

1 **Systematic bias in network proximity Z-scores: A**
2 **comparative robustness audit using *Hypericum perforatum***
3 **constituents**

4 Antony Bevan^{1,*}

¹Department of Pharmacognosy, Madurai Medical College, Madurai, India

*Corresponding author: antonybevan04@gmail.com

5 **Abstract**

6 Network-based metrics are widely used to identify associations between compounds and diseases, as-
7 suming that proximity within a protein–protein interaction network reflects functional relevance. How-
8 ever, these metrics are often reported as Z-scores, which we demonstrate are fundamentally sensitive to
9 the number of targets a compound possesses. This dependency introduces a systematic bias where com-
10 pounds with broad polypharmacology appear statistically significant due to null distribution tightening
11 (the Law of Large Numbers) rather than physical network reachability. Here, we systematically evaluate
12 this bias using the human liver interactome and a controlled comparison of two constituents from *Hy-*
13 *pericum perforatum*. We show that conventional proximity Z-scores yield unstable rankings that reverse
14 depending on network construction parameters. While a compound with many targets may achieve a
15 higher Z-score, it can remain physically more distant from the disease module than a compound with
16 fewer, high-leverage targets. We resolve this by utilizing random walk–based influence propagation and
17 introducing a size-normalized metric: perturbation efficiency. Our results show that influence-based
18 rankings are stable across varied network thresholds and correctly identify high-leverage modulators that
19 proximity metrics miss. This study provides a methodological template for identifying and correcting
20 statistical artifacts in network medicine, enabling more reliable risk assessment in complex biological
21 systems.

22 **Keywords:** network propagation, proximity metrics, metric robustness, drug-induced liver injury,
23 polypharmacology, Z-score confounding, perturbation efficiency.

24 1 Introduction

25 Network-based prioritization is a cornerstone of modern systems biology and drug discovery, assuming
26 that the topological proximity between compound targets and disease genes within a protein–protein in-
27 teraction (PPI) network reflects functional relevance [1–4]. Because raw network distances are sensitive
28 to local topology and degree distribution, they are typically reported as Z-scores relative to degree-
29 matched null models. While these Z-scores successfully quantify statistical significance in most applica-
30 tions [3], we demonstrate they can be confounded by large asymmetries in target set size. As the number
31 of seed nodes increases, the variance of the null distribution decreases (the Law of Large Numbers),
32 leading to deterministic significance inflation for compounds with broad polypharmacology. Identifying
33 whether such results reflect true biological influence, or whether they represent systematic artifacts of
34 distance-based inference, is essential for the reliability of network medicine.

35 Using the human liver interactome as a model system, we investigate this confounding effect through
36 a controlled comparison of two constituents from *Hypericum perforatum* (St. John’s Wort). These con-
37 stituents—Hyperforin and Quercetin—exhibit highly asymmetric target set sizes: Hyperforin possesses
38 10 validated targets, while Quercetin has over 60 [5–7]. This asymmetry provides a sharp stress test
39 for network metrics. Conventional proximity Z-scores predict greater disease-associated significance for
40 the broad-spectrum modulator, even when it is physically more distant from the disease module than the
41 high-leverage modulator. This reversal indicates that proximity-based prioritization is unstable across
42 network construction parameters and susceptible to sample-size artifacts.

43 Here, we evaluate the robustness of proximity-based and influence-based metrics for comparative
44 prioritization. We demonstrate that proximity Z-scores yield unstable, threshold-dependent rankings
45 driven by null-distribution tightening rather than physical reachability. To resolve this instability, we
46 utilize random walk–based influence propagation, which integrates over the entire network topology and
47 captures signal amplification through regulatory hubs [8]. We introduce a normalized metric, perturba-
48 tion efficiency, to account for target set size and ensure unbiased comparisons. Our results show that
49 influence-based propagation provides a stable, theoretically consistent framework for network pharma-
50 cology that correctly identifies high-leverage perturbations where traditional proximity metrics fail.

51 2 Results

52 2.1 Proximity Z-scores are confounded by target set size

53 We first established network context by quantifying target count and shortest-path proximity to 82 DILI-
54 associated genes (Figure 1). Quercetin engages 62 targets in the liver-expressed largest connected
55 component; Hyperforin engages 10. At STRING confidence ≥ 900 , Hyperforin targets are physically
56 closer to DILI genes ($d_c = 1.30$) than Quercetin targets ($d_c = 1.68$; Table 1). However, the proxim-
57 ity Z-scores yield the opposite ranking: Quercetin achieves $Z = -5.44$ ($p < 0.001$), while Hyperforin
58 achieves $Z = -3.86$ ($p < 0.001$). All reported associations survived Benjamini–Hochberg FDR correc-
59 tion ($q < 0.05$).

60 This discrepancy highlights a fundamental confounder in proximity Z-scores: the law of large num-
61 bers. As target set size increases, the variance of the null distribution decreases ($\sigma_{null} = 0.09$ for
62 Quercetin vs 0.24 for Hyperforin), inflating the significance of broader target sets despite greater physical
63 distance. This statistical artifact suggests that Quercetin poses greater risk, whereas the physical topology
64 favors Hyperforin. This effect occurs independently of network construction parameters and represents
65 a fundamental statistical property of averaged distributions: larger samples inherently produce more pre-
66 cise (narrower) null distributions, artificially inflating Z-score magnitude regardless of actual topological
67 proximity.

68 2.2 Influence-based rankings are stable and resolve the confound

69 Random walk with restart (RWR) stabilizes this ranking by integrating over all paths (Figure 2). Hyper-
70 forin achieves influence $Z = +10.12$ ($p < 0.001$); Quercetin achieves $Z = +4.55$ ($p < 0.001$; Table 1).
71 Unlike proximity, influence Z-scores correctly reflect the topological advantage of Hyperforin’s regu-
72 latory hub occupancy. This ranking aligns with the known biological ground truth: Hyperforin is the
73 hepatotoxic constituent responsible for drug-drug interactions via PXR activation, while Quercetin has
74 no documented hepatotoxicity and may be hepatoprotective. The ranking remains consistent across
75 topology-only and expression-weighted analyses, demonstrating that influence propagation is less sus-
76 ceptible to sample-size artifacts than shortest-path distance.

77 2.3 Expression weighting refines the signal

78 To assess whether the RWR signal persists under tissue-specific constraint, we applied expression-
79 weighted influence propagation (EWI), weighting transitions by destination-node liver expression (Fig-

80 ure 3).

81 The Z-score differential narrows but remains substantial under expression weighting: Hyperforin
82 $Z = +8.98$ ($p < 0.001$); Quercetin $Z = +5.79$ ($p < 0.001$). Hyperforin's advantage is driven primarily
83 by the PXR–CYP master regulatory axis, which remains highly active in liver tissue (e.g., CYP3A4 at
84 335 TPM). Quercetin's influence is moderated by its broad, diffuse target profile, which includes several
85 high-expression nodes (e.g., CFB at 1,115 TPM) that do not converge on a DILI effector hub.

86 **2.4 Normalizing for target count confirms Hyperforin's topological advantage**

87 To resolve the target-count paradox, we compared the average network influence of each individual
88 target, reframing polypharmacology as an efficiency problem rather than a coverage problem (Figure 4;
89 Table 2).

Compound	Targets	Eff. (RWR)	Eff. (EWI)	RWR Ratio*	EWI Ratio†
Hyperforin	10	0.1138	0.1330	—	—
Quercetin	62	0.0322	0.0493	—	—
Fold difference	—	—	—	3.5× (3.7×)	2.7× (2.8×)

90 *RWR Ratio: observed ratio (robust ratio in parentheses). †EWI Ratio: observed ratio (robust ratio in paren-
91 theses).

