Received: 21 March 2026 | Revised: 27 April 2026 | Accepted: 17 May 2026 | Published online: 21 May 2026

International Journal of Advance in Clinical Science Research, Volume 5 2026
https://h-tsp.com/
https://doi.org/10.5281/zenodo.20321578

A Systematic Framework for Analyzing Drug Mechanisms in the Era of Biomedical Big Data

Siqi Lan*

Tianjin Medical University, Tianjin, China

*Correspondence Author

Abstract: ***Biomedical big data explosion has caused a paradigm shift in the approach to drug research that had been based on hypothesis until the development of data intensive systemic analysis. Conventional reductionist methods do not generally explain the polypharmacological character of drugs and complexity of interrelation of multifactorial diseases. The present paper suggests an integrative computational model that deciphers the drug mechanisms of actions based on building hierarchical tripartite drug-gene-disease networks. The framework uses a more or less the topographic measure of the therapeutic influence in form of a propagation matrix and network topology features, including the degree centrality and shortest path distance, to determine pivotal regulatory hubs, and mechanistic proximity. A case simulation study with simulated interactions illustrates how the framework has been able to prioritize drug to disease interactions as well as identifies necessary bridge genes that mediate systemic effects. Additionally, we cover the possibility of incorporating this model into AI-based drug discovery pipelines, in particular, in property prediction and generative design sub- modules. Although the existing restrictions are that it depends on static and linear modeling, the framework offers a scalable and explainable basis of systems pharmacology. The data in the future will combine dynamic multi-omics and non-linear deep learning architecture to make drug mechanism studies more interpretable and predictive.***

Keywords: Network Pharmacology, Tripartite Graph, Mechanism Propagation, Biomedical Big Data, Systems Biology.

1. Introduction

High-throughput sequencing and digital healthcare have led to the development of biomedical big data [1-2], which has shifted the focus of drug research towards data-intensive systemic research rather than the hypothesis-driven approach of drug research research. Nevertheless, the reductionist framework of one drug, one target and one disease is becoming more and more insufficient to describe polypharmacological [3], it is also a fundamental assumption underpinning the interlocking nature of multifactorial diseases. The task of massive, heterogeneous datasets translation into comprehensible mechanistic findings is a severe problem. As a result, an integrative approach, system-oriented method of analysis is urgently required to get a more comprehensive picture of the efficacy and adverse reactions of drugs as system phenomena [4]. The notion of network pharmacology, systems biology and integration of multi-omics is finding increasing support, in that a nonlinear causes and effects model should be replaced by a nonlinear and networked data representation of biological activities [5].

Simulating drugs, genes, and diseases as part of an integrated system enables biologists to discover how interactions spread throughout molecular systems and how these interactions affect phenotypic changes [6]. Network schemes also offer quantitative measures to characterize the topological stature, pathway connections, and global influence tendencies thus supplying a more multifaceted insight into the mechanism of drugs.

It is on this basis that the current study presents a logical framework of modeling the drug mechanisms in the biomedical big data setting. The important research questions that are addressed in this paper are the modeling of drug action mechanisms model in a multi-source data environment, the use of network structures and mathematical modeling to quantify the relationship between drugs, targets and diseases, and to increase the explainability and interpretability of the mechanism analysis using a computational modeling. To meet these aims, we present a drug-gene-disease tripartite network, which entails a heterogenous biomedical network with the incorporation of various biomedical data sources. The propagation model, based on the use of matrices, is presented to measure the possibility of drug-disease relationship, and the topological measurements are used to determine meaningful regulatory genes and important pathways. However, this research does not aim at large-scale empirical validation, but instead insists on the conceptual modeling and demonstrative analysis to provide the representation of the possible correctness of a systems-level computational framework. The proposed solution will offer a scalable and explainable basis of drug mechanism studies into the age of biomedical big data by integrating medical expertise with quantitative network modeling.

2. System Framework for Systematic Drug Mechanism Modeling

This paper suggests an integrative computational approach to the systems of decoding the mechanism of action of drugs within the complex biomedical big data environment. This framework will be used to close the gap between disjointed biological information and mechanistic understanding. The framework allows the systematic analysis of the mechanism of interaction between drugs and targets and downstream phenotypic consequences by converting heterogeneous biomedical data on drug-target interactions to a single, quantifiable format. The general structure is structured in four hierarchical levels of: (1) Data Integration, (2) Network Construction, (3) Mechanism Propagation, and (4) Interpretative Analysis.

