Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/7d8etw49

A COMPARATIVE STUDY OF SYSTEMS PHARMACOLOGY AND GENE EXPRESSION MODELS FOR PREDICTING DRUG TARGETS

Dr. Amar Jeeth Raja R*

*Department of Pharmacology Assistant Professor Sri Lakshmi Narayana Institute of Medical Sciences, Osudu, Agaram Village, Villianur Commune, Kudapakkam Post, Puducherry- 605502
*Email:ID; amarjeeth@gmail.com

ABSTRACT

Background: Accurate prediction of drug targets remains a fundamental challenge in pharmaceutical research and drug discovery. Systems pharmacology and gene expression-based approaches represent two predominant computational strategies for target identification, yet comprehensive comparisons of their predictive performance are lacking. This study aimed to systematically compare the accuracy, sensitivity, and specificity of systems pharmacology models versus gene expression models in predicting validated drug targets.

Methods: This comparative computational study analyzed 1,247 drug-target interactions from established pharmacological databases. Systems pharmacology models incorporating network topology, pathway enrichment, and protein-protein interaction data were compared against gene expression models utilizing differential expression analysis and co-expression network approaches. Performance metrics including accuracy, sensitivity, specificity, positive predictive value, and area under the receiver operating characteristic curve (AUC-ROC) were calculated. Validation was performed using an independent dataset of 312 experimentally confirmed drug-target pairs.

Results: Systems pharmacology models demonstrated superior overall accuracy (78.4 ± 5.2%) compared to gene expression models (71.6 ± 6.8%; $p<0.001$). The AUC-ROC values were 0.847 ± 0.042 for systems pharmacology and 0.793 ± 0.051 for gene expression approaches. However, gene expression models showed higher sensitivity for detecting novel targets (82.3% vs. 74.6%; $p=0.008$). Integration of both approaches yielded optimal performance, achieving 86.2 ± 4.1% accuracy and AUC-ROC of 0.912 ± 0.031. The integrated model successfully predicted 89.4% of validated targets in the independent dataset.

Conclusion: While systems pharmacology models demonstrate superior overall accuracy for drug target prediction, gene expression approaches offer complementary advantages in novel target discovery. Integrated computational frameworks combining both methodologies provide optimal predictive performance and should be prioritized in drug discovery pipelines.

Keywords: Drug target prediction; systems pharmacology; gene expression; computational pharmacology; network analysis; drug discovery

INTRODUCTION

The identification and validation of drug targets represents a critical bottleneck in pharmaceutical research, with significant implications for therapeutic development success rates and associated costs **[1]**. Traditional experimental approaches for target discovery are time-consuming, resource-intensive, and frequently yield high attrition rates during clinical development phases **[2]**.

Consequently, computational methodologies have emerged as indispensable tools for prioritizing potential drug targets and guiding experimental validation efforts **[3]**.

Systems pharmacology has evolved as a comprehensive framework for understanding drug action within the context of biological networks **[4]**. This approach integrates diverse data types, including protein-protein interactions, signaling pathways, and metabolic networks, to model the complex relationships between drugs, targets, and disease phenotypes **[5]**. Network-based methods have demonstrated considerable success in identifying novel therapeutic targets and predicting drug opportunities **[6]**. The seminal work by Hopkins and colleagues repurposing established the theoretical foundations for network pharmacology, emphasizing the importance of polypharmacology and multi-target drug design **[7]**.

ntly, gene expression profiling has revolutionized target identification by enabling genome-wide assessment of transcriptional responses to pharmacological perturbations **[8]**. The Connectivity Map project pioneered the systematic use of gene expression signatures to establish connections between drugs, genes, and diseases **[9]**. Subsequent investigations have refined expression-based approaches, incorporating machine learning algorithms and advanced statistical methods to improve predictive accuracy **[10]**. Differential expression analysis and co-expression network construction have become standard methodologies for inferring drug mechanisms and identifying target candidates **[11]**.

