Integrative Network Pharmacology and Transcriptomic Analysis for Discovering Potential Therapeutic Combinations in Cardiovascular Diseases

Felicidad Abascal
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
contactfelicidad@colostate.edu

Norah Spreeuw
School of Computing, Clemson University, Clemson, SC, USA.
norah.spreeuw@clemson.edu

Abstract

Cardiovascular diseases remain the largest single contributor to premature mortality, and much of contemporary secondary prevention already depends on drug combinations whose joint behaviour is anticipated rather than predicted. Network pharmacology and transcriptomic connectivity analysis each supply part of the evidence needed to prioritise combinations rationally, but they are usually applied in isolation, and their scores are concatenated without a principled account of how mechanistic proximity, transcriptional directionality, and target-set redundancy should trade against one another. This paper develops an integrative scoring framework for cardiovascular combination discovery that treats the two evidence layers as complementary measurements of the same underlying perturbation. Disease modules are defined on a confidence-filtered human interactome, drug target sets are scored by degree-preserving network proximity, and compound-induced transcriptional signatures are scored by a normalised reversal statistic computed against a tissue-matched disease signature. A complementarity term quantifies module coverage net of target overlap, and the three components enter a logistic aggregation whose parameters are design weights rather than fitted coefficients. The framework contributes explicit null models for both layers, a Lipschitz sensitivity bound that converts component uncertainty into a bound on ranking stability, and a constrained selection step that enforces mechanistic diversity across a shortlist. Boundary conditions are examined in detail: interactome incompleteness, tissue specificity of cardiovascular signatures, the additivity assumption implicit in combining transcriptional profiles, and the gap between computational synergy priors and the Loewe, Bliss, and Chou–Talalay definitions used experimentally. The result is a reproducible analytical protocol that specifies what evidence a candidate combination must accumulate before it warrants experimental testing in cardiomyocyte or organoid systems.

Keywords

network pharmacology; transcriptomic connectivity; drug combination synergy; cardiovascular disease; disease module; constrained ranking

1. Introduction

Cardiovascular diseases (CVDs) remain the leading contributor to premature mortality worldwide, with ischaemic heart disease and cerebrovascular disease accounting for a large

share of the global burden and a growing proportion of it in low- and middle-income settings [1]. Contemporary prevention and treatment already rest on combination pharmacotherapy: statins with antiplatelet agents after acute coronary syndrome, renin–angiotensin system blockade with beta-adrenergic blockade and mineralocorticoid receptor antagonism in heart failure with reduced ejection fraction, and increasingly early combination of lipid-lowering and glucose-lowering agents in patients with cardiometabolic disease [2]. These regimens were assembled through large outcome trials in which the joint effect of the components was observed rather than mechanistically anticipated. That empirical route is expensive, slow, and poorly suited to the far larger space of plausible pairs and triples now available.

Two computational traditions have grown up around the problem of anticipating drug action before trial. Network pharmacology treats a drug as a perturbation of a protein interaction system and asks where its targets sit relative to the molecular neighbourhood in which a disease phenotype is generated [3]. Network medicine formalised this intuition, showing that disease-associated genes are not randomly scattered through the interactome but cluster into connected modules, and that the topology of those modules carries information about comorbidity and mechanism [4]. The drug–target bipartite network supplied an early quantitative language for promiscuity and target sharing [5]. A second tradition reads drug action through transcriptional state. The Connectivity Map established that compound-induced expression signatures can be compared across compounds, genes, and diseases, and that reversal of a disease signature is a usable proxy for therapeutic direction [6]; the L1000 platform scaled this idea to more than a million profiles and made signature-level matching a routine analytical operation [7].

Combination prediction has drawn on both traditions but rarely on both at once. Network-based approaches have modelled combinations as joint target sets and predicted efficacy from the proximity of the union to a disease module, with strong performance in oncology settings [8]. Community-based formulations extend this logic to complex multi-component systems, identifying combinations whose target sets bridge distinct network communities rather than saturating a single one [9]. Transcriptomic approaches have complemented these by asking whether the combined perturbation reverses the disease signature more completely than either component alone. Each family of methods produces a defensible but partial ranking, and the two rankings do not agree reliably enough to be merged by simple averaging.

