Network bioinformatics analysis provides insight into drug repurposing for COVID-2019

Xu Li^{1,*}, Jinchao Yu², Zhiming Zhang¹, Jing Ren¹, Alex E Peluffo², Wen Zhang¹, Yujie Zhao¹, Kaijing Yan¹, Daniel Cohen² and Wenjia Wang^{1,2}

¹GeneNet Pharmaceuticals, Tianjin, 300410, China

²Pharmext, Issy-les-Moulineaux, 92130, France

* tsl-lixu@tasly.com

ABSTRACT

The COVID-2019 disease caused by the SARS-CoV-2 virus (aka 2019-nCoV) has raised significant health concerns in China and worldwide. While novel drug discovery and vaccine studies are long, repurposing old drugs against the COVID-2019 epidemic can help identify treatments, with known preclinical, pharmacokinetic, pharmacodynamic, and toxicity profiles, which can rapidly enter Phase 3 or 4 or can be used directly in clinical settings. In this study, we presented a novel network based drug repurposing platform to identify potential drugs for the treatment of COVID-2019. We first analysed the genome sequence of SARS-CoV-2 and identified SARS as the closest disease, based on genome similarity between both causal viruses, followed by MERS and other human coronavirus diseases. Using our AutoSeed pipeline (text mining and database searches), we obtained 34 COVID-2019-related genes. Taking those genes as seeds, we automatically built a molecular network for which our module detection and drug prioritization algorithms identified 24 disease-related human pathways, five modules and finally suggested 78 drugs to repurpose. Following manual filtering based on clinical knowledge, we re-prioritized 30 potential repurposable drugs against COVID-2019 (including pseudoephedrine, andrographolide, chloroquine, abacavir, and thalidomide). We hope that this data can provide critical insights into SARS-CoV-2 biology and help design rapid clinical trials of treatments against COVID-2019.

Introduction

The COVID-2019 disease outbreak caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), formerly named “2019 novel coronavirus” (2019-nCoV), is spreading at a high rate throughout China and beyond. As of February 24, 2020, more than 77,000 cases and 2,595 deaths were confirmed in China. New cases spiked outside of China, especially in the Republic of Korea and Italy, where the number of new daily cases is now larger than in China, except for the Hubei province where the outbreak originated¹. While the infection risk has become high at the global level, and the disease is rapidly spreading, no effective therapy exists for COVID-2019. Furthermore, no vaccine is likely to become available within 12 - 18 months² and novel drug discovery is known to take several years. Thus, drug repurposing appears as the best strategy to yield efficient therapies against COVID-2019 rapidly.

Drug repurposing can yield new therapies at a faster rate than novel drug discovery when the safety profiles of the drugs being repurposed have been evaluated in the context of drug development for another disease, and at an even faster rate when the drugs have been approved for other diseases and postmarketing safety surveillance data are available^{3,4}. By relying on already known preclinical, pharmacokinetic, pharmacodynamic and toxicity profiles of the drugs being repurposed, one can dramatically increase the rapidity of the response against a disease with unmet clinical needs, especially for an epidemic disease, where drug proven safe can be immediately tested in trials or administered to patients as compassionate treatment. In this context, multiple repurposed drugs have already been tested against COVID-2019, such as Remdesivir (Gilead Sciences) in Phase 3 worldwide (<https://www.gilead.com/news-and-press/press-room/press-releases/2020/2/gilead-sciences-initiates-two-phase-3-studies-of-investigational-antiviral-remdesivir-for-the-treatment-of-covid-19>), Chloroquine phosphate in Phase 4 in China (<http://www.chictr.org.cn/showprojen.aspx?proj=49592>), Carrimycin in Phase 4 in China (<http://www.chictr.org.cn/showprojen.aspx?proj=49514>). As of February 15, more than ten repurposed drugs were ongoing trials against COVID-19 disease⁵.

In silico methods offer a way to methodically and rapidly yield additional repurposing candidates⁶. For instance, when drug targets associated with the disease of interest are known, and when their protein structures or that of close homologs are available, it is possible to use structural bioinformatics to virtually screen (e.g., using molecular docking) a library of existing drugs against this known targets⁷. A study published on February 27, 2020, relied on this approach, using the predicted structure of all SARS-CoV-2 proteins based on their homology with other known coronavirus protein structures, and identified several compounds with potential anti-viral activity⁸.

