

# Between viral targets and differentially expressed genes in COVID-19: the sweet spot for therapeutic intervention

Carme Zambrana<sup>1</sup>, Alexandros Xenos<sup>1</sup>, René Böttcher<sup>1</sup>, Noël Malod-Dognin<sup>1,2</sup>, and Nataša Pržulj<sup>1,2,3,\*</sup>

<sup>1</sup>Barcelona Supercomputing Center, Barcelona, Spain.

<sup>2</sup>Department of Computer Science, University College London, London, WC1E 6BT, United Kingdom.

<sup>3</sup>ICREA, Pg. Lluís Companys 23, Barcelona, Spain.

\*natasha@bsc.es

## ABSTRACT

The COVID-19 pandemic is raging. It revealed the importance of rapid scientific advancement towards understanding and treating new diseases. To address this challenge, we adapt an explainable artificial intelligence algorithm for data fusion and utilize it on new omics data on viral-host interactions, human protein interactions, and drugs to better understand SARS-CoV-2 infection mechanisms and predict new drug-target interactions for COVID-19. We discover that in the human interactome, the human proteins targeted by SARS-CoV-2 proteins and the genes that are differentially expressed after the infection have common neighbors central in the interactome that may be key to the disease mechanisms. We uncover 185 new drug-target interactions targeting 49 of these key genes and suggest re-purposing of 149 FDA-approved drugs, including drugs targeting VEGF and nitric oxide signaling, whose pathways coincide with the observed COVID-19 symptoms. Our integrative methodology is universal and can enable insight into this and other serious diseases.

## 1 Introduction

2 The ongoing COVID-19 pandemic exposed the shortcomings of healthcare systems and devastated the economy<sup>1-3</sup>. A major  
3 issue is the lack of adequate medications. This has mostly been addressed by extrapolating drug targets from related viruses and  
4 assessing the efficacy of approved drugs<sup>4-7</sup>. Once an effective vaccine has been developed, immunizing most of the population  
5 will pose serious other challenges, including economic and logistic ones. Thus, treatment options for patients is a key issue that  
6 will remain relevant.

7 The COVID-19 disease is caused by a betacoronavirus termed severe acute respiratory syndrome coronavirus 2 (SARS-  
8 CoV-2). This virus reproduces in the upper respiratory tract and is highly infectious due to asymptomatic carrier transmission<sup>8,9</sup>.  
9 As a (+)RNA virus, SARS-CoV-2 completely depends on infected host cells to replicate and thus, interactions with the host  
10 molecular network are crucial in avoiding the host immune response and reprogramming the cell to enforce its reproduction<sup>10</sup>.  
11 SARS-CoV-2 binds to a cellular receptor to enter a host cell, the exopeptidase angiotensin converting enzyme 2 (ACE2)<sup>11</sup>.  
12 Upon ACE2 binding, transmembrane protease, serine 2 (TMPRSS2), is required to prime the viral spike protein and allow the  
13 virus to enter the host cell via endocytosis<sup>12,13</sup>. Once a cell has been infected, the synthesized viral proteins can interact with a  
14 number of host factors to perform viral functions, likely by modulating cellular processes ranging from vesicle trafficking to  
15 regulating gene-expression and ubiquitination<sup>5</sup>. An inflammatory response to the SARS-CoV-2 infection had been revealed  
16 by 1,910 differentially expressed host genes (DEGs) in infected lung tissue<sup>14</sup>. Elevated glucose levels and glycolysis have  
17 been shown to promote SARS-CoV-2 replication and cytokine production in monocytes<sup>15</sup>. Thus, targeting metabolic pathways  
18 may provide new strategies to treat COVID-19 disease. While many studies focused on ACE2, TMPRSS2 and other direct  
19 viral interaction targets as candidates for treating SARS-CoV-2<sup>12,16,17</sup>, few studies have investigated the positioning of the  
20 protein targets in the host molecular interactome and the possible impacts of such positioning<sup>18,19</sup>. Interestingly, the DEGs  
21 from Blanco-Melo *et al.* 2020<sup>14</sup> show little overlap with human proteins that directly interact with the viral ones (described by  
22 Gordon *et al.* 2020<sup>5</sup>). Thus, the underlying molecular mechanisms, from the proteins targeted by the virus to the ones altered  
23 once the infection is onset, is not fully understood.

24 Novel insights have been found by integrating several different molecular interaction network types, by using data fusion  
25 algorithms, such as finding cellular wiring patterns specific to disease (“rewired genes” in disease compared to control), that  
26 can also be used for predicting new cancer-related genes<sup>20,21</sup>. These data fusion algorithms are based on Non-Negative Matrix

27 Tri-Factorization (NMTF), which approximates a high-dimensional data matrix that contains relations between two entities (e.g.  
28 between proteins and drugs) as a product of three low-dimensional, non-negative matrices (called factors)<sup>22</sup>. NMTF based  
29 methods were initially proposed for dimensionality reduction and co-clustering due to its relatedness to k-means clustering<sup>22–24</sup>.  
30 The clustering information is encoded in the low-dimensional matrix factors, named cluster indicator matrices. Moreover,  
31 NMTF is also an intermediate data integration method, by which several relational matrices can be decomposed simultaneously  
32 sharing a matrix factor across the decompositions, that directly integrates all datasets through the inference of a single joint  
33 model; it can also be used for predicting new entries in the input matrices due to the matrix completion property<sup>20,21,25,26</sup>.  
34 Unlike other artificial intelligence algorithms, this method is interpretable and its predicted values are traceable, which are  
35 essential properties when mining biological data. These inherent features of NMTF, interpretability, dimensionality reduction,  
36 co-clustering and prediction of new entries (via matrix completion), have thus far been used for predicting disease associations<sup>25</sup>,  
37 protein-protein interactions<sup>23</sup> and gene functions<sup>27</sup>, as well as for discovering disease-related genes<sup>28</sup>. Moreover, for cancer,  
38 this data fusion framework successfully uncovered patient subgroups with different prognostic survival outcomes, predicted  
39 novel cancer-related genes and proposed drugs for re-purposing<sup>21</sup>. Thus, to predict candidate target genes and the existing drugs  
40 that could be re-purposed for treating COVID-19, we adapt our versatile data fusion framework to fuse viral host interactions,  
41 human protein interactions, and drug data<sup>21</sup>. Among the predicted drug-target interactions (DTIs), we observed that one third of  
42 the targeted genes directly connect the host proteins targeted by the viral proteins (we termed them viral interactors (VIs)) and  
43 the host differentially expressed genes (DEGs). Thus, we decided to further explore how the VIs and the DEGs are connected  
44 in the host interactome.

45 The host molecular interactome is usually modeled by using networks, where biological entities (genes, or equivalently in  
46 this study, proteins, as gene products) and the interactions between them are represented as network nodes and edges (links),  
47 respectively. Networks are widely applicable and are frequently used for representing: physical interactions of proteins via  
48 protein-protein interaction (PPI) networks; metabolic interactions (MI) that correspond to known metabolic pathways; or  
49 functional associations between genes, such as epistasis via genetic interaction (GI) networks. To obtain information about  
50 the organization of a network and the wiring patterns of its nodes, various network properties are being used, ranging from  
51 the basic node degree (the number of edges incident to the node; the higher the degree, the more “degree central” the node)  
52 to several other measures of network centrality<sup>29</sup>. Furthermore, local and global network topology can be assessed by using  
53 graphlets, small, connected, non-isomorphic, induced subgraphs of large networks, that provide a quantitative measure of the  
54 wiring pattern around a node in the network and thus have been used for various applications, including for node centrality and  
55 network distance measures<sup>30,31</sup>. In particular, Graphlet Degree Vectors (GDVs)<sup>30</sup> capture the local wiring patterns for each  
56 node in a network. Studying the topology (structure) of molecular interaction networks revealed that genes (or proteins) with  
57 similar biological functions are either neighbors in the network that tend to form clusters, or are characterized by similar wiring  
58 patterns, independent of being neighbors in the network<sup>30,32</sup>. Thus, to investigate the interplay between the human proteins that  
59 are viral interactors (VI) and those human genes (or equivalently proteins) that are differentially expressed after the infection  
60 (DEGs), we study these two protein sets and their neighbors in the context of the human interactome. We use a holistic view of  
61 the human interactome by merging the PPI, GI and MI networks in the molecular interaction network (MIN). We find that the  
62 neighbors in the human MIN of these two sets have a large overlap (we term the genes in the overlap the “common neighbors”)  
63 containing central genes (with larger node degree in the MIN). Moreover, we find that they are enriched in viral processes, and  
64 hence, they might be involved in the COVID-19 mechanisms.

65 We find new drug-target interactions that open new ways for potential COVID-19 treatments. Firstly, we predict candidate  
66 target genes and the existing drugs that could be re-purposed for treating COVID-19 by disrupting the disease mechanisms.  
67 Moreover, we observe that one third of the targeted genes directly connect the viral interactors (VIs) and the host differentially  
68 expressed genes (DEGs). Secondly, we uncover that in the human interactome VIs and DEGs, while mostly disjoint, are  
69 indirectly connected by their neighbors (common neighbor genes). Furthermore, we find that the common neighbor genes  
70 might be key to the infection mechanisms used by the virus since they are enriched in various viral processes. Finally, we  
71 investigate the biological mechanisms that the predicted candidate target genes are involved in and their relevance for treating  
72 COVID-19. We discover that the targeted genes participate in two molecular pathways, nitric oxide and VEGF signaling, whose  
73 functions strongly correlate with several observed COVID-19 symptoms.

## 74 Results

### 75 The data fusion framework predicts novel DTIs for SARS-CoV-2

76 To predict candidate target genes and the existing drugs that could be re-purposed for treating COVID-19, we adapt our data  
77 fusion framework based on graph-regularized non-negative matrix tri-factorization (GNMTF). This framework was successfully  
78 utilized for ovarian cancer to uncover patient subgroups with different prognostic survival outcomes, predict novel cancer-related  
79 genes and propose drugs for re-purposing<sup>21</sup> (for more details see section “Data fusion framework tailored to SARS-CoV-2” in  
80 Methods).

81 Using this framework, we fuse two heterogeneous networks, viral-host interactions (VHIs) and previously known drug-target  
82 interactions (DTIs), containing three different data types: SARS-CoV-2 proteins, human genes and drugs (either FDA-approved  
83 and experimental) (Figure 1A). In addition, drug relations were captured based on the similarity of their chemical structures,  
84 creating a Drug Chemical Similarity (DCS) network. To have a holistic view of the relationship between genes, we merged the  
85 protein-protein interaction (PPI) network with the genetic interaction (GI) and metabolic interaction (MI) networks. We term  
86 the network created by merging these networks the molecular interaction network (MIN). The PPI network contains the physical  
87 interactions between the products of the genes (proteins), the GI network contains the genetic interactions between genes and  
88 the MI network connects the genes that participate in the same metabolic pathway. In particular, we add the MI network, since  
89 it has been demonstrated that metabolic processes, such as glycolysis, promote SARS-CoV-2 replication, and hence, targeting  
90 metabolic pathways might be key for treating COVID-19<sup>15</sup> (for more details see section “Datasets, pre-processing and matrix  
91 construction” in Methods). We validate that after merging the three networks, the topology (structure) of the PPI network  
92 dominates the MIN by comparing commonly used network properties and the wiring patterns of the constituent networks  
93 and the MIN (for more details see section “Comparison of the molecular interaction network and its constituent networks” in  
94 Supplementary Materials).

