Genetic mechanisms of critical illness in Covid-19

Abstract

The subset of patients who develop critical illness in Covid-19 have extensive inflammation affecting the lungs.[1] Critically-ill patients appear have opposite reponses to immunosuppressive therapy compared with non-critical cases.[2] Since susceptibility to life-threatening infections and immune-mediated diseases are both strongly heritable traits, we reasoned that host genetic variation may identify mechanistic targets for therapeutic development in Covid-19.[3]

The GenOMICC (Genetics Of Mortality In Critical Care, genomicc.org) study is a global collaborative study to understand the genetic bases of critical illness. Here we report results from patients recruited with Covid-19 across 208 intensive care units in the UK (>95% of all ICU beds). We performed a genome-wide association study in 2790 critically-ill patients with life-threatening Covid-19 compared to random controls from three distinct UK population studies and replicated our results in an independent dataset.

We identify and replicate several novel genome-wide significant associations including variants chr19p13.3 (rs2109069, $P = 3.98 \times 10^{-12}$), within the gene encoding dipeptidyl peptidase 9 (*DPP9*), and at chr21q22.1 (rs2236757, $P = 4.99 \times 10^{-8}$) in the interferon receptor *IFNAR2*. Consistent with our focus on extreme disease in younger patients with less comorbidity, we detect a stronger signal at the known 3p21.31 locus than previous studies (rs73064425, $P = 1.2 \times 10^{-27}$).

Using Mendelian randomisation we found compelling evidence in support of a causal link between low expression of *IFNAR2* and life-threatening disease. We detected genome-wide significant gene-level associations for genes with central roles in viral restriction (*OAS1*, *OAS2*, *OAS3*). Transcriptome-wide association in lung tissue revealed that high expression of the monocyte/macrophage chemotactic receptor *CCR2* is associated with severe Covid-19. These results identify specific loci associated with life-threatening disease, and identify potential host-directed therapeutic targets for which existing, licensed drugs are available.

Introduction

Critical illness in Covid-19 is caused, in part, by inflammatory injury affecting the lungs and lung blood vessels.[4,5] There are therefore at least two distinct biological components to mortality risk: susceptibility to viral infection, and propensity to develop harmful pulmonary inflammation. Susceptibility to life-threatening infections[6] and immune-mediated diseases are both strongly heritable. In particular, susceptibility to respiratory viruses[7] such as influenza[8] is heritable and known to be associated with specific genetic variants.[9] In Covid-19, one genetic locus, 3p21.31 has been repeatedly associated with hospitalisation.[10,11] As with other viral illnesses,[12] there are several examples of loss-of-function variants affecting key immune processes that lead to severe disease in young people: for example *TLR7* defects among 4 cases with severe disease.[13] Understanding the molecular mechanisms of critical illness in Covid-19 may reveal new therapeutic targets to modulate this host immune response to promote survival.[3]

There is now strong evidence that critical illness caused by Covid-19 is qualitatively different from mild or moderate disease, even among hospitalised patients. There are multiple distinct disease phenotypes with differing patterns of presenting symptoms[14] and marked differential responses to immunosuppressive therapy.[2] In patients without respiratory failure, there is a trend towards harm from treatment with corticosteroids, whereas among patients with critical respiratory failure, there is a very substan-

tial benefit.[2] On this basis, we can consider patients with Covid-19 respiratory failure to have distinct pathopysiology.

In the UK, the group of patients admitted to critical care is relatively homogeneous, with profound hypoxaemic respiratory failure being by far the most common presentation.[15] The active disease process in these patients is strikingly responsive to corticosteroid therapy[16] and is characterised by pulmonary inflammation including diffuse alveolar damage, macrophage/monocyte influx, mononuclear cell pulmonary artery vasculitis and microthrombus formation.[4,5]

Host-directed therapies have long been a target for the treatment of severe disease caused by respiratory viruses.[17] Identification of genetic loci associated with susceptibility to Covid-19 may lead to specific targets for repurposing or drug development.[3]

The GenOMICC (Genetics Of Mortality In Critical Care, genomicc.org) study has been recruiting patients with critical illness syndromes, including influenza, sepsis, and emerging infections, for 5 years. We recruited patients with life-threatening Covid-19 and performed a genome-wide association study comparing them to controls from three population genetic studies in the UK.

Results

Demographic and summary clinical characteristics of the cohort are described in Table xx. Cases were representative of the UK critically-ill population.[15]

Table 1: Baseline characteristics of patients included. Functionally-limiting comorbid illness was defined at the discretion of the treating clinicians.

Characteristics	GenOMICC (n=2109)	ISARIC (n=134)
Female sex	624 (29.8%)	46 (34.1%)
Age (yrs, mean ± sd)	57.3 ± 12.1	57.3 ± 2.9
European ancestry	1573 (74.6%)	103 (76.3%)
African ancestry	174 (8.2%)	8 (5.9%)
East Asian ancestry	143 (6.8%)	6 (4.4%)
South Asian ancestry	219 (10.4%)	18 (13.3%)
Functionally-limiting comorbidity	396 (18.8%)	0(0%)
Unknown	49 (2.3%)	134(100%)
Invasive ventilation	1557 (73.8%)	25 (18.5%)
Unknown	35 (1.6%)	31 (22.9%)
Died (60 days)	459 (21.8%)	22 (16.3%)
Unknown	338 (16%)	30 (22.4%)

DNA was extracted from whole blood and genome-wide genotyping and quality control were performed according to standard protocols (Materials & Methods). Briefly, genetic ancestry was inferred for unrelated individuals passing quality control using ADMIXTURE and reference individuals from the 1000 Genomes project. Imputation was performed using the TOPMed reference panel.[18] Ancestry-matched controls not having Covid-19 PCR tests were selected from the large population-based cohort UK Biobank.