Another approach to repurposing is the construction of so-called “disease-related molecular networks,” i.e. interactions between gene products (sometimes together with cellular metabolites) involved in the aetiology and symptoms of that disease⁹. There exist several ways to identify disease-related genes, whether using genomic data (e.g., Genome-Wide Association Studies), gene

expression data (e.g., RNAseq differential expression analysis) or data directly collected from the scientific literature (e.g., text mining or expert curation, either analysed in-house or via recognised structured databases). Compared to virtual screening, where the candidate targets are known from the start, network biology methods can identify additional, unanticipated targets, which are part of the same molecular pathways than previously known targets for the disease of interest^{6,10}.

In this study, we performed network bioinformatics analyses to repurpose existing drugs, which are at the completed Phase 2 stage or later, against the now pandemic COVID-2019. We first relied on genome sequence alignment of 2019-nCoV (SARS-CoV-2) to identify SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus) as the most similar virus, followed by MERS-CoV (Middle East Respiratory Syndrome Coronavirus) and other less similar human coronaviruses. We then applied our AutoSeed program, which performed text mining against all NCBI PubMed abstracts (referenced before January 2020) and a systematic database research, which led to 34 COVID-2019-related genes, including ACE2.

To study these disease genes and their role at the systems level, we used an iterative network-building algorithm, AutoNet, that expands, prunes and merges subnetworks, leading to a human COVID-2019 disease network composed of 1,344 genes. In total, 24 enriched pathways were identified in five topological network modules (i.e., community structure, a region where nodes are more densely connected, more likely being related to the same function or a disease¹¹). We scanned this network for known drug-target interactions and applied proximity-based topology analysis¹² to obtain a list of 78 drugs repurposable against COVID-2019. Finally, we manually filtered this list based on the criteria of the drugs' mechanisms of action, their adverse effects, and clinical approvals to yield a total of 30 drugs. In this study, we also discuss the repurposing and mechanisms of thalidomide in particular, since, after sharing our findings with multiple institutions and hospitals in China, one care unit reported the remission of a patient treated with this drug together with low-dose glucocorticoids.

Results

Genome sequence analysis suggests SARS as the most similar disease

After performing a BLASTn search using the SARS-CoV-2 (a.k.a. 2019-nCoV at the time of the analysis) genome sequence against the NCBI GenBank database (see Methods), representative sequences from top results, all being coronaviruses either in humans or other animals, were selected to build a phylogenetic tree using the neighbour-joining method (Figure 1). We find SARS-CoV to be the evolutionarily closest sequence to SARS-CoV-2, with an 80% sequence identity. Among all other human coronaviruses, MERS-CoV is evolutionarily closest to SARS-CoV-2, with a 50% sequence identity. Importantly, we performed this analysis in January 2020, when the virus was less known and studied. Since then, multiple additional sequencing studies have been performed for SARS-CoV-2, including a landmark preprint, which suggested renaming 2019-nCoV to SARS-CoV-2 on the basis of results similar to ours¹³.

Text mining and database searches yield a list of 34 seed genes

In this step, our aim was to identify a list of human genes that are involved in the COVID-2019 disease (Figure 2A). Considering SARS-CoV as the closest virus to SARS-CoV-2, we used SARS as the first keyword for text mining against the database of NCBI PubMed abstracts, downloaded in December 2019. We searched for all human genes co-occurring with the keyword "SARS" (abbreviations, full names, or synonyms) within any sentence (a.k.a "sentence co-occurrence" in NLP methods). We then ranked all genes based on their SARS co-occurrences count.

In order to enrich our text mining results, we added four other terms: "MERS", "coronavirus", "viral pneumonia", and "HIV" (Human Immunodeficiency Viruses). We chose MERS because of its close similarity to SARS-CoV-2 (Figure 1) and the fact that it has been studied for long. "Coronavirus" and "viral pneumonia" were selected because of the nature and symptoms of SARS-CoV-2. Although HIV does not belong to coronaviruses, "HIV" was used as keyword because it was previously reported that HIV and SARS share similar viral protein structures¹⁴ and that HIV drugs can be effective against SARS¹⁵. In addition, there exist extensive research and publication record on HIV, which can enrich our text mining analysis. For these four additional terms, the same co-occurrence analysis was performed, except that only the top 10% of each resulting list was retained. Therefore, the final text-mining-based list was made from the full SARS-related gene output list combined to these four top-10%-retained-gene lists (See Table 1 for sources of extracted texts and papers).

In addition to our in-house text mining analysis, to enrich our search for SARS-related genes, we also searched for the five keywords aforementioned in reference databases including DisGeNET, DrugBank, KEGG, MalaCards, eDGAR, and the GWAS-Catalog, because these databases integrate text-mining results with expert-curated information, from different aspects, including pathways, genetic factors, and animal models (see Methods).

