

## GWAS and meta-analysis identifies multiple new genetic mechanisms underlying severe Covid-19.

Erola Pairo-Castineira<sup>‡,1,2</sup>, Konrad Rawlik<sup>‡,1</sup>, Lucija Klaric<sup>2</sup>, Athanasios Kousathanas<sup>3</sup>, Anne Richmond<sup>2</sup>, Jonathan Millar<sup>1</sup>, Clark D Russell<sup>1,4</sup>, Tomas Malinauskas<sup>5</sup>, Ryan Thwaites, Alex Stuckey<sup>3</sup>, Christopher A Odhams<sup>3</sup>, Susan Walker<sup>3</sup>, Fiona Griffiths<sup>1</sup>, Wilna Oosthuyzen<sup>1</sup>, Kirstie Morrice<sup>6</sup>, Sean Keating<sup>7</sup>, Alistair Nichol<sup>8</sup>, Malcolm G Semple<sup>9,10</sup>, Julian Knight<sup>5</sup>, Manu Shankar-Hari<sup>4</sup>, Charlotte Summers<sup>11</sup>, Charles Hinds<sup>12</sup>, Peter Horby<sup>13</sup>, Lowell Ling<sup>14</sup>, Danny McAuley<sup>15,16</sup>, Hugh Montgomery<sup>17</sup>, Peter J.M. Openshaw<sup>18,19</sup>, Timothy Walsh<sup>7</sup>, Albert Tenesa<sup>1,2,20</sup>, GenOMICC Investigators<sup>\*</sup>, SCOURGE Consortium<sup>\*</sup>, ISARIC4C Investigators<sup>\*</sup>, 23andMe<sup>\*</sup>, Richard H Scott<sup>3,21</sup>, Mark J Caulfield<sup>3,22</sup>, Loukas Moutsianas<sup>3</sup>, Chris P Ponting<sup>2</sup>, James F Wilson<sup>2,20</sup>, Veronique Vitart<sup>2</sup>, Alexandre C Pereira<sup>23</sup>, Andre Luchessi<sup>24,25</sup>, Esteban Parra<sup>26</sup>, Raquel Cruz-Guerrero<sup>27</sup>, Angel Carracedo<sup>27,28</sup>, Angie Fawkes<sup>6</sup>, Lee Murphy<sup>6</sup>, Kathy Rowan<sup>29</sup>, Andy Law<sup>1</sup>, Sara Clohisey Hendry<sup>1</sup>, J. Kenneth Baillie<sup>‡,1,2,4,7</sup>.

‡ - These authors contributed equally

\* - A list of authors and their affiliations appears at the end of the paper.

† - to whom correspondence should be addressed: [j.k.baillie@ed.ac.uk](mailto:j.k.baillie@ed.ac.uk)

<sup>1</sup> Roslin Institute, University of Edinburgh, Easter Bush, Edinburgh, EH25 9RG, UK

<sup>2</sup> MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK

<sup>3</sup> Genomics England, London UK

<sup>4</sup> Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh, UK

<sup>5</sup> Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

<sup>6</sup> Edinburgh Clinical Research Facility, Western General Hospital, University of Edinburgh, EH4 2XU, UK

<sup>7</sup> Intensive Care Unit, Royal Infirmary of Edinburgh, 54 Little France Drive, Edinburgh, EH16 5SA, UK

<sup>8</sup> Clinical Research Centre at St Vincent's University Hospital, University College Dublin, Dublin, Ireland

<sup>9</sup> NIHR Health Protection Research Unit for Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences University of Liverpool, Liverpool, L69 7BE, UK

<sup>10</sup> Respiratory Medicine, Alder Hey Children's Hospital, Institute in The Park, University of Liverpool, Alder Hey Children's Hospital, Liverpool, UK

<sup>11</sup> Department of Medicine, University of Cambridge, Cambridge, UK

<sup>12</sup> William Harvey Research Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK

<sup>13</sup> Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Drive, Oxford, OX3 7FZ, UK

<sup>14</sup> Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong, China

<sup>15</sup> Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, UK

<sup>16</sup> Department of Intensive Care Medicine, Royal Victoria Hospital, Belfast, Northern Ireland, UK

<sup>17</sup> UCL Centre for Human Health and Performance, London, W1T 7HA, UK

<sup>18</sup> National Heart and Lung Institute, Imperial College London, London, UK

<sup>19</sup> Imperial College Healthcare NHS Trust:London,London,UK

<sup>20</sup> Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, Teviot Place, Edinburgh EH8 9AG, UK

<sup>21</sup> Great Ormond Street Hospital, London UK

<sup>22</sup> William Harvey Research Institute, Queen Mary University of London, Charterhouse Square, London EC1 6BQ

<sup>23</sup> Heart Institute, University of Sao Paulo, Brazil

<sup>24</sup> Department of Clinical Analysis and Toxicology, Federal University of Rio Grande do Norte, Brazil

<sup>25</sup> Department of Anthropology, University of Toronto Mississauga, Canada

<sup>26</sup> Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario, Canada

<sup>27</sup> CIBERER- Genomic Medicine Group, University of Santiago de Compostela, Santiago de Compostela, Spain

<sup>28</sup> Instituto de Investigacion Sanitaria de Santiago (IDIS), 15706 Santiago de Compostela, Spain

