

# 1 Network Insights into Improving Drug Target Inference Algorithms

2 Muying Wang<sup>1</sup>, Heeju Noh<sup>2,3</sup>, Ericka Mochan<sup>1,4</sup> and Jason E. Shoemaker<sup>1,5,6\*</sup>

3 <sup>1</sup>Department of Chemical and Petroleum Engineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA,

4 <sup>2</sup>Department of Systems Biology, Columbia University, New York, NY 10032, USA, <sup>3</sup> Institute for Chemical and Bioengineering, ETH Zurich, Zurich 8093, Switzerland, <sup>4</sup> Department of Mathematics and Data Analytics, Carlow University, Pittsburgh, PA 15213, USA, <sup>5</sup> Department of

5 Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA, <sup>6</sup> The McGowan Institute for  
6 Regenerative Medicine, Pittsburgh, PA 15213, USA, \* Correspondence: [jason.shoemaker@pitt.edu](mailto:jason.shoemaker@pitt.edu)

## 8 9 Abstract

10 To improve the efficacy of drug research and development (R&D), a better understanding of drug  
11 mechanisms of action (MoA) is needed to improve drug discovery. Computational algorithms, such as  
12 ProTINA, that integrate protein-protein interactions (PPIs), protein-gene interactions (PGIs) and gene  
13 expression data have shown promising performance on drug target inference. In this work, we evaluated  
14 how network and gene expression data affect ProTINA's accuracy. Network data predominantly determines  
15 the accuracy of ProTINA instead of gene expression, while the size of an interaction network or selecting  
16 cell/tissue-specific networks have limited effects on the accuracy. However, we found that protein network  
17 betweenness values showed high accuracy in predicting drug targets. Therefore, we suggested a new  
18 algorithm, TREAP (<https://github.com/ImmuSystems-Lab/TREAP>), that combines betweenness values  
19 and adjusted *p*-values for target inference. This algorithm has resulted in higher accuracy than ProTINA  
20 using the same datasets.

## 21 Key words

22 drug target inference, network topology, gene expression, protein-protein interactions, betweenness  
23 centrality

## 24 Introduction

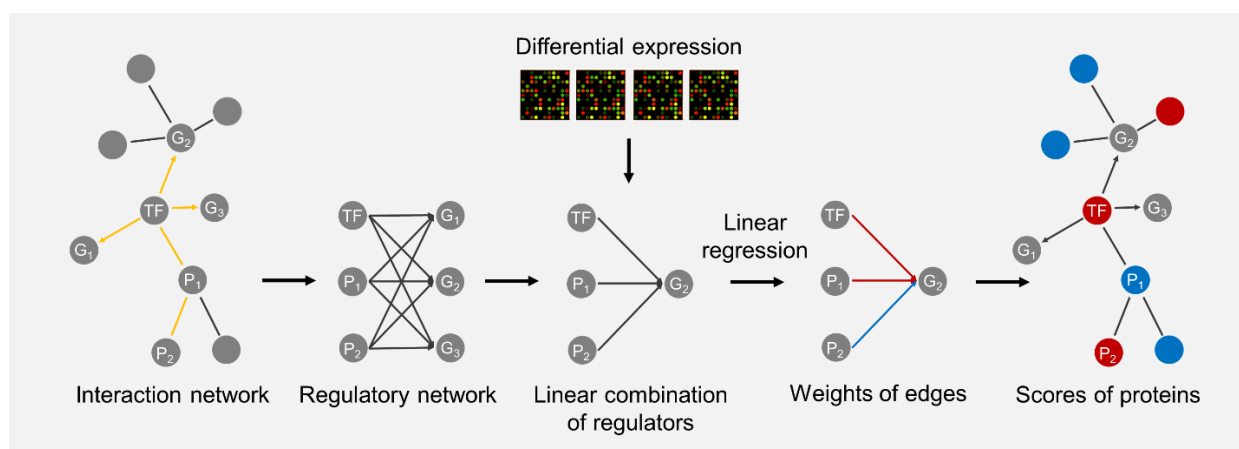
25 The innovation of treatments for diseases remains a challenging task [1-6]. The efficiency of pharmaceutical  
26 research and development (R&D), quantified by the number of new drugs per billion US dollars spent,  
27 dramatically declined from 1950 to 2010 [2]. A large group of drug candidates fail in clinical trials because  
28 they are not effective or safe in humans [2, 5, 7-9]. A major reason is that the systematic effects of drug  
29 candidates are not well studied or modelled in the drug discovery process, and a better understanding of  
30 their mechanisms of action (MoA) can help improve the efficiency of drug R&D [2, 10-12].

31 Two types of computational approaches have been reported to study MoAs of drugs by modeling high-  
32 throughput biological data: comparative analysis and network-based algorithms [13, 14]. Comparative  
33 analysis approaches, such as the Connectivity Map [15], have been used to predict molecular targets of  
34 drugs and assist in drug repurposing [15-20]. They utilize expression profiles as drug signatures and  
35 compare with drugs having known targets, assuming that drugs with high similarities share the same targets.  
36 These approaches much rely on prior knowledge of drugs, thus have limitations in predicting de novo  
37 targets.

38 Network-based algorithms predict drug or disease targets by combining network information and  
39 transcriptomic data [14, 21-27]. Two recent representatives, DeMAND [22] and ProTINA [14], model the  
40 systemic dysregulation of regulatory network caused by a drug treatment, connecting molecular interactions  
41 with differential expression (DE). The regulatory network is generated by using protein-protein interactions

42 (PPIs) and protein-gene interactions (PGIs) obtained from self-curated or public databases, such as  
43 STRING [28] and CellNet [29]. Similar to what has been reported by Noh et al [14], our preliminary  
44 research showed that ProTINA outperforms DeMAND when tested by the same gene expression and  
45 network datasets (Figure S1). Therefore, in this work, we focused on studying ProTINA's performance.

46 For ProTINA, a regulatory network, directing from proteins or transcriptional factors (TFs) to regulated  
47 genes, is generated from input PPIs and PGIs based on certain rules (Figure 1) [14]. The assumption is that  
48 the log fold change (LFC) of a gene is the linear combination of the LFCs of all proteins and TFs that  
49 regulate it. The weights are computed by linear regression methods and then integrated into a score for each  
50 regulator, a protein or a TF. Different from DeMAND, ProTINA may result in negative or positive scores,  
51 representing attenuation or enhancement, respectively. Regulators of larger magnitudes are more likely to  
52 be targets (red nodes in Figure 1). Showing promising results in predicting *in vitro* datasets, DeMAND and  
53 ProTINA have provided a new direction in identifying drug targets and toxicity [14, 22].



54  
55 Figure 1. An overview of ProTINA algorithm. Each node refers to a transcription factor (TF), a non-TF  
56 protein (P) or a gene (G). Arrows present the directions of interactions or edges. The significance of an  
57 edge or protein (including TFs) is color coded, where red refers to high significance while blue refers to  
58 low significance.

