



OPEN

## Drug repositioning by merging active subnetworks validated in cancer and COVID-19

Marta Lucchetta<sup>1,2</sup> & Marco Pellegrini<sup>1</sup>✉

Computational drug repositioning aims at ranking and selecting existing drugs for novel diseases or novel use in old diseases. *In silico* drug screening has the potential for speeding up considerably the shortlisting of promising candidates in response to outbreaks of diseases such as COVID-19 for which no satisfactory cure has yet been found. We describe DrugMerge as a methodology for preclinical computational drug repositioning based on merging multiple drug rankings obtained with an ensemble of disease active subnetworks. DrugMerge uses differential transcriptomic data on drugs and diseases in the context of a large gene co-expression network. Experiments with four benchmark diseases demonstrate that our method detects in first position drugs in clinical use for the specified disease, in all four cases. Application of DrugMerge to COVID-19 found rankings with many drugs currently in clinical trials for COVID-19 in top positions, thus showing that DrugMerge can mimic human expert judgment.

Among the strategies pharmacological science adopts for tackling diseases (either new or old), there are: the development of new drugs, the development of vaccines, and the repurposing of existing drugs. While the development of vaccines and novel drugs are often effective, they are also expensive and time-consuming. Thus repurposing of existing drugs is a cheaper and faster route to explore when one wishes to hedge one's bets<sup>1</sup>.

As data about drugs and diseases accumulate in databases comprising many aspects: genomic, transcriptomic, phenotypic, clinical, and epidemiological, it becomes feasible and desirable to employ this body of accumulated knowledge within a computerized system so as to have a pre-screen in silico of the candidate drugs to be repurposed for a specific disease<sup>2</sup>. These shortlisted drugs are in reduced number with respect to the initial pool of drugs and thus are more amenable to the subsequent steps of the full pipeline (including *in vitro* and *in vivo* experiments, clinical trials, and eventual deployment in clinical practice). For existing drugs often their safety profile for humans is already known, therefore once the initial *in vitro* and *in vivo* assays are positive, they can be moved directly to clinical trials of phase II or phase III (thus skipping phases 0 and I), and accelerating the drug approval lifecycle.

The recent pandemic sparked by the SARS-CoV-2 virus has placed drug repurposing in the spotlight because of the pressure on the public health systems to find cures for acute cases of patients infected by SARS-CoV-2. At the moment of writing, remdesivir has been approved by regulatory bodies to treat acute COVID-19 cases needing hospitalization. No other drug or treatment has passed all phases of clinical trials leading to full approval by the Federal Drug Administration (FDA) or the European Medicines Agency (EMA). Initial drug investigations on COVID-19 were based on the accumulated experience of the efficacy of drugs *in vitro* and *in vivo* experiments for SARS-CoV, MERS-CoV, and other coronaviruses. Computerized drug repurposing however aims at a more systematic approach to candidate drug ranking and selection. Moreover, while having accrued data from other viruses belonging to the same family is a bonus, we should not always rely on having such data handy. In a longer perspective, both the possible variations in the behavior of SARS-CoV-2, due to genetic mutations, or the emergence of new human viral diseases due to cross-species transmission, imply that an easy to use, accurate, and fast pre-screening and ranking of repurposable drugs aimed at countering any new viral threats is needed. Our study is a contribution in this direction. Network-based drug repurposing is a recent new approach to drug repurposing that may increase the effectiveness of drug repurposing pipelines. In particular network-based drug repurposing can merge many 'omics' data sets in a unified framework, thus such approach may be more robust and capable of including systemic biological effects<sup>3</sup>.

In Lucchetta et al.<sup>4</sup> we established Core&Peel as a competitive method for building Disease Active subnetworks (DAS) compared to several state-of-the-art algorithms in several known benchmark diseases. Moreover,

<sup>1</sup>Institute of Informatics and Telematics (IIT), CNR, Pisa 56124, Italy. <sup>2</sup>Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena 53100, Italy. ✉email: marco.pellegrini@iit.cnr.it

we applied Core&Peel and the other algorithms to build a collection of active subnetworks from gene expression data of human tissues infected by COVID-19. Subsequent pathway enrichment analysis revealed that in a COVID-19 infection are activated pathways similar to those activated in other diseases (mostly viral infections). These observations had implications for the choice of existing drugs likely to have an impact on COVID-19 to be short-listed for in vitro experiments and eventually go on to clinical trials. Although often the connection between a disease-related pathway and a repurposable drug was obvious in the specific context, sometimes the connection was hard to establish since, for example, for a relevant disease-related pathway too many drugs are known to be effective. Moreover, there was no quantitative way to prioritize the potentially repurposable drugs obtained from this type of qualitative analysis.

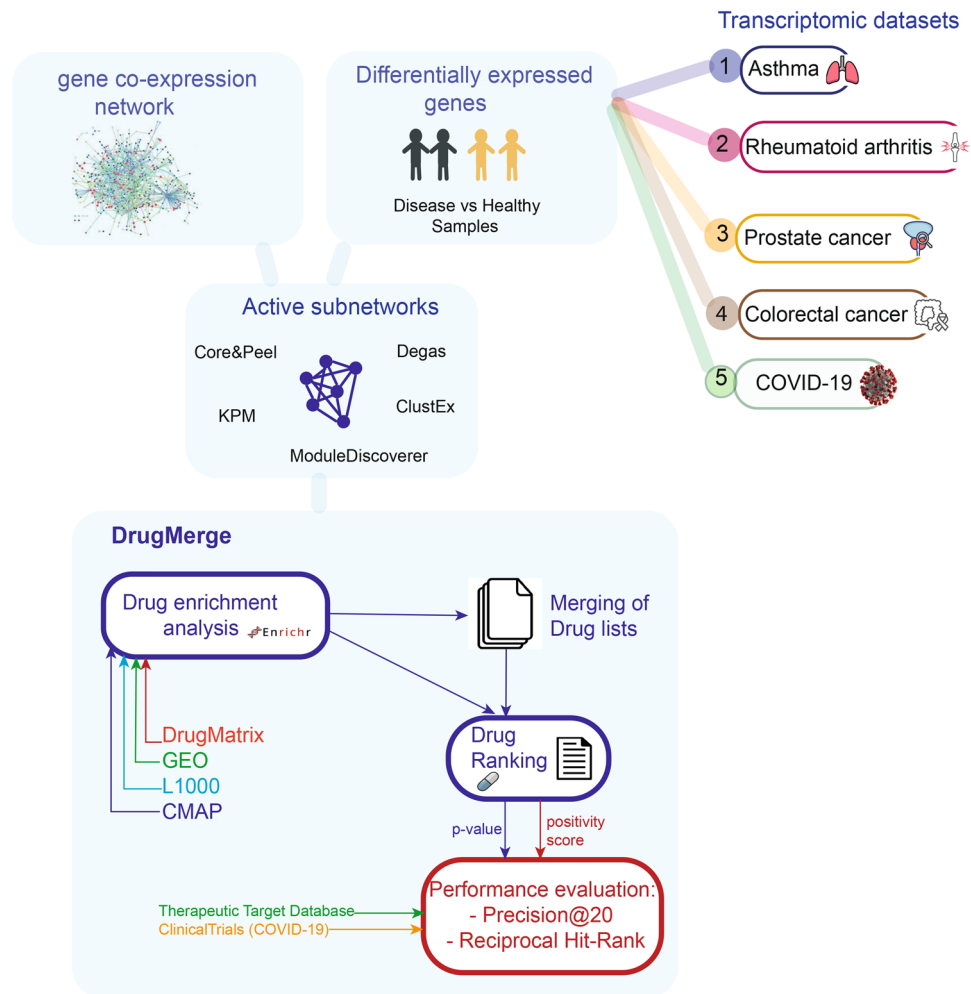
In this paper, building upon Lucchetta et al.<sup>4</sup>, we describe a systematic way to rank repurposable drugs for a specific disease, taking advantage of multiple active subnetwork algorithms and multiple gene expression data sets. We exploit several drug databases to identify drugs that can affect the expression of the genes included in each disease active subnetwork and thus can contrast the disease. Firstly, the drug ranking is based on drug significance (i.e. p-value). After that, we explore the ranking based on 'positivity score', which measures the antagonist activity of the drug onto the disease. We also measure the improvements attainable by merging more drug lists that come from different DAS algorithms.

In our strategy we espouse the point of view that it is convenient to use generalist, off-the-shelf, Disease Active subnetworks (DAS) algorithms rather than developing ad-hoc algorithms for specific networks or specific data. Generalist methods have the advantage of extensive prior use and validation in a variety of tasks, not necessarily related to drug discovery. We do not believe that a single approach or algorithm to building DAS can be a 'silver bullet' for all possible drug repositioning tasks, thus our aim is not that of indicating a priori a winner among a pool of competing algorithms and their combinations. Our methodology is based on using golden standard data (e.g. drugs already validated in clinical use) as a benchmark to determine the best performing algorithm or combination of algorithms to produce a list with validated drugs in high ranking position. The hunch is that the other not-yet-validated drugs high in such a ranking are good drug candidates to shortlist for further downstream analysis. Among the chosen methods, Core&Peel and ModuleDiscoverer are special since they produce a covering of the Disease Active subnetworks by a collection of dense ego-networks which are, in turns, used to determine the positivity score. The remaining three methods do not have this topological property. We have chosen a pool of well known and usable algorithms as a proof of concept. It is likely that DrugMerge will be improved in future releases by adding carefully chosen additional methods for DAS, preferably based on novel principles and insights.

We use known clinically approved drugs for benchmark diseases, and trial drugs for COVID-19 as golden standard. We apply two measures (called precision@20 and reciprocal hit-rank) to assess how well DrugMerge performs with respect to the golden standard. We show that our method has better measurable performance than C-map<sup>5</sup>, a widely used drug ranking portal, as well as COVID-19 specific drug rankings in Taguchi et al.<sup>6</sup> and Mousavi et al.<sup>7</sup>. The quality of our ranking results on COVID-19 is arguably comparable to those in Gysi et al.<sup>8</sup> and Zhou et al.<sup>9</sup>, which use however quite different strategies. The whole workflow of our method is shown in Fig. 1.

Ruiz et al.<sup>10</sup> observe that "...a drug's effectiveness can often be attributed to targeting genes that are distinct from disease-associated genes but that affect the same functional pathways", in contrast to "... existing approaches assume that, for a drug to treat a disease, the proteins targeted by the drug need to be close to or even need to coincide with the disease-perturbed proteins". Ruiz et al. thus build a multi-layered interaction graph comprising proteins, drugs, diseases, and functional pathways (extracted from the GO repository), and devise a random-walk diffusion algorithm to connect the drug nodes and the disease nodes of the model through paths that include the functional nodes of the graph.

