Analysis of genetic interactions of medicines indicated for COVID-19

Gustavo Simões Carnivali
UFMG

Diego Simões Carnivali
UFJF
The viral disease COVID-19 first emerged in China in December 2019 and has already spread worldwide. Due to its high transmission rate and considerable number of deaths, this disease has since become a major topic of scientific studies, such as those that suggest the use of existing drugs to treat COVID-19. The aim of this study is to analyze and characterize the genetic interactions of drugs indicated by other studies for the treatment of this disease. Based on a gene co-expression network (GCN), we propose parameters that assess the connection between the drugs studied. These parameters allow researchers to identify drugs with similar functionality and to better understand the performance of these drugs when combined. Finally, this study presents two tables with the calculated measurements between all drugs, as well as previous analyses on the results found. This study contributes to increase the assertiveness in the prescription of more than one medicine for the treatment of COVID-19.

**Keywords:** COVID-19, Sars-Cov-2, Gene Co-Expression Network, Pharmacogenomics.
INTRODUCTION

At the end of December 2019, the outbreak of a disease caused by a newly discovered coronavirus (COVID-19, formerly known as 2019-nCoV, caused by the Sars-Cov-2 virus) was reported in Wuhan, China [1,2]. Currently, several countries around the globe have already been affected. In general, COVID-19 is a disease with symptomatic treatment, but it can become acute and deadly. A severe onset of the disease can lead to death due to progressive respiratory failure [2,4]. By March 12, 2020, the disease already had 125,048 confirmed cases worldwide, with an estimated mortality rate of approximately 3.7% [5].

Despite the high number of deaths worldwide, the disease caused by coronavirus does not yet have an efficient treatment [6]. This scenario stimulated the development of several studies on drug repurposing [7,8,9,10]. Drug repurposing consists of using an existing drug (already used for other diseases) to treat a new disease. This technique allows new treatments to be recommended in less time and with lower economic cost than necessary for the development of a new drug. [11]

Studies on drug repurposing usually suggest the use of several drugs in combination. Since these studies employ different methods, a significant number of drugs can be indicated to treat the same disease. Because of that, all studies on drug repurposing for COVID-19 require further research to clinically test the indicated drugs [7,8,9,10]. Therefore, the many studies that indicate that several drugs can be used to treat a given disease lead to more clinical trials and the cost of distribution increases.

This study aims to assist clinical trials by indicating which properties of certain drugs make them a priority for future tests. To this end, a prior genetic analysis of drugs indicated by other studies was conducted. More specifically, this study focused on the genetic similarities of several drugs in order to determine the efficacy of treatments that use more than one drug from the studied list. Based on a graph that determines relationships between genes expressed in the human body, named Gene Co-expression Network (GCN), this study theoretically determines the distance between genes linked to the main effect and side effect of the indicated drugs. This process identifies medicines with similar phenotype processes.

Thus, this study provides a summary and a brief analysis of current articles on drug repurposing for COVID-19. In addition, it presents tables with characteristic measurements of 26 medicines for consultation, use and future pharmaceutical applications for the treatment of COVID-19, as well as analyses of the results found.
MATERIALS AND METHODS

Gene Co-expression Network (GCN)

Gene expression can be interpreted as the process in which DNA nucleotide sequences are transcribed into either RNA or proteins, functional genetic products [12, 13]. The speed of gene expression may differ and be influenced by several factors [14].

A change in gene expression may increase or decrease the expression of other genes [15,16]. For example, an increase in the expression of a gene caused by an anomaly can generate an increase in the expression of another gene that was not initially associated with this anomaly, but is involved in its phenotypic effects [17]. Finding these connections allows the identification of unknown metabolic pathways about a disease or anomaly [18].

This correlation between genes is well represented by a graph [19]. Graphs or complex networks are mathematical structures widely used to represent relationships between elements. A graph $G = (V, E)$ is a non-empty set of vertices $V$ and a set of edges $E$ [20]. In this article, the set of vertices $V$ represents the genes that are expressed in the human body while the set of edges $E$ represents the relationships between these genes.