<sup>29</sup> Intensive Care National Audit & Research Centre, London, UK

## Abstract

Pulmonary inflammation drives critical illness in Covid-19,<sup>1;2</sup> creating a clinically homogeneous extreme phenotype, which we have previously shown to be highly efficient for discovery of genetic associations.<sup>3;4</sup> Despite the advanced stage of illness, we have found that immunomodulatory therapies have strong beneficial effects in this group.<sup>1;5</sup> Further genetic discoveries may identify additional therapeutic targets to modulate severe disease.<sup>6</sup> In this new data release from the GenOMICC (Genetics Of Mortality in Critical Care) study we include new microarray genotyping data from additional critically-ill cases in the UK and Brazil, together with cohorts of severe Covid-19 from the ISARIC4C<sup>7</sup> and SCOURGE<sup>8</sup> studies, and meta-analysis with previously-reported data. We find an additional 14 new genetic associations. Many are in potentially druggable targets, in inflammatory signalling (JAK1, PDE4A), monocyte-macrophage differentiation (CSF2), immunometabolism (SLC2A5, AK5), and host factors required for viral entry and replication (TMPRSS2, RAB2A). As with our previous work, these results provide tractable therapeutic targets for modulation of harmful host-mediated inflammation in Covid-19.

## Introduction

Critical illness in Covid-19 is a relatively narrow clinical phenotype characterised by hypoxaemic respiratory failure<sup>9</sup> in which lung injury is, at least in part, caused by host-mediated inflammation.<sup>1;2</sup> The international GenOMICC (Genetics of Mortality in Critical Care) consortium has shown that critical illness is a highly efficient phenotype for discovery of genetic variants associated with Covid-19,<sup>3;4</sup> contributing the largest signal to recent meta-analyses<sup>10</sup> despite only including a relatively small number of patients. This is consistent with our previous predictions for other respiratory viral illnesses.<sup>11;12</sup> New discoveries in the first GenOMICC study (implicating OAS1, TYK2, IFNAR2, and DPP9),<sup>3</sup> completed only 5 months after the first patient presented to a participating intensive care unit, led to new understanding of the host:pathogen interaction,<sup>3;13;14</sup> and, in part, led to the discovery of a new effective treatment for severe Covid-19.<sup>15</sup>

Here, we provide an updated analysis of the international GenOMICC study comprising a combination of microarray genotype data from 11,325 critically ill cases in the UK (9,279 cases) and Brazil (2,186), combined with other studies recruiting hospitalised patients with a strong focus on severe and critical disease: ISARIC4C (655 cases) and SCOURGE consortium (5,934 cases)(Table 1). We put these new results in context by completing a meta-analysis of the new GenOMICC GWAS results with previously-published data.

## Results

The design of the GenOMICC study has been previously described.<sup>3;4</sup> Briefly, patients with confirmed Covid-19 requiring continuous cardiorespiratory monitoring or organ support (a generalisable definition for admission to a critical care area) were recruited in 2020-21. Using participants recruited and genotyped in GenOMICC and ISARIC4C since the first reported GenOMICC GWAS,<sup>3</sup> we performed ancestry-specific GWAS largely following the methods we previously described<sup>3;4</sup> (see Material and Methods, and Supplementary Material). Using the results of these GWAS, previously reported results obtained using GenOMICC participants with whole genome sequencing data<sup>4</sup>, and data from GenOMICC Brazil, we performed trans-ancestry and -platform meta analyses, within the GenOMICC study, for a critically ill Covid-19 phenotype and a hospitalised Covid-19 phenotype (see Material and Methods). Results of these GenOMICC-only meta analyses are illustrated in the lower parts of Figure 1 and Supplementary Figure 3 for the critically ill and hospitalised phenotypes respectively.

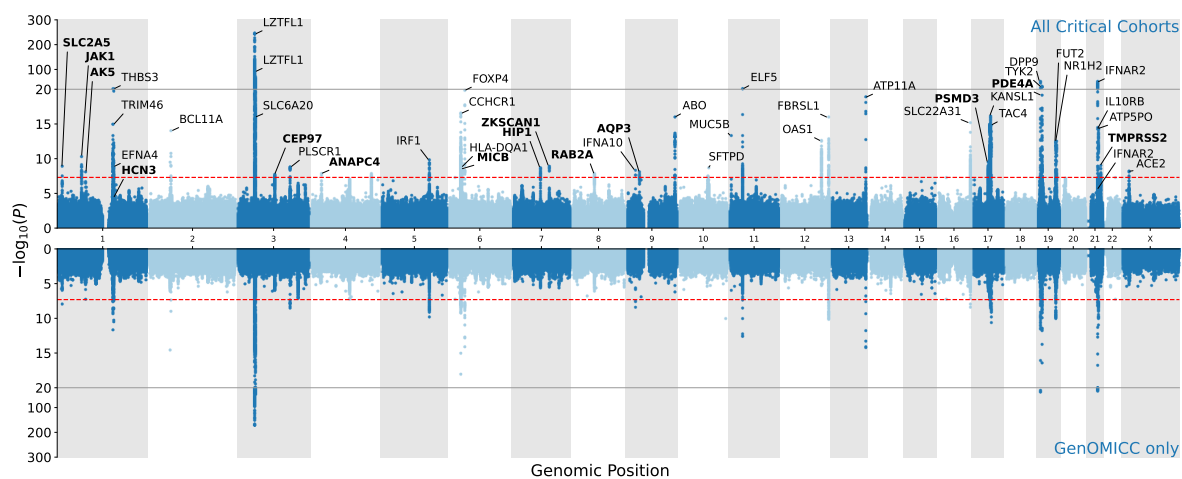


Figure 1: Miami plot showing meta-analysis results obtained using all critical phenotype cohorts (top) and using GenOMICC data only (bottom). Independent lead variants in the all critical cohorts analysis are annotated with nearby or plausible genes, with new associations from the present study in bold.

