



# Network medicine framework shows that proximity of polyphenol targets and disease proteins predicts therapeutic effects of polyphenols

Italo F. do Valle<sup>1</sup>, Harvey G. Roweth<sup>2,3</sup>, Michael W. Malloy<sup>2,3</sup>, Sofia Moco<sup>4</sup>, Denis Barron<sup>4</sup>, Elisabeth Battinelli<sup>2,3</sup>, Joseph Loscalzo<sup>3,5</sup> and Albert-László Barabási<sup>1,6,7</sup> ✉

**Polyphenols, natural products present in plant-based foods, play a protective role against several complex diseases through their antioxidant activity and by diverse molecular mechanisms. Here we develop a network medicine framework to uncover mechanisms for the effects of polyphenols on health by considering the molecular interactions between polyphenol protein targets and proteins associated with diseases. We find that the protein targets of polyphenols cluster in specific neighbourhoods of the human interactome, whose network proximity to disease proteins is predictive of the molecule's known therapeutic effects. The methodology recovers known associations, such as the effect of epigallocatechin-3-O-gallate on type 2 diabetes, and predicts that rosmarinic acid has a direct impact on platelet function, representing a novel mechanism through which it could affect cardiovascular health. We experimentally confirm that rosmarinic acid inhibits platelet aggregation and  $\alpha$ -granule secretion through inhibition of protein tyrosine phosphorylation, offering direct support for the predicted molecular mechanism. Our framework represents a starting point for mechanistic interpretation of the health effects underlying food-related compounds, allowing us to integrate into a predictive framework knowledge on food metabolism, bioavailability and drug interaction.**

Diet plays a defining role in human health. Indeed, while poor diet can substantially increase the risk for coronary heart disease and type 2 diabetes mellitus (T2D), a healthy diet can play a protective role, even mitigating genetic risk for coronary heart disease<sup>1</sup>. Polyphenols are a class of compounds present in plant-based foods, including fruits, vegetables, nuts, seeds, beans (for example coffee and cocoa), herbs, spices, tea and wine, that play a well-documented protective role as antioxidants, which affect several diseases, from cancer to T2D, cardiovascular and neurodegenerative diseases<sup>2,3</sup>. Previous efforts have profiled over 500 polyphenols in more than 400 foods<sup>4,5</sup> 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<sup>6</sup> by binding to proteins and modulating their activity<sup>7,8</sup>, interacting with digestive enzymes<sup>9</sup> and modulating gut microbiota growth<sup>10,11</sup>. 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<sup>12,13</sup>. We therefore need a framework to interpret the evidence present in the literature and to offer in-depth mechanistic predictions of the molecular pathways responsible for the health implications of polyphenols present in the diet. Ultimately, these insights could help us provide evidence on causal diet–health associations as well as guidelines of food consumption for different individuals and help to develop novel diagnostic and therapeutic strategies that could 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 targets, unveiling their relationship to complex diseases. The developed framework is based on the human interactome, a comprehensive subcellular network consisting of all known physical interactions between human proteins that has been validated previously as a platform for understanding disease mechanisms<sup>14,15</sup>, rational drug target identification and drug repurposing<sup>16,17</sup>.

We find that the proteins to which polyphenols bind form identifiable neighbourhoods in the human interactome, allowing us to demonstrate that the proximity between polyphenol targets and proteins associated with specific diseases is predictive of the known therapeutic effects of polyphenols. Finally, we unveil the potential therapeutic effects of rosmarinic acid (RA) on vascular diseases (VD), predicting that its mechanism of action is related to modulation of platelet function. We confirm this prediction by experiments that indicate

<sup>1</sup>Network Science Institute and Department of Physics, Northeastern University, Boston, MA, USA. <sup>2</sup>Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>3</sup>Harvard Medical School, Boston, MA, USA. <sup>4</sup>Nestlé Institute of Health Sciences, Lausanne, Switzerland. <sup>5</sup>Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. <sup>6</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>7</sup>Department of Network and Data Science, Central European University, Budapest, Hungary. ✉e-mail: [barabasi@gmail.com](mailto:barabasi@gmail.com)

that RA modulates platelet function *in vitro* by inhibiting tyrosine protein phosphorylation. Altogether, our results demonstrate that the network-based relationship between disease proteins and polyphenol targets offers a tool to systematically unveil the health effects of polyphenols.

## Results

**Polyphenol targets cluster in specific functional neighbourhoods of the interactome.** We mapped the targets of 65 polyphenols (Methods) to the human interactome, consisting of 17,651 proteins and 351,393 interactions (Fig. 1a,b). We find that 19 of the 65 polyphenols have only one protein target, while a few polyphenols have an exceptionally large number of targets (Fig. 1c). We computed the Jaccard index (JI) of the protein targets of each polyphenol pair, finding only a limited similarity of targets among different polyphenols (average JI = 0.0206) (Supplementary Fig. 1a). Even though the average JI is small, it is still significantly higher ( $Z = 147$ , Supplementary Fig. 1b) than the JI expected if the polyphenol targets were randomly assigned from the pool of all network proteins with degrees matching the original set. This finding suggests that while each polyphenol targets a specific set of proteins, their targets are confined to a common pool of proteins, likely determined by commonalities in the polyphenol-binding domains of the three-dimensional structure of the protein targets<sup>18</sup>. Gene ontology enrichment analysis recovers existing mechanisms<sup>8</sup> and also helps identify new processes related to polyphenol protein targets, such as post-translational protein modifications, regulation and xenobiotic metabolism (Fig. 1d). The enriched gene ontology categories indicate that polyphenols modulate common regulatory processes, but the low similarity in their protein targets, illustrated by the low average JI, indicates that they target different processes within the same process.

We next asked whether the polyphenol targets cluster in specific regions of the human interactome. We focused on polyphenols with more than two targets ( $n = 46$ , Fig. 2) and measured the size and significance of the largest connected component (LCC) formed by the targets of each polyphenol. We found that 25 of the 46 polyphenols have a larger LCC than expected by chance ( $Z$ -score  $> 1.95$ ) (Fig. 1e and Fig. 2). In agreement with experimental evidence documenting the effect of polyphenols on multiple pathways<sup>19</sup>, we find that ten polyphenols have their targets organized in multiple connected components of size  $> 2$ .

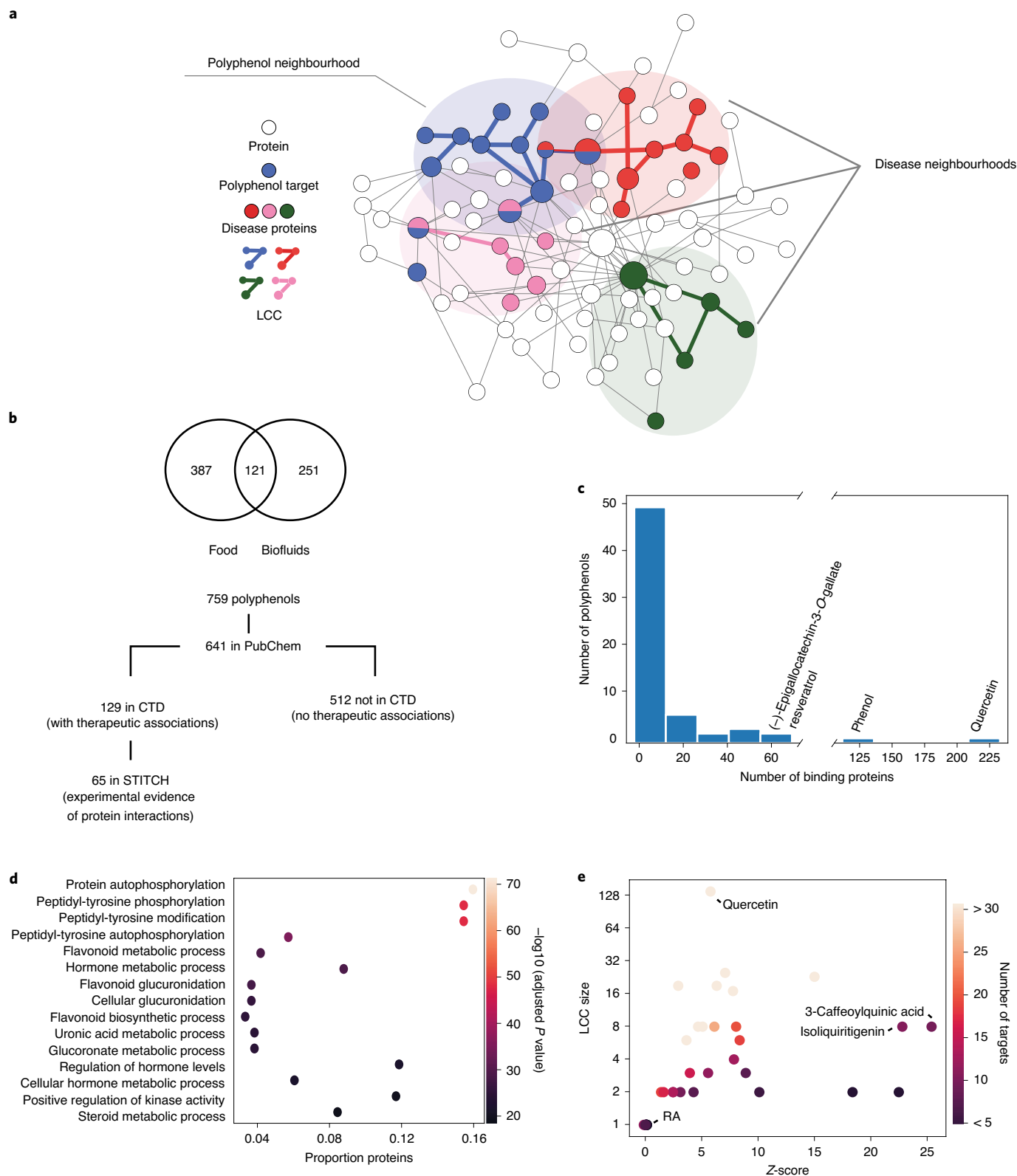
These results indicate that the targets of polyphenols modulate specific well-localized neighbourhoods of the interactome (Fig. 2 and Supplementary Fig. 1c). This prompted us to explore whether the interactome regions targeted by the polyphenols reside within network neighbourhoods associated with specific diseases, thereby seeking a network-based framework to unveil the molecular mechanisms through which specific polyphenols modulate health.

