**Title:**

Drug repurposing to target COVID-19: a big data driven approach

建立用於新冠肺炎研究及藥物開發的知識圖譜平台

**Keywords:** COVID-19, knowledge graph, drug repurposing

**Project Duration:** 24 months

**Funding requested:** $2M-$10M

PC: Dr. Esther Chan

Co-PIs: Prof. Reynold CK Cheng, Prof. Ian CK Wong, Dr. Ruibang Luo, Dr. Chuan Wu, Prof. Benjamin Cowling, Prof. Ivan FN Hung

Co-Is: Dr. Xiaodong Li, Dr. Bo Tang, Ms. Min Ou, Mr. Vincent Yan

**2. Project Objectives**

The objective of the proposed project is to develop an open knowledge-graph-based platform to facilitate global COVID-19 research, and to utilize the platform for discovering potential drug candidates for COVID-19 treatment.

Specifically:

1. To develop an open COVID-19 knowledge graph that will enable researchers worldwide to uncover hidden relationships between COVID-19, drugs, genes, proteins and diseases/symptoms, providing a platform for hypothesis generation and future research on COVID-19.
2. To identify drugs and combinations of drugs that have potential to be repurposed for COVID-19 treatment using a COVID-19 knowledge graph.

**3. Proposed Outline of Research Plan, Methodology** **and Target Achievement**

COVID-19 has emerged as a severe global epidemic, with millions of people infected and causing tens of thousands of deaths worldwide. Yet to date, there is no effective drug treatment or vaccine available to treat or prevent COVID-19 and its complications. There is an urgent need for efficient tools and data platform for COVID-19 drug repurposing and research. Here, we propose two interlinked projects to address these unmet needs in COVID-19 research.

**Project 1: Building an open knowledge graph for COVID-19 research**

**Rationale:**

Knowledge graphs enable us to identify valuable information regarding the large-scale, complex relationships between drug, disease, virus, proteins and genes. Modern computational algorithms can be used to extract hidden linkages from knowledge graphs to generate useful insights and testable hypotheses. Some of the applications possible with the use of knowledge graphs include drug repurposing, identification of potential adverse effect of drugs, and disease subtyping.(1)

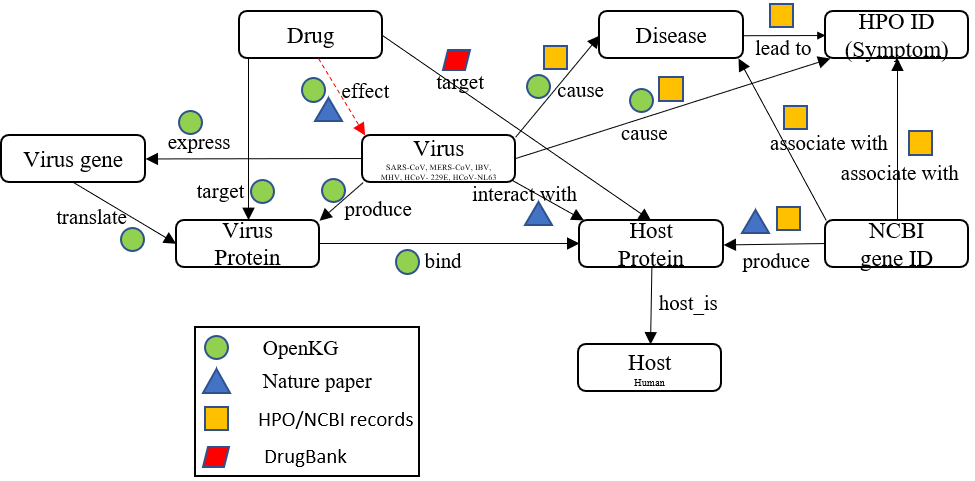
However, to date, there is no unified COVID-19 knowledge graph openly available for use in COVID-19 research. While there have been recent publications on knowledge graphs linking COVID-19 publications, case statistics and genes, (2) they are generally limited in coverage and does not include drug and protein relationships which are essential for discovery of potential drug treatment. Individual data sources are also available providing information on drugs, drug-protein, gene-disease and virus-disease relationships, (3, 4) yet these data sources are all independent and have not be linked into a unified knowledge graph. A comprehensive, unified knowledge graph is needed to enable efficient extraction of high-level linkages, such as existing drugs repurposable for another disease. These high-level linkages provide promising leads and enable faster discovery of potential treatments and unknown intrinsic characteristics of new diseases.