92 Each Hyperforin target contributes $3.7\times$ more DILI-directed influence than each Quercetin target (robust ra-
93 tio). This disparity indicates that Hyperforin's target positions are substantially higher leverage than those of
94 Quercetin, achieving greater perturbation efficiency despite a 6-fold smaller target set. The efficiency ratio re-
95 mains stable within a narrow range: $3.7\times$ at STRING ≥ 700 , $3.5\times$ at ≥ 900 (a 5% variation despite 29% reduction
96 in network density). In contrast, proximity Z-score rankings undergo complete reversal between these thresholds.
97 This demonstrates that perturbation efficiency is a parameter-invariant comparative metric where distance-based
98 Z-scores are not.

100 **2.5 Bootstrap resampling excludes target-selection bias**

101 To rule out the possibility that Hyperforin's advantage arises from favorable target selection rather than strategic
102 network positioning, we performed bootstrap sensitivity analysis (Figure 5). 100 random 10-target subsets were
103 sampled without replacement from Quercetin's 62-target pool and scored by RWR.

104 Hyperforin's observed influence (0.1138) exceeds the entire bootstrap distribution from Quercetin (mean =
105 0.0308, 95% CI = [0.0160, 0.0542]; Table 3). The fold difference between Hyperforin and the bootstrap mean is
106 $3.7\times$. This confirms that Hyperforin's advantage is not an artifact of target count; even when sampling equalized
107 subsets from Quercetin's pool, no configuration matches Hyperforin's influence.

108 **2.6 Ranking stability across network thresholds**

109 The influence ranking is stable across network confidence thresholds (Table 6). Hyperforin ranks first in all RWR
110 and EWI configurations at both ≥ 700 and ≥ 900 thresholds. Notably, the proximity ranking reverses between
111 thresholds: at ≥ 700 , Hyperforin is physically closer ($d_c = 0.60$ vs 1.34) and more "significant" ($Z = -6.04$ vs
112 -5.46). At ≥ 900 , Quercetin appears more "significant" ($Z = -5.44$ vs -3.86) despite being physically more dis-
113 tant (1.68 vs 1.30). This instability in proximity Z-scores—while influence rankings remain stable—demonstrates
114 that influence-based metrics are more robust to network construction parameters.

115 **2.7 Chemical similarity excludes structural confounding**

116 To exclude the possibility that Hyperforin's network signal reflects structural similarity to known hepatotoxins,
117 we performed chemical similarity analysis against the DILrank reference set (Figure 6). Morgan fingerprints
118 (ECFP4) revealed that neither compound exceeds the 0.4 Tanimoto threshold for structural analog detection. No-
119 tably, Quercetin exhibits higher structural similarity to DILI reference drugs yet lower network influence, reinforc-
120 ing that the observed asymmetry is driven by network topology rather than chemical features.

121 **3 Discussion**

122 **3.1 Ranking stability and the Z-score confound**

123 The results of this study highlight a potential limitation in the use of network proximity Z-scores when comparing
124 compounds with asymmetric target set sizes. While proximity is a standard prioritization criterion, our analysis
125 demonstrates that its significance rankings can be influenced by the target count rather than topological distance
126 alone. As the number of targets increases, the variance of the null distribution decreases (a manifestation of the Law
127 of Large Numbers), which can lead to inflated significance levels for compounds with broad polypharmacology. In
128 our controlled comparison, this effect causes a reversal of proximity-based rankings between network thresholds,
129 failing to accurately reflect the physical distance advantage of a high-leverage modulator.

130 Influence-based metrics (RWR and EWI) appear less sensitive to this particular artifact. By integrating over the
131 entire network topology, these methods capture signal propagation through regulatory hubs, providing rankings that
132 remain stable across different network construction parameters. This relative stability suggests that influence-based
133 propagation may offer a more robust framework for comparative network medicine, particularly in the presence of
134 incomplete or asymmetric pharmacological data.

135 The mechanistic explanation for this robustness is that RWR integrates over *all* paths, capturing how signals
136 amplify through hubs like PXR and AKT1. Shortest-path proximity, by contrast, is a descriptive metric for min-
137 imum reachability; treating it as an inferential surrogate for functional impact conflates topological context with
138 biological consequence.

139 **3.2 Relationship to prior work**

140 Our findings do not contradict the foundational work of Guney et al. (2016), but rather identify a specific failure
141 mode that their study design did not stress-test. Guney et al. evaluated network proximity as a classifier for drug-
142 disease associations across 238 drugs with a mean of 3.5 targets per drug—a relatively homogeneous dataset.
143 They reported that proximity is “not biased with respect to the number of targets a drug has” and found that the
144 closest-distance measure (d_c) outperformed a diffusion kernel measure (d_k) for binary classification [3].

145 Our study addresses a fundamentally different question: *comparative ranking* of two compounds with highly
146 asymmetric target counts (10 vs. 62). In this regime, the variance-shrinkage artifact becomes a first-order problem.
147 Guney’s kernel benchmark (d_k) is related to but distinct from random walk with restart; d_k sums contributions from
148 all weighted paths, whereas RWR iteratively propagates probability mass with a restart factor that anchors the walk
149 to seed nodes. More critically, neither d_c nor d_k provides a principled normalization for target set size.

150 The core innovation of this study is *perturbation efficiency*: the average influence exerted per target. This
151 metric is not proposed by Guney et al. and resolves the target-count paradox regardless of whether the underlying
152 propagation method is shortest-path, kernel, or random walk. By framing polypharmacology as an efficiency prob-
153 lem rather than a coverage problem, we provide a bias-corrected comparative framework that survives robustness
154 checks where raw Z-scores fail. While Guney et al. found that a diffusion kernel underperformed closest distance
155 for binary classification of known drug-disease pairs, our task differs fundamentally: we address comparative rank-
156 ing under extreme target-count asymmetry (10 vs. 62 targets). RWR’s restart mechanism enforces locality absent
157 in pure diffusion kernels, and our empirical results demonstrate stable rankings that align with biological ground
158 truth—a criterion not assessed in Guney’s benchmark.

159 **3.3 Expression weighting as a biological constraint**

160 Expression-weighted influence (EWI) constrains signal propagation to liver-active nodes. By attracting signal to
161 highly expressed proteins (destination-node weighting), we ensure that the network propagation reflects tissue-
162 specific biology. Under this constraint, the Hyperforin advantage persists, demonstrating that its topological ef-
163 ficiency is not an artifact of an unconstrained PPI network but is supported by the expression profile of the liver.
164 Attenuation of signal is expected when walks are constrained to active pathways; the fact that the ranking remains
165 stable provides positive evidence for the biological relevance of the PXR axis.

166 **3.4 Perturbation efficiency vs. topological coverage**

167 By normalizing total influence for target set size (where the restart vector is already $|T|$ -weighted), we provide
168 a more balanced comparison of perturbation efficiency. Our results show that a single Hyperforin target exerts
169 3.7-fold more influence on the DILI module than a Quercetin target.

170 This efficiency claim is further validated by bootstrap sensitivity analysis. Even when sampling size-matched
171 10-target subsets from Quercetin’s pool, none reached the influence level achieved by Hyperforin. This demon-

172 strates that the advantage is not due to target count, but to the strategic network position of Hyperforin's tar-
173 gets—specifically their convergence on the PXR master regulator and downstream CYP effectors.

174 **3.5 Mechanistic context: The PXR axis**

175 The stability of the influence ranking aligns with the well-characterized PXR–CYP master regulatory axis. Hyper-
176 forin's primary target, NR1I2 (PXR), induces the expression of major xenobiotic metabolism enzymes including
177 CYP3A4 and CYP2C9 [6, 9]. In our network analysis, these effectors are part of the target set and the DILI
178 module, creating a high-connectivity hub structure that enables efficient propagation. Quercetin's 62 targets, while
179 numerous, are distributed across redundant or peripheral pathways that do not converge on a regulatory bottleneck.
180 Furthermore, clinical evidence indicates that Quercetin is not associated with hepatotoxicity and may exhibit hep-
181 atoprotective properties [7, 10]. Recent experimental studies have corroborated that St. John's wort exacerbates
182 hepatotoxicity through precisely this PXR-mediated bioactivation mechanism [11].