95 In the data fusion framework, the VHI and the DTI networks are represented by their adjacency matrices ( $R_{12}$  and  $R_{23}$ ,  
96 respectively) while the MIN and the DCS networks are represented by their Laplacian matrices ( $L_2$  and  $L_3$ ). The VHI and  
97 the DTI matrices are jointly decomposed in low-dimensional matrix factors  $R_{12} \approx G_1 H_{12} G_2^\top$  and  $R_{23} \approx G_2 H_{23} G_3^\top$ . The  
98 network structure (topology) information from the MIN and DCS networks are incorporated into the data fusion by using  
99 two regularization terms ( $tr(G_2^\top L_2 G_2)$  and  $tr(G_3^\top L_3 G_3)$ , respectively) (Figure 1B; for more details see section “Data fusion  
100 framework tailored to SARS-CoV-2” in Methods).

101 After obtaining the factor matrices, we cluster the genes and drugs by applying hard clustering to the corresponding matrix  
102 factors,  $G_2$  and  $G_3$ , respectively (for more details see section “Extracting clusters of genes and drug” in Methods). As a  
103 validation step, we assess whether the genes and the drugs of the formed clusters are biologically and mechanistically related,  
104 respectively, by performing enrichment analysis with the genes in the  $G_2$  matrix annotated with Gene Ontology (GO) terms  
105 and the drugs in the  $G_3$  matrix annotated with their associated DrugBank “Drug Category” (DC) (for more details see section  
106 “Enrichment analysis of gene and drug clusters” in Methods). As shown in Figure 2, more than 80% of the gene clusters show  
107 enrichments for each of the three GO domains (i.e. Biological Process, Cellular Component, Molecular Function), while the  
108 percentage of enriched genes in the clusters is at least 15%. Similarly, 90% of the drug clusters are enriched in DC (Figure 2).  
109 To assess if an observed enrichment is greater than or equal to an enrichment that may be by chance, we perform a permutation  
110 test (for more details see section “Enrichment analysis of gene and drug clusters” in Methods). The enrichments of both the  
111 gene and the drug clusters are statistically significant compared to randomly generated clusters ( $p$ -value  $\leq 0.01$ ), confirming  
112 that the joint decomposition of VHIs and DTIs successfully extracts meaningful information from these data and is capable of  
113 predicting novel drug-target relations.

114 To predict new, previously unobserved drug-gene interactions, we use the matrix completion property of the reconstructed  
115 drug-target relation matrix,  $\widehat{R}_{23} \approx G_2 H_{23} G_3^\top$ . Each entry of the reconstructed matrix, corresponding to a drug-gene pair,  
116 contains an association score,  $s_A$ , which can be interpreted as a relative measure of confidence for each drug-gene association.  
117 For assessing the accuracy of our predictions, we create precision-recall (PR) and receiver operating characteristic (ROC)  
118 curves, by using as ground truth the known DTIs (for more details see section “Prediction of new drug-target interactions for  
119 drug re-purposing” in Methods). As shown in Figure 3, these curves confirm the ability of the framework to capture known  
120 DTIs accurately ( $PR-AUC = 0.696$ ;  $ROC-AUC = 0.997$ ). To select a threshold for predicting new DTIs, we utilize the F1-score,  
121 which is the harmonic mean of the precision and recall. The best F1-score ( $F_1 = 0.729$ ) is associated with a threshold of  
122  $s_A = 0.296$ , yielding 814 newly predicted DTIs with 565 drugs targeting 172 genes (Supplementary Table S2). These predicted  
123 DTIs include not only FDA-approved drugs but also experimental compounds. Due to the urgent need for finding a treatment  
124 for COVID-19, we will focus on those DTIs that include only FDA-approved drugs. Thus, focusing only on FDA-approved  
125 drugs, we ended up with 573 newly predicted DTIs with 369 drugs targeting 143 genes.

126 To assess the improvement when using the holistic view of the relationship between genes (i.e., MIN) instead of the PPI  
127 network (the topology of which dominates the MIN, as established above), we applied the same framework only using the PPI  
128 network (for more details see section “The holistic view of the human interactome increases the number of predicted DTIs” in  
129 Supplementary Materials). We find that only 61.43% of the DTIs predicted by using the MIN were also predicted by using the  
130 PPI. Moreover, only 33 out of the 533 DTIs predicted by using the PPI were not predicted by using the MIN (23 of which were  
131 targeted by FDA-approved drugs and 10 by experimental ones). In particular, these 33 DTIs have small scores (i.e., they are at  
132 the bottom of the list). Therefore, we obtain more putative DTIs and of higher scores by utilizing all three interaction data sets,  
133 PPI, GI and MI.

134 To validate the DTI obtained by using the MIN, we use four external databases that contain previously validated DTIs,  
135 namely Drug Central, Comparative Toxicogenomics Database (CTD), PharmaGKB and Therapeutic Targets Database (TTD),

136 which we did not use in the data fusion to be able to use them to validate our predictions. Overall, 187 out of the 573 (32.64%)  
137 predicted DTIs are present in at least one of these databases (Supplementary Table S2).

138 Interestingly, among the 143 genes targeted in the predicted DTIs obtained by our data fusion only one is a host protein  
139 targeted by the viral proteins; it is HDAC2 targeted by cannabidiol. To explore the other 142 genes and their possible relations  
140 with SARS-CoV-2 infection, we study their connection to the host proteins that interact with the viral proteins (we termed them  
141 viral interactors (VIs)) in the context of the MIN. We find that 58 drug targeted genes obtained by the data fusion are direct  
142 neighbors of the VIs and the remaining 84 genes are at distance 2 or 3 in the MIN from the VIs (79 are at distance 2 and 5 are  
143 at distance 3). In addition, to further explore the relation of the genes targeted by COVID-19 proteins after the infection, we  
144 study the connection of the drug targeted genes obtained by our data fusion with the differentially expressed genes (DEGs)  
145 in COVID-19 infection described by Blanco-Melo *et al.* 2020<sup>14</sup> in the context of MIN. We find that 10 out of the 143 drug  
146 targeted genes obtained by our data fusion are DEGs, 100 out of the 143 genes are neighbors of the DEGs and the rest (33 out  
147 of the 143) are at distance 2 from the DEGs in COVID-19 infection. Furthermore, we find that 49 out of the 143 genes are at  
148 the same time neighbors of the VIs and DEGs. These genes that connect the VIs and DEGs might be key targets for disrupting  
149 the disease mechanisms.

150 In summary, we adapt a data fusion framework that by jointly decomposing the viral-host protein interactions and the  
151 host-drug interactions successfully predicts new DTIs between the human targets and existing drugs that could be re-purposed.  
152 Moreover, we validate through external databases one third of the predicted DTIs. Lastly, when focusing on the targeted  
153 proteins in the predicted DTIs, we find that one third of the targeted proteins directly connect the host proteins that interact  
154 with the viral proteins and the host proteins coded by the differentially expressed genes in COVID-19 infection (i.e. they are  
155 neighbors of both), hence indicating that our predicted DTIs may hit the human interactome at the points that can disrupt  
156 the viral mechanisms going from the binding of the SARS-CoV-2 viral protein to the host protein towards the differentially  
157 expressed host gene in COVID-19 infection (detailed below).

## 158 **Topological analysis of the human interactome reveals key genes for explaining the molecular mecha-** 159 **nisms of SARS-CoV-2**

160 After finding that one third of the human targets in the predicted DTIs directly connect in the MIN to both the human proteins  
161 that interact with the viral proteins (viral interactors, VIs) and those corresponding to differentially expressed genes (DEGs) in  
162 COVID-19 infection, we further explore how the VIs and the DEGs are connected in the human interactome, in particular in  
163 the above described MIN. Our reasoning is that neighboring genes can act as links between the signal inputs, VIs, and the  
164 observed outputs, such as dysregulated genes, and may thereby be involved in the disease mechanisms.

165 We use the 332 host genes reported by Gordon *et al.* 2020<sup>5</sup> as the set corresponding of viral interactors (we term this gene  
166 set the “VI”). For the DEG set, we use the 1,910 DEGs identified by Blanco-Melo, D. *et al.* 2020<sup>14</sup> in lung tissue samples  
167 from 2 infected patients (see section “Datasets, pre-processing and matrix construction” in Methods). Furthermore, since  
168 previous studies showed that disease genes tend to form densely connected communities<sup>33</sup> in the MIN, we identify direct  
169 network neighbors of both of the above described gene sets (we term these two new gene sets the “VI neighbors” and “DEG  
170 neighbors”). As shown in Figure 4A, these two sets have 52.30% of overlap (statistically significant with p-value = 0, using  
171 hypergeometric test; for more details see section “Analysis of the molecular interaction network and its wiring patterns” in  
172 Methods) and hence, we also explore this overlap as a separated gene set (termed the “common neighbors”). Thus, VI and  
173 DEG genes, while mostly disjoint, are largely (52.30%) indirectly connected by their neighbors. To fully explore the entire set  
174 of neighbors in the MIN network of proteins participating in VIs and the protein products of DEGs in COVID-19 disease, we  
175 study separately those VI neighbor and DEG neighbor genes that overlap and those that do not overlap, and within those that do  
176 not overlap, we term the neighbors of only VIs the “VI-unique neighbors” and the neighbors of only DEGs the “DEG-unique  
177 neighbors”. The rest of the genes in the MIN that are not present in any of these five gene sets (VI, DEGs, VI-unique neighbors,  
178 DEG-unique neighbors, common neighbors) are term “background genes”.

179 To establish whether a SARS-CoV-2 infection affects proteins that are central in the MIN, we analyze the above described  
180 gene sets by the following commonly used network properties: four centrality measures (degree, eigenvector, betweenness and  
181 closeness centralities) and the clustering coefficient (for more details see section “Analysis of the molecular interaction network  
182 and its wiring patterns” in Methods). As shown in Figure 4B, VI and DEG genes show significantly higher degree centralities  
183 ( $p < 0.0001$ ) compared to the background genes, indicating their importance in the MIN. In addition, genes in both of these  
184 sets have a higher clustering coefficient than the background genes, indicating their higher tendency to form clusters (Table  
185 1). Notably, the common neighbor gene set exceeds both VI and DEG genes in all of these measures except for closeness  
186 centrality. Thus, common neighbor genes are likely to participate in many functions, since they are central in the MIN. The  
187 VI-unique and DEG-unique neighbor genes have lower centralities compared to the VI, DEG and common neighbor genes,  
188 which confirms the relevance of the common neighbors followed by the VI-unique and DEG-unique neighbor genes. Therefore,  
189 common neighbor genes are highly connected and central genes that, in particular, connect the proteins targeted by the virus to

190 the ones deregulated after the infection, and hence, they might be key for understanding the underlying molecular mechanism  
191 of COVID-19.

	Average Degree	Eigenvector Centrality	Clustering Coefficient	Betweenness Centrality	Closeness Centrality
VI	65.67	0.006282	0.137887	0.000194	<b>0.359875</b>
DEG	48.77	0.004282	0.14323	0.000168	0.340381
Common Neighbors	<b>78.02</b>	<b>0.006764</b>	<b>0.186346</b>	<b>0.00027</b>	0.358132
VI-unique Neighbors	10.04	0.00095	0.156097	0.000009	0.318445
DEG-unique Neighbors	19.01	0.001446	0.152142	0.000028	0.326536
Background	3.57	0.000291	0.096368	0.000003	0.293636

**Table 1.** Network properties of molecular interaction network (MIN), focusing on the following gene sets: viral interactors (VI), differentially expressed genes after infection (DEG), overlap of the direct network neighbors in the MIN of these two sets (common neighbors), neighbors of the VI and DEG gene set that were not in the common neighbor genes set (VI-unique neighbors and DEG-unique neighbors), and the rest of the genes in the MIN (background genes).