A gene co-expression network (GCN) is a graph representing the co-expression relationships between genes [19]. A GCN is commonly used to find a set of genes associated with an anomaly and to generate a conceptual detail about the phenotypic effects that each gene produces when in contact with a disease [21]. An example of a GCN with human genes is presented in Figure 1. The String platform offers a database with thousands of genes expressed in the human body that allows the creation of a GCN. String’s database was used in this study for the creation of the GCNs used.
Data

The GCN used in this study contains a set of genes whose expression is altered if one of the drugs selected is used. The DrugBank database used in this study offers a list of 103,080 interactions between drugs and several genes expressed in the human body [22]. This list shows which genetic metabolic pathways are altered by drugs.

The change in the expression of a human gene caused by a drug can indicate its main effect and side effect, such as indications for different treatments or side effects [22]. DrugBank and other similar databases were used to indicate medicines by some of the studies analyzed [7,10].

In the sequence, we present the 4 conclusive studies found, which suggest medicines to treat COVID-19. Only drugs approved by the FDA (Food and Drug Administration) and analyzed by DrugBank were considered for the applicability of the methodology.

Network-based drug repurposing for novel coronavirus 2019-nCoV / SARS-CoV-2 [7]: this study approached drug repurposing for COVID-19 by connecting a database of already known drugs with genes linked to Sars-CoV-2 through a gene co-expression network (GCN). The genes related to COVID-19 were obtained through analysis of similarities between COVID-19 and other diseases already studied. Based on this methodology, the study indicated with high reliability 16 FDA approved drugs: Sirolimus, Paroxetine, Mesalazine, Melatonin,
Mercaptopurine, Dactinomycin, Irbesartan, Carvedilol, Colchicine, Equilin, Oxymetholone, Emodin, Quinacrine, Toremifene, and Eplerenone.

**Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 Protease against COVID-19** [8]: This article analyzed the functional properties of Sars-CoV-2 along with previous studies on COVID-19. The study indicated and analyzed the functionality of a set of 3 drugs, previously used to treat other diseases, in the treatment of COVID-19. Based on known properties of the three drugs analyzed, the study determined if they are efficacy to treat COVID-19. Through this methodology, the article indicated with high reliability 3 drugs approved by FDA: Lopinavir, Oseltamivir, and Ritonavir.

**Teicoplanin: an alternative drug for the treatment of coronavirus COVID-19** [9]: this study analyzed applications of a set of 5 drugs suggested for the treatment of COVID-19 based on in vitro tests, and indicated the use of a new drug not yet tested for COVID-19. Based on previous applications of the drug, the authors believe that this medicine is efficient in the treatment of COVID-19. Through this methodology, the article indicated with high reliability 6 FDA-approved drugs: Chloroquine, Remdesivir, Lopinavir, Ribavirin, Ritonavir, and Teicoplanin.

**SARS-CoV-2 protein interaction map reveals targets for drug repurposing** [10]: based on mass spectrometry, this study identified a set of interactions between human proteins and Sars-CoV-2. Of the human proteins identified, 66 presented some regulatory drug associated with them. Among the drugs found to regulate proteins linked to Sars-CoV-2, the ones approved by the FDA are: Rapamycin, Ponatinib, Migalastat, Mycophenolic acid, Ribavirin, Valproic Acid, Apicidin, Chloroquine, Loratadine, Haloperidol, Indomethacin, Entacapone, Metformin, Dabrafenib, Minoxidil, Tacrolimus, Midostaurin, Ruxolitinib, Daunorubicin, S-verapamil, Nafamostat, Chloramphenicol, Linezolid, Tigecycline, Captopril, Lisinopril, and Camostat.

All drugs from the 4 studies mentioned above were added to the DrugBank database, and the ones that influenced any gene present in the database were analyzed. Considering only the drugs that present some genetic definition in DrugBank and removing repetitions of drugs from different studies, we analyzed a set of 26 drugs: Sirolimus, Paroxetine, Mesalazine, Melatonin, Mercaptopurine, Dactinomycin, Irbesartan, Carvedilol, Colchicine, Equilin, Oxymetholone, Emodin, Ritonavir, Chloroquine, Ribavirin, Mycophenolic Acid, Valproic Acid, Loratadine, Haloperidol, Indomethacin, Metformin, Tacrolimus, Ruxolitinib, Daunorubicin, Chloramphenicol, and Lisinopril.
Experimental Methodology

The methodology used to analyze and characterize the 26 drugs mentioned in the previous subsection consists of two measurements that assess the functional distance between drugs. The genes that had their expressions altered by each drug were stored. The genes found were used as part of the GCN constructed with the genes expressed in the human body. Finally, two distance measurements were calculated: the first assessed the number of vertices reached by another vertex in the graph, while the second complemented the first by assessing the distance between two genes influenced by a drug and contained in the GCN. The two measurements are presented below:

