European Journal of Epidemiology
https://doi.org/10.1007/s10654-023-01020-5

COVID-19

Genome-wide epistasis study highlights genetic interactions influencing severity of COVID-19

Shiqi Lin[1,2] · Xingjian Gao[1,3] · Frauke Degenhardt[4] · Yu Qian[1,2] · Tianzi Liu[1] · Xavier Farre Ramon[5] · Syed Sibte Hadi[6] · Manuel Romero-Gómez[7,8,9,10] · Javier Fernández[11,12] · Agustín Albillos[7,13] · Maria Buti Ferret[7,14,15] · Luis Bujanda[7,16] · Antonio Julià[17] · Rafael de Cid[18] · Rosanna Asselta[18,19] · Andre Franke[4] · Fan Liu[1,2,6]

Received: 16 October 2022 / Accepted: 25 May 2023

Abstract

Coronavirus disease 2019 (COVID-19) caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) may lead to life-threatening respiratory symptoms. Understanding the genetic basis of the prognosis of COVID-19 is important for risk profiling of potentially severe symptoms. Here, we conducted a genome-wide epistasis study of COVID-19 severity in 2243 patients with severe symptoms and 12,612 patients with no or mild symptoms from the UK Biobank, followed by a replication study in an independent Spanish cohort (1416 cases, 4382 controls). Our study highlighted 3 interactions with genome-wide significance in the discovery phase, nominally significant in the replication phase, and enhanced significance in the meta-analysis. For example, the lead interaction was found between rs9792388 upstream of *PDGFRL* and rs3025892 downstream of *SNAP25*, where the composite genotype of rs3025892 CT and rs9792388 CA/AA showed higher risk of severe disease than any other genotypes ($P = 2.77 \times 10^{-12}$, proportion of severe cases = 0.24 ~ 0.29 vs. 0.09 ~ 0.18, genotypic OR = 1.96 ~ 2.70). This interaction was replicated in the Spanish cohort ($P = 0.002$, proportion of severe cases = 0.30 ~ 0.36 vs. 0.14 ~ 0.25, genotypic OR = 1.45 ~ 2.37) and showed enhanced significance in the meta-analysis ($P = 4.97 \times 10^{-14}$). Notably, these interactions indicated a possible molecular mechanism by which SARS-CoV-2 affects the nervous system. The first exhaustive genome-wide screening for epistasis improved our understanding of genetic basis underlying the severity of COVID-19.

Keywords COVID-19 · Severity · Genome-wide association study · Epistasis · UK Biobank · Gene–gene interaction

Introduction

The clinical manifestations of the Coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 infection range from asymptomatic to severe and life-threatening. Indeed, most infected people are asymptomatic or develop mild to moderate illness and recover without hospitalization, whereas severe patients develop acute respiratory distress syndrome, sepsis, up to multi-organ failure.

The COVID-19 Host Genetics Initiative (COVID-19 HGI, https://www.covid19hg.org/) represents a global network of academics to investigate the role of human genetics in SARS-CoV-2 infection and COVID-19 severity. To date, a series of genome-wide association studies (GWASs), whole-genome and whole-exome sequencing studies, as well as meta-analyses have been conducted to map host genetic factors in susceptibility and/or severity of the SARS-CoV-2 virus pandemic [1]. These studies together pointed a number of loci, in many instances coding for protein involved in all the stages of the SARS-CoV-2 infection (e.g., the implication of *ACE2*) and in the subsequent disorder (e.g., *SLC6A20* and *LZTFL1*) [2]. Among these loci, the association at 3q21.31 represents the most significant and robust finding, where the top-associated and likely causal single nucleotide polymorphism (SNP) rs10490770 demonstrated an allelic odds ratio (OR) up to 2.7 (95% confidence interval (CI) = 1.8–3.9) for the carriers of the C allele with an age < 60 years to progress into the severe form of the disease [3, 4].

Shiqi Lin, Xingjian Gao and Frauke Degenhardt have contributed equally to this work.