**Proximity between polyphenol targets and disease proteins reveals their therapeutic effects.** Polyphenols can be viewed as drugs in that they bind to specific proteins, affecting their ability to perform their normal functions. We therefore hypothesized that we can apply the network-based framework used to predict the efficacy of drugs in specific diseases<sup>16,17</sup> to also predict the therapeutic effects of polyphenols. The closer the targets of a polyphenol are to disease proteins, the more likely that the polyphenol will affect the disease phenotype. We therefore calculated the network proximity between polyphenol targets and proteins associated with 299 diseases using the closest measure,  $d_c$ , representing the average shortest path length between each polyphenol target and the nearest disease protein (Methods). Consider, for example, (–)-epigallocatechin-3-*O*-gallate (EGCG), a polyphenol abundant in green tea. Epidemiological studies have found a positive relationship between green tea consumption and reduced risk of T2D<sup>20,21</sup>,

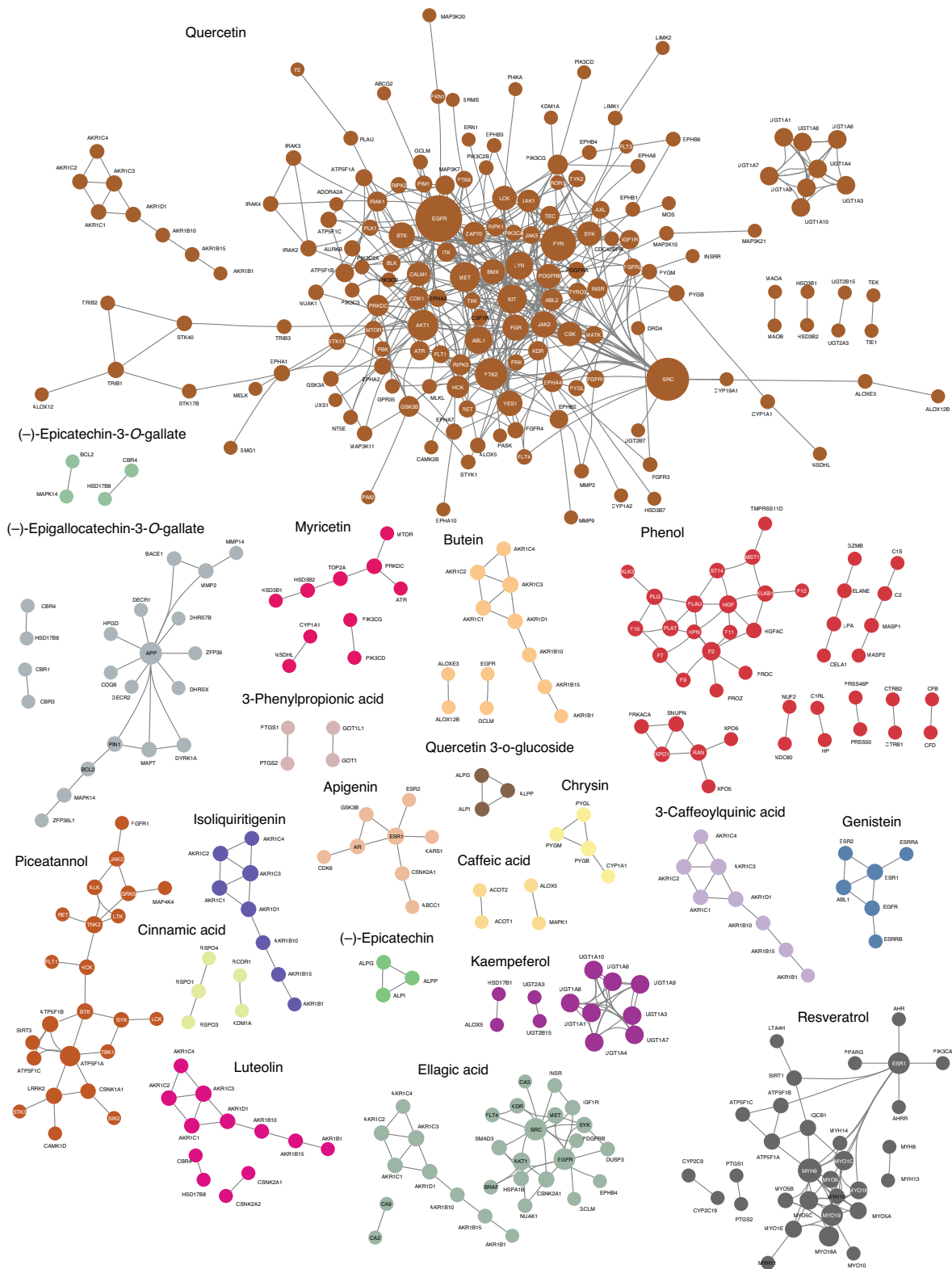
and physiological and biochemical studies have shown that EGCG presents glucose-lowering effects in both *in vitro* and *in vivo* models<sup>22,23</sup>. We identified 54 experimentally validated EGCG protein targets and mapped them to the interactome, finding that the EGCG targets form an LCC of 17 proteins ( $Z$ -score = 7.61) (Fig. 3a). We also computed the network-based distance between EGCG targets and 83 proteins associated with T2D, finding that the two sets are significantly proximal to each other. We ranked all 299 diseases based on their network proximity to the EGCG targets to determine whether we could recover the 82 diseases in which EGCG has known therapeutic effects according to the comparative toxicogenomics database (CTD)<sup>24</sup>. By this analysis, we were able to recover 15 previously known therapeutic associations among the top 20 ranked diseases (Table 1), confirming that network proximity can discriminate between known and unknown disease associations for polyphenols, as previously confirmed for drugs<sup>16,17</sup>.

We expanded these methods to all polyphenol–disease pairs to predict diseases for which specific polyphenols might have therapeutic effects. For this analysis, we grouped all 19,435 polyphenol–disease associations between 65 polyphenols and 299 diseases into known (1,525) and unknown (17,910) associations. The known polyphenol–disease set was retrieved from the CTD, which is limited to manually curated associations for which there is literature-based evidence. For each polyphenol, we tested how well network proximity discriminates between the known and unknown sets by evaluating the area under the curve (AUC) of the receiving operating characteristic curve. For EGCG, network proximity offers good discriminative power (AUC = 0.78, CI = 0.70–0.86) between diseases with known and unknown therapeutic associations (Table 1). We find that network proximity ( $d_c$ ) offers predictive power with an AUC  $> 0.7$  for 31 polyphenols (Fig. 3b). The methodology recovers many associations well-documented in the literature, such as the beneficial effects of umbelliferone on colorectal neoplasms<sup>25,26</sup>. In Table 2, we summarize the top 10 polyphenols for which the network medicine framework offers the best predictive power of therapeutic effects, limiting the entries to those with predictive performance of AUC  $> 0.6$  and where the precision of the performance of the top predictions is greater than 0.6. Given the lack of data on true negative examples, we considered unknown associations as negative cases, observing the same trend when we used an alternative performance metric that does not require true negative labels (that is, AUC of the Precision–Recall curve) (Supplementary Fig. 2).

Finally, we performed multiple robustness checks to exclude the role of potential biases in the input data. To test whether the predictions are biased by the set of known associations retrieved from the CTD, we randomly selected 100 papers from PubMed containing medical subject headings (MeSH) terms that tag EGCG to diseases. We manually curated the evidence for EGCG's therapeutic effects for the diseases discussed in the published papers, excluding reviews and non-English language publications. The dataset was processed to include implicit associations (Methods), resulting in a total of 113 diseases associated with EGCG, of which 58 overlap with the associations reported by the CTD (Fig. 3c). We observed that the predictive power of network proximity was unaffected by whether we considered the annotations from the CTD, the manually curated list or the union of both (Fig. 3d). To test the role of potential biases in the interactome, we repeated our analysis using only high-quality polyphenol–protein interactions retrieved from ligand–protein three-dimensional resolved structures (Supplementary Fig. 1d) and a subset of the interactome derived from an unbiased high-throughput screening (Supplementary Fig. 1f). We found that the predictive power was largely unchanged, indicating that the literature bias in the interactome does not affect our findings. Finally, we retested the predictive performance by considering not only the therapeutic polyphenol–disease associations, but also the



