nature immunology

MATTERS ARISING

https://doi.org/10.1038/s41590-022-01228-9

Check for updates

Reply to: Hultström et al., Genetic determinants of mannose-binding lectin activity predispose to thromboembolic complications in critical COVID-19. Mannose-binding lectin genetics in COVID-19

Rosanna Asselta[1,2,17], Elvezia Maria Paraboschi[1,2,17], Matteo Stravalaci[1,2], Pietro Invernizzi[3,4], Paolo Bonfanti[5], Andrea Biondi[6], Isabel Pagani[7], Mattia Pedotti[8], Andrea Doni[1], Francesco Scavello[1], Sarah N. Mapelli[1], Marina Sironi[1], Chiara Perucchini[1], Luca Varani[8], Milos Matkovic[8], Andrea Cavalli[8,9], Daniela Cesana[10], Pierangela Gallina[10], Nicoletta Pedemonte[11], Valeria Capurro[11], Nicola Clementi[12], Nicasio Mancini[12], Rafael Bayarri-Olmos[13], Peter Garred[13], Rino Rappuoli[14,15], Stefano Duga[1,2], Barbara Bottazzi[1], Mariagrazia Uguccioni[8,2], Elisa Vicenzi[7], Alberto Mantovani[1,2,16] ✉ and Cecilia Garlanda[1,2] ✉

REPLYING TO M. Hultström et al. *Nature Immunology* https://doi.org/10.1038/s41590-022-01227-w (2022)

Prompted by our report on the role of mannose-binding lectin (MBL) in resistance to COVID-19 (ref. [1]), Hultström and colleagues[2] conducted a genetic and biochemical analysis of this fluid-phase pattern recognition molecule in 426 patients of the SweCovid Swedish initiative and in data extracted from summary statistics of the COVID-19 Host Genetics Initiative (HGI)[3]. Our study had reported that MBL binds to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike protein from variants of concern and inhibits the virus, and that genetic variants mapping in the MBL gene (*MBL2*) region predispose to severe COVID-19 (ref. [1]). In apparent contrast to our genetic study, Hultström and colleagues did not find a significant association between *MBL2* single-nucleotide polymorphisms (SNPs) and hospitalization or intensive care admission due to COVID-19 (data extracted from summary statistics of the COVID-19 HGI[3]). They found that *MBL2* haplotypes, composed of functional variants mapping within the gene (alleles named C, B, D, X/Y and L/H, according to the legacy nomenclature[4]) had a dual, U-type, impact on the risk for thrombotic complications in critically ill COVID-19 patients.

Hultström and colleagues[2] did not find any association between severe COVID-19 and D, B and C alleles (rs5030737-A, rs1800450-T and rs1800451-T, all inducing a lowering effect on MBL levels) or other SNPs in the *MBL2* region. Indeed, we had also found borderline associations in the single-variant analysis (none surviving the correction for multiple testing, as declared). However, a significant predisposing effect was observed in those individuals carrying two disruptive alleles among rs5030737, rs1800450 and rs1800451. In addition, a significant protective effect was observed for the haplotype lacking the predisposing rs5030737 allele.

The best association signals stemmed from haplotype analysis performed on the entire *MBL2* region (500 kb upstream and downstream of the gene)[1]. Those haplotypes are in proximity to *MBL2* regulatory regions, do not include the functional variants mapping within the gene and show a highly significant *P* value of association (accounting for multiple testing correction)[1]. In this frame, it would have been interesting to see a comparison with similar analyses on the Swedish cohort described by Hultström et al.[2]. The latter study did not include a case–control single-SNP association analysis on their patients and only results based on the HGI data were reported. This choice was probably based on the advantage of using the huge statistical power given by the HGI summary statistics (obtained from the analysis of at least 8,800 cases and 1 million controls). However,