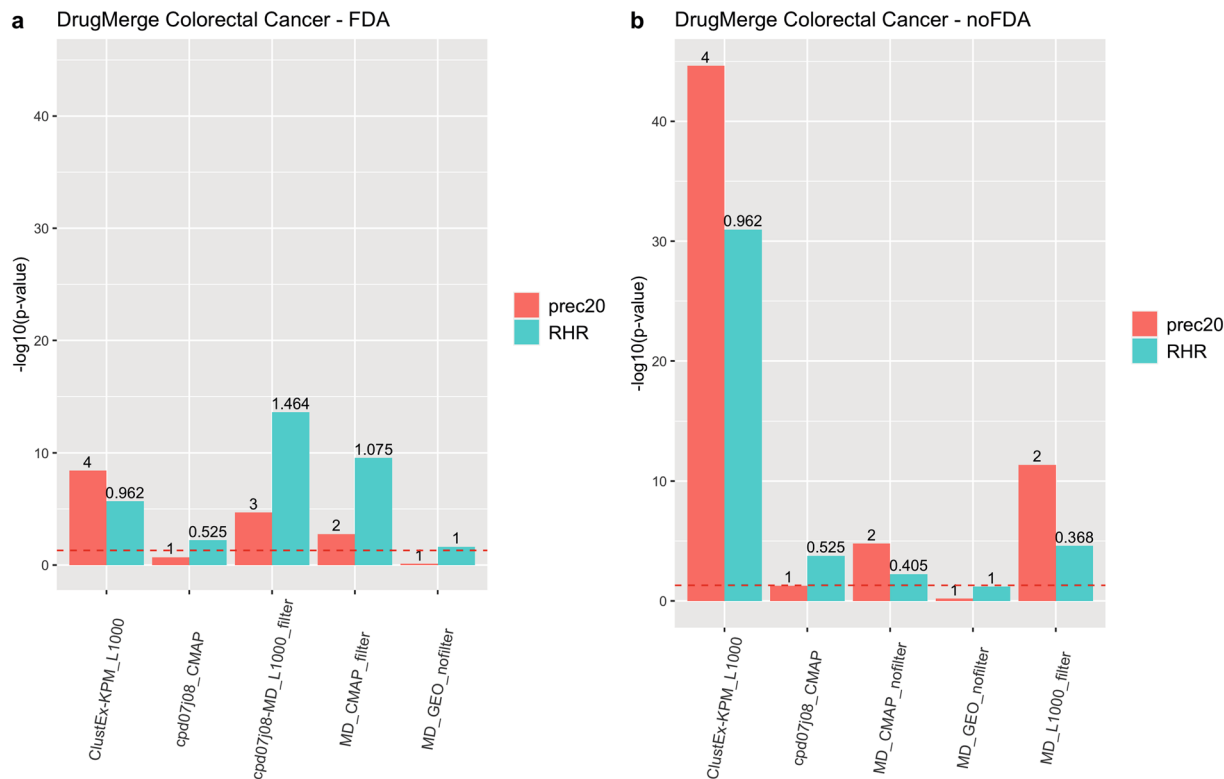
Our approach is intermediate between the functional-based method advocated by Ruiz et al.<sup>10</sup> and protein proximity based methods<sup>8</sup>. We do not represent GO pathways as functional nodes in the model, instead, we rely on a large and comprehensive gene co-expression network to provide the functional association in an implicit way. Dense subgraphs of the co-expression network will be formed by genes that have a coherent pattern of being up/down expressed, and thus are likely to be involved in specific cellular processes, either directly, by expressing binding proteins, or indirectly through cascading regulatory action chains, possibly involving also ncRNA and other regulatory elements not explicitly modelled. All algorithms used in this paper to build Active Network will take advantage of the dense neighborhood of the co-expression network, either directly (e.g. Core&Peel, ModuleDiscoverer) or indirectly via the augmented local graph connectivity (KeyPathwayMiner, Degas, ClustEx). In contrast with the proximity-based methods, we do not define any ad hoc explicit distance function over a network, since the Active Network algorithms already provide a clear-cut demarcation for the part of the global gene network onto which a perturbation should act, thus we can use directly indices for measuring module enrichment or incidence of the drug's differentially expressed genes (DEG) onto the disease's Active Network. We follow a network-based approach to drug repositioning and we focus mainly on the human transcriptional response to the disease, rather than the interaction between the disease agent (e.g. a virus) and the transcriptional response to it. In the case of COVID-19, also it is known that an over-response by the immune system is often the main cause of death in acute cases<sup>11,12</sup>. Our strategy for COVID-19 is thus to prioritize drugs that act on a disease active subnetwork embedded into a global human gene co-expression network, to counter the virus' global effects. This approach is complementary to a different one that aims at drugs interfering with the interactions of the host-proteins with the viral-protein. While the former approach aims at controlling the global transcriptional response of human tissues to the virus, the latter favors drugs blocking key aspects of the virus actions in entering the host cells, and activating the cells' biochemical machinery for reproduction.



**Figure 1.** Overview of the workflow. The first part of the figure shows how we generated the data inputs of DrugMerge. These data have been created in Lucchetta et al.<sup>4</sup>, and we reported them here for sake of completeness. We used several transcriptomic data to identify differentially expressed genes and the gene co-expression network to find one active subnetwork for each method used (Core&Peel, ModuleDiscoverer, Degas, ClustEx, and KeyPathwayMiner). For each of them, we performed the drug enrichment analysis on four drug perturbation databases: DrugMatrix, GEO, L1000, and CMAP. After that, we ranked the drugs according to the p-value or the positivity score. We also merged more drug lists, which come from different algorithms for the same disease and we applied the same ranking strategy. Finally, we evaluated the DrugMerge performance through the Precision@20 and the Reciprocal Hit-Rank using the golden-standard drugs, collected in the Therapeutic Target Database, and ClinicalTrials for COVID-19. The image has been created using Adobe Illustrator.

## Results

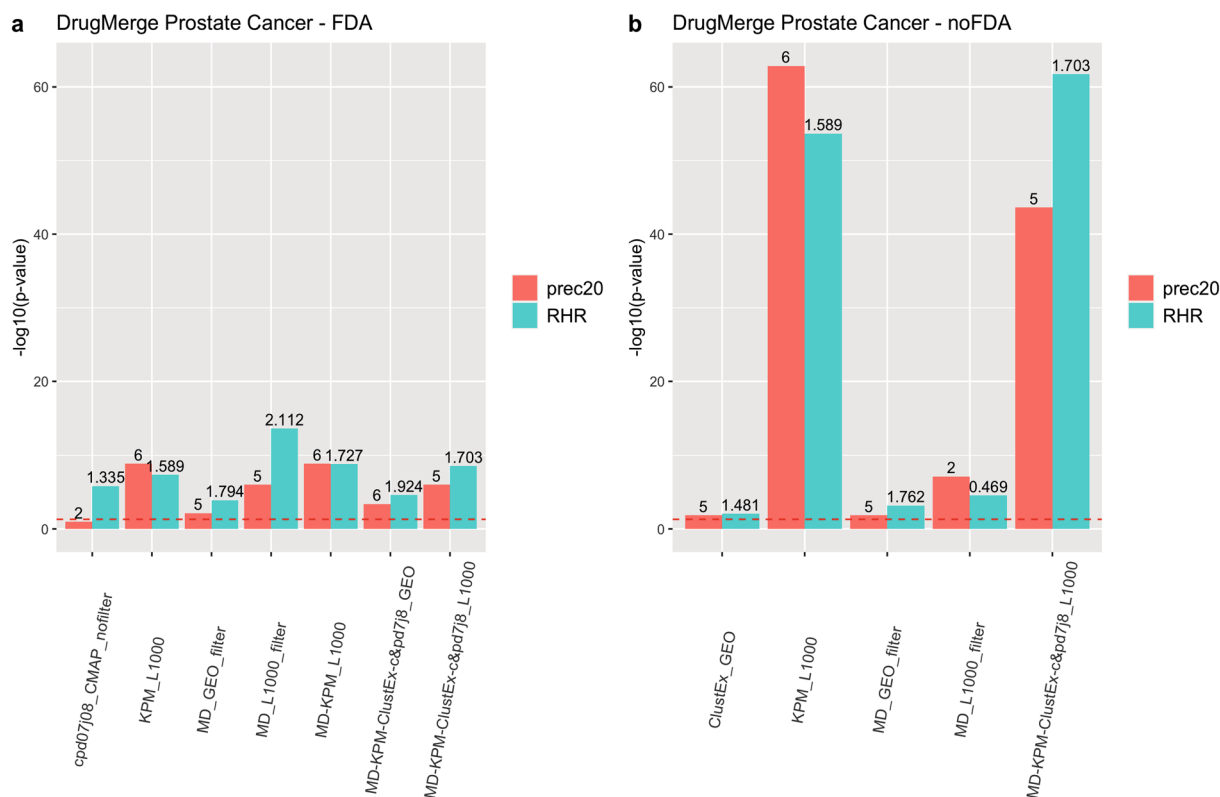
We compute as main quality measures the reciprocal hit ranking (RHR) and the precision at 20 (precision@20) of a ranking with respect to a golden standard: the *Therapeutic Target Database* (TTD)<sup>13</sup> records for four benchmark diseases (asthma, rheumatoid arthritis, colorectal cancer, and prostate cancer), and the trials.gov records for COVID-19. Each measure is then normalized via a z-score, and corresponding p-value, referred to the distribution of quality measures given by taking repeatedly random permutations of the input drug data set. The results are shown in Figs. 2, 3, 4, 5, 6 and 7, Supplementary Figs. S1 and S2 (in Sects. 4 and 5 of Supplementary Material), and Supplementary Tables S1–S8. For DrugMerge we report only significant results at  $p\text{-value} \leq 0.05$  for at least one of the two quality measures. The DrugMerge method could attain significant results for all four benchmark diseases on at least one Drug Database (GEO, DrugMatrix, CMAP, L1000), and often on more than one. Moreover, as reported in the actual rankings in Supplementary Table S7, for the four benchmark data sets the reported drug ranking always has a drug in clinical use at the top of the list for the chosen Drug data set (attaining the best p-value). In Fig. 4 we compare, using the p-value, the performance of DrugMerge versus the CMAP algorithm<sup>5</sup> on the four benchmark diseases (see “Description of CMAP algorithm” section in Supplementary Material). DrugMerge has always a better performance in terms of p-value, even when the same CMAP drug database is used also by DrugMerge. Next, we discuss in detail the results for colorectal cancer,



**Figure 2.** DrugMerge performance on colorectal cancer data. The bars represent the  $-\log_{10}(p\text{-value})$  with respect to the precision@20 (red), and the RHR (light blu). On the x-axis, different algorithms or combinations of them are shown. The numbers on the top of the bars show the absolute values of precision@20 or RHR. The dotted red line represents the limit of significance ( $-\log_{10}(0.05)$ ). All the bars above the line show a significant p-value. The (a) panel represents the DrugMerge performance when only FDA-approved drugs are considered; (b) when all drugs (without any FDA filtering) are considered. The plot has been generated by the *ggplot2*<sup>81</sup> R package.

prostate cancer, and COVID-19. The results of asthma and rheumatoid arthritis are reported in Supplementary Material (Sects. 4, 5).