**Objectives:** To develop an open COVID-19 knowledge graph which will enable researchers worldwide to uncover hidden relationships between COVID-19, drugs, genes, proteins and diseases/complications, providing a big data platform for hypothesis generation and future research on COVID-19.

**Methods:**

The final COVID-19 knowledge graph will have the schema as shown in the figure, covering linkages between COVID-19, drugs, genes, proteins and diseases. A variety of existing data sources would be used to build this knowledge graph, including but not limited to:

* OpenKG: OpenKG provides open knowledge graphs for specific subtopics and themes, including graphs that link existing drugs to viruses, drugs to viral proteins, viruses to diseases, viruses to viral proteins, viral proteins to host proteins, among other linkages.
* DrugBank: The DrugBank database (4) is a comprehensive, freely accessible database containing information on drugs and drug targets. It currently contains 13,575 drug entries and is widely used by industry, medical practitioners and the general public. It has enabled the discovery and repurposing of a number of existing drugs to treat rare and newly identified illnesses and serve as a source of drug-related data in this knowledge graph.
* Human phenotype ontology (HPO): HPO (3) is central in medical genetics and genomics. It provides a comprehensive bioinformatic resource for analysis of human diseases and phenotypes and serves as a computational bridge between genome biology and clinical medicine. It will be used to link disease and gene data in the knowledge graph.
* Literature: existing publications which released integrated datasets would also be incorporated, if relevant, in building the knowledge graph.



**Target achievement and impact:** By linking a variety of scattered data sources into a unified knowledge graph, previously unexplored relationships between COVID-19, diseases, drugs, genes, and patient demographics, could be efficiently extracted. These hidden linkages may provide insights on previously unknown relations of how underlying diseases, concomitant medications or genetic profile may influence a patient’s severity or chance of COVID-19 infection, and to explore any potential differences among different sex or age groups such as elderly. Subsequently this would shed light on ways to prevent or reduce severity of COVID-19 in specific patient populations.

**Project 2: Graph-based drug repurposing for COVID-19 treatment**

**Rationale:**

To date, there is no specific drug treatment nor vaccine available to treat or prevent COVID-19 and its complications. As such, there is an urgent need for developing effective strategies for prevention and treatment of COVID-19. While conventional structure-based screening methods such as protein docking analyses are traditionally used for *de novo* drug discovery, repurposing of existing drugs provide a more cost and time efficient way to discover treatment for new diseases.(5)

Globally, antimalarial and antiviral agents are currently under trial for treatment of COVID-19.(6, 7) While these may have shed hope on treating COVID-19, yet from experience of other diseases, only a small proportion of trialed drug candidates would actually succeed to show efficacy and be approved for treatment. Indeed, preliminary results from some of these trials suggest that certain candidates may not be as promising as speculated. For instance, in a randomized trial from China of 237 patients with severe COVID-19, remdesivir and placebo did not show significant difference in times to clinical improvement (median 21 versus 23 days) and mortality rates (14 versus 13 percent).(8) Interim analysis of an unpublished trial comparing remdesivir with placebo in >1000 patients with confirmed COVID-19 and pulmonary involvement suggested that there was no significant difference in mortality (8 versus 11.6 percent with placebo) despite faster time to recovery (median 11 versus 15 days).(9)

On the other hand, antimalarial and antiviral agents are generally associated with undesirable adverse effect profile and drug-drug interactions. For instance, hydroxychloroquine and chloroquine are known to prolong QTc interval. In an observational study where 84 patients received hydroxychloroquine, 10% had electrocardiographic changes which required discontinuation.(10) Hence, efficient discovery of more drug candidates, especially safer alternatives, is essential.