59  
60 However, as target inference algorithms become more complicated, it is unclear what roles gene expression  
61 and network data play. A recent study has shown that an accurate description of network topology is able  
62 to cover 65% of the perturbation patterns predicted by a full biochemical model with kinetic parameters  
63 [30]. Several studies have shown that proteins associated with disease and proteins that are drug targets  
64 have significantly different positions within biological networks [31-34]. For target inference algorithms,  
65 it remains an open question as to which kind of biological data most affects the accuracy. Furthermore,  
66 algorithms can infer drug targets in a cell/tissue type-specific manner [14, 22, 27], and it is unknown how  
67 efficient or meaningful cell/tissue type-specific network data is for target inference. Answering these  
68 questions can provide us with insights into future algorithm improvement.

69 In this work, we evaluated the impact of gene expression and network data, using the human B cell  
70 microarray data from the DREAM challenge (referred to as DP14) as our benchmark dataset [35], and  
71 introduced a new algorithm to predict drug targets. Firstly, we found that ProTINA's scores are mostly  
72 determined by network data through permutation tests on gene expression. Secondly, we tested how the  
73 selection of networks affects prediction accuracy. Surprisingly, the effects of size or cell type are negligible.

74 Next, our analysis suggested that network betweenness values can accurately predict drug targets. The  
75 performance is comparable with ProTINA and is consistent regardless of the network size. Lastly, we  
76 proposed TREAP to combine betweenness values and adjusted  $p$ -values from DE for target inference,  
77 which has outperformed ProTINA in accuracy. Moreover, the simplicity of the algorithm makes it more  
78 tractable to users who are not experts in systems and network biology. Our future work will focus on better  
79 balancing both types of data and trying other methods, such as machine learning, to improve prediction  
80 accuracy.

## 81 **Materials and Methods**

### 82 **Gene expression data used in the analysis**

83 The microarray data of human Diffuse Large B-Cell Lymphoma (DLBCL) OCI-LY3 cell line treated with  
84 14 different drugs under diverse doses at 3 time points, 6, 12 and 24 hours post treatment were obtained  
85 from the NCI-DREAM challenge drug synergy dataset, DP14 (GEO accession: GSE51068) [35]. Three  
86 samples treated with ‘Aclacinomycin A’ under a lower dose were dropped due to less significance. The  
87 microarray data of human liver cell line HepG2 treated with 62 genotoxic or non-genotoxic chemicals at  
88 12, 24 and 48 hours post treatment were obtained from literature, referred to as HepG2 in this work (GEO  
89 accession: GSE28878) [36]. The microarray data of mouse pancreatic cells treated with 29 chromatin-  
90 targeting compounds were also obtained from GEO database, referred to as MP (GEO accession:  
91 GSE36379) [37]. For all three datasets, raw data were normalized using the RMA function from the “*affy*”  
92 R package [38]. The  $\log_2$  fold change (LFC) values and Benjamini–Hochberg adjusted  $p$ -values (adjusted  
93  $p$ -values) were calculated by the “*limma*” R package [39]. Probes were mapped to gene symbols by using  
94 the “*hgu219.db*” R package for human microarray data and “*moe430a.db*” for mouse data. Those with the  
95 lowest average BH-adjusted  $p$ -value across all samples were chosen when multiple probes were mapped to  
96 the same gene.

### 97 **Networks used in the analysis and calculation of topological features**

98 Human or mouse PPIs and their associated confidence scores were obtained from the STRING database  
99 [28]. Interactions with experimental proof or from curated databases (the channels of ‘experiments’ and  
100 ‘databases’) were extracted. Interactions transferred from other species or duplicated entries were excluded.  
101 Subnetworks were obtained by applying thresholds ranging from 0.4 to 0.9 to the PPI network, referred to  
102 as PPI04, PPI05, PPI06, PPI07, PPI08 and PPI09, respectively in this paper.

103 Human PGIs and their confidence scores were obtained from the Regulatory Circuits, a database of  
104 predicted, cell/tissue type-specific PGIs [40]. PGI networks of 8 different cell/tissue types were studied in  
105 this analysis: ‘lymphocytes of B lineage’, ‘lymphocytes’, ‘lymphoma’, ‘myeloid leukocytes’, ‘lung’, ‘heart’,  
106 ‘epithelial cells’ and ‘hepatocellular carcinoma cell line’. The network of ‘lymphocytes of B lineage’ were  
107 predicted by samples including those from DLBCL, the same cell line with DP14 [40], thus was chosen as  
108 a reference for analysis of DP14. PGI subnetworks for each cell/tissue type, namely PGI05, PGI10, PGI15  
109 and PGI20, were obtained by thresholds ranging from 0.05 to 0.20, respectively. Mouse PGIs were  
110 compiled from two manually curated databases of transcriptional regulatory networks: TRRUST (version  
111 2) [41] and RegNetwork [42]. These interactions are not cell/tissue type-specific, and no threshold was  
112 applied to them prior to analysis of MP.

113 Degree or betweenness values were calculated by the “*igraph*” R package [43]. PPIs or the combination of  
114 PPIs and PGIs (PPI+PGI) were treated as undirected graphs, while PGIs were treated as directed graphs.

### 115 **Reference drug targets**

116 The reference targets of each chemical were extracted from STITCH database (version 5.0) [44, 45] for  
117 analyses of DP14, HepG2 and MP. Only targets with experimental proof or from curated databases were  
118 collected as shown in Table S1.

### 119 **Prediction of drug targets by ProTINA**

120 LFC values, PPI and PGI subnetworks were analyzed by “*protina*” R package [14]. Slope matrices of each  
121 time point were calculated following the user manual. For samples with only two timepoints, control  
122 samples served as 0hr post treatment to calculate associated slope matrices. Samples from different doses  
123 for the same drug were treated as separate groups.

### 124 **Prediction of drug targets by TREAP**

125 For target prediction by TREAP, the assumption is that genes with high betweenness values or low adjusted  
126  $p$ -values are more likely to be drug targets. Adjusted  $p$ -values and PPI+PGI betweenness values were  
127 calculated as explained in the former sections. Ranks of genes were obtained by sorting betweenness values  
128 and adjusted  $p$ -values, respectively, and genes with the same betweenness or adjusted  $p$ -value shared the  
129 same rank. Final scores were calculated by summing up the ranks from both metrics for each gene. In this  
130 work, all analyses on TREAP used 0.9 as the threshold for human or mouse PPIs and 0.20 for human PGIs.  
131 No threshold was applied to mouse PGIs.

### 132 **Calculation and comparison of AUROC values**

133 Area under the receiving operator characteristics (AUROC) values in this paper were calculated by  
134 comparing scored proteins with reference drug targets through the “*pROC*” R package [46]. As ProTINA  
135 scores can be positive or negative, the absolute scores were used to calculate AUROC. The median AUROC  
136 across all drugs in each dataset was calculated to represent accuracy of a whole test. For drugs having more  
137 than one doses, the AUROC values of low doses were excluded. In terms of topological features, degree or  
138 betweenness values were directly used to calculate AUROC values without pre-processing. TREAP scores  
139 were directly used for calculation of AUROC without preprocessing. Difference in any pair of chosen tests  
140 were computed by performing pairwise  $t$ -test between their AUROC values. A  $p$ -value less than 0.05 were  
141 regarded as significantly different.