A final list of seed genes was built by overlapping text-mining and database results (see Methods). This list contains 34 genes (shown as a network in Figure 3 and listed in Supplementary Table 1). Among them, 23 genes are directly linked to SARS. Two genes, CRP and TNF, connect to all keywords. Seven genes STAT1, CCL5, ACE2, IRF3, CXCL10, CTSL and TMPRSS2 are linked to four keywords (including SARS).

Network bioinformatics approach helps to predict 30 repurposable drugs

In order to contextualize and better understand, at a systems level, the molecular and physiological role of the COVID-2019-related genes we found, we applied an in-house developed algorithm to build a molecular (i.e., protein) network taking these 34 genes as seeds. This algorithm repeats subnetwork expanding, merging, and pruning in an iterative manner, controlled by pathway enrichment analysis (see Figure 2.B and Methods). In this way, we obtained a final network of 1,344 genes and 24 enriched pathways (Supplementary Table 3). The Newman greedy heuristic module detection algorithm was applied on the network, leading to five modules, representing the T cell receptor signalling pathway, JAK-STAT signalling pathway, C-type lectin receptor signalling pathway, Chemokine signalling pathway and Endocytosis (Figure 2C). At last, DrugBank's drug-target interactions were added to the network over which proximity-based network analysis¹² identified a list of 78 repurposable drugs (Supplementary Table 3).

Having obtained these 78 drugs, we looked for more information, including clinical drug status, drug category, and adverse effect in the Yaozh (<https://data.yaozh.com/>) and DrugBank¹⁶ databases. The former database provides increased China-related information, including clinical trials in China, traditional Chinese medicine usage and theory, approvals by NMPA (National Medical Products Administration, formerly known as CFDA – China Food and Drug Administration) in China, and studies only published in Chinese, while the latter reports approval process by the U.S FDA (Food and Drug Administration), known targets, therapeutic effects as well as basic chemical information¹⁶.

Through a literature review, we identified a list of important symptoms and mechanisms linked to SARS-CoV-2, including fever, fatigue, cough¹⁷, breathing difficulty, septic shock, viral proliferation, immunodeficiency and pulmonary fibrosis¹⁸ (Figure 4). We manually removed a drug from our list if it did not have any reported effect on any of these key symptoms and mechanisms. We also removed a drug from our list if it had strong reported side effects, was not marketed (either in China or U.S.). We also filtered out drugs for which there is little scientific knowledge. After removing these drugs deemed unfit for rapid repurposing, we obtained a list of 30 drugs (Table 1).

Results sharing and case analysis

In order to help fight COVID-2019 as fast as possible, we first publicly shared our list of 78 drugs (Supplementary Table 3) and our list of 24 enriched pathways (Supplementary Table 2) and we very briefly explained our approach with healthcare professionals and hospitals, via GeneNet's WeChat Chinese blog, on February 12, 2020. At the time, we put forward pseudoephedrine, andrographolide, chloroquine, abacavir, baricitinib, and quercetin as repurposing candidates from our list, because there were other researches also suggesting or predicting these drugs, mainly based on our Yaozh database search and literature review. Pseudoephedrine was ranked first by our algorithm. It is an active compound in the Ephedra herb, widely used in as an herb in TCMs to counter flu and colds. Furthermore, Ephedra has been suggested as a TCM treatment against COVID-2019¹⁹. Andrographolide, ranked second by our algorithm, is an active compound from the *Andrographis paniculata* plant, widely used as an antidiabetic herbal compound in traditional medicine. Andrographolide itself has also been proved to be promising in countering pneumonia in animal models²⁰ and predicted by docking simulation²¹. Chloroquine has been considered as one of the most promising repurposed drugs and is currently being tested against COVID-2019 by more than ten clinical trials⁵. Abacavir was also predicted to treat COVID-2019 by two separate studies²². Baricitinib was suggested by the BenevolentAI company using their knowledge graph technology¹⁹. Finally, Quercetin was predicted by a virtual screening study of Chinese herbal medicines²³ and was recently suggested, by Canadian researchers, as a drug to test against COVID-2019 (<https://www.macleans.ca/news/canada/a-made-in-canada-solution-to-the-coronavirus-outbreak/>).

In a second exchange with partner experts from Chinese institutions and care units, via a webinar organized on February 22, we also put forward thalidomide as an interesting repurposing candidate as it was well ranked by our algorithm, the sole drug with anti-fibrosis effect in our list, while it was neither predicted nor tested by another research group. Later, a successful use of thalidomide combined with low-dose glucocorticoid (methylprednisolone) was reported for a 45-year old Chinese woman who had unsuccessfully been treated with ofloxacin (a fluoroquinolone antibiotic), oseltamivir (a.k.a Tamiflu, an anti-viral medication used to treat influenza A and influenza B) and lopinavir + ritonavir (a combination of anti-viral drugs used to treat HIV)²⁴. None of these drugs are in our list of 78 repurposable drugs. Before being treated with thalidomide + methylprednisolone, the patient showed an increase in C-reactive protein (CRP) and cytokine levels, including interleukin 6 (IL-6), interleukin 10 (IL-10) and interferon gamma (IFN-gamma) together with reduced CD4+ and CD8+ T cells counts. The authors reported that these abnormally high interleukin levels and abnormally low T cell levels returned to normal after three days of their combinatorial treatment.

It was previously shown that thalidomide enhances TCR-mediated T cells activation by by-passing T cell need for co-stimulation by accessory molecules, such as the B7 protein together with the CD28 protein, and therefore can overcome T cell deficiency²⁵. In addition, previous work suggests that lenalidomide, a derivative of thalidomide, can restore T cells motility leading to their activation²⁶. Finally, it was also reported that thalidomide prevents NF-kB from binding to the promoters of its target genes, including TNF-alpha and IFN-Gamma thereby reducing excessive inflammatory response^{27,28}. Altogether, based on these previous studies, the reported successful use of thalidomide by Chen et al.²⁴, and our analysis, we hypothesize that thalidomide

can be effective against COVID-2019 by modifying favourably immune response of the infected patients against the virus (Figure 5).

Discussion

In this study, we applied a network bioinformatics approach to repurpose drugs for COVID-2019. Different from structure-based repurposing methods which rely on several known targets, our approach can prioritize potential drug targets and existing drugs at the systems level in response to this global infectious disease threats. Our seed genes (i.e disease related genes) resulted from our AutoSeed program -- a systematic text mining and database search, while our protein network was built by AutoNet, mainly based on knowledge of pathways, protein-protein interaction and graph theory. Combining these results with module detection and proximity analysis algorithms allowed us to identify 78 old drugs repurposable for COVID-2019 disease. Finally, drug database search and manual curation helped shorten our first list to a final list of 30 rapidly repurposable drugs for COVID-2019. To our knowledge, until now, two other studies have investigated the COVID-2019 disease using network-based repurposing. The first one took advantage of a knowledge graph (another type of network comprising different entity types, such as gene, protein, organism and disease, and relationship types, such as interacting with, phosphorylating, belonging to, etc.) technology to suggest baricitinib as potential treatment²⁹. A second study recently deposited on MedRxiv, used, in part, similar network techniques than reported in this study, although the main difference is that we relied on text-mining and database search for seed genes identification while they essentially relied on the use of transcriptomic data for enrichment analysis³⁰.

While our research succeeded in predicting some promising drugs, we emphasize that it is our first time applying our network bioinformatics pipeline, which was developed for common diseases, in our CloudPhar platform (<http://tcm.tasly.com>), to a virus epidemic, because of our wish to respond to the global emergency caused by COVID-2019. This implies that there is extensive room for improvement. For example, latter studies could benefit from modern text mining technique which have now improved accuracies thanks to the latest deep-learning-powered natural language processing (NLP) technologies. In the present study, we used a relatively simple technique named sentence co-occurrence, because virus names were not yet implemented into our NLP system at the time of the outbreak. Otherwise, NLP could accurately detect virus entity and gene/protein entity and their relationships.³¹ Network building, and analysis steps can also be improved. While we had mainly studied common diseases (vascular heart diseases in particular) and trained deep learning models, this resource could not directly be used for COVID-2019, given substantial differences between them and the virus epidemic. However, machine learning-based network approaches have improved our capacity to analyse big data at the systems level³², and we could build, in the future, a viral-disease-dedicated analysis pipeline using virus-related data for training.