Despite the widespread application of both systems pharmacology and gene expression approaches, direct comparative evaluations of their predictive performance remain surprisingly limited. Previous studies have typically focused on individual methodologies or specific disease contexts, precluding generalizable conclusions regarding relative strengths and limitations **[12]**. Furthermore, the potential synergies between these complementary approaches have not been systematically explored, representing a significant gap in the computational pharmacology literature.

The selection of appropriate computational strategies for target prediction has substantial practical implications. Pharmaceutical companies and academic research groups must allocate finite resources to experimental validation, making accurate prioritization essential **[13]**. Understanding the comparative performance characteristics of available methodologies enables informed decision-making and potentially accelerates drug discovery timelines.

This study aimed to comprehensively compare systems pharmacology and gene expression models for predicting drug targets, evaluating their respective accuracies, sensitivities, and specificities across diverse pharmacological contexts. Additionally, we investigated whether integrating both approaches could enhance predictive performance beyond either methodology alone.

Materials and Methods

Study Design and Data Sources

This comparative computational study utilized drug-target interaction data compiled from multiple established pharmacological databases. Primary data sources included Drug Bank, Therapeutic Target Database (TTD), and ChEMBL, accessed between January 2011 to December 2012. A total of 1,247 validated drug-target interactions encompassing 423 unique drugs and 687 unique protein targets were assembled for model development and evaluation.

Inclusion and Exclusion Criteria

Drug-target pairs were included if they met the following criteria: (1) experimental validation through binding assays, functional studies, or clinical observations; (2) target protein with available structural and functional annotation; (3) drug compound with known chemical structure and pharmacological classification; (4) availability of corresponding gene expression data in public repositories. Exclusion criteria comprised: (1) interactions based solely on computational predictions without experimental confirmation; (2) targets with incomplete pathway annotation; (3) drugs withdrawn from market due to safety concerns unrelated to target engagement.

Systems Pharmacology Model Construction

The systems pharmacology framework integrated three complementary analytical components.

Network topology analysis utilized protein-protein interaction data from STRING database (confidence score >0.7) to construct drug-target-disease networks. Centrality measures including degree, betweenness, and closeness centrality were calculated for each potential target. Pathway enrichment analysis employed KEGG and Reactome databases to assess functional associations between candidate targets and disease-relevant pathways. A composite systems pharmacology score (SPS) was calculated using weighted combinations of network and pathway metrics, with weights optimized through cross-validation procedures.

Gene Expression Model Development

Gene expression models were constructed using transcriptional data from the Gene Expression Omnibus (GEO) repository. Differential expression analysis was performed using limma package, with significance thresholds of adjusted p-value <0.05 and absolute log2 fold-change >1.0. Co-expression networks were generated using weighted gene co-expression network analysis (WGCNA), identifying modules associated with drug response. A gene expression prediction score (GEPS) was derived from differential expression magnitude, co-expression module membership, and connectivity metrics.

Integrated Model Framework

An integrated prediction model combining systems pharmacology and gene expression features was developed using random forest classification. Feature selection employed recursive feature elimination with 10-fold cross-validation. The integrated model incorporated 15 systems pharmacology features and 12 gene expression features, selected based on variable importance rankings.

Validation Dataset

An independent validation dataset comprising 312 experimentally confirmed drug-target pairs, published subsequent to model training, was utilized for external validation. This dataset included 156 positive interactions (confirmed targets) and 156 negative controls (experimentally validated non-targets).

Statistical Analysis

Model performance was evaluated using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC-ROC). Comparisons between models employed paired t-tests for continuous metrics and McNemar's test for classification concordance. All analyses were performed using R statistical software (version 4.2.1). Statistical significance was defined as $p<0.05$, with Bonferroni correction applied for multiple comparisons.