183 **3.6 Limitations**

184 Several limitations warrant consideration. First, network influence is a measure of topological reach and pertur-
185 bation potential, not a direct surrogate for toxicological outcomes. This model is dose-independent and does not
186 account for pharmacokinetics, binding affinity, or saturation kinetics. A high influence score indicates that a com-
187 pound's targets are strategically positioned to modulate a disease module, but the actual biological effect depends
188 on the molecular mechanism of action (e.g., agonism vs. antagonism) and the kinetic context.

189 Second, while we demonstrate that proximity Z-scores are confounded by target set size, influence-based Z-
190 scores are not entirely immune to this effect. As the number of seed nodes increases, the variance of the null
191 distribution for influence sums also decreases, though less severely than for distance-based metrics. Critically,
192 our core claims do not rest on absolute Z-score comparisons. We demonstrate that influence-based *rankings* are
193 stable across network thresholds, while proximity rankings are not. We further resolve the size-dependence by
194 introducing perturbation efficiency (influence per target), which explicitly normalizes for target count and provides
195 a bias-corrected comparative metric.

196 Third, our case study is limited to a single botanical with two contrasting constituents. While this provides a
197 controlled minimal model, generalization to larger compound libraries will require further validation.

198 **3.7 Conclusions**

199 In this study, we utilized *H. perforatum* as a known toxicological model to validate the *reliability* of network
200 metrics; the biological ground truth (Hyperforin-mediated PXR activation) allowed us to confirm that influence
201 propagation correctly identifies high-leverage perturbations where proximity metrics fail. The methodological
202 conclusion is that proximity Z-scores are susceptible to sample-size confounding and should be used descriptively
203 rather than for comparative inference across compounds with differing target counts. Influence-based propagation,

204 combined with per-target normalization, provides a more stable framework that survives robustness checks and
205 aligns better with mechanistic reality.

206 More broadly, this work provides a methodological template for identifying and resolving metric artifacts
207 in network toxicology. By integrating signed edge weights and transcriptometric data, future iterations of this
208 framework could investigate phenotype-specific associations, linking topological influence on specific biological
209 sub-modules to discrete clinical outcomes.

210 4 Methods

211 4.1 Data sources

212 4.1.1 Protein–protein interaction network

213 Human protein–protein interactions were obtained from STRING v12.0 [12]. Combined confidence scores were
214 computed per STRING methodology (text mining, experiments, databases, co-expression, neighborhood, gene
215 fusion, co-occurrence). Only edges with combined confidence ≥ 900 (highest confidence tier) were retained. Raw
216 network: 11,693 genes, 100,383 edges.

217 4.1.2 Liver expression data

218 Gene expression data were obtained from the Genotype-Tissue Expression Project (GTEx) v8 [13]. Median tran-
219 scripts per million (TPM) values for liver tissue were extracted from the 2017-06-05 release (RNASeQCv1.1.9).
220 Genes with liver TPM ≥ 1 were retained. Result: 13,496 liver-expressed genes.

221 4.1.3 Drug-induced liver injury gene set

222 DILI-associated genes were obtained from DisGeNET [14] curated gene-disease associations. Query: UMLS
223 concept identifier C0860207 (Drug-Induced Liver Injury). Inclusion criterion: genes with curated evidence linking
224 to DILI. Raw DILI gene count: 127 genes.

225 4.1.4 Hyperforin targets

226 Hyperforin targets were curated from primary literature sources [6, 9]. Sources included studies of PXR activation,
227 CYP induction, and ABC transporter modulation. Raw target count: 14 proteins (Table 7).

228 4.1.5 Quercetin targets

229 Quercetin targets were retrieved programmatically from ChEMBL v31 [15] via REST API. Query: CHEMBL159
230 (Quercetin). Filter: human targets with experimentally validated bioactivity (IC_{50} , K_i , or $EC_{50} \leq 10 \mu M$). Raw
231 target count: 122 proteins.

232 **4.2 Target processing**

233 Protein identifiers were mapped to HUGO gene symbols using STRING info files and UniProt [16]. Non-human
234 proteins (mouse, rat, bacterial, viral) were excluded. Gene symbols were standardized (e.g., MDR1 → ABCB1).
235 Processed target counts: Hyperforin = 14, Quercetin = 87.

236 **4.3 Network construction**

237 The STRING network was filtered to genes with liver expression ≥ 1 TPM (GTEx v8). The largest connected
238 component (LCC) was extracted using NetworkX [17]. Compound targets and DILI genes not present in the LCC
239 were excluded. Final network: 7,677 nodes, 66,908 edges. Final target counts: Hyperforin = 10, Quercetin = 62.
240 Final DILI gene count: 82.

241 Five genes are targeted by both compounds: ABCG2, AKT1, CYP3A4, MMP2, MMP9. These were retained
242 in both target sets.

243 **4.4 Shortest-path proximity (descriptive)**

244 Mean minimum shortest-path distance from compound targets T to DILI genes D :

$$d_c = \frac{1}{|T|} \sum_{t \in T} \min_{d \in D} \text{dist}(t, d) \quad (1)$$

245 where $\text{dist}(t, d)$ is the unweighted shortest-path length in the LCC. Shortest-path proximity is a descriptive metric.
246 It was used to provide network context, not to test influence.

247 **4.5 Random walk with restart**

248 Influence propagation was quantified using random walk with restart (RWR), a global network propagation al-
249 gorithm that captures both direct and indirect associations by simulating the diffusion of signal from seed nodes
250 [3, 8]. Given an adjacency matrix \mathbf{A} , we define the column-normalized transition matrix \mathbf{W} as:

$$W_{ij} = \frac{A_{ij}}{\sum_k A_{kj}} \quad (2)$$

251 The steady-state probability vector \mathbf{p} is solved iteratively until convergence:

$$\mathbf{p}^{(k+1)} = (1 - \alpha)\mathbf{W}\mathbf{p}^{(k)} + \alpha\mathbf{p}_0 \quad (3)$$

252 where:

253 • $\alpha = 0.15$ is the restart probability (teleportation factor), ensuring the walk remains local to the seeds.

254 • \mathbf{p}_0 is the restart (seed) vector, with $p_0(i) = 1/|T|$ for $i \in T$ (targets) and 0 otherwise.

- 255 • Convergence is defined as the L_1 norm of the difference between successive iterations being $< 10^{-6}$.
- 256 All computations reached convergence within 100 iterations. The total influence I on the DILI module D is the
 257 sum of steady-state probabilities at disease nodes: $I = \sum_{d \in D} p(d)$.

258 **4.6 Permutation testing and degree matching**

259 To assess whether the observed influence I is significantly greater than what would be expected by chance, we
 260 performed permutation testing ($n = 1,000$). To account for the bias where high-degree nodes (hubs) naturally
 261 accumulate more influence, we utilized a degree-preserving sampling strategy. For each target $t \in T$, a random
 262 surrogate node was sampled from the network such that its degree k_{rand} was within $\pm 25\%$ of the original target's
 263 degree k_t . This ensures that the null distribution reflects the connectivity profile of the original target set. Random
 264 seeds were fixed to 42 for reproducibility. Z-scores were computed as $Z = (x_{obs} - \mu_{null})/\sigma_{null}$, and empirical
 265 P -values were derived from the null distribution.

266 **4.7 Expression-weighted influence**

267 Edge weights were modified by destination-node liver expression:

$$W'_{ij} = \frac{A_{ij} \cdot e_i}{\sum_k A_{kj} \cdot e_k} \quad (4)$$

268 where e_i is the normalized liver expression for gene i (GTEx v8 liver). Liver TPM values were log-transformed
 269 ($\log_2(\text{TPM} + 1)$) and min-max normalized to [0, 1] across the network. A minimum expression floor of 0.01 was
 270 applied to ensure all nodes remained reachable. Attracting signal to highly-expressed nodes constrains RWR
 271 propagation to biologically active pathways in the liver. All other RWR parameters were identical. Random seed:
 272 42.

