Network machine learning maps phytochemically-rich "Hyperfoods" to fight COVID-19

Ivan Laponogov
  Imperial College London

Guadalupe Gonzalez
  Imperial College London

Madelen Shepherd
  Imperial College London

Ahad Qureshi
  Imperial College London

Dennis Veselkov
  Imperial College London

Georgia Charkoftaki
  Yale University School of Public Health

Vasilis Vasiliou
  Yale University School of Public Health   https://orcid.org/0000-0003-2552-3118

Jozef Youssef
  Kitchen Theory

Reza Mirnezami
  Royal Free Hospital

Michael Bronstein
  Imperial College London

Kirill Veselkov (✉ kirill.veselkov04@imperial.ac.uk)
  Imperial College London   https://orcid.org/0000-0002-1164-0359

Network machine learning maps phytochemically-rich "Hyperfoods" to fight COVID-19

Ivan Laponogov[1+] and Guadalupe Gonzalez[2+], Madelen Shepherd[1], Ahad Qureshi[1], Dennis Veselkov[2,3], Georgia Charkoftaki[4], Vasilis Vasiliou[4], Jozef Youssef[5], Reza Mirnezami[6], Michael Bronstein[2,7,8], Kirill Veselkov[1,4*]

[1]Department of Surgery and Cancer, Faculty of Medicine, Imperial College, London SW7 2AZ, UK; [2]Department of Computing, Faculty of Engineering, Imperial College, London SW7 2AZ, UK; [3]Intelligify Limited, 160 Kemp House, City Road, London, United Kingdom, EC1V 2NX; [4]Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA; [5]Kitchen Theory, London EN5 4LG, UK; [6]Department of Colorectal Surgery, Royal Free Hospital, Hampstead, London NW3 2QG, UK; [7]Twitter, UK; [8]University of Lugano, Switzerland.

[+]Equal contribution

[*]To whom correspondence should be addressed. Email:
kirill.veselkov04@imperial.ac.uk

ABSTRACT

In this paper, we introduce a network machine learning method to identify potential bioactive anti-COVID-19 molecules in foods based on their capacity to target the SARS-CoV-2-host gene-gene (protein-protein) interactome. Our experiments were run using a supercomputing DreamLab App platform, harnessing the idle computational power of thousands of smartphones. We first demonstrate that the proposed method can predict anti-COVID-19 candidates among experimental and clinically approved drugs (5658 in total) targeting COVID-19 interactomics with the balanced classification accuracy of 80-85% in 5-fold cross-validated settings. This identified the most promising drug candidates that can be potentially "repurposed" against COVID-19 including common drugs used to combat cardiovascular and metabolic disorders, such as simvastatin, atorvastatin and metformin. A database of 7694 bioactive food-based molecules was run through the calibrated machine-learning algorithm, which identified 52 biologically active molecules, from varied chemical classes, including flavonoids, terpenoids, coumarins and indoles predicted to target SARS-CoV-2-host interactome networks. This in turn was used to construct a "food map" with the theoretical anti-COVID-19 potential of each ingredient estimated based on the diversity and relative levels of candidate compounds with antiviral properties. We expect this *in-silico* predicted food map to play an important role in future clinical studies of precision nutrition interventions against COVID-19 and other viral diseases.

Keywords: machine learning, anti-viral, COVID-19, SARS-CoV-2, drug repositioning, food, interactomics, gene-gene networks

46 **BACKGROUND**

47

48 The rapid and continued spread of severe acute respiratory syndrome coronavirus 2

49 (SARS-CoV-2) is resulting in persistent outbreaks of novel coronavirus disease 2019

50 (COVID-19) across the world [1]. This in turn is having damaging effects on global

51 economies and healthcare systems, wellbeing, mental health and on societal

52 dynamics, as a whole. In the absence of effective curative treatments and validated

53 vaccines, there is an urgent need for innovative solutions. Combining conventional

54 medical treatments with nutritional interventions represents one such solution,

55 which is gaining traction [2, 3]. Considerable recent efforts have been directed

56 towards identifying new purposes, or alternative uses, for existing drugs (so-called

57 "drug repurposing") [4, 5]. This offers an attractive way to circumvent the slow and

58 costly pathway to new drug development and regulatory approval. Several

59 examples of repurposed drugs have been tested or are currently being tested in

60 clinical trials for deployment against COVID-19 [6]. In particular, the randomised

61 controlled trials of the corticosteroid dexamethasone have confirmed its capacity to

62 reduce mortality by up to a third in COVID-19 patients admitted to hospital for

63 respiratory support [7]. However, there are no clinically approved drugs or other

64 antiviral therapeutics for COVID-19 prevention, or for the treatment of non-

65 hospitalized symptomatic patients. These patients are typically discharged home

66 with basic advice, but remain at risk of personal clinical deterioration (especially

67 those with underlying comorbidities) and also pose an ongoing risk to close

68 contacts.

69

70   The human diet is rich with molecules that have been shown to play a role in both

71   the prevention and treatment of viral diseases, by interacting with drugs to

72   enhance their potency, or by acting as "medicines" themselves [8]. Of particular

73   relevance are plant-based foods which possess a complex profile of molecules of

74   varied chemical classes such as alkaloids, flavonoids, coumarins, terpenoids, and

75   indoles [9]. Laboratory studies have revealed multiple mechanisms of action by

76   which these dietary compounds exert their action against functionally and

77   genetically diverse viruses [10]. These include disruption of biochemical pathways

78   and gene networks involved in viral entry, replication, spread and shedding, as well

79   as broader anti-inflammatory and antioxidant properties. These molecules are not

80   generally monitored by national nutritional agencies such as the United States

81   Department of Agriculture (USDA) and can be considered the "dark matter" of

82   nutrition beyond currently traced micronutrients (i.e. vitamins, minerals) and

83   macronutrients (i.e. carbohydrates, proteins, fibre and fats) [11]. Furthermore,

84   there is a growing body of evidence that poor dietary habits and diet-related

85   comorbidities such as obesity, diabetes, and cardiovascular disease are at least

86   partially responsible for disparities in adverse outcomes from COVID-19 across the

87   globe [12, 13]. One possible explanation for this could be poor gut microbiome

88   health and pre-existing pro-inflammatory state leading to a dysregulated cytokine

89   storm among vulnerable COVID-19 patients that is associated with the high

90   mortality of such cases [14].

91

92   Identification of dietary constituents and consequent design of phytochemically rich

93   "Hyperfoods" with disease-beating properties can be a safe and cost-effective

94   method for developing tailored nutrition-based therapeutic strategies against many

95   diseases, including COVID-19 [15]. However, it is vitally important to appreciate

96   that the modern era of molecular gastronomy has resulted in a growing expectation

97   for food to fulfil taste, aesthetic, sensory and health-centered requirements. For

98   these reasons, the design of such "Hyperfoods" requires multi-faceted optimisation,

99   taking into account not only pro-health benefits but also considering visual

100   aesthetics (e.g. colour, texture) and sensory (e.g. taste mouthfeel) characteristics

101   [15]. At present, the landscape of potential drug-like molecules in food is

102   unimaginably vast. Thanks to advances in high-throughput mass spectrometry

103   technologies and machine learning, identification and molecular networking of

104   thousands of these molecules from various food sources has become possible [16].

105   Investigating the influence of a single drug or food component on any particular

106   viral infection takes months to years of experimental research. Examples of

107   experimentally derived phytochemicals with antiviral properties include hesperedin

108   and naringin in citrus foods, tannic acid in black tea, emodin in rhubarb, and

109   myristicin in dill and parsley [17, 18]. Given the vast molecular space, the

110   traditional practicalities of investigating the influence of a single molecule or food

111   component would take far too long to have an impact on the current COVID-19

112   crisis.

113

114   Coronaviruses cannot survive or replicate without host assistance. In fact, all

115   viruses have naturally evolved a sophisticated array of molecular strategies

116   designed to exploit the host's cellular machinery to promote viral survival and

117   replication. These strategies rely on a complex network of physical interactions

118 between viral and host genes and proteins (so-called 'virus-host interactome

119 networks', here and further due to the specifics of the existing interaction datasets,

120 "gene" and "protein" terms can be used interchangeably) [19]. The conventional

121 antiviral drug development paradigm assumes that one drug targets one viral

122 protein [20]. In this regard, molecular docking computational simulations have

123 been extensively performed to discover plant-based bioactive molecules for specific

124 SARS-CoV-2 protein targets [21]. This approach has multiple drawbacks among

125 which is the robustness of complex virus-host interaction networks to individual

126 protein perturbations. The putative effects of vaccines and drugs against SARS-

127 CoV-2 specific gene or protein targets can also be complicated by escaped viral

128 mutants [22].

129

130 Here, we hypothesise that an effective anti-COVID-19 preventative or therapeutic

131 intervention should target multiple biochemical networks implicated in virus entry

132 and pathogenesis such as Angiotensin-Converting Enzyme-2 (ACE2)/G protein Mas

133 receptor (MasR) axis, Mitogen-Activated Protein Kinase (MAPK) cascade, and toll-

134 like receptor signalling pathways [23]. Building on our previous work on cancer-

135 beating molecules from food sources [15] and other recent network medicine

136 studies for computational drug repurposing against COVID-19 [24], we have

137 combined network-based machine learning methods, mobile supercomputing, and

138 interactomics data to identify food-based bioactive molecules targeting SARS-CoV-

139 2-human interactome networks. The discovered molecules/sources were used to

140 compile a list of antiviral "Hyperfoods" weighted by the highest diversity and levels

141 of antiviral molecules against SARS-CoV-2-human interactome networks. We

142   envision that the list of phytochemically rich "Hyperfoods" revealed in this work will

143   serve as a fundamental pillar in the design of a precision nutrition intervention

144   strategy for improved clinical outcomes of COVID-19. Further clinical validations of

145   our findings are needed in a randomised double-blind placebo-controlled trial on

146   clinical outcomes among non-hospitalised individuals with COVID-19 and their

147   households.