This paper addresses that integration problem directly. The central contribution is a four-layer scoring framework in which disease-module proximity, transcriptional reversal, target-set complementarity, and mechanistic redundancy are estimated on commensurate scales, combined through an explicitly parameterised aggregation, and then filtered by a diversity-constrained selection rule. The framework is not presented as a validated screen; it is presented as an analytical protocol whose assumptions, null models, sensitivity properties, and failure modes are made explicit so that it can be applied to cardiovascular datasets and audited afterwards. Particular attention is given to the properties that distinguish cardiovascular applications from oncology applications: the tissue specificity of myocardial signatures, the chronic and multifactorial character of the relevant disease modules, and the fact that many cardiovascular drugs act on neurohumoral signalling rather than on a discrete oncogenic driver.

2. Mechanistic and Computational Foundations

2.1 Disease modules and network proximity

The premise of module-based reasoning is empirical rather than axiomatic. Disease-associated gene products show a consistent tendency to lie closer to one another in the human interactome than degree-matched random gene sets, and this tendency holds across most major disease classes when the interactome is sufficiently complete [11]. The observation licenses a proximity statistic: if a drug's targets fall inside or adjacent to the disease module, the drug is more likely to perturb the relevant system than if its targets lie in an unrelated region. Guney and colleagues operationalised this idea as an efficacy screen and demonstrated that proximity between a drug's target set and a disease module separates known efficacious drugs from matched non-indications across a range of diseases [10]. The same reasoning underlies combination hypotheses, with the added nuance that a pair may be attractive precisely because the two target sets enter the module from different directions.

2.2 Transcriptional connectivity

Signature reversal rests on a different evidential base. If a disease state is characterised by a vector of expression changes relative to healthy tissue, then a compound whose own transcriptional effect points in the opposite direction is a candidate therapeutic, and one that reinforces the disease signature is a candidate for harm or for adverse remodelling [6]. The logic scales because signature comparison does not require the compound's direct targets to be known, which is valuable for natural products and for drugs with uncharacterised polypharmacology. It also carries an important asymmetry: reversal is directional evidence about phenotype, not evidence about mechanism, and a compound can reverse a signature through an off-target effect that is irrelevant or deleterious in the myocardium.

2.3 What synergy means computationally

Combination analysis inherits a definitional problem from pharmacology. Loewe additivity defines the expected effect of a combination as the effect of a hypothetical single agent with dose-equivalent potency, and departure from that expectation defines synergy or antagonism [12]. Bliss independence instead compares the observed combined effect with the product of the individual effects under an assumption of independent action [13]. The Chou–Talalay combination index generalises the Loewe framework to dose–response surfaces and remains the most widely used experimental summary [14]. These definitions are not interchangeable, and they diverge most sharply for agents that share a downstream pathway. A computational framework cannot resolve which definition applies; what it can do is state which quantity its score is intended to approximate and avoid implying that an in silico score is measurement of synergy. In what follows, the network and transcriptional terms are treated as priors on the probability that a pair will show favourable interaction under a Loewe-type expectation at the phenotypic level, not as estimates of a combination index.

3. Integrative Framework: Data Layers and Network Construction

The framework takes four inputs: a disease expression matrix with case and control samples from a cardiovascular tissue, a curated drug–target resource, a confidence-scored human interactome, and a pathway or gene-set collection. Drug–target annotations are drawn from a curated database that records both approved and investigational targets and distinguishes primary from secondary assignments [15]. The interactome is assembled from a consolidated protein association resource, retaining only edges above a confidence threshold and restricting analysis to the largest connected component to avoid path-length artefacts caused by disconnected fragments [16].

Differential expression is estimated with moderated linear models for microarray or count-based data, with empirical Bayes variance shrinkage and Benjamini–Hochberg control of the false discovery rate [18, 19]. Where the disease tissue is heterogeneous, co-expression module detection is applied before differential testing so that cell-composition-driven covariation can be separated from coordinated pathway-level change [17]. The disease module is then the set of genes passing joint significance and effect-size criteria:

$$D = \{g \in V : q_g \leq \tau_q, |\beta_g| \geq \tau_\beta\}$$

Here V is the vertex set of the interactome, q_g is the adjusted p-value for gene g, β_g is its estimated log fold change, and τ_q and τ_β are pre-specified thresholds. The joint criterion matters: significance alone admits genes with negligible effect in a large cohort, and effect size alone admits noisy genes with unstable estimates. Genes passing the filter but absent from the interactome are retained for the transcriptional layer and excluded from the network layer, and this discrepancy should be reported rather than silently resolved.