In order to put these new findings in the context of existing knowledge, we performed comprehensive meta analyses, drawing on further GWAS results for the two phenotypes of interest, including data kindly shared by the SCOURGE consortium and published data from the the Covid-19 Human Genetics Initiative (HGIv6, 2021). Characteristics of contributing studies to the meta analyses are summarised in Table 1 and Supplementary Table 1 for the critically ill and hospitalised phenotypes, with further details on each study provided in the Supplementary Material. Since leave-one-out analyses from the HGIv6 release were not available, we used a mathematical subtraction approach, as in our previous work,<sup>4</sup> to remove the GenOMICC signals from HGIv6, yielding an independent data set. This was essential because previous GenOMICC data contribute 20 to 50% of each genetic association signal in the HGI analyses. The results of these All Critical Cohorts and All Hospitalised Cohorts meta analyses are illustrated in Figure 1 and Supplementary Figure 3. Independent lead variants for the critically ill phenotype are summarised in Table 2 with more details provided in Supplementary Table 2, while Supplementary Table 4 contains details of lead variants from the analysis of the hospitalised phenotype.

Since no replication cohorts are available for these meta-analysis, we use heterogeneity across studies to assess the reliability of individual findings (see Supplementary Table 2 and 4). Due to the unusually extreme phenotype in the GenOMICC study, there is heterogeneity detected for the strongest associations from our previous work when compared to studies with more permissive inclusion criteria. Importantly, significant heterogeneity is not detected for the new findings we report here (Supplementary Table 2).

Study	Dataset	Ancestry	Case definition	Control definition	Median age(IQR)	Female(perc)	N cases	Ncontrols
GenOMICC	Whole-genome Sequence	EUR	Critical care	General population and mild cases	60(11.87)	32.17%	5989	42981
GenOMICC	Whole-genome Sequence	EAS	Critical care	General population and mild cases	54(11.27)	40.87%	274	366
GenOMICC	Whole-genome Sequence	SAS	Critical care	General population and mild cases	57(13.22)	25.63%	788	3793
GenOMICC	Whole-genome Sequence	AFR	Critical care	General population and mild cases	57(13.06)	35%	440	1350
GenOMICC	Genotype	EUR	Critical care	General population	58(13.49)	32.4%	1347	6735
GenOMICC	Genotype	SAS	Critical care	General population	58.5(16.10)	28.5%	200	1000
GenOMICC	Genotype	AFR	Critical care	General population	55(16.65)	44.5%	101	505
ISARIC4C	Genotype	EUR	ICU/CPAP/NIV	General population	59(12.75)	30%	140	700
HGIv6	Genotype	ALL	Critical (A2)	General population	NA	NA	8779	1001875
SCOURGE	Genotype	EUR	SCOURGE severe 3 -4	General population	69.13(12.93)*	27.5%	3502	5455
23 and me	Genotype	EUR	Respiratory Support	General population	53(17.5)**	50.7% **	495	680440

Table 1: Description of cohorts included in the critical meta-analysis with median age and percentage of females for the cases. NA: information not available for this cohort. \* mean was provided rather than the median. \*\* data was available for the hospitalisation cohort but not for the critically-ill patients.

In order to infer the effect of genetically-determined variation in gene expression on disease susceptibility, we performed a TWAS for the critical ill phenotype using gene expression data (GTExv8<sup>21</sup>) for two disease-relevant tissues, lung and whole blood. We found significant associations with  $P < 1 \times 10^{-6}$

between critical Covid-19 and predicted expression in lung (25), blood (18) and all-tissue meta-analysis (90; Supplementary file TWAS.xlsx)

## Discussion

In this latest analysis of the GenOMICC study of critical Covid-19, building on the foundations of previously published data from our group and others, we report a total of 45 genetic associations with severe Covid-19, of which 14 have not been reported previously (Table 2). After conditional analysis we find two extra secondary signals in known loci (see details in Materials and methods). In this brief report we provide a superficial description of selected findings. Our results are provided in full in Supplementary Tables 2 & 4.

We find a strong association in a key intracellular signalling kinase, JAK1, which is stimulated by numerous cytokines including Type I interferons and IL-6. Along with TYK2, which we previously reported to be associated with severe Covid-19,<sup>3</sup> JAK1 is the target for JAK inhibitors, which have recently been shown to be effective therapeutics in Covid-19.<sup>15</sup> Although the direction of effect is not clear from our genetic data for either gene, the therapeutic signal is consistent across multiple trials, providing proof-of-concept for target identification using genetics in critical illness.