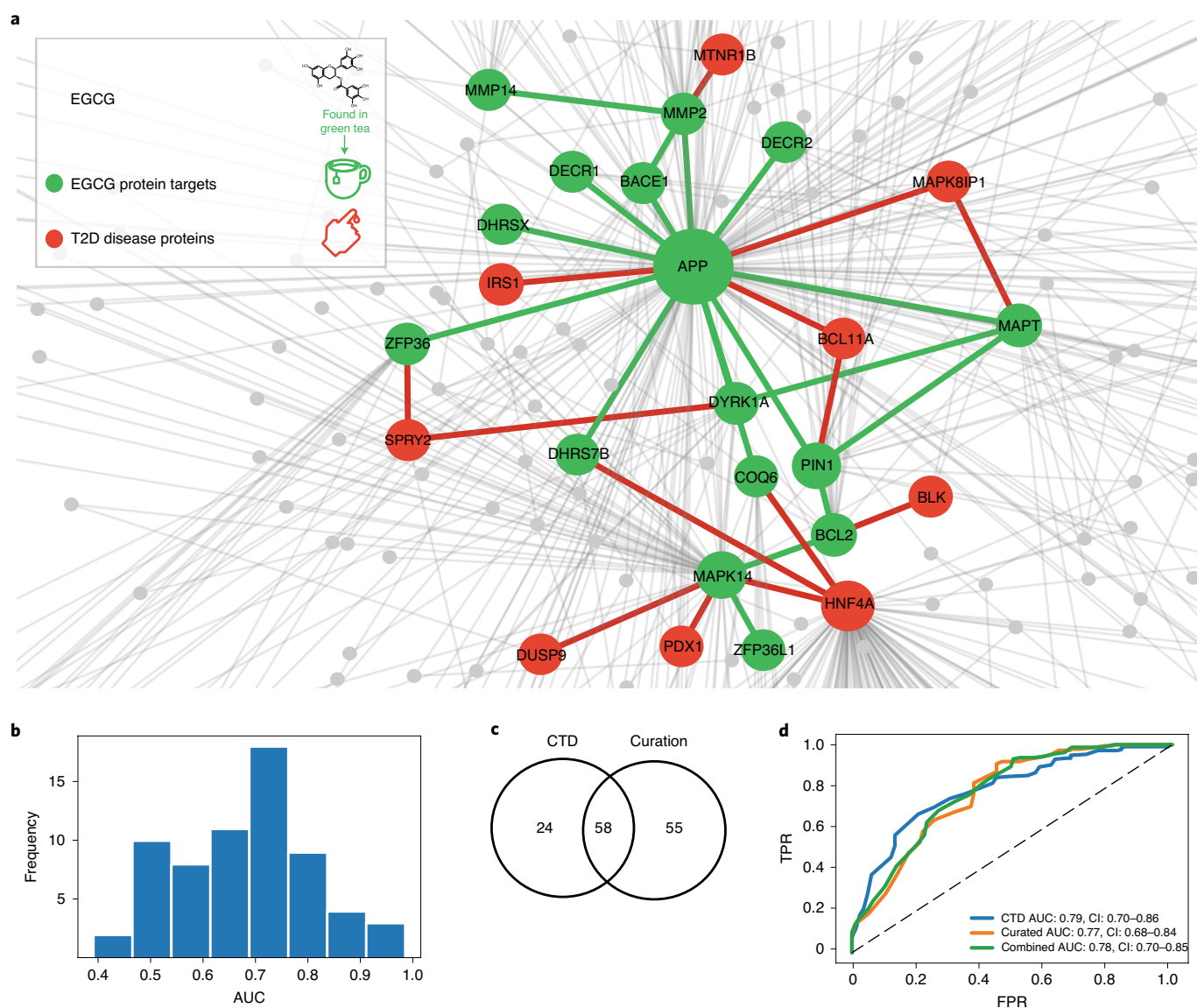
**Fig. 1 | Properties of polyphenol protein targets.** **a**, Schematic representation of the human interactome, highlighting regions where polyphenol targets and disease proteins are localized. **b**, Diagram showing the selection criteria for the polyphenols evaluated in this study. **c**, Distribution of the number of polyphenol protein targets mapped to the human interactome. **d**, Top ( $n=15$ ) enriched gene ontology terms (Biological Process) among all polyphenol protein targets. The x axis shows the proportion of targets mapped to each pathway. **e**, Size of the LCC formed by the targets of each polyphenol in the interactome and the corresponding significance (Z-score).



**Fig. 2 | Protein-protein interactions of polyphenol targets.** The 23 polyphenols whose targets form connected components in the interactome and their respective subgraphs. For example, piceatannol targets form a unique connected component of 23 proteins, while quercetin targets form multiple connected components, the largest having 140 proteins. Polyphenol targets that are not connected to any other target are not shown in the figure. Colours distinguish connected components of different polyphenols.

marker/mechanism ones (another type of curated association available in the CTD) finding that the predictive power remains largely unchanged (Supplementary Notes and Supplementary Fig. 3).

**Network proximity predicts gene expression perturbation induced by polyphenols.** To validate that network proximity reflects the biological activity of polyphenols observed in experimental



**Fig. 3 | Proximity between polyphenol targets and disease proteins is predictive of the therapeutic effects of the polyphenol. a**, Interactome neighbourhood showing the EGCG protein targets and their interactions with T2D-associated proteins. **b**, Distribution of AUC values of the predictions of therapeutic effects for 65 polyphenols. **c**, Comparison of the EGCG-disease associations for the CTD and the in-house database derived from the manual curation of the literature. **d**, Comparison of the prediction performance of the known EGCG-disease associations from the CTD, the in-house manually curated database or the combined datasets. CI, confidence interval; TPR, true positive rate, FPR, false positive rate; dashed line, random expectation.

data, we retrieved expression perturbation signatures from the Connectivity Map database<sup>27</sup> for the treatment of the breast cancer MCF7 cell line with 21 polyphenols (Supplementary Table 1 and Supplementary Fig. 4). We investigated the relationships between the extent to which polyphenols perturb the expression of disease genes, the network proximity between the polyphenol targets and disease proteins and their known therapeutic effects (Fig. 4a). For example, we observe different perturbation profiles for gene pools associated with different diseases: for treatment with genistein (1  $\mu$ M, 6 hours) we observed 10 skin disease genes with perturbation scores greater than 2, while we observed only one highly perturbed cerebrovascular disorder gene (Fig. 4b). Indeed, network proximity indicates that skin disease is closer to the genistein targets than cerebrovascular disorder, suggesting a relationship between network proximity, gene expression perturbation and the therapeutic effects of the polyphenol (Fig. 4a). To test this hypothesis, we computed an enrichment score that measures the over-representation

of disease genes among the most perturbed genes (Methods), finding 13 diseases that have their genes significantly enriched among the genes most deregulated by genistein, of which four have known therapeutic associations. We find that these four diseases are significantly closer to the genistein targets than the nine diseases with unknown therapeutic associations (Fig. 4c). We observed a similar trend for treatments with other polyphenols, whether we use the same concentration (1  $\mu$ M, Fig. 4c) or different ones (100 nM to 10  $\mu$ M, Supplementary Fig. 5). This result suggests that changes in gene expression caused by a polyphenol are indicative of its therapeutic effects, but only if the observed expression change is limited to proteins proximal to the polyphenol targets (Fig. 4a).

Consequently, network proximity should also be predictive of the overall gene expression perturbation caused by a polyphenol on the genes of a given disease. To test this hypothesis, in each experimental combination defined by the polyphenol type and its concentration, we evaluated the maximum perturbation among

**Table 1 | Top 20 predicted therapeutic associations between EGCG and human diseases**

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

Diseases were ordered according to the network distance ( $d_c$ ) of their proteins to EGCG targets, and diseases with relative distance  $Z_{d_c} > -0.5$  were removed. References reported in the CTD for curated therapeutic associations are shown. NA indicates diseases with no documented therapeutic association in CTD.

genes for each disease. We then compared the magnitude of the observed perturbation between diseases that were proximal ( $d_c$  below the 25th percentile,  $Z_{d_c} < -0.5$ ) or distal ( $d_c$  above the 75th percentile,  $Z_{d_c} > -0.5$ ) to the polyphenol targets. Figure 5a,b and Supplementary Fig. 6 show the results for the genistein treatment (1  $\mu$ M, 6 hours), which indicate that diseases proximal to the polyphenol targets show higher maximum perturbation values than distal diseases. The same trend is observed for other polyphenols when we use different  $d_c$  and  $Z_{d_c}$  thresholds for defining proximal and distant diseases (Fig. 5b and Supplementary Figs. 6–9), confirming that the impact of a polyphenol on cellular signalling pathways is localized in the network space, being greater in the vicinity of the polyphenol targets than in neighbourhoods remote from these targets. We also considered gene expression perturbations in the network vicinity of the polyphenol targets, regardless of whether the proteins were disease proteins, and observed higher perturbation scores for proximal proteins in 12 out of 21 polyphenols tested at 10  $\mu$ M (Supplementary Fig. 10). Finally, we found that the enrichment score of perturbed genes among disease genes was not as predictive of the polyphenol therapeutic effects as network proximity (Supplementary Fig. 11).

Altogether, these results indicate that network proximity offers a mechanistic interpretation for the gene expression perturbations induced by polyphenols on disease genes. They also show that network proximity can indicate when gene expression perturbations

result in therapeutic effects, suggesting that future studies could integrate gene expression (whenever available) with network proximity as they aim to more accurately prioritize polyphenol-disease associations.

**Experimental evidence confirms that RA modulates platelet function.** To demonstrate how the network-based framework can facilitate the mechanistic interpretation of the therapeutic effects of selected polyphenols, we next focus on VD. Of 65 polyphenols evaluated in this study, we found 27 to have associations to VD, as their targets were within the VD network neighbourhood (Supplementary Table 3). We therefore inspected the targets of 15 of the 27 polyphenols with 10 or fewer targets. The network analysis identified direct links between biological processes related to vascular health and the targets of three polyphenols: gallic acid, RA and 1,4-naphthoquinone (Supplementary Fig. 12 and Supplementary Notes). The network neighbourhood containing the targets of these polyphenols suggests that gallic acid activity involves thrombus dissolution processes, RA acts on platelet activation and antioxidant pathways through FYN and its neighbours and 1,4-naphthoquinone acts on signalling pathways of vascular cells through MAP2K1 activity (Supplementary Fig. 12 and Supplementary Notes).

To validate the developed framework, we set out to obtain direct experimental evidence of the predicted mechanistic role of RA in VD. The RA targets are in close proximity to proteins related to platelet function, forming the RA/VD-platelet module: a connected component formed by the RA target FYN and the VD proteins associated with platelet function PDE4D, CD36 and APP (Fig. 6a). We therefore asked whether RA influenced platelet activation in vitro. As platelets can be stimulated through different activation pathways, RA effects can, in principle, occur in any of them. To test these different possibilities, we pretreated platelets with RA and then either (1) activated glycoprotein VI by collagen or collagen-related peptide (CRP/CRPXL); (2) activated protease-activated receptors-1,4 by thrombin receptor activator peptide-6 (TRAP-6); (3) activated prostanoid thromboxane receptor by the thromboxane  $A_2$  analogue (U46619) or (4) activated P2Y1/12 receptor by adenosine diphosphate (ADP)<sup>28</sup>. When we compared the network distance between each stimulant receptor and the RA/VD-platelet module (Fig. 6a), we observed that the receptors for CRP/CRPXL, TRAP-6 and U46619 are closer than would be expected for a random distribution, while the receptor for ADP is more distant (Fig. 6b). We expected that platelets would be most affected by RA when treated with stimulants whose receptors are most proximal to the RA/VD-platelet module, that is, CRP/CRPXL, TRAP-6 and U46619, and as a control, we expect no effect for the distant ADP receptor. The experiments confirm this prediction: RA inhibits collagen-mediated platelet aggregation (Fig. 6c) and impairs dense granule secretion induced by CRPXL, TRAP-6 and U46619 (Supplementary Fig. 13). RA-treated platelets also displayed dampened  $\alpha$ -granule secretion (Fig. 6d) and integrin  $\alpha_{IIb}\beta_3$  activation (Supplementary Fig. 13) in response to U46619. As expected, RA did not affect platelet function when we used an agonist whose receptor is distant from the RA/VD-platelet module (that is, ADP). These findings suggest that RA impairs basic hallmarks of platelet activation via strong network effects, supporting our hypothesis that the proximity between RA targets and the neighbourhood associated with platelet function (Fig. 6a) could in part explain RA's impact on VD.