192 To assess whether the genes participating in the aforementioned sets have similar biological functions in the MIN network,  
193 we compare their wiring patterns, by using their Graphlet Degree Vectors (GDVs)<sup>30</sup> (for more details see section “Analysis of  
194 the molecular interaction network and its wiring patterns” in Methods). Previous molecular networks analyses revealed that  
195 genes with similar biological functions tend to group together and have similar wiring patterns in molecular networks<sup>34</sup>. As  
196 shown in Figure 4C, GDV of the common neighbor genes is different from the GDVs of the rest of the gene sets, except for the  
197 GDV of the VI. We verify this by computing the Mann-Whitney U test (with a significance level of 0.05) for each pair of orbits  
198 (Supplementary Table S4). Only five orbit counts are not statistically significantly different between the common neighbor  
199 genes and the VIs. Namely the orbits 1, 4, 5, 8 and 9 (Figure 4C orbits marked with a circle; Supplementary Table S4 marked  
200 in bold). Thus, the common neighbor genes have different wiring patterns compared to the other gene sets, and only show some  
201 similarities with the wiring patterns of VIs genes. This indicates that the common neighbors might have similar biological  
202 functions that could be related to SARS-CoV-2 infection.

203 To investigate whether the biological functions of the common neighbor genes in the MIN are related to COVID-19, we  
204 perform a functional enrichment analysis across multiple functional annotation databases: Gene Ontology (GO), KEGG,  
205 REACTOME and CORUM (for more details see section “Enrichment analysis of gene and drug clusters” in Methods). Among  
206 the significantly enriched terms (see Supplementary Table S5), many are related to viral infections (viral process, p-val =  
207 1.732946e-49; viral transcription, p-val = 4.641211e-15; viral gene expression, p-val = 3.299512e-14; viral life cycle, p-val  
208 = 1.278676e-04; positive regulation of viral process, p-val = 3.235499e-03; positive regulation of viral life cycle, p-val =  
209 1.485025e-02). We perform the same enrichment analysis for the rest of the gene sets and find that VI-unique neighbor,  
210 DEG-unique neighbor and background genes are not enriched in viral processes (see Supplementary Table S6, Supplementary  
211 Table S7, and Supplementary Table S8). These results indicate that the common neighbor genes participate in SARS-CoV-2  
212 infection and hence, they might be potential drug targets to treat COVID-19.

213 Based on these results, we conclude that SARS-CoV-2 proteins mainly interact with central human proteins, or influence the  
214 expression of host proteins that are central in the MIN. Moreover, we find that the neighbors of these two gene sets (common  
215 neighbor genes of the VIs and the DEGs) are also central in the MIN. Interestingly, the common neighbor genes are enriched  
216 in viral related processes, while the VI-unique neighbor, DEG-unique neighbor and background genes are not. Thus, these  
217 common neighbor genes (listed Supplementary Table S9) are likely to be involved in COVID-19 disease and they might be key  
218 for explaining the mechanisms that go from the host proteins targeted by the viral proteins to the differentially expressed genes  
219 resulting from the COVID-19 infection.

## 220 Predicted DTIs involving FDA-approved drugs targeting common neighbor genes disrupt biological mechanisms relevant for COVID-19

221 After discovering that the common neighbor genes (those that directly connect the host proteins that interact with the viral  
222 proteins and the proteins corresponding to differentially expressed genes in COVID-19 infection) are likely to be important  
223 in SARS-CoV-2 infection, we focus on the predicted DTIs that target these common neighbor genes; we term these DTIs  
224 “common neighbor DTIs”. The common neighbor DTIs contain 185 DTIs targeting 49 common neighbor genes with 149 drugs  
225 (see Supplementary Table S10). First, we investigate how many of the 149 drugs targeting the common neighbors are currently  
226 studied in COVID-19 context. Then, to investigate which biological mechanisms are targeted by the common neighbor DTIs,  
227 we perform a functional enrichment analysis of the 49 genes targeted in these DTIs. Finally, we manually check the enriched  
228 pathways and discuss their relevance in the context of COVID-19.  
229

230 We check whether any of these 149 drugs targeting common neighbor genes have been investigated for treating COVID-19;  
231 we use the CORona Drug InTEractions (CORDITE) database (<https://cordite.mathematik.uni-marburg.de>). Also, we ask  
232 whether they are part of interventional clinical trials currently being conducted (retrieved from <https://clinicaltrials.gov>). As  
233 shown in Supplementary Table S11, 17.44% and 11.40% of the drugs involved in the common neighbor DTIs are listed in  
234 CORDITE and subject to at least one active clinical trial on COVID-19, respectively. These results demonstrate the relevance  
235 of the predicted DTIs.

236 We perform an enrichment analysis across multiple functional annotation databases: Gene Ontology (GO), KEGG,  
237 REACTOME and CORUM (for more details see section “Enrichment analysis of gene and drug clusters” in Methods). As  
238 shown in Supplementary Table S10, the 49 genes involved in the common neighbor DTIs are enriched in several GO terms  
239 in all three GO domains (i.e. Biological Process, Cellular Component, Molecular Function). Namely, they are terms related  
240 to: G protein-coupled receptors; tyrosine kinase-mediated activation of MAPK signaling, in particular VEGF and ERK1/2;  
241 cAMP/cGMP signaling; lipid metabolism and blood circulation; ion channel activity and response to amine ligand-binding,  
242 particularly serotonin and dopamine. Likewise, when testing for the enrichment of KEGG and REACTOME pathway terms,  
243 we find enrichments of cellular response pathways (PI3K-AKT, Ras, MAPK, cAMP, VEGF) and terms linked with amine  
244 ligand-binding receptors, cytokine and nitric oxide (NO) signaling.

245 Upon closer inspection, many of these pathways are either directly or indirectly tied to NO and VEGF signaling, which  
246 are also connected to each other (see Figure 5). For instance, KDR (VEGFR-2) is required for VEGF-A mediated induction  
247 of NOS2 and NOS3, leading to the production of the signaling molecule NO by macrophages (NOS2) and endothelial  
248 cells (NOS3)<sup>35</sup>. Increased NO also directly affects inflammatory signaling by regulating cytokine (IL-6, IL-8) and PGE(2)  
249 production<sup>36,37</sup> as well as PTGS2 (COX-2) activation<sup>38</sup>. It is recognized as a key regulator of both VEGF synthesis and platelet  
250 aggregation<sup>39,40</sup>. Lastly, NO is also tied to hypoxia signaling by direct interaction with key components such as HIF-1-alpha,  
251 which in turn regulates VEGF signaling<sup>41,42</sup>.

252 Notably, striking similarities between these NO and VEGF signaling-related functions and COVID-19 symptoms can be  
253 observed. Vascular complications are common in COVID-19 patients<sup>43</sup>. In particular, recent studies on COVID-19 patients  
254 have reported an increased in VEGF levels and platelet activity, as well as extensive blood clotting and endothelial injury as  
255 a sign of direct infection of endothelial cells<sup>44-47</sup>. Moreover, cytokine storms and IL-6 have been related to severe disease  
256 COVID-19<sup>48,49</sup>, with macrophages being potential key players<sup>50</sup>. Finally, neurological symptoms have also been recognized in  
257 COVID-19 patients, and hypoxic injury is one of the possible explanations for the observed tissue damage<sup>51,52</sup>.

258 NO signaling might be central in understanding the disease, since the anatomic sites of COVID-19 symptoms, lung, heart,  
259 circulatory system and brain, also correlate with the expression patterns found for the three known human NO synthases: NOS1  
260 (neural NOS; expressed in peripheral neurons), NOS2 (endothelial NOS; expressed in endothelial cells, cardiac myocytes,  
261 cardiac conduction tissue) and NOS3 (cytokine-inducible NOS; expressed in endothelial cells, myocytes, macrophages)<sup>53</sup>.  
262 Therefore, we propose to further investigate the well tolerated drugs that modulate NO signaling and its related pathways. A  
263 potential candidate from our list of common neighbor DTIs is triflusal, which is known to interact with NFKB, NOS2, PDE10A  
264 as well as PTGS1, and for which we predict PTGS2 and NOS3 as additional target genes. Triflusal is a trifluoromethylated  
265 analogue of acetylsalicylic acid, which is not yet under investigation as COVID-19 treatment, unlike acetylsalicylic acid. Of  
266 note, both triflusal and acetylsalicylic acid act as anticoagulants and a recent study associated anticoagulation with lower  
267 mortality and intubation rates for hospitalized COVID-19 patients, providing further evidence for the validity of our findings<sup>54</sup>.

268 Related to VEGF-signaling, we suggest as a putative target gene KDR (VEGFR-2), which appears in the common neighbor  
269 DTIs targeted by tyrosine kinase inhibitors, such as Imatinib, Dasatinib, Pexidartinib. These drugs are cancer related drugs with  
270 high level toxicity, thus they must be reserved for critically ill cases. Finally, another group of candidate genes from the common  
271 neighbor DTIs worth mentioning are phosphodiesterases. Phosphodiesterases are responsible for regulating cAMP/cGMP  
272 signaling and hence, they have an interplay with both NO and VEGF<sup>55-57</sup>. Our framework predicted that phosphodiesterases  
273 (e.g. PDE4D), could be inhibited by xanthine derivatives such as theophylline.

274 In summary, by focusing on predicted drug-target interactions involving genes located in the common neighborhood of  
275 SARS-CoV-2 VIs and DEGs, we propose a list of 185 DTIs (common neighbor DTIs). For the drugs targeting the common  
276 neighbor DTIs, we validate that some of them have been investigated in COVID-19 related studies, or are currently in clinical  
277 trials for COVID-19 treatment. For the targeted genes in the common neighbor DTIs, we identify functional enrichments related  
278 to cardiovascular integrity, stress signaling and inflammation, all of which can be linked to NO and VEGF signaling. Moreover,  
279 both the molecular functions of NO signaling and the expression patterns of NO synthases correlate with reported COVID-19  
280 symptoms, making it a principal target for further study and potentially drug intervention. Finally, our predicted DTIs provide a  
281 list of FDA-approved drugs that may be used to target genes related to both the VEGF and NO signaling pathways.

## Discussion

In this work, we adapt our GNMTF-based data fusion framework to predicted candidate target genes and existing drugs that could be re-purposed for treating COVID-19. Moreover, we investigate within the human interactome the interplay between the human proteins that are directly targeted by the SARS-CoV-2 proteins and those genes that are differentially expressed after COVID-19 infection. Our study reveals that the host proteins targeted by viral proteins and the differentially expressed genes are indirectly connected by their neighbors (we termed common neighbor genes). Furthermore, we find that the common neighbors are enriched in various viral processes and hence, might be key to the infection mechanisms used by the virus. By focusing on the predicted drug-target interactions involving FDA-approved drugs and targeting the common neighbor genes, we utilize our integrative framework to predict novel drug-target interactions for genes related to the disease-affected pathways. In particular, we find NO and VEGF signaling as potential molecular pathways whose functions are very similar with several observed COVID-19 symptoms.

The framework we adapt in this study differs from other network-based computational studies for drug re-purposing applied to COVID-19 (such as Gysi *et al.* 2020<sup>19</sup> and Sadegh *et al.* 2020<sup>18</sup>) in the following: we do not only predict drugs to be re-purposed but also new candidate target genes. In particular, Gysi *et al.* 2020<sup>19</sup> ranked candidate drugs by aggregating the predictions of three different network-based methods: proximity, diffusion and AI network, based on their efficacy for COVID-19. The approach of Sadegh *et al.* 2020<sup>18</sup> is based on a group of seed nodes, which can be viral proteins and/or human genes, and then creating a subnetwork containing the seeds (using Steiner Tree algorithm), as well as ranking the drugs targeting the seeds using a centrality measure (degree, closeness, betweenness, or TrustRank). In contrast, the framework we adapt in this study is based on the fusion of several data sources, including chemical similarity of the drugs. Furthermore, the molecular interaction network that we generated for the host offers a more complete representation of the cell, as it includes information from several systems-level molecular interaction networks (protein-protein, genetic and metabolic interactions)<sup>21</sup>, whereas Gysi *et al.* 2020<sup>19</sup> and Sadegh *et al.* 2020<sup>18</sup> based its host molecular interactome only on the PPI network.

