Doi: 10.32481/djph.2026.05.17

Knowledge Graph-Driven AI in Biohealth:

From Biomedical Discovery to Health Risk Prediction

Chuming Chen, PhD;[1,2] Manju Anandakrishnan, PhD;[2] Cathy H. Wu, PhD[1,2]

1. Department of Computer and Information Sciences, University of Delaware

2. Center for Bioinformatics and Computational Biology, University of Delaware

Abstract

Knowledge graphs (KGs) have emerged as a powerful tool for knowledge discovery. In this perspective paper, we present a framework for KG construction, graph representation learning, and predictive modeling towards AI-driven discovery in biohealth. We illustrate this through two case studies: (1) Protein Knowledge Network (ProKN) and KSMoFinder, a KG embedding-based model that predicts protein kinase and phosphorylation site associations with state-of-the-art accuracy by learning from biological context in a biomedical knowledge network for drug discovery; (2) Social Determinants of Health (SDoH) KG, built from synthetic data of Veteran Health Administration with a veteran suicide-risk prediction model that uncovers latent, multifactorial risk patterns. These use cases spanning biomedical and population health research, demonstrate how KG-driven AI can bridge the gap between molecular and population level studies. We highlight how such open, interoperable knowledge networks offer a reusable framework for accelerating discovery and addressing complex health challenges. Finally, we provide targeted recommendations for Delaware's health innovation ecosystem to leverage this paradigm for public health strategy, clinical decision-making, and translational research.

The Imperative for Connected Data in Life Sciences

The life sciences are experiencing a transformative shift, driven by the rapid expansion of data across genomics, electronic health records (EHRs), and public health domains. Artificial Intelligence (AI) has become indispensable for finding patterns in life sciences data, enabling advances in diagnostic imaging, genomic interpretation, and predictive analytics. However, conventional methods often treat data as independent features, failing to capture the rich, relational fabric of biological and health systems. Graph-based AI approaches explicitly model entities (genes, diseases, patients) and their interactions as networks, enabling the learning of higher-order dependencies. Graph representation learning has been shown to effectively harness molecular interaction networks, disease comorbidity graphs, and multimodal clinical data, providing a more holistic representation of complex biological and healthcare systems.

Knowledge graphs (KGs) are semantic network structures that explicitly model entities and their relationships in a structured, queryable format, forming a basis for context-aware reasoning and inference. Early semantic integration efforts used linked data to unify diverse biological information, such as the Semantic Web for Health Care and Life Sciences (HCLS). KGs have been used to integrate knowledge from heterogeneous resources such as literature and curated databases into unified representations for disease research and discovery.[1] Large-scale biomedical KGs, such as Hetionet[2] for systematic drug repurposing, may consist of an integrative network of millions of relationships among compounds, diseases, genes, pathways, and phenotypes. Recent advances explore multimodal KG constructs that support integration

across text, images, and structured sources for richer inference tasks. For example, PrimeKG[3] integrates phenotypic, molecular, and clinical data at scale to support precision medicine through multimodal KG learning. A large, open-source life science KG ecosystem has been developed to fuse multi-omics, clinical, and biomedical knowledge and support analytical and inferential workflows.[4] They illustrate how contemporary KGs link diverse biomedical entities to generate mechanistic hypotheses and predictive insights.

Knowledge graphs are also applied in clinical and public health research to represent Social Determinants of Health (SDoH), non-medical factors like socioeconomic conditions and environmental exposures, alongside clinical data, revealing how these factors relate to health outcomes. For example, SDoH-enriched KGs integrate social factors from electronic health records and biomedical data to enable link prediction and uncover associations between social determinants and biological entities in diseases like Alzheimer's.[5] Other studies build KGs from population-level data to examine how SDoH concepts like employment and housing connect to health outcomes, uncovering relational patterns that remain hidden when social factors are treated as static and independent variables.[6] These graph-based models replace traditional static representations with dynamic network structures, enabling exploration of how social and clinical variables cluster together, interact with one another, and amplify risk across populations. Moreover, SDoH KGs can leverage graph-based AI techniques, including graph neural networks (GNNs) and link prediction, to infer missing relationships, detect risk-factor communities, and generate context-aware predictions.[7,8]