### 142 **Permutation tests on gene expression**

143 The null hypothesis for permutation tests in this work is that the median AUROCs of randomized gene  
144 expression are smaller than that of nonrandomized gene expression, and the  $p$ -values were calculated  
145 accordingly. For ProTINA, gene labels for DP14 that refer to the rows of LFC and associated slope matrices  
146 were randomly shuffled for 1000 times. Randomized data were applied to ProTINA under the same network  
147 setup, PPI09 and PGI20. For TREAP, gene labels of adjusted  $p$ -values for each dataset, namely DP14,  
148 HepG2 and MP were randomly shuffled for 1000 times, respectively, and drug targets were predicted using  
149 PPI09 and PGI20 (for MP no threshold was applied). AUROC and median values were calculated as  
150 explained in the former section.

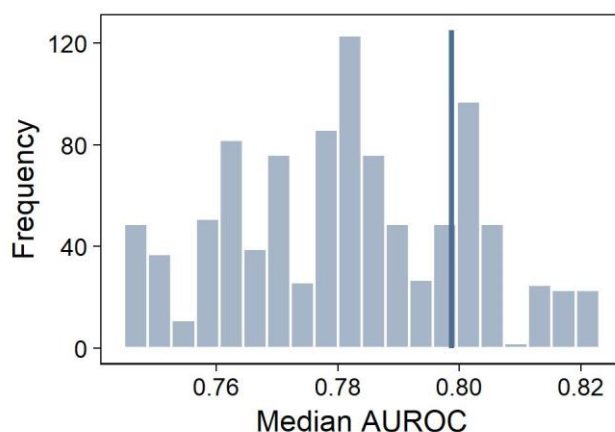
## 151 **Results**

### 152 **Permutation tests show that ProTINA is predominantly determined by network data**

153 To understand how much network or gene expression data contribute to ProTINA’s accuracy, we performed  
154 1000 permutation tests by randomizing the LFC gene expression values (Materials and Methods). To  
155 shorten the computation time on ProTINA, the smallest PPI and PGI subnetworks (PPI09 and PGI20) were  
156 chosen for this analysis (discussed more in the following section). For prediction scores obtained from

157 ProTINA, the area under the receiver operating characteristics (AUROC) values were calculated per drug,  
158 and the median AUROC across all drugs is used as a metric for each test.

159 The median AUROC obtained by nonrandomized LFCs is 0.799. As shown in Fig. 2, 220 of 1000  
160 permutation tests have higher median AUROC values than that (one-tailed  $p$ -value = 0.221). Most tests  
161 have similar accuracy to the original test. This indicates that randomizing LFCs does not diminish  
162 ProTINA's accuracy significantly, therefore, network data determines most of ProTINA's performance.



163  
164 Figure 2. 1000 Permutation tests were performed by randomizing the gene expression and calculating the  
165 median AUROCs. The blue vertical line refers to the median AUROC obtained by nonrandomized gene  
166 expression.

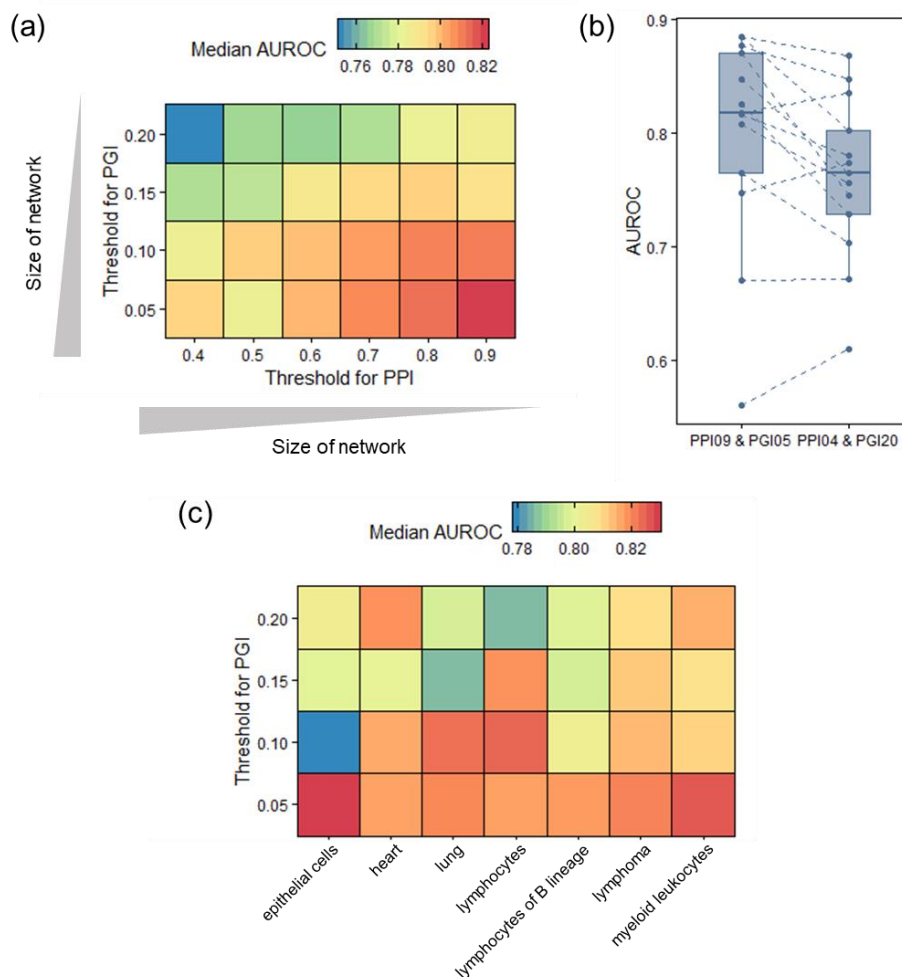
### 167 Selection of networks has limited effects in the prediction accuracy of ProTINA

168 Next, we studied how the selection of PPIs and PGIs affects target inference accuracy. PPI and PGI  
169 subnetworks of different sizes or cell/tissue types were tested using the same gene expression data from  
170 DP14 (Materials and Methods). Similar to permutation tests, the median AUROC represents the accuracy  
171 for each PPI-PGI combination.

172 In total, 24 PPI-PGI combinations of different sizes were tested on ProTINA. As shown in Figure 3a,  
173 ProTINA favors small PPI and large PGI subnetworks. The combination of PPI09 and PGI05 shows the  
174 highest accuracy, and its median AUROC is 0.821. As the threshold increases from 0.4 to 0.9, the number  
175 of interactions in PPI subnetwork ranges from 380375 to 281357 (Figure S2), and the median AUROC  
176 increases for most tests. For example, the median AUROC increases from 0.785 (PPI04) to 0.811 (PPI09)  
177 for analyses under PGI10. But there are PPI subnetworks that do not follow this trend. Under PGI05, PPI05  
178 shows lower median AUROC than PPI04 (0.784 and 0.796, respectively). For PGI subnetworks, as the  
179 number of interactions range from 123394 to 5932 (Figure S2), the median AUROC shows an opposite  
180 trend to that of PPI subnetworks (Figure 3a). An example is that the median AUROC decreases from 0.821  
181 (PGI05) to 0.785 (PGI20) when using PPI09. Most tests show a consistent trend except for those using  
182 PPI05. A possible reason is that new proteins and associated interactions are included in the network as the  
183 threshold for PGIs changes from 0.05 to 0.10. So that the network topology is changed, and predictions  
184 from ProTINA are affected accordingly.

