



Identification of Causal Genes of COVID-19 Using the SMR Method

Yan Zong and Xiaofei Li*

Department of Infectious Diseases, Yiwu Central Hospital, Jinhua, China

Since the first report of COVID-19 in December 2019, more than 100 million people have been infected with SARS-CoV-2. Despite ongoing research, there is still limited knowledge about the genetic causes of COVID-19. To resolve this problem, we applied the SMR method to analyze the genes involved in COVID-19 pathogenesis by the integration of multiple omics data. Here, we assessed the SNPs associated with COVID-19 risk from the GWAS data of Spanish and Italian patients and lung eQTL data from the GTEx project. Then, GWAS and eQTL data were integrated by summary-data-based (SMR) methods using SNPs as instrumental variables (IVs). As a result, six protein-coding and five non-protein-coding genes regulated by nine SNPs were identified as significant risk factors for COVID-19. Functional analysis of these genes showed that UQCRH participates in cardiac muscle contraction, PPA2 is closely related to sudden cardiac failure (SCD), and OGT, as the interacting gene partner of PANO1, is associated with neurological disease. Observational studies show that myocardial damage, SCD, and neurological disease often occur in COVID-19 patients. Thus, our findings provide a potential molecular mechanism for understanding the complications of COVID-19.

Keywords: SMR, COVID-19, eQTL, GWAS, UQCRH, PPA2, OGT, PANO1

OPEN ACCESS

Edited by:

Liang Cheng,
Harbin Medical University, China

Reviewed by:

Juan Wang,
Inner Mongolia University, China
Sheng Li,
Wuhan University, China

*Correspondence:

Xiaofei Li
xiaofeil2000@163.com

Specialty section:

This article was submitted to
Statistical Genetics and Methodology,
a section of the journal
Frontiers in Genetics

Received: 02 April 2021

Accepted: 07 May 2021

Published: 05 July 2021

Citation:

Zong Y and Li X (2021)
Identification of Causal Genes
of COVID-19 Using the SMR Method.
Front. Genet. 12:690349.
doi: 10.3389/fgene.2021.690349

INTRODUCTION

In December 2019, SARS-CoV-2 was first reported to lead to the respiratory disease coronavirus disease 2019 (COVID-19) (Zhou et al., 2020). Subsequently, COVID-19 quickly spread to all parts of the world and became a worldwide public health event. As of February 17, 2021, more than 100 million people had been infected, and more than 2.4 million people had died of COVID-19. At present, a total of seven types of coronaviruses that can infect humans have been discovered, including SARS-CoV, SARS-CoV-2, and MERS-CoV, which have high case fatality rates (CFRs) (Gussow et al., 2020). The other four coronaviruses, HCoV-HKU1, HCoV-NL63, HCoV-OC43, and HCoV-229E, only cause mild symptoms in humans. Although the CFR of SARS-CoV-2 is relatively lower than those of SARS-CoV and MERS-CoV, it is still highly infectious.

Exploring the origin of the virus would help to increase the understanding of SARS-CoV-2 (Cui et al., 2019; Narang et al., 2019; Benetti et al., 2020; Meng et al., 2020; Wan et al., 2020; Qi et al., 2021). Previous studies have shown that bats are the natural host of the evolved coronavirus (Jiang et al., 2020; Kwon et al., 2020). Based on the alignment of the reference genome sequence, a phylogenetic tree was constructed, indicating that the genes are very similar between SARS-CoV-2 and members of the bat Sarbecovirus subgenus Betacoronavirus (Wu et al., 2020). According to sequence mapping, the whole-genome sequence of SARS-CoV-2 has the highest similarity with SARS-CoV BatCoV RaTG13, reaching over 96%. The similarity between SARS-CoV-2 and the

coronaviruses SARS-CoV and MERS-CoV is only 79 and 50%, respectively (Andersen et al., 2020). Although the specific route of transmission of SARS-CoV-2 from its natural host to humans is not yet clear, researchers have discovered that the key functional sites of the SARS-CoV-2 spike protein are almost the same as those of the virus isolated from pangolins. Therefore, a pangolin coronavirus may have provided part of the spike gene for SARS-CoV-2.