The presented data fusion framework exhibits robust performance, as exemplified by its capability to identify previously predicted DTIs involving drugs under current clinical investigation. Beyond its application in this work, the framework is highly versatile and has been successfully applied to identify of cancer driver genes, patient stratification and drug re-purposing<sup>21</sup>. To exploit further this flexibility in the context of viral infections, the framework could be extended to search for the existing drugs with broad-spectrum antiviral activities by including information about host proteins targeted by more than one virus<sup>58,59</sup>. A recent example of such re-purposing is Remdesivir, developed initially against the hepatitis C virus and currently investigated as potential COVID-19 treatment<sup>7</sup>. Besides being economically more efficient, broad spectrum antivirals are by definition likely to act on commonly exploited host pathways that tend to be indispensable for viral replication. Thus, targeting such pathways will pose a higher evolutionary hurdle for the formation of viral resistance, which may circumvent the problems faced when designing highly virus-specific drugs<sup>60</sup>.

## Methods

### Datasets, pre-processing and matrix construction

We obtained the protein-protein interaction (PPI), genetic interaction (GI) and virus-host interaction (VHI) networks from the BioGRID database (version 3.5.183)<sup>61</sup>. VHIs were based on the dataset reported by Gordon *et al.* 2020<sup>5</sup>, with  $n_1 = 26$  viral proteins interacting with 332 host genes. We constructed the human PPI network with all physical interactions between human proteins reported by at least one of the following experiments: Two-hybrid, Affinity Capture-Luminescence, Affinity Capture-Western, Affinity Capture-MS; this resulted in 16,431 proteins (nodes) connected by 272,232 interactions (edges). We constructed the GI network with all the genetic interactions reported in BioGRID; this resulted in 3,302 genes connected by 8,333 interactions. We merged these two networks with the metabolic interaction (MI) network from the KEGG database (accessed in May 2020)<sup>62</sup>. We constructed the MI network by connecting all the genes that participate in the same metabolic pathway. In particular, we retrieved as metabolic pathways all the pathways in KEGG that contain at least one of the following keywords: metabolism, metabolic, glycolysis, TCA, oxidative phosphorylation, fatty acid, pentose, degradation or biosynthesis; this resulted in 1,530 genes connected by 56,564 interactions. The resulting network from merging the PPI, GI and MI networks comprised 336,159 interactions among  $n_2 = 16,872$  genes. We termed this networks the Molecular Interaction Network (MIN) (see Supplementary Figure S1A-B for the overlap of genes and interactions of the three networks). Due to the small number of the host proteins interacting with the viral proteins (332 out of the 16,872), the relational matrix,  $R_{12}^{n_1 \times n_2}$ , containing VHIs is highly sparse. Following our previous data fusion framework<sup>21</sup>, we applied a pre-processing step based on network propagation to smoothen this matrix. The procedure consisted of iteratively updating the  $R_{12}^{n_1 \times n_2}$  using the following update rule:  $R_{12}^{t+1} = \alpha R_{12}^t \bar{A}_2 + (1 - \alpha) R_{12}^0$  where  $\bar{A}_2$  is the normalized adjacency matrix of the MIN network computed as  $\bar{A}_2 = A_2 D_2^{-1}$ ,  $R_{12}^0$  is the initial  $R_{12}$  and  $\alpha$  is a tuning parameter that controls the distance of diffusion through the MIN network. We used  $\alpha = 0.6$  and  $|R_{12}^{t+1} - R_{12}^t| < 10^{-6}$  as convergence criterion to obtain the final network-smoothened matrix,  $\bar{R}_{12}^{n_1 \times n_2}$ .

We obtained the data related to the drugs from the DrugBank database (version 5.1.3)<sup>63</sup>. Drug-Target Interactions (DTIs) between the retrieved  $n_3 = 8,279$  drugs (FDA-approved and experimental) and the  $n_2 = 16,872$  genes in our MIN were captured by the relation matrix  $R_{23}^{n_2 \times n_3}$ . This matrix is quite sparse as the known DTIs involve only 4,420 drugs targeting 2,241 genes. We used the Simplified Molecular-Input Line-Entry System (SMILES) information of these  $n_3$  drugs to create the Drug Chemical Similarity (DCS) network. First, we converted this simplified notation of the chemical structure to a binary vector in which each coordinate represents a particular substructure from the set of all known sub-structures. Then, we computed the chemical similarity between two drugs based on the similarity between their vectors using Tanimoto similarity coefficient<sup>64</sup>. Once the similarity between all drug pairs is computed, we created a network containing the top 5% most similar drug pairs, which resulted in 1,727,436 links.

### Data fusion framework tailored to SARS-CoV-2

We considered three different data types in our analyses: SARS-CoV-2 proteins, human genes and drugs and two relation types among them. SARS-CoV-2 proteins and human genes are related to each other by VHIs, which are captured in a smoothed high-dimensional relation matrix,  $R_{12}^{n_1 \times n_2}$ , with  $n_1$  viral proteins and  $n_2$  human genes (see section ‘‘Datasets, pre-processing and matrix construction’’ for more details); DTIs indicate relationships between human genes and drugs and are captured in a sparse high-dimensional binary relation matrix,  $R_{23}^{n_2 \times n_3}$ , for  $n_2$  human genes and  $n_3$  drugs, where its entries represent whether the product of a gene is targeted by a drug (1) or not (0). In addition to the relations among different data types, the relations between genes were captured by the MIN (for more details see section ‘‘Datasets, pre-processing and matrix construction’’), containing the known PPIs, GIs and MIs among them, whereas drugs relations were captured based on the similarity of their chemical structures, creating a DCS network. Both of these networks were represented by their Laplacian matrix,  $L$ , computed as:  $L = D - A$ , where  $A$  is the adjacency matrix and  $D$  is the diagonal degree matrix (i.e., whose entries on the diagonal are row sums of  $A$  and all other entries in  $D$  are zeros). Thus,  $L_2^{n_2 \times n_2}$  and  $L_3^{n_3 \times n_3}$  represent the MIN and DCS Laplacians, respectively. Figure 1A shows a schematic illustration of the datasets used in this study.

Following our previous data fusion methodology<sup>21</sup>, we used Graph-regularized non-negative matrix tri-factorization (GNMTF) to simultaneously decompose each of the two relation matrices into a product of three non-negative low-dimensional matrices while preserving the network structure of the MIN and DCS. The two decompositions,  $R_{12} \approx G_1 H_{12} G_2^\top$  and  $R_{23} \approx G_2 H_{23} G_3^\top$ , share the matrix factor  $G_2$  fusing the data via simultaneously decomposing the VHI and DTI networks. The network structure of the MIN and DCS is preserved by adding two regularization terms ( $tr(G_2^\top L_2 G_2)$  and  $tr(G_3^\top L_3 G_3)$ , respectively), so that  $G_2$  favors grouping together genes that interact in the MIN and that  $G_3$  favors grouping together drugs that are chemically similar in the DCS network. Figure 1B shows an illustration of the GNMTF. Briefly, the low dimensional matrices can be obtained by solving the optimization problem shown in equation 1:

$$\min_{\substack{G_i \geq 0 \\ (1 \leq i \leq 3)}} J = \min_{\substack{G_i \geq 0 \\ (1 \leq i \leq 3)}} \left( \|R_{12} - G_1 H_{12} G_2^\top\|_F^2 + \|R_{23} - G_2 H_{23} G_3^\top\|_F^2 + tr(G_2^\top L_2 G_2) + tr(G_3^\top L_3 G_3) \right) \quad (1)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm and  $tr$  denotes the trace of a matrix. The objective function,  $J$ , is heuristically minimized with an iterative method, starting from an initial solution and using multiplicative update rules to converge towards a locally optimal solution<sup>65</sup>. The final decomposition (used for predicting novel DTIs) was obtained by using the Singular Value Decomposition (SVD) as an initial solution and  $\frac{|J_{n+1} - J_n|}{|J_n|} < 10^{-5}$  as the convergence criterion.

### Choosing the number of clusters

The number of clusters,  $k_1$ ,  $k_2$  and  $k_3$ , are key parameters of the GNMTF. However, there is no gold standard procedure to find a suitable values of these  $k$ 's. We used the procedure inspired by Brunet *et al.* 2004<sup>66</sup>, consisting of choosing the parameter based on its cluster stability measured by the dispersion coefficient. In particular, the hard clustering procedure was applied to the corresponding matrix factor  $G_i$ , obtaining a clustering encoded in a connectivity matrix  $C_i$ , which is defined as a binary matrix where its rows and columns are the clustered entities (viral proteins, human genes or drugs) and 1 means that both entities belong to the same cluster. By applying this procedure with Random AcOl initialization, we computed the average of the obtained  $C_i$ 's,  $\bar{C}_i$ , and measured the stability of these clusterings according to the dispersion coefficient:  $\rho_{k_i} = \frac{1}{n^2} \sum_{l=1}^n \sum_{j=1}^n 4(C_{lj} - \frac{1}{2})^2$ . The idea is to choose the value of  $k_1$ ,  $k_2$  and  $k_3$  such that the obtained clusters are the most stable, i.e. for which the mean of  $\rho_{k_1}, \rho_{k_2}, \rho_{k_3}$ ,  $mean_{\rho_{k_1}, \rho_{k_2}, \rho_{k_3}} = \frac{\rho_{k_1} + \rho_{k_2} + \rho_{k_3}}{3}$ , is at its maximum.

The stability of the obtained clusters depends on the size of the cluster, smaller clusters will be more stable, but without much biological meaning, with the extreme case being when we obtain as many clusters as there are molecules. Thus, we decided to focus the grid search around the rule of thumb,  $k_i^{RT} = \sqrt{\frac{n_i}{2}}$ , which is a heuristic to determine a fair number of clusters given the number of points  $n_i$  that we need to cluster<sup>67</sup>. According to this heuristic, the number of clusters for each dataset is

383  $k_1^{RT} \approx 3$ ,  $k_2^{RT} \approx 90$ , and  $k_3^{RT} \approx 60$ , corresponding to  $n_1 = 26$  viral proteins,  $n_2 = 16,872$  human genes and  $n_3 = 8,279$  drugs.  
 384 Therefore, we performed a grid search for the following values:  $k_1 \in \{3, 5\}$ ,  $k_2 \in \{60, 80, 100, 120\}$  and  $k_3 \in \{40, 60, 80\}$ . The  
 385 most stable clustering was achieved by  $k_1 = 3$ ,  $k_2 = 120$  and  $k_3 = 80$  ( $mean_{\rho_{k_1 \cdot \rho_{k_2} \cdot \rho_{k_3}}} = 0.661$ ), which are the values that we  
 386 used for the presented results (Supplementary Figure S4).

### 387 Extracting clusters of genes and drugs

388 The matrix factors  $G_2^{n_2 \times k_2}$  and  $G_3^{n_3 \times k_3}$ , from GNMTF decomposition, are the cluster indicators of genes and drugs, respectively;  
 389 based on their entries,  $n_2$  genes are assigned to  $k_2$  clusters and  $n_3$  drugs are assigned to  $k_3$  clusters, respectively. In particular,  
 390 the hard clustering procedure of Brunet *et al.* 2004<sup>66</sup>, was used to cluster the genes of the matrix factor  $G_2^{n_2 \times k_2}$ . The columns of  
 391  $G_2^{n_2 \times k_2}$  correspond to the  $k_2$  clusters and each gene is assigned to the cluster that has the largest entry in the gene's row. The  
 392 clusters can be represented by a binary connectivity matrix,  $C_2^{n_2 \times n_2}$ , where its rows and columns are the genes and 1 means that  
 393 both genes belong to the same cluster. Similarly, we clustered the drugs of the matrix factor  $G_3^{n_3 \times k_3}$  obtaining a connectivity  
 394 matrix  $C_3^{n_3 \times n_3}$  representing the clusters of drugs.