[1]IRCCS Humanitas Research Hospital, Milan, Italy. [2]Department of Biomedical Sciences, Humanitas University, Milan, Italy. [3]Division of Gastroenterology, Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy. [4]European Reference Network on Hepatological Diseases, San Gerardo Hospital, Monza, Italy. [5]Department of Infectious Diseases, San Gerardo Hospital, University of Milano-Bicocca, Monza, Italy. [6]Pediatric Department and Centro Tettamanti-European Reference Network PaedCan, EuroBloodNet, MetabERN-University of Milano-Bicocca-Fondazione MBBM-San Gerardo Hospital, Monza, Italy. [7]Viral Pathogenesis and Biosafety Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [8]Institute for Research in Biomedicine, Università della Svizzera italiana, Bellinzona, Switzerland. [9]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [10]San Raffaele Telethon Institute for Gene Therapy, IRCCS, San Raffaele Scientific Institute, Milan, Italy. [11]UOC Genetica Medica, IRCCS Istituto Giannina Gaslini, Genoa, Italy. [12]Laboratory of Microbiology and Virology, IRCCS Scientific Institute and Vita-Salute San Raffaele University, Milan, Italy. [13]Laboratory of Molecular Medicine, Department of Clinical Immunology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [14]Monoclonal Antibody Discovery Lab, Fondazione Toscana Life Sciences, Siena, Italy. [15]Faculty of Medicine, Imperial College London, London, UK. [16]The William Harvey Research Institute, Queen Mary University of London, London, UK. [17]These authors contributed equally: Rosanna Asselta, Elvezia Maria Paraboschi. ✉e-mail: alberto.mantovani@humanitasresearch.it; cecilia.garlanda@humanitasresearch.it

Table 1 | Population frequencies for functional coding variants in the *MBL2* gene

rs ID	rs1800451	rs1800450	rs5030737
Position and alleles	chr10:52771466:C:T	chr10:52771475:C:T	chr10:52771482:G:A
Legacy name	rs1800451-T = C	rs1800450-T = B	rs5030737-A = D
Effect	p.Gly57Glu	p.Gly54Asp	p.Arg52Cys
Population			
European (non-Finnish)	0.01831	0.1445	0.07350
Swedish	0.01814	0.1470	0.08377
North-western European	0.01848	0.1399	0.07234
Southern European	0.01931	0.1511	0.06861
Bulgarian	0.01386	0.1551	0.06704
Estonian	0.01515	0.1532	0.05968
European (Finnish)	0.008323	0.1366	0.05773
Ashkenazi Jewish	0.02451	0.1343	0.1024
African/African American	0.2302	0.03188	0.01122
Latino/Admixed American	0.01778	0.1688	0.02579
East Asian	0.0001002	0.1691	0.0002506
Japanese	0	0.1974	0
Korean	0	0.2026	0
South Asian	0.03290	0.1401	0.06232

Population frequencies for the B, C and D *MBL2* variants are available on the GnomAD database (https://gnomad.broadinstitute.org; database v.2.1.1; data on 125,748 exomes and 15,708 genomes).

this approach suffers from limitations: (1) haplotype-based analyses cannot be performed using the summary statistics; and (2) it is not possible to finely dissect signals coming from different populations. This is not trivial especially for those SNPs showing minor allele frequencies (MAFs) that differ enormously across ethnic groups. Table 1 summarizes MAFs in different populations for the three functional SNPs of the *MBL2* gene: allele frequencies vary considerably across populations, with SNPs rs1800451 and rs5030737 even being monomorphic in east Asians. Due to standard quality check steps adopted in preparing datasets for association analysis (aimed at discarding monomorphic SNPs or SNPs with extremely low MAFs)[3], it is possible that the SNPs rs1800451 and rs5030737 were not even analyzed in some HGI populations.