Results

Dataset Characteristics

The final analytical dataset comprised 1,247 drug-target interactions involving 423 drugs across 12 therapeutic categories and 687 protein targets spanning 24 functional classes. Kinases represented the largest target class (23.4%), followed by G-protein coupled receptors (18.7%) and ion channels (12.3%). Oncology drugs constituted the predominant therapeutic category (28.6%), followed by cardiovascular agents (19.2%) and central nervous system drugs (16.8%).

Performance Comparison of Individual Models

Systems pharmacology models demonstrated significantly higher overall accuracy compared to gene expression models across the primary dataset. The mean accuracy for systems pharmacology was 78.4 ± 5.2%, compared to 71.6 ± 6.8% for gene expression approaches ($p<0.001$). Specificity was notably higher for systems pharmacology models (81.2 ± 4.8% vs. 68.4 ± 7.2%; $p<0.001$), whereas gene expression models exhibited superior sensitivity (82.3 ± 5.6% vs. 74.6 ± 6.1%; $p=0.008$).

Table 1: Performance Metrics of Systems Pharmacology and Gene Expression Models

Performance Metric	Systems Pharmacology	Gene Expression	p-value
Accuracy (%)	78.4 ± 5.2	71.6 ± 6.8	<0.001
Sensitivity (%)	74.6 ± 6.1	82.3 ± 5.6	0.008
Specificity (%)	81.2 ± 4.8	68.4 ± 7.2	<0.001
PPV (%)	79.8 ± 5.4	72.1 ± 6.5	0.003
NPV (%)	76.4 ± 5.9	79.6 ± 5.2	0.142
AUC-ROC	0.847 ± 0.042	0.793 ± 0.051	<0.001

Values expressed as mean ± SD; PPV: Positive Predictive Value; NPV: Negative Predictive Value; AUC-ROC: Area Under Receiver Operating Characteristic Curve

Target Class-Specific Performance

Model performance varied substantially across different target classes. Systems pharmacology demonstrated superior performance for kinases (accuracy: 84.2 ± 4.1%) and nuclear receptors (accuracy: 82.7 ± 3.8%), while gene expression models performed better for transporters (accuracy: 78.4 ± 5.2%) and metabolic enzymes (accuracy: 76.9 ± 4.8%).

Table 2: Model Accuracy by Target Protein Class

Target Class	n	Systems Pharmacology (%)	Gene Expression (%)	Integrated Model (%)
Kinases	161	84.2 ± 4.1	73.8 ± 5.9	89.4 ± 3.2
GPCRs	129	79.6 ± 5.3	74.2 ± 6.4	86.8 ± 4.1
Ion Channels	85	76.4 ± 6.2	69.8 ± 7.1	83.5 ± 4.8
Nuclear Receptors	72	82.7 ± 3.8	71.4 ± 5.8	88.2 ± 3.6
Transporters	68	71.8 ± 5.9	78.4 ± 5.2	84.6 ± 4.4
Metabolic Enzymes	94	73.2 ± 6.1	76.9 ± 4.8	85.3 ± 3.9
Proteases	78	77.8 ± 5.4	70.6 ± 6.7	85.9 ± 4.2

Values expressed as mean ± SD; GPCRs: G-Protein Coupled Receptors

Integrated Model Performance

The integrated model combining systems pharmacology and gene expression features demonstrated significantly improved performance compared to either individual approach. Overall accuracy reached 86.2 ± 4.1%, representing improvements of 7.8 percentage points over systems pharmacology alone and 14.6 percentage points over gene expression alone (both $p<0.001$). The AUC-ROC for the integrated model was 0.912 ± 0.031, significantly exceeding both individual approaches.

External Validation Results

In the independent validation dataset of 312 drug-target pairs, the integrated model correctly classified 279 interactions (89.4% accuracy). Systems pharmacology alone achieved 76.9% accuracy, while gene expression models achieved 69.2% accuracy in this validation cohort.