**DrugMerge results on colorectal cancer.** Figure 2 and Supplementary Table S3 show that DrugMerge finds significant rankings in several drug databases (L1000, CMAP, GEO), using several algorithms, both when all drugs are considered or when only FDA-approved drugs are considered. For the L1000 dataset Supplementary Table S7, we find in the first position the clinically relevant drug erlotinib. Other drugs in clinical use according to the TTD records that appear in the top 20 positions are lapatinib and sunitinib. Among the top positions in the ranking, we find drugs with known effects in colorectal cancer patients, related animal models, or cell lines, and one of these went into clinical trial stage: mitoxantrone<sup>14</sup>, sorafenib<sup>15</sup>, sunitinib (in clinical trial <https://clinicaltrials.gov/ct2/show/NCT00961571>), dasatinib<sup>16</sup>, sirolimus (rapamycin)<sup>17</sup>, azacitidine<sup>18</sup>, imatinib<sup>19</sup>, tamoxifen<sup>20</sup>, doxorubicin<sup>21</sup>, naltrexone<sup>22</sup>, and digoxin<sup>23</sup>. Mitoxantrone has been tested in 13 patients with advanced or metastatic colon carcinoma and the authors observed that mitoxantrone was moderately effective<sup>14</sup>. Sorafenib is a kinase inhibitor drug approved for the treatment of some cancers. It is known that sorafenib has a strong antitumor and antiangiogenic effect on the colorectal cancer cell line<sup>15</sup>. Scott et al.<sup>16</sup> demonstrated that dasatinib had significant anti-proliferative activity in a subset of a colorectal cancer cell line, even if dasatinib is currently being studied in combination with chemotherapy, as its use as a single agent appears limited. Wagner et al.<sup>17</sup> studied the effects of rapamycin in mice and showed that it was able to suppress advanced stage colorectal cancer. Azacitidine has been tested in combination with entinostat in metastatic colorectal cancer patients, resulting in a tolerable therapy<sup>18</sup>. Imatinib is an oral chemotherapy medication used to treat cancer and it has been proved that it inhibits proliferation of colon cancer cells<sup>19</sup>. Kuruppo et al.<sup>20</sup> demonstrated that tamoxifen has a potent inhibitory action on colorectal cancer metastases in a murine model. Doxorubicin is a chemotherapy medication used to treat cancer. In Lee et al.<sup>21</sup> the authors propose doxorubicin-loaded oligonucleotides attached to gold nanoparticles as a drug delivery system for cancer chemotherapy. They tested this approach in colon cancer cell lines, demonstrating the drug's efficacies such as in vitro cytotoxicity. Naltrexone is primarily used to manage alcohol or opioid dependence. Ma et al.<sup>22</sup> explored the inhibitory effect of low-dose naltrexone on colorectal cancer progression in vivo and in vitro. Digoxin is used to treat various heart conditions such as atrial fibrillation, and heart failure. A recent study has examined the effects of digoxin on two kinds of colorectal cancer cells. This study suggests that digoxin has the potential to be applied as an antitumor drug via inhibiting proliferation and metastasis<sup>23</sup>.



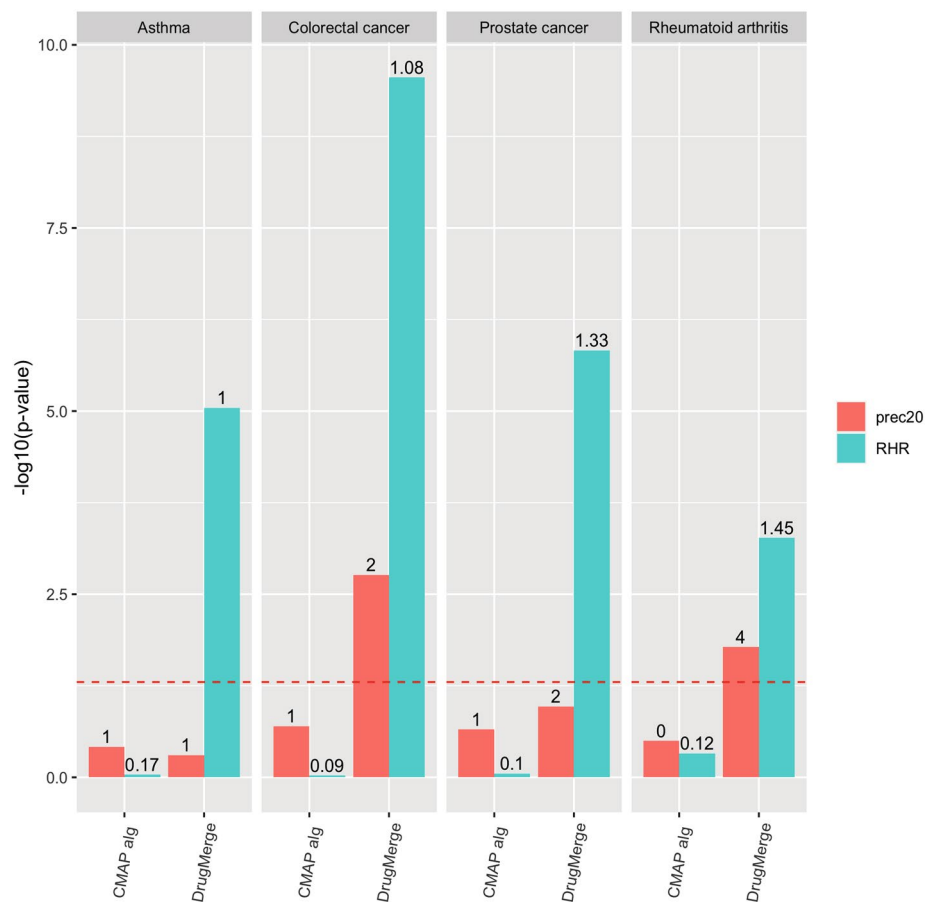
**Figure 3.** DrugMerge performance on Prostate Cancer data. The bars represent the  $-\log_{10}(p\text{-value})$  with respect to the precision@20 (red), and the RHR (light blue). On the x-axis, different algorithms or combinations of them are shown. The numbers on the top of the bars show the absolute values of precision@20 or RHR. The dotted red line represents the limit of significance ( $-\log_{10}(0.05)$ ). All the bars above the dotted line show a significant p-value. The (a) panel represents the DrugMerge performance when only FDA-approved drugs are considered; (b) when all drugs (without any FDA filtering) are considered. The plot has been generated by the *ggplot2*<sup>81</sup> R package.

For the remaining drugs in the top twenty positions, we focus on their action mechanism (i.e. we determine the main pathways affected by the drug having the desired pharmacological effect), and we find some links with colorectal cancer: epirubicin<sup>24</sup>, everolimus<sup>25</sup>, itavastatin<sup>26</sup>, and dactinomycin<sup>27</sup>. Epirubicin is a chemotherapy agent, intercalating into DNA and inhibiting topoisomerase II, thereby inhibiting DNA replication. In Ref.<sup>24</sup> the authors demonstrate topoisomerase IIa gene and protein alterations in colorectal cancer and show that increased expression is associated with pathologically advanced disease. Everolimus is a medication used as an immunosuppressant and in the treatment of several tumors. Its main mechanism of action is mTOR inhibition. Thiem et al.<sup>25</sup> demonstrated that the mTORC1 inhibition reduces the inflammation associated with the gastrointestinal tumor. Itavastatin (or pitavastatin) belongs to the class of statins, which are inhibitors of HMG-CoA reductase, the enzyme that catalyzes the first step of cholesterol synthesis. A meta-analysis work<sup>26</sup> showed that statin use is a protective factor for colorectal cancer prognosis, and it might be significantly associated with lower overall mortality. Finally, dactinomycin is a chemotherapy medication and can inhibit transcription. Recent studies have revealed crucial roles of transcription regulation in colorectal cancer development. In particular, deregulation of transcription factors is pretty frequent, and drastic changes in gene expression profiles play fundamental roles in the multistep process of tumorigenesis<sup>27</sup>.

**DrugMerge results on prostate cancer.** Figure 3 and Supplementary Table S4 show that DrugMerge finds significant rankings in several drug databases (L1000, GEO) for prostate cancer, using several algorithms, both when all drugs are considered or when only FDA approved drugs are considered. For the L1000 data set Supplementary Table S7, we find in the first position the clinically relevant drug mitoxantrone. Other drugs in clinical use according to the TTD records that appear in the top twenty positions are dasatinib, decitabine, menadione, and doxorubicin hydrochloride. Among the top twenty positions in the ranking we find drugs with known effects in prostate cancer patients, related animal models, or cell lines, and one of these went into clinical trial stage: erlotinib<sup>28</sup>, sorafenib<sup>29</sup>, lapatinib<sup>30</sup>, sirolimus<sup>31</sup>, sunitinib<sup>32</sup>, azacitidine<sup>33</sup>, everolimus<sup>34</sup>, gemcitabine (in clinical trial <https://clinicaltrials.gov/ct2/show/NCT00276549>), ciclosporine<sup>35</sup>, and auranofin<sup>36</sup>.

A study of 30 patients with advanced or metastatic prostate cancer has been conducted to test the effect of erlotinib<sup>28</sup>, and it demonstrated an improvement in clinical benefit. Sorafenib has been tested in patients with hormone-refractory prostate cancer<sup>29</sup>, indicating that the treatment may be associated with good outcomes in



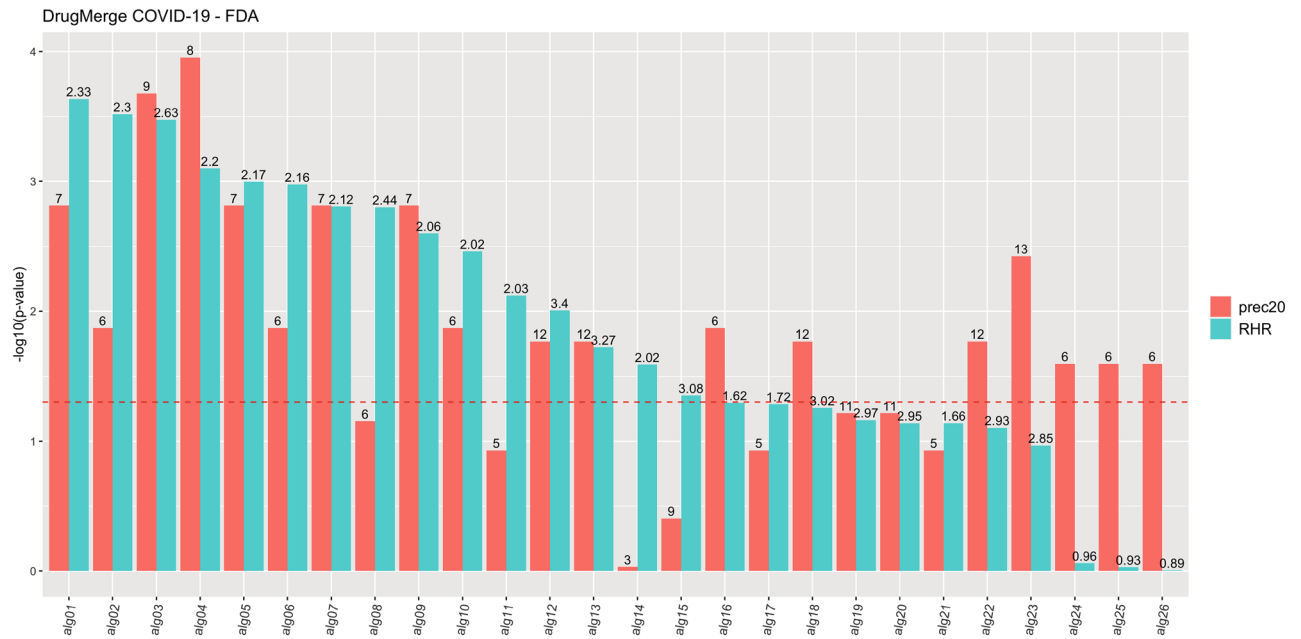


**Figure 4.** Comparison of performance of DrugMerge and the CMAP algorithm across the four benchmark diseases. For Colorectal Cancer and Prostate Cancer both methods use CMAP data. For Asthma and Rheumatoid Arthritis DrugMerge uses GEO data. The bars represent the  $-\log_{10}(p\text{-value})$  with respect to the precision@20 (red), and the RHR (light blue). The numbers on the top of the bars show the absolute values of precision@20 or RHR. The dotted red line represents the limit of significance ( $-\log_{10}(0.05)$ ). All the bars above the dotted line show a significant p-value. The plot has been generated by the *ggplot2*<sup>81</sup> R package.