We report a new lead variant within CSF2, the gene encoding granulocyte-macrophage colony stimulating factor (GM-CSF), a key cytokine in production and differentiation of myeloid cells including monocytes, macrophages and neutrophils. We previously reported that circulating levels of GM-CSF are associated with severe disease,<sup>22</sup> supporting a role for therapeutic targeting in severe Covid. In addition, we show that low expression of *PDE4A* is associated with critical Covid-19. This phosphodiesterase regulates production of multiple inflammatory cytokines by myeloid cells and is targeted by several existing drugs for treatment of a range of inflammatory diseases.<sup>23;24</sup>

Multiple genes implicated in viral entry are associated with severe disease. In addition to ACE2,<sup>19</sup> we see for the first time a genome-wide significant association in TMPRSS2, a key host protease which facilitates viral entry, which we have previously studied as a candidate gene.<sup>25</sup> This effect may be lineage-specific.<sup>26</sup> A strong GWAS association is seen in RAB2A (Table 2), with TWAS evidence more expression of this gene is associated with worse disease (Table 3). This gene is highly ranked in our previous MAIC<sup>27</sup> meta-analysis of host genes implicated in SARS-CoV-2 interaction using in vitro and clinical data,<sup>28</sup> and is consistent with CRISPR screen evidence showing that RAB2A is required for viral replication.<sup>29</sup>

TWAS provides a broad assessment of the effect on illness severity of genotype-predicted expression of specific transcripts in relevant tissues. TWAS results suggest a range of strongly-plausible genes, with direction of effect for gene expression effects in lung, whole blood and cross-tissue metaTWAS (Table 3, Supplemental file twas.xlsx). These include intriguing opposing effect estimates for predicted expression of various chemokine receptors (CCR1, CCR2, CCR9 vs CCR3 and CCR5) interferon- $\alpha$  subtypes (IFNA10 vs IFNA8) and intercellular adhesion molecules (ICAM1 vs ICAM3, ICAM5). Caution is needed in interpretation of these expression effects since the molecular mechanism underlying association is not known.

Although our focus on critical illness enhances discovery power, it has the disadvantage of combining genetic signals for multiple stages in disease progression including viral exposure, infection and replication, and development of inflammatory lung disease. From these data alone we cannot identify when in disease progression, or where in the body, the causal effect happens.

Because we performed a meta-analysis of multiple studies which may have slightly different definitions of the phenotype, effect sizes differ between studies. This, together with ancestry-specific effects,<sup>3</sup> may explain heterogeneity in strong GWAS signals, such as the LZTFL1 signal in table 2. Different studies also have sets of variants which are not completely overlapping, causing P-values between variants in high LD to be more different than expected. Although most of the studies contain individuals from multiple ancestries, a large majority of the individuals are from European ancestry. In future work, there is a scientific and moral imperative to include the full diversity of human populations.

Together, these results deepen our understanding of disease pathogenesis and highlights new biological mechanisms of disease, some of which have the potential for therapeutic targeting.