We next sought to clarify the molecular mechanisms involved in the impact of RA on platelets. Given that platelet activation is coordinated by several kinases, we hypothesized that RA inhibits platelet function by blocking agonist-induced protein tyrosine phosphorylation. We observed that RA-treated platelets demonstrated a dose-dependent reduction in total tyrosine phosphorylation in response to CRPXL, TRAP-6 and U46619 (Fig. 6e). Given that RA caused a substantial decrease in phosphorylation of proteins with

**Table 2 | Top-ranked polyphenols**

Polyphenol	AUC	AUC CI <sup>a</sup>	Precision <sup>b</sup>	Concentration in blood <sup>c</sup> (μM)	No. of 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]	18	6
Ellagic acid	0.79	[0.63–0.92]	0.6		42	19
EGCG	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]	216	140
(–)-Epicatechin	0.65	[0.49–0.80]	0.8	0.625	11	3

Polyphenols for which prediction of therapeutic effects by network proximity to diseases was most successful. The table shows polyphenols with AUC > 0.6 and precision > 0.6. <sup>a</sup>CI<sub>s</sub> calculated with 2,000 bootstraps with replacement and sample size of 50% of the diseases (150/299). <sup>b</sup>Precision was calculated based on the top 10 polyphenols after their ranking based on the distance ( $d_i$ ) of their targets to the disease proteins and considering only predictions with Z-score < -0.5. <sup>c</sup>Concentrations of polyphenols in blood were retrieved from the Human Metabolome Database (HMDB).

atomic mass between 50 and 60 kDa (Fig. 6e), we hypothesized that RA may reduce phosphorylation of FYN (59 kDa) or other similarly sized members of the same protein family (that is, SFKs). To test this, we measured the level of phosphorylation within the activation domain (amino acid 416) of SFKs, finding that RA reduced collagen induced phosphorylation of FYN as well as basal tyrosine phosphorylation of SFKs (Fig. 6f). This indicates that RA perturbs the phospho-signalling networks that regulate platelet response to extracellular stimuli.

Taken together, these findings support our prediction that RA modulates platelet activation and function. They also support the observation that its mechanism of action involves reduction of phosphorylation at the activation domain of the protein tyrosine kinase FYN (Fig. 6a) and the inhibition of general tyrosine phosphorylation. Finally, while polyphenols are usually associated with their antioxidant function, here we illustrate another mechanistic pathway through which they could benefit health.

## Discussion

Here, we propose a network-based framework to predict the therapeutic effects of dietary polyphenols in human diseases. We find that polyphenol protein targets cluster in specific functional neighbourhoods of the interactome, and we show that the network proximity between polyphenol targets and disease proteins is predictive of the therapeutic effects of polyphenols. We demonstrate that diseases whose proteins are proximal to polyphenol targets tend to have significant changes in gene expression in cell lines treated with the respective polyphenol, while such changes are absent for diseases whose proteins are distal to polyphenol targets. Finally, we find that the network neighbourhood around the RA targets and VD proteins are related to platelet function. We validate this mechanistic prediction by showing that RA modulates platelet function through inhibition of protein tyrosine phosphorylation. These observations suggest a role of RA on prevention of VD by inhibiting platelet activation and aggregation.

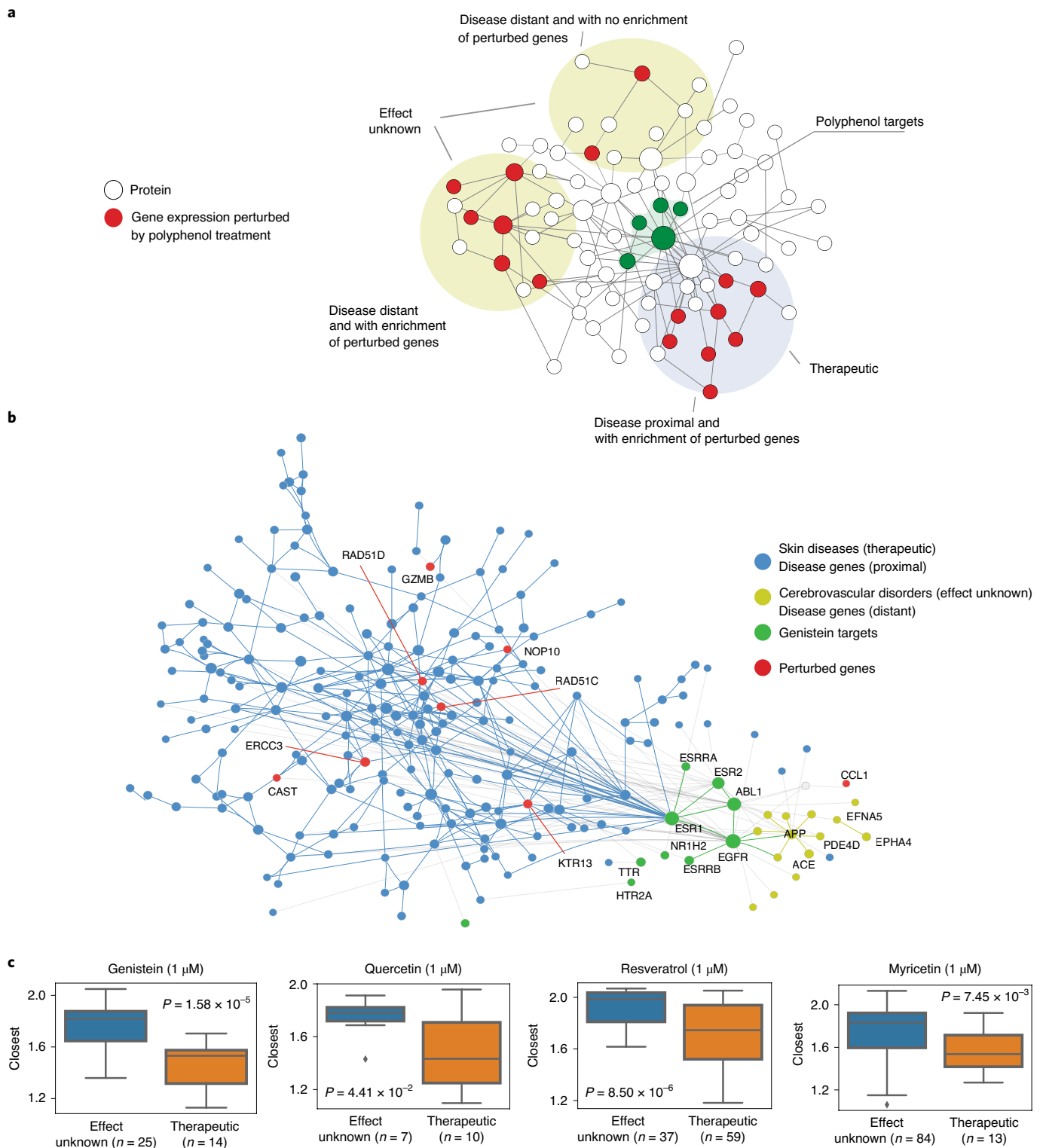
The observed results also suggest multiple avenues through which our ability to understand the role of polyphenols could be improved. First, some of the known health benefits of polyphenols might be caused not only by the native molecules but also by their metabolic byproducts<sup>29,30</sup>. However, we lack data concerning colonic degradation, liver metabolism, bioavailability and interaction with proteins of specific polyphenols or their metabolic byproducts. Future experimental data on protein interactions with polyphenol byproducts and conjugates could be incorporated in the proposed framework, further improving the accuracy of our predictions.

The lack of these data does not invalidate the findings presented here, since previous studies report the presence of unmetabolized polyphenols in blood<sup>31–33</sup> and it has been hypothesized that, in some instances, deconjugation of liver metabolites occurs in specific tissues or cells<sup>34–36</sup>. Therefore, the lack of data concerning specific polyphenols and the fact that other mechanisms exist through which they can affect health (for example, antioxidant activity and microbiota regulation) explain why this methodology might still miss a few known relationships between polyphenols and diseases. Second, considering that several experimental studies of polyphenol bioefficacy have been observed in *in vitro* and *in vivo* models, the proposed framework might help us interpret literature evidence, possibly even allowing us to exclude chemical candidates when considering the health benefits provided by a given food in epidemiological association studies.

Our assumption that network proximity recovers therapeutic associations is based on its predictive performance on a ground-truth dataset for observed therapeutic effects and also relies on previous observations about the effect of drugs on diseases<sup>16,17,37</sup>. While the proposed methodology offers a powerful prioritization tool to guide future research, the real effect of polyphenols on diseases might still be negative, given other unmet factors such as dosage, comorbidities and drug interactions, which can only be ruled out by preclinical and clinical studies. Gene expression perturbation profiles, such as the ones provided by the Connectivity Map, can also be integrated with network proximity to further highlight potential beneficial or harmful effects of chemical compounds<sup>38,39</sup>.