2.1 Data Integration and Matrix Representation

The initial stage focuses on the abstraction of heterogeneous biomedical resources—ranging from genomic associations and gene expression profiles to clinical disease annotations—into a unified relational format. Let the drug set be $D = \{d_1, d_2, \dots, d_n\}$, the gene set $G = \{g_1, g_2, \dots, g_m\}$, and the disease set $S = \{s_1, s_2, \dots, s_k\}$. These entities constitute the fundamental nodes of a tripartite biological system.

The relationships between these entities are encoded into structured adjacency matrices to facilitate high-throughput computation. Specifically, the drug–gene interaction matrix $A \in \mathbb{R}^{n \times m}$ is defined as:

$$A_{ij} = \begin{cases} 1, & if\ drug\ d_i\ targets\ gene\ g_j \\ 0, & otherwise \end{cases}$$

Similarly, the gene–disease association matrix $B \in \mathbb{R}^{m \times k}$ represents the clinical or functional links between gene g_j and disease s_k. By representing biological knowledge in this matrix-based format, the framework can accommodate diverse data sources while ensuring mathematical consistency for subsequent modeling [7].

2.2 Tripartite Network Construction

In the network construction layer, we establish a tripartite graph $\mathcal{G} = (V, E)$, where the vertex set $V = D \cup G \cup S$ represents the union of all biological and chemical entities. The edge set E encompasses all validated and predicted drug–gene and gene–disease connections. Unlike traditional reductionist models that focus on isolated molecular events, this network-based representation captures the topological complexity and systemic nature of biological regulation. This holistic structure allows for the identification of indirect interactions and compensatory pathways that are often overlooked in single-target studies, providing a more realistic simulation of how drugs perturb cellular networks.

2.3 Mechanism Propagation and Quantification

To quantify how therapeutic signals propagate from molecular targets to clinical phenotypes, we introduce a mechanism propagation model. The potential influence of a drug on a specific disease is captured by the drug–disease influence matrix F, initially derived through matrix multiplication:

$$F = A \times B$$

In this formulation, F_{ik} represents the cumulative influence strength of drug d_i on disease s_k, mediated by all intermediate gene interactions. To further refine this model and account for the varying importance of different genes, we introduce a diagonal weighting matrix $W \in \mathbb{R}^{m \times m}$:

$$F = A \times W \times B$$

The weight W_{jj} can be assigned based on biological relevance, such as the gene's differential expression significance, network centrality, or functional confidence score [8]. This weighted approach enhances the interpretability of the model by emphasizing key regulatory nodes that drive the therapeutic process.

2.4 Interpretative Analysis Layer

The final layer employs network topology metrics and path analysis to extract actionable mechanistic insights. We utilize Degree Centrality, defined as $C_d(v) = deg(v)$, to pinpoint "hub" genes or multi-target drugs that exert broad influence across the network. Furthermore, we define the Mechanism Distance $L(d, s)$ as the shortest path length between a drug node d and a disease node s:

$$L(d, s) = min(\, path\ length\ from\ d \rightarrow g \rightarrow s)$$

Shorter pathways are usually considered to be direct, primary mechanisms of therapy whereas longer paths are deemed to be complicated, indirect regulatory processes or systemic actions. This framework offers a scalable and powerful approach to the deciphering of multi-level actions of drugs in the age of systems pharmacology by incorporating numerical scoring and topological interpretation.

3. Mechanism Mining and Illustrative Analysis

3.1 Network Topology and Centrality Analysis

According to the suggested tripartite drug–gene disease network, the mechanism mining is performed in order to determine essential regulatory aspects and measure the possible drug-disease relationships. This system-focused technique is instead of traditional single target analysis because it focuses on propagating interactions and provides a system effect on multi-level. Topological properties are also studied after the construction of a network in order to identify the crucial nodes in the triple structure. Regulatory hubs can be very complex biological systems that are highly connected with and as a result

of which several signaling pathways can take place. In order to identify these hubs, we make use of Degree Centrality which is defined as:

$$C_d(v) = deg(v)$$

where $deg(v)$ represents the number of direct edges incident to node v. Genes with high degree centrality are considered potential bottleneck nodes in the regulatory network, playing critical roles in transmitting therapeutic effects from drugs to disease phenotypes. Furthermore, to characterize the mechanistic proximity between biological entities, we define the Mechanism Distance $L(d, s)$ as the shortest path length between a drug node d and a disease node s:

$$L(d, s) = min\{\, path\ length \mid d \to g \to s\}$$

A shorter distance suggests a more direct mechanistic association, whereas longer paths indicate indirect regulation or systemic downstream effects.