148

149   **RESULTS AND DISCUSSION**

150

151   **Genome-wide network-based machine learning for predicting drug and**

152   **food molecules targeting SARS-CoV-2-host interactome**

153

154   We have used the random walk propagation algorithm to learn the effects of SARS-

155   CoV-2 on human interactome networks governing regulatory and biochemical

156   pathways. The SARS-CoV-2 virus exploits human biomolecular network machinery

157   to promote viral entry, survival, replication, spread and shedding. The propagated

158   SARS-CoV-2-host interactome profile was subjected to the Gene Set Enrichment

159   Analysis (GSEA), which highlighted multiple potential mechanisms by which the

160   coronavirus exerts its activity on the host (Additional file 1). These include

161   membrane surface proteins (ACE2), regulation of programmed cell death pathways

162   (caspase 8 and p38/MAPK signalling), genomic replication pathways (RNA

163   polymerase pathways), immune-modulatory signalling circuits (toll-like receptors,

164   the Nuclear Factor-kB (NF-kB), JAK/STAT signalling pathways) and inflammatory

165   axes (e.g. interleukin pathways; see GSEA pathway analysis for additional details).

166

167   Drug and food molecules were ranked based on their potential interaction capability

168   with COVID-19, which in turn has been derived from their respective effects on the

169   human protein-protein (or gene-gene) interaction network, commonly referred to

170   as the interactome. The main assumption here is that for a given molecule to have

171   an effect against coronavirus, it should target the same pathways and cellular

172   mechanisms targeted by the disease but with the opposite regulatory effect. This

173   action does not necessarily imply that gene/protein targets have direct effect, and

174   the effect can be indirectly exerted through other neighbouring proteins in the

175   network, via gene-gene (protein-protein) interaction. This approach permits

176   modeling the systemic genome-wide response to the disease and drug/food

177   intervention and identifying drug/food-based compounds with the highest

178   probability of being effective against COVID-19 (see Figure 1). Similar network

179   propagation approaches have been applied in cancer research for drug repurposing

180   [24], mutation-driven population stratification [25], and, in our earlier work, for

181   drug repurposing and food-based anti-cancer molecular therapeutics [15]. Although

182   there are other approaches being developed for drug repurposing using multi-omics

183   and phenotypic data [26], these mandate additional datasets that are usually not

184   available for food-based molecules.

185

186   The machine learning algorithm hyperparameters were calibrated for predicting

187   experimentally validated drugs against COVID-19 in a cross-validation setting (see

188   the Method section on parameter optimization, accuracy estimation, and results

189   aggregation). The optimal balanced classification accuracy in the range of 80-

190 84.9% was achieved using an ensemble of parameter settings (3609 models for

191 aggregated interactome (see Additional file 2) and 15 models for manually curated

192 interactome derived from a biological pathway database of COVID-19 WikiPathways

193 (see Additional file 3). Practically, this resulted in approximately 8 out of 10 drugs

194 being correctly classified into their respective classes (i.e potentially anti-COVID-19

195 vs others). For each parameter combination achieving balanced accuracy above

196 80%, a ranked list of compounds (drugs and food molecules) was generated with

197 compounds ranked by the decreasing correlation between compound and disease

198 profiles. The consensus list of top ranked compounds with the highest antiviral

199 ranking and probability is summarised in Additional file 4. For each candidate

200 molecule, we also provided a putative mechanism of action and literature reference

201 where available.

202

203 **Drug repositioning candidates against COVID-19**

204

205 Our analysis identified imiquimod as the top ranked drug with anti-COVID-19

206 potential. Imiquimod acts as an agonist of toll-like receptor 7, which is crucial in

207 recognising single-stranded RNA viruses, such as SARS-CoV-2. Toll-like receptors

208 generate anti-viral immunity and act to induce favourable Type I interferon

209 response, which in turn induces the expression of interferon stimulated genes

210 leading to the inhibition of viral replication [27].

211

212 Several widely used chemotherapeutic agents were found to exert potential anti-

213 COVID-19 effect, including doxorubicin, fluorouracil, and gemcitabine. Doxorubicin

214 is commonly used in the treatment of advanced breast cancer, bladder cancer, and
215 lymphoma, as well as a number of other malignancies. A previous study has
216 indicated that SARS-CoV-2 contains residues that are vulnerable to the reactive
217 glycating agent methylglyoxal, cellular levels of which are increased by doxorubicin
218 [28]. Fluorouracil is a fluoropyrimidine used for the treatment of a number of solid
219 organ tumours. It is a precursor of deoxythymidine triphosphate and uridine-5'-
220 triphosphate (UTP) during biogenesis and interferes with both DNA and RNA
221 metabolism. This drug is preferentially incorporated into RNA instead of UTP, which
222 interferes with RNA processing and protein synthesis and this in turn can lead to
223 the disruption of viral RNA replication and elicit an antiviral effect [29]. Gemcitabine
224 has also been shown to inhibit SARS-CoV-2 replication. It is hypothesised that this
225 effect occurs through targeting of pyrimidine biosynthesis salvage pathways and
226 stimulation of the innate immune system [30]. Although chemotherapy and other
227 anti-cancer treatments may result in significant immune compromise in patients,
228 rendering them more susceptible to viral and other infectious illnesses [31], the
229 findings presented here also highlight a *double-edged* phenomenon, whereby they
230 may actually exert potential beneficial effects against COVID-19 infection.

231

232 Statins are considered a clinically important breakthrough in the prevention and
233 treatment of cardiovascular disease. Simvastatin and atorvastatin were found to
234 offer significant anti-COVID-19 potential. The hypothesis is that statins in general
235 reduce COVID-19 infectivity through the removal of cholesterol used by SARS-CoV-
236 2 to infect cells [32], and reduce risk of cardiovascular complications that are
237 symptomatic of severe COVID-19 infection. In addition, they may enhance innate

238   immune responses to viral infections through inhibition of the Myeloid

239   differentiation primary response 88 signalling pathway. Correspondingly, a recent

240   meta-analysis of data from multiple studies reported a 30% reduction in fatal or

241   severe disease course in patients with confirmed COVID-19 infection who were

242   taking statins [33].

243

244

245   Metformin is globally regarded as one of the key pharmacotherapies in the

246   management of diabetes mellitus. Of note, it was originally introduced as an anti-

247   influenza drug, with glucose-lowering capability regarded as a side-effect of

248   treatment, rather than desired primary endpoint. The many pleiotropic effects of

249   metformin together with its widespread utility in modern medicine have earned it

250   the name "the aspirin of the 21st century" [34]. It activates the AMP-activated

251   protein kinase, resulting in the phosphorylation of angiotensin converting enzyme II

252   (ACE2), which leads to conformational and functional changes to ACE2 that are

253   thought to inhibit SARS-CoV-2 binding and/or entry [35]. In support of these

254   suggestions, a recent meta-analysis demonstrated a reduced risk of mortality in

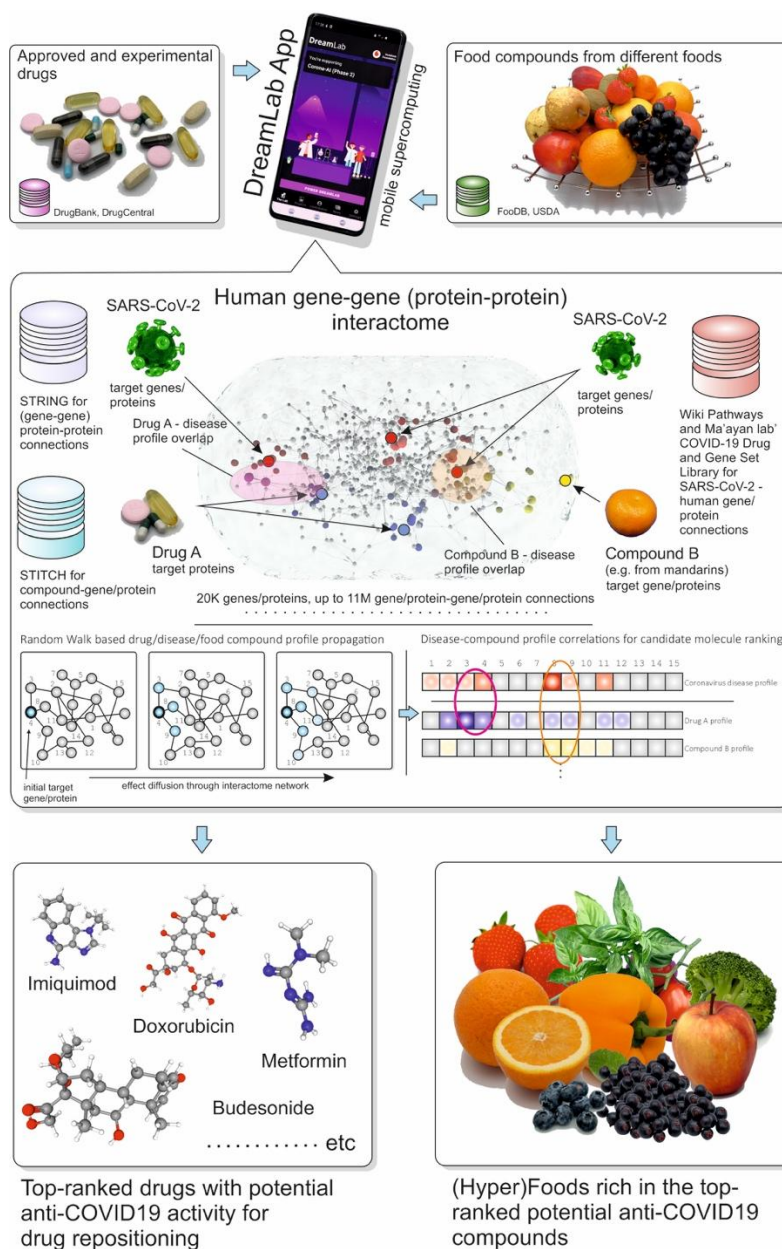255   COVID-19 patients receiving metformin [36].