Extended author information available on the last page of the article

Published online: 26 June 2023

Epistasis, also known as "gene–gene interaction", is a phenomenon in which different genetic variants jointly contribute to the disease susceptibility, with the epistatic effect departing from the sum of individual variants' marginal effects. Here, we performed a genome-wide epistasis study (GWES) for COVID-19 severity in British individuals of the UK Biobank (UKBB) by considering 2,243 patients hospitalized or dead within three months from the infection as cases, and 12,612 infected individuals with no or mild symptoms as controls. We conducted 1.57 trillion multiple linear regressions to exhaustively test pair-wise interactions between 1.77 million common variants over the genome, followed by a replication study in 5798 European individuals from a Spanish cohort (1416 severe cases and 4382 healthy individuals).

Methods

The UKBB discovery cohort

The UKBB is a prospective cohort study that collected deep phenotype and genotype data from 500,000 individuals aged 40–69 years between 2006 and 2010 [5]. During the COVID-19 pandemic, Public Health England offered COVID-19 nucleic-acid testing weekly for UKBB participants living in England. Our samples were obtained from 14,855 individuals having suffered from SARS-CoV-2 infection and included in the UKBB (accessed on June 26, 2021). To avoid confounding factors, we only analyzed British-ancestry samples with imputed genotype data (GRCh37). Health outcomes for these individuals, such as cancer diagnoses, are continuously accumulated through the UK National Registry and hospital records. According to the International Classification of Diseases, Tenth Revision, we collected seven main comorbidities i.e., ischemic heart diseases, heart failure, cerebrovascular diseases, hypertensive diseases, diabetes mellitus, chronic kidney diseases, and malignant neoplasms (Supplementary material Table S1). We considered the 2243 patients hospitalized or dead within three months of infection as severe cases and the remaining 12,612 infected patients with no or mild symptoms as controls. The female-male ratio in cases was 0.63, with an average age of 72.43 (standard deviation (SD) = 7.48; age range: 51–83); the female-male ratio in controls was 1.20, with an average age of 64.57 (SD = 8.38; age range: 50–83). All participants provided the UKBB with informed consent. Details on genome-wide genetic data collection are described in [5]. In brief, genotype calling was carried out by two closely related purpose-designed arrays, the UK BiLEVE Axiom array and the UK Biobank Axiom array. Then genotypes were imputed using the IMPUTE4 program (https://jmarchini.org/software/) combined with the Haplotype Reference Consortium (HRC) and UK10K haplotype resources. After removing SNPs with INFO < 0.8, minor allele frequency (MAF) < 0.01, call rate < 0.97 and Hardy–Weinberg equilibrium test (HWE) $P < 1 \times 10^{-4}$, a total of 7,873,610 SNPs were available for subsequent genetic association analysis. All cases and controls passed individual-wise quality controls, i.e., relatedness check using the genomic relationship matrix (GRM) and individual-level call rate (> 0.97).

The Spanish replication cohort

The replication cohort contained 5,798 individuals from the ancestry-matched Spanish cohort (1416 cases, 4382 controls). Here, cases were recruited from intensive care units and general wards at 6 different centers from Spain and were defined as hospitalization only (mild patients) or with respiratory failure (severe cases). They all had a confirmed SARS-CoV-2 viral RNA polymerase-chain-reaction (PCR) test from relevant biologic fluids. Controls were collected from 3 centers and were healthy blood donors. Details regarding sample characteristics and microarray genotyping can be found in [3].

GWAS

The GWAS was conducted using a logistic regression approach in PLINK v1.9 using as covariates: biological sex, age, age × age, the first five genomic principal components, and the seven main comorbidities (i.e., ischemic heart diseases, heart failure, cerebrovascular diseases, hypertensive diseases, diabetes mellitus, chronic kidney diseases, and malignant neoplasms). Allelic ORs and 95%CI for the minor allele were reported. We considered 5×10^{-8} as our genome-wide significance threshold, and 1×10^{-6} as our genome-wide suggestive significance threshold. To examine the overlap between our GWAS signals and previously established COVID-19 loci, we downloaded "B1_ALL_leave_23andme" file from HGI, which is the GWAS summary statistics for "Hospitalized covid vs. not hospitalized covid", released in round 7 containing 16,512 Hospitalized COVID-19 cases and 71,321 not hospitalized COVID-19 controls.