Meanwhile, federated data ecosystems such as the NIH Common Fund Data Ecosystem (CFDE)[9] advance FAIR (Findable, Accessible, Interoperable, Reusable) principles by harmonizing metadata across NIH Common Fund programs into integrated models, enhancing data discoverability and interoperability. Programs like the NSF Prototype Open Knowledge Network (Proto-OKN)[10] further invest in open, shared knowledge graph infrastructure that links data across different domains, enabling AI-driven discovery and actionable insights in health, science, and society. A recent perspective article[11] discussed six desiderata for a biomedical knowledge network resulting from an NIH workshop participated by thought leaders in the field.

We propose that KG-driven AI represents a paradigm shift for translational research, one that unifies biological and social context into a continuous knowledge fabric. We demonstrate this through projects that apply a common KG-AI pipeline that are generally applicable to molecular and population level studies: (1) predicting kinase-substrate associations via protein-level relationships in Protein Knowledge Network (ProKN),[12] and (2) modeling veteran suicide risk using an SDoH KG.[13] Together, these case studies illustrate how KG-driven AI can offer a reusable, interpretable framework for accelerating discovery in biohealth.

Case Study 1: The Protein Knowledge Network (ProKN) and KSMoFinder - Predicting Molecular Interactions

The Challenge: Protein phosphorylation, a crucial cellular process mediates signaling through kinase-driven modification of substrate proteins, and dysregulation of phosphorylation is observed in many diseases and targets for drug development. Despite advances in phosphoproteomics, experimentally validated kinase-substrate relationships remain sparse and biased toward well-studied kinases, limiting large-scale reconstruction of signaling networks. Computational approaches have traditionally focused on local sequence features surrounding phosphorylation sites, such as kinase recognition motifs, achieving utility for site-level

prediction but largely treating substrates in isolation and failing to capture broader biological context, including protein-protein interactions, functional annotations, cellular localization, and pathway membership.[14] Recent work has reframed kinase-substrate prediction as a network completion or link prediction problem, leveraging heterogeneous biological knowledge encoded as graphs.[15,16] Knowledge graph-based machine learning approaches, including graph embeddings and graph neural networks, integrate sequence, functional, and interaction data to capture long-range dependencies inaccessible to motif-centric models, demonstrating improved coverage and generalization for sparsely annotated kinases.[12]

Our KG-Driven Approach: We developed the Protein Knowledge Network (ProKN), an open knowledge graph that integrates protein-centric information from UniProt, iPTMnet, Reactome, and the CFDE. ProKN represents proteins as interconnected entities based on functional annotations, pathway participation, cellular localization, and membership in molecular complexes, resulting in a comprehensive graph of approximately 4.8 million triples.[12]

KSMoFinder - The Predictive Model: Built on ProKN, KSMoFinder is a predictive framework that leverages knowledge graph embeddings to learn contextual representations of kinases, substrates, and their phosphorylation motifs.[12] In contrast to protein language models that primarily capture sequence patterns, KSMoFinder explicitly encodes biological semantics and relational context, integrating functional annotations, pathway information and motif specificities. A unique contribution of KSMoFinder is its "substrate_motif" prediction level, which combines the functional characteristics of the substrate protein with the local amino acid sequence surrounding a phosphosite. The model provides comprehensive coverage of 430 human kinases across nine major groups, including Atypical, AGC, and CMGC kinases etc.

Key Findings and Translational Impact:

KSMoFinder achieved a ROC-AUC of 0.851, outperforming models based on advanced sequence-only embeddings as shown in Table 1. Ablation studies confirmed that removing the biological relationship data from the KG caused a significant performance drop, proving that context is a critical predictive signal.[12] The model provides biologically plausible rationales. For example, it assigns high probability to CDK19 phosphorylating substrates like MED14 and MED26 because they participate in shared transcription pathways and nuclear localization, in addition to sequence specificity.[12] The model employs a biologically motivated negative generation strategy, pairing kinases with non-interacting proteins and experimentally derived unfavored motifs, which reduces false-negative rates common in random sampling methods.[12]