256

257 **Figure 1**. Schematic diagram of overall workflow. The random walk with restarts algorithm operating

258 within a mobile supercomputing DreamLab App is used to simulate how drug and food-based compounds

259 interact with COVID-19 associated viral gene/protein networks. This has been extrapolated from human

260 genome-wide gene-gene (protein-protein) interactome data and based on known COVID-19 human

261 proteome viral targets (i.e. human genes/proteins interacting with different stages of the virus life cycle to

262 facilitate replication and/or enhance viral potency). Both disease and molecular compound impacts are

263 propagated through the interactome network to model the overall cellular response/interactome

264 perturbation. The resulting compound and disease profiles are then correlated to rank compounds

265 according to their network "overlap" with "reference" viral profiles. This approach is based on the

266 assumption that to have an effect, candidate compounds should target the same network component(s)

267 as the one(s) disrupted by the virus. Therapeutic effect can be direct, or indirect, for example where

268 compounds are found to interact with neighbouring network nodes, resulting in subsequent effect

269 propagation to the desired target.

270

271 **Prediction of "dark matter" of food biochemistry with anti-COVID-19**

272 **properties**

273

274 In addition to minerals, vitamins and micronutrients, all plant-based foods contain

275 phytochemicals that are non-nutritive components in the diet but can exert

276 protective or disease-beating effects. This phytochemistry has been exploited

277 extensively for the development of antiviral drugs with more acceptable side effect

278 profiles, compared to synthetically generated drugs [37]. The network-based

279 analysis presented here identified 52 food-based molecules based on their

280 capability to target SARS-CoV-2-host interactomes. These molecules belong to a

281 variety of chemical classes including (iso)flavonoids, terpenoids, phenols, and

282 indoles (see Figure 2). As highlighted, the presence and abundance of these

283 molecules are not typically monitored by national nutritional agencies, which

284 conventionally focus on minerals, vitamins and macronutrients. These compounds

285 can be regarded as the "dark matter" of nutritional science. Because of their bitter

286 taste, it is interesting to note that the food industry routinely removes some of

287 these compounds through selective breeding and a variety of debittering processes

288 to improve taste [38]. This has even led to the suggestion by some cancer research

289 groups that foods possessing more bitter taste may actually offer greater health

290 benefits [38].

291

292 The (poly)phenolic classes of molecules such as flavonoids, coumarins, stilbenes,

293 indoles, and phenolic acids make up the majority of anti-COVID-19 bioactive

294 compounds identified by our network-based machine learning algorithm. These

295 include flavonols (*e.g.* quercetin, kaempferol, and myricetin), flavones (*e.g.* luteolin

296 and apigenin), flavanols (*e.g.* procyanidin B2), flavanones (naringin), isoflavonoids

297 (daidzein, genistein, and legumelin) as well as stilbenes (trans-resveratrol), indoles

298 (3-indole-carbinol) and phenolic acids (gallic acid). In edible plants such as fruits

299 and vegetables, phenolic molecules are widespread and contribute to their aroma,

300 taste, and colour. These compounds are synthesized in abundance by plants in

301 response to environmental stimuli and play an indispensable role in defence against

302 pathogens (including viruses) and insects [39]. Their ability to disrupt the life cycle

303 of SARS-CoV-2 is partially achieved via interference with viral proteins. For

304 example, among our top-ranked molecules epigallocatechin 3-gallate was

305 demonstrated experimentally to inhibit 3-chymotrypsin-like protease (3CLpro) [38];

306 quercetin demonstrated binding affinity to inhibit 3CLpro and papain-like protease

307 (PLpro) [40], while trans-resveratrol inhibits nucleocapsid (N) proteins [18].

308

309 In addition, the identified compounds appear to mitigate against various patho-

310 physiological processes that develop in response to COVID-19. For example,

311 regulation of the renin-angiotensin system (RAS) and expression of angiotensin-

312 converting enzyme 2 (ACE2), stimulation of immune system, downregulation of

313 pro-inflammatory cytokine release, and amplification of cytotoxic T lymphocyte

314 (CTLs) and natural killer (NK) immune cell pools. The putative mechanism of action

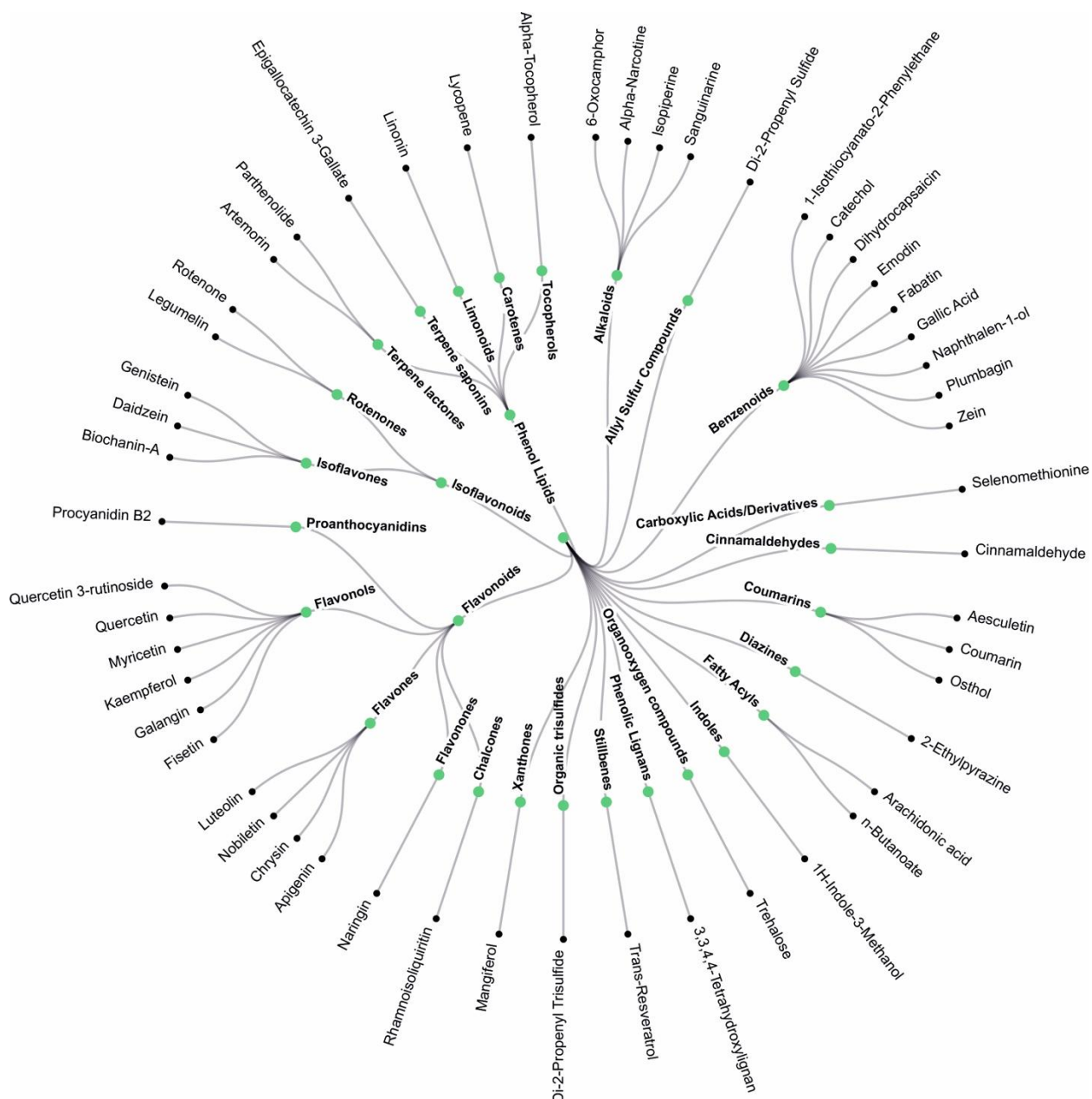315 for each of the identified compounds is summarised in Additional file 4.