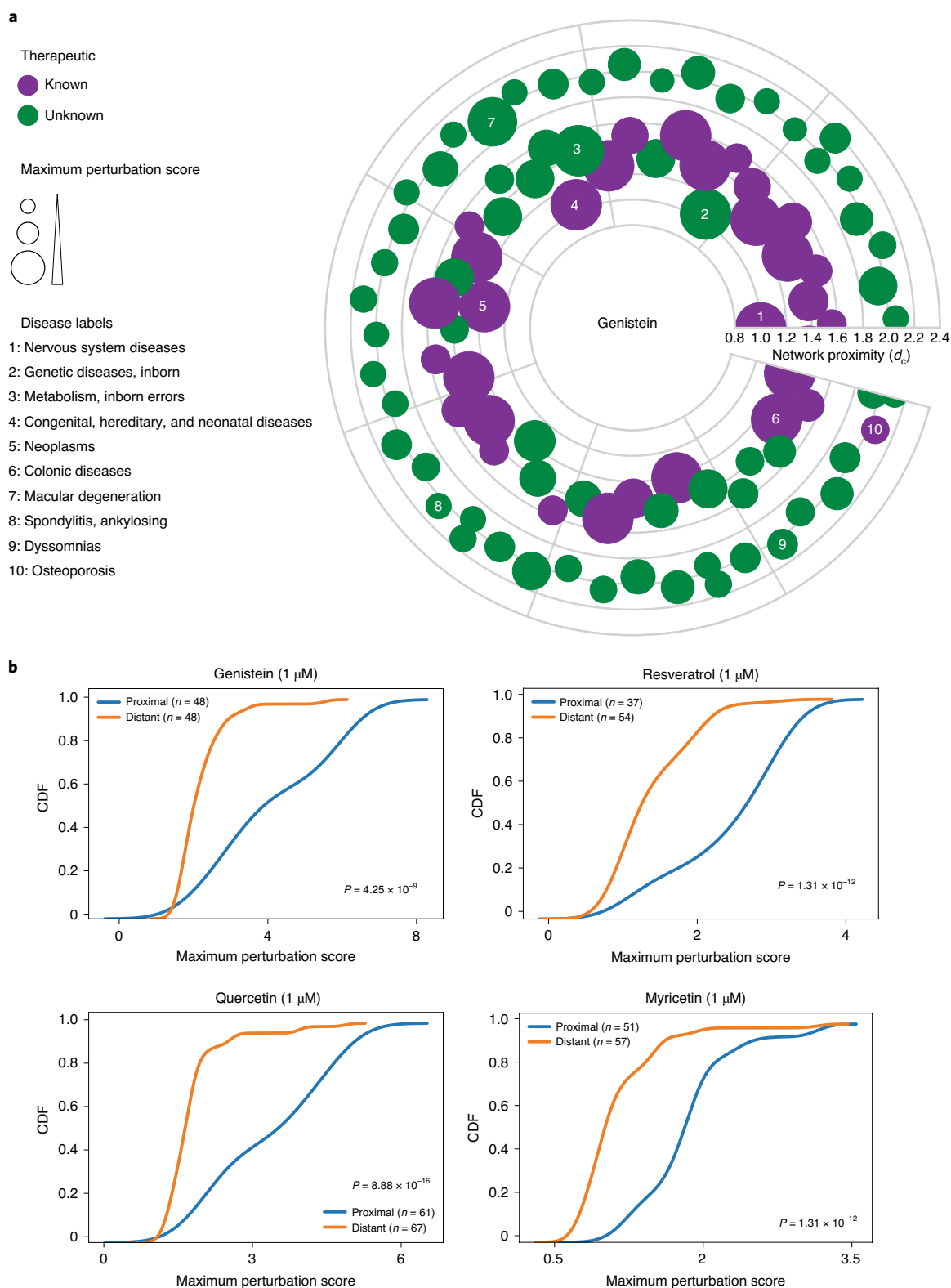
The low bioavailability of some polyphenols in food might still present challenges when considering the therapeutic utility of these molecules. However, 48 of the 65 polyphenols we explored here are predicted to have high gastrointestinal absorption (Supplementary Table 2) and different methodologies are available to increase bioavailability of natural compounds<sup>40,41</sup>. Additionally, in the same way that the polyphenol phlorizin led to the discovery of new strategies for disease treatment resulting in the development of new compounds with higher efficacy<sup>42</sup>, we believe that the present methodology can help us identify polyphenol-based candidates for drug development.

The methodology introduced here offers a foundation for the mechanistic interpretation of alternative pathways through which polyphenols can affect health, such as the combined effect of different polyphenols<sup>37,43</sup> and their interactions with drugs<sup>44</sup>. To address such synergistic effects, we need ground-truth data on these aspects. The developed methodology can be applied to other food-related chemicals, providing a framework by which to understand their

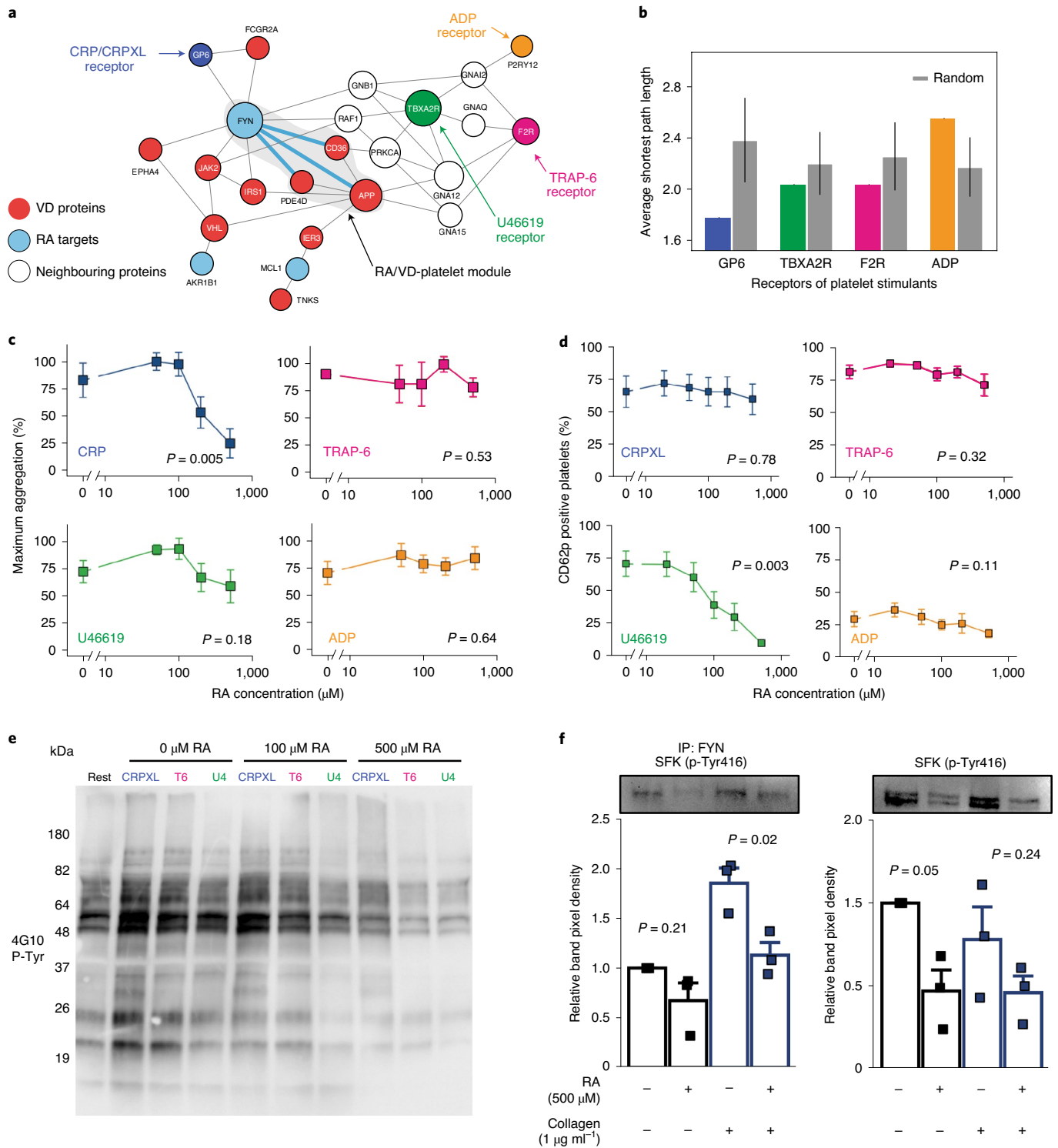


**Fig. 4 | Relationships among gene expression perturbation, network proximity and the therapeutic effects of polyphenols on diseases. a**, Schematic representation of the relationships between the extent to which a polyphenol perturbs disease genes expression, its proximity to the disease genes and its therapeutic effects. **b**, Interactome neighbourhood showing the modules of skin diseases, genistein and cerebrovascular disorders. The skin diseases module has 10 proteins with high perturbation scores ( $>2$ ) in the treatment of the MCF7 cell line with  $1\mu\text{M}$  of genistein. Genes associated with skin disease are significantly enriched among the most differentially expressed genes, and the maximum perturbation score among disease genes is higher in skin disease than cerebrovascular disorders. **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. Boxplots show the median (horizontal line), 25th and 75th percentiles (lower and upper boundaries, respectively). Whiskers extend to data points that lie within 1.5 interquartile ranges of the 25th and 75th quartiles; and observations that fall outside this range are displayed independently.





**Fig. 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  $\mu$ M, 6 hours). Distance from the origin represents the network proximity ( $d_c$ ) to genistein targets. Purple nodes represent diseases for which the therapeutic association was previously known. **b**, Cumulative distribution function (CDF) of the maximum perturbation scores of genes from diseases that are distal or proximal to polyphenol targets for the polyphenols genistein, quercetin, resveratrol and myricetin (1  $\mu$ M, 6 hours). Statistical significance was evaluated with the Kolmogorov–Smirnov test.



**Fig. 6 | RA modulates platelet function.** **a**, Interactome neighbourhood showing RA targets and the RA/VD-platelet module (that is, the connected component formed by the RA target FYN and the VD proteins associated with platelet function (PDE4D, CD36 and APP)) and the receptor for platelet agonists used in our experiments (collagen/CRPXL, TRAP-6, U46619 and ADP). **b**, Average shortest path length from each platelet agonist receptor and the RA/VD-platelet module formed by the proteins FYN, PDE4D, CD36 and APP. Error bars represent standard deviation of that same measure over 1,000 iterations of random selection of nodes in a degree-preserving fashion. **c-f**, Platelet-rich plasma (PRP) or washed platelets were pretreated with RA for 1 hour before stimulation with either collagen (1  $\mu\text{g ml}^{-1}$ ), collagen-related peptide (CRPXL, 1  $\mu\text{g ml}^{-1}$ ), thrombin receptor activator peptide-6 (TRAP-6, 20  $\mu\text{M}$ ), U46619 (1  $\mu\text{M}$ ) or ADP (10  $\mu\text{M}$ ). Platelets were assessed for either aggregation (**c**) or  $\alpha$ -granule secretion (**d**). Platelet lysates were also probed for either non-specific tyrosine phosphorylation (p-Tyr) of the whole cell lysate (**e**) or site-specific phosphorylation of Src family kinases (SFKs) and FYN at residue 416 (**f**).  $n = 3-6$  separate blood donations. In **c** and **d**, points represent the mean and error bars represent  $\pm$  s.e.m. In **f**, points represent values from biological replicates and error bars represent s.e.m.  $P$  values were determined by the Kruskal-Wallis test in **c** and **d** and by unpaired  $t$ -tests in **f**.

health effects. Future research may help us also account for the way food-related chemicals affect endogenous metabolic reactions, impacting not only signalling pathways but also catabolic and anabolic processes. Finally, the methodology provides a framework to interpret and find causal support for associations identified in observational studies. Taken together, the proposed network-based framework has the potential to systematically reveal the mechanism of action underlying the health benefits of polyphenols, offering a logical, rational strategy for mechanism-based drug development of food-based compounds.