Table 1. Prediction Performance of Kinase-Substrate Models Developed Using Embeddings from the KGE Model and Other Protein-Language Models[12]

Embedding source	**ROC-AUC**	**PR-AUC**
KSMoFinder-KGE	0.851 ± 0.008	0.839 ± 0.008
ProtT5	0.752 ± 0.007	0.726 ± 0.001
ESM2	0.691 ± 0.009	0.659 ± 0.01
ESM3	0.501 ± 0.003	0.5 ± 0.002

Random	0.498 ± 0.004	0.5 ± 0.002

Systems Biology Impact and Drug Discovery: By leveraging graph-based representations and relational embeddings, this framework transforms kinase prediction from simple sequence correlation into context-aware inference, capturing functional, structural, and network-level dependencies that are inaccessible to motif-centric models. This approach not only accelerates the generation of testable hypotheses in cancer signaling and other disease contexts, but also facilitates the prioritization of drug targets, the functional interpretation of disease-associated genetic variants, and the systematic exploration of understudied kinases across the human proteome. By integrating multiple scales of biological evidence into a unified predictive model, it provides a scalable and interpretable platform for network-driven discovery in systems biology.

Case Study 2: The BioHealthKG OKN and SDoH KG - Predicting Population Health Risks in Veterans

The Challenge: Social determinants of health, including factors like housing instability, economic security, and social connection, are widely recognized as fundamental drivers of health outcomes across populations, yet their complex and synergistic effects are difficult to capture using traditional statistical approaches that typically examine single risk factors in isolation rather than interacting systems of influence. Public health frameworks and reports have documented how socioeconomic conditions and material resources shape patterns of morbidity and mortality, highlighting the need for analytic frameworks that move beyond simple associations to understand interdependent influences on health.[17] This challenge is especially acute among vulnerable subgroups such as U.S. military veterans, where multiple social needs, including housing instability, unemployment, and limited social support, are prevalent and have been linked to worse mental health outcomes, including elevated symptoms of Post-traumatic stress disorder (PTSD), depression and suicide.[18,19]

Our KG-Driven Approach: We built a patient-centric SDoH KG as part of the BioHealthKG OKN, using a privacy-preserving synthetic cohort of 111,000 veterans generated via the MDClone platform from Veterans Health Administration electronic health records that demonstrated high statistical fidelity against real data.[20] The graph linked over 800,000 nodes across 5.9 million relationships,[12] modeling the clinical and SDoH factors. Using BioCypher[21] and the BioLink ontology[22] for standardization, we organized SDoH factors into the five Healthy People 2030[23] domains: Economic Stability, Education Access, Healthcare Access, Neighborhood and Built Environment, and Social and Community Context.

Key Findings: We used Fast Random Projection (FastRP) to generate graph embeddings for patients and SDoH factors, employing them as features to train an XGBoost model for suicide risk prediction, achieving superior performance (AUC-ROC: 0.996, F1-score: 0.937) over a tabular-feature model (AUC-ROC: 0.963, F1-score: 0.742).[13] The detailed evaluation results are shown in Table 2. This demonstrates the profound predictive capability in connected data. Topological link prediction further identified statistically significant latent connections, such as pathways between "Lack of Housing" and psychosocial circumstances (Normalized Score: 0.396, p=0.00021) and links from literacy challenges to suicidal ideation,[13] quantifying the structural interplay of social and clinical risk.

Doi: 10.32481/djph.2026.05.17

Table 2. Performance of Baselines vs KG-Based Models in Suicide Risk Prediction[13]

Data	Model	Accuracy [CI]	F1-Score [CI]	AUC-ROC [CI]
Tabular	Logistic Regression	0.915 ± 0.002	0.618 ± 0.004	0.903 ± 0.003
	Random Forest	0.932 ±0.001	0.721 ± 0.003	0.957 ± 0.001
	XGBoost	0.935 ± 0.001	0.742 ± 0.002	0.963 ± 0.001
KG Embeddi ng	Logistic Regression	0.965 0.002	0.885 ± 0.004	0.993 ± 0.001
	Random Forest	0.960 0.001	0.850 ± 0.003	0.990 ± 0.001
	XGBoost	0.983 0.001	0.937 ± 0.002	0.996 ± 0.001