Prompted by the study of Hultström and colleagues[2], we repeated association/haplotype analyses in a larger case–control Italian cohort, including the original 332 cases and 1,668 controls (described in ref. [1]) plus 195 cases and 1,522 controls. As in our original study[1], newly enrolled cases were recruited at the Humanitas Clinical and Research Center IRCCS (in Rozzano, Milan, Italy) and San Gerardo Hospital (in Monza, Italy), and were defined as severe COVID-19 cases, all with respiratory failure requiring hospitalization and a confirmed SARS-CoV-2 viral RNA PCR test. Overall, our 527 severe cases belong to the first SARS-CoV-2 wave (March–May 2020), and are therefore extremely homogeneous with regard to their phenotype, treatments, absence of vaccination and viral strain infection. New controls were, again, from the general Italian population with unknown COVID-19 status. We essentially confirmed all our previous results: single-SNP association analysis only revealed borderline associations not surviving the multiple testing correction (data not shown), whereas multiallelic analysis, investigating the previously identified haplotypes mapping in the *MBL2* region[1], evidenced significant associations with severe COVID-19 (Table 2). Considering that the identified haplotypes could be, at least in part, a reflection of single-marker association signals in their proximity, we extracted from the HGI repository, v.6 (not yet available at the time of ref. [1]) all SNPs showing a *P* value of association $<5\times10^{-3}$. Extended Data Fig. 1 reports on these signals, considering three of the analyses performed in the frame of the HGI project: A2 (defined as 'Very severe respiratory confirmed COVID-19 versus population'), B2 ('Hospitalized COVID-19 versus population') and C2 ('COVID-19 versus population'). We are well aware that all these hits are above the genome-wide threshold for significance. Nevertheless, we observed clusters of SNPs corresponding exactly to the haplotypes we identified (Table 2). More interestingly, we noticed that the stronger signals of associations ($10^{-4}<P<10^{-5}$; top signal rs1877134, $P=3.8\times10^{-5}$) correspond to the A2 analysis and map in a genomic region of 18 kb, exactly encompassing the *MBL2* gene. Although suggestive, and given the lack of association signals at the genome-wide level in the COVID-19 HGI results, we should also consider the possibility that the *MBL2* locus does not have significantly consistent COVID-19 risk in every population analyzed. A possible explanation for discrepancies in effect sizes could be that host genetics risk may not be identical across the analyzed cohorts, and between the present study and that of Hultström et al.[2] or more in general among populations. Virus strains, different pandemic waves, vaccinations and medical care strategies and drugs could also have played a role.

Hultström and colleagues[2] report the association between haplotypes determining activity of MBL and the risk for thrombotic complications in severe COVID-19 patients. In particular, genetically determined MBL activity levels were shown to confer risk for pulmonary embolism in a U-shaped fashion (that is, haplotypes LXA/LXA, HYA/0 and LYA/0—associated with intermediate MBL activity—are indeed protective). These results are based on a restricted number of cases (4 of 123 individuals bearing 'intermediate' MBL activity haplotype versus 33 of 231 and 8 of 72 carrying 'high' and 'deficient' haplotypes, respectively). In an effort to replicate these findings, we stratified our 207 severe COVID-19 patients, for whom data on presence/absence of thromboembolic events during hospitalization were available, on the basis of the same MBL activity haplotypes ('high': HYA/HYA, LYA/LYA, HYA/LYA, HYA/LXA and LYA/LXA; 'intermediate': LXA/LXA, HYA/0 and

Table 2 | Locus-wide haplotype analysis

Haplotype	Frequency in cases	Frequency in controls	OR	CI	P[a]	SNPs[b]
ATCGCAA	0.006	0.025	0.21	0.093–0.48	0.0012	6 SNPs, rs11344513|rs7071467
CCC	0.005	0.041	0.11	0.046–0.27	2.96×10^{-5}	3 SNPs, rs17662822|rs1159798|rs1912619
TCCCC	0.0019	0.014	0.13	0.032–0.53	0.011	5 SNPs, rs2204344|rs12218074|rs80035245|rs7935712|rs10824836
TA	0.12	0.087	1.43	1.16–1.75	0.00025	2 SNPs, rs10824844|rs10824845
ATCCCCGCATTGA	0.000	0.021	<1.00	nc	0.059	9 SNPs, rs57504125|chr10:5308418:G:A
AGATCCCCGCGCGTGCAACGGCTGCGGA	0.22	0.18	1.29	1.097–1.51	0.0026	24 SNPs, rs71032688|rs7092597