273 **4.8 Quantifying perturbation efficiency**

274 By defining the restart vector as $\mathbf{p}_0(i) = 1/|T|$ (Eq. 75), the total steady-state probability mass \mathbf{p} is inherently
 275 partitioned among the target set. Consequently, the summed influence I on the DILI module (Eq. 81) represents
 276 the average perturbation efficiency per target. This normalization serves as an effect-size adjustment that allows
 277 for a direct comparison of the per-unit impact of compounds with asymmetric target sets. Hereafter, we refer to
 278 this as the perturbation efficiency.

279 **4.9 Bootstrap sensitivity analysis**

280 To assess whether target count explains the observed ranking: 100 random 10-target subsets were sampled without
 281 replacement from Quercetin's 62-target pool. Each subset was scored by standard RWR. Summary statistics: mean,

282 standard deviation, 95th percentile. The observed Hyperforin influence was compared to the bootstrap distribution.
283 Random seed: 42.

284 **4.10 Chemical similarity analysis**

285 Structural similarity to known hepatotoxins was assessed to exclude confounding by chemical class. Morgan
286 fingerprints (ECFP4; radius = 2, 2048 bits) were generated using RDKit v2023.03 [18]. Reference set: DILIrank
287 2.0 drugs with retrievable SMILES (542 DILI-positive, 365 DILI-negative). SMILES were retrieved via PubChem
288 REST API. Tanimoto coefficient:

$$\text{Tanimoto}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

289 Maximum similarity across the reference set was reported for each compound. Structural analog threshold: Tani-
290 moto > 0.4 [19].

291 **4.11 Software and reproducibility**

292 Python 3.10, NetworkX 3.1 [17]; R 4.3, igraph 1.5. All random seeds fixed at 42. Target lists sorted alphabetically
293 before processing.

294 **4.12 Code and data availability**

295 All code: <https://github.com/antonybevan/h-perforatum-network-tox>

296 Data sources:

- 297 • STRING v12.0: <https://string-db.org>
- 298 • GTEx v8: <https://gtexportal.org>
- 299 • ChEMBL v31: <https://www.ebi.ac.uk/chembl>
- 300 • DILIRank 2.0: <https://www.fda.gov/science-research/ltrkb>

301 **Data availability**

302 All data and code supporting this study are publicly available at: <https://github.com/antonybevan/h-per->
303 **foratum-network-tox**.

304 Source data for all figures and tables are provided in the Supplementary Information. Raw data were obtained
305 from the following public repositories:

- 306 • STRING v12.0: <https://string-db.org>
- 307 • GTEx v8: <https://gtexportal.org>
- 308 • ChEMBL v31: <https://www.ebi.ac.uk/chembl>

- 309 • DisGeNET: <https://www.disgenet.org>
- 310 • DILIrank 2.0: <https://www.fda.gov/science-research/ltrkb>

311 **References**

- 312 [1] Andrew L Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology*,
313 4(11):682–690, 2008. doi: 10.1038/nchembio.118.
- 314 [2] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based ap-
315 proach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011. doi: 10.1038/nrg2918.
- 316 [3] Emre Guney, Jörg Menche, Marc Vidal, and Albert-László Barabási. Network-based in silico drug efficacy
317 screening. *Nature Communications*, 7:10331, 2016. doi: 10.1038/ncomms10331.
- 318 [4] Jörg Menche, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, and
319 Albert-László Barabási. Uncovering disease-disease relationships through the incomplete interactome. *Sci-
320 ence*, 347(6224):1257601, 2015. doi: 10.1126/science.1257601.
- 321 [5] Adolf Nahrstedt and Veronika Butterweck. Biologically active and other chemical constituents of the herb of
322 Hypericum perforatum L. *Pharmacopsychiatry*, 30(S2):129–134, 1997. doi: 10.1055/s-2007-979533.
- 323 [6] Linda B Moore, Bryan Goodwin, Stacey A Jones, G Bruce Wisely, Connie J Serabjit-Singh, Timothy M
324 Willson, John L Collins, and Steven A Kliewer. St. John’s wort induces hepatic drug metabolism through
325 activation of the Pregnen X Receptor. *Proceedings of the National Academy of Sciences*, 97(13):7500–7502,
326 2000. doi: 10.1073/pnas.130155097.
- 327 [7] Agnes W Boots, Guido RMM Haenen, and Aalt Bast. Health effects of quercetin: from antioxidant to
328 nutraceutical. *European Journal of Pharmacology*, 585(2-3):325–337, 2008. doi: 10.1016/j.ejphar.2008.03
329 .008.
- 330 [8] Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N Robinson. Walking the interactome for prior-
331 itization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008. doi:
332 10.1016/j.ajhg.2008.02.013.
- 333 [9] Reginald E Watkins, G Bruce Wisely, Linda B Moore, John L Collins, Millard H Lambert, Shawn P Williams,
334 Timothy M Willson, Steven A Kliewer, and Matthew R Redinbo. The human nuclear xenobiotic receptor
335 PXR: structural determinants of directed promiscuity. *Science*, 292(5525):2329–2333, 2001. doi: 10.1126/
336 science.1060762.
- 337 [10] National Institute of Diabetes and Digestive and Kidney Diseases. LiverTox: Clinical and research informa-
338 tion on drug-induced liver injury [internet]. quercetin. <https://www.ncbi.nlm.nih.gov/books/NBK548281/>, 2020. Updated July 10, 2020.

- 340 [11] Siyu Chen, Xue Wang, Xinran Ye, Qinjin Wang, Xin Sun, Chunyan Ma, Zhidong Yuan, and Yang Yu. St.
341 John's wort exacerbates acetaminophen-induced liver injury through PXR and CYP-mediated bioactivation.
342 *Toxicological Sciences*, 190(1):68–80, 2022. doi: 10.1093/toxsci/kfac098.
- 343 [12] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja
344 Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, Peer Bork, Lars J Jensen, and
345 Christian von Mering. The STRING database in 2023: protein–protein association networks and functional
346 enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D483–D489,
347 2023. doi: 10.1093/nar/gkac1000.
- 348 [13] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*,
349 369(6509):1318–1330, 2020. doi: 10.1126/science.aaz1776.
- 350 [14] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno,
351 Ferran Sanz, and Laura I Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update.
352 *Nucleic Acids Research*, 48(D1):D845–D855, 2020. doi: 10.1093/nar/gkz1021.
- 353 [15] David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula
354 Magaña, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, Maria Gordillo-Marañón, Fiona Hunter,
355 Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis Atkinson, Nicolas Bosc, Chris J
356 Radoux, Aldo Segura-Cabrera, Anne Hersey, and Andrew R Leach. ChEMBL: towards direct deposition
357 of bioassay data. *Nucleic Acids Research*, 47(D1):D930–D940, 2019. doi: 10.1093/nar/gky1075.
- 358 [16] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51
359 (D1):D483–D489, 2023. doi: 10.1093/nar/gkac1052.
- 360 [17] Aric A Hagberg, Daniel A Schult, and Pieter J Swart. Exploring network structure, dynamics, and function
361 using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th*
362 *Python in Science Conference (SciPy 2008)*, pages 11–15, 2008.
- 363 [18] RDKit. RDKit: Open-source cheminformatics. <https://www.rdkit.org>, 2023. Version 2023.03.
- 364 [19] Gerald Maggiora, Martin Vogt, Dagmar Stumpfe, and Jürgen Bajorath. Molecular similarity in medicinal
365 chemistry. *Journal of Medicinal Chemistry*, 57(8):3186–3204, 2014. doi: 10.1021/jm401411z.
- 366 [20] R Scott Obach. Inhibition of human cytochrome P450 enzymes by constituents of St. John's Wort, an herbal
367 preparation used in the treatment of depression. *Journal of Pharmacology and Experimental Therapeutics*,
368 294(1):88–95, 2000. doi: 10.1124/jpet.294.1.88.
- 369 [21] Bernard J Komoroski, Shuyan Zhang, Steven A Wrighton, Stephen C Strom, Raman Venkataraman, and
370 Erin G Schuetz. Induction and inhibition of cytochromes P450 by the St. John's wort constituent hyperforin
371 in human hepatocytes. *Drug Metabolism and Disposition*, 32(5):512–518, 2004. doi: 10.1124/dmd.32.5.512.