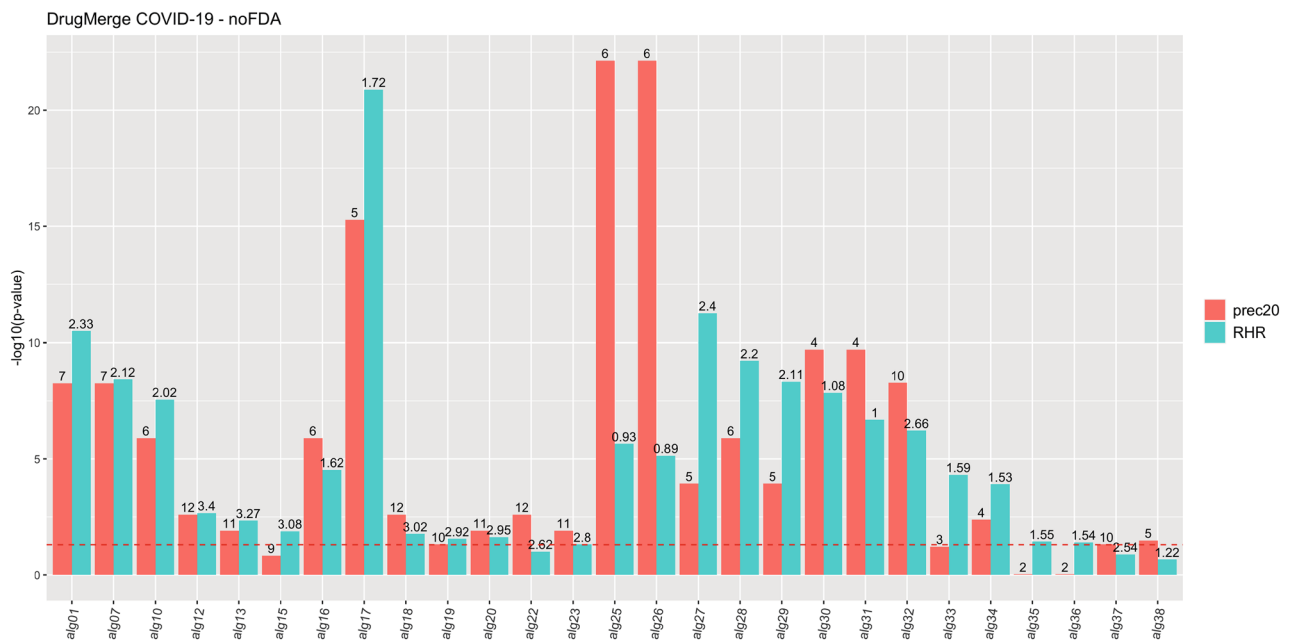
terms of overall survival. Lapatinib is a dual tyrosine kinase inhibitor of the epidermal growth factor 1 (EGFR) and 2 (HER2). In a study of twenty-nine castration-resistant prostate cancer patients, lapatinib showed single-agent activity in a small subset of unselected patients<sup>30</sup>. Nandi et al.<sup>31</sup> investigated the effect of sirolimus encapsulated liposomes in two different prostate cancer cell lines, and noticed an antiproliferative effect. Sunitinib, an inhibitor of tyrosine kinase receptors, has been studied in men with castration-resistant prostate cancer<sup>32</sup>, indicating that it is well tolerated and may have modest benefit in this patient population. Azacitidine, a demethylating agent, has been tested in combination with docetaxel, and prednisone in patients with metastatic castration-resistant prostate cancer and showed to be active in this subset of population<sup>33</sup>. Everolimus, an mTOR inhibitor, has been assessed in mice, resulting in an inhibition of prostate cancer growth<sup>34</sup>.

For the remaining drugs, we focus on their action mechanism, and we find some links with prostate cancer for imatinib<sup>37</sup>, epirubicin<sup>38</sup>, and mitomycin C<sup>39</sup>. Imatinib is an oral chemotherapy medication and functions as a tyrosine kinase inhibitor. Vicentini et al.<sup>37</sup> investigated the effects of the epidermal growth factor receptor tyrosine kinase inhibitor on human prostatic cancer cell lines, suggesting that it may have the potential in blocking tumor growth and progression. Epirubicin has been already found in colorectal cancer and as mentioned before, it is a topoisomerase II inhibitor. An in vitro cell-based assay<sup>38</sup> demonstrated that MHY336, a topoisomerase II inhibitor, significantly inhibited the proliferation of three prostate cancer cell lines. Finally, mitomycin C is used as a chemotherapeutic agent and it is a potent DNA crosslinker. DNA crosslinking has useful merit in chemotherapy and targeting cancerous cells for apoptosis<sup>40</sup>. It is known that evading cell death is one hallmark of cancer, in particular, prostate cancer cells develop multiple apoptosis blocking strategies during the various stages of tumor progression<sup>39</sup>.

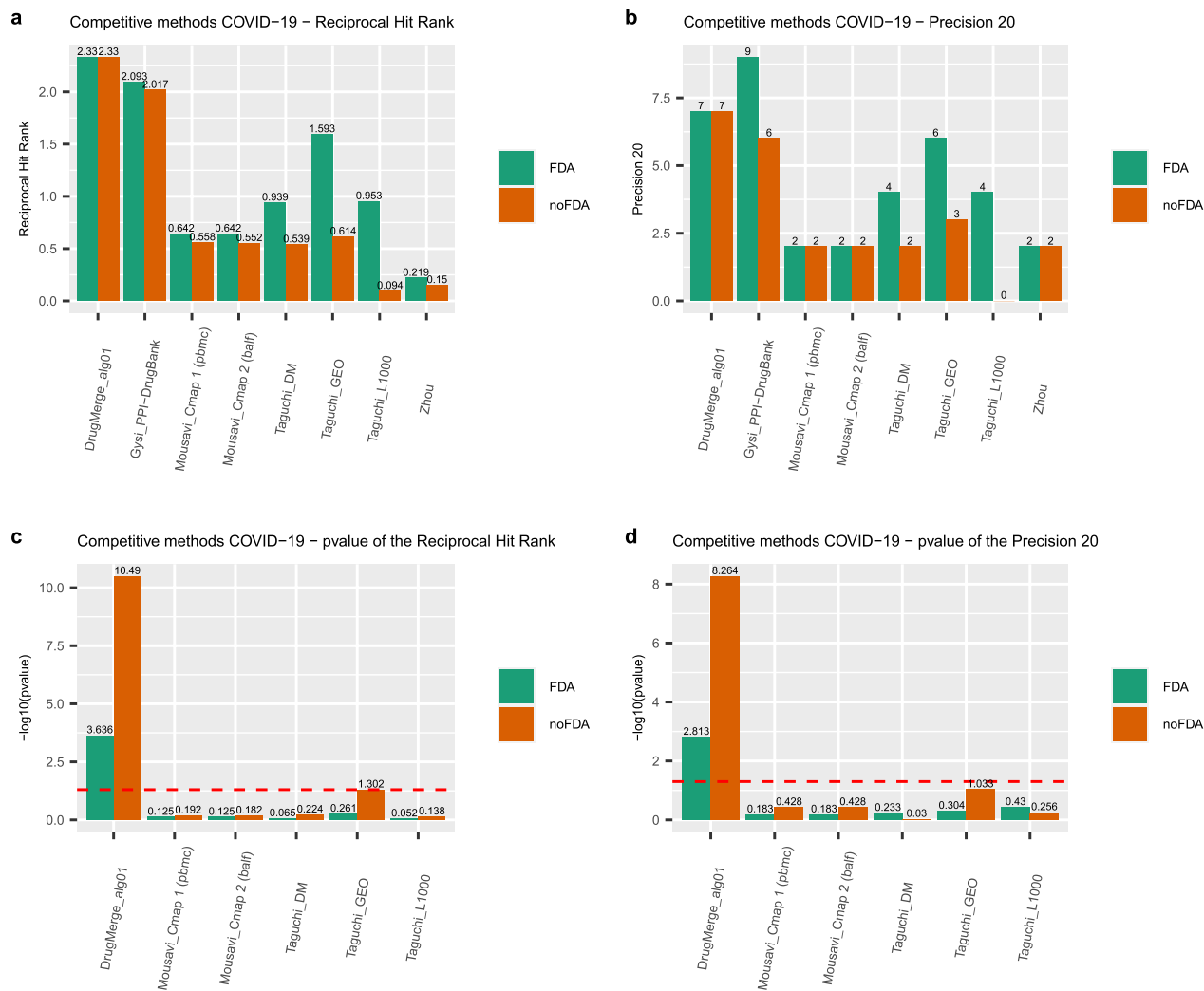
**DrugMerge results on COVID-19.** Figures 5 and 6 and Supplementary Table S5 show that DrugMerge finds significant rankings in all drug databases (L1000, CMAP, GEO, DrugMatrix), using a variety of algorithms and their combinations, both when all drugs are considered and when only FDA approved drugs are considered. For COVID-19 we have a choice of three disease data sets: (i) human adenocarcinomic alveolar basal epithe-



**Figure 5.** DrugMerge performance on COVID-19 data restricted to drugs with FDA approval. The bars represent the  $-\log_{10}(p\text{ value})$  with respect to the precision 20 (red), and the RHR (light blu). On the x-axis, different algorithms or combinations of them are indexed with aliases, since the complete algorithm names are too long. The mapping between them is on Table S9. The numbers on the top of the bars show the absolute values of precision@20 or RHR. The dotted red line represents the limit of significance ( $-\log_{10}(0.05)$ ). All the bars above the dotted line show a significant p-value. The plot has been generated by the *ggplot2*<sup>81</sup> R package.



**Figure 6.** DrugMerge performance on COVID-19 data without FDA approval restriction. The bars represent the  $-\log_{10}(p\text{ value})$  with respect to the precision@20 (red), and the RHR (light blue). On the x-axis, different algorithms or combinations of them are indexed with aliases, since the complete algorithm names are too long. The mapping between them is on Table S9. The numbers on the top of the bars show the absolute values of precision@20 or RHR. The dotted red line represents the limit of significance ( $-\log_{10}(0.05)$ ). All the bars above the dotted line show a significant p-value. The plot has been generated by the *ggplot2*<sup>81</sup> R package.



