysts, electrochemical devices, molecular separation technologies, and metal extraction from waste streams.

Major discoveries: However, in situ investigations of atomically dispersed metals in various chemical environments have been prevented by insufficient imaging resolution and solvent incompatibility. In this study, we combined a specimen design that provides atomic resolution in liquid-phase electron microscopy with deep learning-enabled analysis to explore the interactions between gold adatoms, graphite support, and the solvent collectively.

Likely impact: Matched keywords: deep learning.

Abstract excerpt: The structure and dynamics of adsorbed atoms (adatoms) at solid-liquid interfaces determine the performance of advanced catalysts, electrochemical devices, molecular separation technologies, and metal extraction from waste streams. However, in situ investigations of atomically dispersed metals in various chemical environments have been prevented by insufficient imaging resolution and solvent incompatibility. In this study, we combined a specimen design that provides atomic resolution in liquid-phase electron microscopy with deep learning-enabled analysis to explore the interactions between gold adatoms, graphite support, and the solvent collectively. We tracked the locations of >106 graphite-supported gold adatoms, dimers, and larger clusters in five solvents.

16. AlphaFold as a prior: experimental structure determination conditioned on a pretrained neural network

Journal: Nature Methods **Date:** 2026-04-01 **DOI:** 10.1038/s41592-026-03047-4

Authors: Alisia Fadini; Minhuan Li; Airlie J. McCoy; Suresh Banjara; Hiroki Okumura; Eve Napier; Pietro Fontana; Amir R. Khan; Luca Jovine; Thomas C. Terwilliger; Randy J. Read; Doeke R. Hekstra

Matched generic keywords: alphafold

Impact score: 13.50 (journal=9.50, generic=4.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Methods.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: alphafold. Potentially important for structural bioinformatics.

17. Senescence in cancer: Hallmarks, paradoxes, and therapeutic promise

Journal: Cell **Date:** 2026-04-01 **DOI:** 10.1016/j.cell.2026.03.005

Authors: Clemens Hinterleitner; Hailey V. Goldberg; Domhnall McHugh; Valentin J.A. Barthet; Aveline Filliol; Scott W. Lowe

Matched generic keywords: llm

Impact score: 13.00 (journal=10.00, generic=3.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: llm. Likely important for AI-driven bioinformatics.

18. Assessing uncertainty of sequence representations generated by protein language models

Journal: Nature Methods **Date:** 2026-04-01 **DOI:** 10.1038/s41592-026-03027-8

Matched generic keywords: language model

Impact score: 13.00 (journal=9.50, generic=3.50, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Methods.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: language model. Likely important for AI-driven bioinformatics.

19. Bridging the gap between hybrid and sequence-only protein language models

Journal: Nature Methods **Date:** 2026-03-30 **DOI:** 10.1038/s41592-026-03049-2

Matched generic keywords: language model

Impact score: 13.00 (journal=9.50, generic=3.50, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Methods.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: language model. Likely important for AI-driven bioinformatics.

20. Exome sequencing and analysis of 44,028 British South Asians enriched for high autozygosity

Journal: Nature Genetics **Date:** 2026-03-27 **DOI:** 10.1038/s41588-026-02553-7

Authors: Hye In Kim; Christopher DeBoever; Klaudia Walter; Georgios Kalantzis; Chen Li; Sahar V. Mozaffari; Kousik Kundu; Benjamin M. Jacobs; Pedrum Mohammadi-Shemirani; Anthony M. Musolf; Jonathan M. Davitte; Melis A. Aksit

Matched generic keywords: resource

Impact score: 12.40 (journal=9.20, generic=2.20, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Genes & Health (G&H) is a biomedical study of adult British Pakistani and Bangladeshi research volunteers enriched for autozygosity.

Major discoveries: Here we performed whole-exome sequencing in 44,028 G&H participants, establishing a large publicly available South Asian exome resource linked to longitudinal electronic health records. We performed exome-wide association analyses for 645 electronic health record-derived traits under additive and recessive models, and meta-analyses of 33 cardiometabolic traits with UK Biobank, finding more than 100 novel gene-phenotype associations.

Likely impact: Matched keywords: resource. Introduces a reusable community resource.

Abstract excerpt: Genes & Health (G&H) is a biomedical study of adult British Pakistani and Bangladeshi research volunteers enriched for autozygosity. Here we performed whole-exome sequencing in 44,028 G&H participants, establishing a large publicly available South Asian exome resource linked to longitudinal electronic health records. We performed exome-wide association analyses for 645 electronic health record-derived traits under additive and recessive models, and meta-analyses of 33 cardiometabolic traits with UK Biobank, finding more than 100 novel gene-phenotype associations. We identified 2,991 genes with rare biallelic predicted loss-of-function ("knockout") genotypes, 546 of which had not been previously reported.