4. Scoring Model for Candidate Combinations

4.1 Network proximity of a target set

For a compound c with target set T, proximity to the disease module is the mean shortest-path distance from each target to its nearest module gene:

$$d(T, D) = (1/|T|) \sum_\{t \in T\} min_\{g \in D\} dist(t, g)$$

The distance dist(t,g) is the number of edges on the shortest path between t and g, and targets absent from the largest connected component are treated as maximally distant. The mean rather than the minimum is used because a single well-placed target should not carry an otherwise poorly positioned drug; the min term inside the sum allows each target its own nearest entry point into the module, which accommodates multi-target drugs whose targets engage different subregions.

Raw distances are not comparable across drugs because target sets differ in size and in degree profile. Normalisation against a degree-preserving null model is therefore required:

$$z_net(T, D) = \frac{\big(d(T, D) - \mu_0(T, D)\big)}{\sigma_0}(T, D)$$

The null mean μ₀ and standard deviation σ₀ are estimated by resampling target sets of the same cardinality from interactome vertices matched on degree and on expression variance in the disease tissue. Matching on expression variance is a deliberate departure from purely topological randomisation: highly expressed, highly variable genes occupy denser regions of the network, and unmatched nulls produce systematically deflated z-scores for drugs whose targets happen to be well studied.

4.2 Transcriptional reversal

Let S be the disease signature, taken as the union of the module D and the leading genes of dysregulated hallmark and pathway sets relevant to the tissue [20, 21], each gene weighted by w_g, a decreasing function of its adjusted p-value. Let β_g be the disease log fold change and x_{c,g} the normalised compound-induced log fold change for compound c in a matched cell context. Reversal is the negative normalised inner product:

$$R(c) = -\sum_\{g \in S\} w_g \, \beta_g \, x_\{c, g\} \,/\, \sum_\{g \in S\} w_g \, |\beta_g \, x_\{c, g\}|$$

The normalisation bounds R(c) in the interval [−1, 1], with positive values indicating reversal and negative values indicating reinforcement of the disease signature. The denominator vanishes only if no gene in S responds to the compound, in which case the score is undefined and the compound is excluded rather than assigned a neutral value; a neutral assignment would let unresponsive compounds compete on network evidence alone.

4.3 Complementarity and redundancy

A pair is more interesting when the two target sets reach different parts of the module. Let N_1 and N_2 denote the one-step interactome neighbourhoods of the two target sets. Complementarity combines module coverage with a penalty for overlap:

$$C = |(N_1 \cup N_2) \cap D| / |D| - \lambda |N_1 \cap N_2| / |N_1 \cup N_2|$$

The first term is the fraction of the disease module reached by at least one partner; the second is the Jaccard overlap of the two neighbourhoods. The parameter λ in [0, 1] sets the price of redundancy, and at λ = 1 a fully overlapping pair receives a complementarity of zero regardless of coverage. This term encodes the mechanistic claim that two drugs hitting the same nodal neighbourhood are unlikely to produce more than dose additivity, whereas two drugs entering a module at distinct points can produce genuine pathway-level complementarity, a rationale consistent with systems-pharmacology accounts of multi-target action [22].

4.4 Aggregation

The three standardised components enter a logistic aggregation. For a pair (c_1, c_2), the network term is the weaker of the two partners, z_pair = min(z_net(T_1), z_net(T_2)), on the reasoning that a combination inherits the proximity of its less well-positioned arm, and the transcriptional term is the mean of the individual reversal scores under an assumption of signature-level additivity:

$$\pi(c_1, c_2) = \Lambda(\theta_0 + \theta_1 z_net, min + \theta_2 R + \theta_3 C)$$

Here Λ(η) = 1/(1 + e^(−η)) is the logistic function and θ_0 through θ_3 are design weights, fixed a priori from the relative reliability of each layer or estimated by nested cross-validation in a prospective application. The framework does not assume specific values, and the scores should be read as ordinal priorities rather than calibrated probabilities unless calibration data exist.