**First measure:** First measure: The reach measurement assesses whether there is a connection between two distinct genes (i.e., vertices) whose expressions were altered by a particular drug. We want to determine, with this measure, functional similarities between drugs. Thus, the reach between two drugs is given by how many genes directly affected by a particular drug connect to other genes directly affected by another drug. The reach of one drug in relation to another determines a functional approximation between them. For example, Figure 2 shows a direct path connecting genes G1 and G4, showing a possible strong similarity between medicine 1 and medicine 2. In the example, different paths connect genes G1, G2 and G3 to gene G4, showing high connectivity between medicines 1 and 2, but no path connecting the genes G1, G2, G3 and G4 to the gene G6 however there is a connection with the gene G5. The value of reach between medicine 1 in and medicine 2 is 3 (G1, G2, and G3 reach G4), the same value between medicine 1 and medicine 3 is also 3 (G1, G2, and G3 reach G5) and the reach between medicine 2 and medicine 3 is 1 (G4 reaches G5), showing a possibly greater similarity in the effects of medicines 1 and 2 than between medicines 1 and 2 with medicine 3. This measurement is commutative, that is, for any medicine, the reach between medicine 1 and medicine 2 is equal to the reach between medicine 2 and medicine 1.

**Second measure:** The distance measurement used in this study measured the number of intermediate genes in the shortest path between two genes in a connected graph. We want to determine with this measure functional similarities between medicines. Medicines with a low distance value possibly have significant functional similarities. In Figure 2, for example, G1 and G4 genes have distance 0 because there is a path that does not cross other vertices connecting them.

Vertices G1 and G5, on the other hand, have distance 2, because the shortest path that connects them crosses two other vertices (G9 and G10), the passage through other vertices can weaken the connection showing a low similarity of the effects of the genes. The distance between two drugs is given by the average of all the minimum distance measurements.
connecting the genes directly influenced by the drugs. In Figure 2, for example, the distance between medicine 1 and medicine 2 is given by: 

\[
(0 \times D(G_1, G_4) + 2 \times D(G_2, G_4) + 3 \times D(G_3, G_4)) / 3 = 1.666.
\]

This measurement is also commutative, that is, for any medicine, the distance between medicine 1 and medicine 2 is equal to the distance between medicine 2 and medicine 1.

**Figure 2.** Example of a GCN with genes whose expressions are altered by 3 drugs. The connections in red represent variations in expression obtained by DrugBank, and do not effectively represent connections present and considered in GCN.

We chose to present the measurements of reach and distance in graphs because they allow us to evaluate and differentiate the genes indicated for use in a disease. Genes with high reach value and low distance value are part of common metabolic pathways and possibly present similar phenotypic characteristics among several of the indicated medicines. Genes with such characteristics should not be combined, as they possibly participate in similar cellular and organic processes. Genes with non-significant values possibly affect different phenotype factors in a disease, thus their combined use can be indicated to modify the symptomatic fronts in the disease.

The GCN used to calculate the aforementioned measurements is calculated at its edges; that is, all edges of the network have a value between 0 and 1. This value represents a probability of interaction if A and B genes are connected in a GCN. The value of the edge that connects them represents the reliability that if the expression of one gene is altered, the expression of the connected gene is also altered. [20] To increase the reliability of the analyses and measurements, we only considered connections with values above 0.8: highly reliable according to String [20].
To increase reliability, we only considered paths smaller than 4 because the connections present in the GCN may not represent a total dependency between the genes. In Figure 2, the path that connects G3 to G5 through G13, G11, and G12 was considered because it is a size 3 path; however, if the only path connecting the vertices G3 to G5 was the one that crosses G13, G11, G9, G10, and G12, in this order, it would not be considered because it is a size 5 path.

**Results**

The results are presented in tables 1 and 2 (attached). The tables show the results of the calculations of reach (Table 1) and distance (Table 2), as well as the sum of each measurement for each gene. According to the measurements, the two tables are symmetrical and present the results for the 676 tuples of drugs studied.

Table 1 shows that several medicines have no connection with one another. This may be due to the low density of network connections or due to the properties of the studied drugs, which, despite being indicated for the treatment of the same disease, have different metabolic action pathways. However, some medicines have high connectivity with others, such as Colchicine, which has a high sum and a high connectivity with the other medicines studied.