316

317

318 **Figure 2**. Hierarchical classification of the top 52 predicted antiviral molecules targeting SARS-CoV-2
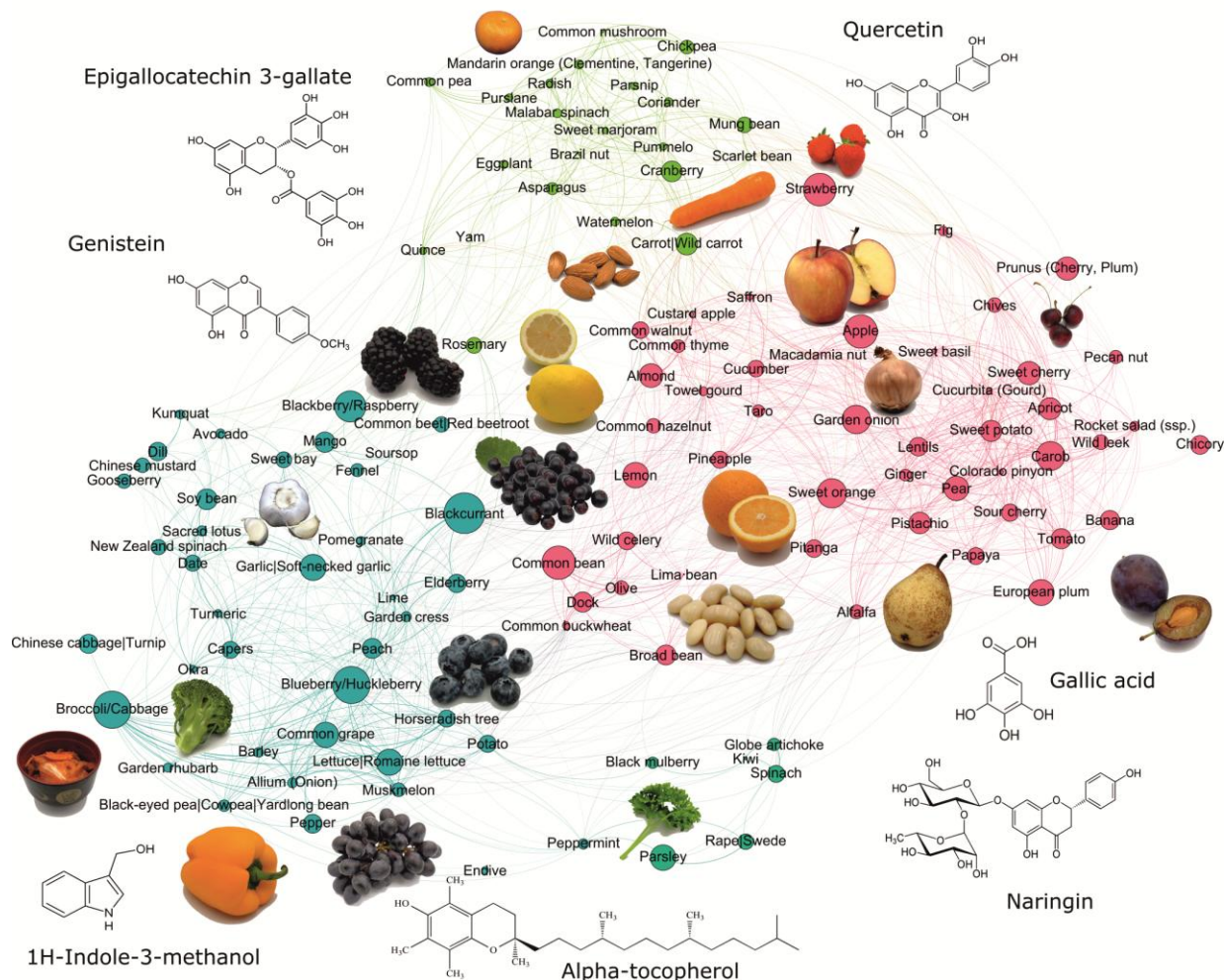
319 human interactome networks.

320

Construction of anti-COVID-19 Food Map

322

323 The potential for particular foods to exert COVID-19 preventative and/or

324 therapeutic effect depends upon the bioavailability and diversity of bioactive

325 molecules with antiviral properties contained therein [41]. A key limitation of the

326 existing literature on food-based compounds is the largely over-simplified view that

327 is commonly taken, whereby studies have tended to focus on specific molecular

328 components in isolation, for example specific flavonoids such as quercetin [42].

329 However, when candidate antiviral agents acting in isolation have been evaluated in

330 clinical studies, they have failed to consistently confer the same level of benefit

331 [43]. It seems more plausible that consumption of whole foods, with their

332 associated phytochemicals *en masse* may provide greater health benefits, due to

333 molecular additive or synergistic effects. It therefore follows that the antiviral

334 properties of a given food will be governed by two key factors: (1) the additive,

335 antagonistic, and synergistic actions of their individual components and (2) the way

336 in which these simultaneously modulate different intracellular pathways involved in

337 SARS-CoV-2 pathogenesis.

338

339 Based on these assumptions, we have constructed a food map with the theoretical

340 anti-COVID-19 capacity of each ingredient ranked according to an "enrichment

341 score" derived from the diversity and relative levels of candidate compounds with

342 antiviral properties (Figure 3). To identify putative mechanisms responsible for the

343 anti-COVID-19 properties of predicted foods, we have simulated the effects of a

344 phytochemical profile of a given food item on human interactome pathways and
345 sub-networks, using the random walker algorithm and gene set enrichment
346 analysis. The analysis showed that the most influential impacted pathways by
347 predicted phytochemically enriched foods with anti-COVID-19 properties exhibited a
348 statistically significant overlap with SARS-CoV-2 disrupted pathways (Additional file
349 5). This implies that a phytochemical profile of food ingredients, rather than
350 individual molecules, exert a combined effect across multiple host pathways
351 affected by SARS-CoV-2. (See the GSEA pathway analysis in Methods for more
352 detail).

353

354 The top ranked phytochemically rich food sources (called "Antiviral Hyperfoods")
355 include different berries (blackcurrant, cranberry, and blueberry), cruciferous
356 vegetables (cabbage, broccoli), apples, citrus fruits (sweet orange and lemon),
357 onions, garlic and beans. A recent study highlighted the potential of cabbage and
358 fermented vegetable consumption in minimising adverse outcomes in COVID-19,
359 supporting our results [44]. The present analysis has demonstrated that this is
360 potentially due to a profile of anti-COVID-19 compounds from various molecular
361 classes rather than individual molecules as previously suggested (see Additional file
362 5). Similarly, the complex antiviral molecular profile of berries such as blackcurrant
363 and blueberries may explain their experimentally observed potency against
364 genetically and phenotypically diverse viruses [45], though their ability to protect
365 specifically against COVID-19 is yet to be evaluated in clinical trials.

366

367

Figure 3. The contained profiles of compounds within specific foods, with predicted effectiveness in targeting SARS-CoV-2-host interactome networks. Each node in the figure denotes a particular food item and node size in each case is scaled by the derived enrichment score based on the diversity and relative levels of molecules with predicted anti-COVID-19 properties. The links between nodes reflect the pairwise correlation ("similarity") antiviral profiles in foods, thus the clusters of foods illustrate molecular commonality between them.

CONCLUSIONS

Current methods for prevention, treatment and containment of COVID-19 have not been effective in curbing the rate of transmission. Figures across the world show a sustained rise in cases. Non-hospitalised patients are discharged home where they continue to pose a risk to close contacts, and where they are at ongoing risk of clinical deterioration (especially those with comorbidities such as diabetes, obesity, and cardiovascular diseases). For this group of patients, there is a critical need for innovative and cost-effective out-of-hospital treatment. The use of precision nutrition strategies is safe and highly promising in this context. Using a network-based machine learning method, we have shown that certain plant-based foods such as berries, cruciferous vegetables, apples, citrus fruits, onions, garlic and beans are most enriched in terms of the diversity and relative abundance of bioactive molecules targeting the SARS-CoV-2-human interactome.

We acknowledge that the present work is subject to a number of limitations. Firstly, the cultivation, storage and cooking methods may influence bioactive molecular composition in foods. Secondly, it remains unclear whether these compounds would be present in sufficient levels to exert beneficial biological activity. Thirdly, the identified phenolic compounds can be filtered out by food producers because of their bitter taste to enhance palatability and taste experience. This raises interesting practical issues for "Hyperfoods" because increasing the content of bitter phytonutrients for health benefits may not be entirely compatible with consumer acceptance. Fourthly, the proposed methodology only accounts for interactions

402   between bioactive food compounds and SARS-CoV-2-human related molecular

403   networks, without necessarily defining the directionality of these relationships.

404   Fifthly, the methods described here do not take into account specific COVID-19

405   individual molecular phenotypic characteristics. Finally, drug-food interactions have

406   not been evaluated in this study; as such it is not clear whether these will lead to

407   synergistic or antagonistic effects where they act on common molecular networks,

408   or whether this combination will disrupt drug metabolism itself. Nevertheless, these

409   considerations notwithstanding, we expect this *in-silico* predicted food map to play

410   an important role in future clinical studies of precision nutrition interventions

411   against COVID-19. In the near future the goal will be to develop a personalised

412   'food passport' for each patient, designed to provide "smarter" food choices with

413   the ability to reduce susceptibility to COVID-19 infection and mitigate against

414   severe forms of the disease.

415

METHODS

416

Corona-AI/DreamLab mobile cloud supercomputing

417

418   The results presented in this manuscript were derived from the Corona-AI: Phase I

419   project for interactome driven drug and food compound search for the potential

420   anti-COVID-19 treatment. Working with Vodafone Foundation, the 'Corona-AI'

421   project used the freely available DreamLab app, which runs calculations using a

422   smartphone's computing power while its user sleeps. Tens to hundreds of

423   thousands smartphones combined together can be used to crunch scientific data at

424   scale rivalling available supercomputers and by far exceeding the capabilities of the

425   normal desktop PCs. The DreamLab App can be freely downloaded by anyone

426   willing to donate the unused computational power of their smartphone to cancer

427   and coronavirus research. The DreamLab App runs when a smartphone is being

428   charged: it loads a small portion of scientific data from the cloud, performs the

429   computations and sends the results back to the cloud for scientists to analyse them.

430   This way anyone can become a citizen scientist and contribute to global research.

431   **Learning propagation profiles of drugs, foods and diseases**

432   The main assumption of the methodology used in this paper was that the

433   drugs/food molecules which were effective at treatment of the particular disease

434   would have a similar pattern of affected genes/proteins to the pattern of the

435   genes/proteins affected by the disease. Due to gene-gene (or protein-protein)

436   interactions within the cell, disease and drug do not necessarily have to affect

437   exactly the same genes/proteins - their effects can be exerted on different, but

438   interlinked proteins and propagated through protein-protein networks. For that the

439   aggregated 20256 genes/proteins were represented as an array of floats where

440   each value represents how strongly the protein was affected/perturbed by the

441   disease or the drug (further referred to as drug or disease profiles). Zero value

442   would mean no effect or a normal unperturbed state.