3.2 Quantitative Influence Propagation

To quantify the latent influence of drugs on disease modules, the mechanism propagation model is applied as follows [8]:

$$F = A \times W \times B$$

In this formulation:

$A \in \mathbb{R}^{n \times m}$ represents the drug–gene interaction matrix.

$B \in \mathbb{R}^{m \times k}$ denotes the gene–disease association matrix.

$W \in \mathbb{R}^{m \times m}$ is a diagonal weighting matrix, where each element W_{jj} reflects the biological importance (e.g., functional confidence or expression significance) of gene g_j.

The resulting influence matrix F provides a quantitative score F_{ik}, representing the inferred strength of the effect of drug d_i on disease s_k through the mediated gene layer. Higher F_{ik} values imply stronger mechanistic relevance and potential therapeutic efficacy.

3.3 Illustrative Case Study and Results

To demonstrate the feasibility of the framework, we constructed an illustrative network comprising 5 drugs ($D_1 - D_5$), 10 genes ($G_1 - G_{10}$), and 3 diseases ($S_1 - S_3$). Adjacency relationships were simulated to reflect biologically plausible interactions. The calculated drug–disease influence scores are summarized in Table 1.

Table 1: Quantitative Influence Scores and Mechanistic Rankings

Drug	Primary Target Genes	Disease S1	Disease S2	Disease S3
D1	G1,G3,G5	0.82	0.34	0.21
D2	G2,G4	0.45	0.76	0.28
D3	G3,G6,G7	0.67	0.4	0.59
D4	G8	0.12	0.18	0.71
D5	G1,G9	0.54	0.22	0.33

As shown in the analysis, Drug D_1 demonstrates the highest influence score (0.82) on Disease S_1. Further topological dissection indicates that this effect is heavily mediated by Gene G_3, which functions as a

high-degree "regulatory bridge" connecting multiple drug nodes to disease modules. The high centrality of G_3 suggests it may be a critical hub gene where multiple signaling pathways converge, making it a high-confidence target for systemic intervention. In contrast, Drug D_4 shows a highly specific but localized effect on Disease S_3 (0.71), suggesting a targeted mechanism with potentially fewer systemic side effects compared to the multi-target profile of D_3.

This descriptive analysis confirms the main strengths of the suggested framework, specifically offering a numerical background to rank and prioritize drug-disease associations and get out of qualitative descriptions. The framework is able to combine the concept of network topology and propagation modeling to locate vital regulatory nodes that initiate therapeutic responses. Moreover, it identifies interpretable biological pathways, which filled the gap between molecular-level targets and clinical phenotypes. Altogether, the matrix-based propagation and network analysis merge effectively so as to provide a more formal representation of drug activities and, as a result, demonstrates that the construct has the potential to convert intricate, multi-source biomedical information into practical and scalable mechanistic understanding.

4. Discussion

The tripartite network framework and propagation model proposed will give a solid computational approach to deconstructing drug mechanisms in the biomedical big data era. This study meets the challenge of pharmaceutical action complexity and polypharmacology of contemporary therapeutics through a shift in the downward perspective of the reductionist single-target model to an integrative perspective based on systems.

4.1 Advantages of the Multi-Source Integrative Framework

The main advantage with this framework is the possibility to use the heterogeneous data (high-throughput molecular profiles and clinical phenotype associations) and to reconcile them in a single mathematical representation. The traditional models tend to underestimate systemic ripple effects which take place when a drug can disturb a biological network. These effects are well described by our matrix-based propagation model ($F = A \times W \times B$): this model is used to compute the distance that a signal propagates through a layer of weights on the laid down genes to disease phenotypes.

Of great importance is the introduction of the weighting matrix W that enables the incorporation of biological priors like significance of expression of the different genes or tissue-relevance and increases the physiological ability to the mechanism scores obtained through the process.

