

## *Analizando características da rede genética gerada por genes vinculados ao Covid-19*

Gustavo Simões Carnivali<sup>1</sup>

### RESUMO

Uma doença viral com alta mortalidade e índice de proliferação surgiu na China em dezembro de 2019 e já se espalhou por todo mundo (Covid 19). Devido a tal mortalidade, vários estudos analisando esse vírus estão sendo realizados. Entre esses estudos, um recente apresentou um conjunto de 94 genes expressos no corpo humano mas com influência sobre essa doença. Este trabalho pretende realizar uma análise sobre a conectividade dos genes encontrados pelo trabalho anterior. A partir de uma rede de expressão gênica com milhares de genes expressos no corpo humano, foi possível caracterizar os genes de interesse. Aplicando a metodologia proposta, foi encontrado um conjunto de genes, dentre os 94, que mais se conectam a outros genes expressos no corpo humano. Dentre esses genes, os 3 mais significativos tiveram suas propriedades analisadas, todos possuem efeitos reguladores de mecanismos celulares, essas características podem explicar vários efeitos secundários causados pelo Covid 19.

**Palavras-Chave:** Covid 19, Sars Cov 2, Gene Co-expression Network

### INTRODUÇÃO

Desde o final de dezembro de 2019, o surto de uma doença causada por um novo coronavírus (Covid 19, anteriormente conhecido como 2019-nCoV) foi relatado em Wuhan China [1,2], Atualmente vários países do globo já foram afetados. Em geral, o Covid 19 é uma doença com tratamento porém ela pode ser aguda e mortal. O início grave da doença pode resultar em morte devida a insuficiência respiratória progressiva [2,4]. Em 12 de março de 2020, a doença foi confirmada em 125.048 pessoas em todo o mundo, com uma mortalidade de aproximadamente 3,7% [5].

Apesar da sua alta mortalidade e ocorrência mundial, a doença causada pelo coronavírus ainda não possui um tratamento eficiente conhecido [6]. Esse cenário gera uma necessidade de conhecer essa doença a fim de gerar metodologias de tratamento para ela. Esse trabalho objetiva desenvolver uma análise genética dos efeitos do coronavírus sobre o conjunto de genes humanos. Será explicitado informações

genéticas entre o parasito e o hospedeiro a fim de permitir que tais informações, além de trazerem maior compreensão sobre a doença, permitam também o apoio e desenvolvimento de propostas de tratamento para o Covid 19.

O trabalho [7], a partir de análises diversas e do conjunto de genes comuns ao vírus estudado, apresentou 94 genes expressos no corpo humano que possuem conexões com o vírus causador do Covid 19. Especificamente, este trabalho objetiva analisar fatores genéticos desses 94 genes. Para isso, será desenvolvida uma rede de co-expressão gênica de vários genes expressos no corpo humano (entre eles os 94 que se conectam ao coronavírus). Após a criação da rede, os 94 genes de interesse serão analisados quanto às suas conexões com outros genes do corpo humano. O objetivo com isso é encontrar vias de conexões genéticas a fim de conhecer mais a doença ou determinar, entre os 94 genes, os que mais se relacionam e, com isso,

<sup>1</sup>Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil. Correspondência para: GS CARNIVALI. E-mail: <gustavocarnivali@gmail.com>.

determinar os que são mais propensos a causarem efeitos fenotípicos típicos da doença ou de serem inibidos por tratamentos diversos.