443

444   Gene-gene (protein-protein) connections and drug-gene/protein connections were

445   filtered according to their confidence level (from STRING and STITCH databases)

446   before disease and drug profile generation and propagation (these thresholds were

447 among the adjustable parameters). Optionally, the number of drug-gene/protein

448 connections was also capped at specific value equalizing compounds with vastly

449 different numbers of known connections. If this option was used, the top N

450 connections are taken for each compound and the minimal connection score is

451 established from them. Then the connections are thresholded according to this

452 minimal connection score. This allowed one to include connections with the same

453 score as the last one in top N.

454

455 Genes/proteins directly affected by the drug or disease receive the initial value of

456 1.0 (or the score weight in case of "score_5_weighted" target protein selection).

457 Then the array of gene/protein perturbations was normalized to the sum of 1.0.

458

459 Random walk algorithm with restarts was used to propagate the perturbation of the

460 genes/proteins through the network. In brief, it models the probability of transition

461 from the starting node in the graph to another linked node based on the network

462 connectivity and the edge weights. By default, the graph was treated as undirected

463 and all edges had a weight of 1.0. Then outgoing connection weights were

464 normalized to the sum of weights to give the total probability of transition from the

465 node of 1.0 resulting in the transition probability matrix W. Parameter c (0.0-1.0)

466 controls the probability of "jumping" back to the original node. *I.e., c* = 0.0 means

467 unlimited wandering through the network resulting in a far-reaching propagation of

468 the initial perturbation and a very smooth profile, while c = 1.0 means a hard

469 restart to the initial position thus resulting in no signal being propagated beyond

470 the starting nodes (i.e. the nodes directly affected by the drug or the disease).

471 Random walk is an iterative algorithm and for each step the new perturbation

472 profile $\mathbf{p}_i$ is given by the following equation:

473

474 $$\mathbf{p}_i = \mathbf{p}_{i-1} * \boldsymbol{W} * (1.0 - c) + c * \mathbf{p}_0,$$

475

476 where $\mathbf{p}_0$ is the starting profile, $\mathbf{p}_{i-1}$ - previous iteration profile and $\boldsymbol{W}$ is the

477 transition probability matrix. Propagation algorithm was iterated until convergence

478 when $|\mathbf{p}_i - \mathbf{p}_{i-1}|$ is less than a set tolerance value for each element.

479

480 The random walk with restarts was applied to simulate the perturbations of direct

481 virus-host protein targets on the whole human interactome. It transforms a short

482 list of genes/proteins directly targeted by the virus into a genome-wide profile of

483 gene scores based on their network proximity to target candidates (referred as the

484 "SARS-CoV-2 genome-wide response profile"). The same random walker algorithm

485 was then used to get the activity profiles of candidate molecules, i.e. drugs or food-

486 based compounds. The Pearson correlation coefficient between propagation profiles

487 of food/drug compounds and COVID-19 disease was used to rank compounds that

488 target SARS-CoV-2-host interactome networks. The Pearson correlation coefficient

489 is defined as:

$$\boldsymbol{\rho}_{j,covid} = \frac{cov(\boldsymbol{p}_j, \boldsymbol{p}_{covid})}{\sigma_j \sigma_{covid}}$$

490

491 Where $cov()$ is covariance between molecule $\boldsymbol{p}_j$ and SARS-CoV-2 $\boldsymbol{p}_{covid}$ perturbation

492 profiles, σ_j and σ_{covid} are the standard deviations of perturbation profiles $\boldsymbol{p}_j$ of

493 molecule $j\square$ and $\boldsymbol{p}^{covid}$ perturbation profile of SARS-CoV-2.

494

495   The parameter settings for interactomes and diffusion processes for compound

496   ranking were optimized as described in the section of "parameter optimization,

497   accuracy estimation and results aggregation".

498   **Compound-protein and protein-protein interactome construction**

499   The interactome used in this study was constructed as was described previously in

500   [15]. In brief, a human genome was constructed from gene or protein sequences

501   from COSMIC [46], NCBI Gene [47], STRING [48] and UniProt [49] databases.

502   15,911 protein sequences matched exactly between databases, 1,532 protein

503   sequences were matched as subsections of larger sequences and 1,686 proteins

504   were matched allowing up to 5% amino acid mismatch. 1,127 mismatched

505   sequences were also included in the final unified set of 20256 gene encoded

506   proteins. The list of genes/proteins (these two terms are used interchangeably with

507   regard to the interactome analysis) was further populated with different gene IDs

508   and synonyms including Ensembl and HGNC. Protein-protein interactions were

509   obtained from STRING (~11M connections) and BioPlex (~100K connections)

510   databases [50] and supplemented with confidence scores (0-999) from STRING.

511

512   Drug-protein interactions were obtained from STITCH database [51], scored by the

513   confidence level of 0-999 for drugs from DrugBank [52] and DrugCentral [53]

514   databases as well as food molecules from FooDB [www.foodb.ca]. Indications for

515   drugs and FDA approval status were extracted from DrugCentral.

516

517   **Coronavirus target proteins aggregation:**

518   Two recent sources for the coronavirus-affected sets of human genes/proteins were

519   used in this work:

520   1) The COVID-19 Drug and Gene Set Library which provides a collection of drug

521   and gene sets related to COVID-19 research aggregated from multiple sources

522   using natural language processing techniques (downloaded on 24/09/2020) [54].

523   This set of genes/proteins is further referred to as the "Aggregated" set. In this set

524   human genes/proteins are scored by the number of times they have been reported

525   as related to COVID-19 with the top score of 88 assigned to STAT1 gene. We

526   generated several sub-selections of genes with different cut-offs for the scores: 40,

527   30, 25 and 20 (counting 72, 143, 248 and 457 genes) referred to as "score_40",

528   "score_30", "score_25" and "score_20" respectively. In these subselections the

529   genes are all initially equally weighted when propagated through interactome and

530   the chosen score threshold serves as an adjustable model parameter. We also

531   included a set of 5,000 top genes (with a minimal score of 5, referred to as

532   "score_5_weighted") with each gene weighted by its score for the propagation and

533   a minimal entry set (consisting of CTSB, CTSL, TMPRSS2 and ACE2 genes [55], first

534   reported in the literature as involved in the initial entry of the virus) referred to as

535   "entry_only".

536   2) COVID-19 Pathways Portal on WikiPathways [23] was used to create a subset of

537   423 coronavirus-affected human genes/proteins referred to as "score_wiki".

538

539

540

541 **Compound selection:**

542 SARS-CoV-2 is a relatively new pathogen and there is a very limited number of

543 experimentally validated drugs which were shown to be effective against it. We

544 manually curated a list of such compounds from literature search (see Additional

545 file 6). Special care was taken to include only compounds with experimental rather

546 than predicted evidence and explainable mechanism of action. Preference was given

547 to the compounds already included in clinical trials. This resulted in a list of 49

548 "positive" class compounds. Drugs were putatively classified into two subgroups -

549 acting through cellular mechanisms directly against the virus ("Direct-Cell") or

550 having symptomatic effects, e.g. anti-inflammatory ("Symptomatic"). This

551 classification is not strict as drugs may have overlapping functions. Finally, drugs

552 were checked for direct target overlaps with the "Aggregated" set of COVID-19-

553 related genes. Drugs with very few to no overlaps in the top 100 genes were

554 marked as less "reliable". This resulted in four sub selections for the "positive" class

555 compounds which were tried in the model parameter optimization stage: 1)

556 "Target_Cell": all drugs acting directly on the host-viral interactome, 27 in total; 2)

557 "Target_Cell_Strict": same as above, but only the most "reliable" drugs included,

558 19 in total; 3) "Target_Cell_Sympt": both symptomatic and host-viral interactome

559 targeting drugs included, 49 in total; 4) "Target_Cell_Sympt_Strict" - same as

560 above, but only the most "reliable" drugs included, 28 in total.

561

562 For the "negative" class both approved and experimental drugs from DrugBank

563 were selected. The antiviral drugs designed to target specific viral protein targets

564 (such as remdesivir, tenofovir and taribavirin) were designated as neutral ("0")

565 class and were excluded from the model calibration. This is because the primary

566 objective here is to target SARS-CoV-2 host interactome networks rather than

567 individual viral proteins.

568

569 All available compounds from DrugBank and FooDB which were not included in the

570 "positive" and "negative" classes were not used at the model calibration and

571 parameter optimization stage. 6593 compounds formed the input "negative" class,

572 however, depending upon the specific parametrization settings the final number of

573 negative class compounds varied between 1181 and 4260 due to the drugs with no

574 connections being automatically removed.

575

576 **Parameter optimization, accuracy estimation and results aggregation:**

577 Pearson correlation coefficients between each drug and disease propagated profiles

578 were calculated for the drugs/food molecules and for coronavirus affected gene

579 sets. Ranges of parameters such as restart probability c, drug-protein connection

580 thresholds and protein-protein interaction thresholds were explored to find the

581 optimal parameters for drug/food molecule ranking.