### 395 Enrichment analysis of gene and drug clusters

396 To compute the functional enrichments of the common neighbor genes, either for the whole list of genes, or for the 49 common  
 397 neighbor genes that were predicted to be targeted by FDA-approved drugs, we used the gprofiler Python package v.1.0.0  
 398 (parameters: organism="hsapiens", source=c("GO", "KEGG", "REAC", "CORUM"))<sup>68</sup>. We used this software for its capability  
 399 to perform the enrichment analysis across multiple functional annotation databases.

400 To assess the quality of the obtained clusters of genes and drugs, we computed the enrichment of biological annotations in  
 401 the clusters. For each gene (or equivalently, protein, as a gene product) in the network, we used the most specific experimentally  
 402 validated Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) annotations present in the Gene  
 403 Ontology (GO)<sup>69</sup>, while for each drug we used the "Drug Categories"(DC) from DrugBank<sup>63</sup>. The probability that an annotation  
 404 is enriched in a cluster was computed by using a hypergeometric test, i.e., sampling without replacement strategy shown in  
 405 equation 2:

$$p = 1 - \sum_{i=0}^{X-i} \frac{\binom{K}{i} \binom{M-K}{N-i}}{\binom{M}{N}} \quad (2)$$

406 where  $N$  is the number of annotated genes (drugs) in the cluster,  $X$  is the number of genes (drugs) in the cluster that are  
 407 annotated with the given annotation,  $M$  is the number of annotated genes (drugs) in the network and  $K$  is the number of genes  
 408 (drugs) in the network that are annotated with the annotation in question. Annotations with a Benjamini-Hochberg adjusted  
 409 p-value<sup>70</sup> of  $p \leq 0.05$  were considered to be statistically significantly enriched. We measured the quality of the clustering by  
 410 computing the percentage of genes (drugs) having at least one of their annotations enriched in their clusters over all annotated  
 411 (drugs), the percentage of GO annotations (Drug Categories) enriched and the percentage of the clusters having at least one  
 412 of their genes (drugs) enriched. To assess if an observed enrichment is greater than or equal to an enrichment by chance, we  
 413 randomly shuffled (permuted) the values in the drug and gene matrix factors respectively and we used the permutation test:  
 414  $p = \frac{r+1}{n+1}$ , where  $r$  is the number of permutations that have an enrichment greater than or equal to the observed enrichment  
 415 and  $n = 100$  is the number of permutations that we used. We consider an enrichment to be statistically significant if the  
 416 corresponding p-value is lower than or equal to 0.05.

### 417 Prediction of new drug-target interactions for drug re-purposing

418 To predict new drug-target interactions, we exploited the matrix completion property of the GNMTF framework. This property  
 419 consists of reconstructing the drug-target relational matrix from the obtained low-dimensional factors as  $\widehat{R}_{23} = G_2 H_{23} G_3^T$ .  
 420 We showed that the reconstructed matrix can be used for extracting previously unobserved drug-target relations and thereby  
 421 finding new drug candidates for re-purposing in the context of cancer<sup>21</sup>. Hence, we used the reconstructed drug-target relation  
 422 matrix,  $\widehat{R}_{23} = G_2 H_{23} G_3^T$ , to predict new DTIs that were not present in the original  $R_{23}$  and to postulate new candidates for drug  
 423 re-purposing in the context of COVID-19. For each drug-gene pair, the matrix  $\widehat{R}_{23}$  contains an association score  $s_A$  representing  
 424 the strength of the association. To extract strongly associated drug-target pairs, we computed the precision-recall (PR) and  
 425 receiver operating characteristic (ROC) curves, using all previously known drug-targets in  $R_{23}$ , the original DTIs, as true  
 426 positives and the pairs in  $\widehat{R}_{23}$  that were not present in the initial dataset, new drug-gene pairs, as false positives (Figure 3). To  
 427 define the threshold for selecting the new DTIs from the reconstructed matrix, we used the association score  $s_A$  related to the  
 428 maximum F1-score, which is the harmonic mean of precision and recall ( $F_1 = 0.729$ ;  $s_A = 0.296$ ) and considered all drug-gene  
 429 pairs whose association score exceeded the threshold and are not the original DTIs to be new predicted DTIs.

## Analysis of the molecular interaction network and its wiring patterns

To compute whether the overlap between the viral interactors (VIs) neighbor gene set and the differentially expressed genes (DEGs) neighbor gene set is significant, we performed a Hypergeometric Test (see equation 2) where  $N$  is the number of genes that are the neighbors of VI genes,  $X$  is the number of genes that are both the neighbors of DEGs and the neighbors of VIs,  $M$  is the total number of genes in the network and  $K$  is the number of genes that are the neighbors of DEGs. Thus,  $p$  is the probability that the number of genes in the overlap is obtained by chance.

We analyzed the MIN using the following network properties: four centrality measures (degree, eigenvector, betweenness and closeness centrality) and the clustering coefficient (for more details see Pržulj *et al.* 2019<sup>29</sup>). The degree of a node is defined as the number of edges connected to the node and indicates the number of interactions in which the node is involved. The eigenvector centrality of a node is based on the importance of its neighbors, which is computed using the spectrum of the network and thus, identifies nodes connected to many highly connected nodes. The betweenness centrality of a node is the ratio of the number of shortest paths from all vertices to all others that pass through the node over all shortest paths and thus, nodes with high betweenness centrality are bottlenecks in the network, meaning that these nodes are more crucial in linking dense regions of the network. The closeness centrality quantifies how close a node is to all other nodes by computing the average of the lengths of the shortest paths from the node to all other nodes in the network. The clustering coefficient is the fraction of triangles that touch the node over all possible triangles in its neighborhood of the node and it captures whether the neighbors of a given node tend to cluster. We used these statistics to compare the relevant sets of genes for COVID-19 (VI, DEG, VI-unique neighbors, DEG-unique neighbors, common neighbors and background genes) and tested for statistically significant ( $p < 0.05$ ) differences in the network statistics of these node sets by using a two-sided Mann-Whitney-Wilcoxon test.

The most sensitive measures capturing the local wiring patterns around nodes in networks are based on graphlets. Graphlets are defined as connected, non-isomorphic, induced subgraphs of large networks<sup>31</sup>. Different topological positions within graphlets are characterized by different symmetry groups of nodes, called automorphism orbits<sup>71</sup>. Orbits are used to generalize the notion of the node degree: the graphlet degrees of a node are the numbers of times a node is found at each orbit position. Yaveroğlu *et al.* 2014<sup>32</sup> proved the existence of redundancies and dependencies between these orbits and proposed a set of 11 non-redundant orbits for 2- to 4-node graphlets (Supplementary Figure S5). Thus, the wiring patterns of each node in the network can be represented by using the 11-dimensional vector, called Graphlet Degree Vector (GDV), or Graphlet Degree Vector Signature, which captures the 11 non-redundant graphlet degrees of a node<sup>30</sup>. To compare the wiring patterns of the different sets of nodes (VIs, DEGs, common and unique neighbors), we therefore calculated the GDV signature for each set of nodes and compared the average signatures of the different sets.

## References

1. Ciotti, M. *et al.* COVID-19 Outbreak: An Overview. *Chemotherapy* **64**, 215–223, DOI: [10.1159/000507423](https://doi.org/10.1159/000507423) (2020).
2. Hiscott, J. *et al.* The global impact of the coronavirus pandemic. *Cytokine Growth Factor Rev.* **53**, 1–9, DOI: [10.1016/j.cytogfr.2020.05.010](https://doi.org/10.1016/j.cytogfr.2020.05.010) (2020).
3. Li, Q. *et al.* Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New Engl. J. Medicine* **382**, 1199–1207, DOI: [10.1056/NEJMoa2001316](https://doi.org/10.1056/NEJMoa2001316) (2020).
4. Guy, R. K., DiPaola, R. S., Romanelli, F. & Dutch, R. E. Rapid repurposing of drugs for covid-19. *Science* **368**, 829–830, DOI: [10.1126/science.abb9332](https://doi.org/10.1126/science.abb9332) (2020).
5. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468, DOI: [10.1038/s41586-020-2286-9](https://doi.org/10.1038/s41586-020-2286-9) (2020).
6. Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 14, DOI: [10.1038/s41421-020-0153-3](https://doi.org/10.1038/s41421-020-0153-3) (2020).
7. Wang, Y. *et al.* Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet* **395**, 1569–1578, DOI: [10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9) (2020).
8. Zou, L. *et al.* SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *New Engl. J. Medicine* **382**, 1177–1179, DOI: [10.1056/NEJMc2001737](https://doi.org/10.1056/NEJMc2001737) (2020).
9. Bai, Y. *et al.* Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* **323**, 1406–1407, DOI: [10.1001/jama.2020.2565](https://doi.org/10.1001/jama.2020.2565) (2020).
10. Ahlquist, P., Noueiry, A. O., Lee, W.-M., Kushner, D. B. & Dye, B. T. Host Factors in Positive-Strand RNA Virus Genome Replication. *J. Virol.* **77**, 8181–8186, DOI: [10.1128/jvi.77.15.8181-8186.2003](https://doi.org/10.1128/jvi.77.15.8181-8186.2003) (2003).
11. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273, DOI: [10.1038/s41586-020-2012-7](https://doi.org/10.1038/s41586-020-2012-7) (2020).