At this stage, our hope is that these results can be helpful in rapidly designing and implementing clinical trials to treat COVID-2019. We also hope that, in the near future, some of the prioritized drugs could be used in combination to provide an even better, and possibly personalized, treatment against COVID-2019 as it is known that combinatorial therapies can be more efficient against many types of diseases⁶, including viral diseases such as HIV³³. In the bioinformatics field, while it is still a challenging task to predict synergistic therapies (some available methods have been recently studied for cancers)³⁴, a simple approach is to combine drugs with different therapeutic effects or affecting different pathways or biological functions, and we hope that our study (Figure 4) together with existing experimental and clinical results could provide helpful hints. At last, we would like to emphasise that Traditional Chinese Medicines (TCMs), which are by nature combination therapies, are now being used to treat COVID-2019 in China^{19,35}. It is noteworthy that, for instance, the TCMs Ma-Xing-Shi-Gan-Tang (MXSGT) or Qing-Fei-Pai-Du-Tang made with Ephedra extracts contains Pseudoephedrine, a molecule with a top rank in our study. Multiple clinical trials (e.g. <http://www.chictr.org.cn/showproj.aspx?proj=50248>) in China are testing MXSGT, which was previously shown to be efficient against lung microvessel hyperpermeability and inflammatory reaction in rat³⁶. In addition, this TCM has been officially suggested by the China's National Health Council and the National Administration of TCMs (source in Chinese: <http://m.gxfin.com/article/finance/gu/default/2020-02-09/5189141.html>) against COVID-2019

Methods

Genome sequence analysis

From NCBI GenBank, the complete genome of Wuhan-Hu-1 (NC_045512.2) was downloaded as the 2019-nCoV sequence. This genome sequence was used to search for closely related viruses, against the whole database using BLASTn (default parameters except that we obtained more results than 100 by default). Among the BLASTn results, we extracted the following complete genome sequences as representative to build a phylogenetic tree: Bat coronavirus BM48-31/BGR/2008 (NC_014470.1), Bat coronavirus CDPHE15/USA/2006 (NC_022103.1), NL63-related bat coronavirus strain BtKYNL63-9a (NC_032107.1), Bat coronavirus HKU8 (NC_010438.1), Bat coronavirus HKU5-1 (NC_009020.1), SARS-CoV (NC_004718.3), MERS-CoV (KX034094.1), Bat coronavirus 1A (NC_010437.1), Human coronavirus OC43 (KX344031.1), Human coronavirus NL63 (JX504050.1), Human coronavirus HKU1 (KF686340.1), Human coronavirus 229E (KY996417.1), Bat coronavirus HKU4-1 (NC_009019.1), Scotophilus bat coronavirus 512 (NC_009657.1), Rousettus bat coronavirus HKU9 (MG762674.1), Rousettus bat coronavirus HKU10 (NC_018871.1), Bat coronavirus HKU2 (NC_009988.1) and Porcine coronavirus HKU15 (NC_039208.1). Multiple sequence alignment was calculated by EMBL-EBI's MSA tool (<https://www.ebi.ac.uk/Tools/msa/>) using default parameters. A tree was built using the neighbour-joining method with the MEGA-X software³⁷, using the maximum composite likelihood model and 1000 bootstraps. The resulting tree was represented using the phylogram format (i.e., a tree

branch lengths are proportional to the amount of inferred evolutionary change)³⁸.

Related genes identification

PubMed (version 2019-12) was downloaded from its FTP site. Note that no article mentioning 2019-nCoV or SARS-CoV-2 had been published before that date, meaning that our text mining analysis did not directly consider the COVID-2019 disease. Instead, it aims at predicting the network base on closely related viruses and their physiology. More than 29 million abstracts were processed for sentence and word tokenization by the natural language processing tool Spacy (v2). Inputted keywords of interest (SARS, MERS, coronavirus, viral pneumonia, and HIV) were extracted by exact matches to detect abbreviations or regression expressions in order to detect full names or synonyms. Entity recognition for genes was proceeded by mapping gene names and unambiguous synonyms from the HGNC database. Co-occurrence numbers were counted by the number of papers where a pair of gene and an input entity was in one sentence. A list of related genes ranked by sentence co-occurrence numbers was obtained for each of the five input entities. The final text-mining resulting list (the network shown in Figure 2) was built from the whole list for SARS and the top 10% of each of the other four lists.

Database search for related genes was performed by a program developed in-house, AutoSeed, which can search for disease-related genes in the following databases: DisGeNET³⁹, DrugBank¹⁶, KEGG⁴⁰, Malacards⁴¹, eDGAR⁴², NHGRI-EBI GWAS-Catalog⁴³. Note that this program was developed for all types of diseases, and not specifically for viral diseases. Its function is to interrogate all of these databases automatically and to return a list of related genes sorted by the number of times they occur in those databases. Although the GWAS-Catalog is one of the resources of AutoSeed, for SARS and MERS, because there are no published GWAS, the findings in that category are, as expected, null. The final database-based list was composed of the whole list for SARS and the top 10% of each of the other four lists.

Network building

Network building was performed automatically by another of our in-house program, AutoNet, originally developed for our drug discovery cloud platform. This algorithm is illustrated by a schematic diagram in Figure 2B.

Data for this step includes a local meta-pathway database for pathway enrichment analysis and a meta-PPI database to grow the network. The meta-pathway database is made of human pathways in KEGG⁴⁰ and Reactome (v70)⁴⁴ databases, after removing small pathways (less than five genes) and pathways which enrich too easily, such as hsa05200: Pathways in cancer. The meta-PPI database is composed of protein-protein interaction databases HPRD⁴⁵, BioGrid⁴⁶ (excluding genetic interactions), and STRING⁴⁷ (excluding PPIs with confidence score < 0.7).

The building process repeats network expanding – merging and pruning in an iterative manner. At the initial state, all seed genes are considered as positive nodes (nodes which will be kept at last, except that we remove all nodes from a subnetwork which cannot be merged with any other subnetworks from the beginning to the end), each seed gene is a subnetwork composed of one node and a dynamic pathway collection for network building is initiated by pathway enrichment analysis (adjusted p-value < 0.001) against our meta-pathway database. In each expanding step, direct partners of any positive nodes are added as temporary nodes using our meta-PPI database. In each merging step, only the pair of subnetworks that share the most positive nodes and temporary nodes are merged, while the other subnetworks wait to be merged in the next iterations. In the subnetwork pruning step, those temporary nodes which are not in any of the pathway collection in the current state are removed. Remaining nodes become positive nodes, and the dynamic pathway collection is updated using pathway enrichment analysis.

Sub-networks are grown until they cannot be further merged. If more than one subnetwork remains, only the biggest one is kept as the final network.

Network-based drug repurposing

After the network was built, core modules were detected (Figure 2C), using the Newman greedy heuristic algorithm⁴⁸, implemented in *igraph* package (v1.2.4.2) in the R language (version 3.5.3). Potential drugs were then mapped to the COVID-19 network through drug-target interactions (source from DrugBank). As shown in figure 2D, different drugs can be linked to one or more different modules (shown as colored areas) in the network. In order to find the maximum effective coverage of the core functional modules for each drug, we used a proximity method with each drug proximity distance calculated as the mean value of the shortest distances between any drug and each of the core modules in the space (equation shown in Figure 2D)¹².

References

1. WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 35. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200224-sitrep-35-covid-19.pdf?sfvrsn=1ac4218d_2 (2020).
2. The Lancet. COVID-19: fighting panic with information. *Lancet* **395**, 537 (2020).

3. Ashburn, T. T. & Thor, K. B. Drug repositioning: Identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* **3**, 673–683 (2004).
4. Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* (2018) doi:10.1038/nrd.2018.168.
5. Harrison, C. Coronavirus puts drug repurposing on the fast track. *Nat. Biotechnol.* (2020) doi:10.1038/d41587-020-00003-1.
6. Nabirotkin, S. *et al.* Next-generation drug repurposing using human genetics and network biology. *Curr. Opin. Pharmacol.* 1–15 (2020) doi:10.1016/j.coph.2019.12.004.
7. Glaab, E. Building a virtual ligand screening pipeline using free software: A survey. *Brief. Bioinform.* **17**, 352–366 (2016).
8. Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* (2020) doi:https://doi.org/10.1016/j.apsb.2020.02.008.
9. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: A network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
10. Hopkins, A. L. Network pharmacology. *Nat. Biotechnol.* **25**, 1110–1111 (2007).
11. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8577–8582 (2006).
12. Guney, E., Menche, J., Vidal, M. & Barabási, A. L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 1–13 (2016).
13. Gorbalenya, A. E. *et al.* Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. 1–15 (2020) doi:10.1101/2020.02.07.937862.
14. Kliger, Y. & Levanon, E. Y. Cloaked similarity between HIV-1 and SARS-CoV suggests an anti-SARS strategy. *BMC Microbiol.* **3**, 1–7 (2003).
15. Yamamoto, N. *et al.* HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochem. Biophys. Res. Commun.* **318**, 719–725 (2004).
16. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
17. Xu, Y.-H. *et al.* Clinical and computed tomographic imaging features of Novel Coronavirus Pneumonia caused by SARS-CoV-2. *J. Infect.* (2020) doi:10.1016/j.jinf.2020.02.017.
18. Shi, H. *et al.* Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet. Infect. Dis.* **3099**, 1–10 (2020).
19. Jin, Y. H. *et al.* A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil. Med. Res.* **7**, 1–23 (2020).
20. Zhang, G. *et al.* Treatment with andrographolide sulfonate provides additional benefits to imipenem in a mouse model of *Klebsiella pneumoniae* pneumonia. *Biomed. Pharmacother.* **117**, 109065 (2019).
21. Wu, C. *et al.* Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* (2020) doi:https://doi.org/10.1016/j.apsb.2020.02.008.
22. Beck, B. R., Shin, B., Choi, Y., Park, S. & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model. *bioRxiv* 2020.01.31.929547 (2020) doi:10.1101/2020.01.31.929547.
23. Zhang, D., Wu, K., Zhang, X., Deng, S. & Peng, B. In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *J. Integr. Med.* (2020) doi:https://doi.org/10.1016/j.joim.2020.02.005.
24. Chen, C. *et al.* Thalidomide Combined with Low-dose Glucocorticoid in the Treatment of COVID-19 Pneumonia. 1–6 (2020).
25. Bartlett, J. B., Dredge, K. & Dalglish, A. G. The evolution of thalidomide and its IMiD derivatives as anticancer agents. *Nat. Rev. Cancer* **4**, 314–322 (2004).
26. Riches, J. C. & Gribben, J. G. Immunomodulation and Immune Reconstitution in Chronic Lymphocytic Leukemia. *Semin. Hematol.* **51**, 228–234 (2014).
27. Keddie, S. *et al.* Clinical perspectives into the use of thalidomide for central nervous system tuberculosis. *Eur. J.*