185 However, median AUROC values vary in a small range while the size of PPI or PGI subnetworks changes  
186 significantly (Figure 3a, 3b and Figure S3). The highest and lowest median AUROC values for ProTINA  
187 are 0.821 and 0.753, with a difference less than 0.1 (Figure 3b). In addition, most of these differences are

188 insignificant. When comparing with the test using PPI09 and PGI20, none of the 24 tests are significant in  
189 AUROC values ( $p$ -value  $> 0.05$  for all tests), although the combination of PPI04 and PGI15 has resulted in  
190 a pairwise  $p$ -value of 0.057. Furthermore, we studied the effects on standard deviations (SDs) of AUROC  
191 values across 12 drugs for each test. All of them maintain at a low level below 0.13. In summary, we  
192 conclude that the size of networks has limited effects on the prediction accuracy of ProTINA, while small  
193 PPI and large PGI networks tend to improve the accuracy.



194  
195 Figure 3. Prediction accuracy of ProTINA using networks of different sizes or cell/tissue types. (a) PPI or  
196 PGI subnetworks of different sizes were tested on ProTINA to predict targets for DP14. The axes refer to  
197 the confidence thresholds for PPI (x axis) and PGI (y axis) subnetworks, and the median AUROC values  
198 are the metric for prediction accuracy. Among all PPI-PGI subnetwork combinations, PPI09-PPI05 and  
199 PPI04-PPI20 have the highest and lowest accuracy in terms of median AUROC values, respectively. Panel  
200 (b) shows the boxplot of these two groups. Each dot represents the AUROC of a drug. (c) PGI subnetworks  
201 of 7 cell/tissue types were applied to ProTINA for target prediction.

202  
203 To analyze the performance of cell/tissue type-specific networks, 28 tests using PGIs from 7 cell/tissue  
204 types were performed on ProTINA using the same PPI subnetwork, PPI09. Most tests counterintuitively  
205 show similar median AUROC values regardless of cell/tissue types (Figure 3c, Figure S4). In theory, the

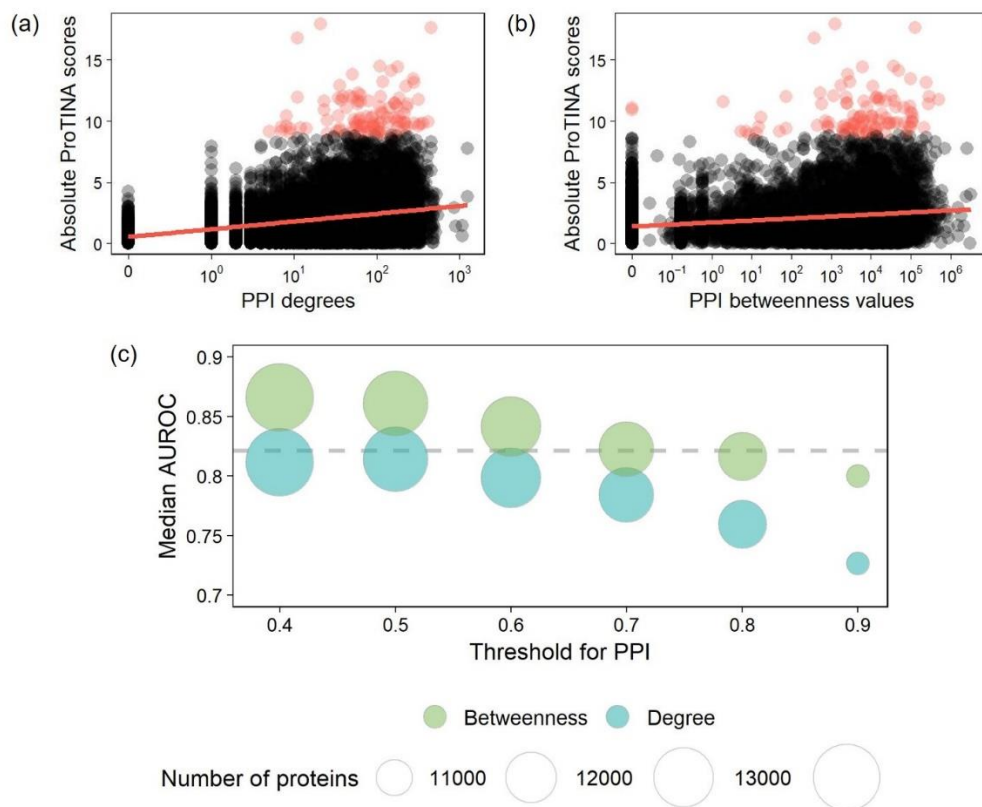
206 PGI subnetworks for immune cells should have higher accuracy than non-immune cell types, and those for  
207 ‘lymphocytes of B cell lineage’ should outperform other immune cells. This is because that samples from  
208 DLBCL, the same cell line with DP14, were used to predict the interactions for ‘lymphocytes of B lineage’  
209 [40] (Materials and Methods). However, using the AUROC values from PGI20 for ‘lymphocytes of B cell  
210 lineage’ as a reference, no other cell/tissue types are significantly different (pairwise  $p$ -values  $> 0.05$  for  
211 all) under the same network setup. In conclusion, we have shown that the selection of PPIs or PGIs in terms  
212 of either the size or cell/tissue type is not the key factor to prediction accuracy of ProTINA.

### 213 **Topological features have similar prediction accuracy to ProTINA, and protein betweenness** 214 **outperforms degree**

215 Our findings have shown that ProTINA depends on network topology more than gene expression, and that  
216 it has consistent performance regardless of the network size or the cell/tissue type the network represents.  
217 These suggest that ProTINA is probably determined by some network topological feature that remains  
218 relatively stable across different PPI or PGI subnetworks, such as protein degree or betweenness. The  
219 degree of a protein is the number of proteins/genes with which it interacts, while the betweenness is a  
220 measure of bottleneckedness, e.g. the amount of information flowing through the proteins that connect the  
221 rest of the network. Analyses of these features and their effects on drug target prediction may provide  
222 meaningful insights on improving prediction accuracy.

223 To test our hypotheses, we studied degree and betweenness values for PPIs, PGIs and PPI-PGI  
224 combinations (referred to as PPI+PGI in the following text). Firstly, for PPIs, we compared scores obtained  
225 from ProTINA (using PPI09 and PGI20) with their associated protein degrees or betweenness values in  
226 PPI09 for each drug. The majority of the drugs show a weak but evident correlation between absolute  
227 ProTINA scores and protein degrees, however, the correlation for betweenness is much lower (Table S2).  
228 For instance, the correlation coefficient is 0.211 for ‘Rapamycin’ (Figure 4a), while the correlation of  
229 betweenness values is smaller than that of degrees, which is 0.085 for the same drug (Figure 4b). Notice  
230 that a large portion of the top 100 proteins scored by ProTINA (red points in Figure 4a, b) lie in the group  
231 of high degree or betweenness values.