- 481 **12.** Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven  
482 Protease Inhibitor. *Cell* **181**, 271–280.e8, DOI: [10.1016/j.cell.2020.02.052](https://doi.org/10.1016/j.cell.2020.02.052) (2020).
- 483 **13.** Iwata-Yoshikawa, N. *et al.* TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine  
484 Models after Coronavirus Infection. *J. Virol.* **93**, DOI: [10.1128/jvi.01815-18](https://doi.org/10.1128/jvi.01815-18) (2019).
- 485 **14.** Blanco-Melo, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **181**,  
486 1036–1045.e9, DOI: [10.1016/j.cell.2020.04.026](https://doi.org/10.1016/j.cell.2020.04.026) (2020).
- 487 **15.** Codo, A. C. *et al.* Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-  
488  $1\alpha$ /Glycolysis-Dependent Axis. *Cell Metab.* **32**, 437–446.e5, DOI: [10.1016/j.cmet.2020.07.007](https://doi.org/10.1016/j.cmet.2020.07.007) (2020).
- 489 **16.** Xia, S. *et al.* Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion  
490 inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **30**, 343–355, DOI:  
491 [10.1038/s41422-020-0305-x](https://doi.org/10.1038/s41422-020-0305-x) (2020).
- 492 **17.** Monteil, V. *et al.* Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble  
493 Human ACE2. *Cell* **181**, 905–913.e7, DOI: [10.1016/j.cell.2020.04.004](https://doi.org/10.1016/j.cell.2020.04.004) (2020).
- 494 **18.** Sadegh, S. *et al.* Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nat. Commun.* **11**, 1–9,  
495 DOI: [10.1038/s41467-020-17189-2](https://doi.org/10.1038/s41467-020-17189-2) (2020).
- 496 **19.** Gysi, D. M. *et al.* Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19. Preprint  
497 at <http://arxiv.org/abs/2004.07229> (2020).
- 498 **20.** Malod-Dognin, N. *et al.* Towards a data-integrated cell. *Nat. Commun.* **10**, 1–13, DOI: [10.1038/s41467-019-08797-8](https://doi.org/10.1038/s41467-019-08797-8)  
499 (2019).
- 500 **21.** Gligorijević, V., Malod-Dognin, N. & Pržulj, N. Patient-specific data fusion for cancer stratification and personalised  
501 treatment. In *Pacific Symposium on Biocomputing*, vol. 21, 321–332, DOI: [10.1142/9789814749411\\_0030](https://doi.org/10.1142/9789814749411_0030) (World Scientific  
502 Publishing Co. Pte Ltd, 2016).
- 503 **22.** Ding, C., Li, T., Peng, W. & Park, H. Orthogonal nonnegative matrix tri-factorizations for clustering. In *Proceedings*  
504 *of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 2006, 126–135, DOI:  
505 [10.1145/1150402.1150420](https://doi.org/10.1145/1150402.1150420) (ACM Press, New York, New York, USA, 2006).
- 506 **23.** Wang, H., Huang, H., Ding, C. & Nie, F. Predicting protein-protein interactions from multimodal biological data sources  
507 via nonnegative matrix tri-factorization. *J. Comput. Biol.* **20**, 344–358, DOI: [10.1089/cmb.2012.0273](https://doi.org/10.1089/cmb.2012.0273) (2013).
- 508 **24.** Ding, C., He, X. & Simon, H. D. On the equivalence of nonnegative matrix factorization and spectral clustering.  
509 In *Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005*, 606–610, DOI: [10.1137/1.  
510 9781611972757.70](https://doi.org/10.1137/1.9781611972757.70) (Society for Industrial and Applied Mathematics Publications, 2005).
- 511 **25.** Žitnik, M., Janjić, V., Larminie, C., Zupan, B. & Pržulj, N. Discovering disease-disease associations by fusing systems-level  
512 molecular data. *Sci. Reports* **3**, 1–9, DOI: [10.1038/srep03202](https://doi.org/10.1038/srep03202) (2013).
- 513 **26.** Žitnik, M. & Zupan, B. Data fusion by matrix factorization. *IEEE Transactions on Pattern Analysis Mach. Intell.* **37**,  
514 41–53, DOI: [10.1109/TPAMI.2014.2343973](https://doi.org/10.1109/TPAMI.2014.2343973) (2015).
- 515 **27.** Gligorijević, V., Janjić, V. & Pržulj, N. Integration of molecular network data reconstructs Gene Ontology. *Bioinformatics*  
516 **30**, 594–600, DOI: [10.1093/bioinformatics/btu470](https://doi.org/10.1093/bioinformatics/btu470) (2014).
- 517 **28.** Hwang, T. *et al.* Co-clustering phenome-genome for phenotype classification and disease gene discovery. *Nucleic Acids*  
518 *Res.* **40**, e146–e146, DOI: [10.1093/nar/gks615](https://doi.org/10.1093/nar/gks615) (2012).
- 519 **29.** Pržulj, N. (ed.) *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical*  
520 *and Computational Scientists* (Cambridge University Press, 2019).
- 521 **30.** Milenković, T. & Pržulj, N. Uncovering biological network function via graphlet degree signatures. *Cancer Informatics* **6**,  
522 257–273, DOI: [10.4137/cin.s680](https://doi.org/10.4137/cin.s680) (2008).
- 523 **31.** Pržulj, N., Corneil, D. G. & Jurisica, I. Modeling interactome: Scale-free or geometric? *Bioinformatics* **20**, 3508–3515,  
524 DOI: [10.1093/bioinformatics/bth436](https://doi.org/10.1093/bioinformatics/bth436) (2004).
- 525 **32.** Yaveröa Lu, Ö. N. *et al.* Revealing the hidden Language of complex networks. *Sci. Reports* **4**, 1–9, DOI: [10.1038/srep04547](https://doi.org/10.1038/srep04547)  
526 (2014).
- 527 **33.** Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev.*  
528 *Genet.* **12**, 56–68, DOI: [10.1038/nrg2918](https://doi.org/10.1038/nrg2918) (2011).

- 529 **34.** Davis, D., Yaverolu, Ö. N., Malod-Dognin, N., Stojmirovic, A. & Pržulj, N. Topology-function conservation in protein-  
530 protein interaction networks. *Bioinformatics* **31**, 1632–1639, DOI: [10.1093/bioinformatics/btv026](https://doi.org/10.1093/bioinformatics/btv026) (2015).
- 531 **35.** Kroll, J. & Waltenberger, J. VEGF-A induces expression of eNOS and iNOS in endothelial cells via VEGF receptor-2  
532 (KDR). *Biochem. Biophys. Res. Commun.* **252**, 743–746, DOI: [10.1006/bbrc.1998.9719](https://doi.org/10.1006/bbrc.1998.9719) (1998).
- 533 **36.** Vuolteenaho, K. *et al.* Leptin enhances synthesis of proinflammatory mediators in human osteoarthritic cartilage-Mediator  
534 role of NO in leptin-induced PGE 2, IL-6, and IL-8 Production. *Mediat. Inflamm.* **2009**, DOI: [10.1155/2009/345838](https://doi.org/10.1155/2009/345838)  
535 (2009).
- 536 **37.** Basudhar, D. *et al.* Coexpression of NOS2 and COX2 accelerates tumor growth and reduces survival in estrogen receptor-  
537 negative breast cancer. *Proc. Natl. Acad. Sci. United States Am.* **114**, 13030–13035, DOI: [10.1073/pnas.1709119114](https://doi.org/10.1073/pnas.1709119114)  
538 (2017).
- 539 **38.** Kim, S. F., Huri, D. A. & Snyder, S. H. Medicine: Inducible nitric oxide synthase binds, S-nitrosylates, and activates  
540 cyclooxygenase-2. *Science* **310**, 1966–1970, DOI: [10.1126/science.1119407](https://doi.org/10.1126/science.1119407) (2005).
- 541 **39.** Komori, K. *et al.* Nitric oxide synthesis leads to vascular endothelial growth factor synthesis via the no/cyclic guanosine  
542 3,5-monophosphate (CGMP) pathway in human corpus cavernosal smooth muscle cells. *J. Sex. Medicine* **5**, 1623–1635,  
543 DOI: [10.1111/j.1743-6109.2008.00772.x](https://doi.org/10.1111/j.1743-6109.2008.00772.x) (2008).
- 544 **40.** Förstermann, U. & Münzel, T. Endothelial nitric oxide synthase in vascular disease: From marvel to menace. *Circulation*  
545 **113**, 1708–1714, DOI: [10.1161/CIRCULATIONAHA.105.602532](https://doi.org/10.1161/CIRCULATIONAHA.105.602532) (2006).
- 546 **41.** Jeffrey Man, H. S., Tsui, A. K. & Marsden, P. A. Nitric oxide and hypoxia signaling. In Litwack, G. (ed.) *Nitric Oxide*,  
547 vol. 96 of *Vitamins Hormones*, 161–192, DOI: [10.1016/B978-0-12-800254-4.00007-6](https://doi.org/10.1016/B978-0-12-800254-4.00007-6) (Academic Press Inc., 2014),  
548 vitamins and hormones edn.
- 549 **42.** Forsythe, J. A. *et al.* Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1.  
550 *Mol. Cell. Biol.* **16**, 4604–4613, DOI: [10.1128/mcb.16.9.4604](https://doi.org/10.1128/mcb.16.9.4604) (1996).
- 551 **43.** Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D. & Patel, A. N. Cardiovascular Disease, Drug Therapy, and Mortality in  
552 Covid-19. *New Engl. J. Medicine* **382**, e102, DOI: [10.1056/NEJMoa2007621](https://doi.org/10.1056/NEJMoa2007621) (2020).
- 553 **44.** Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**,  
554 497–506, DOI: [10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).
- 555 **45.** Manne, B. K. *et al.* Platelet gene expression and function in patients with COVID-19. *Blood* **136**, 1317–1329, DOI:  
556 [10.1182/blood.2020007214](https://doi.org/10.1182/blood.2020007214) (2020).
- 557 **46.** Rapkiewicz, A. V. *et al.* Megakaryocytes and platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in  
558 COVID-19: A case series. *EClinicalMedicine* **24**, DOI: [10.1016/j.eclinm.2020.100434](https://doi.org/10.1016/j.eclinm.2020.100434) (2020).
- 559 **47.** Ackermann, M. *et al.* Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *New Engl. J. Medicine*  
560 **383**, 120–128, DOI: [10.1056/NEJMoa2015432](https://doi.org/10.1056/NEJMoa2015432) (2020).
- 561 **48.** Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet* **395**, 1033–1034,  
562 DOI: [10.1016/S0140-6736\(20\)30628-0](https://doi.org/10.1016/S0140-6736(20)30628-0) (2020).
- 563 **49.** Aziz, M., Fatima, R. & Assaly, R. Elevated interleukin-6 and severe COVID-19: A meta-analysis. *J. Med. Virol.* **92**,  
564 2283–2285, DOI: [10.1002/jmv.25948](https://doi.org/10.1002/jmv.25948) (2020).
- 565 **50.** Merad, M. & Martin, J. C. Pathological inflammation in patients with COVID-19: a key role for monocytes and  
566 macrophages. *Nat. Rev. Immunol.* **20**, 355–362, DOI: [10.1038/s41577-020-0331-4](https://doi.org/10.1038/s41577-020-0331-4) (2020).
- 567 **51.** Ahmed, M. U. *et al.* Neurological Manifestations of COVID-19 (SARS-CoV-2): A Review. *Front. Neurol.* **11**, 518, DOI:  
568 [10.3389/fneur.2020.00518](https://doi.org/10.3389/fneur.2020.00518) (2020).
- 569 **52.** Wu, Y. *et al.* Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain, Behav. Immun.*  
570 **87**, 18–22, DOI: [10.1016/j.bbi.2020.03.031](https://doi.org/10.1016/j.bbi.2020.03.031) (2020).
- 571 **53.** Andrew, P. J. & Mayer, B. Enzymatic function of nitric oxide synthases. *Cardiovasc. Res.* **43**, 521–531, DOI: [10.1016/  
572 S0008-6363\(99\)00115-7](https://doi.org/10.1016/S0008-6363(99)00115-7) (1999).
- 573 **54.** Nadkarni, G. N. *et al.* Anticoagulation, Mortality, Bleeding and Pathology Among Patients Hospitalized with COVID-19:  
574 A Single Health System Study. *J. Am. Coll. Cardiol.* **76**, 1815–1826, DOI: [10.1016/j.jacc.2020.08.041](https://doi.org/10.1016/j.jacc.2020.08.041) (2020).
- 575 **55.** Gewaltig, M. T. & Kojda, G. Vasoprotection by nitric oxide: Mechanisms and therapeutic potential. *Cardiovasc. Res.* **55**,  
576 250–260, DOI: [10.1016/S0008-6363\(02\)00327-9](https://doi.org/10.1016/S0008-6363(02)00327-9) (2002).