GWES

For GWES, we focused on 1,770,996 SNPs with MAF > 0.02 to avoid false positives due to rare composite genotypes in the epistasis analysis. The initial screening for interactions was carried out in an exhaustive manner testing potential epistasis between 1.57 trillion pairs of SNPs based on linear regression, $Y \sim \beta_1 SNP_1 + \beta_2 SNP_2 + \beta_3 SNP_1 SNP_2 + covariates$, where Y represents the case–control status but treated as a

continuous variable to reduce the computation burden of logistic regression, and SNPs were coded as the number of minor alleles plus one. We test if β_3 is significantly deviated from zero. We considered $P < 5 \times 10^{-12}$ as our study-wide suggestive significance threshold, and $P < 1 \times 10^{-13}$ as our study-wide significance threshold.

The initial screen was carried out in a heterogeneous supercomputing environment. The analysis was programed in Heterogeneous Interface for Portability (HIP) on GPU and used the Radeon Open Computing platform (ROCm) to call Basic Linear Algebra Subprograms (BLAS) to accelerate the matrix solving process. The computational efficiency was optimized by using Multipoint Interface (MPI) master/slave mode to adaptively split the computing tasks and allocate them to CPUs and GPUs on multiple computing nodes for asynchronous execution. This analysis was conducted using 80 compute nodes, 2560×86 CPU cores and 320 GPU accelerators and was completed in 21.3 h with a double precision computation performance of 15.5 Tera Floating-point Operations Per Second (TFLOPS).

After the initial screen, study-wide suggestively significant interactions were re-examined using logistic regression. The signals remaining study-wide suggestively significant were further passed to the Spanish cohort for replication. We used sex, age, age × age, and the first five principal components of ancestry as covariates in the replication analysis. Then a meta-analysis of UKBB and Spanish association data was carried out for study-wide suggestively significant interactions using the METAL package (http://csg.sph.umich.edu/abecasis/metal/).

Gene ontology (GO) enrichment analysis

Focusing on the 220 pairs of potentially interacting SNPs identified from our GWES, we obtained a set of 363 nearest genes (overlap a 1-kb upstream and downstream region), which was passed to the GO enrichment analysis considering all genes over the genome as the background using the Metascape web service (https://metascape.org). The enriched GO terms were identified by cumulative hypergeometric test and then hierarchically clustered into a tree based on Kappa-statistical similarities among their gene memberships. Then 0.3 kappa score was applied as the threshold to cast the tree into term clusters. For the two SNPs (rs9792388 and rs3025892) with the strongest evidence of epistasis effect, we identified their nearest genes (*PDGFRL*, *SNAP25*) and obtained 20 additional genes showing interactions with these two genes using GeneMANIA (http://genemania.org/). GO enrichment analysis of the 22 genes was conducted considering all interacting genes with GO annotations over the genome as the background. GO terms and q-values for FDR-corrected hypergeometric enrichment tests are reported.

Results

The discovery GWAS included a total of 14,855 SARS-CoV-2 infected British individuals present in the UKBB database updated on 26 June 2021. We considered 2243 patients hospitalized or dead within three months of infection as cases and the remaining 12,612 infected patients with no or mild symptoms as controls (Supplementary material Table S1). The association analysis confirmed the previously well-established locus at 3q22.1, with 32 SNPs in the region surpassing the genome-wide significance (lead SNP rs73064425 near *LZTFL1*, allelic OR for T allele = 1.49, 95%CI 1.31–1.68, $P = 4.27 \times 10^{-10}$; Supplementary material Figure S1 and Table S2). No other locus was genome-wide significant. The lead SNP rs10490770 from COVID-19 HGI was in high LD with rs73064425 ($r^2 = 0.99$) and resulted genome-wide significant also in this study ($P = 5.76 \times 10^{-10}$). Besides 3q22.1, 36 SNPs at three loci (2q13, top hit rs75302730; 6q15, top hit rs4321803; 15q25.1, top hit rs102807) were suggestively significant ($P < 1 \times 10^{-6}$) with largely consistent allele effects as COVID-19 HGI (Supplementary material Figure S1).