232 We next tried to predict drug targets by using PPI degree or betweenness values without considering gene  
233 expression or PGIs. The assumption is that proteins with higher degree or betweenness values are more  
234 likely to be targets. As shown in Figure 4c, the median AUROC values for PPI degree or betweenness  
235 values are close to those for ProTINA. What’s more, betweenness values perform better than degrees. The  
236 highest and lowest median AUROC values for degrees are 0.814 and 0.727 (Figure 4c), and those for  
237 betweenness values are 0.866 and 0.800, even higher than associated median AUROC values for ProTINA  
238 under the same network setup. As the size of PPI subnetworks shrinks, the median AUROCs for degrees  
239 decreases accordingly (Figure 4c), with a correlation coefficient value of -0.950 between the medians and  
240 thresholds. While the decrease of network size also diminishes the accuracy of betweenness, the median  
241 AUROCs remain higher than those of degrees and decrease relatively slower.



242

243 Figure 4. (a) Degree or (b) betweenness values of proteins in PPIs were compared with associated ProTINA  
244 scores for ‘Rapamycin’. The correlation coefficient is 0.211 for absolute ProTINA scores and degrees, and  
245 that for betweenness values is 0.085. Red points refer to the top 100 proteins scored by ProTINA. (c) The  
246 degree and betweenness values were used to predict drug targets assuming higher scores are more likely to  
247 be targets. Each point shows the median AUROC value and the number of proteins under a PPI threshold.  
248 For reference, the grey dashed line refers to the highest median AUROC achieved by ProTINA, which was  
249 obtained from using PPI09 and PGI05.

250

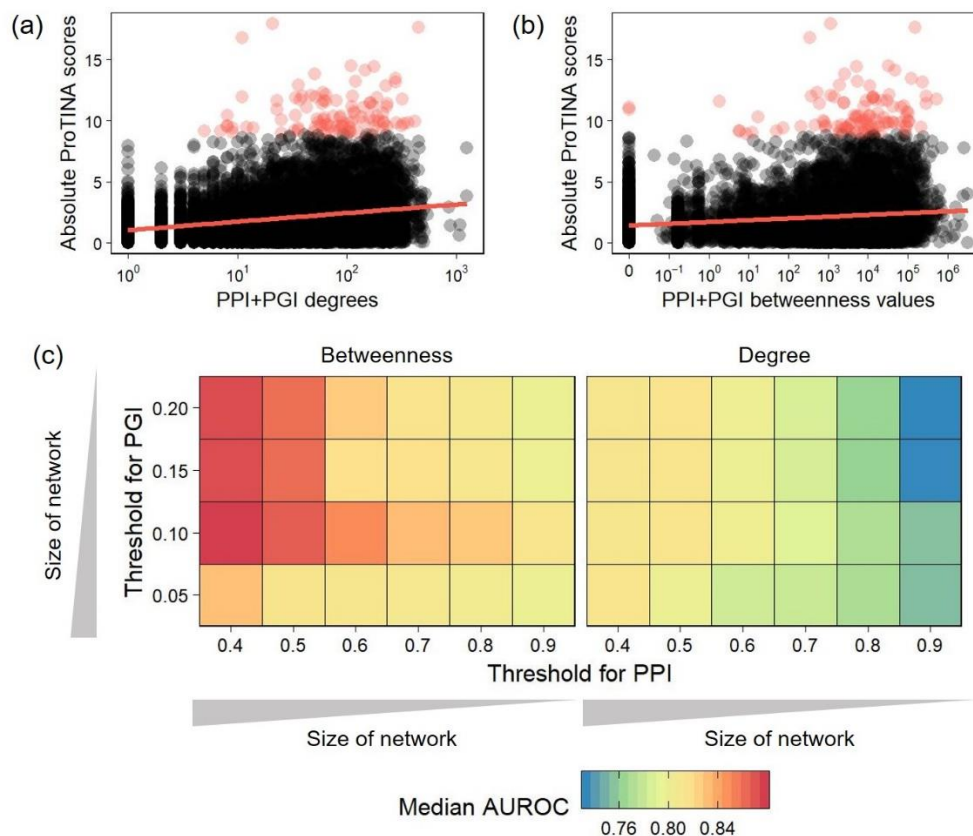
251 Secondly, the degree and betweenness values of PGI subnetworks were also compared with associated  
252 ProTINA scores, however, there are no clear trends between them (Table S2). In addition, drug target  
253 prediction based on PGI degree or betweenness values are not comparable with that by PPI topological  
254 features (Figure S5). This might be related to the limited amount of PGI interactions.

255 Lastly, we calculated topological features for PPI+PGI and compared them with ProTINA scores. As  
256 expected, they show the same trend with PPIs (Table S2). The correlation coefficient between degrees and  
257 ProTINA scores is 0.208 for ‘Rapamycin’, while that for betweenness values is 0.079 (Figure 5a, b).

258 Predicting drug targets by PPI+PGI degree or betweenness values results in higher median AUROC values  
259 than those for PPIs. In addition, for all thresholds of PPIs or PGIs applied to this analysis, betweenness  
260 values outperform degrees. The accuracy for PPI+PGI degrees ranges from 0.833 to 0.733 in terms of  
261 median AUROC values, and that for betweenness values ranges from 0.878 to 0.782 (Figure 5c). As the  
262 size of PPI or PGI subnetworks decreases, the median AUROC values for PPI+PGI degrees also decreases.  
263 PPI+PGI betweenness values have the same behavior as the size of PPIs changes, while for PGIs the trend



264 is less evident. PGI10 has the best performance in parallel comparisons. In summary, betweenness values  
265 well predict drug targets and show even higher accuracy than ProTINA.



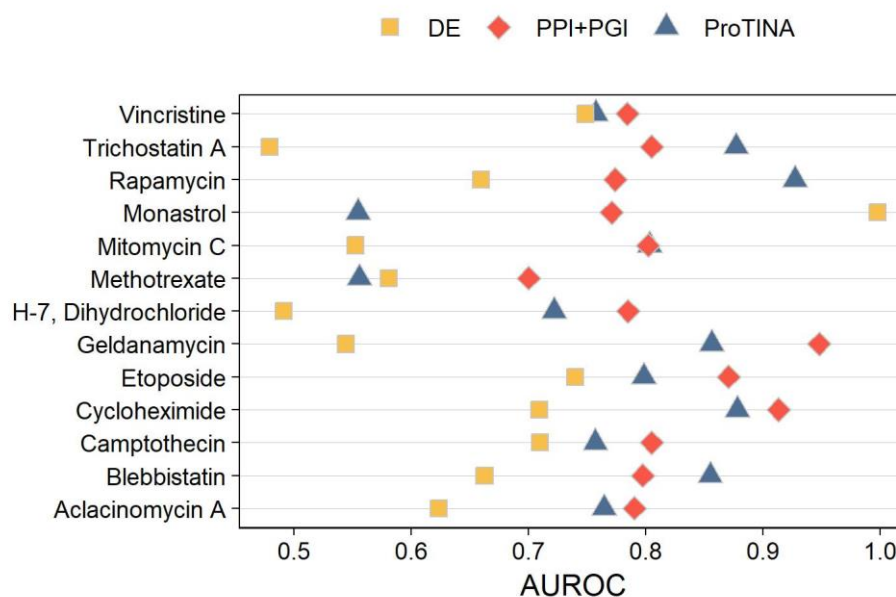
266  
267 Figure 5. (a) PPI+PGI degree or (b) betweenness values were compared with associated ProTINA scores  
268 for ‘Rapamycin’. The correlation coefficient for absolute ProTINA scores versus degrees and absolute  
269 ProTINA scores versus betweenness values is 0.208 and 0.079, respectively. Red points refer to the top 100  
270 proteins scored by ProTINA. (c) The degree and betweenness values were used as measures to predict drug  
271 targets, and the median AUROC values were calculated for each prediction. The axes refer to the confidence  
272 thresholds for PPI (x axis) and PGI (y axis) subnetworks.