Chr	Pos(b38)	rsid	OR	OR <sub>CI</sub>	P	P <sub>cond</sub>	Gene	Citation
1	9067157	rs2478868	0.92	0.9-0.95	1.5e-09	4.82E-13	SLC2A5	GenOMICC <i>new</i>
1	64948270	rs12046291	1.1	1.07-1.13	5.1e-11		JAK1	GenOMICC <i>new</i>
1	77501822	rs71658797	1.1	1.08-1.18	9.8e-09		AK5	GenOMICC <i>new</i>
1	155066988	rs114301457	2.4	1.81-3.18	1.5e-09		EFNA4	GenOMICC <sup>16</sup>
1	155175305	rs7528026	1.4	1.25-1.45	1.3e-15		TRIM46	GenOMICC <sup>16</sup>
1	155197995	rs41264915	1.2	1.16-1.26	1.9e-21		THBS3	HGI <sup>17</sup>
1	155278322	rs11264349	0.94	0.92-0.97	7.3e-05		HCN3	GenOMICC <i>new</i>
2	60480453	rs1123573	1.1	1.09-1.15	1e-14		BCL11A	GenOMICC <sup>16</sup>
3	45796521	rs2271616	1.2	1.14-1.23	1.1e-16		SLC6A20	HGI <sup>17</sup>
3	45818159	rs17713054 <sup>‡</sup>	2.1	1.97-2.15	7.6e-243		LZTFL1	SCGG <sup>18</sup>
3	45873093	rs35482426	0.53	0.5-0.57	6.1e-91		LZTFL1	SCGG <sup>18</sup>
3	101737122	rs11706986	1.1	1.05-1.11	1.1e-08		CEP97	GenOMICC <i>new</i>
3	146517122	rs343320	1.2	1.11-1.23	1.8e-09		PLSCR1	GenOMICC <sup>16</sup>
4	25446871	rs7664615	1.1	1.07-1.14	1.5e-08		ANAPC4	GenOMICC <i>new</i>
4	167824478	rs1073165	1.1	1.05-1.1	1.7e-08		.	GenOMICC <i>new</i>
5	132441275	rs10066378	0.89	0.86-0.92	1.3e-10		IRF1	GenOMICC <sup>16</sup>
6	31153455	rs111837807	0.83	0.8-0.87	3.2e-17		CCHCR1	GenOMICC <sup>3</sup>
6	31513129	rs2516401	0.92	0.9-0.95	2.2e-09		MICB	GenOMICC <i>new</i>
6	32702531	rs2856717	1.1	1.06-1.11	2.2e-09		HLA-DQA1	GenOMICC <sup>16</sup>
6	41522644	rs41435745	1.4	1.31-1.51	1.5e-20		FOXP4	HGI <sup>17</sup>
7	75623396	rs1179620	0.92	0.9-0.95	2.3e-09		HIP1	GenOMICC <i>new</i>
7	100032719	rs2897075	1.1	1.05-1.11	1.1e-09		ZKSCAN1	GenOMICC <i>new</i>
8	60532539	rs13276831	1.1	1.05-1.1	2e-08		RAB2A	GenOMICC <i>new</i>
9	21206606	rs28368148	0.59	0.49-0.7	5.3e-09	IFNA10	GenOMICC <sup>3</sup>	
9	33425186	rs60840586	1.1	1.07-1.14	9.7e-09	AQP3	SCOURGE <sup>8</sup>	
9	133271182	rs879055593	1.1	1.1-1.16	1e-16	ABO	SCGG <sup>18</sup>	
10	79946568	rs721917	0.93	0.91-0.95	5.1e-09	SFTPD	HGI <sup>19</sup>	
11	1219991	rs35705950	0.85	0.82-0.89	5.5e-14	MUC5B	HGI <sup>19</sup>	
11	34482745	rs61882275	0.89	0.86-0.91	6.8e-21	ELF5	GenOMICC <sup>16</sup>	
12	112919637	rs2660	1.1	1.07-1.13	3.9e-13	OAS1	GenOMICC <sup>3</sup>	
12	132481571	rs11614702	1.1	1.08-1.14	1.5e-16	FBRSL1	GenOMICC <sup>16</sup>	
13	112881427	rs12585036	1.1	1.11-1.18	4e-19	ATP11A	GenOMICC <sup>16</sup>	
16	89196249	rs117169628	1.2	1.12-1.2	7.2e-16	SLC22A31	GenOMICC <sup>16</sup>	
17	40003082	rs12941811	0.92	0.9-0.95	4.5e-10	PSMD3	GenOMICC <i>new</i>	
17	46085231	rs8080583	0.88	0.86-0.91	8e-17	KANSL1	<sup>20</sup>	
17	49863303	rs77534576	1.4	1.25-1.45	1.5e-15	TAC4	<sup>20</sup>	
19	4717660	rs12610495	0.8	0.77-0.82	9.1e-51	DPP9	GenOMICC <sup>3</sup>	
19	10352442	rs34536443	1.5	1.39-1.61	2.2e-28	TYK2	GenOMICC <sup>3</sup>	
19	10414696	rs142770866	1.2	1.19-1.3	7.6e-20	PDE4A	GenOMICC <i>new</i>	
19	48702888	rs516316	0.91	0.89-0.93	3.7e-13	FUT2	GenOMICC <sup>16</sup>	
19	50374423	rs35463555	1.1	1.07-1.13	4.2e-13	NR1H2	HGI <sup>19</sup>	
21	33229937	rs188401375	0.74	0.66-0.84	3e-05	IFNAR2	GenOMICC <sup>3</sup>	
21	33237639	rs9636867	82	0.80-0.84	8.7e-49	IFNAR2	GenOMICC <sup>3</sup>	
21	33287378	rs8178521	1.1	1.1-1.17	6.2e-15	IL10RB	GenOMICC <sup>16</sup>	
21	33949755	rs75912728	1.2	1.13-1.22	3.6e-15	ATP5PO	GenOMICC <sup>16</sup>	
21	41479527	rs915823	1.1	1.07-1.14	1.7e-09	TMPRSS2	GenOMICC <i>new</i>	
23	15523993	rs35697037	1	1.03-1.06	6.8e-09	ACE2	HGI <sup>19</sup>	

Table 2: Genome-wide significant associations with critical covid. Chr: chromosome, Pos(b38): position on human genome build 38, OR: odds ratio, OR<sub>CI</sub>: confidence interval, Gene: nearest or most plausible nearby gene, Citation: first demonstration of association. <sup>‡</sup> indicates significant heterogeneity across studies. P<sub>cond</sub>: P-value in conditional analysis in variants with P > 5 × 10<sup>-8</sup> Details about conditional analysis are in Supplementary Table 3