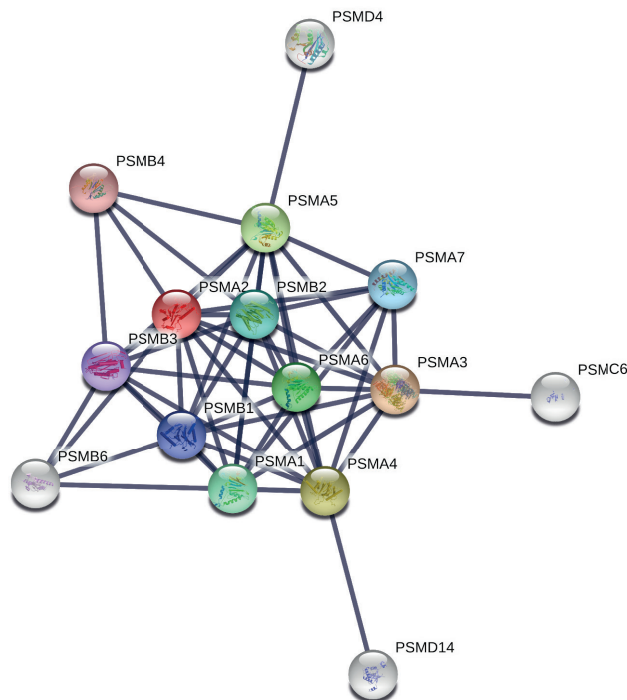
## METODOLOGIA

Expressão pode ser interpretada como o processo em que sequências de nucleotídeos de DNA são transcritas em RNA ou proteínas, em um produto genético funcional [8, 9]. A velocidade em que os genes são expressos pode ser diferente e pode ser influenciada por vários fatores [10].

A alteração da expressão de um gene pode aumentar ou diminuir a expressão de outros genes [11,12]. Por exemplo, o aumento da expressão de um gene, causado por uma anomalia, pode gerar um

aumento da expressão de outro gene que inicialmente não era associado a esta anomalia, mas que participa de seus efeitos fenotípicos [13]. Encontrar essas relações permite descobrir caminhos metabólicos desconhecidos sobre uma doença ou anomalia [14].

A correlação entre genes é bem representada por um grafo [15]. Grafos ou redes complexas são estruturas matemáticas muito usadas para representar relacionamentos entre instâncias. Um grafo  $G = (V, E)$  é uma estrutura formada por um conjunto não vazio de vértices  $V$  e um conjunto de arestas  $E$  [16]. Neste trabalho, o conjunto de vértices  $V$  representa os genes que participam dos efeitos fenotípicos de uma doença e o conjunto de arestas  $E$  as relações existentes entre esses genes.



**Figura 1.** Exemplo de GCN com genes humanos.

Uma rede de co-expressão de genes, GCN, é um grafo que representa as relações de co-expressão entre genes [15]. Uma GCN é comumente usada para encontrar um conjunto de genes associados a uma anomalia e gerar um detalhamento conceitual sobre os efeitos fenotípicos que cada gene gera em uma doença [17]. Um exemplo visual de GCN com genes humanos pode ser visto na Figura 1.

Um interactoma foi obtido com interações entre

SARS-CoV-2 e proteínas humanas [18]. Em resumo, o interactoma contém interações proteína-proteína do repositório NCBI Virus [19] vírus SARS-CoV-2 (<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>) com proteínas humanas conhecidas. O interactoma possui um total de 125 proteínas (31 proteínas virais e 94 proteínas humanas).

A ferramenta String oferece um banco de dados com milhares de genes expressos no corpo

humano (entre eles os 94 genes vinculados ao Covid 19 encontrados no estudo [18]) que permite a criação de uma GCN. A GCN desenvolvida será utilizada para determinar propriedades fundamentais dos 94 genes de interesse.

### **Análise do grau dos genes vinculados ao Covid 19**

A análise que será desenvolvida neste trabalho verifica o número de conexões totais que um gene possui em uma rede (i.e., grau do vértice) [16]. A análise do grau de um vértice determina a conectividade de vértices em uma rede, em uma GCN, genes com alto grau possuem alta relevância na expressão de outros genes. Caso tenham sua expressão alterada, genes com alto grau causarão alterações em cascata em vários outros genes, tendo várias implicações fenotípicas secundárias causadas por essas alterações.

A Rede de co-expressão gênica utilizada é valorada em suas arestas, isto é, cada aresta possui um valor entre 0 e 1. Este valor determina o nível de confiabilidade estatístico da existência da conexão, isso é, arestas com valores próximos de 1 determinam que se a expressão de um gene for alterada existe uma alta probabilidade da expressão do gene conectado a ele também ter sua expressão alterada. Neste trabalho, este valor será chamado de Score.