- 372 [22] M Hennessy, D Kelleher, JP Lloyd, A Alrajhi, O Meenaghan, C McDonald, F Mulcahy, JP Spiers, and
373 J Feely. St John's wort increases expression of P-glycoprotein: implications for drug interactions. *British*
374 *Journal of Clinical Pharmacology*, 53(1):75–82, 2002. doi: 10.1046/j.1365-2125.2002.01512.x.
- 375 [23] H Assefa and V Butterweck. The role of hyperforin in the metabolic and transport-mediated drug interactions
376 of St. John's wort. *Planta Medica*, 70(4):291–300, 2004. doi: 10.1055/s-2004-818938.
- 377 [24] Er-Jia Wang, Mary Barecki-Roach, and William W Johnson. Quantitative characterization of direct P-
378 glycoprotein inhibition by St John's wort constituents hypericin and hyperforin. *Journal of Pharmacy and*
379 *Pharmacology*, 56(11):1451–1456, 2004. doi: 10.1211/0022357044736.
- 380 [25] C Quiney, C Billard, A M Faussat, C Salanoubat, and J P Kolb. Hyperforin directly inhibits AKT1 kinase
381 activity and promotes apoptosis in AML cells. *Leukemia*, 21(10):2101–2111, 2007. doi: 10.1038/sj.leu.240
382 4834.
- 383 [26] C Quiney, C Billard, C Salanoubat, J D Fourneron, and J P Kolb. Hyperforin inhibits MMP-9 secretion by
384 B-cell chronic lymphocytic leukemia cells. *Leukemia*, 20(8):1514–1521, 2006. doi: 10.1038/sj.leu.2404283.
- 385 [27] Kristian Leuner, Viacheslav Kazanski, Marina Müller, Kirill Essin, Britta Henke, Martina Gassen, Christopher
386 Koch, Christina Bulut, Karola Silbermann, Annette Kopp-Schneider, Gerald Thiel, Vladimir Laketa,
387 Inna Gorshkova, Valentina Przetheskikh, Christian Harteneck, Wolfgang F Graier, Vadym Degtiar, Peter
388 Lipp, Axel Lückhoff, and Walter E Müller. Hyperforin—a key constituent of St. John's wort specifically
389 activates TRPC6 channels. *The FASEB Journal*, 21(14):4101–4111, 2007. doi: 10.1096/fj.07-8110com.
- 390 [28] Katarina Hostanska, J Reiher, S Jessenmeyer, J Reichling, and R Saller. Hyperforin and hypericin: synergistic
391 cytotoxicity and induced apoptosis in human malignant cell lines. *European Journal of Pharmaceutics and*
392 *Biopharmaceutics*, 55(3):301–310, 2003. doi: 10.1016/s0939-6411(03)00021-3.
- 393 [29] Vikas Kumar, Alexander Mdzinarishvili, Thomas Kiewert, Maria P Abbracchio, Annalisa Pinna, Renata
394 Ciccarelli, Walter E Müller, and Jochen Klein. NMDA receptor-antagonistic properties of hyperforin, a
395 constituent of St. John's wort. *Journal of Pharmacological Sciences*, 102(1):47–54, 2006. doi: 10.1254/jphs
396 .fp06041.

397 **Figure Legends**

398 **Figure 1. Network context: target count and physical proximity to DILI genes.** (A) Target count in the
399 liver-expressed largest connected component. Quercetin: 62 targets; Hyperforin: 10 targets. (B) Shortest-path
400 proximity (d_c) to 82 DILI-associated genes. Hyperforin is physically closer ($d_c = 1.30$) than Quercetin ($d_c = 1.68$).
401 Z-scores represent deviation from degree-matched null expectation ($n = 1,000$ permutations). Quercetin: $Z =$
402 -5.44 ($p < 0.001$); Hyperforin: $Z = -3.86$ ($p < 0.001$). Negative Z-scores indicate closer-than-random proximity.
403 Network: STRING v12.0 (confidence ≥ 900), GTEx v8 (liver TPM ≥ 1).

404 **Figure 2. Instability of proximity Z-scores.** Dumbbell plot showing the dissociation between shortest-path
405 proximity (left) and random walk influence (right) at STRING confidence ≥ 900 . At this threshold, Quercetin
406 appears more "significant" in Z-score but is physically more distant (1.68 vs 1.30) from DILI genes. Hyperforin:
407 proximity $Z = -3.86$, influence $Z = +10.12$ ($p < 0.001$). Quercetin: proximity $Z = -5.44$, influence $Z = +4.55$
408 ($p < 0.001$). Influence quantified by random walk with restart (RWR; $\alpha = 0.15$). $n = 1,000$ degree-matched
409 permutations per compound.

410 **Figure 3. Expression weighting refines influence propagation.** Waterfall decomposition of Z-score changes
411 under expression-weighted influence (EWI). Initial Hyperforin advantage: $\Delta Z = +5.57$ (RWR). Hyperforin change:
412 -1.14 (attenuation of signal through liver-active hubs). Quercetin change: $+1.24$ (gain from high-expression
413 nodes like CFB). Residual Hyperforin advantage: $\Delta Z = +3.19$. Both compounds remain significant under EWI:
414 Hyperforin $Z = +8.98$ ($p < 0.001$); Quercetin $Z = +5.79$ ($p < 0.001$). Expression weighting from GTEx v8 liver
415 tissue.

416 **Figure 4. Average network influence quantifies efficiency disparity.** Phase plot of total influence versus
417 target count. Horizontal lines represent efficiency tiers (Efficiency/average influence = constant). Hyperforin
418 occupies a higher efficiency region despite fewer targets. Efficiency/average influence values: Hyperforin = 0.1138
419 (RWR), 0.1330 (EWI); Quercetin = 0.0322 (RWR), 0.0493 (EWI). Efficiency difference: $3.7 \times$ (based on bootstrap
420 mean comparison). The observed influence represents an effect-size normalization (total steady-state mass on DILI
421 genes); no independent permutation test was performed.

422 **Figure 5. Bootstrap sensitivity analysis excludes target-count confounding.** Density distribution of RWR
423 influence scores from 100 random 10-target samples drawn from Quercetin's 62-target pool. Shaded region: 95%
424 confidence interval (0.0160–0.0542). Vertical line: Hyperforin observed influence (0.1138). Hyperforin exceeds
425 the entire bootstrap distribution ($3.7 \times$ fold vs. mean). This confirms that Hyperforin's advantage is not attributable
426 to favorable target count. Bootstrap is a robustness control; it does not provide independent statistical evidence.

427 **Figure 6. Chemical similarity control excludes structural confounding.** Maximum Tanimoto similarity to
428 DILrank reference drugs. Reference set: 542 DILI-positive, 365 DILI-negative drugs. Hyperforin: max = 0.15
429 (DILI+), 0.20 (DILI-). Quercetin: max = 0.21 (DILI+), 0.22 (DILI-). Dashed line: 0.4 threshold for structural
430 analog detection [19]. Neither compound is a structural analog of known hepatotoxins. This orthogonal analysis
431 excludes chemical class as an explanation for the observed network signal. Fingerprints: Morgan (ECFP4), radius
432 2, 2048 bits.

433 **Tables**

Table 1: Network metrics reveal the instability of proximity Z-scores. While Quercetin achieves more significant proximity Z-scores due to tighter null distributions, Hyperforin is physically closer (d_c) to DILI genes. Influence-based metrics resolve this confounding and stably prioritize Hyperforin. Network: STRING v12.0 LCC (confidence ≥ 900) filtered to liver-expressed genes.