582

583 The following parameter ranges have been used: 1) Propagation parameter c for

584 compounds: 0.0001, 0.0002, 0.0004, 0.0007, 0.0009, 0.001, 0.005, 0.01, 0.02,

585 0.04, 0.07, 0.1, 0.2, 0.4, 0.7; 2) Propagation parameter c for coronavirus-host

586 profiles - 0.0001, 0.0002, 0.0004, 0.0007, 0.0009, 0.001, 0.005, 0.01, 0.02, 0.04,

587 0.07, 0.1, 0.2, 0.4, 0.7, 1.0; 3) STITCH confidence threshold/

588 minimum_connections/maximum_top_connections: 200/10/9999, 400/10/9999,

589   500/10/9999, 600/5/9999, 700/5/9999, 0/10/15, 0/10/25, 0/10/35, 0/10/50,

590   0/10/100, 0/10/200, 0/10/300; 4) STRING protein-protein confidence threshold: 0,

591   100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750; 5) Compound

592   "positive" class target groups: "Target_Cell", "Target_Cell_Strict",

593   "Target_Cell_Sympt", "Target_Cell_Sympt_Strict" 6) Target host gene sets for

594   SARS-CoV-2: "score_wiki", "score20", "score25", "score30", "score40",

595   "score_5_weighted", "entry_only".

596

597   The best parameters were established through cross-validation in 5 repeats of 5-

598   fold stratified *k*-fold splitting for each parameter combination. Drugs were ranked

599   by their profile correlations with the disease profile. Class separation threshold was

600   set as the one resulting in the minimal difference between sensitivity and

601   specificity. Balanced accuracy was used for establishing the best parameter

602   combinations due to high class imbalance.

603

604   Median ranked *(r*-values) and their MADs were calculated for compounds

605   independently for "Aggregated" and "WikiPathway" SARS-CoV-2 host interactomes.

606   The ensemble of parameter settings in the range of balanced classification

607   accuracies of 80-84.9% was used to provide consensus ranking of drug and food

608   molecule candidates. The final ranking list for the two parameter sets was

609   calculated using geometric mean of the *r*-values and MADs to guarantee that only

610   the candidates scored highly using both sources of SARS-CoV-2 target genes would

611   be at the top of the list. *r*-values were calculated for each compound as the sum of

612   compounds in the "negative" class with the correlation coefficient higher than that

613 of a given compound divided by the total number of the "negative" class

614 compounds.

615

616 The toxic compounds were removed from the presented lists using literature and

617 T3DB. For food molecules, we have also excluded compounds which are present in

618 trace amounts (e.g. minerals) and/or are of non-natural origin.

619

620 **GSEA section:**

621 Pathway analytics was performed using gene set enrichment analysis (GSEA) via

622 the Python GSEAPY package [56]. We used the random walk propagation algorithm

623 on the initial SARS-CoV-2 host interactome to simulate the effects of SARS-CoV-2

624 on human interactome networks. This simulated genomic profile was used as input

625 for the PreRank module of GSEA to find statistically significant enriched

626 pathways/gene sets. KEGG v7.2 and Reactome v74 were used as default gene sets.

627

628 As a means of validating food predictions, we built genomic perturbation profiles of

629 the predicted foods with anti-COVID-19 properties and subjected them to gene set

630 enrichment analysis. Food genomic perturbation profiles were built by aggregating

631 genomic perturbation profiles of their constituent molecules which represent the

632 interactions between food molecules and the human interactome. Profiles of food

633 molecules were weighted by their concentration in each of the food items (which

634 can be found in the Additional file 7) and added together. The final food

635 perturbation profiles were multiplied by the SARS-CoV-2 simulated genomic

636 perturbation profile before being used as inputs to the PreRank module of GSEA.

637 The statistical significance of the pathway overlap between each of the anti-COVID-

638 19 predicted foods and SARS-CoV-2 was measured using Fisher exact test

639 (Additional file 5). A threshold of FDR=0.05 was used to define the statistically

640 significant enriched pathways in both groups.

641

642 **Food map construction:**

643 The final food selection was based on the highest number and, where available,

644 quantity of the anti-SARS-CoV-2 food compounds and is provided in the Additional

645 file 7. Concentration of compounds within foods were extracted from the USDA

646 Special Interest Database on Flavonoids [57].

647

648 An enrichment score for each food item was calculated as a weighted sum of the

649 number of different molecules with anti-COVID-19 properties (phytochemical

650 "diversity") and their relative abundance where the experimental concentration

651 data of molecules was available across all foods studied here. The enrichment score

652 is defined as:

653

$$ES = \sum_i \log\left(\frac{\max(\boldsymbol{c}_i)}{st(\boldsymbol{c}_i)} + 1\right) + b_i$$

654

655 Where $\boldsymbol{c}_i$ is a vector with the concentrations of food molecule i across several

656 samples of the food of interest, $st(\cdot)$ denotes standard deviation, and b_i is a binary

657 indicator of the molecule-food association.

Supplementary information

Additional file 1. Enriched pathways in SARS-CoV-2

Additional file 2. Best parameters model with aggregated interactome

Additional file 3. Best parameters model with wiki pathways

Additional file 4. Anti-SARS-CoV-2 predicted drugs and foods.

Additional file 5. GSEA of anti-SARS-CoV-2 predicted foods

Additional file 6. Manually curated anti SARS-CoV-2 compounds

Additional file 7. Foods and compound concentrations

DECLARATIONS

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Supporting data and software can be found in

https://bitbucket.org/iAnalytica/corona-ai-correlator/src/master/

Competing interests

Authors declare no competing interests

687 **Author's contributions**

688 K.V. designed the concept and supervised the study. I.L. and GG implemented the

689 computational workflow and aggregated the datasets. K.V., I.L., J.Y. and M.B.

690 devised the applications and use cases. K.V., I.L., G.G., M.S., G.C., D.V., V.V. and

691 A.Q. interpreted the drug repositioning and food-molecule discovery findings. G.G.,

692 K.V. and I.L. benchmarked the algorithms. K.V. and I.L. prepared the figures. K.V.

693 and R.M. wrote the manuscript along with J.Y., M.B., D.V., J.Y. and I.L. All authors

694 contributed to results interpretation.

695

705 science community for their computing contribution as part of the CORONA-

706 AI/DRUGS DreamLab projects.

707

REFERENCES

709

710   1.   Wang, C., et al., *A novel coronavirus outbreak of global health concern.* Lancet,
711   2020. **395**(10223): p. 470-473.

712   2.   Fernandez-Quintela, A., et al., *Key Aspects in Nutritional Management of*
713   *COVID-19 Patients.* J Clin Med, 2020. **9**(8).

714   3.   Di Matteo, G., et al., *Food and COVID-19: Preventive/Co-therapeutic Strategies*
715   *Explored by Current Clinical Trials and in Silico Studies.* Foods, 2020. **9**(8).

716   4.   Chong, C.R. and D.J. Sullivan, Jr., *New uses for old drugs.* Nature, 2007.
717   **448**(7154): p. 645-6.

718   5.   Zhou, Y., et al., *Artificial intelligence in COVID-19 drug repurposing.* Lancet Digit
719   Health, 2020.

720   6.   Guy, R.K., et al., *Rapid repurposing of drugs for COVID-19.* Science, 2020.
721   **368**(6493): p. 829-830.

722   7.   Group, R.C., et al., *Dexamethasone in Hospitalized Patients with Covid-19 -*
723   *Preliminary Report.* N Engl J Med, 2020.

724   8.   Scalbert, A., et al., *The food metabolome: a window over dietary exposure.* Am J
725   Clin Nutr, 2014. **99**(6): p. 1286-308.

726   9.   Panche, A.N., A.D. Diwan, and S.R. Chandra, *Flavonoids: an overview.* J Nutr
727   Sci, 2016. **5**: p. e47.

728   10.   Naithani, R., et al., *Antiviral activity of phytochemicals: a comprehensive review.*
729   Mini Rev Med Chem, 2008. **8**(11): p. 1106-33.

730 11. Barabási, A.-L., G. Menichetti, and J. Loscalzo, *The unmapped chemical*

731   *complexity of our diet.* Nature Food, 2020. **1**(1): p. 33-37.

732 12. Belanger, M.J., et al., *Covid-19 and Disparities in Nutrition and Obesity.* N Engl J

733   Med, 2020. **383**(11): p. e69.

734 13. Foldi, M., et al., *Obesity is a risk factor for developing critical condition in COVID-*

735   *19 patients: A systematic review and meta-analysis.* Obes Rev, 2020. **21**(10): p.

736   e13095.

737 14. Zuo, T., et al., *Depicting SARS-CoV-2 faecal viral activity in association with gut*

738   *microbiota composition in patients with COVID-19.* Gut, 2020.

739 15. Veselkov, K., et al., *HyperFoods: Machine intelligent mapping of cancer-beating*

740   *molecules in foods.* Sci Rep, 2019. **9**(1): p. 9237.

741 16. Aksenov, A.A., et al., *Auto-deconvolution and molecular networking of gas*

742   *chromatography-mass spectrometry data.* Nat Biotechnol, 2020.

743 17. Boozari, M. and H. Hosseinzadeh, *Natural products for COVID-19 prevention*

744   *and treatment regarding to previous coronavirus infections and novel studies.*

745   Phytother Res, 2020.

746 18. Xian, Y., et al., *Bioactive natural compounds against human coronaviruses: a*

747   *review and perspective.* Acta Pharm Sin B, 2020. **10**(7): p. 1163-1174.

748 19. Gordon, D.E., et al., *A SARS-CoV-2 protein interaction map reveals targets for*

749   *drug repurposing.* Nature, 2020. **583**(7816): p. 459-468.

750 20. de Chassey, B., et al., *Virus-host interactomics: new insights and opportunities*

751   *for antiviral drug discovery.* Genome Med, 2014. **6**(11): p. 115.

752 21. Basu, A., A. Sarkar, and U. Maulik, *Molecular docking study of potential*

753   *phytochemicals and their effects on the complex of SARS-CoV2 spike protein*

754   *and human ACE2.* Sci Rep, 2020. **10**(1): p. 17699.

755   22.   Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that*
756   *D614G Increases Infectivity of the COVID-19 Virus.* Cell, 2020. **182**(4): p. 812-
757   827 e19.

