Predicting the Health Impact of Dietary Polyphenols Using a Network Medicine Framework — Source link ⬈

Italo Faria do Valle, Harvey G. Roweth, Harvey G. Roweth, Michael W. Malloy ...+10 more authors

Institutions: Northeastern University, Harvard University, Brigham and Women's Hospital, Nestlé ...+1 more institutions

Related papers:

- Network Medicine Framework Shows Proximity of Polyphenol Targets and Disease Proteins is Predictive of the Therapeutic Effects of Polyphenols
- The Impact of Polyphenols on Human Health
- Role of dietary polyphenols in the platelet aggregation network - a review of the in vitro studies
- Polyphenols and Gene Expression
- miRNA as molecular target of polyphenols underlying their biological effects

1 **Network Medicine Framework Shows Proximity of Polyphenol Targets and**

2 **Disease Proteins is Predictive of the Therapeutic Effects of Polyphenols**

3

4 Italo F. do Valle[1], Harvey G. Roweth[2,3], Michael W. Malloy[2,3], Sofia Moco[4], Denis

5 Barron[4], Elisabeth Battinelli[2,3], Joseph Loscalzo[3,5], Albert-László Barabási[1,6,7]

6

7 [1] Network Science Institute and Department of Physics, Northeastern University, Boston, MA, USA

8 [2] Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

9 [3] Harvard Medical School, Boston, MA, USA

10 [4] Nestle Institute of Health Sciences, Lausanne, Switzerland

11 [5] Department of Medicine Brigham and Women's Hospital

12 [6] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical

13 School, Boston, MA, USA

14 [7] Department of Network and Data Science, Central European University, Budapest, Hungary

15

16 **Abstract**

17

18 Polyphenols, natural products present in plant-based foods, play a protective role

19 against several complex diseases through their antioxidant activity and by diverse

20 molecular mechanisms. Here we developed a network medicine framework to uncover

21 the mechanistic roles of polyphenols on health by considering the molecular interactions

22 between polyphenol protein targets and proteins associated with diseases. We find that

23 the protein targets of polyphenols cluster in specific neighborhoods of the human

24 interactome, whose network proximity to disease proteins is predictive of the molecule's

25 known therapeutic effects. The methodology recovers known associations, such as the

26 effect of epigallocatechin 3-O-gallate on type 2 diabetes, and predicts that rosmarinic

27 acid (RA) has a direct impact on platelet function, representing a novel mechanism

28 through which it could affect cardiovascular health. We experimentally confirm that RA

29 inhibits platelet aggregation and alpha granule secretion through inhibition of protein

30 tyrosine phosphorylation, offering direct support for the predicted molecular mechanism.

31 Our framework represents a starting point for mechanistic interpretation of the health

32 effects underlying food-related compounds, allowing us to integrate into a predictive

33 framework knowledge on food metabolism, bioavailability, and drug interaction.

1

Introduction

Diet plays a defining role in human health. Indeed, while poor diet can significantly increase the risk for coronary heart disease (CHD) and type 2 diabetes mellitus (T2D), a healthy diet can play a protective role, even mitigating genetic risk for CHD[1]. Polyphenols are a class of compounds present in plant-based foods, from fruits to vegetables, nuts, seeds, beans (e.g. coffee, cocoa), herbs, spices, tea, and wine, with well documented protective role as antioxidants, which affect several diseases, from cancer to T2D, cardiovascular, and neurodegenerative diseases[2,3]. Previous efforts profiled over 500 polyphenols in more than 400 foods[4,5] and have documented the high diversity of polyphenols to which humans are exposed through their diet, ranging from flavonoids to phenolic acids, lignans, and stilbenes.

The underlying molecular mechanisms through which specific polyphenols exert their beneficial effects on human health remain largely unexplored. From a mechanistic perspective, dietary polyphenols are not engaged in endogenous metabolic processes of anabolism and catabolism, but rather affect human health through their anti- or pro-oxidant activity[6], by binding to proteins and modulating their activity[7,8], interacting with digestive enzymes[9], and modulating gut microbiota growth[10,11]. Yet, the variety of experimental settings and the limited scope of studies that explore the molecular effects of polyphenols have, to date, offered a range of often conflicting evidence. For example, two clinical trials, both limited in terms of the number of subjects and the intervention periods, resulted in conflicting conclusions about the beneficial effects of resveratrol on glycemic control in T2D patients[12,13]. We, therefore, need a framework to interpret the evidence present in the literature, and to offer in-depth mechanistic predictions on the molecular pathways responsible for the health implications of polyphenols present in diet. Ultimately, these insights could help us provide evidence on causal diet-health associations, guidelines of food consumption for different individuals, and help to develop novel diagnostic and therapeutic strategies, which may lead to the synthesis of novel drugs.

Here, we address this challenge by developing a network medicine framework to capture the molecular interactions between polyphenols and their cellular binding

65   targets, unveiling their relationship to complex diseases. The developed framework is

66   based on the human interactome, a comprehensive subcellular network consisting of all

67   known physical interactions between human proteins, which has been validated

68   previously as a platform for understanding disease mechanisms[14,15], rational drug target

69   identification, and drug repurposing[16,17].

70   We find that the proteins to which polyphenols bind form identifiable

71   neighborhoods in the human interactome, allowing us to demonstrate that the proximity

72   between polyphenol targets and proteins associated with specific diseases is predictive

73   of the known therapeutic effects of polyphenols. Finally, we unveil the potential

74   therapeutic effects of rosmarinic acid (RA) on vascular diseases (V), predicting that its

75   mechanism of action is related to modulation of platelet function. We confirm this

76   prediction by experiments that indicate that RA modulates platelet function *in vitro* by

77   inhibiting tyrosine protein phosphorylation. Altogether, our results demonstrate that the

78   network-based relationship between disease proteins and polyphenol targets offers a

79   tool to systematically unveil the health effects of polyphenols.

80

81   **Results**

82

83   Polyphenol Targets Cluster in Specific Functional Neighborhoods of the Interactome

84

85   We mapped the targets of 65 polyphenols (see Methods) to the human

86   interactome, consisting of 17,651 proteins and 351,393 interactions (Fig 1a,b). We find

87   that 19 of the 65 polyphenols have only one protein target, while a few polyphenols

88   have an exceptional number of targets (Fig 1c). We computed the Jaccard Index (JI) of

89   the protein targets of each polyphenol pair, finding only a limited similarity of targets

90   among different polyphenols (average JI = 0.0206) (Supplementary Figure 1a). Even

91   though the average JI is small, it is still significantly higher (Z = 147, Supplementary

92   Figure 1b) than the JI expected if the polyphenol targets were randomly assigned from

93   the pool of all network proteins with degrees matching the original set. This finding

94   suggests that while each polyphenol targets a specific set of proteins, their targets are

95   confined to a common pool of proteins, likely determined by commonalities in the

3

96   polyphenol binding domains of the three-dimensional structure of the protein targets[18].

97   Gene Ontology (GO) Enrichment Analysis recovers existing mechanisms[8] and also

98   helps identify new processes related to polyphenol protein targets, such as post-

99   translation protein modifications, regulation, and xenobiotic metabolism (Fig 1d). The

100 enriched GO categories indicate that polyphenols modulate common regulatory

101 processes, but the low similarity in their protein targets, illustrated by the low average JI,

102 indicates that they target different processes within the same process.

103   We next asked whether the polyphenol targets cluster in specific regions of the

104 human interactome. We focused on polyphenols with more than two targets (n=46, Fig

105 2), and measured the size and significance of the largest connected component (LCC)

106 formed by the targets of each polyphenol. We found that 25 of the 46 polyphenols have

107 a larger LCC than expected by chance (Z-score > 1.95) (Fig 1e, Fig 2). In agreement

108 with experimental evidence documenting the effect of polyphenols on multiple

109 pathways[19], we find that ten polyphenols have their targets organized in multiple

110 connected components of size > 2.

111   These results indicate that the targets of polyphenols modulate specific well

112 localized neighborhoods of the interactome (Fig 2, Supplementary Figure 1c). This

113 prompted us to explore if the interactome regions targeted by the polyphenols reside

114 within network neighborhoods associated with specific diseases, seeking a network-

115 based framework to unveil the molecular mechanism through which specific

116 polyphenols modulate health.

117

118 Proximity Between Polyphenol Targets and Disease Proteins Reveals their Therapeutic

119 Effects

120

121 Polyphenols can be viewed as drugs in that they bind to specific proteins, affecting their

122 ability to perform their normal functions. We, therefore, hypothesized that we can apply

123 the network-based framework used to predict the efficacy of drugs in specific

124 diseases[16,17] to also predict the therapeutic effects of polyphenols. The closer the

125 targets of a polyphenol are to disease proteins, the more likely that the polyphenol will

126 affect the disease phenotype. We, therefore, calculated the network proximity between

4

127 polyphenol targets and proteins associated with 299 diseases using the closest

128 measure, d_c, representing the average shortest path length between each polyphenol

129 target and the nearest disease protein (see Methods). Consider for example (-)-

130 epigallocatechin 3-O-gallate (EGCG), a polyphenol abundant in green tea.

131 Epidemiological studies have found a positive relationship between green tea

132 consumption and reduced risk of T2D[20,21], and physiological and biochemical studies

133 have shown that EGCG presents glucose-lowering effects in both *in vitro* and *in vivo*

134 models[22,23]. We identified 54 experimentally validated EGCG protein targets and

135 mapped them to the interactome, finding that the ECGC targets form an LCC of 17

136 proteins (Z = 7.61) (Fig 3a). We also computed the network-based distance between

137 EGCG targets and 83 proteins associated with T2D, finding that the two sets are

138 significantly proximal to each other. We ranked all 299 diseases based on the network

139 proximity to the ECGC targets in order to determine whether we can recover the 82

140 diseases in which ECGC has known therapeutic effects according to the CTD database.

141 By this analysis, we were able to recover 15 previously known therapeutic associations

142 among the top 20 ranked diseases (Table 1), confirming that network-proximity can

143 discriminate between known and unknown disease associations for polyphenols, as

144 previously confirmed among drugs[16,17].