## Methods

**Building the interactome.** The human interactome was assembled from 16 databases containing six different types of protein–protein interactions (PPIs): (1) binary PPIs tested by high-throughput yeast two-hybrid (Y2H) experiments<sup>45</sup>; (2) kinase–substrate interactions from literature-derived low-throughput and high-throughput experiments from KinomeNetworkX<sup>46</sup>, Human Protein Resource Database (HPRD)<sup>47</sup> and PhosphositePlus<sup>48</sup>; (3) carefully literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS) and from literature-derived low-throughput experiments from InWeb<sup>49</sup>, BioGRID<sup>50</sup>, PINA<sup>51</sup>, HPRD<sup>52</sup>, MINT<sup>53</sup>, IntAct<sup>54</sup> and InnateDB<sup>55</sup>; (4) high-quality PPIs from three-dimensional protein structures reported in Instruct<sup>55</sup>, Interactome3D<sup>56</sup> and INSIDER<sup>57</sup>; (5) signalling networks from literature-derived low-throughput experiments as annotated in SignaLink2.0 (ref. <sup>58</sup>) and (6) protein complexes from BioPlex2.0 (ref. <sup>59</sup>). The genes were mapped to their Entrez ID based on the National Center for Biotechnology Information (NCBI) database as well as their official gene symbols. The resulting interactome includes 351,444 PPIs connecting 17,706 unique proteins (Supplementary Data 1). The LCC has 351,393 PPIs and 17,651 proteins.

**Polyphenols, polyphenol targets and disease proteins.** We retrieved 759 polyphenols from the PhenolExplorer database<sup>1</sup>. The database lists polyphenols with food composition data or that have been profiled in biofluids after interventions with polyphenol-rich diets. For our analysis, we only considered polyphenols that (1) could be mapped in PubChem IDs, (2) were listed in the Comparative Toxicogenomics (CTD) database<sup>24</sup> as having therapeutic effects on human diseases and (3) had protein-binding information present in the STITCH database<sup>60</sup> with experimental evidence (Fig. 1a). After these steps, we considered a final list of 65 polyphenols, for which 598 protein targets were retrieved from STITCH (Supplementary Table 1). We considered 3,173 disease proteins corresponding to 299 diseases retrieved from Menche et al.<sup>15</sup>. Gene ontology enrichment analysis of protein targets was performed using the Bioconductor package clusterProfiler with a significance threshold of  $P < 0.05$  and Benjamini–Hochberg multiple testing correction with  $Q < 0.05$ .

**Polyphenol–disease associations.** We retrieved the polyphenol–disease associations from the Comparative Toxicogenomics Database (CTD). We considered only manually curated associations labelled as therapeutic. By considering the hierarchical structure of diseases along the MeSH tree, we expanded explicit polyphenol–disease associations to also include implicit associations. This procedure was performed by propagating associations in the lower branches of the MeSH tree to consider diseases in the higher levels of the same tree branch. For example, a polyphenol associated with heart diseases would also be associated with the more general category of cardiovascular diseases. By performing this expansion, we obtained a final list of 1,525 known associations between the 65 polyphenols and the 299 diseases considered in this study.

### Network proximity between polyphenol targets and disease proteins.

The proximity between a disease and a polyphenol was evaluated using a distance metric that takes into account the shortest path lengths between polyphenol targets and disease proteins<sup>16</sup>. Given  $S$ , the set of disease proteins,  $T$ , the set of polyphenol targets and  $d(s, t)$ , the shortest path length between nodes  $s$  and  $t$  in the network, we define:

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

We also calculated a relative distance metric ( $Z_{d_c}$ ) that compares the absolute distance  $d_c(s, T)$  between a disease and a polyphenol with a reference distribution describing the random expectation. The reference distribution corresponds to the expected distances between two randomly selected groups of proteins matching the size and degrees of the original disease proteins and polyphenol targets in the network. It was generated by calculating the proximity between these two randomly selected groups across 1,000 iterations. The mean  $\mu_{d_c(s, T)}$  and standard deviation  $\sigma_{d_c(s, T)}$  of the reference distribution were used to convert the absolute distance  $d_c$  into the relative distance  $Z_{d_c}$ , defined as:

$$Z_{d_c} = \frac{d - \mu_{d_c(s, T)}}{\sigma_{d_c(s, T)}} \quad (2)$$

We performed a degree-preserving random selection, but due to the scale-free nature of the human interactome, we avoid repeatedly choosing the same (high-degree) nodes by using a binning approach in which nodes within a certain degree interval were grouped together such that there were at least 100 nodes in the bin. Supplementary Data 2 reports the proximity scores  $d_c$  and  $Z_{d_c}$  for all pairs of diseases and polyphenols.

**Area under receiving operating characteristic curve analysis.** For each polyphenol, we used the AUC to evaluate how well the network proximity distinguishes diseases with known therapeutic associations from all the others in the set of 299 diseases. The set of known associations (therapeutic) retrieved from the CTD were used as positive instances and all unknown associations were defined as negative instances, and the area under the receiving operating characteristic curve was computed using the implementation in the Scikit-learn Python package. Furthermore, we calculated 95% CIs using the bootstrap technique with 2,000 resamplings with sample sizes of 150 each. Considering that the AUC provides an overall performance, we also searched for a metric to evaluate the top-ranking predictions. For this analysis, we calculated the precision of the top 10 predictions, considering only the polyphenol–disease associations with relative distance  $Z_{d_c} < -0.5$  (ref. <sup>16</sup>).

**Analysis of network proximity and gene expression deregulation.** We retrieved perturbation signatures from the Connectivity Map database (<https://clue.io/>) for the MCF7 cell line after treatment with 21 polyphenols. These signatures reflect the perturbation of the gene expression profile caused by the treatment with that particular polyphenol relative to a reference population, which comprises all other treatments in the same experimental plate<sup>27</sup>. For polyphenols having more than one experimental instance (such as time of exposure, cell line and dose), we selected the one with highest `distil_cc_q75` value (75th quantile of pairwise Spearman correlations in landmark genes, <https://clue.io/connectopedia/glossary>). We performed gene set enrichment analysis<sup>61</sup> to evaluate the enrichment of disease genes among the top deregulated genes in the perturbation profiles. This analysis offers enrichment scores that have small values when genes are randomly distributed among the ordered list of expression values and high values when they are concentrated at the top or bottom of the list. The enrichment score significance is calculated by creating 1,000 random selections of gene sets with the same size as the original set and calculating an empirical  $P$  value by considering the proportion of random sets resulting in enrichment scores smaller than the original case. The  $P$  values were adjusted for multiple testing using the Benjamini–Hochberg method. The network proximity  $d_c$  of disease proteins and polyphenol targets for diseases with significant enrichment scores were compared according to their therapeutic and unknown-therapeutic associations using the Student's  $t$ -test. The relevant code for calculating the network proximity, AUCs and enrichment scores can be found at <https://github.com/italodovale/polyphenols>.

**Platelet isolation.** Human blood collection was performed as previously described in accordance with the Declaration of Helsinki and ethics regulations with Institutional Review Board approval from Brigham and Women's Hospital (P001526). Healthy volunteers did not ingest known platelet inhibitors for at least 10 days prior. Citrated whole blood underwent centrifugation with a slow brake (177 × g, 20 minutes), and the PRP fraction was acquired for subsequent experiments. For washed platelets, PRP was incubated with 1 μM prostaglandin E<sub>1</sub> (Sigma, P5515) and immediately underwent centrifugation with a slow brake (1,000 × g, 5 minutes). Platelet-poor plasma was aspirated, and pellets were resuspended in platelet resuspension buffer (PRB; 10 mM HEPES, 140 mM NaCl, 3 mM KCl, 0.5 mM MgCl<sub>2</sub>, 5 mM NaHCO<sub>3</sub>, 10 mM glucose, pH 7.4).

**Platelet aggregometry.** Platelet aggregation was measured by turbidimetric aggregometry as previously described<sup>62</sup>. Briefly, PRP was pretreated with RA for 1 hour before adding 250 μl to siliconized glass cuvettes containing magnetic stir bars. Samples were placed in Chrono-Log Model 700 Aggregometers before the addition of various platelet agonists. Platelet aggregation was monitored for 6 minutes at 37 °C with a stir speed of 1,000 r.p.m. and the maximum extend of aggregation recorded using AGGRO/LINK8 software. In some cases, dense granule release was simultaneously recorded by supplementing samples with Chrono-Lume (Chrono-Log, 395) according to the manufacturer's instructions.

**Platelet α-granule secretion and integrin α<sub>IIb</sub>β<sub>3</sub> activation.** Changes in platelet surface expression of P-selectin (CD62P) or binding of Alexa Fluor 488-conjugated fibrinogen were used to assess α-granule secretion and integrin α<sub>IIb</sub>β<sub>3</sub> activation, respectively. First, PRP was preincubated with RA for 1 hour, followed by stimulation with various platelet agonists under static conditions at 37 °C for 20 minutes. Samples were then incubated with APC-conjugated anti-human CD62P antibodies (BioLegend, 304910) and 100 μg ml<sup>-1</sup> Alexa Fluor 488-Fibrinogen (Thermo Scientific, F13191) for 20 minutes before fixation in 2% (v/v) paraformaldehyde (Thermo Scientific, AAJ19945K2). For each sample,

50,000 platelets were processed using a Cytex Aurora spectral flow cytometer. Percent positive cells were determined by gating on fluorescence intensity compared to unstimulated samples.

**Platelet cytotoxicity.** Cytotoxicity was tested by measuring lactate dehydrogenase (LDH) release by permeabilized platelets into the supernatant<sup>63</sup>. Briefly, washed platelets were treated with various concentrations of RA for 1 hour, before isolating supernatants via centrifugation (15,000×g, 5 min). A Pierce LDH Activity Kit (Thermo Scientific, 88953) was then used to assess supernatant levels of LDH.