Table 3: External Validation Performance

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC-ROC
Systems Pharmacology	76.9	73.1	80.8	0.824
Gene Expression	69.2	79.5	59.0	0.768
Integrated Model	89.4	88.5	90.4	0.926

Feature Importance Analysis

Within the integrated model, network centrality measures from systems pharmacology contributed most substantially to predictive performance. Betweenness centrality ranked highest in variable importance (mean decrease in Gini: 0.142), followed by pathway enrichment score (0.128) and differential expression magnitude (0.116). Co-expression module connectivity ranked fourth (0.098), indicating the complementary contributions of both methodological approaches.

Discussion

This comprehensive comparative study demonstrates distinct performance characteristics of systems pharmacology and gene expression models for drug target prediction, while establishing the superior accuracy of integrated approaches. These findings have significant implications for computational drug discovery strategies and resource allocation in target validation efforts.

The observed superiority of systems pharmacology models in overall accuracy aligns with theoretical expectations regarding the information content of network-based approaches **[14]**. By incorporating protein-protein interaction topology and pathway relationships, systems pharmacology captures functional context that pure expression-based methods may overlook. The particularly strong performance for kinase and nuclear receptor targets likely reflects the well-characterized network properties of these signaling proteins, enabling accurate inference from connectivity patterns **[15]**.

Conversely, the higher sensitivity of gene expression models for detecting novel targets suggests their utility in discovery-oriented applications. Transcriptional responses to pharmacological perturbations provide direct readouts of cellular effects, potentially identifying targets through mechanism-agnostic approaches **[16]**. The superior performance of expression models for transporters and metabolic enzymes may reflect the strong transcriptional regulation of these protein classes, making them particularly amenable to expression-based detection **[17]**.

The substantial performance improvement achieved through model integration corroborates emerging paradigms emphasizing multi-modal data integration in computational biology **[18]**. The complementary information captured by network topology and transcriptional responses yields synergistic predictive power exceeding either approach alone. Similar integration strategies have proven successful in related domains, including drug response prediction and disease gene prioritization **[19]**.

Our findings regarding target class-specific performance have practical implications for methodology selection. When investigating kinase or nuclear receptor targets, systems pharmacology approaches merit prioritization, whereas expression-based methods may be preferable for transporter and enzyme targets **[20]**. Such stratified analytical strategies could optimize resource utilization in drug discovery programs.

The robust validation performance of the integrated model (89.4% accuracy) in an independent dataset provides confidence in generalizability. This external validation addresses potential overfitting concerns inherent to computational model development and supports prospective application in target prioritization efforts **[21]**.

Several limitations warrant acknowledgment. The reliance on previously validated drug-target interactions may introduce biases toward well-characterized target classes. Additionally, gene expression data availability varied across drug categories, potentially influencing comparative assessments. Future investigations incorporating additional data modalities, including proteomic and

metabolomic profiles, may further enhance predictive performance **[22]**.
The computational frameworks developed herein are readily implementable using publicly available databases and open-source software tools, facilitating widespread adoption. Integration of these approaches into drug discovery pipelines could accelerate target identification and reduce experimental validation costs **[23]**.

Conclusion

This comparative study establishes that systems pharmacology and gene expression models possess distinct and complementary strengths for drug target prediction. Systems pharmacology demonstrates superior overall accuracy and specificity, particularly for well-characterized signaling proteins including kinases and nuclear receptors. Gene expression approaches offer advantages in sensitivity and perform optimally for metabolic targets subject to strong transcriptional regulation. The integration of both methodologies yields substantially improved predictive performance, achieving nearly 90% accuracy in external validation. These findings support the adoption of integrated computational frameworks in drug discovery pipelines, enabling more efficient prioritization of candidate targets for experimental validation. Future developments incorporating additional omics data modalities may further enhance predictive capabilities, accelerating the identification of novel therapeutic targets.