145   We expanded these methods to all polyphenol-disease pairs, to predict diseases

146 for which specific polyphenols might have therapeutic effects. For this analysis, we

147 grouped all 19,435 polyphenol-disease associations between 65 polyphenols and 299

148 diseases into known (1,525) and unknown (17,910) associations. The known

149 polyphenol-disease set was retrieved from CTD, which is limited to manually curated

150 associations for which there is literature-based evidence. For each polyphenol, we

151 tested how well network proximity discriminates between the known and unknown sets

152 by evaluating the area under the Receiving Operating Characteristic (ROC) curve

153 (AUC). For EGCG, network proximity offers good discriminative power (AUC = 0.78, CI:

154 0.70 - 0.86) between diseases with known and unknown therapeutic associations (Table

155 1). We find that network proximity (d_c) offers predictive power with an AUC > 0.7 for 31

156 polyphenols (Fig 3b). The methodology recovers many associations well documented in

157 the literature, like the beneficial effects of umbelliferone on colonrectal neoplasms[24,25].

158   In Table 2 we summarize the top 10 polyphenols for which the network medicine

159   framework offers the best predictive power of therapeutic effects, limiting the entries to

160   predictive performance of AUC > 0.6 and performance over top predictions with

161   precision > 0.6. Given the lack of data on true negative examples, we considered

162   unknown associations as negative cases, observing the same trend when we used an

163   alternative performance metric that does not require true negative labels (i.e. AUC of

164   the Precision-Recall curve) (Supplementary Figure 2).

165   Finally, we performed multiple robustness checks to exclude the role of potential

166   biases in the input data. To test if the predictions are biased by the set of known

167   associations retrieved from CTD, we randomly selected 100 papers from PubMed

168   containing MeSH terms that tag EGCG to diseases. We manually curated the evidence

169   for EGCG's therapeutic effects for the diseases discussed in the published papers,

170   excluding reviews and non-English language publications. The dataset was processed

171   to include implicit associations (see Methods), resulting in a total of 113 diseases

172   associated with EGCG, of which 58 overlap with the associations reported by CTD (Fig

173   3c). We observed that the predictive power of network proximity was unaffected by

174   whether we considered the annotations from CTD, the manually curated list, or the

175   union of both (Fig 3d). To test the role of potential biases in the interactome, we

176   repeated our analysis using only high-quality polyphenol-protein interactions retrieved

177   from ligand-protein 3D resolved structures (Supplementary Figure 1d) and a subset of

178   the interactome derived from an unbiased high-throughput screening (Supplementary

179   Figure 1f). We find that the predictive power was largely unchanged, indicating that the

180   literature bias in the interactome does not affect our findings. Finally, we re-tested the

181   predictive performance by considering not only the therapeutic polyphenol-disease

182   associations, but also the marker/mechanism ones - another type of curated association

183   available in CTD - finding that the predictive power remains largely unchanged

184   (Supplementary Notes, Supplementary Figure 3).

185

186   Network Proximity Predicts Gene Expression Perturbation Induced by Polyphenols

187

6

188   To validate that network proximity reflects biological activity of polyphenols observed in

189   experimental data, we retrieved expression perturbation signatures from the

190   Connectivity Map database[26] for the treatment of the breast cancer MCF7 cell line with

191   21 polyphenols (Supplementary Table 1, Supplementary Figure 4). We investigated the

192   relationship between the extent to which polyphenols perturb the expression of disease

193   genes, the network proximity between the polyphenol targets and disease proteins, and

194   their known therapeutic effects (Fig 4a). For example, we observe different perturbation

195   profiles for gene pools associated with different diseases: for treatment with genistein (1

196   μM, 6 hours) we observe 10 skin disease genes with perturbation score > 2, while we

197   observe only one highly perturbed cerebrovascular disorder gene (Fig 4b). Indeed,

198   network proximity indicates that skin disease is closer to the genistein targets than

199   cerebrovascular disorder, suggesting a relationship between network proximity, gene

200   expression perturbation, and the therapeutic effects of the polyphenol (Fig 4a). To test

201   this hypothesis, we computed an enrichment score that measures the

202   overrepresentation of disease genes among the most perturbed genes (see Methods),

203   finding 13 diseases that have their genes significantly enriched among the most

204   deregulated genes by genistein, of which 4 have known therapeutic associations. We

205   find that these four diseases are significantly closer to the genistein targets than the

206   nine diseases with unknown therapeutic associations (Fig 4c). We observed a similar

207   trend for treatments with other polyphenols, whether we use the same (1μM, Fig 4c) or

208   different (100nM to 10μM, Supplementary Figure 5) concentrations. This result suggests

209   that changes in gene expression caused by a polyphenol are indicative of its therapeutic

210   effects, but only if the observed expression change is limited to proteins proximal to the

211   polyphenol targets (Fig 4a).

212   Consequently, network proximity should also be predictive of the overall gene

213   expression perturbation caused by a polyphenol on the genes of a given disease. To

214   test this hypothesis, in each experimental combination defined by the polyphenol type

215   and its concentration, we evaluated the maximum perturbation among genes for each

216   disease. We then compared the magnitude of the observed perturbation between

217   diseases that were proximal (d_c < 25th percentile, Z_{d_c} < -0.5) or distal (d_c > 75th

218   percentile, Z_{d_c} > -0.5) to the polyphenol targets. Figures 5a-b and Supplementary Figure

219   6 show the results for the genistein treatment (1µM, 6 hours), indicating that diseases

220   proximal to the polyphenol targets show higher maximum perturbation values than distal

221   diseases. The same trend is observed for other polyphenols when we use different d_c

222   and Z_{d_c} thresholds for defining proximal and distant diseases (Figs 5b, Supplementary

223   Figures 6-9), confirming that the impact of a polyphenol on cellular signaling pathways

224   is localized in the network space, being greater in the vicinity of the polyphenol targets

225   compared to neighborhoods remote from these targets. We also considered gene

226   expression perturbations in the network vicinity of the polyphenol targets, regardless of

227   whether the proteins were disease proteins or not, observing higher perturbation scores

228   for proximal proteins in 12 out 21 polyphenols tested at 10µM (Supplementary Figure

229   10). Finally, we find that the enrichment score of perturbed genes among disease genes

230   is not as predictive of the polyphenol therapeutic effects as network proximity

231   (Supplementary Figure 11).

232   Altogether these results indicate that network proximity offers a mechanistic

233   interpretation for the gene expression perturbations induced by polyphenols on disease

234   genes. They also show that network proximity can indicate when gene expression

235   perturbations result in therapeutic effects, suggesting that future studies could integrate

236   gene expression (whenever available) with network proximity as they aim to more

237   accurately prioritize polyphenol-disease associations.

238

239   Experimental Evidence Confirms that Rosmarinic Acid Modulates Platelet Function

240

241   To demonstrate how the network-based framework can facilitate the mechanistic

242   interpretation of the therapeutic effects of selected polyphenols, we next focus on

243   vascular diseases (V). Of 65 polyphenols evaluated in this study, we found 27 to have

244   associations to V, as their targets were within the V network neighborhood

245   (Supplementary Table 3). We, therefore, inspected the targets of 15 of the 27

246   polyphenols with 10 or less targets. The network analysis identified direct links between

247   biological processes related to vascular health and the targets of three polyphenols:

248   gallic acid, rosmarinic acid, and 1,4-naphthoquinone (Supplementary Figure 12,

249   Supplementary Notes). The network neighborhood containing the targets of these

250 polyphenols suggests that gallic acid activity involves thrombus dissolution processes,

251 rosmarinic acid acts on platelet activation and antioxidant pathways through FYN and its

252 neighbors, and 1,4-naphthoquinone acts on signaling pathways of vascular cells

253 through MAP2K1 activity (Supplementary Figure 12, Supplementary Notes).

254   To validate the developed framework, we set out to obtain direct experimental

255 evidence of the predicted mechanistic role of rosmarinic acid (RA) in V. The RA targets

256 are in close proximity to proteins related to platelet function, forming the RA-V-platelet

257 module: a connected component formed by the RA target FYN and the V proteins

258 associated with platelet function PDE4D, CD36, and APP (Fig 6a). We, therefore, asked

259 whether RA influenced platelet activation *in vitro*. As platelets can be stimulated through

260 different activation pathways, RA effects can, in principle, occur in any of them. To test

261 these different possibilities, we pretreated platelets with RA and then activated with: 1)

262 glycoprotein VI by collagen or collagen-related peptide (CRP/CRPXL); 2) protease-

263 activated receptors-1,4 by thrombin receptor activator peptide-6 (TRAP-6); 3)

264 prostanoid thromboxane receptor by the thromboxane A_2 analogue (U46619); and 4)

265 P2Y1/12 receptor by adenosine diphosphate (ADP)[27]. When we compared the network

266 distance between each stimulant receptor and the RA-V-platelet module (Fig 6a), we

267 observed that the receptors for CRP/CRPXL, TRAP-6, and U46619 are closer than

268 random expectation, while the receptor for ADP is more distant (Fig 6b). We expected

269 that platelets would be most affected by RA when treated with stimulants whose

270 receptors are most proximal to the RA-V-platelet module, i.e., CRP/CRPXL, TRAP-6,

271 and U46619, and as a control, we expect no effect for the distant ADP receptor. The

272 experiments confirm this prediction: RA inhibits collagen-mediated platelet aggregation

273 (Fig 6c) and impairs dense granule secretion induced by CRPXL, TRAP-6, and U46619

274 (Supplementary Figure 13). RA-treated platelets also displayed dampened alpha-

275 granule secretion (Fig 6d) and integrin $\alpha IIb\beta 3$ activation (Supplementary Figure 13) in

276 response to U46619. As expected, RA did not affect platelet function when we used an

277 agonist whose receptor is distant from the RA-V-module, i.e., ADP. These findings

278 suggest that RA impairs basic hallmarks of platelet activation via strong network effects,

279 supporting our hypothesis that the proximity between RA targets and the neighborhood

280 associated with platelet function (Fig 6a) could in part explain RA's impact on V.

281   We next searched to clarify the molecular mechanisms involved in the impact of

282   RA on platelets. Given that platelet activation is coordinated by several kinases, we

283   hypothesized that RA inhibits platelet function by blocking agonist-induced protein

284   tyrosine phosphorylation. We observed that RA-treated platelets demonstrated a dose-

285   dependent reduction in total tyrosine phosphorylation in response to CRPXL, TRAP-6

286   and U46619 (Fig 6e). Given that RA caused a substantial decrease in phosphorylation

287   of proteins with atomic mass between 50-60 KDa (Fig 6e), we hypothesized that RA

288   may reduce phosphorylation of FYN (59 KDa), or other similarly sized members of the

289   same protein family (*i.e.* src family kinases, SFKs). To test this, we measured the level

290   of phosphorylation within the activation domain (amino acid 416) of SFKs, finding that

291   RA reduced collagen induced phosphorylation of FYN as well as basal tyrosine

292   phosphorylation of SFKs (Fig 6f). This indicates that RA perturbs the phospho-signaling

293   networks that regulate platelet response to extracellular stimuli.