[a]*P* values are presented as noncorrected for multiple testing; threshold for significance taking into account multiple testing (six tests) is $P=0.0083$; *P* values are corrected for sex, age, sex × age, age × age and the first ten PCs of ancestry. [b]The number of SNPs composing the haplotype is indicated. All the SNPs forming the haplotype are shown for short haplotypes (including a maximum of five SNPs). For more complex haplotypes (including >5 SNPs) only the first and the last SNPs are indicated.

LYA/0; and 'deficient': LXA/0 and 0/0). This analysis did not show any association with thromboembolic events (χ^2 test, $P=0.36$). Indeed, we observed a higher frequency of thromboembolic events in individuals carrying the 'intermediate' haplotype (9 of 59 individuals bearing 'intermediate' MBL activity haplotype versus 11 of 118 and 2 of 30 carrying 'high' and 'deficient' haplotypes, respectively). In the regression model, correcting for covariates, we confirmed these negative results.

Finally, we stratified our entire cohort of 3,717 individuals for these high/intermediate/deficient haplotypes, and used this marker in an association analysis for severity: again, no association signal emerged ($P=0.33$). However, we observed a moderate predisposing effect toward severe COVID-19 for those individuals carrying two disruptive 0 alleles ($P=0.067$ in the regression model accounting for covariates).

MBL recognition of SARS-CoV-2 spike protein was found to trigger complement activation via the lectin pathway[1]. However, the antiviral activity of MBL observed in vitro with three different cellular models occurred in serum-free conditions, and thus were irrespective of complement activity[1]. Molecular modeling indicates the MBL-binding site spans across the S1 and S2 regions of SARS-CoV-2 spike protein, and suggests a fusion neutralization mechanism. In addition, inhibition of virus entry into host cells might depend on competition with C-type lectins, which act as entry receptors or co-receptors[5]. The mechanism responsible for the antiviral activity of MBL remains to be fully defined, but it is clearly complement independent. Functional assays aimed at investigating the relevance of the three structural *MBL2* polymorphisms in the study by Hultström and colleagues[2] are based on complement activation. The relevance of rs5030737-associated protein in MBL anti-SARS-CoV-2 activity cannot be investigated using this readout, because antiviral activity is complement independent.

In conclusion, the apparent discrepancy between the two reports may well be accounted for by the breadth of the genetic analysis, substantial differences in the frequency of MBL genotypes in different human populations (Table 1) and other factors that usually make genetic association studies during a global pandemic extremely challenging (new emerging virus strains, different pandemic waves, vaccinations and different medical care approaches). The possibility that MBL-triggered complement activation may amplify advanced tissue-damaged disease, including thromboembolic complications, is biologically plausible and deserves further studies.

Online content

Any methods, additional references, Nature Research reporting summaries, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-022-01228-9.

Received: 9 March 2022; Accepted: 26 April 2022;
Published online: 27 May 2022

References

1. Stravalaci, M. et al. Recognition and inhibition of SARS-CoV-2 by humoral innate immunity pattern recognition molecules. *Nat. Immunol.* **23**, 275–286 (2022).
2. Hultström, M. et al. Genetic determinants of mannose-binding lectin activity predispose to thromboembolic complications in critical COVID-19. *Nat. Immunol.* https://doi.org/10.1038/s41590-022-01227-w (2022).
3. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
4. Madsen, H. O. et al. Interplay between promoter and structural gene variants control basal serum level of mannan-binding protein. *J. Immunol.* **155**, 3013–3020 (1995).
5. Lempp, F. A. et al. Lectins enhance SARS-CoV-2 infection and influence neutralizing antibodies. *Nature* **598**, 342–347 (2021).