The GWES exhaustively tested 1.57 trillion interactions between 1.77 million SNPs (MAF > 2%) in a pair-wise manner and identified 10 study-wide significant interactions between 20 SNPs from 18 distinct loci (> 600kbp) ($P < 1 \times 10^{-13}$), as well as 220 study-wide suggestively significant interactions between 440 SNPs from 182 distinct loci ($P < 5 \times 10^{-12}$) (Fig. 1a and Supplementary material Table S3). Notably, all identified interactions were determined by two SNPs far apart from each other (> 1Mbp) or located on different chromosomes. The most significant interaction ($P = 7.18 \times 10^{-15}$) was found between 7q21.13 rs117337366 (mapping within the *CDK14* gene) and 9p13.3 rs192517877 (located in a genomic region hosting many genes, the closest to rs192517877 being *TLN1*, *CREB3*, *GBA2*, *RGP1*, *MSMP*, *NPR2*, *SPAG8*, *HINT2*, *TMEM8B*, and *FAM221B*, in a range of only 100 kb).

A GO analysis focusing on the 363 genes surrounding the 440 interacting SNPs(< 1kbp) revealed significant functional enrichments ($P < 10^{-4}$) for "modulation of chemical synaptic transmission", "behavior", "regulation of neurotransmitter levels", and "cell–cell adhesion via plasma-membrane adhesion molecules" terms (Fig. 1b), the last term being possibly related to COVID-19[6].

The replication analysis was conducted for the 220 suggestive interactions in a Spanish cohort consisting of 1,416 patients with severe symptoms and 4,382 healthy controls. Eight interactions were replicated at nominal significance ($P < 0.05$) although none survived strict Bonferroni correction ($P < 4 \times 10^{-4}$, Supplementary material Table S3). A

