

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Supplementary methods

32

33 **Prioritization of candidate causal genes for eight identified genomic loci**

34 To further explore the causal effects of nine index SNPs in eight genomic loci on severe
 35 COVID-19, we performed a prioritization analysis by leveraging the Open Target Genetics (OTG,
 36 <https://genetics.opentargets.org/>) tool [1]. The OTG tool is a web-access integrative resource that
 37 aggregates human GWAS summary statistics, including UK Biobank data and NHGRI-EBI GWAS
 38 catalog, and functional genomic data, including protein quantitative trait loci data (pQTLs),
 39 expression QTL (eQTL) from eQTL Catalogue, eQTLGen and GTEx, and promoter capture hiC
 40 (PCHI-C) from 27 different cell types, and DNase I hypersensitive site (DHS)-gene promoter
 41 correlation and other datasets from a wide range of cell types and tissues. Based on the
 42 Variant2Gene (V2G) algorithm, for the index SNP rs2166172 in 1p22.2, we prioritized *BARHL2*
 43 as a candidate causal gene for severe COVID-19. With the highest V2G score, *CCHCR1* was
 44 highlighted as candidate causal gene for the index SNP rs143334143 in 6p21.33, *VSTM2A* as
 45 candidate causal gene for rs622568 in 7p11.2, *OAS1* as candidate causal gene for rs2269899 in
 46 12q24.13, *DPP9* as candidate causal gene for rs2109069 in 19p13.3, and *IFNAR2* as candidate
 47 causal gene for rs13050728 in 21q22.11. The index SNP of rs13050728 ($P = 1.91 \times 10^{-11}$) in
 48 21q22.11 is highly LD with the reported SNP of rs9976829 ($R^2 = 1$) [2] and rs2236757 ($R^2 = 0.8266$)
 49 [3]. The index SNP of rs2269899 in 12q24.13 ($P = 3.24 \times 10^{-8}$) shows high LD with the reported
 50 SNP of rs10735079 ($R^2 = 0.9282$) [3].

51

52 **Combination of GWAS summary statistics with scRNA-seq data (dataset #1) by using the** 53 **MAGMA algorithm**

54 As an independent technical approach to validate genetics-related peripheral immune cells in
 55 PBMCs implicated in severe COVID-19 identified from the RolyPoly algorithm, we applied a
 56 generalized linear regression (GLR) model in the MAGMA (v1.06,
 57 <https://ctg.cncr.nl/software/magma>) [4, 5]. Conditioning on the 10% most specific genes from the
 58 single cell data for each cell type, we examined gene-level genetic associations of severe COVID-
 59 19-associated immune cell types by adding these variables as covariates for the linear regression
 60 model [5-7]. The specificity of each gene for each cell type was calculated by using the mean gene
 61 expression of cell type divided the sum of mean gene expression across all cell types. A window of
 62 50 kb upstream to 50 kb downstream for a given gene coordinates was set to calculate gene-level

63 association statistics. The 1,000 Genome Project European Phase 3 panel [8] was used as the
64 reference panel for both methods. We restricted the GLR analysis to SNPs in the autosomes, and
65 any SNPs with minor allele frequency < 5% were excluded. The major histocompatibility complex
66 region (Chr6: 25-35 Mbp) was also excluded due to the extensive linkage disequilibrium (LD) in
67 this region.

68

69 **Using the Cell-ID method to assess the enrichment in specific gene signatures of individual** 70 **cells in single cell RNA-seq dataset on severe COVID-19**

71 To deeply investigate the which individual cells in the scRNA-seq dataset are enriched in the
72 gene lists associated to the GWAS hits (N = 34 genes, Additional file 3: Figure S13B), as well as
73 functional terms (Counts > 10 genes as the Cell-ID default) of inflammatory and cytokine genes (N
74 = 324 genes, Additional file 2: Table S10), cytokine-cytokine receptor interactions (N = 294 genes),
75 chemokine signaling pathway (N = 189 genes), T cell activation (GO: 0042110), response to
76 interferon alpha (GO: 0035455), response to interferon beta (GO: 0035456), leukocyte migration
77 (GO: 0050900) we re-performed the analysis by using the Cell-ID tool [9], a clustering-free
78 multivariate statistical method for the robust extraction of per-cell gene signatures from single cell
79 sequencing data. The Cell-ID conducts a dimensionality reduction of gene expression matrix
80 through multiple correspondence analysis (MCA), a statistical algorithm that offers a simultaneous
81 representation of observation (e.g., cells) and variables (e.g., genes) in low-dimensional space [10,
82 11]. Both genes and cells are projected in a common orthogonal space. In such space, a gene is
83 closer to a cell, suggesting the gene is more specific to the cell.

84

85 **References**

- 86 1. Ghousaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, Fumis L,
87 Miranda A, Carvalho-Silva D, Buniello A, et al: **Open Targets Genetics: systematic**
88 **identification of trait-associated genes using large-scale genetics and functional**
89 **genomics**. *Nucleic Acids Res* 2021, **49**:D1311-d1320.
- 90 2. Ma Y, Huang Y, Zhao S, Yao Y, Zhang Y, Qu J, Wu N, Su J: **Integrative Genomics**
91 **Analysis Reveals a 21q22.11 Locus Contributing Risk to COVID-19**. *Hum Mol Genet*
92 2021.
- 93 3. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S,
94 Parkinson N, Fourman MH, Russell CD, et al: **Genetic mechanisms of critical illness in**
95 **COVID-19**. *Nature* 2021, **591**:92-98.
- 96 4. de Leeuw CA, Mooij JM, Heskes T, Posthuma D: **MAGMA: generalized gene-set**
97 **analysis of GWAS data**. *PLoS Comput Biol* 2015, **11**:e1004219.

- 98 5. Bryois J, Skene NG, Hansen TF, Kogelman LJA, Watson HJ, Liu Z, Brueggeman L,
99 Breen G, Bulik CM, Arenas E, et al: **Genetic identification of cell types underlying**
100 **brain complex traits yields insights into the etiology of Parkinson's disease.** *Nat Genet*
101 2020, **52**:482-493.
- 102 6. Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, Giusti-Rodriguez P,
103 Hodge RD, Miller JA, Muñoz-Manchado AB, et al: **Genetic identification of brain cell**
104 **types underlying schizophrenia.** *Nat Genet* 2018, **50**:825-833.
- 105 7. Sheng X, Guan Y, Ma Z, Wu J, Liu H, Qiu C, Vitale S, Miao Z, Seasock MJ, Palmer M,
106 et al: **Mapping the genetic architecture of human traits to cell types in the kidney**
107 **identifies mechanisms of disease and potential treatments.** *Nat Genet* 2021, **53**:1322-
108 1333.
- 109 8. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL,
110 McCarthy S, McVean GA, Abecasis GR: **A global reference for human genetic**
111 **variation.** *Nature* 2015, **526**:68-74.
- 112 9. Cortal A, Martignetti L, Six E, Rausell A: **Gene signature extraction and cell identity**
113 **recognition at the single-cell level with Cell-ID.** *Nat Biotechnol* 2021, **39**:1095-1102.
- 114 10. Aşan Z, Greenacre M: **Biplots of fuzzy coded data.** *Fuzzy sets and Systems* 2011, **183**:57-
115 71.
- 116 11. Rausell A, Juan D, Pazos F, Valencia A: **Protein interactions and ligand binding: from**
117 **protein subfamilies to functional specificity.** *Proceedings of the National Academy of*
118 *Sciences* 2010, **107**:1995-2000.
- 119