4.2 Network Topology and biological interpretability.

In addition to predictive scoring, network topology measures (e.g., Degree Centrality and Mechanism Distance) are considered to provide the model with an interpretation that will be comprehensible by clinical researchers. Indeed, the occurrence of the so-called hub genes such as G_3 in our example study is the basis of a very learnt biological speculation: hubs are regulatory bottlenecks whose disruption is a prerequisite to a successful therapeutic outcome. The $L(d, s)$ analysis also gives the shortest path measurement, which is a measure of mechanistic directness, enabling the researcher to identify the key therapeutic objectives and secondary systemic corrections. This is the duality in quantitative scoring and qualitative pathway mining between black-box computational models and conventional experimental pharmacology.

4.3 Integration into the AI-Driven Drug Discovery Pipeline

The practical use of the suggested framework is most effectively perceived when intermingled into a thorough, AI-based drug discovery cycle as shown in Figure 3 [9]. In this pipeline, our framework is considered a key part of the Property Prediction module, namely, acting on the Systems Models layer. Although classic pipelines tend to deal with only discrete measures, such as Pharmacokinetics (PK) or fundamental efficacy, the model suggested here gives us the required "mechanistic topography" to access the systemic safety and polypharmacology. Specifically, the derived mechanism scores (F_{ik}) and identified hub genes act as Design Criteria, guiding the Generative Molecular Design engine to produce chemical entities with optimized and specific mechanistic profiles [9]. Moreover, the framework can be used to quantify Mechanism Distance to inform the Active Learning module about the most appropriate time to perform simulations or Human-relevant Assays that prove other new drug-disease relationships. This integration does not only increase interpretability of lead optimization but also enables the discovery environment to be a highly closed-loop experimental setup thus considerably decreasing costs of the experiments.

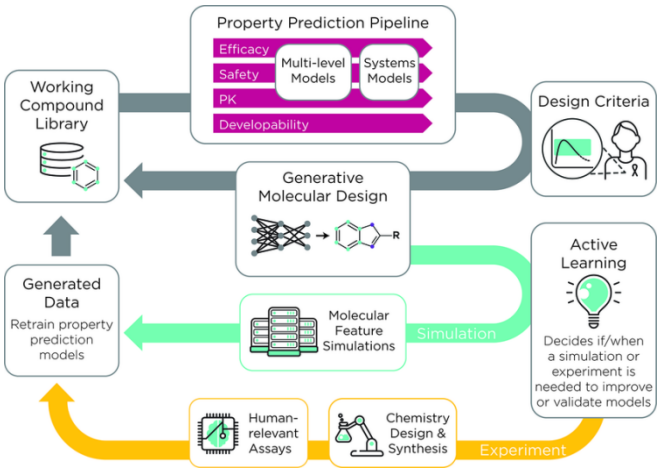

Figure 3: Conceptual integration of the systematic mechanism framework within an AI-augmented discovery cycle.

The diagram illustrates the iterative feedback loop where systems-level mechanistic insights guide molecular design and optimize experimental validation via active learning.

4.4 Limitations and Future Directions.

This study is not without limitations even though it has its positive side. First, the existing system employs a fixed network model, which might not illustrate that the biological systems respond dynamically to drug treatment in a holistic way. Subsequent versions may include time-series multi-omics data that predicts the kinetic development of drug-gene-disease interactions. Second, the matrix-based method is computationally efficient and scalable but propagation of influence is assumed to be linear. The ability of the model to represent a more difficult, non-linear biological regulatory logic could potentially be enhanced by including non-linear graph neural networks (GNNs) or deep learning architectures. Lastly, despite the demonstrative analysis demonstrating the framework is feasible, the major step to take as a follow up measure is the large-scale empirical validation of the framework using actual clinical data as well as laboratory experiments to substantiate the predicted relationships.

5. Conclusion

This paper provides a systematic computational environment of drug drug modeling in the context of biomedical big data. Using multi-source heterogeneous data and assigning them with tripartite drug-gene-disease network, we have indeed shifted to a reductionist approach of viewing a therapeutic action in terms of one drug and one target, to a systems approach of evaluating action in terms of systems. Its results make it clear that the matrix-based propagation model is capable of quantitatively capturing the latent effect of drugs on disease phenotypes, as well as emphasize that the identification of vital regulatory bottlenecks in the form of hub genes is essential. The case study example presented is affirmative that numerical scoring of a topological measure, including degree centrality and mechanism distance, is a way to achieve both quantitative and qualitative prioritizations of the entirety of polypharmacological dynamics.