**Figure 7.** Performance of drug rankings reported in literature for COVID-19 drug repurposing. The (a,b) panels show the absolute values of RHR and precision 20, respectively. The (c,d) show the  $-\log_{10}(p)$  value of the RHR and precision 20, respectively. The dotted red line indicates the 0.05 significance threshold. The analyses have been performed using only FDA-approved drugs (green bars) and without any FDA filtering (red bars). The plot has been generated by the *ggplot2*<sup>81</sup> R package.

lial A549 cells (cells); (ii) bronchoalveolar lavage fluid RNA-Seq (BALF); (iii) infected patient peripheral blood mononuclear cells (PBMC), four Drug datasets (L1000, CMAP, GEO, and DrugMatrix), and several algorithms and weighting functions. We proceed by first running a single algorithm on a single drug data set, collecting the configurations that give the best significant performance measures. Next, we seek improvements by merging the rankings for several algorithms with significant performance on a single disease data set. And finally, we combine the improved significant results by merging the rankings obtained by several significant algorithms on several COVID-19 disease data sets. Figures 5 and 6 and Supplementary Table S5 report the best results in each phase, while Table S9 report the full algorithmic settings.

We focus now on two such rankings in Supplementary File S7 attaining the best p-values on RHR for all drugs or FDA-approved drugs only. The highest p-value on RHR when only FDA-approved drugs are considered is attained on the CMAP Drug dataset using all three disease data sets for COVID-19 and three algorithms (ClustEx, Core&Peel, Degas), (Alg 01 in Fig. 5, see Supplementary Table S9). The top two positions of the ranking report two drugs listed in clinical trial on [www.trials.gov](http://www.trials.gov) for COVID-19: etoposide and mefloquine. Within the first 20 positions, further 5 drugs are in clinical trials for COVID-19: colchicine, chlorpromazine, propofol, tretinoin, and ivermectin. Among the top ten positions in the ranking we find drugs with known effects animal models or cell lines or proposed for further scrutiny in silico screenings: thioridazine<sup>41</sup>, prochlorperazine<sup>42</sup>, primaquine<sup>43</sup>, fluphenazine<sup>44</sup>. The highest p-value on RHR when all drugs are considered is attained by KeyPathwayMiner (KPM) on the L1000 drug data set with the BALF COVID-19 disease data set (Alg 17 in Fig. 6, see Supplementary Table S9). The top position of the ranking reports a drug listed in clinical trial on [www.trials.gov](http://www.trials.gov) for COVID-19: sirolimus. Within the first twenty positions further 4 drugs are in clinical trials for COVID-19: decitabine, tamoxifen, simvastatin, and imatinib. Among the top ten positions in the ranking we find drugs with



known effects in animal models or cell lines or proposed for further scrutiny in *in silico* screenings: lapatinib<sup>45</sup>, dasatinib<sup>46</sup>, mitoxantrone<sup>47</sup>, everolimus<sup>48</sup>, sunitinib<sup>49</sup>.

For the remaining drugs in the top 10 positions in both lists, we determine the main mechanism of action (i.e. we determine the main pathways affected by the drug having the desired pharmacological effect) and we then look for hints in the literature of relevance of these pathways for COVID-19. In the first list, we find methylergometrine, which is a partial agonist and antagonist of serotonin, dopamine, and  $\alpha$ -adrenergic receptors. Costa et al.<sup>50</sup> conjectured that serotonin inhibitors could have a neuroprotective effect in patients affected by COVID-19. In the second list, we find erlotinib, stuent/sunitinib, azacitidine, and sorafenib. Erlotinib is an epidermal growth factor receptor (EGFR) inhibitor in use for the treatment of non-small cell lung cancer (NSCLC) and pancreatic cancer. Epidermal growth factor receptor (EGFR) signaling is known to be involved in the progress of viral infections<sup>51</sup>, including those caused by SARS-CoV<sup>52</sup> and SARS-CoV-2<sup>53</sup>, thus it is a plausible target pathway to consider. Sunitinib (Sutent) multi-targeted receptor tyrosine kinase (RTK) inhibitor used in the treatment of renal cell carcinoma (RCC) and imatinib-resistant gastrointestinal stromal tumor (GIST). Weidberg et al.<sup>46</sup> recently advocated the repurposing of kinase inhibitors for the treatment of COVID-19. Azacitidine has been shown effective against human immunodeficiency virus (HIV) *in vitro*<sup>54</sup> and human T-lymphotropic virus (HTLV)<sup>55</sup>. Its action is mainly at an epigenetic level, inhibiting DNA methyltransferase, causing hypomethylation of DNA. The role of epigenetics in COVID-19 is a growing area of research with large potential for advances in drug target identification<sup>56</sup>. Sorafenib is a kinase inhibitor drug approved for the treatment of advanced primary renal cell carcinoma, advanced primary liver cancer, and forms of resistant advanced thyroid carcinoma. Sorafenib is a protein kinase inhibitor with activity against the Raf/Mek/Erk pathway<sup>57</sup>. Several authors have explored the role of MAP Kinase signaling pathway in the coronavirus family at large and for SARS-CoV-2 in particular<sup>58</sup>. In Supplementary File S8 and Fig. 7 we report performance measures for the proposed COVID-19 drug repurposing rankings from Taguchi et al.<sup>6</sup>, Mousavi et al.<sup>7</sup>, Gysi et al.<sup>8</sup>, and Zhou et al.<sup>9</sup>.

For Taguchi et al. and Mousavi et al. the drug data sets used are overlapping with ours thus it is possible to compute the significance of the performance measures on a common basis. Interestingly as shown in Supplementary File S8 no such rankings attain the significance threshold  $p\text{-value} \leq 0.05$  in any configuration. For Gysi et al. and Zhou et al., the drug data sets used are different from the ones we use, and the two methods are not based on gene expression data, and it is not possible to determine with confidence the initial pool of drug candidates. For this reason, we do not compute the normalized significance, and we report directly the absolute performance measures. The results of Gysi et al. are in absolute value superior to those of Zhou et al. and qualitatively comparable to those obtained by DrugMerge on GEO data for precision@20, and on all four Drug datasets for RHR. In particular, Gysi et al.<sup>8</sup> use data on interactions between viral and human proteins in Gordon et al.<sup>59</sup> to seed their active subnetwork and finish their listing with the help of expert advice, which we do not use. Gysi et al. also have a notion of ‘positive’ drug action akin to the one used in our paper.

## Discussion

As drug repositioning is a vast subject, with implications from several areas of biology and medicine, we comment on the relationship of DrugMerge with several issues arising in the Drug Repositioning literature. Each issue is introduced by a heading.

**Three strategies.** We can identify three main strategies in drug repositioning (DR): the first is disease-oriented DR (i.e. find repurposable drugs for a specified disease), the second is drug-oriented DR (finding a disease for a specified drug), the third is generic DR (find any novel drug-disease pair). The approaches may vary significantly across the three types, and also the validation methodology is strongly influenced by the objective of the strategy. In this study we tackle disease-oriented DR, however, we should point out that the base methodology can in principle be adapted easily to the other two strategies, provided sufficient high-quality data is available.

**DrugMerge in a nutshell.** DrugMerge is a proof-of-concept exploring a novel use of disease active subnetworks in the context of drug repositioning aimed at a specific disease. In particular, DrugMerge uses multiple Disease Active subnetworks (DAS) algorithms and a state-of-the-art rank merging procedure in conjunction with two scoring functions. The first scoring function is directly the  $p$ -value of the enrichment of a drug DE genes within the DAS. The second scoring function, called ‘positivity score’, measures the antagonistic activity of the drug onto the disease, as mediated by a covering of the DAS as a collection of dense ego-networks computed by two methods: Core&Peel and ModuleDiscoverer. Matching the resulting rankings with the golden standard of a disease indicates the configuration attaining the best results, and thus those likely to include good candidates for drug repositioning in top positions (see “Methods” section for more details).

**Performance on real data.** The tests involving four benchmark diseases with drugs in clinical use did confirm the superior performance of the proposed method over the current state of the art (e.g. CMAP) in placing such clinically approved drugs in top-ranking positions. In this case, the evaluation criterion is very strict, however we were able to guess correctly the efficacy and safety of drugs administered to patients basing our analysis only on transcriptomic data from *in vitro* cell lines and tissues. Interestingly, while efficacy is in principle detectable by examining specific types of cells or tissues primarily affected by a disease, safety is a much broader concept involving all human organs and tissues, thus harder to pinpoint with localized experiments either *in vitro* or *in silico*.

Tests on the data from the COVID-19 disease also show, according to our validation criterion, a performance superior to some of the approaches proposed in the literature, while for some others we notice a broadly equivalent outcome, even when using different approaches and different input data. Results on COVID-19 should be

taken with caution, however, as our validation methodology implies that the proposed ranking is able to mimic the best judgment of human experts, which may be based on their prior knowledge and expertise, data available to them, and their assessment of the odds, leading to a shortlisting of such drugs in clinical trials. Thus our validation methodology might indeed be failing us in cases where the disease to be tackled has key features that escape our present general understanding. What can be gained in our exercise is a way to make drug shortlisting for in vitro further assessment (and eventual accession to clinical trial) more systematic, while also using relatively cost effective data sources.

**Use of transcriptomic data.** Besides the global gene co-expression network, which we can assume as background knowledge, our method relies on rather standard and relatively inexpensive RNA-Seq and microarray transcriptomic measurements of differentially expressed genes in infected cells/tissues. Systematization and finding alternatives to pure serendipity (discovery by chance) are key aspects of all in silico based drug repositioning efforts. In the “Related work” section on Supplementary Material, we discuss other related works on drug repositioning. One bonus of this parsimonious data requirement is that it becomes easier to track the evolution of viral Mechanism of Action and disease etiology as the viral agent mutates, thus in principle the drug ranking, and thus follow up decisions, may change rapidly when transcriptomic data on new viral strains of COVID-19 become available.

**Other coronaviruses.** SARS-CoV-2 is a member of the coronavirus family, responsible also for the SARS and MERS epidemics. Many studies, in particular at the end of 2019 and beginning of 2020, make up for the paucity of data on SARS-CoV-2 by making use of genomic, transcriptomic, proteomic, and pharmacological data relative to SARS-CoV and MERS-CoV and other members of the coronavirus family, including those known to be hosted by other animal species.

In our study, we chose not to use data from viruses different from SARS-CoV-2 as primary data. This choice comes from two orders of considerations. The first is that data from close relatives of the pathogen may not always be available, and there is thus a general long term advantage attained by methods that rely just on relatively small and focused data. The second issue is that we focus not on the pathogen in itself but on the human transcriptional response to it, which is known to be rather different across SARS-CoV-2, SARS-CoV, and MERS epidemics. In this line of reasoning, it makes more sense to highlight what makes human transcriptional response to SARS-CoV-2 different from that induced by other coronaviruses, even when the causing pathogens are phylogenetically very close.

**Current limitations of DrugMerge.** We use the list merging procedure to merge lists obtained by several algorithms and several DE gene data sets from different cell lines/tissues but always referred to a single drug data sets, and a single disease. Different drug data sets are often obtained using different initial candidate drugs and use different technologies. Moreover sometimes, due to the different technologies and other experimental settings, even when we restrict the analysis to common drugs, replicability across drug data sets may be low (see e.g.<sup>60,61</sup> reporting low replicability across CMAP and L1000 data using the CMAP algorithm). Moreover the performance measures we use measure the quality of the whole ranking, and the ranking itself gives a way to compare drugs listed by position. However, based on these measures we refrain from comparing (and thus give a relative order) to two drugs in different rankings obtained with different perturbation drug data sets. Such a global ranking would need further insight and technological bias correction, which we leave for future research.