Methods

Patient cohorts and ethical approvals. Approvals for the project were obtained from the relevant ethics committees (for Humanitas hospitals, reference no. 316/20, and for the University of Milano–Bicocca School of Medicine, San Gerardo Hospital, reference no. 84/2020). The requirement for informed consent was waived.

For genetic association analyses, we investigated a total of 3,717 individuals. These included: (1) 527 patients with severe COVID-19, which was defined as hospitalization with respiratory failure and a confirmed SARS-CoV-2 viral RNA PCR test from nasopharyngeal swabs; patients were recruited from intensive care units and general wards at three hospitals in Lombardy, that is, the Humanitas Clinical and Research Center, IRCCS, in Rozzano, Milan, Italy (180 patients), the San Gerardo Hospital, in Monza, Italy (320 patients) and the Humanitas Gavazzeni hospital, in Bergamo, Italy (27 patients); and (2) 3,190 controls from the general Italian population with unknown COVID-19 status.

Data on thromboembolic events were available for 207 cases (29% females and 71% males; mean age: 68.9 ± 12.3 years).

Genotyping and imputation. Details on DNA extraction, array genotyping and quality checks are reported elsewhere[6,7].

Genetic coverage was increased by performing SNP imputation, as described[1]. In the post-imputation steps, we retained only those SNPs with $R^2 \geq 0.6$ and MAF ≥ 1%. Next, we accurately checked cases and controls for solving within-Italian relationships and for testing the possible existence of population stratification within and across batches: to this aim, we performed principal component analysis (PCA), using a linkage disequilibrium-pruned subset of SNPs across chromosome 10 and the Plink v.1.9 package[8]. The final set of analyzed variants comprised 3,452 SNPs, distributed in the *MBL2* region (the gene ± 500 kb).

Statistical analysis. Case–control allele–dose association tests were performed using the PLINK v.1.9 logistic-regression framework for dosage data. Age, sex, age × age, sex × age and the first ten PCs from PCA were introduced in the model as covariates. Analyses were conducted always referring to the minor allele. All *P* values are accompanied by odds ratio (OR) and 95% confidence interval (CI).

Haplotype analysis and reconstruction of functional haplotypes was performed using PLINK v.1.07 (ref. [9]).

The statistical analysis for association of functional haplotypes (high, intermediate and deficient) with severe COVID-19 or thromboembolic complications was performed using a binomial glm model in R (https://www.r-project.org) with the following covariates: age, sex, age × age, sex × age and ten PCs as already calculated for previous analyses.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The dataset for 332 COVID-19 patients is publicly available at the European Bioinformatics Institute (www.ebi.ac.uk/gwas) under accession nos. GCST90000255 and GCST90000256 (ref. [6]), whereas the dataset for 1,668 healthy individuals of the general population is deposited in the Genotypes and Phenotypes database (https://www.ncbi.nlm.nih.gov/gap) under accession no. phs000294.v1.p1 (ref. [7]). As for the remaining 195 patients and 1,522 controls, data are available from the corresponding author on reasonable request, until their publication in a public repository (pending the acceptance of an unrelated manuscript)[10]. Data related to the analyzed *MBL2* locus are available at (https://doi.org/10.5281/zenodo.6452010).

References

6. Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
7. Myocardial Infarction Genetics Consortium et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2009).
8. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
9. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
10. Degenhardt, F. et al. Detailed stratified GWAS analysis for severe COVID-19 in four European populations. Preprint at *MedRxiv*, https://doi.org/10.1101/2021.1107.1121.21260624 (2022).