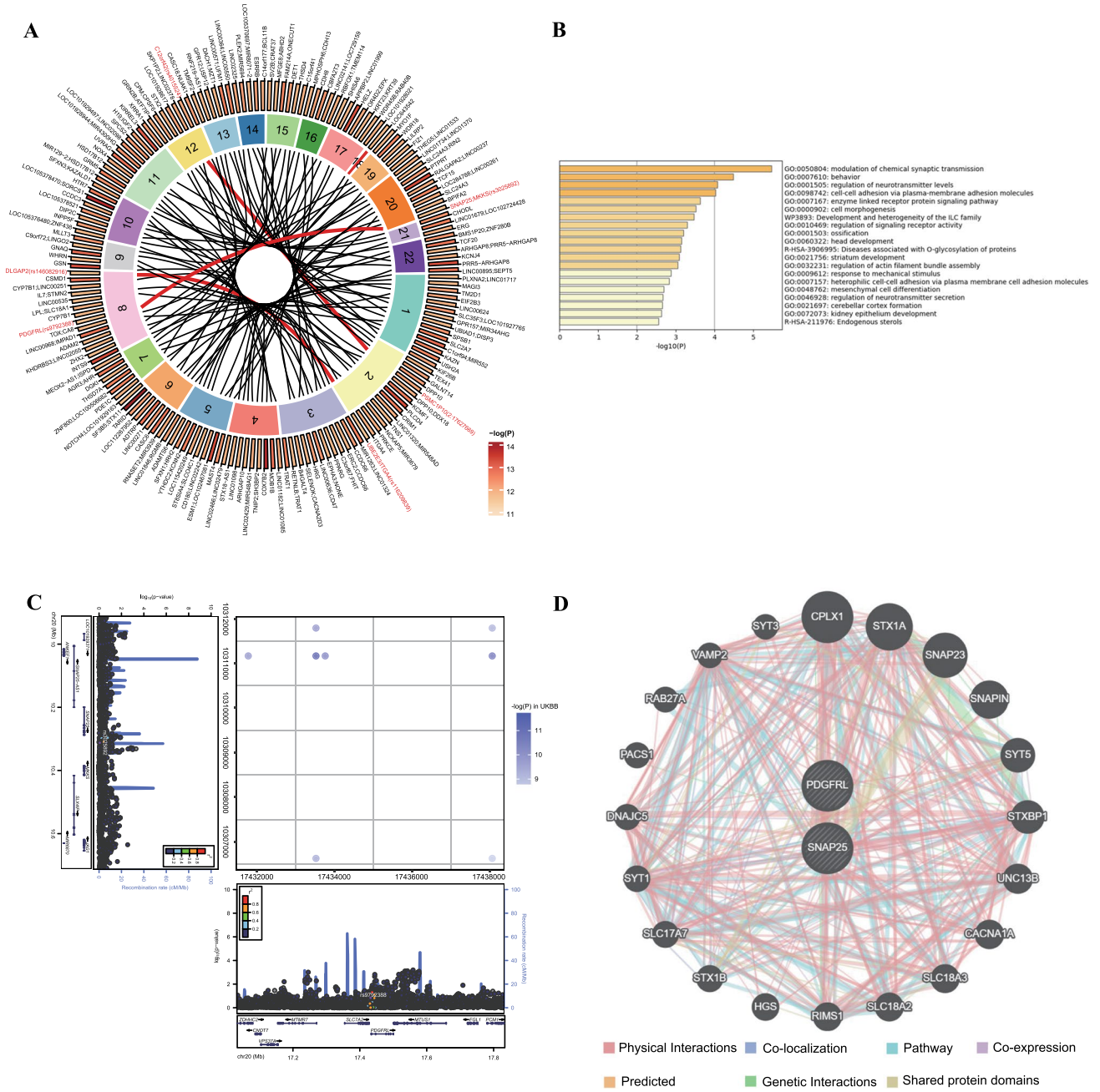

Fig. 1 Genome-wide epistasis study identified 220 pairs of interactions. **A** The redness in the outermost circle denotes the significance of the 220 pairs of SNP interactions ($P<5\times10^{-12}$) and closest genes are annotated outside of the circle. The innermost circle approximates the chromosomal positions where the interacting SNPs are located. Three interactions which showed enhanced significance in the meta-analysis are highlighted in red. **B** Enriched terms from GO analysis of 363 genes surrounding the 220 pairs of interacting SNPs. **C** The regional association patterns between 8p22 rs9792388 and 20p12.2 rs3025892. Both SNPs individually were non-significant (the left and down panels) while the composite genotype of the two SNPs showed highly significant association with the severity of COVID-19 (the middle panel). **D** The functional interaction network of 22 genes from the GeneMANIA analysis of protein interactions between *PDGFRL* and *SNAP25*

meta-analysis of UKBB and Spanish cohorts highlighted 3 interactions with enhanced statistical significance (Fig. 1a). The interaction between 8p22 rs9792388 and 20p12.2 rs3025892 had the most significance in replication analysis and the enhanced significance in the meta-analysis ($P_{GWES}=2.77\times10^{-12}$, $P_{\mathrm{Rep}}=0.003$, $P_{\mathrm{META}}=4.97\times10^{-14}$). Both SNPs individually were not associated with COVID-19 in the discovery dataset (rs9792388 $P=0.14$, rs3025892

$P = 0.85$; Fig. 1c) and had high allele frequencies in UKBB (rs9792388 $f = 0.14$, rs3025892 $f = 0.07$). In UKBB, the composite genotype of rs3025892 CT and rs9792388 CA/AA showed higher risk of severe disease (proportion of severe cases = 0.24 ~ 0.29) than any other genotypes (0.09 ~ 0.18; genotypic OR = 1.96 ~ 2.70; Table 1). This pattern was replicated in the Spanish cohort, i.e., the composite genotype of rs3025892 CT and rs9792388 CA/AA showed higher risk of severe disease (0.30 ~ 0.36) than any other genotypes (0.14 ~ 0.25; genotypic OR = 1.45 ~ 2.37; Table 1). The 8p22 rs9792388 is located upstream of *PDGFRL* which is related to platelet activating factor receptor activity (*PDGFR*) and platelet-derived growth factor beta-receptor (*PDGFRβ*) activity. Previous study has reported *PDGFRβ* as a brain pericyte-specific marker, and the level of soluble PDGFRβ was significantly reduced in the cerebrospinal fluid of SARS-CoV-2 patients involving the nervous system [7]. The other interacting SNP, 20p12.2 rs3025892, is located downstream of *SNAP25*, which is an important regulator involved in neurotransmitter release. Affinity Capture-MS and yeast two-hybrid experiments found that SNAP25 has physical interactions with NSP4, ORF3A, ORF7B of SARS-CoV-2 in the Biological General Repository for Interaction Datasets (BioGRID, https://thebiogrid.org/). *PDGFRL* and *SNAP25* are co-expressed [8, 9] and significantly enriched in neurotransmitter transport with their related genes (*STXBP1*, *SLC18A3*, *SYT1* and *STX1B*, FDR = 1.72×10^{-26}; Fig. 1d). The interaction of *PDGFRL* and *SNAP25* suggests a possible infection pathway of SARS-CoV-2 in brain pericytes.

The second enhanced interaction was found between 2q31.3 rs116209839, located in the intergenic region of *UBE2E3* and *ITGA4*, and 8p23.3 rs146082916, which resides near *DLGAP2* ($P_{GWES} = 3.02 \times 10^{-12}$, $P_{\text{Rep}} = 0.01$, $P_{\text{META}} = 3.62 \times 10^{-13}$; Supplementary material Table S3). The genetic interaction of *DLGAP2* and *UBE2E3* has also been identified by radiation hybrid genotypes [10]. *DLGAP2* encodes a membrane-associated protein important for synapse organization and signaling in neuronal cells. *UBE2E3* is a member of the E2 ubiquitin-conjugating enzyme family, which targets abnormal or short-lived proteins for degradation. Both SNPs individually were not associated with COVID-19 in UKBB ($P > 0.05$). The composite genotype of rs116209839 AG and rs146082916 GA/AA showed a higher risk of severe disease than other genotypes in UKBB (0.42 ~ 0.67 vs. 0.14 ~ 0.24; genotypic OR = 4.25 ~ 11.03; Table 1) as well as in the Spanish cohort (genotypic OR > 1.85; Table 1). The third interaction was between chr2:17,627,669 (Reference allele R: AC, Risk allele D: deletion of C) near *PSMC1P10* and 12q23.3 rs4015524 in *C12orf42* ($P_{GWES} = 8.84 \times 10^{-13}$, $P_{\text{Rep}} = 0.01$, $P_{\text{META}} = 1.91 \times 10^{-13}$). The double heterozygosity for these two SNPs showed a higher risk of severe disease than other genotypes, both in the discovery (genotypic OR = 6.42; Table 1) as well as in the Spanish cohort (genotypic OR = 3.11; Table 1).

These replication signals were generally weaker compared with those observed in traditional GWAS. Although the LD patterns at the significant loci were largely similar between UKBB and the Spanish cohort (Supplementary material Figure S2), minor variations in allele frequencies

Table 1 Risk of severe COVID-19 as a function of composite genotype of interacting SNPs

		UKBB N = 14,855			Spanish N = 5798		
		rs3025892 20p12.2 *SNAP25/MKKS*					
		CC	CT	TT	CC	CT	TT
rs9792388	CC	0.15(9174)	0.12(1425)	0.09(45)	0.25(3676)	0.22(400)	0.22(10)
8p22 *PDGFRL*	CA	0.15(2794)	0.24(442)	0.18(23)	0.24(1317)	0.30(124)	0.14(8)
	AA	0.13(231)	0.29(42)	–(0)	0.25(112)	0.36(14)	–(0)
		rs146082916 8p23.3 *DLGAP2*					
		GG	GA	AA	GG	GA	AA
rs116209839	AA	0.15(12,414)	0.14(622)	0.14(8)	0.25(4684)	0.22(378)	0.40(10)
2q31.3 *UBE2E3/ITGA4*	AG	0.14(1250)	0.42(67)	0.67(3)	0.23(522)	0.35(46)	1.00(1)
	GG	0.24(30)	–(0)	–(0)	0.15(20)	–(0)	–(0)
		rs4015524 12q23.2 *C12orf42*					
		AA	AC	CC	AA	AC	CC
2:17,627,669	RR	0.15(12,787)	0.15(934)	0.05(20)	0.24(5396)	0.23(209)	0.50(2)
2p24.2 *PSMC1P10*	RD	0.13(693)	0.51(46)	–(0)	0.23(248)	0.46(14)	–(0)
	DD	0.20(5)	–(0)	–(0)	–(0)	–(0)	–(0)

Each cell displays the proportion of severe cases among carriers of the composite genotype. Sample size is displayed in brakets

R represents the reference allele of AC and D represents the risk allele where C is deleted

may affect the replication results due to the statistical nature of epistasis analysis, which warrant further investigations in future studies.