**Dose dependency.** Drug datasets report often measurements for a range of different dosages taken at a range of different time points. For the p-valued ranking, we use the dosage/time point giving the lowest p-value. For rankings based on positivity score, we take the average positivity score. These choices are justified for an initial assessment of drug potential to be effective against a disease. Dosage and time points could be taken into account together with available dose-dependent toxicity data in a second level of analysis when additional data is collected in in vitro experiments.

**Tissue specificity.** For this study, we use all available transcriptomic drug and disease data regardless of the cell lines/tissues/species for which the transcriptomic measurements are taken. This choice increases the robustness of the results in general, however, it might miss some more subtle tissue-specific phenomena.

On the one hand, it is known that the SARS-CoV-2 virus infects several human organs including lungs, liver, kidneys, and gut<sup>62</sup>, and the cells transcriptional response may in principle be diverse in each tissue. However, Dudley et al.<sup>63</sup> analyzing a large repository of disease-related gene differential expression data conclude that “molecular signature of disease across tissues is overall more prominent than the signature of tissue expression across diseases”. Our analysis of active subnetworks extracted from three COVID-19 data sets (BALF, PBMC, cells) also supports for COVID-19 a general high similarity of gene expression across the sample types. Differences in the extracted active networks are mainly due to the algorithms used.

Also, drugs may act differently in different cell types. For example, chloroquine has been found to be ineffective in infected human lung cells<sup>64</sup>, while it was found effective in infected Vero cells (a cell line isolated from kidney epithelial cells extracted from an African green monkey)<sup>65</sup>. Similar tissue-specific differences in drug sensitivity are noticed by Dittmar et al.<sup>66</sup>. In principle, if a variety of options are available, one may opt for including only the data drawn from biological samples more similar to the actual tissues affected by the disease in patients.

In our study, we use a large variety of cellular types for drug perturbation analysis. Drug matrix has drug perturbation signatures available on liver, heart, kidney, thigh muscle, and primary hepatocytes (from *Rattus norvegicus*). GEO has drug perturbation signatures for several tissues and species including *homo sapiens*, *mus*

*musculus* and *Rattus norvegicus*. CMAP measures drug perturbations in breast cancer epithelial cell line MCF7 (MCF7/ssMCF7), in prostate cancer epithelial cell line PC3, and the nonepithelial lines HL60 (leukemia) and SKMEL5 (melanoma). L1000 collects drug perturbation data from 98 different cellular contexts including human primary cell lines and human cancer cell lines.

For COVID-19 we report two drug rankings attaining top performance, one is obtained by merging data from all three tissues used to measure DEG, therefore it indicates evidence for drugs that may act on a general systemic level across different human organs. This is complemented by the second ranking which is attained specifically for BALF measurements (bronchoalveolar lavage fluid) and thus it may indicate drugs with a specific action in pulmonary tissues.

**Drug mechanisms of action.** One advantage of our approach, in contrast with many others, is that we do not have an initial explicit guess for a putative target protein for the drug or a putative target pathway (or mechanism of action). However, once relevant drugs are shortlisted we can analyze their likely mechanism of action, either through the literature or through the use of specialized databases.

**On specific drugs in clinical trials.** In this study, we have used four public Drug-related data sets (GEO, DrugMatrix, CMAP, and L1000) representing a sufficient drug basis for our investigation and the validation of the proposed methodology. However, a much larger amount of useful data on drugs is proprietary therefore it is important to devise incentives for sharing these repositories of useful data in a global drug repositioning strategy. As a consequence of the non-complete drug coverage of our data sets we could not assess some drugs whose use in COVID-19 cases has been advocated, such as remdesivir (approved by FDA in October 2020<sup>67</sup>) or lopinavir (not approved by FDA). Cao et al.<sup>68</sup> recently reported the results of a randomized clinical trial involving 199 hospitalized adult patients with severe COVID-19: no benefit was observed with lopinavir-ritonavir treatment over the standard care cohort. Interestingly, we have data from DrugMatrix on ritonavir (not approved by FDA), which fails the positivity score filter, thus our results on ritonavir are consistent with the findings in Cao et al. which are at a clinical level.

**Role of comorbidities.** Comorbidities in patients affected by COVID-19 are important clinical aspects, as it is known that comorbidities are one of the strongest risk factors for such patients, strongly affecting mortality rates<sup>69</sup>. Comorbidities may alter drastically the effect of drugs. The disease gene expression data (for all five diseases) we use in this study do not come from patients with known comorbidities, therefore the results of our study should be interpreted in such a setting. Fisco et al.<sup>70</sup> try to exploit known comorbidities as a positive tool in modeling the relevant features of COVID-19, while other authors, based on protein target analysis, have a more skeptical view<sup>8</sup>: “In summary, we find that the SARS-CoV-2 targets do not overlap with disease genes associated with any major diseases, indicating that a potential COVID-19 treatment can not be derived from the arsenal of therapies approved for specific diseases”. The effect of concomitant comorbidities on repurposed drugs might be better approached when transcriptomic data from these sub-populations of COVID-19 patients become available.

**Drug combinations.** In this study, we do not treat the issue of combinations of drugs. This is a novel area of research in which network-based approaches have been shown to give novel insight and may help in coping with the combinatorial explosion of drug combinations. However, validated data on drug combinations is limited thus making the task of performance evaluation and method validation more complex w.r.t the analysis for single drugs. We plan to incorporate models of multi-drug/disease interactions in our setting as future research<sup>71</sup>.

**Toxicity and adverse events.** Drugs can be toxic or cause adverse events. Toxicity can be detected via ad-hoc in vitro and in vivo experiments or it can be predicted via in silico models<sup>72</sup>. Any of these techniques should be applied downstream to the use of DrugMerge for extra safety. DrugMerge uses two forms of filtering to handle adverse effects and toxicity. The first mechanism is the inclusion of drugs having a contrary effect to the disease as measured by the positivity score (which in turn is similar to a mechanism also used by Gysi et al.). The second filtering is via the possibility to limit the analysis to FDA-approved drugs, for which toxicity and adverse effects are known and have been considered to be outweighed by the benefits in certain conditions. This mild form of filtering is sufficient to increase the performance in our tests and thus mimic human expert judgment, which includes consideration of toxicity and adverse effects. However, toxicity should be always be reconsidered ex-post to finalize any shortlisting decision.

**Role of the disease causes.** Since we base our approach on the transcriptomic responses of human cells/tissues to drugs and diseases, the cause of the disease (viral, bacterial, genetic, metabolic, etc.) is not critical for the application of the methodology. Indeed in our study, we applied it to a variety of cases. Asthma is an inflammatory disease of the airways of the lungs having several triggering causes, including exposure to air pollution and allergens. Rheumatoid arthritis (RA) is a long-term autoimmune disorder that primarily affects joints. Cancer is generally associated with alterations of the genetic code in somatic human cells. Finally, COVID-19 is of viral origin. The advantage of generality is balanced by the possibility that a high-ranking drug is effective against symptoms or side-effects of the diseases, rather than acting on its main mechanism of action. In one of the drug rankings for COVID-19, we find within the top 20 positions prochlorperazine which has both antiviral activity<sup>41</sup> and is also a broad-spectrum antiemetic contrasting COVID-19 symptoms like nausea and vomiting<sup>42</sup>.

At this level of analysis, DrugMerge does not yet distinguish the different actions a drug might have thus further investigations are needed to untangle complex drug actions.

## Conclusions

In this conclusive section we summarize the features of DrugMerge and we comment on possible extensions and applications, as well as additional benefits of drug prioritization in general.

In this study, we describe DrugMerge, a methodology for ranking repurposable drugs where we rank drugs by their ability to affect collections of disease active subnetworks (computed by an ensemble of state-of-the-art algorithms) and contrast the perturbation induced by the disease in such sub-networks. DrugMerge should be considered as a proof-of-concept and its predictions should be considered as indications for shortlisting drugs for further in vitro analysis rather than absolute predictions. In particular, aspects like dosage, tissue specificity, and a more refined analysis of possible adverse side effects will need to be tackled further in developing DrugMerge towards a full-fledged product.

**Network drug-disease model.** The rationale behind our approach is a model in which drugs and diseases interact indirectly by their mutual action on the sub-networks rather than directly by affecting the same target proteins. Moreover, we are able to leverage an ensemble of different algorithms for computing active subnetworks, and then merge the resulting rankings, thus increasing the robustness of the methodology. Results of four benchmark diseases are encouraging since we could find in all four cases drugs in clinical use (thus both safe and effective) in first ranking positions, and several other drugs in clinical use in high ranking positions.

**Benchmarking with prior knowledge.** These results imply that DrugMerge when provided with input data from transcriptional assays, has the potential for coming close to expert human judgment, and guess drugs in actual clinical usage. For COVID-19 we measure how close we come to guessing drugs that have been short-listed for clinical trials according to the FDA records. For most drugs in high ranking positions but not on clinical trials we have found several supporting evidence in the literature or by direct assessment of their likely Mechanism of Action, that makes their indication by DrugMerge biologically plausible.

**Other benefits of drug prioritization.** Prioritizing drugs for in vitro testing and eventual accession to clinical trials is valuable for decision makers in a pandemic situation. Lack of prioritization in drug repurposing efforts has led to some drugs being over-researched, early in the pandemic, even based on dubious evidence (e.g. hydroxychloroquine in COVID-19). Moreover having many small clinical trials has the effect of reducing the pool of eligible patients available for larger and more promising trials ( Generic Drug Repurposing for COVID-19 and Beyond. Susan Athey, Rena Conti, Richard Frank, and Jonathan Gruber Institute for Health System Innovation & Policy. July2020. <https://www.gsb.stanford.edu/faculty-research/publications/generic-drug-repurposing-covid-19-beyond>).

**Sub-populations of patients.** A tool such as DrugMerge aiming at drug ranking might be used in conjunction with an activity of data collection that targets specific sub-populations of patients. For example, most trials to date for repurposed drugs in COVID-19 have focused on hospitalized patients, even though prevention might be more valuable and effective if initiated earlier in disease progression. Also transcriptomic data from patients affected by frequent comorbidities might help towards the study of drugs specifically selected for sub-populations at higher risk of mortality.

**Preclinical nature of this research.** This paper focuses on the preclinical considerations that may indicate a drug to be shortlisted for further preclinical in vitro and in vivo testing, leading to eventual clinical trial. Naturally, all evidence, which may conceivably be noisy and often contradictory, should be gathered and considered before pushing a drug further down the pipeline.

## Methods

**Drug ranking based on p-value.** We use the Core&Peel technique described in Lucchetta et al.<sup>4</sup> and 4 other methods (ClustEx<sup>73</sup>, ModuleDiscoverer<sup>74</sup>, Degas<sup>75</sup>, and KeyPathwayMiner<sup>76</sup>) to produce active subnetworks for a given disease  $D$ , starting from a list of differentially expressed genes in samples affected by the disease w.r.t. healthy controls  $DEG(D)$  and a gene co-expression network. In Lucchetta et al., we employed the gene co-expression network made available by the DREAM challenge<sup>77</sup>, and DEG for seven different datasets: two inflammatory disease microarray experiments (asthma and rheumatoid arthritis), two cancer RNA-Seq studies (prostate and colorectal cancer), and three COVID-19 datasets called BALF (bronchoalveolar lavage fluid RNA-Seq), PBMC (infected patient peripheral blood mononuclear cells) and COVID19 cells (human adenocarcinomic alveolar basal epithelial A549 cells). In Lucchetta et al., we compared the performance of Core&Peel and the other four algorithms in detecting active subnetworks, resulting in that Core&Peel is a competitive method in that purpose. In this work, we use the five active subnetworks detected by all five methods as inputs of DrugMerge. Let  $AN$  be any such active network. For a given Drug perturbation datasets (for us: GEO, Drug-Matrix, CMAP, and L1000; see description in “Drug perturbation databases” section of Supplementary Material) using the *enrichR* R/Bioconductor package, we obtain a listing of drugs enriched in  $AN$  ranked by their adjusted p-value in reverse order (from the smallest to the largest). We cut this list at the adjusted p-value of 0.05, retaining only the significantly enriched drugs.



**Filtering of drug ranking.** We have noticed that a straightforward application of the p-value based drug ranking includes in top ranking positions some molecules and drugs that are known to be either generally toxic or to induce diseases in animal models. For this reason, we introduce the option of applying two filters to a ranked list of drugs. The first filter is used to retain only drugs that are approved by the FDA (we use a list of drugs downloaded from <https://www.accessdata.fda.gov> on September 9th 2020, with 6315 entries). For FDA-approved drugs usually adverse effects and toxicity levels are known and probably acceptable in most cases. The second filter retains drugs that by the directionality analysis appear to have a 'positive' action on the disease, i.e. genes are up/down regulated antagonistically to the disease effect. This list of positive drugs can be obtained with the method described in the next section.

**Drug ranking based on drug positivity scores.** For this analysis, we exploit the fact that two of the methods (Core&Peel and ModuleDiscoverer) we employ to build active subnetworks produce the active sub-network as a union of dense sub-graphs (ego-networks) of a large co-expression network (in detail this co-expression network has been produced for the DREAM challenge on disease module detection<sup>77</sup>). Since such network is built on the correlation of the expression-vectors over thousands of conditions, a high correlation vector implies that a pair of genes are generally both up or both down regulated at any given time. In a dense ego-network, such tendency is coherently common to all genes in the ego-network. We exploit this phenomenon as follows. For an ego-network that is significantly enriched in genes down-regulated by the disease, we subtract the number of up-regulated genes by the drug from the number of down-regulated genes by the drug. If this number is positive we say that the drug has a *positive* effect of the disease (for this ego-network). We do a symmetric calculation of the case of ego-networks that are significantly enriched in genes up-regulated by the disease. Summing these effects (with their sign) over all the ego-networks comprising the active subnetwork, we can compute the directionality of action of the drug and a magnitude (directionality score). We retain the drugs that have a positive directionality score and we rank them by the magnitude of the score. Note that this scheme relies just on counting the number of up/down regulated genes, not on the magnitude of the expression, as long as it is significant. In a variant of this scheme, we can use the size of the ego-networks as a weighting factor. This method is justified by an extension to the ego-networks of the phenomenon Chen et al.<sup>78</sup> have noticed that in preclinical models of breast, liver, and colon cancers, reversal of gene expression between drug and disease correlates well with drug efficacy. Furthermore, Chen et al. validated the efficacy of drugs in xenograft models of the diseases.

**Merging of drug lists.** Given two or more ranked lists of drugs, we wish to aggregate them into a single list that takes into account the ranking of any drug in each of the lists where it occurs. Following Gysi et al.<sup>8</sup> we use the C-rank routine in Zitnik et al.<sup>79</sup> to perform rank aggregation. C-rank is defined in Zitnik et al. to work with any custom-defined ranking of items (thus we do not make use of the graph-based community ranking scores for which c-rank was originally designed, instead we use it here as a generic rank aggregation method). Interestingly this rank aggregation method is oblivious of the origin of the ranking and we can thus easily aggregate lists obtained with different active subnetwork algorithms on different disease-DEG sets. The only caveat is that the lists supplied to c-rank must be of equal length, thus the case of missing drugs must be handled, by fitting the drugs missing from a list at its bottom with a scaled mean score value obtained from the other lists.

**Performance evaluation.** The performance evaluation aims at assessing how well a predicted ranked list gives priorities in accordance to human expert judgment as codified by disease-treatment pairs extracted from the *Therapeutic Target Database*<sup>13</sup> for the four benchmark diseases. Since for COVID-19 currently only one drug has gained acceptance as an approved clinical treatment, we resort to the list of drugs for the treatment of COVID-19 undergoing clinical trial as listed in the portal <https://www.clinicaltrials.gov/> (at Sept 5th, 2020). We use two main measures. The first is precision at 20 (precision@20), that is the number of drugs in the first 20 positions of the proposed list of repurposable drugs that are hits in the TTD-derived listing, or, for COVID-19, in the clinicaltrials.gov listing. The second is the Reciprocal Hit-Rank (RHR)<sup>80</sup>. For a list  $L$  of predictions length  $N$ , let the hits be in positions  $p_1, p_2, \dots, p_h$ . We define:

$$RHR(L) = \sum_{i=1}^h \frac{1}{p_i}$$

Note that this formula does not imply a fixed length for the lists being compared, and it is fair in comparing both long and short lists. A longer list may have more hits by chance, however it will have positive but diminishing returns for hits in its tail.

**Statistical significance.** Let  $D$  be a set of drugs and  $H \subset D$  a golden standard for a disease (e.g. the drugs in  $D$  that are in clinical use for the disease). We generate  $n$  random permutations uniformly at random  $r_i(D)$ , for  $i = 1, \dots, n$  and we take, for a performance measure  $\mu_H()$  depending on  $H$ , the mean  $m$  and standard deviation  $sd$  of the sequence  $M = [\mu_H(r_i(D)) \text{ for } i = 1, \dots, n]$ . For a given drug ranking  $R(D)$ , its *zscore* for  $\mu_H()$  is its deviation from the mean of the above described random distribution normalized by the standard deviation of the random distribution:

$$zscore(R(D)) = \frac{\mu_H(R(D)) - m(M)}{sd(M)}$$

Assuming that the random distribution of the values of  $\mu_H(r_i(D))$  is normal, we obtain the (two sided) p-value by plugging the z-score value into the complementary cumulative distribution function (ccdf) of the normal distribution from the package `scipy.stats` of SciPy.org.

The number  $n$  of random permutations is determined as follows. Starting with  $n = 100,000$ , we increment  $n$  in steps of 10,000, stopping when the value of both  $m$  and  $sd$  are stable, with a displacement from the previous iteration less than 0.01.

## Data availability

Primary data used in this study is available in public repositories. Intermediate results, processing software and scripts are available on GitHub: <https://github.com/martaluc/DrugMerge>.

Received: 20 June 2021; Accepted: 23 September 2021

Published online: 06 October 2021

## References

- Ashburn, T. T. & Thor, K. B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
- Sam, E. & Athri, P. Web-based drug repurposing tools: A survey. *Brief. Bioinform.* **20**, 299–316 (2019).
- Barabasi, A.-L. & Oltvai, Z. N. Network biology: Understanding the cell functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
- Lucchetta, M. & Pellegrini, M. Finding disease modules for cancer and COVID-19 in gene co-expression networks with the core & peel method. *Sci. Rep.* **10**, 17628. <https://doi.org/10.1038/s41598-020-74705-6> (2020).
- Lamb, J. *et al.* The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
- Taguchi, Y. & Turki, T. A new advanced in silico drug discovery method for novel coronavirus (SARS-CoV-2) with tensor decomposition-based unsupervised feature extraction. *Preprints* (2020).
- Mousavi, S. Z., Rahmanian, M. & Sami, A. A connectivity map-based drug repurposing study and integrative analysis of transcriptomic profiling of SARS-CoV-2 infection. *Infect. Genet. Evol.* **86**, 104610 (2020).
- Gysi, D. M. *et al.* Network medicine framework for identifying drug repurposing opportunities for COVID-19. *PNAS* **118**, e2025581118 (2020).
- Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 1–18 (2020).
- Ruiz, C., Zitnik, M. & Leskovec, J. Discovery of disease treatment mechanisms through the multiscale interactome. *BioRxiv*. <https://doi.org/10.1038/s41467-021-21770-8> (2020).
- Moore, J. B. & June, C. H. Cytokine release syndrome in severe COVID-19. *Science* **368**, 473–474 (2020).
- Laing, A. G. *et al.* A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat. Med.* **26**, 1623–1635 (2020).
- Chen, X., Ji, Z. L. & Chen, Y. Z. Ttd: Therapeutic target database. *Nucleic Acids Res.* **30**, 412–415 (2002).
- Cornelia, G. *et al.* Treatment of advanced colorectal cancer with mitoxantrone, high dose folinic acid and fluorouracil. *Tumori J.* **77**, 445–446 (1991).
- Kacan, T. *et al.* Antitumor activity of sorafenib on colorectal cancer. *J. Oncol. Sci.* **2**, 53–57 (2016).
- Scott, A. J. *et al.* Evaluation of the efficacy of dasatinib, a src/abl inhibitor, in colorectal cancer cell lines and explant mouse model. *PLoS ONE* **12**, e0187173 (2017).
- Wagner, M. *et al.* Effective treatment of advanced colorectal cancer by rapamycin and 5-fu/oxaliplatin monitored by timp-1. *J. Gastrointest. Surg.* **13**, 1781–1790 (2009).
- Azad, N. S. *et al.* Combination epigenetic therapy in metastatic colorectal cancer (mcrcc) with subcutaneous 5-azacitidine and entinostat: A phase 2 consortium/stand up 2 cancer study. *Oncotarget* **8**, 35326 (2017).
- Samei, L., Yaling, P., Lihua, Y., Yan, Z. & Shuyan, J. Effects and mechanism of imatinib in inhibiting colon cancer cell proliferation. *Med. Sci. Monit.* **22**, 4126 (2016).
- Kuruppu, D., Christophi, C., Bertram, J. F. & O'Brien, P. E. Tamoxifen inhibits colorectal cancer metastases in the liver: A study in a murine model. *J. Gastroenterol. Hepatol.* **13**, 521–527 (1998).
- Lee, C.-S. *et al.* Doxorubicin-loaded oligonucleotide conjugated gold nanoparticles: A promising in vivo drug delivery system for colorectal cancer therapy. *Eur. J. Med. Chem.* **142**, 416–423 (2017).
- Ma, M., Wang, X., Liu, N., Shan, F. & Feng, Y. Low-dose naltrexone inhibits colorectal cancer progression and promotes apoptosis by increasing m1-type macrophages and activating the bax/bcl-2/caspase-3/parp pathway. *Int. Immunopharmacol.* **83**, 106388 (2020).
- Hou, Y.-Q. *et al.* Multifaceted anti-colorectal tumor effect of digoxin on hct8 and sw620 cells in vitro. *Gastroenterol. Rep.* **8**, 465–475 (2020).
- Coss, A. *et al.* Increased topoisomerase iia expression in colorectal cancer is associated with advanced disease and chemotherapeutic resistance via inhibition of apoptosis. *Cancer Lett.* **276**, 228–238 (2009).
- Thiem, S. *et al.* mtorc1 inhibition restricts inflammation-associated gastrointestinal tumorigenesis in mice. *J. Clin. Investig.* **123**, 767 (2013).
- Li, L. *et al.* Statins use and the prognosis of colorectal cancer: A meta-analysis. *Clin. Res. Hepatol. Gastroenterol.* **45**, 101588 (2021).
- Xu, H. *et al.* Transcription factors in colorectal cancer: Molecular mechanism and therapeutic implications. *Oncogene* **4**, 1–15 (2020).
- Gravis, G. *et al.* Results from a monocentric phase II trial of erlotinib in patients with metastatic prostate cancer. *Ann. Oncol.* **19**, 1624–1628 (2008).
- Zaborowska, M., Szmít, S. & Szczylik, C. Sorafenib in progressive castrate-resistant prostate cancer. Can we talk about a new therapeutic option? *Arch. Med. Sci.* **8**, 528 (2012).
- Whang, Y. E. *et al.* A phase II study of lapatinib, a dual egfr and her-2 tyrosine kinase inhibitor, in patients with castration-resistant prostate cancer. *Urol. Oncol. Semin. Orig. Investig.* **31**, 82–86 (2013).
- Nandi, U., Onyesom, I. & Douroumis, D. Anti-cancer activity of sirolimus loaded liposomes in prostate cancer cell lines. *J. Drug Deliv. Sci. Technol.* **61**, 102200 (2021).
- Michaelson, M. D. *et al.* Phase ii study of sunitinib in men with advanced prostate cancer. *Ann. Oncol.* **20**, 913–920 (2009).
- Singal, R. *et al.* Phase I/II study of azacitidine, docetaxel, and prednisone in patients with metastatic castration-resistant prostate cancer previously treated with docetaxel-based therapy. *Clin. Genitourin. Cancer* **13**, 22–31 (2015).
- Morgan, T. M. *et al.* Rad001 (everolimus) inhibits growth of prostate cancer in the bone and the inhibitory effects are increased by combination with docetaxel and zoledronic acid. *Prostate* **68**, 861–871 (2008).