Nonetheless, the study has some limitations even despite such contributions. The model that is used currently employs a static network representation, which is not necessarily the best at realizing the dynamic nature of biological responses, and assumes that signal propagation takes place linearly, which can be overly simplistic in capturing the non-linear nature of cellular control. Besides, although the viability of the framework was proven by applying an illustrative analysis, it was not empirically validated on a large scale with the use of clinical data in the real world. Further studies need to include time-series multi-omics data to detect kinetic changes and use non-linear GNNs to improve the predictability. After all, it will be critical to confirm these computational predictions by rigorous laboratory work and clinical data to close the gap between in silico programming and precision medicine.

References

[1] Del Campo, L., Sánchez-López, A., Salaices, M., von Kleeck, R. A., Expósito, E., González-Gómez, C., Cussó, L., Guzmán-Martínez, G., Ruiz-Cabello, J., Desco, M., Assoian, R. K., Briones, A. M., & Andrés, V. (2019). Vascular smooth muscle cell-specific progerin expression in a mouse model of Hutchinson–Gilford progeria syndrome promotes arterial stiffness: Therapeutic effect of dietary nitrite. Aging Cell, 18(3), e12936.

[2] Engel, J. L., Zhang, X., Wu, M., Wang, Y., Espejo Valle-Inclán, J., Hu, Q., Woldehawariat, K. S., Sanders, M. A., Smogorzewska, A., Chen, J., Cortés-Ciriano, I., Lo, R. S., & Ly, P. (2024). The Fanconi anemia pathway induces chromothripsis and ecDNA-driven cancer drug resistance. Cell, 187(21), 6055–6070.e22.

[3] Hamczyk, M. R., & Andrés, V. (2019). Vascular smooth muscle cell loss underpins the accelerated atherosclerosis in Hutchinson–Gilford progeria syndrome. Nucleus, 10(1), 28–34.

[4] Kneissig, M., Keuper, K., de Pagter, M. S., van Roosmalen, M. J., Martin, J., Otto, H., Passerini, V., Campos Sparr, A., Renkens, I., Kropveld, F., Vasudevan, A., Sheltzer, J. M., Kloosterman, W. P., & Storchová, Z. (2019). Micronuclei-based model system reveals functional consequences of chromothripsis in human cells. eLife, 8, e50292.

[5] Koblan, L. W., Erdos, M. R., Wilson, C., Cabral, W. A., Levy, J. M., Xiong, Z.-M., Tavarez, U. L., Davison, L. M., Gete, Y. G., Mao, X., Newby, G. A., Doherty, S. P., Narisu, N., Sheng, Q., Krilow, C., Lin, C. Y., Gordon, L. B., Cao, K., Collins, F. S., … Liu, D. R. (2021). In vivo base editing rescues Hutchinson–Gilford progeria syndrome in mice. Nature, 589(7843), 608–614.

[6] Kreienkamp, R., Graziano, S., Coll-Bonfill, N., Bedia-Diaz, G., Cybulla, E., Vindigni, A., Dorsett, D., Kubben, N., Batista, L. F. Z., & Gonzalo, S. (2018). A cell-intrinsic interferon-like response links replication stress to cellular aging caused by progerin. Cell Reports, 22(8), 2006–2015.

[7] Repczynska, A., Ciastek, B., & Haus, O. (2024). New insights into the Fanconi anemia pathogenesis: A crosstalk between inflammation and oxidative stress. International Journal of Molecular Sciences, 25(21), 11619.

[8] Wheaton, K., Campuzano, D., Ma, W., Sheinis, M., Ho, B., Brown, G. W., & Benchimol, S. (2017). Progerin-induced replication stress facilitates premature senescence in Hutchinson–Gilford progeria syndrome. Molecular and Cellular Biology, 37(14), e00659-16.

[9] Zhang, C.-Z., Spektor, A., Cornils, H., Francis, J. M., Jackson, E. K., Liu, S., Meyerson, M., & Pellman, D. (2015). Chromothripsis from DNA damage in micronuclei. Nature, 522(7555), 179–184.