Gene	$Z_{lung}$	$Z_{blood}$	$Z_{meta}$	$Z_{SD}$	$n_{meta}$	$P_{meta}$
CCR9		31	13	11	7	2.2e-224
SLC6A20	5.7		9.2	6.2	33	6.9e-224
IL10RB	8.4	-3.1	7.1	3.9	46	4e-33
CCR2	1.6	0.51	5.9	7.7	35	1.3e-42
TRIM46			5.8	5.1	18	4.7e-22
CDC37			5.1	0.92	27	3.5e-08
MUC1	5.7	7	5	2.7	42	8.2e-24
RAB2A	5	3.9	4.6	0.88	45	1.3e-06
TCF19	5.6	4.6	4.4	2.7	49	1.1e-10
FYCO1	8.9	3.7	4.2	4.1	47	2e-41
GSDMA	5.9	0.56	3.8	2.2	38	3.3e-06
ELF5	5.8		3.7	1.7	9	2.5e-08
ABO	5.1	-1.7	3.7	1.9	48	1.6e-07
CCHCR1	4.3	4.3	3.5	2.1	49	2.9e-12
ATP5O	6.4	2.4	3.3	1.6	47	3.8e-11
RP11-387H17.6	4.3		2.7	2.8	5	8.5e-06
CAT	4.1	3.1	2.7	2.2	49	5.5e-13
TYK2	5.8	7.8	2.7	3.7	46	8.7e-17
HLA-C	4.6	5.9	2.7	2.5	48	2.2e-14
ADAM15	-6.4	3	2.5	2.1	46	8.9e-16
GSTCD	-4.7	2.8	2.3	2.4	47	7.6e-06
MED24	-1.4	4.9	2.2	3.4	46	8.9e-08
HIP1	4.1	-5.8	0.94	1.7	42	1.8e-07
OAS3	0.97	4.5	0.75	4.1	37	9.7e-09
ZNF778	4.2	0.24	0.74	2.1	49	1.7e-06
FUT2	6.6		0.36	5.2	28	2e-10
FOXP4	8.7		0.12	2	28	8e-16
ARHGAP27	-6.6	3.6	0.098	5.2	44	7.8e-10
EFNA1	-2.5	-6.6	0.042	2.6	48	2.2e-07
ACSL6	3.1	4.6	-0.41	3.6	44	1.4e-07
RAVER1	-3.8	5.7	-0.56	2	45	7.2e-10
XCR1	-2.7	5.7	-0.7	6.1	23	1.6e-122
JAK1	-5.3	0.82	-1.7	3.3	47	6.9e-06
CEP97	-5.3	0.86	-1.7	2.8	17	8.8e-06
MAMSTR	-1.7	5.5	-1.8	3.5	45	9.5e-09
NXPE3	-3.5	-4.3	-2.2	2	47	4.2e-06
PSORS1C2	-4.3		-2.4	1.7	39	4.6e-10
IFNAR1	-4.1	-3.6	-2.5	2.2	49	1.5e-06
OAS1	-4.7	-3.3	-2.9	1.8	41	2.7e-09
NTN5	-4.3	3.7	-3.4	3.1	47	9.8e-09
CCR5	-15	2.1	-3.7	3.5	46	9.3e-47
DPP9	-1.2	-5.5	-3.9	3.8	46	4.3e-41
CSF3	-5.6		-4	2	38	7e-07
LINC01301	-5		-4.2	1.2	38	4.4e-06
IFNAR2	-5.4	-4.1	-4.4	7	48	6.7e-52
RASIP1	-5.2	-5.2	-4.6	2	32	3.7e-08
WNT3	-7.6	-7	-4.7	2.3	47	1.2e-10
NAPSA	-0.56	-6.8	-6.1	1.9	39	1.2e-10
CXCR6	-21		-8.1	9.6	38	4.4e-134
CCR3	-6.2	-1.8	-9	6	6	2e-65

Table 3: Selected TWAS results for lung, whole blood and metaTWAS with  $Z > 5$  in any analysis and  $P_{meta} < 1 \times 10^{-5}$  (See Supplementary File twas.xlsx for full results). Gene: Gene symbol,  $Z_{lung}$ : TWAS Z-score in lung,  $Z_{blood}$ : TWAS Z-score in blood,  $Z_{meta}$ : metaTWAS mean Z-score,  $Z_{SD}$ : metaTWAS Z-score standard deviation,  $n_{meta}$ : number of SNPs included in metaTWAS,  $P_{meta}$ : metaTWAS P-value.

## Acknowledgements

We thank the patients and their loved ones who volunteered to contribute to this study at one of the most difficult times in their lives, and the research staff in every intensive care unit who recruited patients at personal risk during the most extreme conditions ever witnessed in most hospitals.

GenOMICC was funded by Sepsis Research (the Fiona Elizabeth Agnew Trust), the Intensive Care Society, a Wellcome Trust Senior Research Fellowship (J.K.Baillie, 223164/Z/21/Z), the Department of Health and Social Care (DHSC), Illumina, LifeArc, the Medical Research Council, UKRI, a BBSRC Institute Program Support Grant to the Roslin Institute (BBS/E/D/20002172, BBS/E/D/10002070 and BBS/E/D/30002275) and UKRI grants MC\_PC\_20004, MC\_PC\_19025, MC\_PC\_1905, and MRNO2995X/1.

This research is supported in part by the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (grant ref MC\_PC.20029). We acknowledge NHS Digital, Public Health England and the Intensive Care National Audit and Research Centre who provided clinical data on the participants. This study owes a great deal to the National Institute for Healthcare Research Clinical Research Network (NIHR CRN) and the Chief Scientist's Office (Scotland), who facilitate recruitment into research studies in NHS hospitals, and to the global ISARIC and InFACT consortia. GenOMICC genotype controls were obtained using UK Biobank Resource under project 788 funded by Roslin Institute Strategic Programme Grants from the BBSRC (BBS/E/D/10002070 and BBS/E/D/30002275) and Health Data Research UK (references HDR-9004 and HDR-9003). The work of LK was supported by an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1).

We would like to thank the research participants and employees of 23andMe for making this work possible. A full list of contributors who have provided data that was collated in the HGI project, including previous iterations, is available at <https://www.covid19hg.org/acknowledgements>.

## Contributions

EP-C, KR, KM, SK, MJC, CPP, JFW, VV, AL, EP, RC-G, AC, AF, LMu, KRo, ALa, SCH and JKB contributed to design. EP-C, KR, LK, AK, AR, AS, CAO, SW, LM, ALa, SCH and JKB contributed to data analysis. EP-C, KR, AK, AR, JM, CDR, AS, CAO, SW, LM, ALa and SCH contributed to bioinformatics. EP-C, KR, AK, JM, CDR, RT, AN, MGS, MJC, LM, SCH and JKB contributed to writing and reviewing the manuscript. SW, FG and WO contributed to project management. FG, WO, KM, SK, AN, MGS, JK, MS-H, CS, CH, PH, LL, DM, HM, PJO, TW, AT, RHS, CPP, AF, LMu, KRo, ALa and SCH contributed to oversight. FG, WO and JKB contributed to ethics and governance. KM, AF and LMu contributed to sample handling and sequencing. MJC and JKB contributed to scientific leadership. CPP, KRo, SCH and JKB contributed to conception. CPP, JFW, VV, AL, EP, RC-G, AC and KRo contributed to reviewing the manuscript. KRo and ALa contributed to clinical data management.