**Immunoprecipitation and western blot.** Washed platelets were pretreated with RA for 1 hour, followed by a 15 minute treatment with Eptifibatid (50 μM). Platelets were then stimulated with various agonists for 5 minutes under stirring conditions (1,000 r.p.m., 37°C). Platelets were lysed on ice with RIPA Lysis Buffer System (Santa Cruz, sc-24948) and supernatants clarified via centrifugation (15,000×g, 10 min, 4°C). For immunoprecipitation of FYN, lysates were first precleared of IgG by incubating with Protein A agarose beads (Cell Signaling Technologies, 9863S) for 30 minutes at 4°C, before isolation of the supernatant via centrifugation (15,000×g, 10 min, 4°C). Supernatants were incubated with anti-FYN antibodies (Abcam, 2A10) overnight at 4°C before incubation with Protein A beads for 1 hour. Beads were then washed five times with NP-40 lysis buffer (144 mM Tris, 518 mM NaCl, 6 mM EDTA, 12 mM Na<sub>2</sub>VO<sub>3</sub>, 33.3% (v/v) NP-40, Halt protease inhibitor cocktail (Thermo, 78429)).

For western blot analysis, total cell lysates or immunoprecipitated FYN were reduced with Laemmli Sample Buffer (Bio-Rad, 1610737) and proteins separated by molecular weight in PROTEAN TGX precast gels (Bio-Rad, 4561084). Proteins were transferred to PVDF membranes (Bio-Rad, 1620174) and probed with either 4G10 (Milipore, 05-321), a primary antibody clone that recognizes phosphorylated tyrosine residues, or primary antibodies that probe for the site-specific phosphorylation of SFKs (p-Tyr416) within their activation loop. Membranes were incubated with horseradish peroxidase-conjugated secondary antibodies (Cell Signaling Technologies, 7074S) to catalyse an electrochemiluminescent reaction (Thermo Scientific, PI32109). Membranes were visualized using a Bio-Rad ChemiDoc Imaging System, and densitometric analysis of protein lanes was conducted using ImageJ (NIH, Version 1.52a).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data supporting the findings of this study are available at <https://github.com/italodovale/polyphenols> and within the paper and its Supplementary Information files.

## Code availability

Computer code is available at <https://github.com/italodovale/polyphenols>.

Received: 27 August 2020; Accepted: 16 February 2021;

Published online: 19 March 2021

## References

- Khera, A. V. et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).
- Arts, I. C. W. & Hollman, P. C. H. Polyphenols and disease risk in epidemiologic studies. *Am. J. Clin. Nutr.* **81**, 317S–325S (2005).
- Wang, X., Ouyang, Y. Y., Liu, J. & Zhao, G. Flavonoid intake and risk of CVD: a systematic review and meta-analysis of prospective cohort studies. *Br. J. Nutr.* **111**, 1–11 (2014).
- Neveu, V. et al. Phenol-Explorer: an online comprehensive database on polyphenol contents in foods. *Database* **2010**, bap024 (2010).
- Pérez-Jiménez, J., Neveu, V., Vos, F. & Scalbert, A. Systematic analysis of the content of 502 polyphenols in 452 foods and beverages: an application of the phenol-explorer database. *J. Agric. Food Chem.* **58**, 4959–4969 (2010).
- Zhang, H. & Tsao, R. Dietary polyphenols, oxidative stress and antioxidant and anti-inflammatory effects. *Curr. Opin. Food Sci.* **8**, 33–42 (2016).
- Boly, R. et al. Quercetin inhibits a large panel of kinases implicated in cancer cell biology. *Int. J. Oncol.* **38**, 833–842 (2011).
- Lacroix, S. et al. A computationally driven analysis of the polyphenol-protein interactome. *Sci. Rep.* **8**, 2232 (2018).
- Hanhineva, K. et al. Impact of dietary polyphenols on carbohydrate metabolism. *Int. J. Mol. Sci.* **11**, 1365–1402 (2010).
- Hervet-Hernández, D. & Goñi, I. Dietary polyphenols and human gut microbiota: a review. *Food Rev. Int.* **27**, 154–169 (2011).
- Zhang, S. et al. Dietary pomegranate extract and inulin affect gut microbiome differentially in mice fed an obesogenic diet. *Anaerobe* **48**, 184–193 (2017).
- Thazhath, S. S. et al. Administration of resveratrol for 5 wk has no effect on glucagon-like peptide 1 secretion, gastric emptying, or glycemic control in type 2 diabetes: a randomized controlled trial. *Am. J. Clin. Nutr.* **103**, 66–70 (2016).
- Bhatt, J. K., Thomas, S. & Nanjan, M. J. Resveratrol supplementation improves glycemic control in type 2 diabetes mellitus. *Nutr. Res.* **32**, 537–541 (2012).
- Sharma, A. et al. A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum. Mol. Genet.* **24**, 3005–3020 (2014).
- Menche, J. et al. Disease networks. Uncovering disease–disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).
- Guney, E., Menche, J., Vidal, M. & Barabási, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 10331 (2016).
- Cheng, F. et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat. Commun.* **9**, 2691 (2018).
- Kovács, I. A. et al. Network-based prediction of protein interactions. *Nat. Commun.* **10**, 1240 (2019).
- Sarkar, F. H., Li, Y., Wang, Z. & Kong, D. Cellular signaling perturbation by natural products. *Cell. Signal.* **21**, 1541–1547 (2009).
- Iso, H. et al. The relationship between green tea and total caffeine intake and risk for self-reported type 2 diabetes among Japanese adults. *Ann. Intern. Med.* **144**, 554–562 (2006).
- Song, Y., Manson, J. E., Buring, J. E., Sesso, H. D. & Liu, S. Associations of dietary flavonoids with risk of type 2 diabetes, and markers of insulin resistance and systemic inflammation in women: a prospective study and cross-sectional analysis. *J. Am. Coll. Nutr.* **24**, 376–384 (2005).
- Keske, M. A. et al. Vascular and metabolic actions of the green tea polyphenol epigallocatechin gallate. *Curr. Med. Chem.* **22**, 59–69 (2015).
- Wolfram, S. et al. Epigallocatechin gallate supplementation alleviates diabetes in rodents. *J. Nutr.* **136**, 2512–2518 (2006).
- Davis, A. P. et al. The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).
- Muthu, R., Selvaraj, N. & Vaiyapuri, M. Anti-inflammatory and proapoptotic effects of umbelliferone in colon carcinogenesis. *Hum. Exp. Toxicol.* **35**, 1041–1054 (2016).
- Muthu, R. & Vaiyapuri, M. Synergistic and individual effects of umbelliferone with 5-fluorouracil on tumor markers and antioxidant status of rat treated with 1,2-dimethylhydrazine. *Biomed. Aging Pathol.* **3**, 219–227 (2013).
- Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e17 (2017).
- Grover, S. P., Bergmeier, W. & Mackman, N. Platelet signaling pathways and new inhibitors. *Arterioscler. Thromb. Vasc. Biol.* **38**, e28–e35 (2018).
- Moco, S., Martin, F. P. J. & Rezzi, S. Metabolomics view on gut microbiome modulation by polyphenol-rich foods. *J. Proteome Res.* **11**, 4781–4790 (2012).
- van Duynhoven, J. et al. Metabolic fate of polyphenols in the human superorganism. *Proc. Natl Acad. Sci. USA* **108**, 4531–4538 (2011).
- Ottaviani, J. I., Heiss, C., Spencer, J. P. E., Kelm, M. & Schroeter, H. Recommending flavanols and procyanidins for cardiovascular health: revisited. *Mol. Aspects Med.* **61**, 63–75 (2018).
- Stalmach, A., Troufflard, S., Serafini, M. & Crozier, A. Absorption, metabolism and excretion of Choladi green tea flavan-3-ols by humans. *Mol. Nutr. Food Res.* **53**, S44–S53 (2009).
- Meng, X. et al. Identification and characterization of methylated and ring-fission metabolites of tea catechins formed in humans, mice, and rats. *Chem. Res. Toxicol.* **15**, 1042–1050 (2002).
- Perez-Vizcaino, F., Duarte, J. & Santos-Buelga, C. The flavonoid paradox: conjugation and deconjugation as key steps for the biological activity of flavonoids. *J. Sci. Food Agric.* **92**, 1822–1825 (2012).
- Shimoi, K. & Nakayama, T. Glucuronidase deconjugation in inflammation. *Methods Enzymol.* **400**, 263–272 (2005).
- Kaneko, A. et al. Glucuronides of phytoestrogen flavonoid enhance macrophage function via conversion to aglycones by β-glucuronidase in macrophages. *Immun. Inflamm. Dis.* **5**, 265–279 (2017).
- Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 (2019).
- Smalley, J. L., Gant, T. W. & Zhang, S.-D. Application of connectivity mapping in predictive toxicology based on gene-expression similarity. *Toxicology* **268**, 143–146 (2010).
- Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
- Amanzadeh, E. et al. Quercetin conjugated with superparamagnetic iron oxide nanoparticles improves learning and memory better than free quercetin via interacting with proteins involved in LTP. *Sci. Rep.* **9**, 6876 (2019).
- Shaikh, J., Ankola, D. D., Beniwal, V., Singh, D. & Kumar, M. N. V. R. Nanoparticle encapsulation improves oral bioavailability of curcumin by at least 9-fold when compared to curcumin administered with piperine as absorption enhancer. *Eur. J. Pharm. Sci.* **37**, 223–230 (2009).
- Chao, E. C. & Henry, R. R. SGLT2 inhibition—A novel strategy for diabetes treatment. *Nat. Rev. Drug Discov.* **9**, 551–559 (2010).
- Caldera, M. et al. Mapping the perturbome network of cellular perturbations. *Nat. Commun.* **10**, 5140 (2019).