Currently, speculation on drug repurposing for COVID-19 treatment had been focused mainly on antiviral and antimalarial agents, which were known to exert antiviral effect on HIV, Ebola and other viruses, but not proven effective for SARS-CoV-2. Antiviral drug combinations, such as cocktail therapy using two or more antiviral agents, are also being explored. However, repurposing of non-antiviral agents had not been adequately explored, less intuitive linkages between non-antiviral agents and COVID-19. Yet, it had been shown that non-antiviral agents could also influence host’s response to viral infections, by reducing the chance of viral entry into cells or suppressing overreaction of the immune system to the virus. For instance, Angiotensin II Receptor Blockers (ARBs) may antagonize the proinflammatory effects of angiotensin II which is increased due to COVID-19 infection.(11) Further, cardiovascular medications may improve survival as COVID-19 have been reported to have cardiovascular sequelae.

Yet, with the vast number of existing drugs available, experimental approaches to find potential repurposable drugs would be costly and time-consuming. Efficient computational methods to extract potentially repurposable drugs and to rank them in terms of clinical relevance are essential to allow timely discovery of potential COVID-19 treatment.

**Objectives:** To identify drugs and combination of drugs that have potential to be repurposed for COVID-19 treatment.

**Methods**: Linkages relevant to drug repurposing will be extracted from a large-scale, comprehensive COVID-19 knowledge graph generated from project 1, using motif-search algorithms. Motifs represent high-order connection patterns of interest in knowledge graphs. Examples of motifs relevant to drug repurposing include drug-disease-drug, drug-gene-virus-gene-drug, drug-gene-disease-symptom-disease-gene-drug and others which will be further defined and explored. Subgraphs that matches the motifs of interest will then be extracted using specific motif-clique discovery algorithms. These subgraphs will provide information on drug candidates potentially repurposable for COVID-19 treatment. Further information regarding these drug candidates, such as potential adverse effects and biological mechanisms supporting their use will be further extracted from the knowledge graph and reviewed from literature to rank them in terms of clinical applicability. Combinations of antiviral agents, immunosuppressants and cardiovascular medications will also be explored in a similar way for their potential of repurposing for COVID-19 treatment.

**Target achievement and impact:** Findings will reveal novel drug candidates and combinations of drugs from different drug classes that have the potential to be repurposed for COVID-19 treatment. The efficacy of these drug candidates for COVID-19 treatment can then be further validated via randomized clinical trials and multi-centre Big Data observational studies using electronic medical records, which can be translated to clinical practice and form the treatment armamentarium for COVID-19.

**References**

1. Li B, Cheng R, Hu J, et al. MC-Explorer: Analyzing and Visualizing Motif-Cliques on Large Networks. IEEE International Conference on Data Engineering 2020. p. 1722-5.

2. The CovidGraph Project. CovidGraph. 2020. Available from: <https://covidgraph.org/>.

3. Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017;45(D1):D865-D76.

4. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-d82.

5. Cheng F. In Silico Oncology Drug Repositioning and Polypharmacology. Methods Mol Biol. 2019;1878:243-61.

6. Sanders JM, Monogue ML, Jodlowski TZ, et al. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020;323(18):1824-36.

7. Hung I, Lung K-C, Tso E, et al. Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. The Lancet. 2020.

8. Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. The Lancet. 2020;395(10236):1569-78.

9. Beigel JH, Tomashek KM, Dodd LE, et al. Remdesivir for the Treatment of Covid-19 — Preliminary Report. N Engl J Med. 2020.

10. Mahévas M, Tran V-T, Roumier M, et al. Clinical efficacy of hydroxychloroquine in patients with covid-19 pneumonia who require oxygen: observational comparative study using routine care data. BMJ. 2020;369.

11. Rothlin RP, Vetulli HM, Duarte M, et al. Telmisartan as tentative angiotensin receptor blocker therapeutic for COVID-19. Drug development research. 2020.

**4. Collaboration Plan** (PC and Co-PIs only)

Experts from pharmacology and pharmacy, computer science, public health and medicine formed a multidisciplinary team to support this application. This collaboration combines the strengths of each discipline by applying modern data, graph and computational techniques to connect the basic sciences in pharmacy and medicine for drug repurposing, to search for potential COVID-19 treatment and provide an open platform to facilitate global research on COVID-19.