5. Analytical Pipeline, Null Models, and Decision Rules

5.1 Composed null models

Each component carries its own null. The network null is degree- and expression-matched resampling; the transcriptional null is permutation of sample labels in the disease cohort with re-estimation of β_g, which propagates differential-expression uncertainty into R(c) rather than treating the signature as fixed. The complementarity term requires a third null in which target neighbourhoods are rewired within the interactome while preserving module size. These nulls must be composed rather than applied independently: a pair that is extreme on two marginally significant components is not thereby jointly significant, and a joint permutation in which all three components are recomputed under a single randomisation scheme is the defensible route. Reported rankings should carry an empirical joint p-value alongside the composite score.

5.2 Sensitivity of the ranking

Because θ is a design choice, the stability of the ranking under perturbation of the weights and of the component scores must be quantified. The logistic derivative satisfies $\Lambda'(\eta) = \Lambda(\eta)(1 - \Lambda(\eta)) \leq 1/4$, so a first-order bound follows directly:

$$|\Delta\pi| \;\leq\; ¼ \sum\nolimits_k |\theta_k|\,|\Delta s_k|$$

The index k runs over the three components s_k with perturbations Δs_k. The bound is conservative but useful in practice: it converts the width of a confidence interval on z_net, on $\bar{R}$, and on C into a guaranteed interval on π, so pairs whose intervals overlap can be flagged as rank-indistinguishable instead of being ordered arbitrarily. Reporting the shortlist as tiers of indistinguishable pairs is more honest than reporting a strict ordering.

5.3 Diversity-constrained selection

Ranking alone yields shortlists dominated by large target neighbourhoods and by well-characterised drugs. A selection step imposes mechanistic diversity by capping the number of selected combinations that reuse the same target, pathway, or gene-set element e:

$$max \sum\nolimits_{k \in K} \pi_k\, y_k \; s.t. \sum\nolimits_{k:\, e \in k} y_k \;\leq\; n_e$$

The binary variable y_k selects candidate pair k from the pool K, and n_e caps reuse of element e. This is a small integer program solvable exactly for realistic pool sizes. Its purpose is not optimality in an abstract sense but coverage of distinct mechanistic hypotheses, which is what an experimental follow-up programme can actually test in parallel.

6. Technical Discussion and Boundary Conditions

Interactome incompleteness is the most consequential limitation of the network layer. Shortest paths computed on a partial interactome overestimate distances, and the overestimation is not uniform: well-studied proteins accumulate edges while under-studied membrane and mitochondrial proteins remain sparsely connected, biasing proximity in favour of established targets [11]. Degree- and expression-matched nulls mitigate but do not remove the bias, and any cardiovascular application should report the fraction of target sets with members outside the largest connected component.

Tissue specificity is the central limitation of the transcriptional layer. A compound's signature in an immortalised line may bear little relation to its effect in adult cardiomyocytes, and a whole-blood disease signature in atherosclerosis captures immune activation rather than plaque biology. Where matched cardiovascular contexts exist, they should be preferred; where they do not, the reversal score should be down-weighted rather than transferred without comment. The additivity assumption embedded in $\bar{R}$ is a further constraint: combining two compounds that share a mechanism, or that interact pharmacodynamically, violates the assumption, and such pairs should be identified through the overlap penalty in C before the transcriptional term is trusted. Transcriptional profiling is nonetheless a productive route to mechanism-of-action assignment and repurposing hypotheses precisely because it does not require prior target knowledge [23].

The relationship between a computational prior and experimental synergy deserves explicit statement. A high composite score asserts that a pair is mechanistically plausible, directionally coherent, and complementary. It does not assert a combination index below unity. Confirmation requires dose–response matrices in a cardiovascular model system and formal testing against Loewe or Bliss expectations [12, 13, 14]. The appropriate claim for a computational study is that the framework concentrates experimental effort on a smaller,

mechanistically annotated set of pairs, and that claim is falsifiable by measuring the hit rate of an experimentally tested shortlist against a randomly drawn control set of comparable size.

Machine-learning extensions should be evaluated with the same discipline. Supervised models trained on combination response data can learn powerful predictors, but they generalise poorly across tissue contexts and are vulnerable to leakage when drugs or scaffolds appear in both training and test partitions [24]. A defensible evaluation splits by compound scaffold and by disease context, not by random row assignment, and reports performance against a trivial baseline that ranks by single-agent efficacy. The framework presented here is deliberately interpretable rather than predictive in the black-box sense, because the practical value of a cardiovascular shortlist lies in the mechanistic annotation that accompanies each pair, not only in its rank.