44. Jensen, K., Ni, Y., Panagiotou, G. & Kouskoumvekaki, I. Developing a molecular roadmap of drug–food interactions. *PLoS Comput. Biol.* **11**, e1004048 (2015).
45. Rolland, T. et al. A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014).
46. Cheng, F., Jia, P., Wang, Q. & Zhao, Z. Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. *Oncotarget* **5**, 3697–3710 (2014).
47. Calçada, D. et al. The role of low-grade inflammation and metabolic flexibility in aging and nutritional modulation thereof: a systems biology approach. *Mech. Ageing Dev.* <http://dx.doi.org/10.1016/j.mad.2014.01.004> (2014).
48. Hornbeck, P. V. et al. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
49. Li, T. et al. A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2016).
50. Chatri-Aryamontri, A. et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* **45**, D369–D379 (2017).
51. Cowley, M. J. et al. PINA v2.0: mining interactome modules. *Nucleic Acids Res.* **40**, D862–D865 (2012).
52. Peri, S. et al. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **32**, D497–D501 (2004).
53. Orchard, S. et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2014).
54. Breuer, K. et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **41**, D1228–D1233 (2013).
55. Meyer, M. J., Das, J., Wang, X. & Yu, H. INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics* **29**, 1577–1579 (2013).
56. Mosca, R., Céol, A. & Aloy, P. Interactome3D: adding structural details to protein networks. *Nat. Methods* **10**, 47–53 (2013).
57. Meyer, M. J. et al. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* **15**, 107–114 (2018).
58. Fazekas, D. et al. SignalLink 2 – a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.* <https://doi.org/10.1186/1752-0509-7-7> (2013).
59. Huttlin, E. L. et al. Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).
60. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
61. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
62. Roweth, H. G. et al. Two novel, putative mechanisms of action for citalopram-induced platelet inhibition. *Sci. Rep.* **8**, 16677 (2018).
63. Roweth, H. G. et al. Citalopram inhibits platelet function independently of SERT-mediated 5-HT transport. *Sci. Rep.* **8**, 3494 (2018).
64. Nath, S., Bachani, M., Harshavardhana, D. & Steiner, J. P. Catechins protect neurons against mitochondrial toxins and HIV proteins via activation of the BDNF pathway. *J. Neurovirol.* **18**, 445–455 (2012).
65. Park, K.-S. et al. (–)-Epigallocatechin-3-O-gallate counteracts caffeine-induced hyperactivity: evidence of dopaminergic blockade. *Behav. Pharmacol.* **21**, 572–575 (2010).
66. Ramesh, E., Geraldine, P. & Thomas, P. A. Regulatory effect of epigallocatechin gallate on the expression of C-reactive protein and other inflammatory markers in an experimental model of atherosclerosis. *Chem. Biol. Interact.* **183**, 125–132 (2010).
67. Han, S. G., Han, S.-S., Toborek, M. & Hennig, B. EGCG protects endothelial cells against PCB 126-induced inflammation through inhibition of AhR and induction of Nrf2-regulated genes. *Toxicol. Appl. Pharmacol.* **261**, 181–188 (2012).
68. Sheng, R., Gu, Z.-L. & Xie, M.-L. Epigallocatechin gallate, the major component of polyphenols in green tea, inhibits telomere attrition mediated cardiomyocyte apoptosis in cardiac hypertrophy. *Int. J. Cardiol.* **162**, 199–209 (2013).
69. Devika, P. T. & Stanely Mainzen Prince, P. (–)-Epigallocatechin gallate protects the mitochondria against the deleterious effects of lipids, calcium and adenosine triphosphate in isoproterenol induced myocardial infarcted male Wistar rats. *J. Appl. Toxicol.* **28**, 938–944 (2008).
70. Yi, Q.-Y. et al. Chronic infusion of epigallocatechin-3-O-gallate into the hypothalamic paraventricular nucleus attenuates hypertension and sympathoexcitation by restoring neurotransmitters and cytokines. *Toxicol. Lett.* **262**, 105–113 (2016).
71. Devika, P. T. & Prince, P. S. M. Preventive effect of (–)-epigallocatechin-gallate (EGCG) on lysosomal enzymes in heart and subcellular fractions in isoproterenol-induced myocardial infarcted Wistar rats. *Chem. Biol. Interact.* **172**, 245–252 (2008).
72. Hushmendy, S. et al. Select phytochemicals suppress human T-lymphocytes and mouse splenocytes suggesting their use in autoimmunity and transplantation. *Nutr. Res.* **29**, 568–578 (2009).
73. Shen, K. et al. Epigallocatechin 3-gallate ameliorates bile duct ligation induced liver injury in mice by modulation of mitochondrial oxidative stress and inflammation. *PLoS ONE* **10**, e0126278 (2015).
74. ZHEN, M. et al. Green tea polyphenol epigallocatechin-3-gallate inhibits oxidative damage and preventive effects on carbon tetrachloride-induced hepatic fibrosis. *J. Nutr. Biochem.* **18**, 795–805 (2007).
75. Yasuka, Y. et al. (–)-Epigallocatechin gallate prevents carbon tetrachloride-induced liver fibrosis by inhibiting the expression of the PDGFR $\beta$  and IGF-1R. *Chem. Biol. Interact.* **182**, 159–164 (2009).
76. Cao, W. et al. iTRAQ-based proteomic analysis of combination therapy with taurine, epigallocatechin gallate, and genistein on carbon tetrachloride-induced liver fibrosis in rats. *Toxicol. Lett.* **232**, 233–245 (2015).
77. Kitamura, M. et al. Epigallocatechin gallate suppresses peritoneal fibrosis in mice. *Chem. Biol. Interact.* **195**, 95–104 (2012).
78. Sakla, M. S. & Lorson, C. L. Induction of full-length survival motor neuron by polyphenol botanical compounds. *Hum. Genet.* **122**, 635–643 (2008).
79. Shimizu, M. et al. (–)-Epigallocatechin gallate inhibits growth and activation of the VEGF/VEGFR axis in human colorectal cancer cells. *Chem. Biol. Interact.* **185**, 247–252 (2010).

## Acknowledgements

This study was supported, in part, by NIH grants 1P01HL132825, HG007690, HL108630 and HL119145; American Heart Association grants 151708 and D700382 and ERC grant 810115-DYNASET. We would like to thank P. Ruppert, G. Menichetti and I. Kovacs for support in this study, F. Cheng for assembling the human interactome and A. Grishchenko for help with data visualization.

## Author contributions

I.F.d.V. and A.-L.B. designed the study. I.F.d.V. performed all computational analyses. H.G.R., M.W.M., E.B. and J.L. designed and performed experimental validation. J.L. guided I.F.d.V. in validation case studies. S.M. and D.B. guided I.F.d.V. in data interpretation and curation of disease associations obtained from the literature. I.F.d.V. and A.-L.B. wrote the paper with input from all authors. All authors read and approved the manuscript.

## Competing interests

J.L. and A.-L.B. are co-scientific founders of Scipher Medicine, Inc., which applies network medicine strategies to biomarker development and personalized drug selection; A.-L.B. is the founder of Datapolis Inc., which explores mobility patterns in urban planning, and Foodome, Inc., which applies data science to health. I.F.d.V. is a scientific consultant for Foodome, Inc.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s43016-021-00243-7>.

**Correspondence and requests for materials** should be addressed to A.-L.B.

**Peer review information** *Nature Food* thanks Dariush Mozaffarian and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples included blood donations from individual, healthy donors, on separate days. Each such blood donation was considered a sample. Sample sizes ranged from 3-6 blood donations, from 3-6 separate healthy donors.
Data exclusions	No data was excluded from analysis.
Replication	Experimental repeats were conducted as biological replicates, where the same test was repeated using isolated platelets from the blood of individual, healthy donors, on separate days. Experiments were repeated 3-6 times, from 3-6 separate blood donors.
Randomization	Samples were not randomized. This was due to the nature of our in vitro experiments, where platelets isolated from healthy donors were treated with either rosmarinic acid or a vehicle control. Therefore, any donor-specific or temporal effects were paired and accounted for.
Blinding	Samples were not blinded. This was due to the nature of our in vitro experiments, where platelets isolated from healthy donors were treated with either rosmarinic acid or a vehicle control. Given that our experimental workflow encompassed automated data acquisition, analysis and interpretation (flow cytometry, aggregometry, densitometry, colorimetry), we believe that blinding the operator to platelets treated with rosmarinic acid or control was not required.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	<ol style="list-style-type: none"> <li>anti-Fyn (Abcam, 2A10)</li> <li>4G10 (Milipore, 05-321)</li> <li>anti-Phospho-Src Family (Tyr416) (Cell Signaling Technology, #2101)</li> </ol>
Validation	<ol style="list-style-type: none"> <li>anti-Fyn            Suitable for: Western blot (WB), enzyme-linked immunosorbent assay (ELISA), Immunohistochemistry-paraffin (IHC-P), Flow Cytometry, Immunocytochemistry (ICC), Immunofluorescence (IF).            Previously validated in: Human breast tissue, human brain tissue, U251 cells and HEK293 cells.            References: Yamada Y et al. Identification of hypo- and hypermethylated genes related to atherosclerosis by a genome-wide analysis of DNA methylation. <i>Int J Mol Med</i> 33:1355-63 (2014).            Product link: <a href="https://www.abcam.com/fyn-antibody-2a10-ab119855.html?productWallTab=ShowAll">https://www.abcam.com/fyn-antibody-2a10-ab119855.html?productWallTab=ShowAll</a> </li> <li>Anti-Phosphotyrosine Antibody, clone 4G10®            Suitable for: WB            Previously validated in: Roweth, H.G., Yan, R., Bedwani, N.H. et al. Citalopram inhibits platelet function independently of SERT-mediated 5-HT transport. <i>Sci Rep</i> 8, 3494 (2018).            Product link: <a href="https://www.emdmillipore.com/US/en/product/Anti-Phosphotyrosine-Antibody-">https://www.emdmillipore.com/US/en/product/Anti-Phosphotyrosine-Antibody-</a> </li> </ol>

clone-4G10,MM\_NF-05-321#documentation

3. anti-Phospho-Src Family (Tyr416)

Suitable for: WB

Previously validated in Roweth, H.G., Yan, R., Bedwani, N.H. et al. Citalopram inhibits platelet function independently of SERT-mediated 5-HT transport. *Sci Rep* 8, 3494 (2018).

Product link: <https://www.cellsignal.com/products/primary-antibodies/phospho-src-family-tyr416-antibody/2101>

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	N/A
Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	N/A

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Healthy male and females ranging in age from 21-60 volunteered freely to participate in the study
Recruitment	The study involved donation of a blood sample and was performed in accordance with an IRB approved protocol (Protocol #: 2012P001526)
Ethics oversight	Institutional Review Board approval from Brigham and Women's Hospital (P001526)

Note that full information on the approval of the study protocol must also be provided in the manuscript.