758   23.   Ostaszewski, M., et al., *COVID-19 Disease Map, building a computational*
759   *repository of SARS-CoV-2 virus-host interaction mechanisms.* Sci Data, 2020.
760   **7**(1): p. 136.

761   24.   Gysi, D.M., et al., *Network Medicine Framework for Identifying Drug Repurposing*
762   *Opportunities for COVID-19.*ArXiv, 2020.

763   25.   Hofree, M., et al., *Network-based stratification of tumor mutations.* Nat Methods,
764   2013. **10**(11): p. 1108-15.

765   26.   Pushpakom, S., et al., *Drug repurposing: progress, challenges and*
766   *recommendations.* Nat Rev Drug Discov, 2019. **18**(1): p. 41-58.

767   27.   Angelopoulou, A., et al., *Imiquimod - A toll like receptor 7 agonist - Is an ideal*
768   *option for management of COVID 19.* Environ Res, 2020. **188**: p. 109858.

769   28.   Al-Motawa, M.S., et al., *Vulnerabilities of the SARS-CoV-2 Virus to*
770   *Proteotoxicity—Opportunity for Repurposed Chemotherapy of COVID-19*
771   *Infection.* Frontiers in Pharmacology, 2020. **11**(1579).

772   29.   Neogi, U., et al., *Feasibility of Known RNA Polymerase Inhibitors as Anti-SARS-*
773   *CoV-2 Drugs.* Pathogens, 2020. **9**(5).

774   30.   Lee, K., et al., *Gemcitabine, a broad-spectrum antiviral drug, suppresses*
775   *enterovirus infections through innate immunity induced by the inhibition of*
776   *pyrimidine biosynthesis and nucleotide depletion.* Oncotarget, 2017. **8**(70): p.
777   115315-115325.

778   31.   Derosa, L., et al., *The immuno-oncological challenge of COVID-19.* Nature
779   Cancer, 2020. **1**(10): p. 946-964.

780   32.   Wang, S., et al., *Cholesterol 25-Hydroxylase inhibits SARS-CoV-2 and other*
781   *coronaviruses by depleting membrane cholesterol.* EMBO J, 2020. **39**(21): p.
782   e106057.

783   33.   Kow, C.S. and S.S. Hasan, *Meta-analysis of Effect of Statins in Patients with*
784   *COVID-19.* Am J Cardiol, 2020. **134**: p. 153-155.

785   34.   Romero, R., et al., *Metformin, the aspirin of the 21st century: its role in*
786   *gestational diabetes mellitus, prevention of preeclampsia and cancer, and the*
787   *promotion of longevity.* Am J Obstet Gynecol, 2017. **217**(3): p. 282-302.

788   35.   Scheen, A.J., *Metformin and COVID-19: From cellular mechanisms to reduced*
789   *mortality.* Diabetes Metab, 2020.

790   36.   Kow, C.S. and S.S. Hasan, *Mortality risk with preadmission metformin use in*
791   *patients with COVID-19 and diabetes: A meta-analysis.* J Med Virol, 2020.

792   37.   Chattopadhyay, D., et al., *Recent advancements for the evaluation of anti-viral*
793   *activities of natural products.* N Biotechnol, 2009. **25**(5): p. 347-68.

794   38.   Drewnowski, A. and C. Gomez-Carneros, *Bitter taste, phytonutrients, and the*
795   *consumer: a review.* Am J Clin Nutr, 2000. **72**(6): p. 1424-35.

796   39.   Mierziak, J., K. Kostyn, and A. Kulma, *Flavonoids as important molecules of*
797   *plant interactions with the environment.* Molecules, 2014. **19**(10): p. 16240-65.

798   40.   Jang, M., et al., *Tea Polyphenols EGCG and Theaflavin Inhibit the Activity of*
799   *SARS-CoV-2 3CL-Protease In Vitro.* Evid Based Complement Alternat Med,
800   2020. **2020**: p. 5630838.

801   41.   Liu, R.H., *Potential synergy of phytochemicals in cancer prevention: mechanism*
802   *of action.* J Nutr, 2004. **134**(12 Suppl): p. 3479S-3485S.

803   42.   Derosa, G., et al., *A role for quercetin in coronavirus disease 2019 (COVID-19).*
804   Phytother Res, 2020.

43. Liu, R.H., *Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals.* Am J Clin Nutr, 2003. **78**(3 Suppl): p. 517S-520S.

44. Bousquet, J., et al., *Cabbage and fermented vegetables: From death rate heterogeneity in countries to candidates for mitigation strategies of severe COVID-19.* Allergy, 2020.

45. Ikuta, K., K. Mizuta, and T. Suzutani, *Anti-influenza virus activity of two extracts of the blackcurrant (Ribes nigrum L.) from New Zealand and Poland.* Fukushima J Med Sci, 2013. **59**(1): p. 35-8.

46. Tate, J.G., et al., *COSMIC: the Catalogue Of Somatic Mutations In Cancer.* Nucleic Acids Res, 2019. **47**(D1): p. D941-D947.

47. Brown, G.R., et al., *Gene: a gene-centered information resource at NCBI.* Nucleic Acids Res, 2015. **43**(Database issue): p. D36-42.

48. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.

49. UniProt, C., *UniProt: a worldwide hub of protein knowledge.* Nucleic Acids Res, 2019. **47**(D1): p. D506-D515.

50. Huttlin, E.L., et al., *The BioPlex Network: A Systematic Exploration of the Human Interactome.* Cell, 2015. **162**(2): p. 425-440.

51. Kuhn, M., et al., *STITCH 2: an interaction network database for small molecules and proteins.* Nucleic Acids Res, 2010. **38**(Database issue): p. D552-6.

52. Wishart, D.S., et al., *DrugBank 5.0: a major update to the DrugBank database for 2018.* Nucleic Acids Res, 2018. **46**(D1): p. D1074-D1082.

53. Ursu, O., et al., *DrugCentral: online drug compendium.* Nucleic Acids Res, 2017. **45**(D1): p. D932-D939.

831   54.   Kuleshov, M.V., et al., *The COVID-19 Drug and Gene Set Library.* Patterns (N

832   Y), 2020. **1**(6): p. 100090.

833   55.   Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2*

834   *and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p.

835   271-280 e8.

836   56.   Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based*

837   *approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U

838   S A, 2005. **102**(43): p. 15545-50.

839   57.   *Bhagwat S., et al., USDA's Expanded Flavonoid Database for the assessment of*

840   *Dietary Intakes. U.S. Department of Agriculture, Agricultural Research Service,*

841   *2014. Nutrient Data Laboratory Home Page:*

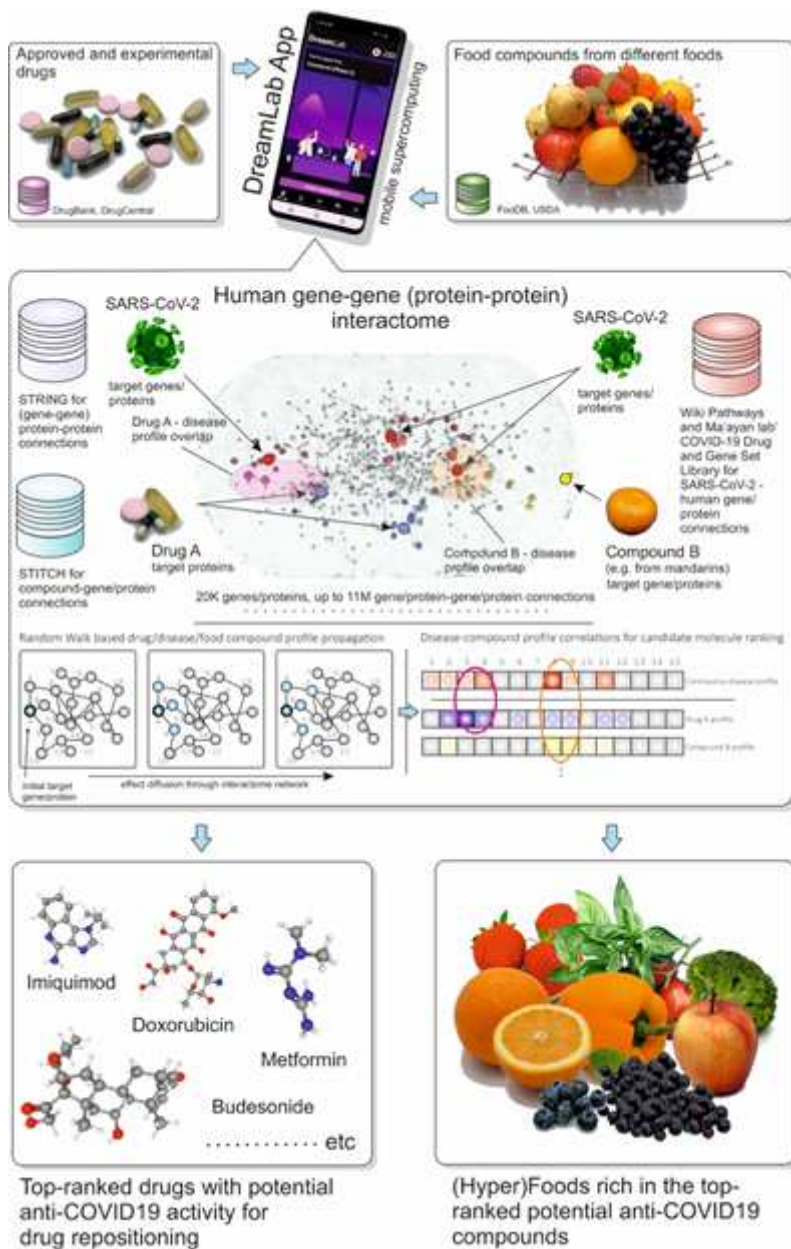842   *http://www.ars.usda.gov/nutrientdata*