Metric	Compound	Targets	Observed	Z	p	Efficiency
<i>Tier 1: Shortest-path proximity</i>						
	Hyperforin	10	$d_c = 1.30$	-3.86	< 0.001*	—
	Quercetin	62	$d_c = 1.68$	-5.44	< 0.001*	—
<i>Instability: Quercetin is physically more distant yet more "significant"</i>						
<i>Tier 2: Random walk influence (RWR)</i>						
	Hyperforin	10	0.1138	+10.12	< 0.001*	0.1138
	Quercetin	62	0.0322	+4.55	< 0.001	0.0322
<i>Resolution: Correctly prioritizes physical proximity and regulatory hub modulation</i>						
<i>Tier 3: Expression-weighted influence (EWI)</i>						
	Hyperforin	10	0.1330	+8.98	< 0.001*	0.1330
	Quercetin	62	0.0493	+5.79	< 0.001	0.0493

*At permutation floor (<1/1,000).

Efficiency = average influence per target; RWR = random walk with restart; EWI = expression-weighted influence; d_c = mean minimum shortest-path distance; DILI = drug-induced liver injury. All associations survived Benjamini–Hochberg FDR correction ($q < 0.05$).

Table 2: Average influence efficiency. Normalization to the total seeding mass quantifies the average influence per target. Hyperforin targets are 3.7-fold more efficient at perturbing the DILI module than Quercetin targets.

Analysis	Hyp. Eff.	Quer. Eff.	Eff. Ratio*	Rob. Ratio†
RWR (topology-only)	0.1138	0.0322	3.5×	3.7×
EWI (expression-weighted)	0.1330	0.0493	2.7×	2.8×

*Efficiency Ratio = Observed average influence ratio. †Robust Ratio = Observed influence / size-matched Bootstrap Mean (N=10). RWR = random walk with restart; EWI = expression-weighted influence.

Table 3: **Bootstrap sensitivity excludes target-count confounding.** Random 10-target subsets ($n = 100$) sampled without replacement from Quercetin’s 62-target pool. Hyperforin’s observed influence exceeds the entire bootstrap distribution.

Statistic	Value	Interpretation
Hyperforin observed	0.1138	Reference
Bootstrap mean	0.0308	Expected if targets equivalent
Bootstrap SD	0.0100	Sampling variability
Bootstrap 95% CI	[0.0160, 0.0542]	2.5th–97.5th percentile
Hyperforin / mean	3.7×	Effect size
Exceeds 95% CI?	Yes	Not attributable to sampling

Random seed: 42. Note: Bootstrap confirms robustness to target selection; it does not constitute independent inferential evidence.

Table 4: **Chemical similarity excludes structural confounding.** Neither compound resembles known hepatotoxins ($\text{Tanimoto} < 0.4$). Quercetin is more similar to DILI-positive drugs yet shows lower network influence.

Compound	Max Tanimoto (DILI+)	Max Tanimoto (DILI-)	Analog?*	Network rank
Hyperforin	0.154	0.202	No	1 (higher influence)
Quercetin	0.212	0.220	No	2 (lower influence)

*Analog threshold: $\text{Tanimoto} > 0.4$ (Maggiora et al., 2014). Morgan fingerprints (ECFP4, radius 2, 2048 bits). DILIRank: 542 DILI+, 365 DILI- drugs.

Table 5: **Hyperforin targets include regulatory hubs.** All 10 Hyperforin targets in the liver-expressed LCC, with liver expression (GTEx v8) and network degree. PXR (NR1I2) is the master regulator; CYP enzymes are downstream effectors.

Gene	Protein	TPM	Degree	Function	DILI link
NR1I2	PXR	43	28	Master regulator	Direct
CYP3A4	CYP3A4	335	89	Xenobiotic metabolism	Direct
CYP2C9	CYP2C9	434	76	Xenobiotic metabolism	Direct
CYP2B6	CYP2B6	125	42	Xenobiotic metabolism	Indirect
AKT1	PKB	33	312	Stress signaling	Indirect
ABCB1	P-gp	7	53	Drug efflux	Direct
ABCC2	MRP2	60	38	Drug efflux	Direct
ABCG2	BCRP	4	31	Drug efflux	Indirect
MMP2	MMP2	5	87	ECM remodeling	Indirect
MMP9	MMP9	1	94	ECM remodeling	Indirect

AKT1 is the highest-degree target (312 neighbors). Five of 10 targets (NR1I2, CYP3A4, CYP2C9, ABCB1, ABCC2) are directly connected to DILI genes. TPM = transcripts per million; DILI = drug-induced liver injury; LCC = largest connected component.

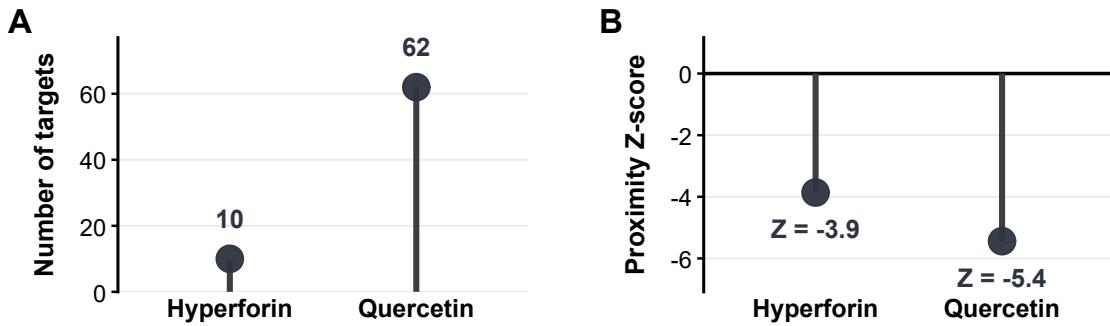
Table 6: **Influence ranking is robust to network construction parameters.** Hyperforin ranks first across all thresholds and influence metrics. Proximity Z-scores are unstable and reverse rankings between thresholds, failing to accurately reflect the physical distance advantage of Hyperforin.

Threshold	Compound	RWR Z	EWI Z	Proximity d_c	Proximity Z
≥ 700 (11,693 nodes)	Hyperforin	+12.08	+11.20	0.60	-6.04
	Quercetin	+5.53	+7.09	1.34	-5.46
≥ 900 (7,677 nodes)	Hyperforin	+10.12	+8.98	1.30	-3.86
	Quercetin	+4.55	+5.79	1.68	-5.44

Note: At ≥ 900 , Quercetin achieves a more "significant" proximity Z-score despite being physically more distant (1.68 vs 1.30) from DILI genes. RWR = random walk with restart; EWI = expression-weighted influence; d_c = mean minimum shortest-path distance; DILI = drug-induced liver injury.

434 **Figures**

Network context: target count and proximity to DILI genes



[DESCRIPTIVE CONTEXT] Target count and shortest-path proximity provide network context but are not used for causal inference. Proximity Z-scores represent deviation from degree-matched random expectation ($n = 1,000$ permutations). Negative values indicate closer-than-random proximity. Data: STRING v12.0 (≥ 900), human liver LCC.

Figure 1: Network context: target count and physical proximity to DILI genes. (A) Target count in the liver-expressed largest connected component. Quercetin: 62 targets; Hyperforin: 10 targets. (B) Shortest-path proximity (d_c) to 82 DILI-associated genes. Hyperforin is physically closer ($d_c = 1.30$) than Quercetin ($d_c = 1.68$). Z-scores represent deviation from degree-matched null expectation ($n = 1,000$ permutations). Quercetin: $Z = -5.44$ ($p < 0.001$); Hyperforin: $Z = -3.86$ ($p < 0.001$). Negative Z-scores indicate closer-than-random proximity. Network: STRING v12.0 (confidence ≥ 900), GTEx v8 (liver TPM ≥ 1).