The three groups of medicines with the highest reach values were: Colchicine and Indomethacin (Score 69), Colchicine and Melatonin (Score 46), and Colchicine and Emodin (Score 35). In the study [27], Melatonin and Colchicine are presented as possible targets for the treatment of the same disease. In the study [28] the similar effect of the drugs Colchicine and Indomethacin are proposed for the control of polychondritis relapsing. Despite being indicated to treat different anomalies, these medicines have anti-inflammatory action [23,24,25,26] and, therefore, should not be indicated combined for the treatment of COVID-19. Bringing also the reliability in the other results found.

Table 2 shows the result of the distance measurement for the 26 drugs studied. In this table, it is possible to notice a high number of zeros in the matrix accordingly to the previous results. Most genes presented a final average lower than 1, which means that most of them are reached by other genes in short paths, increasing the reliability of the results and maintaining the same conclusion made from the analysis of the reach measurements.

For consultation and better analyze a summary of the performance of the most scored drugs are presented: [22]

**Colchicine:** first approved by the FDA in 1961, colchicine is an alkaloid drug commonly used in the management of gout. Other than its use in gout, colchicine has been approved for managing exacerbations of Familial Mediterranean Fever (FMF).
**Indomethacin**: Indometacin, or indomethacin, is a non-steroidal anti-inflammatory drug (NSAID) with anti-inflammatory, analgesic, and antipyretic properties. The pharmacological effect of indomethacin is not fully understood, however, it is thought to be mediated through potent and nonselective inhibition of the enzyme cyclooxygenase (COX). Indometacin has been extensively studied in clinical trials as one of the most potent NSAIDs in blocking prostaglandin synthesis and was among the first NSAIDs to be used in the symptomatic treatment of migraine and for headaches.

**Melatonin**: melatonin regulates the sleep-wake cycle by chemically causing drowsiness and lowering the body temperature. Melatonin is also implicated in the regulation of mood, learning and memory, immune activity, dreaming, fertility and reproduction. Melatonin is also an effective antioxidant.

**Emodin**: this compound belongs to the class of organic compounds known as hydroxyanthraquinones. Emodin is an active component of several plants used in Traditional Chinese Medicine. It has various actions including laxative, antibacterial and antiinflammatory effects, Emodim was also investigated for the treatment of Polycystic Kidney.

In both tables, the main diagonal represents values managed by paths connecting two genes with expressions influenced by the same drug. This result can determine non-plural drugs in their phenotypic effects with diverse but not independent implications.

**Conclusion and Discussions**

This paper presented a genetic analysis of drugs indicated by other studies for the treatment of COVID-19. Based on a connectivity analysis, we characterized a set of 26 drugs indicated by 4 studies. A summary of the 4 papers was presented together with the medicines indicated by them, as well as the data that allowed the analyses of the medicines. Two analyses were conducted in order to determine the drugs that have similar functional implications.

The first analysis consisted of calculating the number of possible connections through the network between two sets of genes that have their expressions altered by two different drugs. This calculation was performed for all 676 possible tuples among the 26 drugs used. The result was presented in tables and may support future applications of these drugs. The results showed that many of the medicines do not connect, but that some drugs have significant similarities and should not be combined. For example, the four most scored drugs have anti-inflammatory actions, the use of more than one medication with the same action should be avoided as they can cause similar damage to the organism, the use therefore must be accompanied by competent professionals.
The second analysis complements the first by presenting a functional distance between connected drugs, allowing us to analyze the statistical reliability of the connections presented. This calculation was also conducted for all 676 possible tuples among the 26 drugs analyzed, presented in a table. The new results reinforce and bring greater reliability to the previous result. These tables can be used by health professionals to supervise possible treatments for Covid 19.

These results can be included in experimental studies that may validate the drugs.

Furthermore, an extended version of this study can be developed with the addition of other drugs that are yet to be indicated, or with the addition of other analyses on the set of available drugs.

ACKNOWLEDGMENTS

The authors would like to thank Capes, Fapemig and UFMG for the funding. In addition, the authors would like to thank the Academic Publishing Advisory Center (Centro de Assessoria de Publicação Acadêmica, CAPA – www.capa.ufpr.br) of the Federal University of Paraná (UFPR) for assistance with English language translation and editing.

REFERÊNCIAS