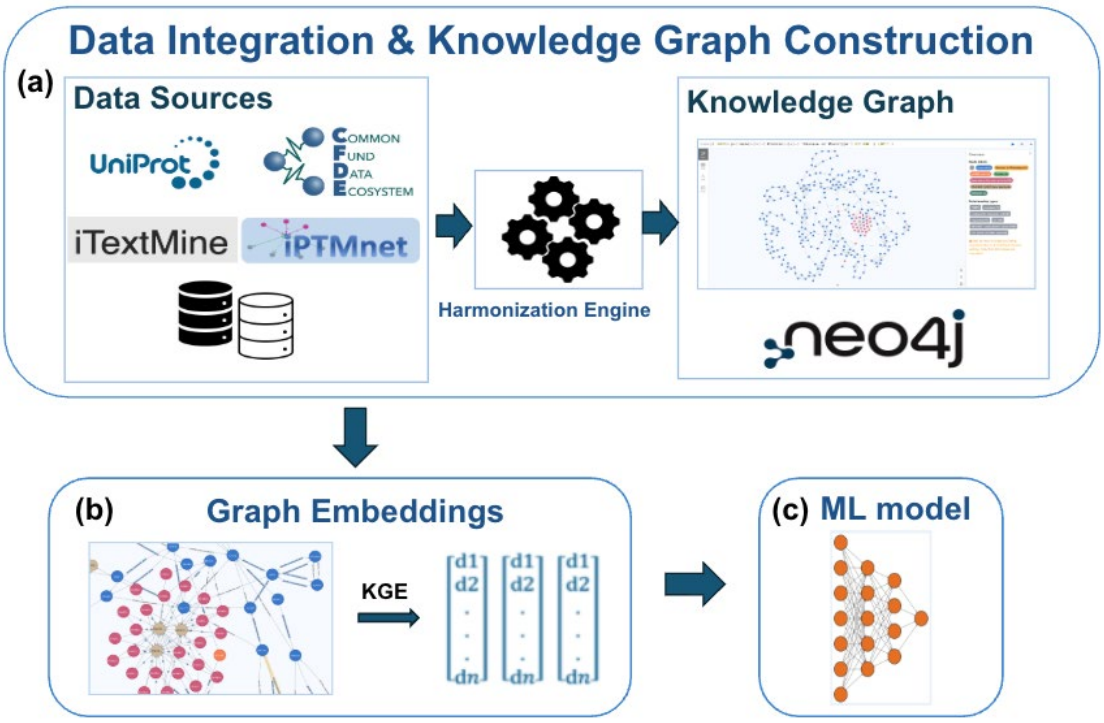
Public Health Impact: This approach can transform a list of risk factors into an interpretable risk map. For Delaware's public health officials, a similar KG could be used to identify how specific combinations of transportation barriers, food insecurity, and social isolation are clustered to amplify risk for diabetes, asthma, or substance use in specific communities. This enables coordinated interventions that target the network of risk, not just isolated symptoms.

Synthesis: A Generalized KG-AI Framework

The two case studies presented above exemplify a coherent and generalizable KG-AI methodology spanning biological to the population level research. This framework can be conceptualized as three reusable components (Figure 1):

Figure 1. KG-AI Framework

Doi: 10.32481/djph.2026.05.17



A unified methodology comprising three reusable components: (a) **Knowledge Graph Construction** integrates heterogeneous data into a semantically structured heterogeneous graph. (b) **Embedding Learning** generates low-dimensional entity embeddings encoding relational context and network topology. (c) **Predictive Modeling** uses graph-native embeddings as features for downstream models enabling accurate predictions.

Knowledge Graph Construction (Figure 1a): Integrate heterogeneous, domain-specific data (e.g., patients and SDoH; proteins and pathways) into a semantically structured graph using frameworks like BioCypher and ontologies such as BioLink. This stage transforms isolated datasets into a connected knowledge fabric, capturing both entities and their interrelationships.