Proximity does not predict influence

Proximity ranking is threshold-dependent; influence ranking is stable



[CORE INFERENCE] The rank reversal demonstrates that shortest-path proximity does not predict functional influence. Lines connect each compound's proximity Z-score with its influence Z-score (random walk with restart, RWI). Both metrics derived from degree-matched permutation null models ($n = 1,000$). Data: STRING v12.0 (≥ 900).

Figure 2: **Instability of proximity Z-scores.** Dumbbell plot showing the dissociation between shortest-path proximity (left) and random walk influence (right) at STRING confidence ≥ 900 . At this threshold, Quercetin appears more "significant" in Z-score but is physically more distant (1.68 vs 1.30) from DILI genes. Hyperforin: proximity $Z = -3.86$, influence $Z = +10.12$ ($p < 0.001$). Quercetin: proximity $Z = -5.44$, influence $Z = +4.55$ ($p < 0.001$). Influence quantified by random walk with restart (RWR; $\alpha = 0.15$). $n = 1,000$ degree-matched permutations per compound.

Expression weighting attenuates but does not reverse the advantage

Gap: +5.6 (RWR) → +3.2 (EWI)



[CONSTRAINT ANALYSIS] The RWR advantage (+5.6) is partitioned under expression-weighted influence propagation:
 (1) Hyperforin's change under expression weighting (-1.1); (2) Quercetin's gain (+1.2, driven by CFB at 1115
 TPM). Residual advantage (+3.2) remains significant (both $p < 10^{-8}$). GTEx v8 liver expression (TPM ≥ 1). STRING
 v12.0 (≥ 900), $n = 1,000$ degree-matched permutations.

Figure 3: Expression weighting refines influence propagation. Waterfall decomposition of Z-score changes under expression-weighted influence (EWI). Initial Hyperforin advantage: $\Delta Z = +5.57$ (RWR). Hyperforin change: -1.14 (attenuation of signal through liver-active hubs). Quercetin change: $+1.24$ (gain from high-expression nodes like CFB). Residual Hyperforin advantage: $\Delta Z = +3.19$. Both compounds remain significant under EWI: Hyperforin $Z = +8.98$ ($p < 0.001$); Quercetin $Z = +5.79$ ($p < 0.001$). Expression weighting from GTEx v8 liver tissue.

Average network influence quantifies perturbation efficiency

Normalization reframes polypharmacology as efficiency, not coverage

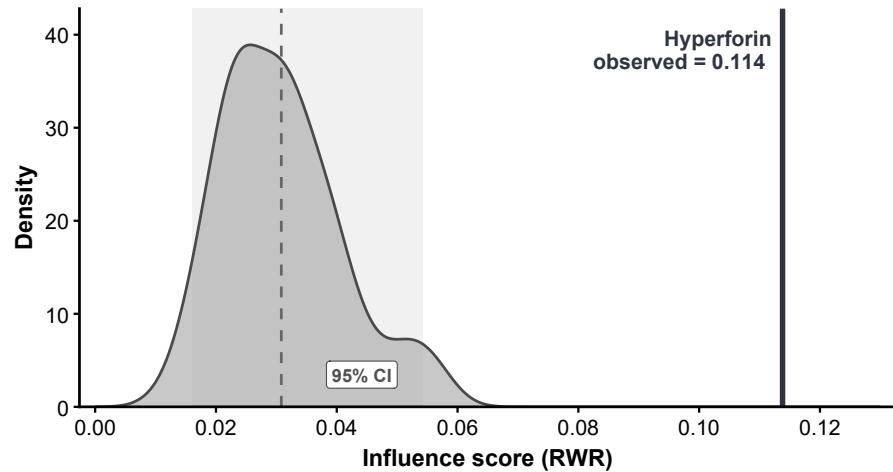


Average influence represents an effect-size normalization (total steady-state mass on DILI genes); no independent permutation test was performed. Horizontal lines represent efficiency tiers (Average Influence = constant). Hyperforin occupies a higher efficiency region despite fewer targets. Data: STRING v12.0 (≥ 900), $n = 1,000$ permutations.

Figure 4: Average network influence quantifies efficiency disparity. Phase plot of total influence versus target count. Horizontal lines represent efficiency tiers (Efficiency/average influence = constant). Hyperforin occupies a higher efficiency region despite fewer targets. Efficiency/average influence values: Hyperforin = 0.1138 (RWR), 0.1330 (EWI); Quercetin = 0.0322 (RWR), 0.0493 (EWI). Efficiency difference: **3.7 \times** (based on bootstrap mean comparison). The observed influence represents an effect-size normalization (total steady-state mass on DILI genes); no independent permutation test was performed.

Bootstrap sensitivity analysis excludes target-count confounding

Distribution of influence scores from random 10-target samples



[ROBUSTNESS CONTROL] Bootstrap sensitivity: 100 random 10-target samples from Quercetin's pool, scored by random walk with restart (RWI). Shaded = 95% CI. Hyperforin (solid line) exceeds entire distribution. Data: STRING v12.0 (≥ 900).

Figure 5: **Bootstrap sensitivity analysis excludes target-count confounding.** Density distribution of RWR influence scores from 100 random 10-target samples drawn from Quercetin's 62-target pool. Shaded region: 95% confidence interval (0.0160–0.0542). Vertical line: Hyperforin observed influence (0.1138). Hyperforin exceeds the entire bootstrap distribution ($3.7 \times$ fold vs. mean). This confirms that Hyperforin's advantage is not attributable to favorable target count. Bootstrap is a robustness control; it does not provide independent statistical evidence.

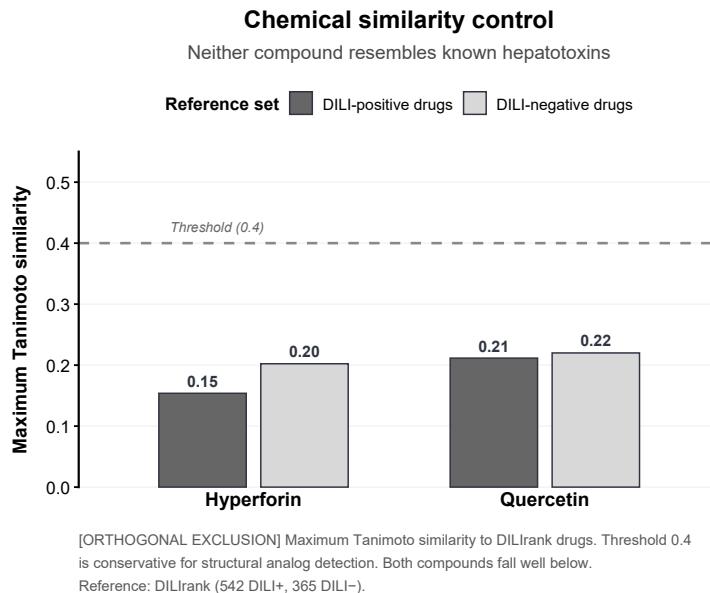


Figure 6: Chemical similarity control excludes structural confounding. Maximum Tanimoto similarity to DILIrack reference drugs. Reference set: 542 DILI-positive, 365 DILI-negative drugs. Hyperforin: max = 0.15 (DILI+), 0.20 (DILI-). Quercetin: max = 0.21 (DILI+), 0.22 (DILI-). Dashed line: 0.4 threshold for structural analog detection [19]. Neither compound is a structural analog of known hepatotoxins. This orthogonal analysis excludes chemical class as an explanation for the observed network signal. Fingerprints: Morgan (ECFP4), radius 2, 2048 bits.

435 **Author contributions**

436 **Antony Bevan:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Cu-
437 ration, Writing - Original Draft, Writing - Review & Editing, Visualization.

438 **Acknowledgements**

439 This research received no external funding. AI-assisted tools were used to assist with code development and
440 statistical analysis. The author takes full responsibility for all content.