## Data accessibility

All data, including downloadable summary data and access applications for individual-level data, can be obtained through the GenOMICC gateway site <https://genomicc.org/data>.

## Code availability

Code to calculate the imputation of  $P$ -value based on LD SNPs is available at [https://github.com/baillielab/GenOMICC\\_GWAS](https://github.com/baillielab/GenOMICC_GWAS)

## Materials and methods

### Ethical approval

All participants, or their representatives where appropriate, gave informed consent. Details of ethical review can be found in Table 4.

Study/cohort	Ethical review committee	Reference Number	Decision
GenOMICC Scotland	Scotland A Research Ethics Committee	15/SS/0110	approved
GenOMICC England/Wales/Northern Ireland	Coventry and Warwickshire Research Ethics Committee	19/WM/0247	approved
GenOMICC Brazil (BraCovid)	National Research Ethics Committee (CONEP) and Ethics Committee for the Analysis of Research Projects at HC FMUSP (CAPPesq)	5025/20/054	approved
ISARIC4C England/Wales/Northern Ireland	South Central Oxford C Research Ethics Committee	13/SC/0149	approved
ISARIC4C Scotland	Scotland A Research Ethics Committee	20/SS/0028	approved
Covid-19 HGI	Multiple ethics committees	see ref <sup>17</sup>	approved
SCOURGE	Galician Ethical Committee	2020/197	approved
23andme	Ethical and Independent Review Services <a href="http://www.eandireview.com">http://www.eandireview.com</a>	N/A	approved

Table 4: Ethical review details for primary cohorts.

### Hospitalisation meta-analysis

The hospitalised phenotype includes patients who were hospitalised with a laboratory-confirmed SARS-Cov2 infection. In this analysis we included GenOMICC, GenOMICC Brazil, ISARIC4C, HGIv6 B2 phenotype with subtraction of GenOMICC data, SCOURGE hospitalised vs population and mild cases, and 23andme broad respiratory phenotype. Summary description of each analysis can be found in Supplementary material and a table with the included studies can be found in Supplementary Table 1.

### Critical illness meta-analysis

The critically ill Covid-19 group included patients who were hospitalized owing to symptoms associated with laboratory-confirmed SARS-CoV-2 infection and who required respiratory support or whose cause of death was associated with Covid-19. In the critical analysis we included GenOMICC, critically-ill patients from ISARIC4C, HGIv6 phenotype A2 with subtraction of GenOMICC data, SCOURGE severity grades 3 and 4 vs population controls, and 23andme respiratory support phenotype. Summary description of each analysis can be found in Supplementary material and a table with the included studies can be found in table 1.

### Meta-analysis

All meta-analysis across studies were performed using an inverse-variance weighting method and control for population stratification in the METAL software<sup>30</sup>. Variants were filtered out if they were not present in one of the three biggest studies: GenOMICC European ancestry, HGIv6, or SCOURGE. Allele frequency was calculated as the average frequency across studies with the METAL option AVERAGEFREQ. Heterogeneity was calculated using a Cochran's Q-test implemented in METAL. For variants in the same position with different REF and ALT alleles across studies, the GenOMICC variant in European population was selected and the rest were removed. Finally, variants with switched ALT and REF alleles between HGIv6 and GenOMICC were also removed based on differences in allele frequency of the alternative allele. Variants were annotated to the closest genes using dbsnp version b151 GRCh38p7 and biomaRt R package version 2.46.3.<sup>31</sup>

### LD clumping

We used plink 1.9 to clump variants that were genome-wide significant for each analysis with  $P1 = 5 \times 10^{-8}$ ,  $P2 = 0.01$ , clump distance 1500Kb and  $r^2 = 0.1$ . As a reference population for clumping we used individuals from European ancestry with whole genome sequence available in the GenOMICC study and whole genomes from 100K Genomics England project<sup>4</sup>.

### Conditional analysis

We performed a step-wise conditional analysis to find independent signals. In the hospitalised analysis we performed an European-specific meta-analysis to find conditionally independent signals. For the



critical illness meta-analysis European-specific data is not available, but European ancestry is largely predominant (87.2% of critically-ill cases) and we performed the conditional analysis using the meta-analysis results of the whole cohort. To perform the conditional analysis we used GCTA 1.9.3 –cojo-slt function<sup>32</sup>. The parameters for the function were  $pval = 5 \times 10^{-8}$ , a distance of 10,000 kb and a co-linear threshold of 0.9<sup>33</sup>, and the reference population for the conditional analysis were individuals from European ancestry with whole genome sequence available in the GenOMICC study and whole genomes from 100K Genomics England project<sup>4</sup>.

## Transcriptome-wide Association Studies

We performed TWAS in the MetaXcan framework and the GTEXv8 eQTL and sQTL MASHR-M models available for download in <http://predictdb.org/>. We first calculated individual TWAS for whole blood and lung with the S-PrediXcan function<sup>34,35</sup>. Then we performed a metaTWAS including data from all tissues to increase statistical power using s-MultiXcan<sup>36</sup>. We applied Bonferroni correction to the results in order to choose significant genes and introns for each analysis.