Embedding Learning (Figure 1b): Apply graph representation learning algorithms[24] to generate low-dimensional embeddings for each entity. By encoding relational context and topological role, these embeddings represent each entity in terms of its position and interactions within the network, complementing information derived from intrinsic attributes.

Predictive Modeling (Figure 1c): Our methodology trains predictive models, such as XGBoost or neural networks, using graph-native embeddings as features, leading to accurate and interpretable results for diverse tasks. This end-to-end framework bridges a critical gap in the KG landscape by scaling across resolutions, from protein molecular functions to population-level social exposomes. By providing a practical, scalable, and generalized KG-AI implementation, our work advances the FAIR and interconnected data vision.[25] The advantages of our approach are threefold:

1. **Accuracy:** Connected data and network topology enhance predictive power, outperforming traditional machine learning on traditional non-graph data in both case studies.

Doi: 10.32481/djph.2026.05.17

2. **Interpretability:** Graph structure enables explanations by tracing predictions to influential nodes and pathways (e.g., a patient's risk linked to housing and social isolation; a kinase's activity linked to shared localization).
3. **Actionability:** Outputs are not just scores, but maps of influence, guiding interventions to central nodes or high-risk pathways for clinicians, public health officials, or biomedical researchers.

For Delaware's health innovation ecosystem, this unified framework establishes a shared competency in graph-based data integration and AI, adaptable to pressing health challenges, from chronic disease disparities to accelerating biotech-driven drug discovery. The future of translational research lies in connected learning across scales, and this KG-AI framework provides a practical, scalable architecture to realize that vision.

Recommendations for Delaware's Health Innovation Ecosystem

With integrated healthcare systems, a growing life sciences sector, and committed public health leadership, Delaware has the infrastructure and institutional capacity to benefit from knowledge graph–driven AI. Based on insights from our case studies, we put forward the following targeted and actionable recommendations.

Public Health Agencies & Policymakers: Delaware should prioritize the development of a Delaware Health Equity Knowledge Graph (DE-HEKG), integrating de-identified, HIPAA-compliant data across health systems, housing, education, and environmental monitoring. Leveraging semantic standards and privacy-preserving techniques, the DE-HEKG would enable identification of population-level risk clusters, for example, links between transportation barriers and pediatric asthma emergency visits, supporting targeted, network-informed public health interventions. Robust governance, formalized through a proposed "Delaware Data Trust," would establish the ethical, technical, and stewardship frameworks necessary to build institutional trust, facilitate sustainable cross-agency data sharing, and realize the full public health potential of integrated data resources.

Healthcare Providers & Systems: Clinical workflows can be enhanced by KG-powered decision support that contextualizes patient data within their social and medical network. By integrating SDoH sub-graphs into EHRs, providers can identify compounding risk factors, e.g., "lives alone + limited transportation + low health literacy" and trigger tailored interventions or referrals. Complementary training in graph-based, interpretable AI ensures clinicians can understand and trust model outputs, moving beyond generic risk flags to actionable, personalized care strategies.

Academic and Industry Researchers: Delaware can accelerate translational research by forming interdisciplinary teams that combine domain experts with graph and AI specialists. Seed grants and shared "Graph AI Labs" can support projects such as applying the KSMoFinder/ProKN framework to cancer, biotechnology, or precision agriculture. Open-source pipelines, visualization tools, and Delaware-centric benchmark tasks will strengthen statewide technical capacity, foster collaboration, and position the state as a contributor to the national open-science ecosystem, enabling scalable, reproducible discoveries across biomedical and life sciences domains.

In summary, investing in the KG-AI paradigm strengthens Delaware's data infrastructure, cross-sector collaboration, and workforce development. It supports precision public health, enhances population health outcomes, and drives innovation in the state's life sciences ecosystem.

Conclusion

The combination of knowledge graphs and artificial intelligence represents a significant step forward in life sciences and data science, moving from isolated correlations to reasoning over connected systems. Our work spans applications from veteran suicide risk to systems biology, demonstrating a unified KG-driven paradigm that produces models that are more accurate, interpretable, and actionable across scales. For Delaware, investing in the necessary data infrastructure, cross-sector partnerships, and workforce expertise provides an opportunity to improve population health, support life sciences innovation, and advance precision public health. By enabling connected learning across datasets and domains, knowledge graphs offer a practical foundation for the future of health discovery.