Acknowledgements

This work was conducted in the framework of, and made possible by, the collective effort of the Humanitas COVID-19 Task Force, the Humanitas Gavazzeni COVID-19 Task Force and the COVID-19 Storm trial. This paper is dedicated to S. Duga, who passed on 10 November 2021 and had made a key contribution to the genetic section of this report. This work was supported by a philanthropic donation by Dolce & Gabbana fashion house (to A.M. and C.G.), by the Italian Ministry of Health for COVID-19 (grant no. COVID-2020-12371640 to A.M. and C.G.), by the Italian Ministry of University and Research (to P.I.), by the Department of Excellence project PREMIA (PREcision MedIcine Approach: bringing biomarker research to the clinic, to P.I.) and by Fondazione Cariplo (Biobanking of COVID-19 patient samples to support national and international research, to A.B.). We also thank the Banca Intesa San Paolo for their generous contribution (to R.A.) and AMAF Monza ONLUS and AIRCS for unrestricted research funding. L.V. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101003650 and from the Swiss National Science Foundation (grant no. 31003A_182270).

Author contributions

A.M., C.G. and R.A. conceived this extension of the original study reported in ref. [1]. M.S. conducted the experimental work related to binding and complement activation. The genetic analysis was conducted by E.M.P. R.A. supervised the analysis. P.I., P.B. and A.B. provided samples for the genetic analysis. All authors contributed to project design and planning, data analysis and interpretation.

Competing interests

A.M., C.G. and B.B. are inventors of a patent (EP20182181) on PTX3 and obtain royalties on related reagents. A.M., C.G., B.B. and E.V. are inventors of two patents (102021000002738 and EP21214373.9) on MBL. R.R. is a full-time employee of the GSK group of companies. The other authors declare no competing interests.

Additional information

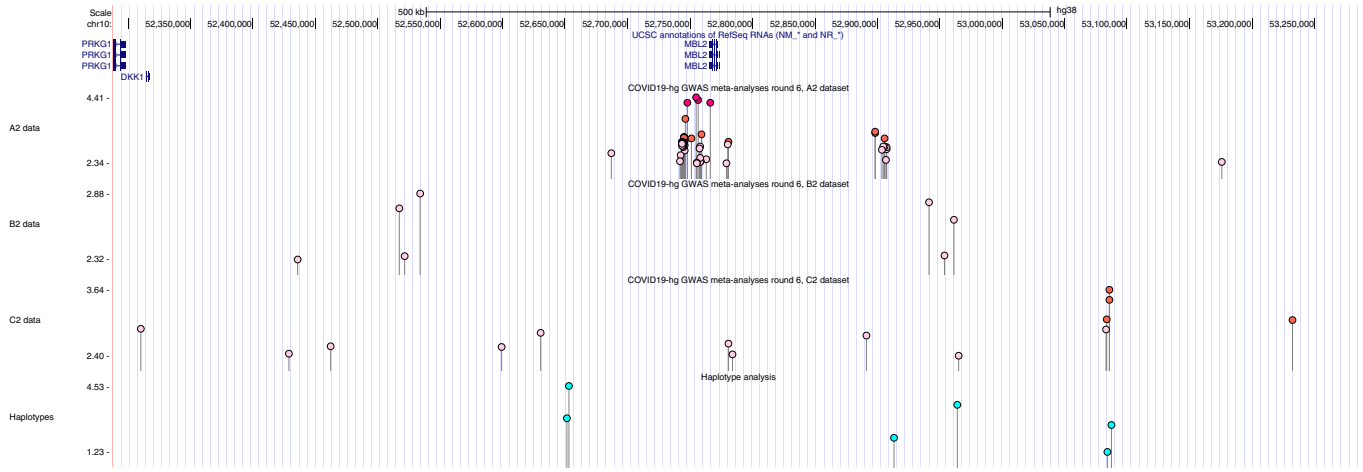
Extended data Extended data are available for this paper at https://doi.org/10.1038/s41590-022-01228-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-022-01228-9.