Neurol. **25**, 1345–1351 (2018).

28. Wen, H. *et al.* Recurrent ECSIT mutation encoding V140A triggers hyperinflammation and promotes hemophagocytic syndrome in extranodal NK/T cell lymphoma. *Nat. Med.* **24**, 154–164 (2018).
29. Richardson, P. *et al.* Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet (London, England)* **395**, e30–e31 (2020).
30. Zhou, Y. *et al.* Network-based Drug Repurposing for Human Coronavirus. *medRxiv* (2020). doi:10.1101/2020.02.03.20020263.
31. Esteva, A. *et al.* A guide to deep learning in healthcare. *Nat. Med.* **25**, 24–29 (2019).
32. Sügis, E. *et al.* HENA, heterogeneous network-based data set for Alzheimer’s disease. *Sci. data* **6**, 151 (2019).
33. Mendoza, P. *et al.* Combination therapy with anti-HIV-1 antibodies maintains viral suppression. *Nature* **561**, 479–484 (2018).
34. Preuer, K. *et al.* DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538–1546 (2018).
35. Luo, H. *et al.* Can Chinese Medicine Be Used for Prevention of Corona Virus Disease 2019 (COVID-19)? A Review of Historical Classics, Research Evidence and Current Prevention Programs. *Chin. J. Integr. Med.* **11655**, 1–8 (2020).
36. Ma, L.-Q. *et al.* Posttreatment with Ma-Xing-Shi-Gan-Tang, a Chinese Medicine Formula, Ameliorates Lipopolysaccharide-Induced Lung Microvessel Hyperpermeability and Inflammatory Reaction in Rat. *Microcirculation* **21**, 649–663 (2014).
37. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
38. Pavlopoulos, G. A., Soldatos, T. G., Barbosa-Silva, A. & Schneider, R. A reference guide for tree analysis and visualization. *BioData Min.* **3**, 1 (2010).
39. Piñero, J. *et al.* DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).
40. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).
41. Rappaport, N. *et al.* MalaCards: An amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* **45**, D877–D887 (2017).
42. Babbi, G. *et al.* eDGAR: A database of disease-gene associations with annotated relationships among genes. *BMC Genomics* **18**, (2017).
43. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
44. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
45. Keshava Prasad, T. S. *et al.* Human Protein Reference Database - 2009 update. *Nucleic Acids Res.* **37**, 767–772 (2009).
46. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541 (2019).
47. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).
48. Newman, M. E. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* **69**, 066133 (2004).

Acknowledgments

We would like to thank Yunhui Hu, Jia Sun and Hairong Wang from GeneNet, Serguei Nabirotkin from Pharnext, Zu Liu for fruitful discussions, and thank Zu Liu for his help in genome analysis.

Author contributions

X.L. conceived the study, J.Y. and A.E.P. wrote the manuscript, X.L., Z.Z. and J.R. conducted the bioinformatics analyses, X.L., J.Y., Z.Z. and J.R. analysed the results, W.W. supervised the study. All authors reviewed the manuscript.