**Expertise in Pharmacology & Pharmacy and Big data**

The PC, Dr. Esther WY Chan is a pharmacist with expertise in medication safety and effectiveness. She has led completed multicentre randomised controlled trials in Hong Kong and Australia (ECS/RGC, NHMRC); and completed two GRF/RGC funded big data projects. She will lead and supervise the research team in all aspects of this study, including literature review, research methodology, interpretation of knowledge graph outputs in a clinical context, generation of research reports and dissemination of findings.

Co-PI, Prof. Ian CK Wong is an expert in using healthcare big data with a strong track record of leading large research programmes on medication use and safety. Together with the PC, he will contribute to the interpretation of results regarding potential drug candidates and provide guidance on study feasibility, refine methodology, and contribute to data interpretation and research dissemination.

**Expertise in Bioinformatics, Health informatics, Big graphs**

Co-PI, Dr Ruibang Luo, is experienced in bioinformatics algorithm and health informatics. His research results have been published in peer-reviewed journals including Nature, Nature Biotechnology, Nature Communications, and Bioinformatics. He made the 2017 list of Forbes 30 Under 30 Asia in Healthcare and Science, and he made the 2019 list of Innovators Under 35 Asia Pacific by MIT Technology Review. He will assist in the application of knowledge graph techniques for COVID-19 research and drug repurposing.

Co-PI, Dr. Chuan Wu is an active researcher in cloud computing and the modeling, analysis, design, and implementation of large-scale network systems. She has served as program committee members and review panels for leading database conferences and journals. She will assist in the implementation and storage of large-scale knowledge graphs.

Co-PI, Prof. Reynold CK Cheng is a Professor in computer science with expertise in large-scale data management and querying and mining of uncertain data. He has led projects involving application of heterogeneous information networks in solving real world problems and developed efficient query algorithms for large-scale graph databases. He has served on the program committees and review panels for leading database conferences and journals like SIGMOD, VLDB, ICDE, and TODS. He will provide the database tools to build large-scale COVID-19 knowledge graph and apply computational techniques for drug repurposing and research using knowledge graphs.

**Expertise in Infectious diseases, Medicine and Public Health**

Co-PI, Prof. Ben Cowling is an expert in infectious disease epidemiology and public health and is leading cutting edge research in COVID-19. He has more than 400 peer-reviewed publications including research articles in top journals in medicine, respiratory system and infectious diseases. He will provide valuable input on research methodology and clinical interpretation of study findings.

Co-PI, Prof. Ivan FN Hung is a specialist in Infectious Disease and leading clinical trialist for COVID-19. He has published more than 150 international peer reviewed original articles, including research articles in the Lancet, the Lancet Infectious Diseases and the Clinical Infectious Diseases and is an expert in innovative treatment of severe influenza. He will provide valuable input on the clinical interpretation and application of study findings.

**5. Pathways to Impact**

COVID-19 has emerged as a severe global epidemic. By the end of May 2020, the number of confirmed cases worldwide had reached over 5 million. The lack of specific drug treatment for COVID-19 had contributed to more than 350,000 deaths worldwide with 100,000 deaths in the United States alone. While the situation in Hong Kong is gradually under control, the situations in Europe and America have not yet shown clear signs of improvement, and the infection is still actively spreading with more than 100,000 new cases per day worldwide. These countries are still hopeful for specific COVID-19 drug treatments to emerge, which could potentially save tens of thousands of lives globally.

Unfortunately, although a number of antiviral and antimalarial agents are currently under trial for evaluating their efficacy as COVID-19 treatment, preliminary results suggest that some of these agents may not be as promising as speculated (8), whereas others are associated with serious adverse effects such as electrocardiographic changes which limits their usage.(10) Hence, effective data platform and tools are essential to enable efficient discovery of new drug candidates and drug combinations, especially safer alternatives.