- 577 **56.** Jäger, R., Groneberg, D. & Friebe, A. Role of NO/cGMP signalling in VEGF-mediated angiogenesis. *BMC Pharmacol.*  
578 **11**, 1–1, DOI: [10.1186/1471-2210-11-s1-p35](https://doi.org/10.1186/1471-2210-11-s1-p35) (2011).
- 579 **57.** Lee, H. T., Chang, Y. C., Tu, Y. F. & Huang, C. C. VEGF-A/VEGFR-2 signaling leading to cAMP response element-binding  
580 protein phosphorylation is a shared pathway underlying the protective effect of preconditioning on neurons and endothelial  
581 cells. *J. Neurosci.* **29**, 4356–4368, DOI: [10.1523/JNEUROSCI.5497-08.2009](https://doi.org/10.1523/JNEUROSCI.5497-08.2009) (2009).
- 582 **58.** García-Serradilla, M., Risco, C. & Pacheco, B. Drug repurposing for new, efficient, broad spectrum antivirals. *Virus Res.*  
583 **264**, 22–31, DOI: [10.1016/j.virusres.2019.02.011](https://doi.org/10.1016/j.virusres.2019.02.011) (2019).
- 584 **59.** Martinez, J. P., Sasse, F., Brönstrup, M., Diez, J. & Meyerhans, A. Antiviral drug discovery: Broad-spectrum drugs from  
585 nature. *Nat. Prod. Reports* **32**, 29–48, DOI: [10.1039/c4np00085d](https://doi.org/10.1039/c4np00085d) (2015).
- 586 **60.** Adalja, A. & Inglesby, T. Broad-Spectrum Antiviral Agents: A Crucial Pandemic Tool. *Expert. Rev. Anti-Infective Ther.*  
587 **17**, 467–470, DOI: [10.1080/14787210.2019.1635009](https://doi.org/10.1080/14787210.2019.1635009) (2019).
- 588 **61.** Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541, DOI:  
589 [10.1093/nar/gky1079](https://doi.org/10.1093/nar/gky1079) (2019).
- 590 **62.** Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways,  
591 diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361, DOI: [10.1093/nar/gkw1092](https://doi.org/10.1093/nar/gkw1092) (2017).
- 592 **63.** Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**,  
593 D1074–D1082, DOI: [10.1093/nar/gkx1037](https://doi.org/10.1093/nar/gkx1037) (2018).
- 594 **64.** Nikolova, N. & Jaworska, J. Approaches to Measure Chemical Similarity - A Review. *QSAR Comb. Sci.* **22**, 1006–1026,  
595 DOI: [10.1002/qsar.200330831](https://doi.org/10.1002/qsar.200330831) (2004).
- 596 **65.** Wang, F., Li, T. & Zhang, C. Semi-supervised clustering via matrix factorization. In *Proceedings of the 2008 SIAM*  
597 *International Conference on Data Mining*, vol. 1, 1–12, DOI: [10.1137/1.9781611972788.1](https://doi.org/10.1137/1.9781611972788.1) (Society for Industrial and  
598 Applied Mathematics Publications, 2008).
- 599 **66.** Brunet, J. P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular pattern discovery using matrix  
600 factorization. *Proc. Natl. Acad. Sci. United States Am.* **101**, 4164–4169, DOI: [10.1073/pnas.0308531101](https://doi.org/10.1073/pnas.0308531101) (2004).
- 601 **67.** Kodinariya, T. M. & Makwana, P. R. Review on determining number of Cluster in K-Means Clustering. *Int. J. Adv. Res.*  
602 *Comput. Sci. Manag. Stud.* **1**, 2321–7782 (2013).
- 603 **68.** Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update).  
604 *Nucleic Acids Res.* **47**, W191–W198, DOI: [10.1093/nar/gkz369](https://doi.org/10.1093/nar/gkz369) (2019).
- 605 **69.** Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29, DOI: [10.1038/75556](https://doi.org/10.1038/75556)  
606 (2000).
- 607 **70.** Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple  
608 Testing. *J. Royal Stat. Soc. Ser. B (Methodological)* **57**, 289–300, DOI: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x) (1995).
- 609 **71.** Pržulj, N. Biological network comparison using graphlet degree distribution. *Bioinformatics* **23**, DOI: [10.1093/](https://doi.org/10.1093/bioinformatics/btl301)  
610 [bioinformatics/btl301](https://doi.org/10.1093/bioinformatics/btl301) (2007).

## 611 Acknowledgements

612 This work was supported by the European Research Council (ERC) Consolidator Grant 770827 and the Spanish State Research  
613 Agency AEI 10.13039/501100011033 grant number PID2019-105500GB-I00.

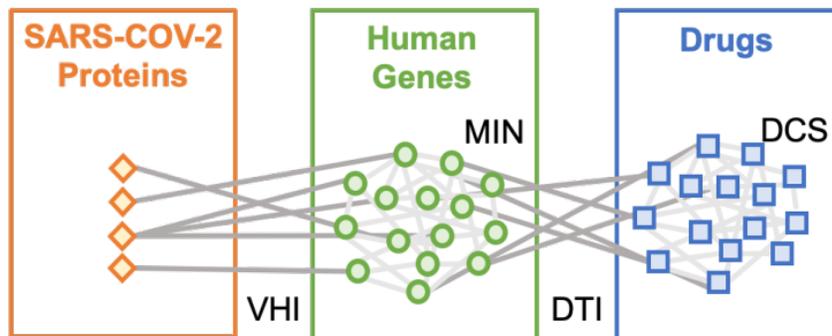
## 614 Author contributions statement

615 C.Z. implemented the data fusion framework, explored the candidate drug-target interactions and wrote the manuscript. A.X.  
616 analyzed the molecular network, explored the candidate drug-target interactions, and contributed to writing the manuscript.  
617 R.B. explored the candidate genes and drug-target interactions, directed the project and contributed to writing the manuscript.  
618 N.M-D. and N.P. conceived and directed the study and contributed to writing the manuscript. All the authors analyzed the  
619 results and reviewed the manuscript.

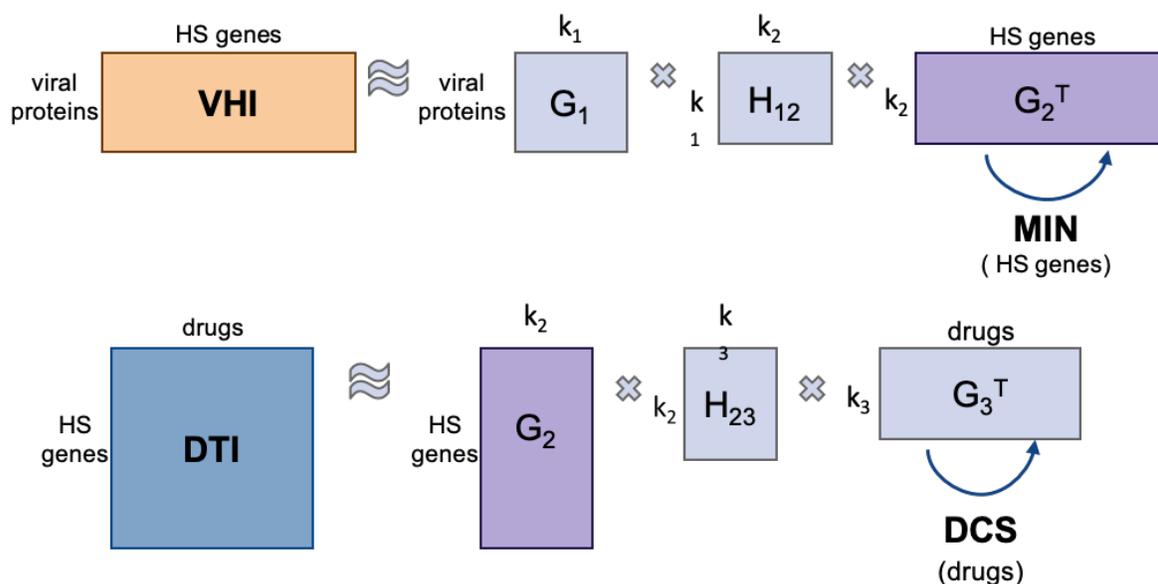
## 620 Additional information

621 The authors declare no competing interests.

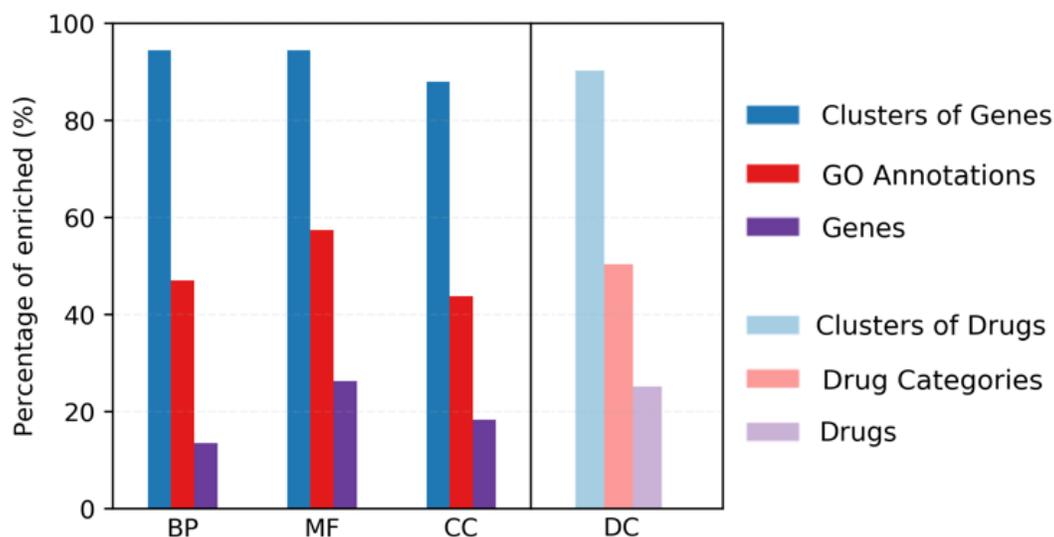
## a) Data and Networks



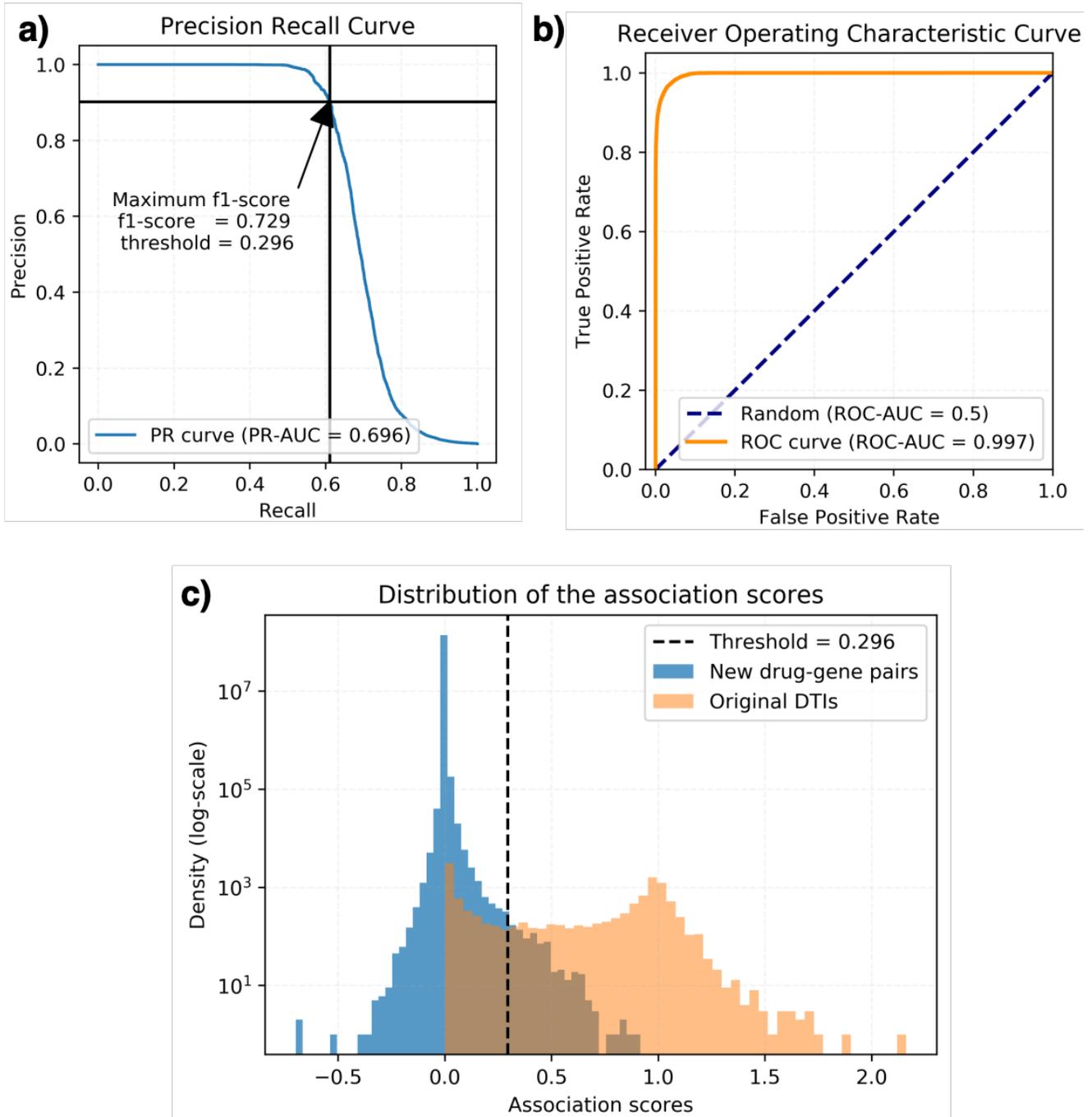
## b) Graph regularized non-negative matrix tri-factorization