294   Altogether, these findings support our prediction that RA modulates platelet

295   activation and function. It also supports the observation that its mechanism of action

296   involves reduction of phosphorylation at the activation domain of the protein-tyrosine

297   kinase FYN (Fig 6a) and the inhibition of general tyrosine phosphorylation. Finally, while

298   polyphenols are usually associated to their antioxidant function, here we illustrate

299   another mechanistic pathway through which they could benefit health.

300

301   **Discussion**

302

303   Here, we proposed a network-based framework to predict the therapeutic effects of

304   dietary polyphenols in human diseases. We find that polyphenol protein targets cluster

305   in specific functional neighborhoods of the interactome, and we show that the network

306   proximity between polyphenol targets and disease proteins is predictive of the

307   therapeutic effects of polyphenols. We demonstrate that diseases whose proteins are

308   proximal to polyphenol targets tend to have significant changes in gene expression in

309   cell lines treated with the respective polyphenol, while such changes are absent for

310   diseases whose proteins are distal to polyphenol targets. Finally, we find that the

311   network neighborhood around the RA targets and vascular disease proteins are related

312   to platelet function. We validate this mechanistic prediction by showing that RA

313   modulates platelet function through inhibition of protein tyrosine phosphorylation. These

314   observations suggest a role of RA on prevention of vascular diseases by inhibiting

315   platelet activation and aggregation.

316   The observed results also suggest multiple avenues through which our ability to

317   understand the role of polyphenols could be improved. First, some of the known health

318   benefits of polyphenols might be caused not only by the native molecules, but also by

319   their metabolic byproducts [28,29]. We, however, lack data about colonic degradation, liver

320   metabolism, bioavailability, and interaction with proteins of specific polyphenols or their

321   metabolic byproducts. Future experimental data on protein interactions with polyphenol

322   byproducts and conjugates can be incorporated in the proposed framework, further

323   improving the accuracy of our predictions. The lack of this data does not invalidate the

324   findings presented here, since previous studies report the presence of unmetabolized

325   polyphenols in blood[30–32]; and it has been hypothesized that, in some instances,

326   deconjugation of liver metabolites occurs in specific tissues or cells[33–35]. Therefore, the

327   lack of data for specific polyphenols and the fact that other mechanisms exist through

328   which they can affect health (e.g. antioxidant activity, microbiota regulation) explain why

329   this methodology might still miss a few known relationships between polyphenols and

330   diseases. Second, considering that several experimental studies of polyphenol

331   bioefficacy have been observed in *in vitro* and *in vivo* models, the proposed framework

332   might help us interpret literature evidence, possibly even allowing us to exclude

333   chemical candidates when considering the health benefits provided by a given food in

334   epidemiological association studies.

335   Our assumption that network proximity recovers therapeutic associations is

336   based on its predictive performance on a ground truth dataset for observed therapeutic

337   effects and also relies on previous observations about the effect of drugs on

338   diseases[16,17,36]. While the proposed methodology offers a powerful prioritization tool to

339   guide future research, the real effect of polyphenols on diseases might still be negative,

340   given other unmet factors such as dosage, comorbidities, and drug interactions, which

341   can only be ruled out by pre-clinical and clinical studies. Gene expression perturbation

342   profiles, such as the ones provided by the Connectivity map, can also be integrated with

343   network proximity to further highlight potential beneficial or harmful effects of chemical

344   compounds[37,38].

345   The low bioavailability of some polyphenols in food might still present challenges

346   when considering the therapeutic utility of these molecules. However, 48 of the 65

347   polyphenols we explored here are predicted to have high gastrointestinal absorption

348   (Supplementary Table 2) and different methodologies are available to increase

349   bioavailability of natural compounds[39,40]. Additionally, in the same way that the

350   polyphenol phlorizin led to the discovery of new strategies for disease treatment

351   resulting in the development of new compounds with higher efficacy[41], we believe that

352   the present methodology can help us identify polyphenol-based candidates for drug

353   development.

354   The methodology introduced here offers a foundation for the mechanistic

355   interpretation of alternative pathways through which polyphenols can affect health, e.g.,

356   the combined effect of different polyphenols[36,42] and their interaction with drugs[43]. To

357   address such synergistic effects, we need ground-truth data on these aspects. The

358   developed methodology can be applied to other food-related chemicals, providing a

359   framework by which to understand their health effects. Future research may help us

360   also account for the way that food-related chemicals affect endogenous metabolic

361   reactions, impacting not only signaling pathways, but also catabolic and anabolic

362   processes. Finally, the methodology provides a framework to interpret and find causal

363   support for associations identified in observational studies. Taken together, the

364   proposed network-based framework has the potential to reveal systematically the

365   mechanism of action underlying the health benefits of polyphenols, offering a logical,

366   rational strategy for mechanism-based drug development of food-based compounds.

367

368   **Methods**

369

370   Building the Interactome

371

372   The human interactome was assembled from 16 databases containing different types of

373   protein-protein interactions (PPIs): 1) binary PPIs tested by high-throughput yeast-two-

12

374   hybrid (Y2H) experiments[44]; 2) kinase-substrate interactions from literature-derived low-

375   throughput and high-throughput experiments from KinomeNetworkX[45], Human Protein

376   Resource Database (HPRD)[46], and PhosphositePlus[47]; 3) carefully literature-curated

377   PPIs identified by affinity purification followed by mass spectrometry (AP-MS), and from

378   literature-derived low-throughput experiments from InWeb[48], BioGRID[49], PINA[50],

379   HPRD[51], MINT[52], IntAct[52], and InnateDB[53]; 4) high-quality PPIs from three-dimensional

380   (3D) protein structures reported in Instruct[54], Interactome3D[55], and INSIDER[56]; 5)

381   signaling networks from literature-derived low-throughput experiments as annotated in

382   SignaLink2.0[57]; and 6) protein complex from BioPlex2.0[58]. The genes were mapped to

383   their Entrez ID based on the National Center for Biotechnology Information (NCBI)

384   database as well as their official gene symbols. The resulting interactome includes

385   351,444 protein-protein interactions (PPIs) connecting 17,706 unique proteins

386   (Supplementary Data 1). The largest connected component has 351,393 PPIs and

387   17,651 proteins.

388

389   <u>Polyphenols, Polyphenol Targets, and Disease Proteins</u>

390

391   We retrieved 759 polyphenols from the PhenolExplorer database[4]. The database lists

392   polyphenols with food composition data or profiled in biofluids after interventions with

393   polyphenol-rich diets. For our analysis, we only considered polyphenols that: 1) could

394   be mapped in PubChem IDs, 2) were listed in the Comparative Toxicogenomics (CTD)

395   database[59] as having therapeutic effects on human diseases, and 3) had protein-

396   binding information present in the STITCH database[60] with experimental evidence (Fig

397   1a). After these steps, we considered a final list of 65 polyphenols, for which 598 protein

398   targets were retrieved from STITCH (Supplementary Table 1). We considered 3,173

399   disease proteins corresponding to 299 diseases retrieved from Menche *et al* (2015)[15].

400   Gene ontology enrichment analysis of protein targets was performed using the

401   Bioconductor package clusterProfiler with a significance threshold of $p < 0.05$ and

402   Benjamini-Hochberg multiple testing correction with $q < 0.05$.

403

404   <u>Polyphenol Disease Associations</u>

405

406   We retrieved the polyphenol-disease associations from the Comparative

407   Toxicogenomics Database (CTD). We considered only manually curated associations

408   labeled as therapeutic. By considering the hierarchical structure of diseases along the

409   MeSH tree, we expanded explicit polyphenol-disease associations to include also

410   implicit associations. This procedure was performed by propagating associations in the

411   lower branches of the MeSH tree to consider diseases in the higher levels of the same

412   tree branch. For example, a polyphenol associated with heart diseases would also be

413   associated with the more general category of cardiovascular diseases. By performing

414   this expansion, we obtained a final list of 1,525 known associations between the 65

415   polyphenols and the 299 diseases considered in this study.

416

417   <u>Network Proximity Between Polyphenol Targets and Disease Proteins</u>

418

419   The proximity between a disease and a polyphenol was evaluated using a distance

420   metric that takes into account the shortest path lengths between polyphenol targets and

421   disease proteins[16]. Given S, the set of disease proteins, T, the set of polyphenol targets,

422   and $d(s,t)$, the shortest path length between nodes s and t in the network, we define:

423

424 $$d_c(S,T) = \frac{1}{||T||} \sum_{t \in T} \min_{s \in S} d(s,t) \qquad (1)$$

425   We also calculated a relative distance metric (Z_{dc}) that compares the absolute distance

426     $d_c(S,T)$ between a disease and a polyphenol with a reference distribution describing the

427   random expectation. The reference distribution corresponds to the expected distances

428   between two randomly selected groups of proteins matching the size and degrees of the

429   original disease proteins and polyphenol targets in the network. It was generated by

430   calculating the proximity between these two randomly selected groups across 1,000

431   iterations. The mean $\mu_{d(S,T)}$ and standard deviation $\sigma_{d(S,T)}$ of the reference distribution

432   were used to convert the absolute distance d_c into the relative distance Z_{dc}, defined as:

433

434 $$Z_{dc} = \frac{d - \mu_{d_c(S,T)}}{\sigma_{d_c(S,T)}} \qquad (2)$$

14

435

436   We performed a degree-preserving random selection, but due to the scale-free nature

437   of the human interactome, we avoid repeatedly choosing the same (high degree) nodes

438   by using a binning approach in which nodes within a certain degree interval were

439   grouped together such that there were at least 100 nodes in the bin. The

440   Supplementary Data 2 reports the proximity scores d_c and Z_{d_c} for all pairs of diseases

441   and polyphenols.

442

443   <u>Area Under ROC Curve Analysis</u>

444

445   For each polyphenol, we used AUC to evaluate how well the network proximity

446   distinguishes diseases with known therapeutic associations from all of the others of the

447   set of 299 diseases. The set of known associations (therapeutic) retrieved from CTD

448   were used as positive instances, all unknown associations were defined as negative

449   instances, and the area under the ROC curve was computed using the implementation

450   in the Scikit-learn Python package. Furthermore, we calculated 95% confidence

451   intervals using the bootstrap technique with 2,000 resamplings with sample sizes of 150

452   each. Considering that AUC provides an overall performance, we also searched for a

453   metric to evaluate the top ranking predictions. For this analysis, we calculated the

454   precision of the top 10 predictions, considering only the polyphenol-disease

455   associations with relative distance $Z_{dc} < -0.5$[16].