Computational methods with omics data have shown strong power in identifying disease-related genes (Zhao T. et al., 2020c). SARS-CoV-2 is a single-stranded RNA virus (Chen et al., 2020). The reference genome shows that it has 29,903 nucleic acid base pairs, including seven conserved unstructured protein domains and four structural protein domains, including the spike protein. The sequence of SARS-CoV-2 has mutated since its discovery. As early as March 2020, researchers analyzed 160 early virus strains and found three important single-point mutations, T8782C, C28144T, and G26144T (Forster et al., 2020). Among them, the mutations T8782C and C28144T are used to distinguish between type A and B viruses, and the mutation G26144T represents a newer virus type (type C). In May 2020, Cheng et al. analyzed more than 1,800 strains of viruses in the Americas, Europe, and Asia and verified that type B viruses are more infectious and have almost replaced the type A viruses in current circulation (Cheng et al., 2021). Through comparison with the reference genome, it was found that each virus mutated at approximately 1.75 sites per month. The overall mutations were silent mutations and would not cause major functional changes in the virus. Cheng et al. proposed to analyze the mutation rules of the virus over time and identified seven dominant mutations. According to functional bioinformatics analysis, these mutations likely caused the virus to decrease in toxicity and increase in infectivity. In addition, mutation clustering information (Zhu X. et al., 2019; Qi et al., 2020; Zhao T. et al., 2020b; Zhao X. et al., 2020; Zou et al., 2020; Zhao et al., 2021) indicated that the virus circulating in the Americas is more similar to RaTG13.

By searching for the origin of SARS-CoV-2, researchers also found that SARS-CoV-2 and SARS-CoV use the same receptor, angiotensin-converting enzyme II (ACE2). SARS-CoV-2 invades human cells by binding to ACE2. By interacting with proteins in human cells, SARS-CoV-2 affects human health. In August 2020, researchers found hundreds of interactions between human proteins and SARS-CoV-2 proteins. Although many proteins that interact with SARS-CoV-2 have been discovered so far, the genes associated with SARS-CoV-2 pathogenesis are still unknown. Since genomics data of many COVID-19 patients have been reported, there is an opportunity to find potential risk genes through analysis and integration of omics data of susceptible populations (Cheng, 2019; Cheng et al., 2019; Li F. et al., 2020; Wang et al., 2020).

SMR is a method used to determine the causal association between genetically determined traits and diseases, and eQTLs are genetic variations related to the expression of traits. Since eQTL data are tissue specific, it is possible to correlate the eQTL data of disease-related tissues with disease GWAS data and use the SMR method to find causal genes for diseases. Therefore, in our study, we used lung eQTL data of GTEx and GWAS data of

COVID-19 patients to identify the genes related to COVID-19 pathogenesis using the SMR method.

MATERIALS AND METHODS

Risk SNPs for Severe COVID-19

According to current knowledge, some people are more susceptible than others to COVID-19. To assess the impact of

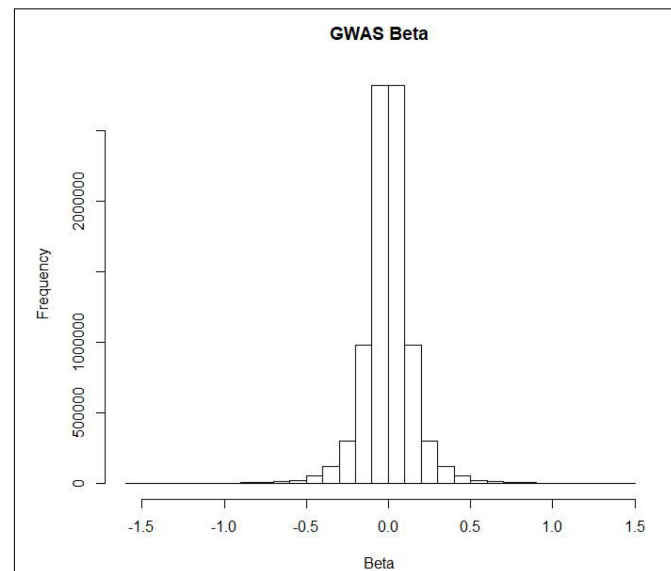


FIGURE 1 | The distribution of GWAS beta in severe COVID-19 patients.

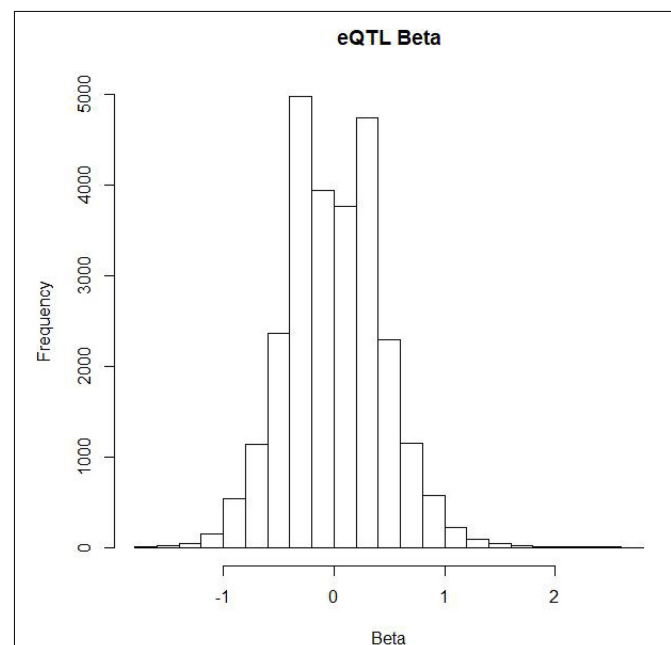


FIGURE 2 | The distribution of eQTL beta on lung SNPs.

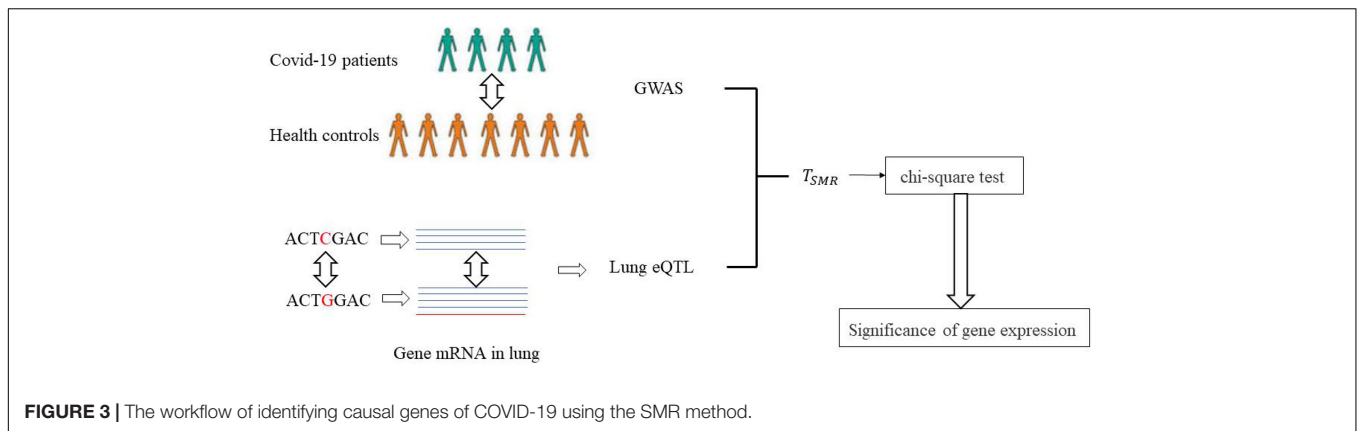


FIGURE 3 | The workflow of identifying causal genes of COVID-19 using the SMR method.

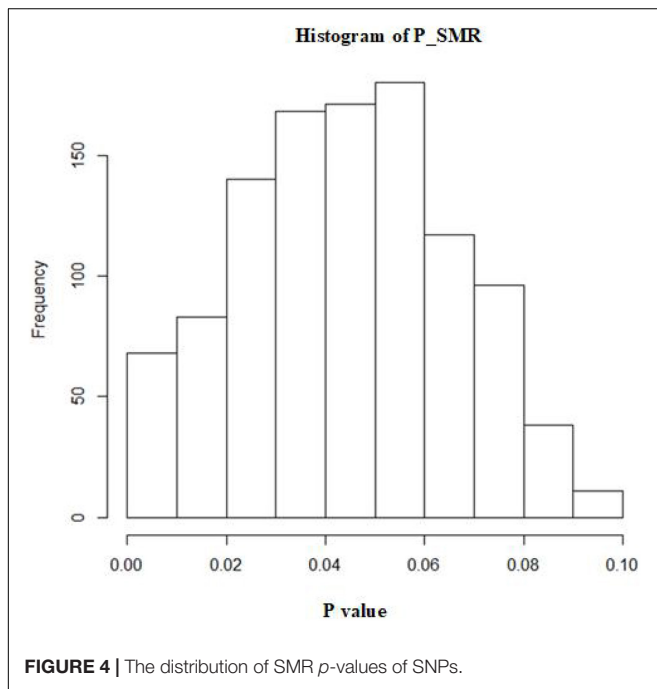


FIGURE 4 | The distribution of SMR p -values of SNPs.

SNPs on the risk of COVID-19, researchers conducted a genome-wide association study (GWAS) on two groups of European patients in seven hospitals in Italy and Spain and performed a meta-analysis of the results of the two groups (Ellinghaus et al., 2020). First, quality control was conducted on 1980 severe COVID-19 patients, and 1,610 patients remained after removing population outliers. Then, the researchers performed a GWAS on 835 Italian patients and 1,255 control group members (Sun et al., 2019), as well as 775 Spanish patients and 950 control group members. As a result, they obtained severe COVID-19 risk data for 8,582,968 SNPs. Finally, the two sets of experimental results were meta-analyzed to obtain the final risk SNPs. **Figure 1** shows the distribution of risk betas of SNPs in COVID-19 patients. Most SNPs had no impact on the risk of COVID-19. In addition, researchers found that COVID-19 high-risk SNPs clustered on 3p21.31.

Lung eQTL

To date, many GWAS have been performed. These studies have identified thousands of disease-related risk SNPs. Since more than 90% of SNP sites exist outside of protein-coding genes, it is difficult to understand the mechanisms by which these SNPs affect diseases. To this end, researchers investigate expressed quantitative trait loci (eQTLs) to reveal the genes regulated by SNPs in the blood, lung and other tissues (Zhao T. et al., 2020a). Therefore, the NIH launched the gene type-tissue expression (GTEx) project, with the goal of establishing the relationships between SNPs and gene expression in different tissues. Currently, the project has accepted more than 900 post-mortem donors. Sequencing of different tissues in the donors has identified a large number of SNP-regulated genes. Summarized GTEx data could be obtained from the project's website. **Figure 2** shows the distribution of beta values of lung SNPs on gene expression.

Identification of Potential Causal Genes of COVID-19 Based on the SMR Method

In the domain of biomedicine, many causal disease associations have been discovered through observational research, such as the association between smoking and lung cancer (Ghosh and Yan, 2020; Li J. et al., 2020; Yuan et al., 2020). However, the associations between phenotypes and diseases found in these observational studies cannot reflect causality. However, because observational studies are usually disturbed by external factors and often face practical problems related to long time frames and high costs, there are large errors in the analysis of pathogenic factors of diseases. Mendelian randomization follows the Mendelian inheritance law of allele separation and free recombination of nonalleles and makes causal inferences based on genetic variation, which does not change with environment or age, so this method is widely used in causal inference of pathogenic factors. Sometimes, due to the influence of confounding factors, the correlation found is not accurate. This greatly limits the development of the field. To solve this problem, the statistician Katan (1986) introduced the concept of Mendelian randomization (MR) in 1986 to study whether low serum cholesterol levels can increase the risk of cancer. MR uses genotype as an instrumental variable (IV) and applies the two-stage least squares method to infer the pathogenicity of diseases.

With the gradual deepening of GWAS research, this method has been widely applied using SNPs as IVs. This method assumes that Z is an instrumental variable (SNP), X represents exposure factors or gene expression levels, and Y represents the disease. According to the two-stage least squares method, the effect of X on Y is evaluated as follows.

$$Beta_{XY} = Beta_{ZX} / Beta_{ZY} \tag{1}$$

Here, $Beta_{ZX}$ represents the least-squares estimate of Z on X, and $Beta_{ZY}$ is the least-squares estimate of Z on Y. Then, we estimate the significance of X on Y as T_{MR} .

$$T_{MR} = (Beta_{XY})^2 / var(Beta_{XY}) \tag{2}$$

Theoretically, GWAS and eQTL need to target the same sample, but GWAS and eQTL are performed independently, so we use the two-sample MR method, and T_{SMR} is obtained as the final evaluation value (Zhao S. et al., 2019; Zhao T. et al., 2019):

$$T_{SMR} = Z_{ZY}^{2*} Z_{ZX}^2 / (Z_{ZY}^2 Z_{ZX}^2) \tag{3}$$

where Z_{zy} represents the Z statistics from the GWAS, and Z_{zx} represents the z statistics from the eQTL study. Since the true

values of $Beta_{ZX}$ and $Beta_{ZY}$ cannot be obtained, we can use estimations to replace them. The T_{SMR} yields an approximate Chi-square test statistic.

Here, we used the SMR method to evaluate the mechanism by which SNPs affect COVID-19 and identify potential pathogenic genes in the lungs. The specific process is shown in **Figure 3**. The GWAS data of COVID-19 patients and the lung eQTL data were obtained from a public dataset. T_{SMR} was calculated based on the two-sample MR method, which was then further evaluated using the chi-square test to identify significant SNPs and genes.

RESULTS

Causal SNPs and Genes of COVID-19

A total of 1,072 SNPs appeared in both GWAS and eQTL data. After application of the SMR method, 1,072 SNPs were determined to regulate the P value distribution of COVID-19 through genes (**Figure 4**). The P values of SMR for most SNPs were concentrated in the range of 0.02–0.08. Here, the threshold was set as 0.003, and then 11 genes (UQCRH,

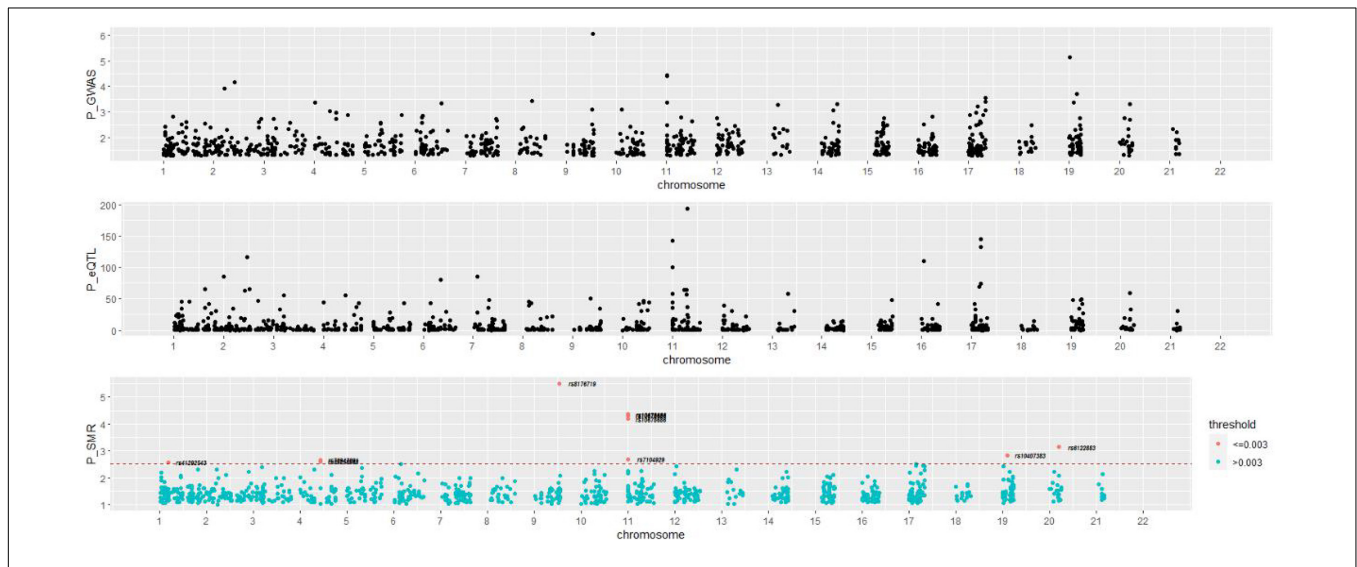


FIGURE 5 | The experimental results based on SMR.

TABLE 1 | Nine Causal SNPs and eleven pathogenic genes of COVID-19.

| SNP | P_SMR | Gene | chr | pos | P_GWAS | P_eQTL |
|------------|----------|---------------|-----|-----------|------------|-------------|
| rs41292543 | 0.002574 | UQCRH | 1 | 46309111 | 0.001553 | 2.65E – 16 |
| rs35258888 | 0.002419 | PPA2 | 4 | 105355205 | 0.001035 | 2.61E – 10 |
| rs70947091 | 0.002154 | PAPSS1 | 4 | 107694523 | 0.001899 | 2.28E – 56 |
| rs8176719 | 3.07E-06 | ABO | 9 | 133257521 | 8.76E – 07 | 1.27E – 34 |
| rs10678686 | 4.68E-05 | AP006621.5 | 11 | 780321 | 3.96E – 05 | 1.28E – 100 |
| rs10678686 | 4.34E-05 | CMB9-55F22.1 | 11 | 780321 | 3.96E – 05 | 9.62E – 143 |
| rs10678686 | 6.40E-05 | AP006621.6 | 11 | 780321 | 3.96E – 05 | 1.12E – 44 |
| rs7104929 | 0.001973 | PANO1 | 11 | 784340 | 3.73E – 05 | 0.0111629 |
| rs10407383 | 0.001423 | CTD-2027119.2 | 19 | 24134099 | 0.000412 | 4.61E – 09 |
| rs6122883 | 0.000702 | LINC01273 | 20 | 50172836 | 0.000494 | 3.43E – 34 |
| rs6151429 | 0.000645 | ARSA | 22 | 50625049 | 0.000522 | 4.70E – 50 |

PPA2, PAPSS1, ABO, AP006621.5, CMB9-55F22.1, AP006621.6, PANO1, CTD-2027I19.2, LINC01273, and ARSA) regulated by nine SNPs (rs41292543, rs35258888, rs70947091, rs8176719, rs10678686, rs10678686, rs10678686, rs7104929, rs10407383, rs6122883, and rs6151429) with P values lower than the threshold were considered to increase the risk of COVID-19 (Table 1). Figure 5 shows the GWAS P value, eQTL P value, and SMR P-value of all SNPs.

Functional Analysis of Causal Genes

Among the identified pathogenic genes, there were a total of six protein-coding genes (UQCRH, PPA2, PAPSS1, ABO, PANO1, and ARSA) and five noncoding genes. We then performed functional analysis of these six protein-coding genes to identify their functions, related diseases and pathways.

UQCRH participates in cardiac muscle contraction. A large number of studies have found that approximately 8–12% of COVID-19 patients have myocardial damage (Lippi et al., 2020). Although heart problems are usually not the most prominent or deadly feature of COVID-19, they are common and are severe enough that most people admitted to the hospital for COVID-19 are now being screened for myocardial damage. There are many potential causes of COVID-19-related myocardial damage, but it is often difficult to determine the specific cause in a specific individual. The UQCRH gene found here may be a potential cause. In addition, according to Gene Ontology annotation, UQCRH has ubiquinol-cytochrome-c reductase activity.

PPA2 is closely related to sudden cardiac failure (SCD). At present, there have been reports of COVID-19 patients who have died of SCD. In July 2020, Samira et al. diagnosed three patients with COVID-19 according to the reverse transcriptase-polymerase chain reaction of nasopharyngeal swabs and radiological examinations (Shirazi et al., 2021) who eventually died of SCD. Thus, the authors recommended that it is necessary to monitor the heart conditions of COVID-19 patients. Although there is no direct causal link between SCD and COVID-19, analysis of current data shows that there is a reasonable link between them. According to the latest studies, the incidence of SCD in the community and hospital environment has increased since the outbreak of COVID-19 (Yadav et al., 2020). Based on our findings, PPA2 can increase the risk of COVID-19, so PPA2 may be a potential factor that results in SCD in COVID-19 patients.

Interaction Between the Causal Gene PANO1 and OGT

We searched for the interactions between causal genes of COVID-19 and other genes. As a result, we found that OGT can interact with PANO1. Then, we further investigated the function of OGT to explore the potential mechanism of PANO1 in the risk of COVID-19. At the start of the COVID-19 epidemic, some patients experienced neurological symptoms, such as feeling confused, being unable to discern direction, and

feeling restless (Marshall, 2020). A total of 0.2% of the patients of two other SARS-CoV-2-related coronaviruses, SARS-CoV and MERS-CoV, have neurological disease. Given the number of COVID-19 patients, hundreds of thousands of patients may have neurological complications. As genes related to neurological disease and risk genes for COVID-19, OGT, and PANO1 must be considered further.

CONCLUSION

In this article, we used the SMR method to analyze the genes involved in COVID-19 pathogenesis. Here, the risk SNPs for COVID-19 were derived from the GWAS data of Spanish and Italian patients. Lung eQTL data were acquired from the GTEx project. In the postgenomic era, MR and SMR methods have been widely used (Liu et al., 2018). Currently, a large number of pathogenic phenotypes and genes have been identified based on these methods. Through SMR, this article discovered six protein-coding genes and five noncoding genes that can increase the risk of COVID-19. Finally, nine SNPs that met the threshold conditions were identified, and the SMR method was used to determine that these SNPs regulated 11 disease-causing genes that could increase the risk of COVID-19. Then, disease pathway enrichment analysis was performed on these genes.

Through functional analysis, we found that UQCRH participates in cardiac muscle contraction, PPA2 is closely related to SCD, and myocardial damage and SCD occurred in patients with COVID-19. Therefore, our findings provide a potential molecular mechanism for these processes. Further analysis revealed an interaction between OGT and PANO1. OGT is associated with neurological disease. This may explain the neurological complications in COVID-19 patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YZ and XL wrote the manuscript and did the experiments. XL provided ideas of this work. YZ analyzed the data. Both authors approved the submitted version.

REFERENCES

- Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nat. Med.* 26, 450–452. doi: 10.1038/s41591-020-0820-9
- Benetti, E., Tita, R., Spiga, O., Cioffi, A., Birolo, G., Bruselles, A., et al. (2020). ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *Eur. J. Hum. Genet.* 28, 1602–1614. doi: 10.1038/s41431-020-0691-z
- Chen, W., Feng, P., Liu, K., Wu, M., and Lin, H. (2020). Computational identification of small interfering RNA targets in SARS-CoV-2. *Viol. Sin.* 35, 359–361. doi: 10.1007/s12250-020-00221-6
- Cheng, L. (2019). Computational and biological methods for gene therapy. *Curr. Gene Ther.* 19, 210–210. doi: 10.2174/156652321904191022113307
- Cheng, L., Han, X., Zhu, Z., Qi, C., Wang, P., and Zhang, X. (2021). Functional alterations caused by mutations reflect evolutionary trends of SARS-CoV-2. *Brief. Bioinform.* 22, 1442–1450. doi: 10.1093/bib/bbab042
- Cheng, L., Zhao, H., Wang, P., Zhou, W., Luo, M., Li, T., et al. (2019). Computational methods for identifying similar diseases. *Mol. Ther. Nucleic Acids* 18, 590–604. doi: 10.1016/j.omtn.2019.09.019
- Cui, J., Li, F., and Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17, 181–192. doi: 10.1038/s41579-018-0118-9
- Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* 383, 1522–1534. doi: 10.1056/nejmoa2020283
- Forster, P., Forster, L., Renfrew, C., and Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9241–9243. doi: 10.1073/pnas.2004999117
- Ghosh, A., and Yan, H. (2020). Stability analysis at key positions of EGFR related to non-small cell lung cancer. *Curr. Bioinform.* 15, 260–267. doi: 10.2174/15748936146661912121212026
- Gussow, A. B., Auslander, N., Faure, G., Wolf, Y. I., Zhang, F., and Koonin, E. V. (2020). Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proc. Natl. Acad. Sci. U.S.A.* 117, 15193–15199. doi: 10.1073/pnas.2008176117
- Jiang, S., Du, L., and Shi, Z. (2020). An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. *Emerg. Microbes Infect.* 9, 275–277. doi: 10.1080/22221751.2020.1723441
- Katan, M. B. (1986). Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* 1, 507–508. doi: 10.1016/s0140-6736(86)92972-7
- Kwon, E., Cho, M., Kim, H., and Son, H. S. (2020). A study on host tropism determinants of influenza virus using machine learning. *Curr. Bioinform.* 15, 121–134. doi: 10.2174/1574893614666191104160927
- Li, F., Luo, M., Zhou, W., Li, J., Jin, X., Xu, Z., et al. (2020). Single cell RNA and immune repertoire profiling of COVID-19 patients reveal novel neutralizing antibody. *Protein Cell* 1–5. doi: 10.1007/s13238-020-00807-6
- Li, J., Chang, M., Gao, Q., Song, X., and Gao, Z. (2020). Lung cancer classification and gene selection by combining affinity propagation clustering and sparse group lasso. *Curr. Bioinform.* 15, 703–712. doi: 10.2174/1574893614666191017103557
- Lippi, G., Lavie, C. J., and Sanchis-Gomar, F. (2020). Cardiac troponin I in patients with coronavirus disease 2019 (COVID-19): evidence from a meta-analysis. *Prog. Cardiovasc. Dis.* 63, 390–391. doi: 10.1016/j.pcad.2020.03.001
- Liu, G. Y., Zhao, Y., Jin, S., Hu, Y., Wang, T., Tian, R., et al. (2018). Circulating vitamin E levels and Alzheimer's disease: a Mendelian randomization study. *Neurobiol. Aging* 72, 189.e1–189.e9. doi: 10.1016/j.neurobiolaging.2018.08.008
- Marshall, M. (2020). How COVID-19 can damage the brain. *Nature* 585, 342–343. doi: 10.1038/d41586-020-02599-5
- Meng, C., Zhang, J., Ye, X., Guo, F., and Zou, Q. (2020). Review and comparative analysis of machine learning-based phage virion protein identification methods. *Biochim. Biophys. Acta Proteins Proteom.* 1868:140406. doi: 10.1016/j.bbapap.2020.140406
- Narang, P., Dangi, M., Sharma, D., Khichi, A., and Chhillar, A. K. (2019). An integrated Chikungunya virus database to facilitate therapeutic analysis: ChkVDb. *Curr. Bioinform.* 14, 323–332. doi: 10.2174/1574893613666181029124848
- Qi, C., Wang, P., Fu, T., Lu, M., Cai, Y., Chen, X., et al. (2021). A comprehensive review for gut microbes: technologies, interventions, metabolites and diseases. *Brief. Funct. Genomics* 20, 42–60. doi: 10.1093/bfpg/ela029
- Qi, R., Ma, A., Ma, Q., and Zou, Q. (2020). Clustering and classification methods for single-cell RNA-sequencing data. *Brief. Bioinform.* 21, 1196–1208. doi: 10.1093/bib/bbz062
- Shirazi, S., Mami, S., Mohtadi, N., Ghaysouri, A., Tavan, H., and Nazari, A. (2021). Sudden cardiac death in COVID-19 patients, a report of three cases. *Future Cardiol.* 17, 113–118. doi: 10.2217/fca-2020-0082
- Sun, L., Liu, G., Su, L., and Wang, R. (2019). HS-MMGKG: a fast multi-objective harmony search algorithm for two-locus model detection in GWAS. *Curr. Bioinform.* 14, 749–761. doi: 10.2174/1574893614666190409110843
- Wan, Y., Shang, J., Graham, R., Baric, R. S., and Li, F. (2020). Receptor recognition by the novel coronavirus from wuhan: an analysis based on decade-long structural studies of SARS Coronavirus. *J. Virol.* 94, e127–e120.
- Wang, P., Jin, X., Zhou, W., Luo, M., Xu, Z., Xu, C., et al. (2020). Comprehensive analysis of TCR repertoire in COVID-19 using single cell sequencing. *Genomics* 113, 456–462. doi: 10.1016/j.ygeno.2020.12.036
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269.
- Yadav, R., Bansal, R., Budakoty, S., and Barwad, P. (2020). COVID-19 and sudden cardiac death: a new potential risk. *Indian Heart J.* 72, 333–336. doi: 10.1016/j.ihj.2020.10.001
- Yuan, F., Lu, L., and Zou, Q. (2020). Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. *Biochim. Biophys. Acta Mol. Basis Dis.* 1866:165822. doi: 10.1016/j.bbadis.2020.165822
- Zhao, S., Jiang, H., Liang, Z. H., and Ju, H. (2019). Integrating multi-omics data to identify novel disease genes and single-nucleotide polymorphisms. *Front. Genet.* 10:1336. doi: 10.3389/fgene.2019.01336
- Zhao, T., Hu, Y., and Cheng, L. (2020a). Deep-DRM: a computational method for identifying disease-related metabolites based on graph deep learning approaches. *Brief. Bioinform.* bbaa212. doi: 10.1093/bib/bbaa212
- Zhao, T., Hu, Y., Peng, J., and Cheng, L. (2020b). DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* 36, 4466–4472. doi: 10.1093/bioinformatics/btaa428
- Zhao, T., Hu, Y., Zang, T., and Cheng, L. (2020c). MRTFB regulates the expression of NOMO1 in colon. *Proc. Natl. Acad. Sci.* 117, 7568–7569. doi: 10.1073/pnas.2000499117
- Zhao, T., Hu, Y., Zang, T., and Wang, Y. (2019). Integrate GWAS, eQTL, and mQTL data to identify Alzheimer's disease-related genes. *Front. Genet.* 10:1021. doi: 10.3389/fgene.2019.01021
- Zhao, T., Lyu, S., Lu, G., Juan, L., Zeng, X., Wei, Z., et al. (2021). SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Res.* 49, D1413–D1419. doi: 10.1093/nar/gkaa838
- Zhao, X., Jiao, Q., Li, H., Wu, Y., Wang, H., Huang, S., et al. (2020). ECFS-DEA: an ensemble classifier-based feature selection for differential expression analysis on expression profiles. *BMC Bioinformatics* 21:43. doi: 10.1186/s12859-020-3388-y
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zhu, X., Li, H.-D., Guo, L., Wu, F.-X., and Wang, J. (2019). Analysis of single-cell RNA-seq data by clustering approaches. *Curr. Bioinform.* 14, 314–322. doi: 10.2174/1574893614666181120095038
- Zou, Q., Lin, G., Jiang, X., Liu, X., and Zeng, X. (2020). Sequence clustering in bioinformatics: an empirical study. *Brief. Bioinform.* 21, 1–10. doi: 10.1093/bib/bby090

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zong and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.