35. Kawahara, T. *et al.* The role of nfatc1 in prostate cancer progression: Cyclosporine a and tacrolimus inhibit cell proliferation, migration, and invasion. *Prostate* **75**, 573–584 (2015).
36. Liu, N. *et al.* Auranofin lethality to prostate cancer includes inhibition of proteasomal deubiquitinases and disrupted androgen receptor signaling. *Eur. J. Pharmacol.* **846**, 1–11 (2019).
37. Vicentini, C. *et al.* Prostate cancer cell proliferation is strongly reduced by the epidermal growth factor receptor tyrosine kinase inhibitor zd1839 in vitro on human cell lines and primary cultures. *J. Cancer Res. Clin. Oncol.* **129**, 165–174 (2003).
38. Patra, N. *et al.* A novel epoxypropoxy flavonoid derivative and topoisomerase II inhibitor, mhy336, induces apoptosis in prostate cancer cells. *Eur. J. Pharmacol.* **658**, 98–107 (2011).
39. Gurumurthy, S., Vasudevan, K. M. & Rangnekar, V. M. Regulation of apoptosis in prostate cancer. *Cancer Metastasis Rev.* **20**, 225–243 (2001).
40. Deans, A. J. & West, S. C. Dna interstrand crosslink repair and cancer. *Nat. Rev. Cancer* **11**, 467–480 (2011).
41. Otręba, M., Kośmider, L. & Rzepecka-Stojko, A. Antiviral activity of chlorpromazine, fluphenazine, perphenazine, prochlorperazine, and thioridazine towards rna-viruses. A review. *Eur. J. Pharmacol.* **887**, 173553 (2020).
42. Kow, C. & Hasan, S. Prochlorperazine for nausea and vomiting accompanied covid-19. *J. Gastroenterol. Hepatol. (Australia)* **36**, 524 (2020).
43. Sachdeva, C. *et al.* In silico potential of approved antimalarial drugs for repurposing against covid-19. *OMICS* **24**, 568–580 (2020).
44. Nazeam, J. *et al.* Based on principles and insights of covid-19 epidemiology, genome sequencing, and pathogenesis: Retrospective analysis of sinigrin and prolixinrx (fluphenazine) provides off-label drug candidates. *SLAS Discov. Adv. Sci. Drug Discov.* **25**, 1123–1140 (2020).
45. Raymonda, M. H. *et al.* Pharmacologic profiling reveals lapatinib as a novel antiviral against sars-cov-2 in vitro. *BioRxiv.* <https://doi.org/10.1101/2020.11.25.398859> (2020).
46. Weisberg, E. *et al.* Repurposing of kinase inhibitors for treatment of COVID-19. *Pharm. Res.* **37**, 1–29 (2020).
47. Lokhande, K. B., Doiphode, S., Vyas, R. & Swamy, K. V. Molecular docking and simulation studies on SARS-CoV-2 mpro reveals metoxantrone, leucovorin, birinapant, and dynasore as potent drugs against COVID-19. *J. Biomol. Struct. Dyn.* **2020**, 1–12 (2020).
48. Terrazzano, G. *et al.* An open question: Is it rational to inhibit the mtor-dependent pathway as COVID-19 therapy? *Front. Pharmacol.* **11**, 856 (2020).
49. Wang, P.-G., Tang, D.-J., Hua, Z., Wang, Z. & An, J. Sunitinib reduces the infection of SARS-CoV, MERS-CoV and SARS-CoV-2 partially by inhibiting ap2m1 phosphorylation. *Cell Discov.* **6**, 1–5 (2020).
50. Costa, L. H., Santos, B. M. & Branco, L. G. Can selective serotonin reuptake inhibitors have a neuroprotective effect during COVID-19? *Eur. J. Pharmacol.* **889**, 173629 (2020).
51. Hondermarck, H., Bartlett, N. W. & Nurcombe, V. The role of growth factor receptors in viral infections: An opportunity for drug repurposing against emerging viral diseases such as COVID-19? *FASEB BioAdv.* **2**, 296–303 (2020).
52. Venkataraman, T. & Frieman, M. B. The role of epidermal growth factor receptor (egfr) signaling in SARS coronavirus-induced pulmonary fibrosis. *Antiviral Res.* **143**, 142–150 (2017).
53. Klann, K. *et al.* Growth factor receptor signaling inhibition prevents SARS-CoV-2 replication. *Mol. Cell* **80**, 164–174 (2020).
54. Dapp, M. J., Clouser, C. L., Patterson, S. & Mansky, L. M. 5-Azacytidine can induce lethal mutagenesis in human immunodeficiency virus type 1. *J. Virol.* **83**, 11950–11958 (2009).
55. Diamantopoulos, P. T. *et al.* Antiretroviral activity of 5-azacytidine during treatment of a htlv-1 positive myelodysplastic syndrome with autoimmune manifestations. *Virol. J.* **9**, 1–5 (2012).
56. Pruijboom, L. Methylation pathways and SARS-CoV-2 lung infiltration and cell membrane-virus fusion are both subject to epigenetics. *Front. Cell. Infect. Microbiol.* **10**, 290 (2020).
57. Liu, L. *et al.* Sorafenib blocks the raf/mek/erk pathway, inhibits tumor angiogenesis, and induces tumor cell apoptosis in hepatocellular carcinoma model plc/prf/5. *Can. Res.* **66**, 11851–11858 (2006).
58. Hemmat, N. *et al.* The roles of signaling pathways in SARS-CoV-2 infection; lessons learned from SARS-CoV and MERS-CoV. *Arch. Virol.* **166**, 1–22 (2021).
59. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 1–13 (2020).
60. Lim, N. & Pavlidis, P. Evaluation of connectivity map shows limited reproducibility in drug repositioning. *BioRxiv* **11**, 845693 (2019).
61. Lin, K. *et al.* A comprehensive evaluation of connectivity methods for 1000 data. *Brief. Bioinform.* **21**, 2194–2205 (2020).
62. Mallapaty, S. Mini organs reveal how the coronavirus ravages the body. *Nature* **583**, 15–16 (2020).
63. Dudley, J. T., Tibshirani, R., Deshpande, T. & Butte, A. J. Disease signatures are robust across tissues and experiments. *Mol. Syst. Biol.* **5**, 307 (2009).
64. Hoffmann, M. *et al.* Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2. *Nature* **585**, 1–5 (2020).
65. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* **30**, 269–271 (2020).
66. Dittmar, M. *et al.* Drug repurposing screens reveal fda approved drugs active against SARS-CoV-2. *BioRxiv.* <https://doi.org/10.1101/2020.06.19.161042> (2020).
67. Rubin, D., Chan-Tack, K., Farley, J. & Sherwat, A. Fda approval of remdesivir—A step in the right direction. *N. Engl. J. Med.* **383**, 2598–2600 (2020).
68. Cao, B. *et al.* A trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. *N. Engl. J. Med.* **382**, 1787–1799 (2020).
69. Sanyaolu, A. *et al.* Comorbidity and its impact on patients with COVID-19. *SN Compr. Clin. Med.* <https://doi.org/10.1007/s42399-020-00363-4> (2020).
70. Fiscon, G., Conte, F., Farina, L. & Paci, P. Saverunner: A network-based algorithm for drug repurposing and its application to COVID-19. *PLoS Comput. Biol.* **17**, e1008686 (2021).
71. Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1–11 (2019).
72. Zhang, L. *et al.* Applications of machine learning methods in drug toxicity prediction. *Curr. Top. Med. Chem.* **18**, 987–997 (2018).
73. Gu, J., Chen, Y., Li, S. & Li, Y. Identification of responsive gene modules by network-based gene clustering and extending: Application to inflammation and angiogenesis. *BMC Syst. Biol.* **4**, 47 (2010).
74. Vlaic, S. *et al.* Module discoverer: Identification of regulatory modules in protein–protein interaction networks. *Sci. Rep.* **8**, 1–11 (2018).
75. Ulitsky, I., Krishnamurthy, A., Karp, R. M. & Shamir, R. Degas: De novo discovery of dysregulated pathways in human diseases. *PLoS ONE* **5**, e13367 (2010).
76. Alcaraz, N., Küçük, H., Weile, J., Wipat, A. & Baumbach, J. Key pathway miner: Detecting case-specific biological pathways using expression data. *Internet Math.* **7**, 299–313 (2011).
77. Choobdar, S. *et al.* Assessment of network module identification across complex diseases. *Nat. Methods* **16**, 843–852 (2019).
78. Chen, B. *et al.* Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat. Commun.* **8**, 1–12 (2017).
79. Zitnik, M. *et al.* Prioritizing network communities. *Nat. Commun.* **9**, 1–9 (2018).
80. Deshpande, M. & Karypis, G. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst. (TOIS)* **22**, 143–177 (2004).
81. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).

## Acknowledgements

This work was supported by M.L.'s doctoral fellowship funded by CNR, Institute of Informatics and Telematics. The research exposed in this article has been conducted as curiosity-driven free research by the authors.

## Author contributions

M.P. designed the study. M.P. and M.L. wrote the code and interpreted the results. M.P. and M.L. wrote and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99399-2>.

**Correspondence** and requests for materials should be addressed to M.P.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021