Serão calculados 3 níveis de grau dos genes vinculados ao Covid 19. O primeiro cálculo será o somatório dos Scores de todas as arestas conectadas ao gene. Genes conectados a outros genes com alto grau também possuem uma alta probabilidade de influência, considerando isso, os níveis 2 e 3 da medida irão considerar o somatório dos Scores de todas as arestas conectadas a um gene mais o Score de todas as arestas conectadas aos seus vizinhos de distância 3 e 5. A distância representa o número de vértices que conectam outros dois vértices no menor caminho possível. Por exemplo, na Figura 1 o gene PSMC6 possui distância 1 do gene PSMA4 e distância 2 do gene PSMD14.

Os 5 genes com maior pontuação são listados abaixo, ordenados de forma crescente para os 3 níveis de medida:

Nível 1: PSMA2, EEF1A1, POLR2B, ENO1, CHEK2

Nível 2: EEF1A1, PSMA2, POLR2B, ENO1, SERPING1

Nível 3: EEF1A1, PSMA2, POLR2B, RPS17, ENO1

Um resumo das propriedades dos genes com maior

pontuação é apresentado abaixo: PSMA2 - Componente

**PSMA2** - Componente do complexo do proteossoma do núcleo 20S envolvido na degradação proteolítica da maioria das proteínas intracelulares. Este complexo desempenha numerosos papéis essenciais dentro da célula, associando-se às diferentes partículas reguladoras. [21]

**EEF1A1** - Este gene está associado a uma proteína que promove a ligação ribossomos durante a biossíntese de proteínas. Com a participação de outros genes, forma um complexo que age como um fator de transcrição específico de célula Th1 e liga o promotor de IFN-gamma para regular diretamente sua transcrição e, portanto, está envolvido de maneira importante na produção de citocinas Th1. [21]

**POLR2B** - Segundo maior componente da RNA polimerase II, que sintetiza precursores de mRNA e muitos RNAs não codificadores funcionais. Propõe-se contribuir para a atividade catalítica da polimerase. [21]

A partir desse resultado, podemos perceber que, irrelevante a forma de calcular o grau dos vértices, houve pouca mudança entre os genes com maior pontuação nas 3 formas de cálculo.

Analisando propriedades dos genes que tiveram maior pontuação nas medidas podemos perceber que os 3 genes considerados possuem um grande papel regulador em diversos processos celulares, isso era esperado, genes com alto nível de conexão em uma GCN necessariamente estão conectados a vários outros genes e conseqüentemente a várias funcionalidades dentro da célula. Assim, esses genes conectam vários outros aos genes que constituem o Covid 19, possuindo necessariamente uma significativa importância dentro da doença e de seus fatores fenotípicos ainda desconhecidos.

Tal resultado ainda pode explicar a alta mortalidade e o alto número de sintomas leves causados pelo Covid 19 [22]. Como o vírus altera a expressão de genes reguladores e de alto grau médio isso implica que sua atuação no corpo humano pode ser diversa e atuante sobre várias vias metabólicas distintas.

### **CONCLUSÃO**

Esse trabalho apresentou uma análise sobre um conjunto de genes humanos vinculados a uma doença

viral. A doença teve início em dezembro de 2019 na China e em poucos meses se espalhou por vários países. Por ser uma doença recente, pouco é conhecido sobre ela, porém, sua alta mortalidade e poder de expansão estimulam o contínuo avanço nos estudos sobre essa doença. [22]. Como o vírus altera a expressão de genes reguladores e de alto grau médio isso implica que sua atuação no corpo humano pode ser diversa e atuante sobre várias vias metabólicas distintas.

A partir de redes de expressão gênica, com o objetivo de analisar os genes vinculados a doença estudada, esse trabalho analisou a conectividade de um conjunto de 94 genes humanos vinculados ao Covid 19 com outro conjunto de milhares de genes humanos não relacionados ao Covid 19. Foram analisados os graus de relacionamento desses genes, com o objetivo de encontrar os genes que mais se conectam e influencia outros genes no corpo humano.

As características de um conjunto limitado de 3 genes com alto grau foi apresentado. Analisando esses genes foi visto que todos participam de fatores reguladores da célula, podendo explicar um conjunto vasto de sintomas causados por essa doença. Esse resultado prévio permite a maior compreensão sobre fatores desconhecidos da doença, ainda que novos alvos terapêuticos sejam indicados para o tratamento dessa doença.

Como trabalhos futuros sugere-se a análise de outras propriedades dessa rede e uma análise mais profunda dos componentes de maior pontuação encontrados. Ainda pode-se considerar a utilização de outros sintetizadores de redes além do String e uma análise de outros genes além dos 94 utilizados.

Os códigos e dados utilizados nos experimentos assim como outras informações do texto estão disponíveis no link: <https://www.gscarnivali.com/publications/analizando-caracteristicas-da-rede-genetica-gerada-por-genes-vinculados-ao-covid-19/>

## AGRADECIMENTOS

Capes, Fapemig e UFMG pelo financiamento.

## REFERÊNCIAS

1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 579.7798 (2020): 265-269.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
3. Sohrabi, Catrin, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery* (2020).
4. Chan, Jasper Fuk-Woo, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 395.10223 (2020): 514-523.
5. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 72. (2020).
6. Zhou, Yadi, et al. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell discovery* 6.1 (2020): 1-18.
7. Srinivasan, Suhas, et al. Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. *Viruses* 12.4 (2020): 360.
8. Shimkets, Richard A., ed. *Gene expression profiling*. Vol. 258. Springer Science & Business Media, 2004.
9. Parmigiani, Giovanni, et al. *The analysis of gene expression data: an overview of methods and software*. The analysis of gene expression data. Springer, New York, NY, 2003. 1-45.
10. Lee, Cheol-Koo, et al. Gene expression profile of aging and its retardation by caloric restriction. *Science* 285.5432 (1999): 1390-1393.
11. van Noort, Vera, Berend Snel, and Martijn A. Huynen. Predicting gene function by conserved co-expression. *TRENDS in Genetics* 19.5 (2003): 238-242.
12. Zhu, Qun, et al. "Enhanced protection against fungal attack by constitutive co-expression of chitinase and glucanase genes in transgenic tobacco. *Bio/technology* 12.8 (1994): 807-812.
13. Emilsson, Valur, et al. Genetics of gene expression and its effect on disease. *Nature* 452.7186 (2008): 423-428.
14. Smeets, Cleo JLM, and D. S. Verbeek. *Cerebellar*

ataxia and functional genomics: identifying the routes to cerebellar neurodegeneration. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1842.10 (2014): 2030-2038.

15. Stuart, Joshua M., et al. A gene-coexpression network for global discovery of conserved genetic modules. *science* 302.5643 (2003): 249-255.

16. Paulo Feofiloff, Yoshiharu Kohayakawa, and Yoshiko Wakabayashi. Uma introdução sucinta à teoria dos grafos. 2011.

17. Smeets, Cleo JLM, and D. S. Verbeek. Cerebellar ataxia and functional genomics: identifying the routes to cerebellar neurodegeneration. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1842.10 (2014): 2030-2038.

18. Srinivasan, Suhas, et al. Structural Genomics of SARS-CoV-2 Indicates Evolutionary Conserved Functional Regions of Viral Proteins. *Viruses* 12.4 (2020): 360.

19. Wheeler, David L., et al. Database resources of the national center for biotechnology information. *Nucleic acids research* 36.suppl\_1 (2007): D13-D21.

20. Snel, Berend, et al. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic acids research* 28.18 (2000): 3442-3444.

21. UniProt Consortium. The universal protein resource (UniProt). *Nucleic acids research* 36.suppl\_1 (2007): D190-D195.

22. Surveillances, Vital. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly* 2.8 (2020): 113-122.