Dr. Chen may be contacted at chenc@udel.edu.

Acknowledgement

This work was partially supported by grants from the National Science Foundation (2333740 and 2438144) and the National Institutes of Health (P20GM103446, U54GM104941, R35GM141873, U24OD038424 and S10OD028725).

References

1. Chen, C., Ross, K. E., Gavali, S., Cowart, J. E., & Wu, C. H. (2021, December 7). COVID-19 Knowledge Graph from semantic integration of biomedical literature and databases. *Bioinformatics (Oxford, England)*, *37*(23), 4597–4598. https://doi.org/10.1093/bioinformatics/btab694 PubMed
2. Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., . . . Baranzini, S. E. (2017, September 22). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, *6*, e26726. https://doi.org/10.7554/eLife.26726 PubMed
3. Chandak, P., Huang, K., & Zitnik, M. (2023, February 2). Building a knowledge graph to enable precision medicine. *Scientific Data*, *10*(1), 67. https://doi.org/10.1038/s41597-023-01960-3 PubMed
4. Callahan, T. J., Tripodi, I. J., Stefanski, A. L., Cappelletti, L., Taneja, S. B., Wyrwa, J. M., . . . Hunter, L. E. (2024, April 11). An open source knowledge graph ecosystem for the life sciences. *Scientific Data*, *11*(1), 363. https://doi.org/10.1038/s41597-024-03171-w PubMed
5. Shang, T., Yang, S., Zhai, T., He, W., Mamourian, E., Zhang, J., . . . Shen, L. (2025, September 23). A novel computational analysis integrating social determinants information from EHR and literature with Alzheimer's disease biological knowledge through large language models and knowledge graphs. *Innovation in Aging*, *9*(Suppl 1), S2–S13. https://doi.org/10.1093/geroni/igaf102 PubMed

6. Bettencourt-Silva, J. H., Mulligan, N., Jochim, C., Yadav, N., Sedlazek, W., Lopez, V., & Gleize, M. (2020, November 23). Exploring the social drivers of health during a pandemic: Leveraging knowledge graphs and population trends in COVID-19. *Studies in Health Technology and Informatics*, *275*, 6–11. https://doi.org/10.3233/SHTI200684 PubMed

7. Johnson, R., Li, M. M., Noori, A., Queen, O., & Zitnik, M. (2024, August). Graph artificial intelligence in medicine. *Annual Review of Biomedical Data Science*, *7*(1), 345–368. https://doi.org/10.1146/annurev-biodatasci-110723-024625 PubMed

8. Li, M. M., Huang, K., & Zitnik, M. (2022, December). Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, *6*(12), 1353–1369. https://doi.org/10.1038/s41551-022-00942-x PubMed

9. Evangelista, J. E., Clarke, D. J. B., Byrd, A. I., Srinivasan, S., Srinivasan, S., Maurya, M. R., . . . Ma'ayan, A. (2026, January 6). The CFDE workbench: Integrating metadata and processed data from common fund programs. *Journal of Molecular Biology*, *169631*, 169631; Advance online publication. https://doi.org/10.1016/j.jmb.2026.169631 PubMed

10. Proto-OKN. (n.d.). *Prototype open knowledge network*. Retrieved April 6, 2026, from https://www.proto-okn.net/

11. Wu, C., Liu, H., Flannick, J., Musen, M. A., Su, A. I., Hunter, L. E., . . . Wu, C. H. (2026, March 20). Desiderata for a biomedical knowledge network: Opportunities, challenges and future directions. *Bioinformatics Advances*, *6*(1), vbag036. https://doi.org/10.1093/bioadv/vbag036 PubMed

12. Anandakrishnan, M., Ross, K. E., Chen, C., Vijay-Shanker, K., & Wu, C. H. (2024). KSMoFinder-knowledge graph embedding of proteins and motifs for predicting kinases of human phosphosites. *Bioinformatics Advances*, *5*(1), vbaf289. https://doi.org/10.1093/bioadv/vbaf289 PubMed