456

457   <u>Analysis of Network Proximity and Gene Expression Deregulation</u>

458

459   We retrieved perturbation signatures from the Connectivity Map database

460   (https://clue.io/) for the MCF7 cell line after treatment with 21 polyphenols. These

461   signatures reflect the perturbation of the gene expression profile caused by the

462   treatment with that particular polyphenol relative to a reference population, which

463   comprises all other treatments in the same experimental plate[26]. For polyphenols having

464   more than one experimental instance (time of exposure, cell line, dose), we selected the

465   one with highest distil_cc_q75 value (75th quantile of pairwise spearman correlations in

466 landmark genes, https://clue.io/connectopedia/perturbagen_types_and_controls). We

467 performed Gene Set Enrichment Analysis[61] to evaluate the enrichment of disease

468 genes among the top deregulated genes in the perturbation profiles. This analysis offers

469 Enrichment Scores (ES) that have small values when genes are randomly distributed

470 among the ordered list of expression values and high values when they are

471 concentrated at the top or bottom of the list. The ES significance is calculated by

472 creating 1,000 random selection of gene sets with the same size as the original set and

473 calculating an empirical p-value by considering the proportion of random sets resulting

474 in ES smaller than the original case. The p-values were adjusted for multiple testing

475 using the Benjamini-Hochberg method. The network proximity d_c of disease proteins

476 and polyphenol targets for diseases with significant ES were compared according to

477 their therapeutic and unknown-therapeutic associations using the Student's t-test. The

478 relevant code for calculating the network proximity, AUCs, and enrichment scores can

479 be found on https://github.com/italodovalle/polyphenols.

480

481 Platelet Isolation

482

483 Human blood collection was performed as previously described in accordance with the

484 Declaration of Helsinki and ethics regulations with Institutional Review Board approval

485 from Brigham and Women's Hospital (P001526). Healthy volunteers did not ingest

486 known platelet inhibitors for at least 10 days prior. Citrated whole blood underwent

487 centrifugation with a slow brake (177 x g, 20 minutes), and the PRP fraction was

488 acquired for subsequent experiments. For washed platelets, PRP was incubated with 1

489 μM prostaglandin E_1 (Sigma, P5515) and immediately underwent centrifugation with a

490 slow brake (1000 x g, 5 minutes). Platelet-poor plasma was aspirated, and pellets

491 resuspended in platelet resuspension buffer (PRB; 10 mM Hepes, 140 mM NaCl, 3 mM

492 KCl, 0.5 mM $MgCl_2$, 5 mM $NaHCO_3$, 10 mM glucose, pH 7.4).

493

494 Platelet Aggregometry

495

16

496 Platelet aggregation was measured by turbidimetric aggregometry as previously

497 described[62]. Briefly, PRP was pretreated with RA for 1 hour before adding 250 µL to

498 siliconized glass cuvettes containing magnetic stir bars. Samples were placed in

499 Chrono-Log® Model 700 Aggregometers before the addition of various platelet agonists.

500 Platelet aggregation was monitored for 6 minutes at 37°C with a stir speed of 1000 rpm

501 and the maximum extend of aggregation recorded using AGGRO/LINK®8 software. In

502 some cases, dense granule release was simultaneously recorded by supplementing

503 samples with Chrono-Lume® (Chrono-Log®, 395) according to the manufacturer's

504 instructions.

505

506 <u>Platelet Alpha Granule Secretion and Integrin $\alpha_{IIb}\beta_3$ Activation</u>

507

508 Changes in platelet surface expression of P-selectin (CD62P) or binding of Alexa

509 Fluor™ 488-conjugated fibrinogen were used to assess alpha granule secretion and

510 integrin $\alpha_{IIb}\beta_3$ activation, respectively. First, PRP was pre-incubated with RA for 1 hour,

511 followed by stimulation with various platelet agonists under static conditions at 37°C for

512 20 minutes. Samples were then incubated with APC-conjugated anti-human CD62P

513 antibodies (BioLegend®, 304910) and 100 µg/mL Alexa Fluor™ 488-Fibrinogen (Thermo

514 Scientific™, F13191) for 20 minutes before fixation in 2% [v/v] paraformaldehyde

515 (Thermo Scientific™, AAJ19945K2). Fifty thousand platelets were processed per

516 sample using a Cytek™ Aurora spectral flow cytometer. Percent-positive cells were

517 determined by gating on fluorescence intensity compared to unstimulated samples.

518

519 <u>Platelet Cytotoxicity</u>

520

521 Cytotoxicity were tested by measuring lactate dehydrogenase (LDH) release by

522 permeabilized platelets into the supernatant[63]. Briefly, washed platelets were treated

523 with various concentrations of RA for 1 hour, before isolating supernatants via

524 centrifugation (15,000 x g, 5 min). A Pierce LDH Activity Kit (Thermo Scientific™,

525 88953) was then used to assess supernatant levels of LDH.

526

17

527   <u>Immunoprecipitation and Western blot</u>

528

529   Washed platelets were pre-treated with RA for 1 hour, followed by a 15 minute

530   treatment with Eptifibatide (50 μM). Platelets were then stimulated with various agonists

531   for 5 minutes under stirring conditions (1000 rpm, 37°C). Platelets were lysed on ice

532   with RIPA Lysis Buffer System® (Santa Cruz®, sc-24948) and supernatants clarified via

533   centrifugation (15,000 x g, 10 min, 4°C). For immunoprecipitation of FYN, lysates were

534   first precleared of IgG by incubating with Protein A agarose beads (Cell Signaling

535   Technologies, 9863S) for 30 minutes at 4°C, before isolation of the supernatant via

536   centrifugation (15,000 x g, 10 min, 4°C). Supernatants were incubated with anti-FYN

537   antibodies (Abcam, 2A10) overnight at 4°C before incubation with Protein A beads for 1

538   hour. Beads were then washed 5 times with NP-40 lysis buffer (144 mM Tris, 518 mM

539   NaCl, 6 mM EDTA, 12 mM Na_2VO_3, 33.3% [v/v] NP-40, Halt^{TM} protease inhibitor

540   cocktail (Thermo, 78429)).

541   For Western Blot analysis, total cell lysates or immunoprecipitated FYN were

542   reduced with Laemmli Sample Buffer (Bio-Rad, 1610737) and proteins separated by

543   molecular weight in PROTEAN TGX^{TM} precast gels (Bio-Rad, 4561084). Proteins were

544   transferred to PVDF membranes (Bio-Rad, 1620174) and probed with either 4G10

545   (Milipore, 05-321), a primary antibody clone that recognizes phosphorylated tyrosine

546   residues, or primary antibodies that probe for the site-specific phosphorylation of src

547   family kinases (SFKs, p-Tyr416) within their activation loop. Membranes were incubated

548   with horseradish peroxidase-conjugated secondary antibodies (Cell Signaling

549   Technologies, 7074S) to catalyze an electrochemiluminescent reaction (Thermo

550   Scientific^{TM}, PI32109). Membranes were visualized using a Bio-Rad ChemiDoc Imaging

551   System and densitometric analysis of protein lanes conducted using ImageJ (NIH,

552   Version 1.52a).

553

554   **Author Contributions**

555

556   I.F.V and A.L.B designed the study. I.F.V. performed all computational analyses. H.G.R,

557   M.W.M., E.B., and J.L designed and performed experimental validation. J.L. guided

18

558   I.F.V. on validation case studies. S.M and D.B guided I.F.V for data interpretation and

559   curation of disease associations obtained from literature. I.F.V and A.L.B wrote the

560   paper with input from all authors. All authors read and approved the manuscript.

561

562   **Acknowledgements**

563

569

570   **Declaration of Interests**

571

572   J.L. and A.L.B are co-scientific founder of Scipher Medicine, Inc., which applies network

573   medicine strategies to biomarker development and personalized drug selection. A.L.B is

574   the founder of Nomix Inc. and Foodome, Inc. that apply data science to health; I.F.V is a

575   scientific consultant for Foodome, Inc.

576

577   **<u>Data Availability</u>**

578

579   The authors declare that all data supporting the findings of this study are available at

580   https://github.com/italodovalle/polyphenols and within the paper and its supplementary

581   information files.

582

583   **Code Availability**

584

585   Computer code is available at https://github.com/italodovalle/polyphenols

586

587

588 **Table 1 – Top 20 Predicted Therapeutic Associations Between EGCG and Human**
589 **Diseases.** Diseases were ordered according to the network distance (d_c) of their
590 proteins to EGCG targets and diseases with relative distance $Z_{dc} > -0.5$ were removed.
591 References reported in CTD for curated 'therapeutic associations' are shown.
592

| Disease	Distance d_c	Significance Z_{d_c}	Known Therapeutic Effect (References)
nervous system diseases	1.13	-1.72	[64,65]
nutritional and metabolic diseases	1.25	-1.45	[23]
metabolic diseases	1.25	-1.41	[23]
cardiovascular diseases	1.27	-2.67	[66–71]
immune system diseases	1.29	-1.31	[72]
vascular diseases	1.33	-3.47	[66,67,70]
digestive system diseases	1.33	-1.57	[73–77]
neurodegenerative diseases	1.37	-1.71	[78]
central nervous system diseases	1.41	-0.54	[78]
autoimmune diseases	1.41	-1.30	[72]
gastrointestinal diseases	1.43	-1.02	[79]
brain diseases	1.43	-0.89	NA
intestinal diseases	1.49	-1.08	[79]
inflammatory bowel diseases	1.54	-2.10	NA
bone diseases	1.54	-1.18	NA
gastroenteritis	1.54	-1.92	NA
demyelinating diseases	1.54	-1.78	NA
glucose metabolism disorders	1.54	-1.58	[23]
heart diseases	1.56	-1.20	[68,69,71]
diabetes mellitus	1.56	-1.66	[23]

593

594 **Table 2 – Top Ranked Polyphenols.** Polyphenols for which network proximity to
595 diseases best predicts their therapeutic effects. Table showing polyphenols with AUC >
596 0.6 and Precision > 0.6. (*) Confidence intervals calculated with 2,000 bootstraps with
597 replacement and sample size of 50% of the diseases (150/299). (**) Precision was
598 calculated based on the top 10 polyphenols after their ranking based on the distance
599 (d_c) of their targets to the disease proteins and considering only predictions with Z-score
600 < -0.5.(***) Concentrations of polyphenols in blood were retrieved from the Human
601 Metabolome Database (HMDB)

| Polyphenol	AUC	AUC CI*	Precision**	Concentration in Blood***	N Mapped Targets	LCC Size
Coumarin	0.93	[0.86 - 0.98]	0.6		7	1
Piceatannol	0.86	[0.77 - 0.94]	0.6		39	23
Genistein	0.82	[0.75 - 0.89]	0.7	[0.006 - 0.525 uM]	18	6
Ellagic acid	0.79	[0.63 - 0.92]	0.6		42	19
(-)-epigallocatechin 3-O-gallate	0.78	[0.70 - 0.86]	0.8		51	17
Isoliquiritigenin	0.75	[0.77 - 0.94]	0.6		10	8
Resveratrol	0.75	[0.66 - 0.82]	1		63	25
Pterostilbene	0.73	[0.61 - 0.84]	0.6		5	2
Quercetin	0.73	[0.64 - 0.81]	1	[0.022 - 0.080 uM]	216	140
(-)-epicatechin	0.65	[0.49 - 0.80]	0.8	0.625 uM	11	3

602

603
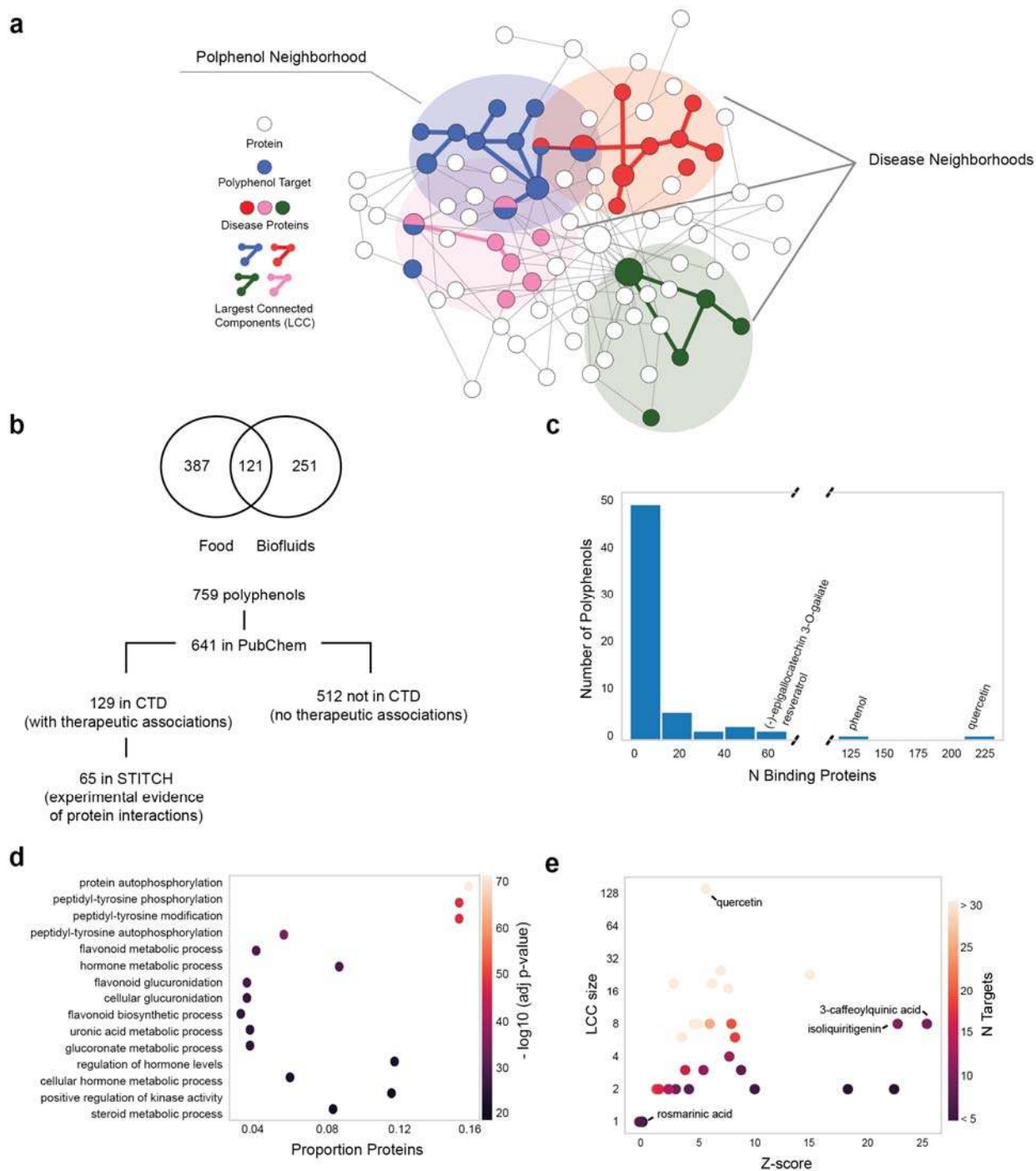
604

605 **Figure 1 – Properties of Polyphenol Protein Targets.** (A) Schematic representation of the human
606 interactome, highlighting regions where polyphenol targets and disease proteins are localized. (B)
607 Diagram showing the selection criteria of the polyphenols evaluated in this study. (C) Distribution of the
608 number of polyphenol protein targets mapped to the human interactome. (D) Top (n=15) enriched GO
609 terms (Biological Process) among all polyphenol protein targets. The X-axis shows the proportion of
610 targets mapped to each pathway. (E) Size of the Largest Connected Component (LCC) formed by the
611 targets of each polyphenol in the interactome and the corresponding significance (z-score).
612

22

23

614 **Figure 2 – Protein-Protein Interactions of Polyphenol Targets.** The 23 polyphenols whose targets
615 form connected components in the interactome and their respective subgraphs. For example, piceatannol
616 targets form a unique connected component of 23 proteins, while quercetin targets form multiple
617 connected components, the largest with 140 proteins. Polyphenol targets that are not connected to any
618 other target are not shown in the figure. Colors distinguish connected component of different polyphenols.
619
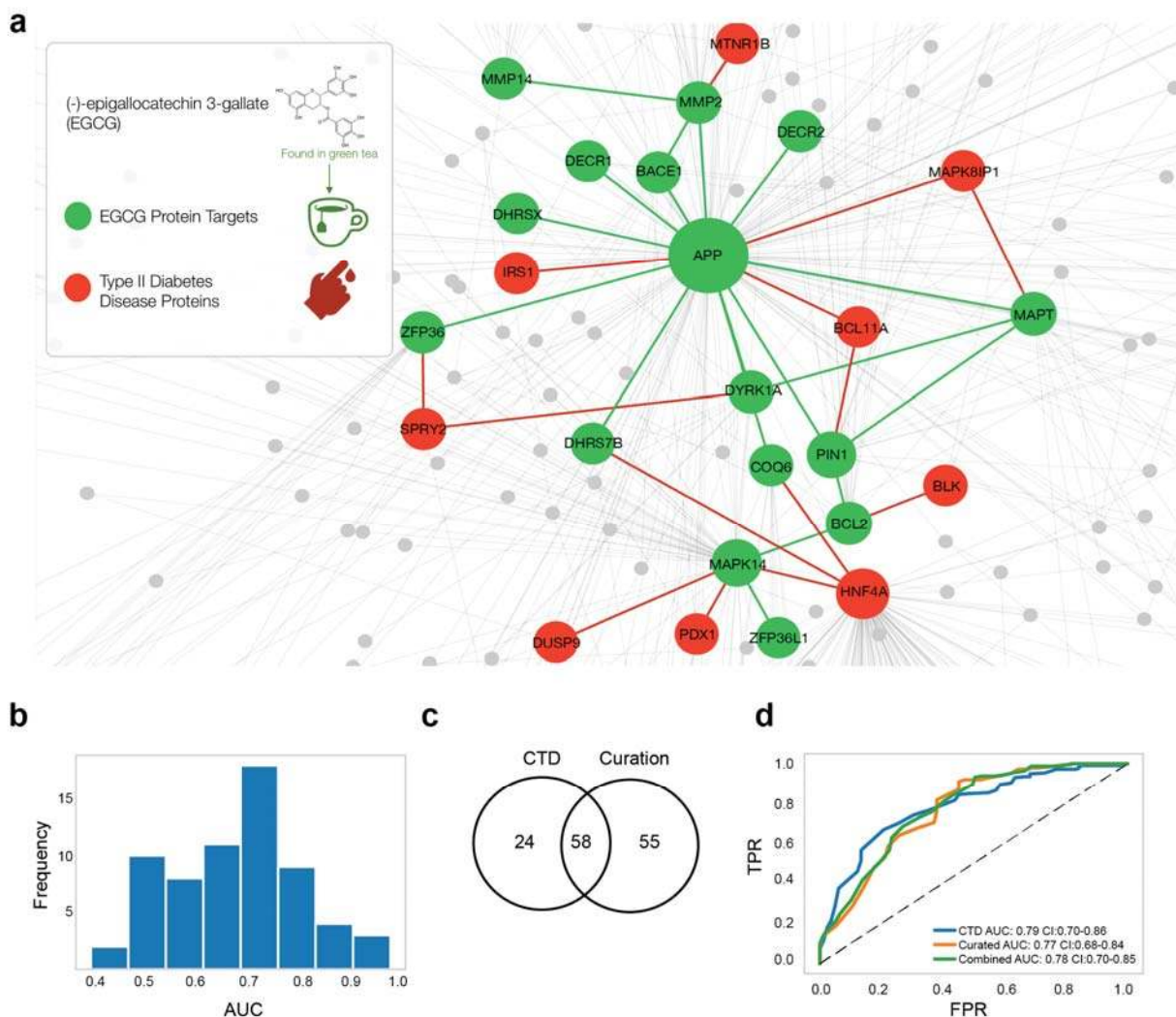
620

621 **Figure 3 – Proximity Between Polyphenol Targets and Disease Proteins is Predictive of**
622 **Therapeutic Effects of the Polyphenol.** (a) Interactome neighborhood showing the EGCG protein
623 targets and their interactions with type 2 diabetes (T2D)-associated proteins. (b) Distribution of AUC
624 values considering the predictions of therapeutic effects for 65 polyphenols. (c) Comparison of the ECGC-
625 disease associations considering the CTD database and the in-house database derived from the manual
626 curation of the literature. (d) Comparison of the prediction performance when considering known EGCG-
627 disease associations from the CTD, in-house manually curated database, or combined datasets.
628

24
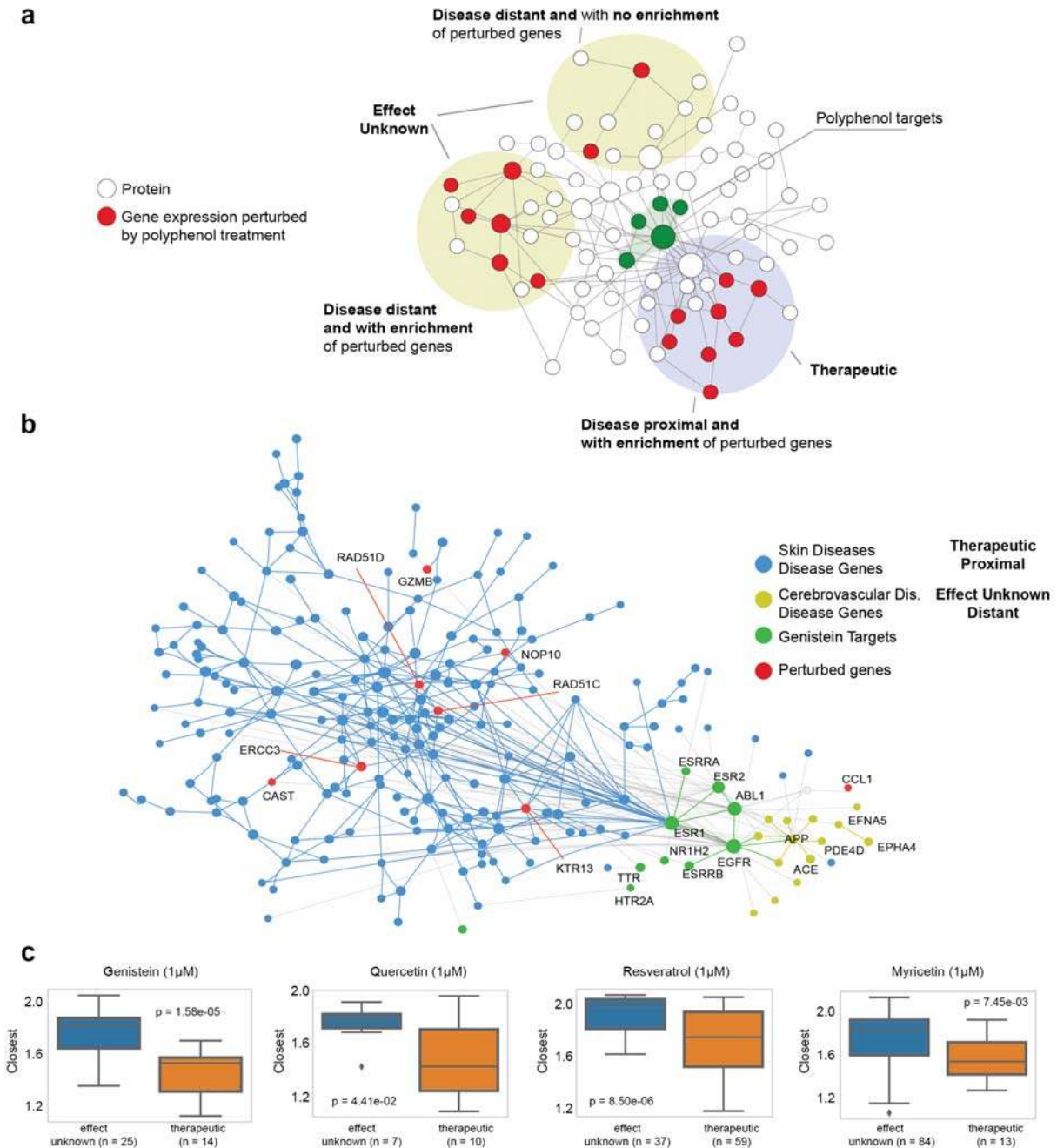
629
Figure 4 – Relationships among Gene Expression Perturbation, Network Proximity, and Therapeutic Effects of Polyphenols on Diseases. (a) Schematic representation of the relationship between the extent to which a polyphenol perturbs disease genes expression, its proximity to the disease genes, and its therapeutic effects. (b) Interactome neighborhood showing the modules of skin diseases (SD), genistein, and cerebrovascular disorders (CD). The SD module has 10 proteins with high perturbation scores (>2) in the treatment of the MCF7 cell line with 1 µM of genistein. Genes associated to SD are significantly enriched among the most differentially expressed genes, and the maximum perturbation score among disease genes is higher in SD than CD. (c) Among the diseases in which genes are enriched with highly perturbed genes, those with therapeutic associations show smaller network distances to the polyphenol targets than those without. The same trend is observed in treatments of the polyphenols quercetin, resveratrol, and myricetin.
641

642
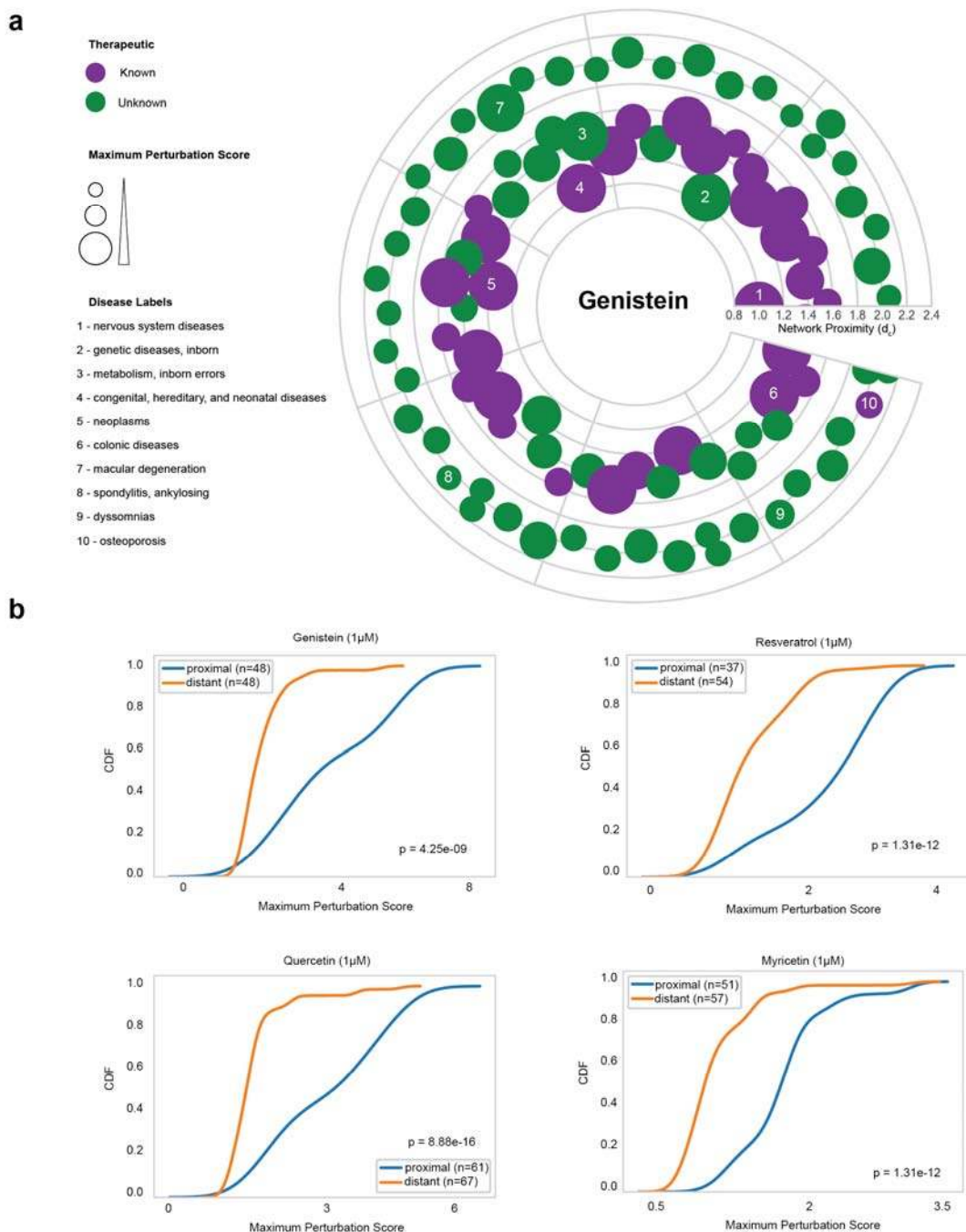
Figure 5 – Diseases Proximal to Polyphenol Targets Have Higher Gene Expression Perturbation Profiles. (a) Proximal and distal diseases in relation to genistein targets. Each node represents a disease and the node size is proportional to the perturbation score after treatment with genistein (1 µM, 6 hours). Distance from the origin represents the network proximity (d_c) to genistein targets. Purple nodes represent diseases in which the therapeutic association was previously known. (c) Cumulative distribution of the maximum perturbation scores of genes from diseases that are distal or proximal to polyphenol targets considering different polyphenols (1 µM, 6 hours): genistein, quercetin, resveratrol, and myricetin. Statistical significance was evaluated with the Kolmogorov-Smirnov test.

651

26

652
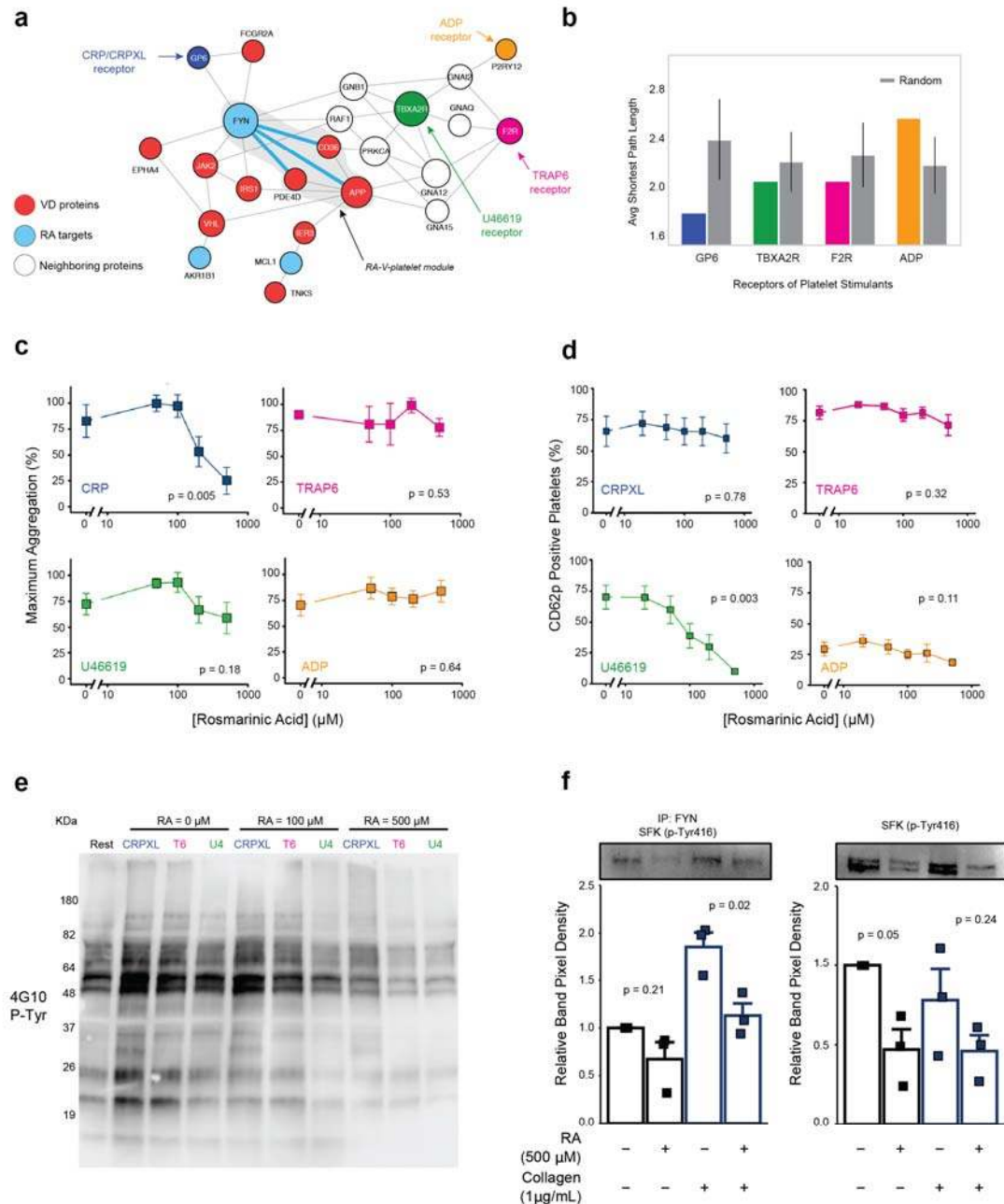
Figure 6 – Rosmarinic Acid Modulates Platelet Function. (a) Interactome neighborhood showing rosmarinic acid (RA) targets and the RA-V-platelet module - the connected component formed by the RA target FYN and the V proteins associated with platelet function PDE4D, CD36, and APP – and the receptor for platelet agonists used in our experiments (collagen/CRPXL, TRAP6, U46619, and ADP). (b) Average shortest path length from each platelet agonist receptor and the RA-V-platelet module formed by the proteins FYN, PDE4D, CD36, APP. Bars represent standard deviation of that same measure over 1000 iterations of random selection of nodes in a degree preserving fashion. c-e) Platelet-rich plasma (PRP) or washed platelets were pre-treated with RA for 1 hour before stimulation with either collagen (1 μg/mL), collagen-related peptide (CRP-XL, 1μg/mL), thrombin receptor activator peptide-6 (TRAP-6, 20 μM), U46619 (1 μM), or ADP (10 μM). Platelets were assessed for either (c) aggregation, (d) alpha

27

663   granule secretion. Platelet lysates were also probed for either (e) non-specific tyrosine phosphorylation
664   (p-Tyr) of the whole cell lysate, or (d) site-specific phosphorylation of src family kinases (SFKs) and FYN
665   at residue 416. n = 3-6 separate blood donations, mean +/- SEM. p-values in (c) and (d) were determined
666   by Kruskal-Wallis test and by unpaired t.tests in (f).
667

668   References

669

670   1.   Khera, A. V. *et al.* Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary
671   Disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).

672   2.   Arts, I. C. W. & Hollman, P. C. H. Polyphenols and disease risk in epidemiologic
673   studies. *Am. J. Clin. Nutr.* **81**, 317S-325S (2005).

674   3.   Wang, X., Ouyang, Y. Y., Liu, J. & Zhao, G. Flavonoid intake and risk of CVD: A
675   systematic review and meta-analysis of prospective cohort studies. *Br. J. Nutr.*
676   **111**, 1–11 (2014).

677   4.   Neveu, V. *et al.* Phenol-Explorer: an online comprehensive database on
678   polyphenol contents in foods. *Database* **2010**, bap024–bap024 (2010).

679   5.   Pérez-Jiménez, J., Neveu, V., Vos, F. & Scalbert, A. Systematic analysis of the
680   content of 502 Polyphenols in 452 foods and beverages: An application of the
681   phenol-explorer database. *J. Agric. Food Chem.* **58**, 4959–4969 (2010).

682   6.   Zhang, H. & Tsao, R. Dietary polyphenols, oxidative stress and antioxidant and
683   anti-inflammatory effects. *Curr. Opin. Food Sci.* (2016)
684   doi:10.1016/j.cofs.2016.02.002.

685   7.   Boly, R. *et al.* Quercetin inhibits a large panel of kinases implicated in cancer cell
686   biology. *Int. J. Oncol.* **38**, 833–842 (2011).

687   8.   Lacroix, S. *et al.* A computationally driven analysis of the polyphenol-protein
688   interactome. *Sci. Rep.* **8**, 2232 (2018).

689   9.   Hanhineva, K. *et al.* Impact of dietary polyphenols on carbohydrate metabolism.
690   *Int. J. Mol. Sci.* **11**, 1365–402 (2010).

691   10.   Hervert-Hernández, D. & Goñi, I. Dietary polyphenols and human gut microbiota:
692   A review. *Food Rev. Int.* **27**, 154–169 (2011).

693   11.   Zhang, S. *et al.* Dietary pomegranate extract and inulin affect gut microbiome
694   differentially in mice fed an obesogenic diet. *Anaerobe* **48**, 184–193 (2017).

695   12.   Thazhath, S. S. *et al.* Administration of resveratrol for 5 wk has no effect on
696   glucagon-like peptide 1 secretion, gastric emptying, or glycemic control in type 2
697   diabetes: a randomized controlled trial. *Am. J. Clin. Nutr.* **103**, 66–70 (2016).

698   13.   Bhatt, J. K., Thomas, S. & Nanjan, M. J. Resveratrol supplementation improves

29

699   glycemic control in type 2 diabetes mellitus. *Nutr. Res.* **32**, 537–41 (2012).

700 14. Sharma, A. *et al.* A disease module in the interactome explains disease

701   heterogeneity, drug response and captures novel pathways and genes in asthma.

702   *Hum. Mol. Genet.* **24**, 3005–3020 (2014).

703 15. Menche, J. *et al.* Disease networks. Uncovering disease-disease relationships

704   through the incomplete interactome. *Science* **347**, 1257601 (2015).

705 16. Guney, E., Menche, J., Vidal, M. & Barabási, A.-L. Network-based in silico drug

706   efficacy screening. *Nat. Commun.* **7**, 10331 (2016).

707 17. Cheng, F. *et al.* Network-based approach to prediction and population-based

708   validation of in silico drug repurposing. *Nat. Commun.* **9**, 1–12 (2018).

709 18. Kovács, I. A. *et al.* Network-based prediction of protein interactions. *Nat.*

710   *Commun.* **10**, 1240 (2019).

711 19. Sarkar, F. H., Li, Y., Wang, Z. & Kong, D. Cellular signaling perturbation by

712   natural products. *Cell. Signal.* **21**, 1541–7 (2009).

713 20. Iso, H. *et al.* The relationship between green tea and total caffeine intake and risk

714   for self-reported type 2 diabetes among Japanese adults. *Ann. Intern. Med.* **144**,

715   554–62 (2006).

716 21. Song, Y., Manson, J. E., Buring, J. E., Sesso, H. D. & Liu, S. Associations of

717   dietary flavonoids with risk of type 2 diabetes, and markers of insulin resistance

718   and systemic inflammation in women: a prospective study and cross-sectional

719   analysis. *J. Am. Coll. Nutr.* **24**, 376–84 (2005).

720 22. Keske, M. A. *et al.* Vascular and metabolic actions of the green tea polyphenol

721   epigallocatechin gallate. *Curr. Med. Chem.* **22**, 59–69 (2015).

722 23. Wolfram, S. *et al.* Epigallocatechin gallate supplementation alleviates diabetes in

723   rodents. *J. Nutr.* **136**, 2512–8 (2006).

724 24. Muthu, R., Selvaraj, N. & Vaiyapuri, M. Anti-inflammatory and proapoptotic effects

725   of umbelliferone in colon carcinogenesis. *Hum. Exp. Toxicol.* **35**, 1041–54 (2016).

726 25. Muthu, R. & Vaiyapuri, M. Synergistic and individual effects of umbelliferone with

727   5-fluorouracil on tumor markers and antioxidant status of rat treated with 1,2-

728   dimethylhydrazine. *Biomed. Aging Pathol.* **3**, 219–227 (2013).

729 26. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and

730   the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).

731   27.   Grover, S. P., Bergmeier, W. & Mackman, N. Platelet Signaling Pathways and

732   New Inhibitors. *Arterioscler. Thromb. Vasc. Biol.* **38**, e28–e35 (2018).

733   28.   Moco, S., Martin, F. P. J. & Rezzi, S. Metabolomics view on gut microbiome

734   modulation by polyphenol-rich foods. *J. Proteome Res.* **11**, 4781–4790 (2012).

735   29.   van Duynhoven, J. *et al.* Metabolic fate of polyphenols in the human

736   superorganism. *Proc. Natl. Acad. Sci.* **108**, 4531–4538 (2011).

737   30.   Ottaviani, J. I., Heiss, C., Spencer, J. P. E., Kelm, M. & Schroeter, H.

738   Recommending flavanols and procyanidins for cardiovascular health: Revisited.

739   *Mol. Aspects Med.* **61**, 63–75 (2018).