### 273 Missing information in network topology can be covered by differential expression

274 Topological features, degree or betweenness values, have shown high prediction accuracy without taking  
275 gene expression into account. Our permutation tests have also indicated that network data has more effects  
276 in ProTINA’s performance than gene expression data. What’s more, ProTINA has much better performance  
277 than differential expression (DE) analysis on drug target prediction according to prior research [14]. All of  
278 the above has raised a question about whether gene expression data can help to predict drug targets. To  
279 address this concern, we compared DE analysis (adjusted *p*-values, Materials and Methods) with two other  
280 target prediction methods in terms of their performance on each drug: PPI+PGI betweenness and ProTINA.

281 We calculated AUROC values for all three methods using the same network setup, PPI09 and PGI20  
282 (Figure 6). For most drugs, such as ‘Mitomycin C’ or ‘Cycloheximide’, PPI+PGI betweenness and  
283 ProTINA have close AUROC values, and they outperform DE analysis. Consistent behaviors between  
284 ProTINA and PPI+PGI further indicates the impact of network topology on ProTINA’s accuracy. In  
285 contrast to these drugs, DE has much higher prediction accuracy than PPI+PGI or ProTINA for ‘Monastrol’

286 (the AUROCs are 0.998, 0.771 and 0.555, respectively). This means that DE analysis of gene expression  
287 data can capture information missing in network topology, and that it is necessary to include gene  
288 expression data for drug target inference and improvement of accuracy.



289  
290 Figure 6. AUROC values of each drug obtained from three different methods: differential expression (DE)  
291 analysis by adjusted  $p$ -values, betweenness values from the combination of PPI09 and PGI20 (PPI+PGI)  
292 and ProTINA analysis by PPI09 and PGI20.

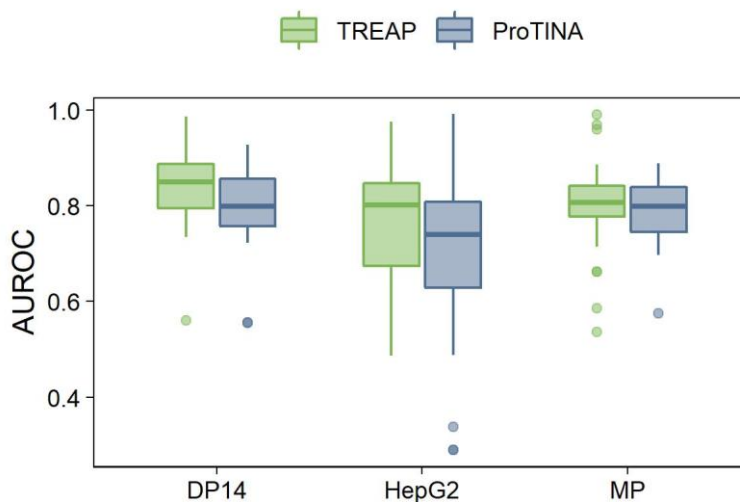
### 293 A novel algorithm that combines network topology and DE analysis for target inference

294 To better combine network topology and DE analysis and improve inference accuracy, we suggest TREAP  
295 (target inference by ranking betweenness values and addjusted p-values) to predict drug targets. There are  
296 three steps for this algorithm. The first step is to calculate PPI+PGI betweenness values and obtain adjusted  
297  $p$ -values from DE analysis. For gene expression profiles with multiple timepoints, DE analysis can be  
298 performed per time point or across all timepoints. The second step is to calculate the ranks of genes by  
299 sorting betweenness values and adjusted  $p$ -values, respectively. Genes with high betweenness values or low  
300 adjusted  $p$ -values are scored with high ranks. The third step is to generate final scores by summing up the  
301 ranks from both metrics for each gene. Genes with higher scores are more likely to be targets for associated  
302 drugs.

303 TREAP was tested by three different gene expression profiles: (i) DP14 [35], (ii) human HepG2 cells treated  
304 with genotoxic or non-genotoxic chemicals, referred to as HepG2 in this work [36] and (iii) mouse  
305 pancreatic cell lines treated with chromatin-targeting compounds, referred to as MP [37]. Human and mouse  
306 PPIs were obtained from STRING [28], and PGIs were obtained from Regulatory Circuits [40], TRRUST  
307 (version 2) [41] and RegNetwork [42]. AUROCs were calculated by comparing scored genes with known  
308 targets for each test as a measurement of accuracy.

309 TREAP shows stable performance and maintains high accuracy for all datasets tested in this study (median  
310 AUROCs  $> 0.800$ , Figure 7). While TREAP takes significantly less computation time than ProTINA, it has  
311 higher median AUROCs when compared with ProTINA under the same dataset (Figure 7). For DP14, the  
312 median AUROC of TREAP is 0.850, higher than that of ProTINA, 0.799 ( $p$ -value = 0.11). Notice that it is

313 also higher than using PPI+PGI betweenness values alone, which is 0.798. TREAP significantly  
314 outperforms ProTINA in HepG2. The median AUROC is significantly improved from 0.739 to 0.801, with  
315 a  $p$ -value of 0.0002. For MP, TREAP and ProTINA have close median AUROCs as 0.806 and 0.799,  
316 respectively ( $p$ -value = 0.39). By integrating betweenness values and adjusted  $p$ -values to represent both  
317 network topology and DE analysis, TREAP is comparable with and sometimes better than ProTINA in  
318 accuracy for all datasets analyzed in this work. In addition, we performed 1000 permutation tests on TREAP  
319 by randomizing the adjusted  $p$ -values for each dataset. Different from ProTINA, TREAP is significant when  
320 compared with permutation tests on the adjusted  $p$ -values from DP14, with the one-tailed  $p$ -value as 0.007  
321 (Figure S6). For HepG2, the one-tailed  $p$ -value is 0.058, while for MP, TREAP is less significant and shows  
322 a one-tailed  $p$ -value as 0.314.



323  
324 Figure 7. AUROC values of TREAP and ProTINA predictions for different gene expression profiles:  
325 human lymphoma cells (DP14), human liver cancer cells (HepG2) and mouse pancreatic cells (MP). ( $p$ -  
326 values = 0.11, 0.0002 and 0.39, respectively)

## 327 Discussion

328 Our analyses have shown that, even though ProTINA requires both gene expression and network data for  
329 inputs, network data predominantly determines accuracy of drug target inference. What's more, the  
330 cell/tissue type or size of a network has limited impact on ProTINA's accuracy, while topology, especially  
331 the degree value, affects the performance of ProTINA more.

332 However, ProTINA has two limitations due to the reliance on network topology alone and the connection  
333 with protein degrees. First, PPIs, a major part of network data, have a known bias toward protein abundance  
334 [33, 47]. It has been reported that interactions obtained from high-throughput experiments have a  
335 correlation between the protein degree and abundance. Second, we have shown that differential analysis of  
336 gene expression data can uncover meaningful information missing in network topology.

337 To address these two limitations, we suggested a new algorithm, TREAP, which combines protein  
338 betweenness values and adjusted  $p$ -values, representing information from both sources of network topology  
339 and DE analysis, to predict drug targets. We chose betweenness values because they are less sensitive to  
340 network sizes and more accurate than degrees in target prediction. TREAP shows more consistent  
341 performance than ProTINA when tested by different gene expression profiles and maintains a median

342 AUROC above 0.800. In addition, TREAP takes significantly less computation time than ProTINA, and its  
343 simplicity makes it more tractable to users who are not experts in systems and network biology. It is also  
344 flexible in dealing with samples of limited or multiple timepoints as adjusted  $p$ -values can be calculated per  
345 timepoint or across all timepoints based on user's needs. However, ProTINA needs at least two timepoints  
346 to fully take advantage of the algorithm [14]. Currently, betweenness values and adjusted  $p$ -values are  
347 weighted equally for TREAP. Future work should focus on better balancing both types of data and trying  
348 other scoring methods to improve prediction accuracy.

349 TREAP is presented here as an alternative approach to ProTINA, but it is worth emphasizing advantages  
350 specific to each algorithm. As stated above, TREAP is significantly faster, and the algorithm is not complex,  
351 enabling users from several branches of research to access the tool and understand the findings. ProTINA,  
352 however, is a mechanistically derived algorithm, which allows users with expertise in computational  
353 biology to dive deeper into the possible mechanisms of a drug's activity. The accuracy of their predictions  
354 is similar when measured using the AUROC, but the permutation tests presented here suggest that TREAP  
355 is more likely to use drug-specific gene expression to make a more accurate prediction.

356 TREAP and its derivatives have potential in a variety of applications for drug innovation. First, it can assist  
357 in selection of drug candidates and serve as a preliminary test of the efficacy or safety by connecting with  
358 databases for functional annotations, e.g. Gene Ontology [48, 49]. Studying predicted targets can help  
359 exclude poorly targeting drug candidates or those causing severe damage to biological systems. Second,  
360 the algorithm can be applied to drug repurposing by exploring published datasets characterizing drug  
361 treatments, assuming that a pair of drugs sharing the same group of predicted targets can be used to treat  
362 the same disease. Last but not the least, the algorithm can help to discover disease mechanisms [14]. Similar  
363 to drug treatments, diseases can also be treated as a type of perturbation to the biological system of interest.  
364 Predicting disease targets may assist in identifying key components of disease mechanisms and pathology,  
365 which is crucial for innovations in disease treatment [10].

366

367 **Author Contributions:** conceptualization, J.E.S. and M.W.; methodology, M.W.; software, M.W.; formal  
368 analysis, M.W.; investigation, M.W. and E.M.; writing—original draft preparation, M.W.; writing—review  
369 and editing, J.E.S., E.M. and N.H.; visualization, M.W.; supervision, J.E.S.

370 **Funding:** This research received no external funding.

371 **Conflicts of Interest:** The authors declare no conflict of interest.

372 **Supplementary Materials:** Figure S1: Comparing the accuracy of ProTINA and DeMAND; Figure S2:  
373 The number of interactions under different thresholds for PPIs and PGIs; Figure S3: Performance of  
374 ProTINA on DP14 using PPIs and PGIs of different sizes; Figure S4: Performance of ProTINA on DP14  
375 using PGIs from different cell/tissue types; Figure S5: Predicting drug targets by degree or betweenness  
376 values of PGIs; Figure S6: Median AUROCs of permutation tests on TREAP. Table S1: Reference drug  
377 targets for each analysis; Table S2: Correlation coefficients between degree or betweenness values and  
378 ProTINA scores for each drug; Table S3: AUROCs of each drug for ProTINA, topological features and  
379 TREAP; Table S4: Degree and betweenness values of proteins under different thresholds.

380

381 **References**

- 382 1. Schuhmacher, A., O. Gassmann, and M. Hinder, *Changing R&D models in research-based*  
383 *pharmaceutical companies*. J Transl Med, 2016. **14**(1): p. 105.
- 384 2. Scannell, J.W., et al., *Diagnosing the decline in pharmaceutical R&D efficiency*. Nat Rev Drug  
385 Discov, 2012. **11**(3): p. 191-200.
- 386 3. Pushpakom, S., et al., *Drug repurposing: progress, challenges and recommendations*. Nat Rev  
387 Drug Discov, 2019. **18**(1): p. 41-58.
- 388 4. Paul, S.M., et al., *How to improve R&D productivity: the pharmaceutical industry's grand*  
389 *challenge*. Nat Rev Drug Discov, 2010. **9**(3): p. 203-14.
- 390 5. DiMasi, J.A., H.G. Grabowski, and R.W. Hansen, *Innovation in the pharmaceutical industry: New*  
391 *estimates of R&D costs*. J Health Econ, 2016. **47**: p. 20-33.
- 392 6. Cook, D., et al., *Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional*  
393 *framework*. Nat Rev Drug Discov, 2014. **13**(6): p. 419-31.
- 394 7. Hay, M., et al., *Clinical development success rates for investigational drugs*. Nat Biotechnol,  
395 2014. **32**(1): p. 40-51.
- 396 8. Hwang, T.J., et al., *Failure of Investigational Drugs in Late-Stage Clinical Development and*  
397 *Publication of Trial Results*. JAMA Intern Med, 2016. **176**(12): p. 1826-1833.
- 398 9. Harrison, R.K., *Phase II and phase III failures: 2013-2015*. Nat Rev Drug Discov, 2016. **15**(12): p.  
399 817-818.
- 400 10. Spagnolo, P. and T.M. Maher, *Clinical trial research in focus: why do so many clinical trials fail in*  
401 *IPF?* Lancet Respir Med, 2017. **5**(5): p. 372-374.
- 402 11. Naci, H. and J.P. Ioannidis, *How good is "evidence" from clinical studies of drug effects and why*  
403 *might such evidence fail in the prediction of the clinical utility of drugs?* Annu Rev Pharmacol  
404 Toxicol, 2015. **55**: p. 169-89.
- 405 12. Gashaw, I., et al., *What makes a good drug target?* Drug Discovery Today, 2012. **17**: p. S24-S30.
- 406 13. Chua, H.N. and F.P. Roth, *Discovering the targets of drugs via computational systems biology*. J  
407 Biol Chem, 2011. **286**(27): p. 23653-8.
- 408 14. Noh, H., J.E. Shoemaker, and R. Gunawan, *Network perturbation analysis of gene transcriptional*  
409 *profiles reveals protein targets and mechanism of action of drugs and influenza A viral infection*.  
410 Nucleic Acids Res, 2018.
- 411 15. Lamb, J., et al., *The Connectivity Map: using gene-expression signatures to connect small*  
412 *molecules, genes, and disease*. Science, 2006. **313**(5795): p. 1929-35.
- 413 16. Ganter, B., et al., *Development of a large-scale chemogenomics database to improve drug*  
414 *candidate selection and to understand mechanisms of chemical toxicity and action*. J Biotechnol,  
415 2005. **119**(3): p. 219-44.
- 416 17. Wang, M., C. Tang, and J. Chen, *Drug-Target Interaction Prediction via Dual Laplacian Graph*  
417 *Regularized Matrix Completion*. Biomed Res Int, 2018. **2018**: p. 1425608.
- 418 18. Vertes, A., et al. *Inferring Mechanism of Action of an Unknown Compound from Time Series*  
419 *Omics Data*. 2018. Cham: Springer International Publishing.
- 420 19. Wolpaw, A.J., et al., *Modulatory profiling identifies mechanisms of small molecule-induced cell*  
421 *death*. Proc Natl Acad Sci U S A, 2011. **108**(39): p. E771-80.
- 422 20. Iorio, F., et al., *Transcriptional data: a new gateway to drug repositioning?* Drug Discov Today,  
423 2013. **18**(7-8): p. 350-7.
- 424 21. Rush, S.T.A. and D. Repsilber, *Capturing context-specific regulation in molecular interaction*  
425 *networks*. BMC Bioinformatics, 2018. **19**(1): p. 539.
- 426 22. Woo, J.H., et al., *Elucidating Compound Mechanism of Action by Network Perturbation Analysis*.  
427 Cell, 2015. **162**(2): p. 441-451.
- 428 23. Chen, E.Y., et al., *Expression2Kinases: mRNA profiling linked to multiple upstream regulatory*  
429 *layers*. Bioinformatics, 2012. **28**(1): p. 105-11.