Finally, given that the 3q22.1 rs73064425 polymorphism showed a significant genome-wide significant association with COVID-19 severity, we also looked up in our results whether any SNPs across the genome could interact with rs73064425. No significant interactions were found at the genome-wide significance level ($P < 5 \times 10^{-8}$, Supplementary material Figure S3).

Discussion

In this work, we show the very first attempt to systematically screen for genetic interactions in COVID-19, pointing to a list of 220 pairs of SNP interactions potentially affecting the severity of the disease. In addition, we highlight 3 interactions that showed highly consistent and significant patterns in an independent Spanish cohort.

As a general consideration, we obtained strong evidence for brain-related pathological changes in COVID-19 although the exact pathways remain elusive. We mapped a set of gene interactions enriched in brain pathology pathways and highlighted a strong interaction between *PDGFRL* and *SNAP25*, potentially involved in SARS-CoV-2 infection pathways in brain pericytes. These findings add to our understanding of the etiology and provide avenues for the management of COVID-19 severity.

We acknowledge some limitations of our work, mainly associated to the relative strength of epistatic signals evidence in the replication stage. This is likely explained by a combination of genetic factors, such as allelic heterogeneity, variation in allele frequencies, and different LD structures between the analyzed populations. The manifestation of genetic interactions in disease risk may also be diluted due to cumulative noises introduced by intermediate factors at epigenomic, proteomic, metabolomic, and other molecular layers. Systematic and integrated analysis of multi-omics data at the genome-wide scale should be carried out in future studies. To perform robust risk profiling for severe forms of COVID-19, fast, trace and mobile DNA genotyping techniques need to be developed besides the current methods that are commonly employed for DNA testing of the virus.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10654-023-01020-5.

Acknowledgements This research has been conducted using the UK Biobank Resource under Application Number 67076. The authors are grateful for the dedication, commitment and contribution of the participants, the general practitioners, pharmacists, and the staff from the UKBB. The contribution of Banca Intesa Sanpaolo is gratefully acknowledged.

Author contributions Conceptualization: FL; Data analysis: SL, XG, FD, XFR; Data acquisition: SL, XG, FD, YQ, TL, SSH, MRG, JF, AA, MB, LB, AJ, RdC, RA, AF; Funding acquisition: FL, SH; Manuscript writing: SL, RA, FL; Supervision: FL, RA, AF; Manuscript commenting and final approval: all authors.

Funding This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB38010400, XDC01000000), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), National Natural Science Foundation of China (81930056), Science and Technology Service Network Initiative of Chinese Academy of Sciences (KFJ-STS-QYZD-2021–08-001, KFJ-STS-ZDTP-079). This work was also supported by Naif Arab University for Security Sciences (Grant No. PR2-10).

Data availability Summary statistics of the GWES (41 K items $P < 1 \times 10^{-6}$) and GWAS (7.87 million items $P < 5 \times 10^{-8}$) are available from GWAS Atlas (https://ngdc.cncb.ac.cn/gwas/browse/GVP000004).

Declarations

Conflict of interest The authors declare no conflicts of interest.