References

1. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008;4(11):682-690. doi:10.1038/nchembio.118. PMID: 18936753
2. Sams-Dodd F. Target-based drug discovery: is something wrong? Drug Discov Today. 2005;10(2):139-147. doi:10.1016/S1359-6446(04)03316-1. PMID: 15718163
3. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. Nat Biotechnol. 2004;22(10):1253-1259. doi:10.1038/nbt1017. PMID: 15470465
4. Cooke EJ, Savage RS, Wild DL. Computational approaches to the integration of gene expression, ChIP-chip and sequence data in the inference of gene regulatory networks. Semin Cell Dev Biol. 2009 Sep;20(7):863-8. doi: 10.1016/j.semcdb.2009.08.004. Epub 2009 Aug 12. PMID: 19682595.
5. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008 Apr;18(4):644-52. doi: 10.1101/gr.071852.107. PMID: 18381899; PMCID: PMC3863981.
6. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007;25(10):1119-1126. doi:10.1038/nbt1338. PMID: 17921997
7. Hopkins AL. Drug discovery: predicting promiscuity. Nature. 2009;462(7270):167-168. doi:10.1038/462167a. PMID: 19907487
8. Lamb J. The Connectivity Map: a new tool for biomedical research. Nat Rev Cancer. 2007;7(1):54-60. doi:10.1038/nrc2044. PMID: 17186018
9. Lamb J, Crawford ED, Peck D, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929-1935. doi:10.1126/science.1132939. PMID: 17008526
10. Iorio F, Bosotti R, Scacheri E, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci USA. 2010;107(33):14621-14626. doi:10.1073/pnas.1000138107. PMID: 20679242
11. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4:Article17. doi:10.2202/1544-6115.1128. PMID: 16646834
12. Keiser MJ, Setola V, Irwin JJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462(7270):175-181. doi:10.1038/nature08506. PMID: 19881490
13. Arrell DK, Terzic A. Network systems biology for drug discovery. Clin Pharmacol Ther. 2010;88(1):120-125. doi:10.1038/clpt.2010.91. PMID: 20520602
14. Csermely P, Agoston V, Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci. 2005;26(4):178-182.

doi:10.1016/j.tips.2005.02.007. PMID: 15808341
15. Mencher SK, Wang LG. Promiscuous drugs compared to selective drugs (promiscuity can be a virtue). BMC Clin Pharmacol. 2005;5:3. doi:10.1186/1472-6904-5-3. PMID: 15854222
16. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004 Aug;3(8):673-83. doi: 10.1038/nrd1468. PMID: 15286734.
17. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. Science. 2008;321(5886):263-266. doi:10.1126/science.1158140. PMID: 18621671
18. Pujol A, Mosca R, Farrés J, Aloy P. Unveiling the role of network and systems biology in drug discovery. Trends Pharmacol Sci. 2010;31(3):115-123. doi:10.1016/j.tips.2009.11.006. PMID: 20117850
19. Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics. 2009 Sep 15;25(18):2397-403. doi: 10.1093/bioinformatics/btp433. Epub 2009 Jul 15. PMID: 19605421; PMCID: PMC2735674.
20. Berger SI, Iyengar R. Network analyses in systems pharmacology. Bioinformatics. 2009;25(19):2466-2472. doi:10.1093/bioinformatics/btp465. PMID: 19648136
21. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics. 2008;24(13):i232-i240. doi:10.1093/bioinformatics/btn162. PMID: 18586719
22. van der Greef J, McBurney RN. Innovation: Rescuing drug discovery: in vivo systems pathology and systems pharmacology. Nat Rev Drug Discov. 2005;4(12):961-967. doi:10.1038/nrd1904. PMID: 16341063
23. Fliri AF, Loging WT, Thadeio PF, Bhati RA. Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. J Med Chem. 2005;48(22):6918-6925. doi:10.1021/jm050494g. PMID: 16250651.