21. Systematic analysis of snRNA genes reveals frequent RNU2-2 variants in dominant and recessive developmental and epileptic encephalopathies

Journal: Nature Genetics **Date:** 2026-03-30 **DOI:** 10.1038/s41588-026-02547-5

Authors: Elsa Leitão; Amandine Santini; Benjamin Cogne; Miriam Essid; Maria Athanasiadou; Christy W. LaFlamme; Pierre Marijon; Virginie Bernard; Kevin Jousselin; Nicolas Chatron; Giulia Barcia; Boris Keren

Matched generic keywords: transcriptomics

Impact score: 12.30 (journal=9.20, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Small nuclear RNAs (snRNAs) are essential components of the spliceosome.

Major discoveries: De novo variants in snRNA genes RNU4-2 (ReNU syndrome), RNU5B-1 and RNU2-2 have been linked to dominant neurodevelopmental disorders (NDDs), revealing a large unexpected contribution of noncoding RNA genes to genetic diseases. Here, through international collaborations, we analyze systematically 200 potentially functional snRNA genes in a French cohort of 34,329 people with rare disorders.

Likely impact: Matched keywords: transcriptomics.

Abstract excerpt: Abstract Small nuclear RNAs (snRNAs) are essential components of the spliceosome. De novo variants in snRNA genes RNU4-2 (ReNU syndrome), RNU5B-1 and RNU2-2 have been linked to dominant neurodevelopmental disorders (NDDs), revealing a large unexpected contribution of noncoding RNA genes to genetic diseases. Here, through international collaborations, we analyze systematically 200 potentially functional snRNA genes in a French cohort of 34,329 people with rare disorders. We report RNU2-2 variants in 141 individuals, including 35 with recurrent dominant pathogenic variants and 91 affected members from 73 families with biallelic variants.

22. An atlas to navigate environmental factors and health

Journal: Nature Medicine **Date:** 2026-03-30 **DOI:** 10.1038/s41591-026-04286-w

Matched generic keywords: atlas

Impact score: 12.20 (journal=8.80, generic=2.40, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=1.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Medicine.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: atlas. Builds an atlas-scale reference or dataset.

23. Large-scale proteomics across neurological disorders uncovers biomarker panel and targets in multiple sclerosis

Journal: Cell **Date:** 2026-04-01 **DOI:** 10.1016/j.cell.2026.01.017

Authors: Jakob Maximilian Bader; Christine Makarov; Sabrina Richter; Maximilian Thomas Strauss; Friederike Held; Maria Wahle; Michael Baggio Lorenz; Lara Pöschl; Patricia Skowronek; Marvin Thielert; Achim Berthele; Wen-Feng Zeng

Matched generic keywords: proteomics

Impact score: 12.10 (journal=10.00, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: proteomics.

24. HorusEye: a self-supervised foundation model for generalizable X-ray tomography restoration

Journal: Nature Computational Science **Date:** 2026-03-27 **DOI:** 10.1038/s43588-026-00973-3

Authors: Yuetan Chu; Longxi Zhou; Gongning Luo; Kai Kang; Suyu Dong; Zhongyi Han; Lianming Wu; Xianglin Meng; Changchun Yang; Xin Guo; Yuan Cheng; Yuan Qi

Matched generic keywords: foundation model

Impact score: 11.90 (journal=8.40, generic=3.50, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Computational Science.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: foundation model. Likely important for AI-driven bioinformatics.

25. A deep joint-learning proteomics model for diagnosis of six conditions associated with dementia

Journal: Nature Medicine **Date:** 2026-03-31 **DOI:** 10.1038/s41591-026-04303-y

Authors: Lijun An; Alexa Pichet Binette; Ines Hristovska; Gabriele Vilkaite; Yu Xiao; Romina Zendehdel; Zijian Dong; Bart Smets; Rowan Saloner; Shinya Tasaki; Ying Xu; Varsha Krish

Matched generic keywords: proteomics

Impact score: 11.90 (journal=8.80, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Co-pathology is a common feature of neurodegenerative diseases that complicates diagnosis, treatment and clinical management.

Major discoveries: However, sensitive, specific and scalable biomarkers for in vivo pathological diagnosis are not available for most neurodegenerative neuropathologies. Here we present Proteomics-based Artificial Intelligence for Dementia Diagnosis (ProtAIDe-Dx), a deep joint-learning model on 17,187 patients and controls (age of 70.3 ± 11.5 years, 53.2% female), that uses plasma proteomics to provide simultaneous probabilistic diagnosis across 6 conditions associated with dementia in aging.

Likely impact: Matched keywords: proteomics.

Abstract excerpt: Abstract Co-pathology is a common feature of neurodegenerative diseases that complicates diagnosis, treatment and clinical management. However, sensitive, specific and scalable biomarkers for in vivo pathological diagnosis are not available for most neurodegenerative neuropathologies. Here we present Proteomics-based Artificial Intelligence for Dementia Diagnosis (ProtAIDe-Dx), a deep joint-learning model on 17,187 patients and controls (age of 70.3 ± 11.5 years, 53.2% female), that uses plasma proteomics to provide simultaneous probabilistic diagnosis across 6 conditions associated with dementia in aging. ProtAIDe-Dx achieves cross-validated balanced classification accuracy of 70–95% and area under the curve of >78% across all conditions.