The proposed project aims to provide a knowledge-graph-based platform and tools that will enable researchers worldwide to efficiently extract previously less intuitive linkages between COVID-19, drugs, genes, proteins and diseases. This will further facilitate efficient discovery of potential COVID-19 treatments that can be tested via clinical trials and Big Data observational studies, and subsequently be translated to clinical use.

The proposed research is expected to have local and international impacts as illustrated below.

1. **Who are the potential beneficiaries of the proposed research in the short (1-3 years), medium (4-10 years) and long term (over 10 years)?**

*Short-term*: healthcare professionals, government, medical researchers worldwide

*Medium-term*: COVID-19 patients, their family and caregivers; healthcare system users and the society as a whole

*Long-term*: the global population, patients with currently incurable or chronic diseases, especially the growing elderly population who are prone to infections and chronic diseases.

1. **How will the potential beneficiaries benefit? What will be the objective demonstrable/measurable benefits beyond academia?**

In the short term, hidden or less obvious linkages between COVID-19, diseases, drugs, genes, and patient demographics, can be discovered. These hidden linkages may provide insight on previously unknown relations of how sex and age, underlying diseases or concomitant medications may influence a patient’s severity or chance of COVID-19 infection. Subsequently this would shed light on ways to prevent or reduce severity of COVID-19 in specific patient populations, through randomized clinical trials and population-based studies. This could potentially reduce the severity of the outbreak and hasten the resumption of normal work and economic productivity, in Hong Kong as well as Europe and America where the outbreak is still not under control. Findings will also help inform treatment decisions, prioritization of healthcare resources, and design and implementation of public health policies by the government.

In the medium-term, new treatments for COVID-19, originating from the drug candidates extracted from our knowledge graph, which have proven to have benefit may have new added indications or approved for COVID-19 treatment and made accessible to patients worldwide. While COVID-19 epidemic seems to come under control in recent months in Hong Kong, the situation in Europe and America are still severe and actively spreading, and patients in these regions are very much in need of new potential treatments. Further, as COVID-19 or its variants could potentially be a recurrent epidemic, availability of new COVID-19 treatment could significantly reduce mortality and morbidity due to COVID-19 worldwide, potentially saving tens of thousands of lives, particularly in older individuals with chronic health conditions.

Furthermore, the outbreak of COVID-19 and the lack of specific drug treatment had led to an unprecedented demand on healthcare system in countries worldwide. Indeed, the saturation of healthcare system capacity due to COVID-19 had also interfered with the normal course of treatment of patients with other diseases such as psychiatry and cancer. Patients with acute conditions requiring emergency treatments, such as those with acute myocardial infarction, suffered significant delays resulting in complicated in-hospital course and worse clinical outcomes. As such, availability of new COVID-19 treatment will immensely benefit not only COVID-19 patients and their caregivers, but also other patients and users of the healthcare system, as well as the society as a whole. Such benefits are demonstrable by shortened hospital length of stay, reduction in service delays and time from symptoms to first medical contact for emergency medical conditions, as well as reduced healthcare costs due to COVID-19 and its complications.

In the long-term, the proposed project provides powerful tools and experience for knowledge-graph-based drug repurposing. These knowledge graphs, tools and techniques could be applied to any infectious disease that emerges in the future, and allow quick identification and testing of new treatments, which can potentially prevent large-scale public health crisis due to new viral or infectious diseases, and allow early control of any potential outbreak. On the other hand, these knowledge-graph based techniques could also be extended to help identify treatments for currently incurable diseases and chronic diseases as well, which could potentially reduce the economic and societal burden associated with these diseases worldwide. These benefits could be demonstrated by a shortened time to discovery of treatments for new diseases and increased number of curative treatments for previously incurable diseases and chronic diseases, as well as reduced economic and healthcare costs associated with these diseases.

1. **What will be done during and / or after the project to increase the likelihood of achieving the identified benefit and reaching the identified beneficiaries?**

During the project, we will engage closely with collaborators to share interim findings and challenges. Findings will be disseminated at international conferences and in respectable peer-reviewed journals, as well as to the general public, policy makers, and international healthcare community through press conferences held at The University of Hong Kong, potentially reaching >50 international media sources; and media channels including newsletters, website and social media to raise public awareness on this topic.