13. Chen, C., Piya, F. L., Rolnick, J. A., Milbourne, S. A., Wu, C., Powers, T. M., . . . Beheshti, R. (2025). Leveraging social determinants of health (SDoH) knowledge graph to identify latent patterns in veteran suicide risk. *In Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2025)*. IEEE. https://openreview.net/forum?id=mHRNk9qzfG

14. Nováček, V., McGauran, G., Matallanas, D., Vallejo Blanco, A., Conca, P., Muñoz, E., . . . Fey, D. (2020, December 3). Accurate prediction of kinase-substrate networks using knowledge graphs. *PLoS Computational Biology*, *16*(12), e1007578. https://doi.org/10.1371/journal.pcbi.1007578 PubMed

15. Gavali, S., Ross, K., Chen, C., Cowart, J., & Wu, C. H. (2022, October 31). A knowledge graph representation learning approach to predict novel kinase-substrate interactions. *Molecular Omics*, *18*(9), 853–864. https://doi.org/10.1039/D1MO00521A PubMed

16. Anandakrishnan, M., Ross, K. E., Chen, C., Shanker, V., Cowart, J., & Wu, C. H. (2023, October 6). KSFinder-a knowledge graph model for link prediction of novel phosphorylated substrates of kinases. *PeerJ*, *11*, e16164. https://doi.org/10.7717/peerj.16164 PubMed

17. Marmot, M. (2005, March 19-25). Social determinants of health inequalities. *Lancet*, *365*(9464), 1099–1104. https://doi.org/10.1016/S0140-6736(05)71146-6 PubMed

18. Holder, N., Holliday, R., Ranney, R. M., Bernhard, P. A., Vogt, D., Hoffmire, C. A., . . . Maguen, S. (2023, October). Relationship of social determinants of health with symptom severity among Veterans and non-Veterans with probable posttraumatic stress disorder or depression. *Social Psychiatry and Psychiatric Epidemiology*, *58*(10), 1523–1534. https://doi.org/10.1007/s00127-023-02478-0 PubMed

19. Mitra, A., Pradhan, R., Melamed, R. D., Chen, K., Hoaglin, D. C., Tucker, K. L., . . . Yu, H. (2023, March 1). Associations between natural language processing-enriched social determinants of health and suicide death among US veterans. *JAMA Network Open*, *6*(3), e233079. https://doi.org/10.1001/jamanetworkopen.2023.3079 PubMed

20. Reiner Benaim, A., Almog, R., Gorelik, Y., Hochberg, I., Nassar, L., Mashiach, T., . . . Beyar, R. (2020, February 20). Analyzing medical research results based on synthetic data and their relation to real data results: Systematic comparison from five observational studies. *JMIR Medical Informatics*, *8*(2), e16492. https://doi.org/10.2196/16492 PubMed

21. Lobentanzer, S., Aloy, P., Baumbach, J., Bohar, B., Carey, V. J., Charoentong, P., . . . Saez-Rodriguez, J. (2023, August). Democratizing knowledge representation with BioCypher. *Nature Biotechnology*, *41*(8), 1056–1059. https://doi.org/10.1038/s41587-023-01848-y PubMed

22. Unni, D. R., Moxon, S. A. T., Bada, M., Brush, M., Bruskiewich, R., Caufield, J. H., . . . Mungall, C. J., & the Biomedical Data Translator Consortium. (2022, August). Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. *Clinical and Translational Science*, *15*(8), 1848–1855. https://doi.org/10.1111/cts.13302 PubMed

23. Office of Disease Prevention and Health Promotion. (n.d.). *Healthy people 2030: Social determinants of health*. U.S. Department of Health and Human Services. Retrieved April 6, 2026, from https://odphp.health.gov/healthypeople/priority-areas/social-determinants-health

24. Khoshraftar, S., & An, A. (2024). A survey on graph representation learning methods. *ACM Transactions on Intelligent Systems and Technology*, *15*(1), 1–55. https://doi.org/10.1145/3633518

25. National Institutes of Health Office of Data Science Strategy. (n.d.). *NIH strategic plan for data science*. Retrieved April 6, 2026, from https://datascience.nih.gov/sites/g/files/mnhszr336/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf