Mechanisms of Ageing and Development 194 (2021) 111433

Contents lists available at ScienceDirect

Mechanisms of Ageing and Development

journal homepage: www.elsevier.com/locate/mechagedev

Association analysis framework of genetic and exposure risks for COVID-19 in middle-aged and elderly adults

Yuan Zhang [a,1], Hongxi Yang [a,1], Shu Li [a], Wei-Dong Li [b], Ju Wang [c], Yaogang Wang [a,*]

[a] School of Public Health, Tianjin Medical University, Tianjin, China
[b] Department of Genetics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China
[c] School of Biomedical Engineering and Technology, Tianjin Medical University, Tianjin, China

ARTICLE INFO

Keywords:
Genetic risks
Exposure risks
Association analysis
COVID-19

ABSTRACT

Coronavirus disease 2019 (COVID-19) is a current pandemic, and studies reported that older people have higher rates of infection and more severe cases. Recently, studies have revealed the involvement of both genetic and exposure factors in the susceptibility of COVID-19. However, the correlation between them is still unclear. Thus, we aimed to investigate the correlation between genetic and exposure factors associated with COVID-19. We retrieved the information of 7362 participants with COVID-19 testing results from the UK Biobank. We identified genetic factors for COVID-19 by genome-wide association studies (GWAS) summary analysis. In this study, 21 single-nucleotide polymorphisms (SNPs) and 15 exposure factors [smoking, alcohol intake, daytime dozing, body mass index (BMI), triglyceride, High Density Lipoprotein (HDL), diabetes, chronic kidney disease, chronic liver disease, dementia, atmosphere NO2 concentration, socioeconomic status, education qualification, ethnicity, and income] were found to be potential risk factors of COVID-19. Then, a gene-exposure (G × E) association network was built based on the correlation among and between these genetic factors and exposure factors. rs140092351, a SNP on microRNA miR1202, not only had the most significant association with COVID-19, but also interacted with multiple exposure factors. Dementia, alcohol consumption, daytime dozing, BMI, HDL, and atmosphere NO_2 concentration were among most significant G × E interactions with COVID-19 infection ($P = 0.001$).

1. Introduction

Coronavirus disease 2019 (COVID-19) is a current pandemic caused by a positive-sense RNA virus named the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)(Zhu et al., 2020). Patients infected with COVID-19 might develop acute respiratory distress syndrome, have a high likelihood of admission to intensive care, and might die.(Huang et al., 2020) In addition, COVID-19 has a very dynamic structure and spreads rapidly. As of Jul 29, 2020, approximately 16.5 million cases and 655,112 deaths have been confirmed worldwide (WHO, 2020). Human genetic and exposure factors may contribute to the extremely high transmissibility of SARS-CoV-2 and to the relentlessly progressive disease observed in a small but significant proportion of infected individuals; yet, these factors are largely unknown. Development of new preventive and/or therapeutic strategies for COVID-19 will be greatly facilitated by systematic identification of exposure factors and gene polymorphisms which modulate the risk of infection and severe illness.

Recently, studies have focused on the characteristics(Lescure et al., 2020; Liu et al., 2020; Shereen et al., 2020), epidemiology(Bi et al., 2020; Zhai et al., 2020; Zhang, 2020), and genomic characterization (Devaux et al., 2020; Ellinghaus et al., 2020; Ovsyannikova et al., 2020) of COVID-19 infection. These studies reported that older people have higher rates of infection and more severe cases. Hou et al. investigated genetic susceptibility to COVID-19 by examining DNA polymorphisms in ACE2 and TMPRSS2 from ~81,000 human genomes, found that ACE2 or TMPRSS2 DNA polymorphisms were likely associated with genetic susceptibility of COVID-19, calling for a human genetics initiative for fighting the COVID-19 pandemic(Hou et al., 2020). However, little is known about the correlation between the genetic and exposure factors associated with the infection of COVID-19.

We hypothesized the existence of nonrandom correlation among and between the genetic and exposure factors associated with COVID-19, based on which an association network of these factors can be built. Then, by examining whether a person fit into such association network

* Corresponding author at: School of Public Health, Tianjin Medical University, Tianjin, 300070, China.
E-mail address: YaogangWANG@tmu.edu.cn (Y. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.mad.2021.111433
Received 25 September 2020; Received in revised form 27 December 2020; Accepted 5 January 2021
Available online 12 January 2021

"pattern" could provide us a more comprehensive assessment of the risks, susceptibility, and treatment responses of COVID-19, and improve our understanding on the etiology of the disease. Thus, in the present study, we aimed to investigate the correlation of genetic and exposure factors associated with COVID-19 in middle-aged and elderly adults, as well as the global phenotype-genotype association framework for the disease.

2. Methods

2.1. Study design and population

Data related to COVID-19 were obtained from the UK Biobank, a health resource for a population-based study of more than 500,000 participants that attended one of 22 assessment centers across the United Kingdom between 2006 and 2010.(Cox, 2018; Sudlow et al., 2015) Participants provided extensive information via questionnaires, interviews, health records, physical measures, blood samples, and genotype results, allowing for linkage of extensive exposure, genetic and clinical data. Recently, COVID-19 testing results for a subset of participants were made available by Public Health England.(Armstrong et al., 2020) In this study, 7362 participants (mean age 69 years) with COVID-19 testing results or with exposure and genetic information were included.

2.2. Assessment of exposure factors

In present study, the exposure factor screening was based on a previously published review(Zhang et al., 2020). Based on the extensive review and analysis of the above-mentioned review, we have enriched, improved, integrated, and assembled the literature on the exposure risk factors, methods, and models of COVID-19, and applied UKB data to analyze, verify and expand. In this study, the exposure factors are organized into five hierarchical levels, including behavior risks, metabolic risks, disease risks, environmental risks, and socio-demographic index.

We used 17 indicators for behavior risks. Briefly, smoking status was categorized as never, previous, or current smoking. Regular physical activity was defined as per week $\geq$150 min of moderate activity, or per week $\geq$ 75 min of vigorous activity (Lloyd-Jones et al., 2010). Alcohol intake (including wine, beer, spirits, and fortified wine) was categorized as <1 g/day, 1–7 g/day, 8–15 g/day, and $\geq$ 16 g/day. All sleep behaviors were self-reported, and we included six sleep factors (chronotype, duration, insomnia, snoring, daytime dozing, and nap during day). All of the UK Biobank participants completed a questionnaire on their usual dietary pattern, most of which asked about the frequency of consumption of main foods and food groups. The questions used in this manuscript are those that asked about the frequency of consumption of fresh fruit, raw vegetables, cooked vegetables, oily fish, non-oily fish, processed meat, beef, lamb, pork, tea, and coffee.

Nine metabolic risk factors were included in our study. Of them, height, weight, waist circumference, and hip circumference were measured directly during a medical examination from which body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared. Non-fasting venous blood, available in a sub-sample, was drawn with assaying conducted at dedicated central laboratory for uric acid, cholesterol, triglyceride, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and Vitamin D.(Elliott et al., 2008)

Eleven disease factors were employed in the study. Vital statuses of each participant were identified chiefly using linkage with hospital admission data. Disease affection statuses were documented, including type 2 diabetes (T2DM), chronic kidney disease, hypertension, depression, dementia, cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD), asthma, chronic liver disease, and cancer. Each disease factor was categorized as undiagnosed, diagnosed < 10 years ago, and diagnosed $\geq$ 10 years ago according to disease duration.

Environmental exposures, were collected by the Small Area Health Statistics Unit as part of the BioSHaRE-EU Environmental Determinants of Health Project (http://www.bioshare.eu/). UK Biobank is a participating biobank in this project. In this study, 4 environmental factors, including PM2.5, PM10, NO2, and NO_x were included into environmental exposures.

We used 4 indicators of socio-demographic index. Total annual household incomes before tax were self-reported and classified into five groups (Less than £18,000, 18,000–30,999, 31,000–51,999, 52,000–100,000, and greater than 100,000). For educational qualifications, we used a seven category variable (College or University degree, A levels/AS levels or equivalent, O levels/GCSEs or equivalent, CSEs or equivalent, NVQ or HND or HNC or equivalent, other professional qualifications, and none of the above). Socioeconomic status categories derived from Townsend deprivation index(Guillaume et al., 2016) quintiles 1, 2–4, and 5, combining information on social class, employment, car availability, and housing. Ethnicity was self-reported and categorized as White, Mixed, Asian or Asian British, Black or Black British, and other ethnic groups.

2.3. Ascertainment of genetic factors

Genetic risk factors for COVID-19 were identified by the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/), (Initiative, 2020) a global initiative to bring together the human genetics community to generate, share, and analyze data to learn the genetic determinants of COVID-19 susceptibility, severity, and outcomes. In this study, 102 single-nucleotide polymorphisms (SNPs) reaching a conventional genome-wide significance threshold of *P*-value $<1 \times 10^{-6}$ were identified (Supplementary Method, Supplementary Table 1).

2.4. Ascertainment of hospitalization for COVID-19

Provided by Public Health England, data on COVID-19 status downloaded on July 6, 2020 covered the period March 16, 2020 until May 31, 2020. Nose and/or throat swabs were taken from hospitalized patients and detection of SARS-CoV-2 can be reported as positive or negative.

2.5. Covariates

All of the models were adjusted for age, sex, ethnicity (white, mixed, Asian, black, and others), qualifications (College degree, A levels/AS levels, O levels/GCESs, CSEs, NVQ or HND or HNC, other professional qualifications, and none of the above), and socioeconomic status (categories derived from Townsend deprivation index(Guillaume et al., 2016) quintiles 1, 2–4, and 5, combining information on social class, employment, car availability, and housing).

2.6. Statistical analyses

Baseline characteristics of the samples were summarized across COVID-19 infection status as percentages for categorical variables and means and standard deviations (SDs) for continuous variables. Student's *t*-test was used to compare the means of continuous variables and normally distributed data; otherwise, the Mann–Whitney *U* test was applied. A Shapiro–Wilk normality test was used to assess the normality of the distribution. Categorical data were assessed by chi-square test. Multivariate logistic regression analyses were used to assess the association of both exposure and genetic factors with the risk of COVID-19. For exposure factors, in Model 1, we conducted univariate logistic regression with unadjusted; in Model 2, we adjusted for age and sex; in Model 3, we selected all the significant variables in the Model 2 to enter the multivariate logistic regression model. For genetic factors, Model 1, in which we included all SNPs with unadjusted for other confounding;

Model 2, we adjusted for age, sex ethnicity, education level, and socioeconomic statuses.

Gene-gene and gene-exposure interactions were analyzed using generalized multifactor dimensionality reduction (GMDR) (http://www.ssg.uab.edu/gmdr/). The best gene-gene, gene-exposure interaction model based on the values arising from 10-fold cross-validation (CV) consistency and accuracy testing were selected. A permutation test with 1000 replications was used to measure the empirical *P* values thereby substantiating the significance of the model. For GMDR method, sex, age, ethnicity, qualification, and socioeconomic status were used to build a score statistic with adjustment for the covariates. In GMDR analysis, a *P* value was corrected for multiple testing by permutation test and a corrected P value < 0.05 (two-tailed) was considered to be statistically significant. For validating the results of GMDR, OR (with 95 % CI) of risk factors were computed by logistic regression analysis. To narrow down the number of possible combinations, only dominant models were subjected to further analysis. Cytoscape (version 3.7.1) was used to layout the association network.(Shannon et al., 2003) All statistical analyses were performed using R (version 3.6.1).

3. Results

3.1. Characteristics of participants

In this analysis, after excluding participants without genetic information or without exposure information, 7362 participants were ultimately included for these samples, the mean age was 69.20 ± 8.68 years, and 3647 (49.54 %) individuals were male. In total, 1485 (20.17 %) participants were positive for COVID-19 infection. The baseline characteristics of the participants are provided in Table 1. Compared to participants negative for COVID-19 infection, the positive participants were more likely to be male, be of Asian or Black ethnic group, have a higher socioeconomic status and income; they were also more likely to have a history of T2DM or dementia, a higher BMI and a lower level of HDL, whereas less likely to consume alcohol, and less likely to have a university degree.

3.2. Genetic and exposure factors associated with COVID-19

From the 102 SNPs identified in the GWAS summary analysis, we obtained 21 SNPs that were associated with COVID-19 (Fig. 1A; Supplementary Table 2). The rs140092351 locus on microRNA MIR1202 yielded the most significant association. For the 45 exposure factors examined (Supplementary Table 3), we found 15 exposure factors associated with COVID-19, including smoking, alcohol intake, daytime dozing, BMI, TG, HDL, diabetes, chronic kidney disease, chronic liver disease, dementia, atmosphere NO_2 concentration, socioeconomic status, education qualifications, ethnicity, and income (Fig. 1B).

3.3. Correlation between the genetic and exposure factors

We also detected associations between the genetic and exposure factors. Altogether, 247 associations among and between the 21 genetic risks and 15 exposure factors of COVID-19 infection were identified, based on which a risk factor association network was conducted (Fig. 2A). Fig. 2B shows the correlation coefficient of SNPs and exposure factors. Among the exposure factors, ethnicity was associated with sixteen genetic loci of COVID-19, and atmosphere NO2 concentration was associated with ten genetic loci of COVID-19, while alcohol intake was associated with nine gene loci of COVID-19. Furthermore, smoking was associated with eight gene loci of COVID-19, and T2DM was associated with four gene loci of COVID-19.

3.4. Gene-exposure interaction

The significance of gene-exposure interaction was further evaluated

Table 1
Characteristics of the participants by COVID-19.

Characteristic	Non-COVID-19	COVID-19	*P* value
No (%)	5877(79.83)	1485(20.17)	
Age, mean (SD), year	69.46(8.52)	68.16(9.23)	<0.001
Male, no. (%)	2863(48.72)	784(52.79)	0.005
Ethnicity, no. (%)			<0.001
White	5455(93.3)	1291(87.53)	
Mixed	40(0.68)	10(0.68)	
Asian or Asian British	135(2.31)	64(4.34)	
Black or Black British	135(2.31)	77(5.22)	
Other	82(1.4)	33(2.24)	
Diabetes, no. (%)			0.004
Undiagnosed	4928(83.85)	1202(80.94)	
Diagnosed < 10 years ago	660(11.23)	180(12.12)	
Diagnosed ≥ 10 years ago	289(4.92)	103(6.94)	
Chronic kidney disease, no. (%)			0.053
Undiagnosed	4701(79.99)	1159(78.05)	
Diagnosed < 10 years ago	1044(17.76)	278(18.72)	
Diagnosed ≥ 10 years ago	132(2.25)	48(3.23)	
Dementia, no. (%)			<0.001
Undiagnosed	5563(94.66)	1296(87.27)	
Diagnosed < 10 years ago	306(5.21)	188(12.66)	
Diagnosed ≥ 10 years ago	8(0.14)	1(0.07)	
Chronic liver disease, no. (%)			0.053
Undiagnosed	5463(92.96)	1402(94.41)	
Diagnosed < 10 years ago	347(5.9)	64(4.31)	
Diagnosed ≥ 10 years ago	67(1.14)	19(1.28)	
Alcohol intake, no. (%)			0.002
<1 g/day	1769(30.24)	509(34.46)	
1 to 7 g/day	989(16.91)	265(17.94)	
8 to 15 g/day	1140(19.49)	245(16.59)	
≥16 g/day	1952(33.37)	458(31.01)	
Smoking, no. (%)			0.133
Current	2800(47.92)	720(48.98)	
Former	2260(38.68)	582(39.59)	
Never	783(13.4)	168(11.43)	
Daytime dozing, no. (%)			0.015
Never/rarely	4176(71.73)	994(67.9)	
Sometimes	1418(24.36)	404(27.6)	
Usually	228(3.92)	66(4.51)	
NO_2 (μg/m^3)			<0.001
1(lowest fifth)	986(16.99)	196(13.4)	
2	1104(19.02)	266(18.18)	
3	1157(19.93)	300(20.51)	
4	1255(21.62)	298(20.37)	
5(highest fifth)	1302(22.43)	403(27.55)	
BMI(kg/m^2)			0.001
<25	1571(27.58)	346(23.96)	
25 to 29.9	2367(41.56)	606(41.97)	
30 to 34.9	1202(21.1)	304(21.05)	
≥35	556(9.76)	188(13.02)	
HDL(mmol/l)			<0.001
Q1(0.8−1.17)	1517(29.51)	434(34.23)	
Q2(1.18−1.40)	1278(24.86)	348(27.44)	
Q3(1.41−1.67)	1200(23.35)	286(22.56)	
Q4(1.68−3.58)	1145(22.28)	200(15.77)	
TG(mmol/l)			0.104
<1.7	3203(57.29)	754(54.28)	
1.7−2.25	1049(18.76)	288(20.73)	
>2.26	1339(23.95)	347(24.98)	
Socioeconomic status, no. (%)			<0.001
1(lowest deprived)	991(16.89)	206(13.88)	
2	1078(18.37)	244(16.44)	
3	1090(18.58)	257(17.32)	
4	1157(19.72)	313(21.09)	
5(highest deprived)	1551(26.44)	464(31.27)	
Qualification, no. (%)			<0.001
College or University degree	1606(27.7)	351(24.06)	
A levels/AS levels or equivalent	590(10.18)	135(9.25)	
O levels/GCSEs or equivalent	1190(20.53)	265(18.16)	
CSEs or equivalent	309(5.33)	103(7.06)	
NVQ or HND or HNC or equivalent	449(7.75)	138(9.46)	
Other professional qualifications	332(5.73)	93(6.37)	
None of the above	1321(22.79)	374(25.63)	
Income, no. (%)			0.094
Less than 18,000	1575(31.68)	408(33.04)	

(*continued on next page*)

Table 1 (*continued*)

Characteristic	Non-COVID-19	COVID-19	*P* value
18,000–30,999	1252(25.19)	308(24.94)	
31,000–51,999	1107(22.27)	293(23.72)	
52,000–100,000	799(16.07)	187(15.14)	
Greater than 100,000	238(4.79)	39(3.16)	

Abbreviations: BMI body mass index; CSE Certificate of Secondary Education; GCSE General Certificate of Secondary Education; HDL high density lipoprotein; HNC Higher National Certificate; HND Higher National Diploma; NVQ National Vocational Qualification; TG triglyceride.

using the GMDR model with age, sex, ethnicity, qualification, and socioeconomic status as covariates (Table 2). Dementia, alcohol consumption, daytime dozing, BMI, HDL, and atmosphere NO2 concentration were among most significant G x E interaction with COVID-19 infection ($P < 0.001$). Furthermore, we assessed the exposure factors selected by GMDR using logistic regression analysis, which incorporated age, sex, ethnicity, qualification, and socioeconomic status as covariates. Results were summarized in Table 3. For example, individuals with allele A+ (GA, AA) of rs3136704 and had dementia were more susceptible (OR, 4.25; 95 % CI, 2.91−6.19) to COVID-19 infection relative to the rest of the study population. While subjects carrying the G^- allele (GG) of rs140092351 or T^+ allele (CT, TT) of rs12950851 with consumption alcohol 8−15 g/day (OR, 0.47; 95 % CI, 0.32−0.67) were in lower risk of COVID-19 infection compared to others.

4. Discussion

We found a significant association framework between and among genetic and exposure factors of COVID-19 infection. The rs140092351 locus on a microRNA miR1202 not only had the most significant association with COVID-19, but also interacted with multiple exposure factors. Dementia, alcohol consumption, daytime dozing, BMI, HDL, and atmosphere NO_2 concentration were among most significant G x E interactions with COVID-19 infection.

Our findings suggested that 15 exposure factors, including diabetes, dementia, chronic kidney disease, chronic liver disease, smoking, alcohol intake, daytime dozing, BMI, TG, HDL, atmosphere NO2 concentration, socioeconomic status, education qualifications, ethnicity, and income were associated with COVID-19 infection. Emerging evidence has suggested that smoking status, black or Asian/Asian British, deprivation (both with a strong gradient), diabetes, reduced kidney function, chronic liver disease, and diabetes are risk factors of COVID-19 and resulting complications.(Huang et al., 2020; Williamson et al., 2020) However, the association of daytime dozing, TG, HDL, atmosphere NO2 concentration, and qualification with COVID-19 have not been reported. In our study, individuals which reported daytime dozing "sometimes", a high level of TG (1.7−2.25 mmol/l), a lower level of HDL (<1.67 mmol/l), a high atmosphere NO2 concentration (>32.5 $\mu g/m^3$), and a low qualification were more vulnerable to COVID-19 infection.

Our results found that there are wide correlations between exposure factors and susceptibility genes for COVID-19 infection. At the population level, the distribution of susceptible genes among individuals is characteristic, and susceptible individuals often have a series of susceptible gene polymorphisms. We called the genotype combination of susceptible genes that are associated with a certain phenotype a "pangenotype." However, despite considering pathway enrichment and pairwise association between genotypes, previous studies have often targeted a single phenotype. In fact, the combination of certain genotypes is related to different phenotypes, and some phenotypes are also associated with each other. Thus, different genotypes and phenotypes form an association networks. In this study, the non-random combinations of genotypes ("genetic signature") clustered in exposure factors associated with COVID-19. Thus, it is important to take systemic look of the multi-dimensional network for COVID-19.

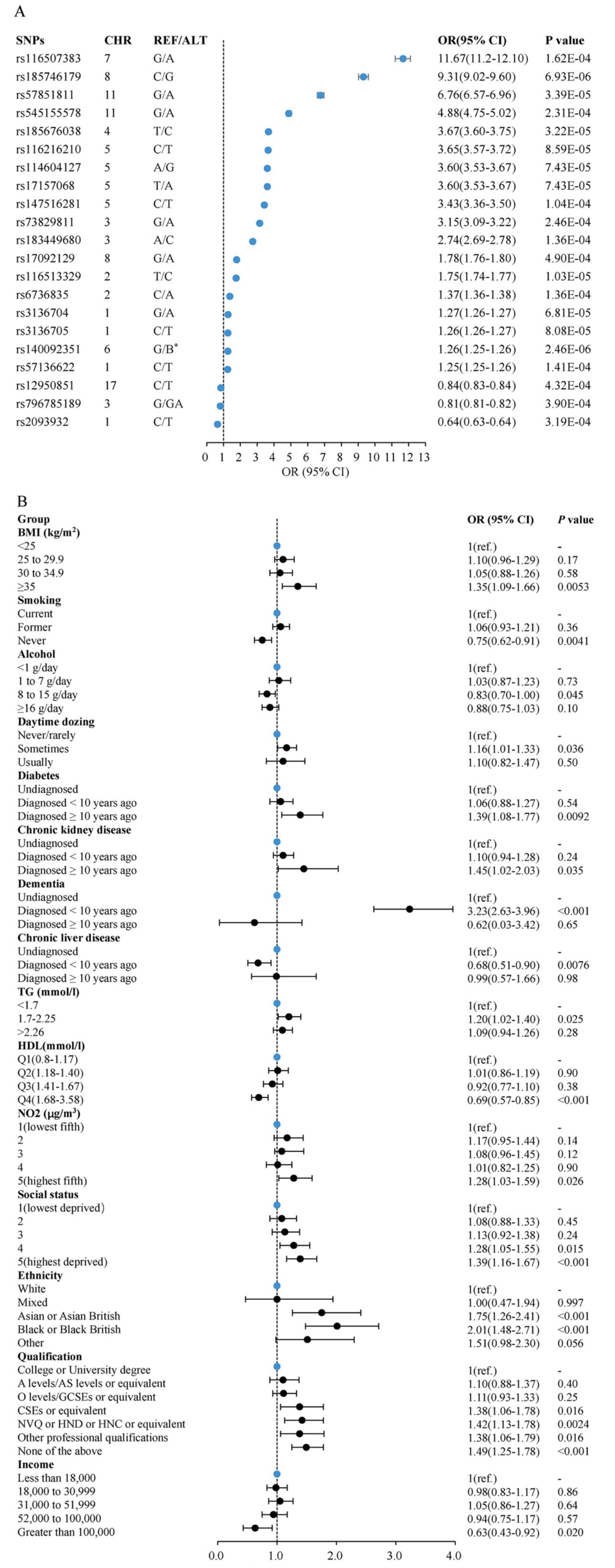

Fig. 1. Association of exposure (A) and genetic (B) factors with COVID-19 infection.
Multivariate logistic model adjusted for age, sex, ethnicity, socioeconomic status, and qualification.*B = GTTTCTCTAGTTTGGA.

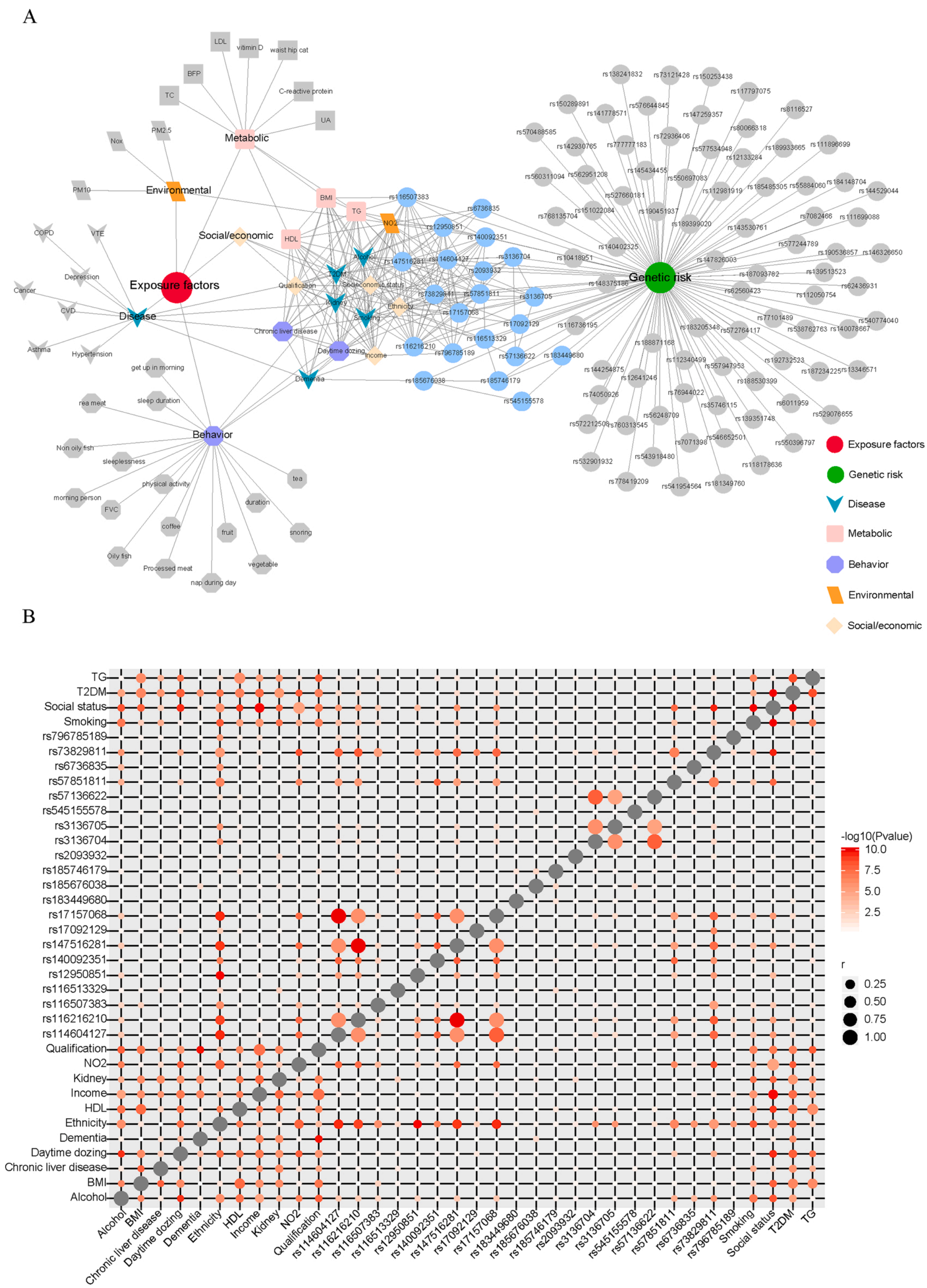

Fig. 2. Association of genetic and exposure factors of COVID-19 infection.
(A) The network of association analysis framework for genetic and exposure factors for COVID-19 infection. Gray shading indicates genetic or exposure factors are not associated with COVID-19 infection. (B) Heatmap of correlation coefficients of 21 genetic and 15 exposure factors.

Dementia, alcohol consumption, daytime dozing, BMI, HDL, and atmosphere NO2 concentration were among the most significant G x E interactions with COVID-19 infection. The rs140092351 locus, located on microRNA miR1202 yielded the most significant association, while multiple G x E interactions were connected by that polymorphism. The function of rs140092351 remains unknown, although several studies have reported that a microRNA that 12Kb upstream of rs140092351, miR-1202, was related to brain tumors, glioma, depression, and neuroinflammation. Juan et al. reported that miR-1202 was abnormally expressed in prefrontal cortex in depressed patients (Lopez et al., 2014). Furthermore, Song et al. claimed that over-expression of miR-1202 could inactivate TLR4/NF-κB related inflammatory signal pathway through targeting its target protein Rab1a to play a protective role in neuroinflammation (Song et al., 2020). The present result regarding G x

Table 2
Best gene–gene/exposure interaction models as identified by GMDR.

Best combination	Testing accuracy	Cross-validation consistency	P value[a]
Gene-Gene interaction			
rs140092351, rs5852036	0.530	8/10	0.001
rs140092351, rs5852036, rs116513329	0.534	6/10	0.001
Gene-dementia interaction			
rs3136704, dementia	0.541	7/10	0.001
rs3136704, rs140092351, dementia	0.557	9/10	0.001
Gene-T2DM interaction			
rs3136704, T2DM	0.518	5/10	0.055
rs3136704, rs6736835, T2DM	0.505	3/10	0.055
Gene-alcohol interaction			
rs140092351, alcohol	0.525	7/10	0.001
rs140092351, rs12950851, alcohol	0.528	6/10	0.011
Gene-daytime dozing interaction			
rs140092351, rs12950851, daytime dozing	0.533	8/10	0.001
Gene-NO_2 interaction			
rs140092351, NO2	0.518	6/10	0.055
rs140092351, rs12950851, NO2	0.550	10/10	0.001
Gene-BMI interaction			
rs3136704, BMI	0.508	3/10	0.055
rs140092351, rs12950851, BMI	0.535	7/10	0.001
Gene-TG interaction			
rs140092351, TG	0.520	3/10	0.055
rs140092351, rs57136622, TG	0.520	6/10	0.055
Gene-smoking interaction			
rs140092351, smoking	0.520	2/10	0.055
rs140092351, rs3136704, smoking	0.530	6/10	0.011
Gene-HDL interaction			
rs12950851, HDL	0.532	6/10	0.001
rs140092351, rs12950851, HDL	0.557	10/10	0.001
Gene-qualification interaction			
rs140092351, qualification	0.526	9/10	0.001
rs140092351, rs3136704, qualification	0.522	4/10	0.055
Gene-income interaction			
rs140092351, income	0.521	7/10	0.055
rs140092351, rs12950851, income	0.529	7/10	0.055
Gene-TDI interaction			
rs140092351, TDI	0.519	4/10	0.055
rs140092351, rs12950851, TDI	0.531	7/10	0.055

[a] GMDR analysis adjusted for age, sex, ethnicity, socioeconomic status, and qualification.

E interaction is an important discovery that may indicate new unreported biological pathways and mechanisms that need to be further verified.

Our present study also had several limitations. We are unable to assess exposure to SARS-CoV-2 in most UKB participants. This has important implications for case–control studies, because we cannot distinguish individuals who have not contracted SARS-CoV-2 following exposure from those who have not been exposed. Furthermore, genetic factors related to exposure factors may not cause COVID-19 by themselves, but likely to increase the susceptibility of the disease by increasing the risk of phenotypic factors (both behavioral and pathophysiologic) that associated with it. In addition, our exposure factors were collected between 2006 and 2010 and may not represent the current state of exposure.

Unlike previous protein-protein or genetic interaction studies, in this study, we conducted a unique association network among phenotypes, lifestyle, environmental, genotypes, disease of associated with COVID-19. The research results of above association analysis framework shows that when genetic factors of COVID-19 cannot be changed, the identification and improvement of genetically-related exposure factors can modulate the infection of COVID-19.

Table 3
Stratified analysis for interaction between gene and gene/environment on COVID-19.

Factors			OR (95 % CI)[a]	P value
Gene-Gene interaction				
rs140092351	rs5852036			
GG	GGA or GAGA		1(ref.)	
GB or BB[b]	GG		1.62 (1.35−1.95)	<0.001
rs140092351	rs5852036	rs116513329		
GG	GGA or GAGA	TT	1(ref.)	
GB or BB	GG	TC or CC	2.62 (1.59−4.25)	<0.001
Gene-dementia interaction				
Dementia	rs3136704			
No	GG		1(ref.)	
Yes	GA or AA		4.25 (2.91−6.19)	<0.001
Dementia	rs3136704	rs140092351		
No	GG	GG	1(ref.)	
Yes	GA or AA	GB or BB	5.82 (3.25−10.47)	<0.001
Gene-alcohol interaction				
Alcohol	rs140092351			
0−7 g/day	GB or BB		1(ref.)	
8−15 g/day	GG		0.67 (0.53−0.84)	<0.001
Alcohol	rs140092351	rs12950851		
0−7 g/day	GB or BB	CC	1(ref.)	
8−15 g/day	GG	CT or TT	0.47 (0.32−0.67)	<0.001
Gene-smoking interaction				
Smoking	rs140092351	rs3136704		
Never	GG	GG	1(ref.)	
Current	GB or BB	GA or AA	2.15 (1.47−3.16)	<0.001
Gene-daytime dozing interaction				
Daytime dozing	rs140092351	rs12950851		
Never/rarely	GG	CT or TT	1(ref.)	
Sometimes	GB or BB	CC	1.70 (1.27−2.25)	<0.001
Gene-BMI interaction				
BMI	rs140092351	rs12950851		
$<25\ kg/m^2$	GG	CT or TT	1(ref.)	
$\geq 35\ kg/m^2$	GB or BB	CC	2.30 (1.40−3.77)	<0.001
Gene-HDL interaction				
HDL	rs140092351	rs12950851		
0.8−1.17 mmol/l	GB or BB	CC	1(ref.)	
1.68−3.58 mmol/l	GG	CT or TT	0.57 (0.36−0.90)	0.019
Gene-NO2 interaction				
NO2	rs140092351	rs12950851		
1(lowest fifth)	GG	CT or TT	1(ref.)	
5(highest fifth)	GB or BB	CC	2.40 (1.41−4.11)	0.0013
Gene-qualification interaction				
Qualification	rs140092351			
College	GG		1(ref.)	
GSEs/NVQ/HND/HNC	GB or BB		1.83 (1.47−2.27)	<0.001