441 **Competing interests**

442 The author(s) declare no competing interests.

443 **Supplementary Tables**

Table 7: **Table S1. Hyperforin target genes and literature sources.** All 14 raw targets with UniProt IDs, gene symbols, and primary literature sources. Targets marked with * are present in the liver-expressed LCC (STRING ≥ 900 , GTEx TPM ≥ 1).

UniProt	Gene	In LCC	Source
O75469	NR1I2 (PXR)	Yes*	[6, 9]
P08684	CYP3A4	Yes*	[6]
P11712	CYP2C9	Yes*	[20]
P20813	CYP2B6	Yes*	[21]
P08183	ABCB1	Yes*	[22]
Q9UNQ0	ABCG2	Yes*	[23]
O15440	ABCC2	Yes*	[24]
P31749	AKT1	Yes*	[25]
P08253	MMP2	Yes*	[25]
P14780	MMP9	Yes*	[26]
Q9Y210	TRPC6	No	[27]
P15692	VEGFA	No	[26]
Q13794	PMAIP1	No	[28]
Q12879	GRIN1	No	[29]

Table 8: **Table S2. Quercetin target curation summary.** Target counts at each processing stage.

Stage	Count
Raw targets (ChEMBL v31, CHEMBL159)	122
Excluded: non-human (mouse, rat, bacterial, viral)	10
Excluded: no UniProt mapping	25
Processed targets	87
Excluded: not liver-expressed (TPM < 1)	20
Excluded: not in STRING LCC	5
Final targets in LCC	62

Table 9: **Table S3. DILI gene set curation.** Genes associated with drug-induced liver injury from DisGeNET (UMLS C0860207).

Stage	Count
Raw DILI genes (DisGeNET)	127
In STRING ≥ 700 liver LCC	84
In STRING ≥ 900 liver LCC	82
Excluded: miRNAs (not in PPI network)	21
Excluded: cytokines (not in LCC)	12
Excluded: other	12

Table 10: **Table S4. Genes targeted by both compounds.** Five genes present in both Hyperforin and Quercetin target sets.

Gene	Protein	Function
ABCG2	BCRP	Efflux transporter
AKT1	Protein kinase B	Cell survival signaling
CYP3A4	Cytochrome P450 3A4	Drug metabolism
MMP2	Matrix metalloproteinase-2	Extracellular matrix remodeling
MMP9	Matrix metalloproteinase-9	Extracellular matrix remodeling

Table 11: **Table S5. Direct DILI gene connectivity.** Hyperforin targets with first-order (distance = 1) connections to DILI genes in the STRING network (≥ 900). DILI neighbors are genes present in the 82-gene DILI set.

Target	DILI Neighbors	N	Function
CYP3A4	NR1I2, CYP2E1, UGT1A9, GSTM1, GSTP1	5	Xenobiotic metabolism
AKT1	MAP3K5, NFE2L2, CTNNB1, IGF1	4	Stress response
MMP9	LCN2, SPP1, MMP2	3	Inflammation/ECM
ABCB1	ABCC2, NR1I2	2	Drug transport
CYP2C9	CYP2E1, NR1I2	2	Xenobiotic metabolism
CYP2B6	NR1I2	1	Xenobiotic metabolism
NR1I2	CYP2E1, ABCC2	2	Master regulator
ABCG2	ABCC2	1	Drug transport
ABCC2	NR1I2, ABCB1	2	Drug transport
MMP2	MMP9, SPP1	2	ECM remodeling
Total unique		12	

Table 12: **Table S6. Quercetin direct DILI gene connectivity summary.** Summary statistics for first-order DILI connections across Quercetin's 62 targets.

Metric	Value
Total targets in LCC	62
Targets with ≥ 1 direct DILI neighbor	18
Total direct DILI connections	31
Mean DILI neighbors per target	0.50
<i>Hyperforin comparison:</i>	
Hyperforin targets with ≥ 1 DILI neighbor	10/10 (100%)
Mean DILI neighbors per Hyperforin target	2.4

Table 13: **Table S7. Quercetin target genes in the liver-expressed network.** All 62 Quercetin targets in STRING v12.0 LCC (confidence ≥ 900) with liver TPM ≥ 1 (GTEX v8). Sorted by descending liver expression.

Gene	TPM	Gene	TPM	Gene	TPM	Gene	TPM
CFB	1115	CYP3A4	335	FN1	229	ALDH2	183
ANPEP	160	PPIA	112	SERPINA5	104	CYP1A2	72
CA2	64	APP	63	PYGL	55	HDAC6	45
ESRRRA	42	MAOA	35	AKR1C2	33	AKT1	33
CTSH	28	XDH	26	CHRNA4	25	PIK3R1	24
PIM1	24	LDLR	23	EGFR	17	ELOVL1	18
PKN1	16	GSK3A	13	YES1	13	MET	12
DAPK1	12	BACE1	11	CSNK2A1	10	FSTL1	9
SIRT6	8	GSK3B	7	CDK7	7	CAV2	7
PTPN2	6	CYP1A1	5	PRMT7	5	MMP2	5
AKR1B1	5	PDE6D	5	PTK2	4	ABCG2	4
IQGAP1	4	ADRB2	3	BRAF	4	KDR	3
SRC	3	ALOX5	3	CYP1B1	3	TLR4	3
NUAK1	3	AXL	2	ADA	2	LCK	2
ABCC1	2	PLK1	1	ACHE	1	MMP9	1
SYK	1	PDZK1IP1	1				

Table 14: **Table S8. DILI gene set (82 genes).** Genes in STRING v12.0 LCC (confidence ≥ 900) with liver TPM ≥ 1 (GTEX v8). Source: DisGeNET (UMLS C0860207). Sorted alphabetically.

82 DILI-Associated Genes							
ABCB1	AHR	ALB	ALDOB	AMBP	APOA1	APOE	APOH
ARG1	ARNT	ATG5	BAX	BTD	C3	CAT	CCL2
CLU	COL3A1	CTNNB1	CXCL1	CXCL10	CYP2A6	CYP2C19	CYP2C9
CYP2E1	DGAT2	ENO1	FGA	FLT1	FMO3	GADD45A	GC
GCLC	GPT	GSN	GSTM1	GSTM2	GSTP1	HLA-A	HLA-B
HLA-DQB1	HLA-DRB1	HMGB1	HMOX1	HPD	HPX	IGF1	IL18
IL1R2	KRT18	LCN2	LGALS3	MAP3K5	MED1	MMP2	MTHFR
NAT2	NFE2L2	NR1H3	NR1H4	NR1I2	NR1I3	PLAT	PLG
PNP	POLG	PON1	PPARA	PRKDC	PTGS2	RBP1	SLPI
SNX18	SOD1	SOD3	SPP1	TALDO1	TBXA2R	TCTN1	TF
TTR	UGT1A9						

Table 15: **Table S9. Null distribution parameters from permutation testing.** Null distribution parameters (mean and standard deviation) from $n = 1,000$ degree-matched permutations. Note the tightening of the Quercetin null distribution as the number of targets increases, which drives the inflation of proximity Z-scores.

Metric	Compound	μ_{null}	σ_{null}	x_{obs}	Z-score
<i>Shortest-path proximity (at ≥ 900)</i>					
	Hyperforin (10)	2.21	0.235	1.30	-3.86
	Quercetin (62)	2.17	0.091	1.68	-5.44
<i>Random walk influence (at ≥ 900)</i>					
	Hyperforin (10)	0.0147	0.0098	0.1138	+10.12
	Quercetin (62)	0.0148	0.0038	0.0322	+4.55
<i>Expression-weighted influence (at ≥ 900)</i>					
	Hyperforin (10)	0.0205	0.0125	0.1330	+8.98*
	Quercetin (62)	0.0209	0.0049	0.0493	+5.79*

*Significance remains high despite tissue-specific attenuation. μ_{null} = null distribution mean; σ_{null} = null distribution standard deviation; x_{obs} = observed metric value.