## References

- [1] Horby, P. *et al.* Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *New England Journal of Medicine* (2020).
- [2] Dorward, D. A. *et al.* Tissue-specific Immunopathology in Fatal COVID-19. *American Journal of Respiratory and Critical Care Medicine* (2020).
- [3] Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in covid-19. *Nature* **591**, 92–98 (2021).
- [4] Kousathanas, A. *et al.* Whole genome sequencing reveals host factors underlying critical Covid-19. *Nature* 1–10 (2022).
- [5] Abani, O. *et al.* Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial. *The Lancet* **397**, 1637–1645 (2021).
- [6] Baillie, J. K. Targeting the host immune response to fight infection. *Science* **344**, 807–808 (2014).
- [7] Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: Prospective observational cohort study. *BMJ* **369** (2020).
- [8] Carracedo, Á. & on Covid-19 (scourge), S. C. t. U. R. o. h. G. A genome-wide association study of COVID-19 related hospitalization in Spain reveals genetic disparities among sexes (2021).
- [9] Docherty, A. B. *et al.* Features of 20 133 uk patients in hospital with covid-19 using the isaric who clinical characterisation protocol: prospective observational cohort study. *BMJ (Clinical research ed.)* **369**, m1985 (2020).
- [10] Niemi, M. E. K. *et al.* Mapping the human genetic architecture of covid-19. *Nature* **600**, 472–477 (2021).
- [11] Horby, P., Nguyen, N. Y., Dunstan, S. J. & Baillie, J. K. The Role of Host Genetics in Susceptibility to Influenza: A Systematic Review. *PLoS ONE* **7**, e33180 (2012).
- [12] Patarčić, I. *et al.* The role of host genetic factors in respiratory tract infectious diseases: Systematic review, meta-analyses and field synopsis. *Scientific Reports* **5** (2015).
- [13] Zhou, S. *et al.* A neanderthal *oas1* isoform protects individuals of european ancestry against covid-19 susceptibility and severity. *Nature medicine* **27**, 659–667 (2021).

- [14] Wickenhagen, A. *et al.* A prenylated dsrna sensor protects against severe covid-19. *Science (New York, N. Y.)* **374**, eabj3624 (2021).
- [15] Horby, P. W. *et al.* Baricitinib in patients admitted to hospital with COVID-19 (RECOVERY): A randomised, controlled, open-label, platform trial and updated meta-analysis (2022).
- [16] Kousathanas, A. *et al.* Whole genome sequencing identifies multiple loci for critical illness caused by covid-19. *medRxiv* (2021). URL <https://www.medrxiv.org/content/early/2021/09/08/2021.09.02.21262965>.
- [17] COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* (2021). URL <https://doi.org/10.1038/s41586-021-03767-x>.
- [18] Ellinghaus, D. *et al.* Genomewide association study of severe covid-19 with respiratory failure. *The New England journal of medicine* **383**, 1522–1534 (2020).
- [19] Initiative, C.-. H. G. & Ganna, A. Mapping the human genetic architecture of COVID-19: An update (2021).
- [20] Degenhardt, F. *et al.* New susceptibility loci for severe COVID-19 by detailed GWAS analysis in European populations (2021).
- [21] Consortium, T. G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). URL <https://science.sciencemag.org/content/369/6509/1318>. Publisher: American Association for the Advancement of Science \_eprint: <https://science.sciencemag.org/content/369/6509/1318.full.pdf>.
- [22] Thwaites, R. S. *et al.* Inflammatory profiles across the spectrum of disease reveal a distinct role for GM-CSF in severe COVID-19. *Science Immunology* **6** (2021).
- [23] Sakkas, L. I., Mavropoulos, A. & Bogdanos, D. P. Phosphodiesterase 4 inhibitors in immune-mediated diseases: Mode of action, clinical applications, current and future perspectives. *Current medicinal chemistry* **24**, 3054–3067 (2017).
- [24] Contreras, S., Milara, J., Morcillo, E. & Cortijo, J. Selective inhibition of phosphodiesterases 4A, b, c and d isoforms in chronic respiratory diseases: Current and future evidences. *Current pharmaceutical design* **23**, 2073–2083 (2017).
- [25] David, A. *et al.* A common TMPRSS2 variant has a protective effect against severe COVID-19. *Current Research in Translational Medicine* **70**, 103333 (2022).
- [26] Meng, B. *et al.* Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts tropism and fusogenicity. *Nature* 1–1 (2022).
- [27] Li, B. *et al.* Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nature Communications* **11**, 164 (2020).
- [28] Parkinson, N. *et al.* Dynamic data-driven meta-analysis for prioritisation of host genes implicated in COVID-19. *Scientific Reports* **10**, 22303 (2020).
- [29] Hoffmann, H.-H. *et al.* Functional interrogation of a sars-cov-2 host protein interactome identifies unique and shared coronavirus host factors. *bioRxiv : the preprint server for biology* (2020).
- [30] Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190–2191 (2010).
- [31] Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature protocols* **4**, 1184–91 (2009).
- [32] Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Gcta: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76–82 (2011). URL <https://pubmed.ncbi.nlm.nih.gov/21167468/>.

- [33] Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369–375 (2012). URL <https://doi.org/10.1038/ng.2213>.
- [34] Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature Communications* **9**, 1825 (2018).
- [35] Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).
- [36] Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genetics* **15**, 1–20 (2019).