843

Figures

Figure 1

Schematic diagram of overall workflow. The random walk with restarts algorithm operating within a mobile supercomputing DreamLab App is used to simulate how drug and food-based compounds interact with COVID-19 associated viral gene/protein networks. This has been extrapolated from human genome-wide gene-gene (protein-protein) interactome data and based on known COVID-19 human proteome viral targets (i.e. human genes/proteins interacting with different stages of the virus life cycle to facilitate replication and/or enhance viral potency). Both disease and molecular compound impacts are propagated through the interactome network to model the overall cellular response/interactome perturbation. The resulting compound and disease profiles are then correlated to rank compounds according to their network "overlap" with "reference" viral profiles. This approach is based on the

assumption that to have an effect, candidate compounds should target the same network component(s) as the one(s) disrupted by the virus. Therapeutic effect can be direct, or indirect, for example where compounds are found to interact with neighbouring network nodes, resulting in subsequent effect propagation to the desired target.
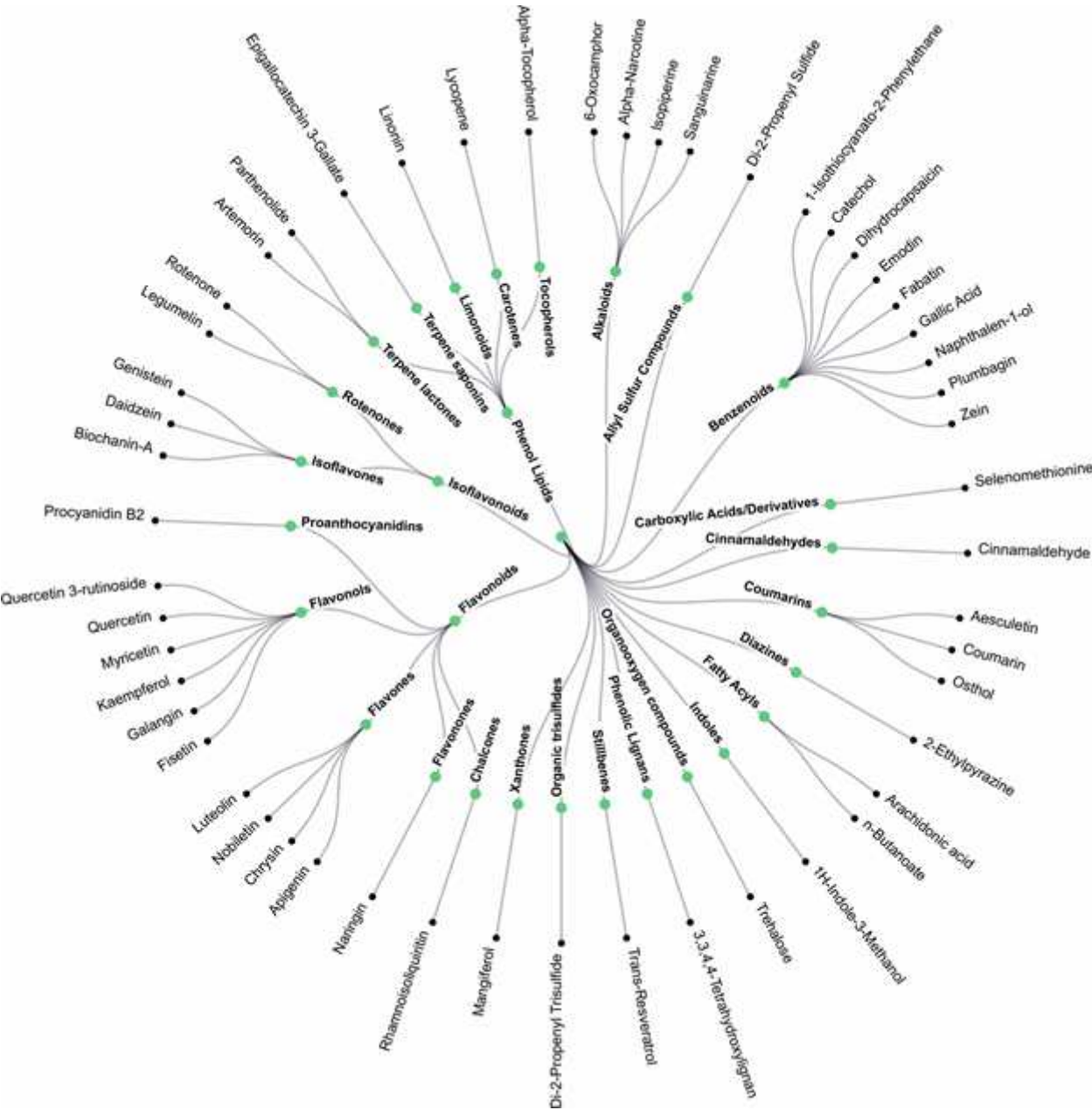

Figure 2

Hierarchical classification of the top 52 predicted antiviral molecules targeting SARS-CoV-2 human interactome networks.
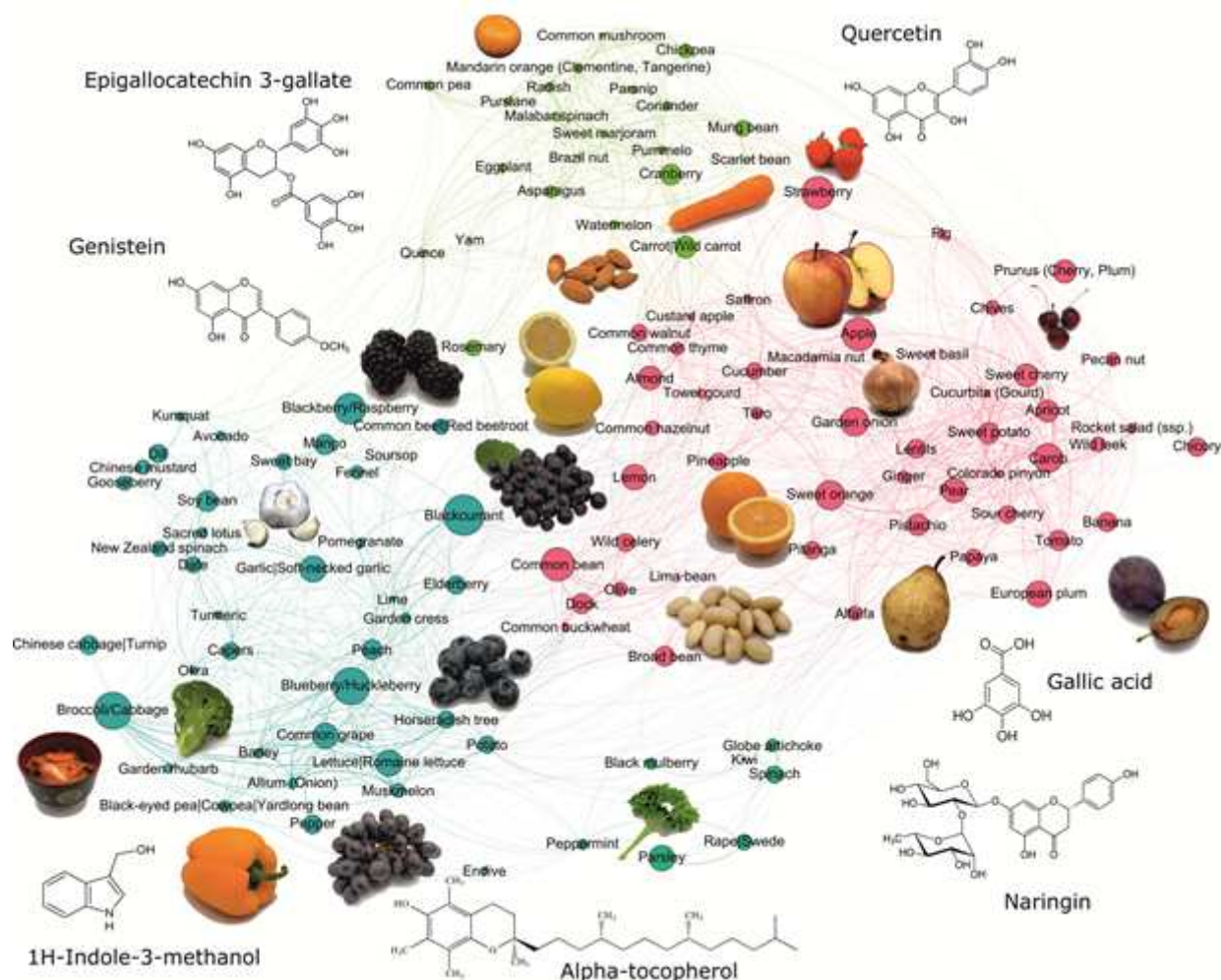
Figure 3

The contained profiles of compounds within specific foods, with predicted effectiveness in targeting SARS-CoV-2-host interactome networks. Each node in the figure denotes a particular food item and node size in each case is scaled by the derived enrichment score based on the diversity and relative levels of molecules with predicted anti-COVID-19 properties. The links between nodes reflect the pairwise correlation ("similarity") antiviral profiles in foods, thus the clusters of foods illustrate molecular commonality between them.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1EnrichedpathwaysSARSCoV2.xlsx
- Additionalfile2bestparametersmodelaggregated.xlsx
- Additionalfile3bestparametersmodelwiki.xlsx
- Additionalfile4AntiSARSCoV2predicteddrugsfoods.xlsx

- Additionalfile5GSEAfoods.xlsx
- Additionalfile6ManuallycuratedantiSARSCoV2.xlsx
- Additionalfile7Foodsandcompoundconcentrations.xlsx