- 430 24. Koido, M., et al., *InDePTH: detection of hub genes for developing gene expression networks*  
431 *under anticancer drug treatment*. *Oncotarget*, 2018. **9**(49): p. 29097-29111.
- 432 25. Ji, X., J.M. Freudenberg, and P. Agarwal, *Integrating Biological Networks for Drug Target*  
433 *Prediction and Prioritization*. *Methods Mol Biol*, 2019. **1903**: p. 203-218.
- 434 26. Cosgrove, E.J., et al., *Predicting gene targets of perturbations via network-based filtering of*  
435 *mRNA expression compendia*. *Bioinformatics*, 2008. **24**(21): p. 2482-90.
- 436 27. Failli, M., J. Paananen, and V. Fortino, *Prioritizing target-disease associations with novel safety*  
437 *and efficacy scoring methods*. *Sci Rep*, 2019. **9**(1): p. 9852.
- 438 28. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage,*  
439 *supporting functional discovery in genome-wide experimental datasets*. *Nucleic Acids Res*, 2019.  
440 **47**(D1): p. D607-D613.
- 441 29. Cahan, P., et al., *CellNet: network biology applied to stem cell engineering*. *Cell*, 2014. **158**(4): p.  
442 903-915.
- 443 30. Santolini, M. and A.L. Barabasi, *Predicting perturbation patterns from the topology of biological*  
444 *networks*. *Proc Natl Acad Sci U S A*, 2018. **115**(27): p. E6375-E6383.
- 445 31. Zhu, M., et al., *The analysis of the drug-targets based on the topological properties in the human*  
446 *protein-protein interaction network*. *Journal of Drug Targeting*, 2009. **17**(7): p. 524-532.
- 447 32. Lopes, T.J., et al., *Identifying problematic drugs based on the characteristics of their targets*.  
448 *Front Pharmacol*, 2015. **6**: p. 186.
- 449 33. Ackerman, E.E., et al., *Network-Guided Discovery of Influenza Virus Replication Host Factors*.  
450 *MBio*, 2018. **9**(6).
- 451 34. Feng, Y., Q. Wang, and T. Wang, *Drug Target Protein-Protein Interaction Networks: A Systematic*  
452 *Perspective*. *Biomed Res Int*, 2017. **2017**: p. 1289259.
- 453 35. Bansal, M., et al., *A community computational challenge to predict the activity of pairs of*  
454 *compounds*. *Nat Biotechnol*, 2014. **32**(12): p. 1213-22.
- 455 36. Magkoufopoulou, C., et al., *A transcriptomics-based in vitro assay for predicting chemical*  
456 *genotoxicity in vivo*. *Carcinogenesis*, 2012. **33**(7): p. 1421-9.
- 457 37. Kubicek, S., et al., *Chromatin-targeting small molecules cause class-specific transcriptional*  
458 *changes in pancreatic endocrine cells*. *Proc Natl Acad Sci U S A*, 2012. **109**(14): p. 5364-9.
- 459 38. Gautier, L., et al., *affy—analysis of Affymetrix GeneChip data at the probe level*. *Bioinformatics*,  
460 2004. **20**(3): p. 307-315.
- 461 39. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and*  
462 *microarray studies*. *Nucleic Acids Research*, 2015. **43**(7): p. e47-e47.
- 463 40. Marbach, D., et al., *Tissue-specific regulatory circuits reveal variable modular perturbations*  
464 *across complex diseases*. *Nat Methods*, 2016. **13**(4): p. 366-70.
- 465 41. Han, H., et al., *TRRUST v2: an expanded reference database of human and mouse transcriptional*  
466 *regulatory interactions*. *Nucleic Acids Res*, 2018. **46**(D1): p. D380-D386.
- 467 42. Liu, Z.P., et al., *RegNetwork: an integrated database of transcriptional and post-transcriptional*  
468 *regulatory networks in human and mouse*. *Database (Oxford)*, 2015. **2015**.
- 469 43. Csardi, G. and T. Nepusz, *The Igraph Software Package for Complex Network Research*.  
470 *InterJournal*, 2005. **Complex Systems**: p. 1695.
- 471 44. Kuhn, M., et al., *STITCH: interaction networks of chemicals and proteins*. *Nucleic acids research*,  
472 2008. **36**(Database issue): p. D684-D688.
- 473 45. Szklarczyk, D., et al., *STITCH 5: augmenting protein-chemical interaction networks with tissue*  
474 *and affinity data*. *Nucleic acids research*, 2016. **44**(D1): p. D380-D384.
- 475 46. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves*.  
476 *BMC Bioinformatics*, 2011. **12**(1): p. 77.

- 477 47. Ivanic, J., et al., *Influence of protein abundance on high-throughput protein-protein interaction*  
478 *detection*. PLoS One, 2009. **4**(6): p. e5815.
- 479 48. The Gene Ontology, C., *The Gene Ontology Resource: 20 years and still GOing strong*. Nucleic  
480 Acids Res, 2019. **47**(D1): p. D330-D338.
- 481 49. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology*  
482 *Consortium*. Nat Genet, 2000. **25**(1): p. 25-9.

483

484