740   31.   Stalmach, A., Troufflard, S., Serafini, M. & Crozier, A. Absorption, metabolism and

741   excretion of Choladi green tea flavan-3-ols by humans. *Mol. Nutr. Food Res.*

742   (2009) doi:10.1002/mnfr.200800169.

743   32.   Meng, X. *et al.* Identification and characterization of methylated and ring-fission

744   metabolites of tea catechins formed in humans, mice, and rats. *Chem. Res.*

745   *Toxicol.* (2002) doi:10.1021/tx010184a.

746   33.   Perez-Vizcaino, F., Duarte, J. & Santos-Buelga, C. The flavonoid paradox:

747   Conjugation and deconjugation as key steps for the biological activity of

748   flavonoids. *J. Sci. Food Agric.* **92**, 1822–1825 (2012).

749   34.   Shimoi, K. & Nakayama, T. Glucuronidase deconjugation in inflammation.

750   *Methods Enzymol.* **400**, 263–272 (2005).

751   35.   Kaneko, A. *et al.* Glucuronides of phytoestrogen flavonoid enhance macrophage

752   function via conversion to aglycones by β-glucuronidase in macrophages. *Immun.*

753   *Inflamm. Dis.* **5**, 265–279 (2017).

754   36.   Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug

755   combinations. *Nat. Commun.* **10**, 1197 (2019).

756   37.   Smalley, J. L., Gant, T. W. & Zhang, S.-D. Application of connectivity mapping in

757   predictive toxicology based on gene-expression similarity. *Toxicology* **268**, 143–

758   146 (2010).

759   38.   Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to

760   connect small molecules, genes, and disease. *Science* **313**, 1929–35 (2006).

31

761 39. Amanzadeh, E. *et al.* Quercetin conjugated with superparamagnetic iron oxide
762   nanoparticles improves learning and memory better than free quercetin via
763   interacting with proteins involved in LTP. *Sci. Rep.* **9**, 1–19 (2019).

764 40. Shaikh, J., Ankola, D. D., Beniwal, V., Singh, D. & Kumar, M. N. V. R.
765   Nanoparticle encapsulation improves oral bioavailability of curcumin by at least 9-
766   fold when compared to curcumin administered with piperine as absorption
767   enhancer. *Eur. J. Pharm. Sci.* **37**, 223–30 (2009).

768 41. Chao, E. C. & Henry, R. R. SGLT2 inhibition-A novel strategy for diabetes
769   treatment. *Nat. Rev. Drug Discov.* **9**, 551–559 (2010).

770 42. Caldera, M. *et al.* Mapping the perturbome network of cellular perturbations. *Nat.*
771   *Commun.* **10**, (2019).

772 43. Jensen, K., Ni, Y., Panagiotou, G. & Kouskoumvekaki, I. Developing a Molecular
773   Roadmap of Drug-Food Interactions. *PLoS Comput. Biol.* **11**, 1–15 (2015).

774 44. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell*
775   **159**, 1212–1226 (2014).

776 45. Cheng, F., Jia, P., Wang, Q. & Zhao, Z. Quantitative network mapping of the
777   human kinome interactome reveals new clues for rational kinase inhibitor
778   discovery and individualized cancer therapy. *Oncotarget* **5**, 3697–710 (2014).

779 46. Calçada, D. *et al.* The role of low-grade inflammation and metabolic flexibility in
780   aging and nutritional modulation thereof: A systems biology approach. *Mech.*
781   *Ageing Dev.* (2014) doi:10.1016/j.mad.2014.01.004.

782 47. Hornbeck, P. V *et al.* PhosphoSitePlus, 2014: mutations, PTMs and
783   recalibrations. *Nucleic Acids Res.* **43**, D512-20 (2015).

784 48. Li, T. *et al.* A scored human protein-protein interaction network to catalyze
785   genomic interpretation. *Nat. Methods* **14**, 61–64 (2016).

786 49. Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2017 update.
787   *Nucleic Acids Res.* **45**, D369–D379 (2017).

788 50. Cowley, M. J. *et al.* PINA v2.0: mining interactome modules. *Nucleic Acids Res.*
789   **40**, D862-5 (2012).

790 51. Peri, S. *et al.* Human protein reference database as a discovery resource for
791   proteomics. *Nucleic Acids Res.* **32**, D497-501 (2004).

32

792   52.   Orchard, S. *et al.* The MIntAct project--IntAct as a common curation platform for
793   11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358-63 (2014).

794   53.   Breuer, K. *et al.* InnateDB: systems biology of innate immunity and beyond--
795   recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228-33 (2013).

796   54.   Meyer, M. J., Das, J., Wang, X. & Yu, H. INstruct: a database of high-quality 3D
797   structurally resolved protein interactome networks. *Bioinformatics* **29**, 1577–9
798   (2013).

799   55.   Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein
800   networks. *Nat. Methods* **10**, 47–53 (2013).

801   56.   Meyer, M. J. *et al.* Interactome INSIDER: a structural interactome browser for
802   genomic studies. *Nat. Methods* **15**, 107–114 (2018).

803   57.   Fazekas, D. *et al.* SignaLink 2 - a signaling pathway resource with multi-layered
804   regulatory networks. *BMC Syst. Biol.* **7**, 7 (2013).

805   58.   Huttlin, E. L. *et al.* Architecture of the human interactome defines protein
806   communities and disease networks. *Nature* **545**, 505–509 (2017).

807   59.   Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2019.
808   *Nucleic Acids Res.* **47**, D948–D954 (2019).

809   60.   Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated
810   over the tree of life. *Nucleic Acids Res.* **43**, D447-52 (2015).

811   61.   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based
812   approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.*
813   *U. S. A.* **102**, 15545–50 (2005).

814   62.   Roweth, H. G. *et al.* Two novel, putative mechanisms of action for citalopram-
815   induced platelet inhibition. *Sci. Rep.* **8**, 1–14 (2018).

816   63.   Roweth, H. G. *et al.* Citalopram inhibits platelet function independently of SERT-
817   mediated 5-HT transport. *Sci. Rep.* **8**, 1–14 (2018).

818   64.   Nath, S., Bachani, M., Harshavardhana, D. & Steiner, J. P. Catechins protect
819   neurons against mitochondrial toxins and HIV proteins via activation of the BDNF
820   pathway. *J. Neurovirol.* **18**, 445–455 (2012).

821   65.   Park, K.-S. *et al.* (-)-Epigallocatethin-3-O-gallate counteracts caffeine-induced
822   hyperactivity: evidence of dopaminergic blockade. *Behav. Pharmacol.* **21**, 572–

33

823   575 (2010).

824   66.   Ramesh, E., Geraldine, P. & Thomas, P. A. Regulatory effect of epigallocatechin

825   gallate on the expression of C-reactive protein and other inflammatory markers in

826   an experimental model of atherosclerosis. *Chem. Biol. Interact.* **183**, 125–32

827   (2010).

828   67.   Han, S. G., Han, S.-S., Toborek, M. & Hennig, B. EGCG protects endothelial cells

829   against PCB 126-induced inflammation through inhibition of AhR and induction of

830   Nrf2-regulated genes. *Toxicol. Appl. Pharmacol.* **261**, 181–8 (2012).

831   68.   Sheng, R., Gu, Z.-L. & Xie, M.-L. Epigallocatechin gallate, the major component of

832   polyphenols in green tea, inhibits telomere attrition mediated cardiomyocyte

833   apoptosis in cardiac hypertrophy. *Int. J. Cardiol.* **162**, 199–209 (2013).

834   69.   Devika, P. T. & Stanely Mainzen Prince, P. (-)-Epigallocatechin gallate protects

835   the mitochondria against the deleterious effects of lipids, calcium and adenosine

836   triphosphate in isoproterenol induced myocardial infarcted male Wistar rats. *J.*

837   *Appl. Toxicol.* **28**, 938–44 (2008).

838   70.   Yi, Q.-Y. *et al.* Chronic infusion of epigallocatechin-3-O-gallate into the

839   hypothalamic paraventricular nucleus attenuates hypertension and

840   sympathoexcitation by restoring neurotransmitters and cytokines. *Toxicol. Lett.*

841   **262**, 105–113 (2016).

842   71.   Devika, P. T. & Prince, P. S. M. Preventive effect of (-)epigallocatechin-gallate

843   (EGCG) on lysosomal enzymes in heart and subcellular fractions in isoproterenol-

844   induced myocardial infarcted Wistar rats. *Chem. Biol. Interact.* **172**, 245–52

845   (2008).

846   72.   Hushmendy, S. *et al.* Select phytochemicals suppress human T-lymphocytes and

847   mouse splenocytes suggesting their use in autoimmunity and transplantation.

848   *Nutr. Res.* **29**, 568–78 (2009).

849   73.   Shen, K. *et al.* Epigallocatechin 3-Gallate Ameliorates Bile Duct Ligation Induced

850   Liver Injury in Mice by Modulation of Mitochondrial Oxidative Stress and

851   Inflammation. *PLoS One* **10**, e0126278 (2015).

852   74.   ZHEN, M. *et al.* Green tea polyphenol epigallocatechin-3-gallate inhibits oxidative

853   damage and preventive effects on carbon tetrachloride–induced hepatic fibrosis.

854   *J. Nutr. Biochem.* **18**, 795–805 (2007).

855   75.   Yasuda, Y. *et al.* (–)-Epigallocatechin gallate prevents carbon tetrachloride-

856   induced rat hepatic fibrosis by inhibiting the expression of the PDGFRβ and IGF-

857   1R. *Chem. Biol. Interact.* **182**, 159–164 (2009).

858   76.   Cao, W. *et al.* iTRAQ-based proteomic analysis of combination therapy with

859   taurine, epigallocatechin gallate, and genistein on carbon tetrachloride-induced

860   liver fibrosis in rats. *Toxicol. Lett.* **232**, 233–245 (2015).

861   77.   Kitamura, M. *et al.* Epigallocatechin gallate suppresses peritoneal fibrosis in mice.

862   *Chem. Biol. Interact.* **195**, 95–104 (2012).

863   78.   Sakla, M. S. & Lorson, C. L. Induction of full-length survival motor neuron by

864   polyphenol botanical compounds. *Hum. Genet.* **122**, 635–643 (2008).

865   79.   Shimizu, M. *et al.* (–)-Epigallocatechin gallate inhibits growth and activation of the

866   VEGF/VEGFR axis in human colorectal cancer cells. *Chem. Biol. Interact.* **185**,

867   247–252 (2010).

868