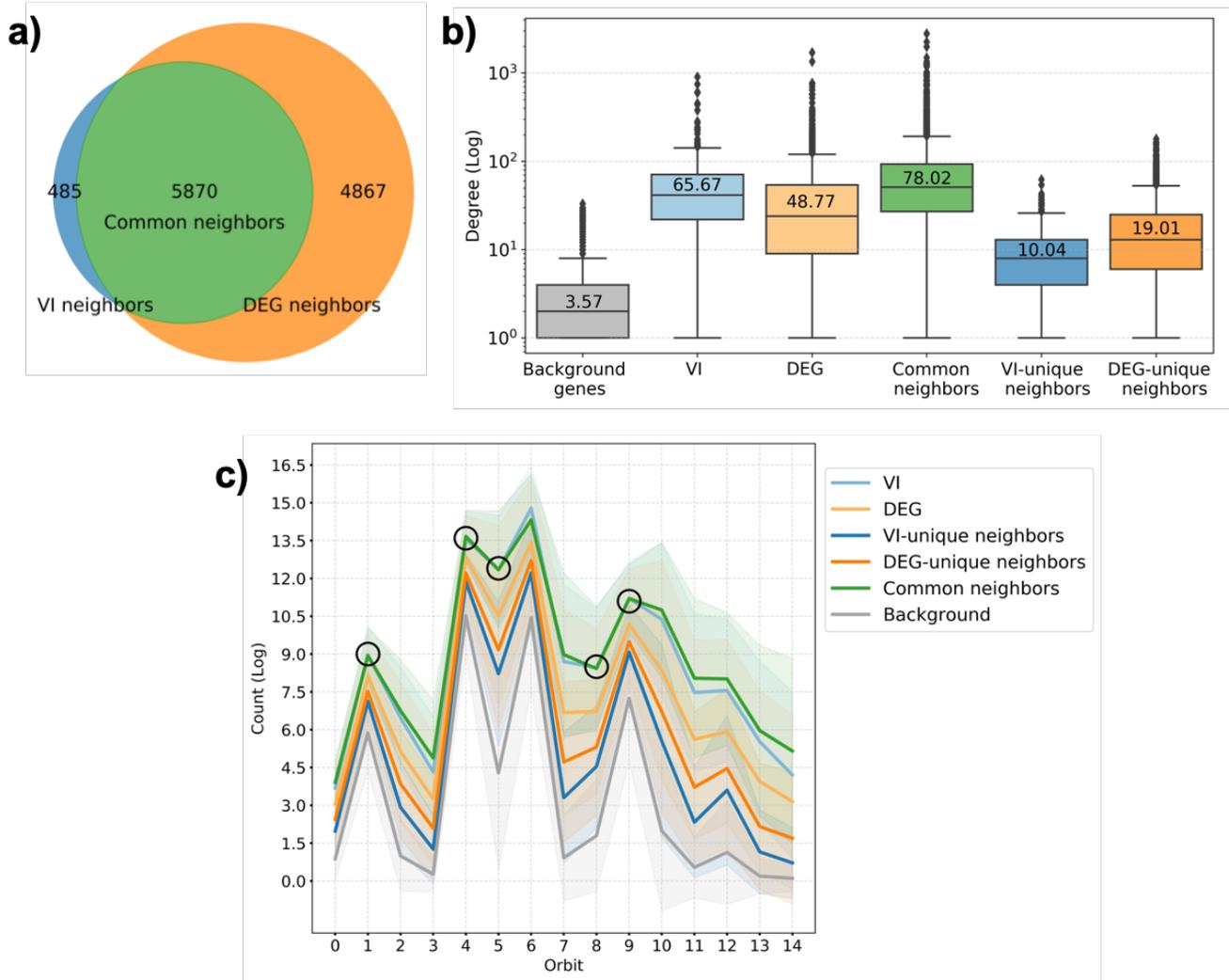
**Figure 1. Illustration of the data and framework.** **a)** Schematic illustration of datasets used in this study. Three data types are represented: SARS-CoV-2 proteins (in orange), human genes (in green) and drugs (in blue). Two relational datasets connect different types of data: virus-host protein-protein interactions (VHIs) and drug-target interactions (DTIs). Network structural knowledge from these data types is contained in the molecular interaction network (MIN) and the drug chemical similarity (DCS) network. **b)** Graph-regularized non-negative matrix tri-factorization (GNMTF) used for fusing the VHIs, DTIs, MIN and DCS networks. The matrix factor  $G_2$  is shared across decompositions to simultaneously decompose the VHI and DTI networks. Network structure (topology) information from the MIN and DCS networks are incorporated into the data fusion by using two regularization terms (illustrated by arcs with arrows). The parameters  $k_1$ ,  $k_2$  and  $k_3$  indicate the numbers of clusters of viral proteins, human genes and drugs, respectively.



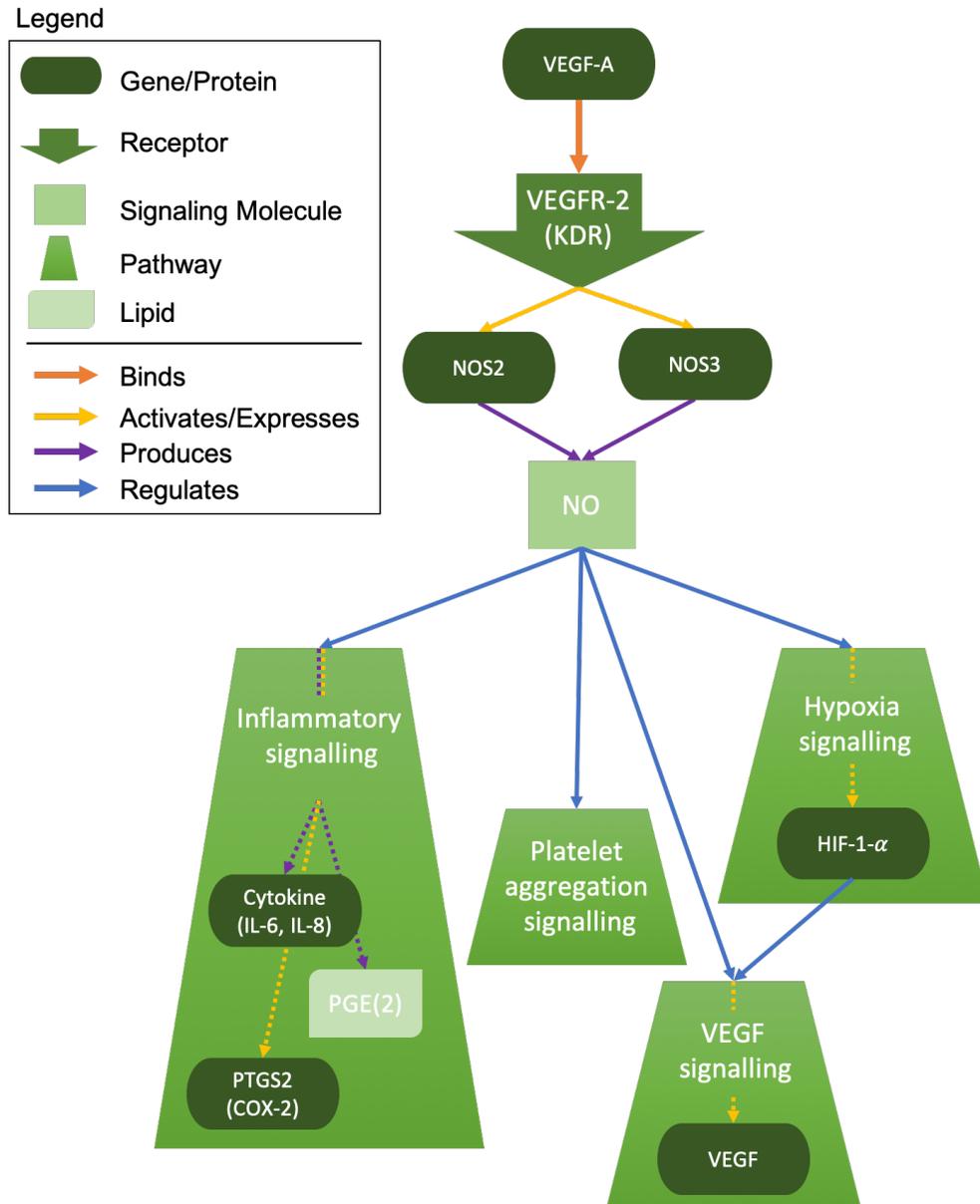
**Figure 2. Enrichment analysis for assessing the functional relevance of the gene and drug clusters obtained by the framework.** The gene clusters are analyzed by using GO term annotations for the three domains: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC); and the drug clusters are analyzed by using “Drug Categories” (DC) from DrugBank (horizontal axis). The probability that an annotation is enriched in a cluster was computed using a hypergeometric test. Then, we computed three types of enrichments: the percentage of clusters of genes (drugs) having at least one of their genes (drugs) enriched (in blue); the percentage of GO annotations (Drug Categories) enriched (in red); and the percentage of genes (drugs) having at least one of their annotations enriched in their clusters over all annotated genes (drugs) (in purple).



**Figure 3. Prediction of new DTIs.** **a)** Precision-Recall (PR) and **b)** Receiver Operating Characteristic (ROC) curves for assessing the ability of the framework to capture the known drug-target interactions accurately. AUC - area under the curve. **c)** Distribution of the association scores of the reconstructed matrix, for the original DTIs (orange) and the new drug-gene pairs obtained by the matrix completion property of GNMTF (blue). New drug-gene pairs on the right side of the threshold (dashed line) were considered to be newly predicted DTIs.



**Figure 4. Network properties of the molecular interaction network (MIN), focusing on the gene sets. a)** Venn Diagram showing the overlap between VI neighbor and DEG neighbor gene sets. **b)** Comparison of the average degrees (given on the vertical axis) for the different gene sets (given on the horizontal axis) based on their degrees in the MIN. All pair-wise comparisons between these are statistically significant with  $p < 0.0001$ , using two-sided Mann-Whitney-Wilcoxon test. **c)** GDV signatures for the VI, DEG, VI-unique neighbors, DEG-unique neighbors, common neighbors and background gene sets. All pair-wise comparisons between the counts (on the vertical axis) of the orbits (denoted by 0 to 14 on the horizontal axis) of common neighbors and the rest of the gene sets are statistically significant with  $p < 0.05$ , using two-sided Mann-Whitney-Wilcoxon test, except for the orbits marked with a black circle between common neighbors and VI gene set.



**Figure 5. Illustration of the biological mechanisms targeted by the common neighbor DTIs.** The illustration shows that many of the pathways are tied to NO and VEGF signaling. NO production is directly related to VEGFR-2 receptor and at the same time NO regulates VEGF signaling pathway among others: inflammatory signaling, hypoxia signaling and platelet aggregation.