26. Quemliclustat and chemotherapy with or without zimberelimab in metastatic pancreatic adenocarcinoma: a randomized phase 1 trial

Journal: Nature Medicine **Date:** 2026-03-30 **DOI:** 10.1038/s41591-026-04283-z

Authors: Zev A. Wainberg; Gulam A. Manji; Nathan Bahary; Susanna V. Ulahannan; Shubham Pant; David R. Spiegel; Nataliya V. Uboha; Paul E. Oberstein; Anwaar Saeed; Brandon Beagle; Ji Yun Kim; Ning Wang

Matched generic keywords: spatial

Impact score: 11.90 (journal=8.80, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=1.00, extra bonus=0.00)

Why it matters: Abstract Quemliclustat potentially inhibits CD73, a key enzyme producing immunosuppressive adenosine.

Major discoveries: In a phase 1b trial (ARC-8), we evaluated safety and efficacy of quemliclustat combined with gemcitabine/nab-paclitaxel (G/nP) with or without zimberelimab (anti-programmed cell death protein 1 (PD-1)) in first-line metastatic pancreatic ductal adenocarcinoma (PDAC). During the dose-escalation phase, 22 patients were enrolled across five dose levels of quemliclustat (25 mg, 50 mg, 75 mg, 100 mg or 125 mg) with G/nP + zimberelimab.

Likely impact: Matched keywords: spatial. Potentially important for spatial omics analysis.

Abstract excerpt: Abstract Quemliclustat potentially inhibits CD73, a key enzyme producing immunosuppressive adenosine. In a phase 1b trial (ARC-8), we evaluated safety and efficacy of quemliclustat combined with gemcitabine/nab-paclitaxel (G/nP) with or without zimberelimab (anti-programmed cell death protein 1 (PD-1)) in first-line metastatic pancreatic ductal adenocarcinoma (PDAC). During the dose-escalation phase, 22 patients were enrolled across five dose levels of quemliclustat (25 mg, 50 mg, 75 mg, 100 mg or 125 mg) with G/nP + zimberelimab. During the dose-expansion phase, 116 patients were enrolled, beginning with a single-arm, non-randomized cohort receiving quemliclustat 100 mg + G/nP + zimberelimab, followed by a randomized cohort in which patients were assigned in a 2:1 ratio to receive quemliclustat 100 mg + G/nP with or without zimberelimab.

27. 3d-OT: a deep geometry-aware framework for heterogeneous slices alignment of spatial multi-omics

Journal: Nature Methods **Date:** 2026-03-27 **DOI:** 10.1038/s41592-026-03034-9

Authors: Bingjie Dai; Litai Yi; Peizhuo Wang; Hanshuang Li; Pengwei Hu; Yancheng Song; Jixiang Xing; Zhenxing Feng; Zhiyuan Yuan; Yongchun Zuo

Matched generic keywords: spatial

Impact score: 11.60 (journal=9.50, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Methods.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: spatial. Potentially important for spatial omics analysis.

28. Identifying maximally informative signal-aware representations of single-cell data using the information bottleneck

Journal: Cell Systems **Date:** 2026-04-01 **DOI:** 10.1016/j.cels.2026.101571

Authors: Serafima Dubnov; Zoe Piran; Amit Alper; Adi Yefroimsky; Hermona Soreq; Mor Nitzan

Matched generic keywords: single-cell

Impact score: 11.40 (journal=8.40, generic=3.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell Systems.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: single-cell. Likely relevant to single-cell analysis workflows.

29. Single-cell morphodynamical trajectories enable prediction of gene expression accompanying cell state change

Journal: Cell Systems **Date:** 2026-04-01 **DOI:** 10.1016/j.cels.2026.101567

Authors: Jeremy Copperman; Ian C. Mclean; Sean M. Gross; Jalim Singh; Vaibhav Murthy; Young Hwan Chang; Alexander E. Davies; Daniel M. Zuckerman; Laura M. Heiser

Matched generic keywords: single-cell

Impact score: 11.40 (journal=8.40, generic=3.00, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Cell Systems.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: single-cell. Likely relevant to single-cell analysis workflows.

30. The potential of wheat spatial omics

Journal: Nature Genetics **Date:** 2026-04-01 **DOI:** 10.1038/s41588-026-02542-w

Authors: Xiao-Yuan Tao; Cong Tan; Yifan Liu; Yuanyuan Wang; Ali Raza; Jing He; Lulu Wang; Keke Xia; Yan Yan; Sha Liao; Wei Jiang; Yue Qu

Matched generic keywords: spatial

Impact score: 11.30 (journal=9.20, generic=2.10, interest=0.00, citation bonus=0.00, abstract bonus=0.00, extra bonus=0.00)

Why it matters: This paper appears to present a potentially high-impact contribution in bioinformatics or computational biology published in Nature Genetics.

Major discoveries: Abstract unavailable from the retrieval sources used by this script.

Likely impact: Matched keywords: spatial. Potentially important for spatial omics analysis.