Additional information

Supplementary information

Competing interests: X.L., Z.Z., J.R., W.Z., Y.Z., K.Y. and W.W. are employees of GeneNet Pharmaceuticals. J.Y., A.E.P. and D.C. are employees of Pharnext.

Figures

Figure 1. Sequence analysis suggests SARS-CoV as the most similar virus to the 2019-nCoV. Nineteen genome sequences were aligned using EMBI-EBI’s MSA tool, and a neighbour-joining phylogenetic tree was built by the MEGA-X tool. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown above the branches. The scale represents 0.10 residue substitutions per site.

(A) Related genes identification (AutoSeed)

(C) Module detection

(B) Network building (AutoNet)

(D) Network-based Drug prioritization

Figure 2. The workflow of our network bioinformatics pipeline for SARS-CoV-2 drug repurposing. The equation shown in (D) represents how the distance was calculated to prioritize drugs based on proximity.

Figure 3. Thirty-four genes related to SARS-CoV-2 identified by text mining and database searches. Each link represents at least one sentence co-occurrence in PubMed abstracts or at least one relationship recorded in one of our searched databases.

Figure 4. Symptoms and mechanisms related to SARS-CoV-2 and the corresponding categories of our 30 suggested drugs.

Figure 5. Thalidomide’s potential Mechanism-of-Action on COVID-2019.

Tables

DrugBank ID	Drug name	Type	Group	Initial rank	Category
DB00852	Pseudoephedrine	small molecule	Approved	1	Antipyretic or analgesic; antiasthmatic; anti-inflammatory
DB05767	Andrographolide	small molecule	investigational	2	Antipyretic or analgesic; antiviral; anti-bacterial; anti-inflammatory
DB05513	Atiprimod	small molecule	investigational	3	Immunomodulator
DB05017	YSIL6	small molecule	investigational	8	Immunomodulator
DB06083	Tapinarof	small molecule	investigational	11	anti-inflammatory
DB00005	Etanercept	biotech drug	Approved, Investigational	12	Antipyretic or analgesic; anti-inflammatory
DB00051	Adalimumab	biotech drug	Approved	13	Antipyretic or analgesic; anti-inflammatory
DB00065	Infliximab	biotech drug	Approved	14	anti-inflammatory
DB00608	Chloroquine	small molecule	approved; investigational; vet_approved	15	anti-bacterial; anti-inflammatory
DB00668	Epinephrine	small molecule	approved; vet_approved	16	Antiasthmatic
DB01041	Thalidomide	small molecule	approved; investigational; withdrawn	17	anti-fibrosis; Immunomodulator
DB01407	Clenbuterol	small molecule	approved; investigational; vet_approved	18	Antiasthmatic
DB01411	Pranlukast	small molecule	investigational	19	Antiasthmatic
DB04956	Afelimomab	biotech drug	Investigational	21	Immunomodulator
DB06674	Golimumab	biotech drug	Approved	32	Antipyretic or analgesic; anti-inflammatory
DB09036	Siltuximab	biotech drug	Approved, Investigational	35	Anti-viral
DB01250	Olsalazine	small molecule	Approved	36	anti-inflammatory
DB12698	Ibalizumab	biotech drug	Approved, Investigational	39	anti-HIV
DB01327	Cefazolin	small molecule	Approved	43	anti-bacterial
DB01048	Abacavir	small molecule	approved; investigational	50	anti-HIV
DB02375	Myricetin	small molecule	experimental	51	anti-inflammatory
DB04464	N-Formylmethionine	small molecule	experimental	52	Immunomodulator
DB06475	Ruplizumab	biotech drug	Investigational	54	Immunomodulator
DB00452	Framycetin	small molecule	Approved	60	anti-bacterial
DB01009	Ketoprofen	small molecule	approved; vet_approved	65	anti-inflammatory
DB04835	Maraviroc	small molecule	approved; investigational	66	anti-HIV
DB06652	Vicriviroc	small molecule	investigational	69	anti-HIV
DB00172	Proline	small molecule	approved; nutraceutical	75	Nutrition
DB04216	Quercetin	small molecule	experimental; investigational	76	Antiasthmatic
DB11638	Arteminol	small molecule	experimental; investigational	78	anti-bacterial

Table 1. Thirty drugs predicted to repurpose for COVID-2019 disease. Type and Group were obtained by querying DrugBank. Initial ranks came from our proximity-based drug prioritization algorithm. Categories were obtained from Youzh (Chinese drug database), DrugBank and manual curation when the data was not available in neither of these databases.