Correspondence and requests for materials should be addressed to Alberto Mantovani or Cecilia Garlanda.

Peer review information *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | The *MBL2* locus: structure and main association signals with severe COVID-19. A screenshot from the UCSC Genome browser (http://genome.ucsc.edu/; release Dec. 2013, GRCh38/hg38) specifically highlighting the 1-Mb region surrounding the *MBL2* gene is shown. The panel reports, in order, the following tracks: i) the ruler with the scale at the genomic level; ii) chromosome 10 nucleotide numbering; iii) the UCSC RefSeq track; iv) COVID-19 risk variants from the COVID-19 HGI GWAS Analysis A2 (8,779 cases, 25 studies, Release 6: June 2021); v) COVID-19 risk variants from the COVID-19 HGI GWAS Analysis B2 (24,274 cases, 43 studies, Release 6: June 2021); vi) COVID-19 risk variants from the COVID-19 HGI GWAS Analysis C2 (112,612 cases, 74 studies, Release 6: June 2021); vii) COVID-19 risk haplotypes, marked by the tagging SNP, from our larger study (lollipops show all haplotypes reported in Table 2). All panels reporting COVID-19 risk variants from the COVID-19 HGI GWAS analysis show only signals at $P < 5 \times 10^{-3}$; the lollipop height and color depends on the -log10(P): dark pink indicates $P < 10^{-4}$, orange indicates $10^{-4} < P < 10^{-3}$, light pink indicates $10^{-3} < P < 5 \times 10^{-3}$.

nature portfolio

Corresponding author(s): DBPR- NI-A32363A

Last updated by author(s): April 12, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	no software was used.
Data analysis	For genetic studies, case-control allele-dose association tests were performed using plink (version 1.07 and version 1.9). Haplotype analysis was performed using beagle (version 3.3 and version 5.1) and R (version 4.5). All of them are freely available. Statistical analysis R software (version 4.5, https://www.r-project.org/) was used for the statistical analyses.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The dataset for 332 COVID-19 patients is publicly available at the European Bioinformatics Institute (www.ebi.ac.uk/gwas) under accession numbers GCST90000255 and GCST90000256 6, whereas the dataset for 1668 healthy individuals of the general population is deposited in the Genotypes and Phenotypes database (https://www.ncbi.nlm.nih.gov/gap/) under the phs000294.v1.p1 accession code 7. As for the remaining 195 patients and 1522 controls, data are available from the

corresponding author on reasonable request, until their publication in public repository (pending the acceptance of an unrelated manuscript) 8. Data related to the analyzed MBL2 locus are available at (https://doi.org/10.5281/zenodo.6452010).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available data were used and no calculation was performed on sample-size.
Data exclusions	n.a.
Replication	n.a.
Randomization	n.a.
Blinding	n.a.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	For genetic association analyses, we investigated a total of 3717 individuals. These included: i) 527 patients with severe COVID-19, which was defined as hospitalization with respiratory failure and a confirmed SARS-CoV-2 viral RNA PCR test from nasopharyngeal swabs. Data on thromboembolic events were available for 207 cases (29% females and 71% males; mean age: 68.9±12.3).
Recruitment	Patients were recruited from intensive care units and general wards at three hospitals in Lombardy, i.e., the Humanitas Clinical and Research Center, IRCCS, in Rozzano, Italy (180 patients); the San Gerardo Hospital, in Monza, Italy (320 patients); and the Humanitas Gavazzeni hospital, in Bergamo, Italy (27 patients); ii) 3190 controls from the general Italian population with unknown COVID-19 status.
Ethics oversight	Approvals for the project were obtained from the relevant ethics committees (for Humanitas hospitals, reference number 316/20; for the University of Milano-Bicocca School of Medicine, San Gerardo Hospital, reference number 84/2020).

Note that full information on the approval of the study protocol must also be provided in the manuscript.