7. Conclusion

The problem addressed in this paper is not a shortage of candidate compounds but a shortage of principled ways to combine heterogeneous evidence about them. Network proximity says where a target set sits relative to a disease module; transcriptional reversal says whether a compound moves the myocardium's expression state in the therapeutic direction; complementarity says whether two compounds reach the module by different routes. Each is informative, none is sufficient, and naive averaging of the three obscures exactly the disagreements that matter most. The framework developed here makes the aggregation explicit, supplies a null model for each component and a composed null for the joint score, bounds the effect of component uncertainty on the final ranking, and adds a diversity constraint so that a shortlist samples distinct mechanistic hypotheses rather than repeatedly rewarding the same well-studied neighbourhood. Its assumptions are stated as assumptions: additivity of transcriptional signatures, comparability of degree-matched targets, and a proximity–efficacy relationship inherited from the network medicine literature rather than re-derived here. Applied to cardiovascular transcriptomes and target annotation, the protocol yields a ranked, annotated, and uncertainty-quantified set of pairs suitable for testing in cardiomyocyte, organoid, or animal models, together with the specific failure modes that experimental follow-up should interrogate.

References

1. Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., … GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study. Journal of the American College of Cardiology, 76(25), 2982–3021.

2. Arnett, D. K., Blumenthal, R. S., Albert, M. A., Buroker, A. B., Goldberger, Z. D., Hahn, E. J., Himmelfarb, C. D., Khera, A., Lloyd-Jones, D., McEvoy, J. W., Michos, E. D., Miedema, M. D., Muñoz, D., Smith, S. C., Virani, S. S., Williams, K. A., Yeboah, J., & Ziaeian, B. (2019). 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: A report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation, 140(11), e596–e646.

3. Hopkins, A. L. (2008). Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology, 4(11), 682–690.

4. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56–68.

5. Yıldırım, M. A., Goh, K.-I., Cusick, M. E., Barabási, A.-L., & Vidal, M. (2007). Drug-target network. Nature Biotechnology, 25(10), 1119–1126.

6. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science, 313(5795), 1929–1935.

7. Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., Gould, J., Davis, J. F., Tubelli, A. A., Asiedu, J. K., Lahr, D. L., Hirschman, J. E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I. C., Lam, D., … Golub, T. R. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell, 171(6), 1437–1452.

8. Cheng, F., Kovács, I. A., & Barabási, A.-L. (2019). Network-based prediction of drug combinations. Nature Communications, 10, 1197.

9. HerbSyner_Finder: a network community-based model for identifying synergistic combinations from herbal medicines and complex systems. Targetome. 2026;2(2):0-0. doi: 10.48130/targetome-0026-001

10. Guney, E., Menche, J., Vidal, M., & Barabási, A.-L. (2016). Network-based in silico drug efficacy screening. Nature Communications, 7, 10331.

11. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A.-L. (2015). Uncovering disease-disease relationships through the incomplete interactome. Science, 347(6224), 1257601.

12. Loewe, S. (1953). The problem of synergism and antagonism of combined drugs. Arzneimittel-Forschung, 3(6), 285–290.

13. Bliss, C. I. (1939). The toxicity of poisons applied jointly. Annals of Applied Biology, 26(3), 585–615.

14. Chou, T.-C., & Talalay, P. (1984). Quantitative analysis of dose-effect relationships: The combined effects of multiple drugs or enzyme inhibitors. Advances in Enzyme Regulation, 22, 27–55.

15. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074–D1082.

16. Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research, 49(D1), D605–D612.

17. Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

18. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43(7), e47.

19. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

20. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database hallmark gene set collection. Cell Systems, 1(6), 417–425.

21. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30.

22. Zhao, S., & Iyengar, R. (2012). Systems pharmacology: Network analysis to identify multiscale mechanisms of drug action. Annual Review of Pharmacology and Toxicology, 52, 505–521.

23. Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., & di Bernardo, D. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. Proceedings of the National Academy of Sciences, 107(33), 14621–14626.

24. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477.