[a] Multivariable logistic model adjusted for age, sex, ethnicity, socioeconomic status, and qualification.
[b] B = GTTTCTCTAGTTTGGA.

5. Conclusion

We found a significant association framework between and among genetic and exposure factors of COVID-19 infection. Phenotype-genotype association were common among genetic and exposure factors. The rs140092351 locus on a microRNA miR1202 not only had the most significant association with COVID-19, but also interacted with multiple exposure factors. Dementia, alcohol consumption, daytime dozing, BMI, HDL, and atmosphere NO_2 concentration were among most significant G x E interactions with COVID-19 infection. Our findings will provide a new perspective for comprehensive prevention and treatment of COVID-19.

Funding

This work was supported by the National Natural Science Foundation of China (grant numbers: 71910107004, 91746205).

Author contributions

YW conceived the idea. YW, YZ and HY designed the study. YW, YZ and HY led the analysis with support from SL. YW and YZ drafted the paper, YW, YZ, WL, and JW finalized the paper. All authors contributed to the analysis, intellectual content, critical revisions to the drafts of the paper and approved the final version.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

We thank the participants of the UK biobank. This research has been conducted using the UK biobank Resource under the project number of 45676. We would like to thank all UK Biobank participants and staff, and all health-care workers involved in the diagnosis and treatment of COVID-19 patients.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.mad.2021.111433.

References

Armstrong, J., et al., 2020. Dynamic linkage of COVID-19 test results between Public Health England's second generation surveillance system and UK Biobank. Microb. Genom.

Bi, Q., et al., 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect. Dis. 20, 911–919.

Cox, N., 2018. UK Biobank shares the promise of big data. Nature 562, 194–195.

Devaux, C.A., et al., 2020. ACE2 receptor polymorphism: susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. J. Microbiol. Immunol. Infect. 53, 425–435.

Ellinghaus, D., et al., 2020. Genomewide association study of severe Covid-19 with respiratory failure. N. Engl. J. Med.

Elliott, P., et al., 2008. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int. J. Epidemiol. 37, 234–244.

Guillaume, E., et al., 2016. Development of a cross-cultural deprivation index in five European countries. J. Epidemiol. Community Health 70, 493–499.

Hou, Y., et al., 2020. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Med. 18, 216.

Huang, C., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506.

Initiative, C.-H.G., 2020. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur. J. Hum. Genet. 28, 715–718.

Lescure, F.X., et al., 2020. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. Lancet Infect. Dis. 20, 697–706.

Liu, H., et al., 2020. Clinical and CT imaging features of the COVID-19 pneumonia: focus on pregnant women and children. J. Infect. 80, e7–e13.

Lloyd-Jones, D.M., et al., 2010. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation 121, 586–613.

Lopez, J.P., et al., 2014. miR-1202 is a primate-specific and brain-enriched microRNA involved in major depression and antidepressant treatment. Nat. Med. 20, 764–768.

Ovsyannikova, I.G., et al., 2020. The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity. Immunol. Rev. 296, 205–219.

Shannon, P., et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Shereen, M.A., et al., 2020. COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. J. Adv. Res. 24, 91–98.

Song, S.H., et al., 2020. MiR-1202 exerts neuroprotective effects on OGD/R induced inflammation in HM cell by negatively regulating Rab1a involved in TLR4/NF-kappa B signaling pathway. Neurochem. Res. 45, 1120–1129.

Sudlow, C., et al., 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12 e1001779.

Williamson, E.J., et al., 2020. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature.

World Health Organization, 2020. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-Ncov) (Acccessed 29 July 2020). https://www.who.int/news-room/detail/30-01-2020-stat-ement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov.

Zhai, P., et al., 2020. The epidemiology, diagnosis and treatment of COVID-19. Int. J. Antimicrob. Agents 55, 105955.

Zhang, X., 2020. Epidemiology of Covid-19. N. Engl. J. Med. 382, 1869.

Zhang, Y., et al., 2020. A network analysis framework of genetic and nongenetic risks for type 2 diabetes. Rev. Endocr. Metab. Disord.

Zhu, N., et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727–733.