Ethics approval Human participants: UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (MREC) to obtain and disseminate data and samples from the participants (http://www.ukbiobank.ac.uk/ethics/), and these ethical regulations cover the work in this study. Written informed consent was obtained from all of the participants.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

1. Initiative C-HG. A first update on mapping the human genetic architecture of COVID-19. Nature. 2022;608(7921):E1-e10. https://doi.org/10.1038/s41586-022-04826-7.
2. Karlsen TH. Understanding COVID-19 through genome-wide association studies. Nat Genet. 2022;54(4):368–9. https://doi.org/10.1038/s41588-021-00985-x.
3. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide association study of severe Covid-19 with respiratory failure. N Engl J Med. 2020;383(16):1522–34. https://doi.org/10.1056/NEJMoa2020283.
4. Nakanishi T, Pigazzini S, Degenhardt F, et al. Age-dependent impact of the major common genetic risk factor for COVID-19 on severity and mortality. J Clin Investig. 2021. https://doi.org/10.1172/jci152386.
5. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9. https://doi.org/10.1038/s41586-018-0579-z.
6. Bauer W, Ulke J, Galtung N, et al. Role of cell adhesion molecules for prognosis of disease development of patients with and without COVID-19 in the emergency department. J Infect Dis. 2021;223(8):1497–9. https://doi.org/10.1093/infdis/jiab042.
7. Bocci M, Oudenaarden C, Sàenz-Sardà X, et al. Infection of brain pericytes underlying neuropathology of COVID-19 patients. Int J Mol Sci. 2021;22(21):11622. https://doi.org/10.3390/ijms222111622.

8. Innocenti F, Cooper GM, Stanaway IB, et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 2011;7(5):e1002078. https://doi.org/10.1371/journal.pgen.1002078.
9. Dobbin KK, Beer DG, Meyerson M, et al. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clin Cancer Res Off J Am Assoc Cancer Res. 2005;11(2 Pt 1):565–72. https://doi.org/10.1158/1078-0432.565.11.2.
10. Lin A, Wang RT, Ahn S, Park CC, Smith DJ. A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. Genome Res. 2010;20(8):1122–32. https://doi.org/10.1101/gr.104216.109.

Authors and Affiliations

Shiqi Lin[1,2] · Xingjian Gao[1,3] · Frauke Degenhardt[4] · Yu Qian[1,2] · Tianzi Liu[1] · Xavier Farre Ramon[5] · Syed Sibte Hadi[6] · Manuel Romero-Gómez[7,8,9,10] · Javier Fernández[11,12] · Agustín Albillos[7,13] · Maria Buti Ferret[7,14,15] · Luis Bujanda[7,16] · Antonio Julià[17] · Rafael de Cid[18] · Rosanna Asselta[18,19] · Andre Franke[4] · Fan Liu[1,2,6]

✉ Rosanna Asselta
rosanna.asselta@hunimed.eu

✉ Fan Liu
fliu@nauss.edu.sa

1 CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing, China

2 College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

3 National Clinical Research Center for Kidney Diseases, Jinling Hospital, 305 East Zhongshan Road, Nanjing, China

4 Institute of Clinical Molecular Biology and University Hospital of Schleswig-Holstein, Christian-Albrechts-University, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany

5 Genomes for Life-GCAT Lab., Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain

6 Department of Forensic Sciences, College of Criminal Justice, Naif Arab University of Security Sciences, 11452 Riyadh, Kingdom of Saudi Arabia

7 Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas (CIBEREHD), Instituto de Salud Carlos III (ISCIII), Madrid, Spain

8 Instituto de Biomedicina de Sevilla (IBIS), Seville, Spain

9 University of Sevilla, Seville, Spain

10 Digestive Diseases Unit, Institute of Biomedicine of Seville, Virgen del Rocio University Hospital, University of Seville, Seville, Spain

11 Hospital Clinic, University of Barcelona, and IDIBAPS, Barcelona, Spain

12 European Foundation for the Study of Chronic Liver Failure (EF-CLIF), Barcelona, Spain

13 Department of Gastroenterology, Hospital Universitario Ramón y Cajal, Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), University of Alcalá, Madrid, Spain

14 Liver Unit, Hospital Universitari Vall d'Hebron, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain

15 Universitat Autònoma de Barcelona, Bellatera, Spain

16 Department of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute – Donostia University Hospital, University of the Basque Country (UPV/EHU), CIBERehd, Ikerbasque, San Sebastian, Spain

17 Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Barcelona, Spain

18 Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20072 Pieve Emanuele, Milan, Italy

19 Humanitas Clinical and Research Center- IRCCS, Via Manzoni 56, 20089 Rozzano, Milan, Italy