Further controls were selected for the European cases from the population-based Generation Scotland cohort to allow validation of associations. GWAS was carried out separately by ancestry group using logistic regression in PLINK and accounting for age, sex, postal code deprivation decile and principal components of ancestry. As well as standard filters for minor allele frequency (>0.01), imputation

quality (0.9) and Hardy-Weinberg equilibrium (10^{-150}), GWAS results were filtered on allele frequency against the genome aggregation database (gnomAD), to avoid biases arising from different imputation panels (and arrays) between cases and controls.

Since no study of critical illness in Covid-19 of sufficient size is available, replication was sought in the Covid-19 Host Genetics Initiative (HGI) hospitalised Covid-19 versus population analysis. Meta-analysis of GenOMICC and HGI was performed in METAL. 13 variants, in 10 distinct genomic loci, were significantly associated with life-threatening Covid-19 in transethnic meta-analysis. Of these, xxx variants replicated in the Covid-19 HGI study.

GWAS results

Table 2: Lead variants from independent genome-wide significant regions.

Location	SNP ID	Genes	p_meta	p_rep
3:45901089_T/C	rs73064425	*FYCO1*	1.72×10^{-27}	
19:4719443	rs2109069	*DPP9*	3.97×10^{-12}	
21:34624917_A/G	rs2236757	*IFNAR2*	1.99×10^{-7}	
12:113380008_A/G	rs10735079	*OAS1*, *OAS2*, *OAS3*	7.25×10^{-8}	
19:10466123	rs11085727	*TYK2*, *ICAM1*, *ICAM3*, *ICAM5*	1.31×10^{-7}	
7:107607902	rs2237698	*LAMB1*	2.61×10^{-8}	
12:103014757	rs10860891	*IGF1*	5.3×10^{-17}	
7:138172471_G/A	rs6467776	*TRIM24*	7.8×10^{-8}	

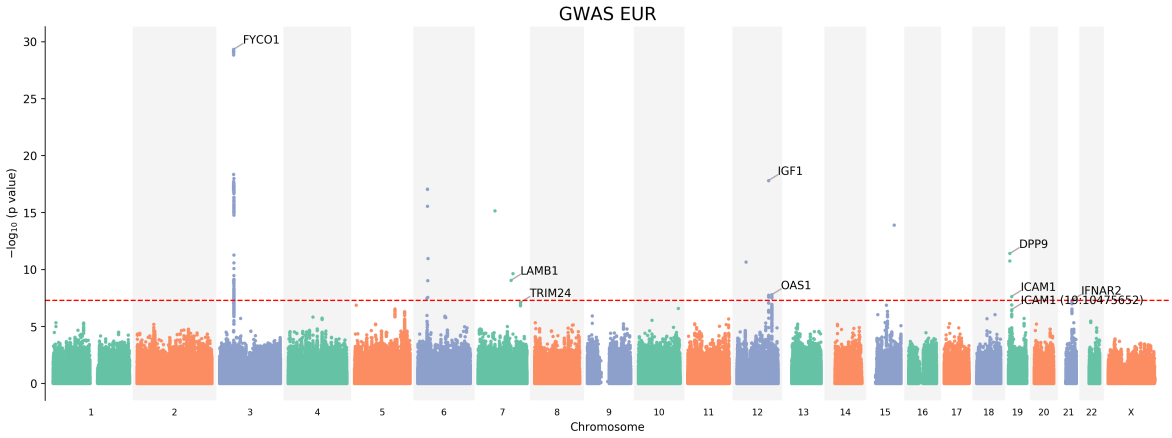

Figure 1: Manhattan plot showing SNP-level p-values from EUR

Manhattan plot showing SNP-level p-values from meta-analysis vs Covid19-hg

Replication

Table 3: Summary of replication statistics. {#tbl: replication}

topSNP	chr:pos (b37)	risk allele	beta	p-value	nearest gene
rs67959919	3:45871908	A	0.6571	1.445e-37	LZTFL1
rs143334143	6:31121426	A	0.316	1.665e-10	CCHCR1
rs9501257	6:33055355	A	-0.3126	4.93e-09	HLA-DPB1
rs622568	7:54647894	A	-0.2889	9.894e-11	
rs2237698	7:107607902	T	0.2661	1.728e-08	LAMB1
rs12705891	7:113317708	T	-0.1793	2.89e-08	
rs10087754	8:122832148	A	-0.177	3.475e-08	
rs10860891	12:103014757	A	-0.2961	1.072e-09	LOC105369944
rs4766664	12:113362997	T	-0.1981	2.643e-09	OAS1
rs11634857	15:79766794	A	-0.2325	8.453e-09	
rs2277732	19:4723670	A	0.2537	2.014e-13	DPP9
rs11085727	19:10466123	T	0.2203	1.868e-10	TYK2
rs13050728	21:34615210	T	0.2112	2.825e-10	IFNAR2

Gene-level association test

Gene level burden of significance was calculated using MAGMA. Twelve genes had a gene level p-value $<5x10^{-6}$.

Table 4: Genes with p<5e-6 in gene-level (MAGMA) analysis.

Gene	P	Number of variants
LZTFL1	7.9751E-14	103
FYCO1	1.7685E-13	191
XCR1	3.5541E-13	46
CXCR6	1.5365E-09	34
OAS3	8.6688E-09	111
CCR1	2.1795E-07	32
OAS2	4.3267E-07	55
IFNAR2	6.253E-07	90
CCR3	7.046E-07	52
CCR2	1.2877E-06	27
OAS1	2.1406E-06	36
TNFSF15	5.3835E-06	22

Of these 12 genes, 7 are found in the 3p21.31 locus: *LZTFL1*, *FYCO1*, *XCR1*, *CXCR6*, *CCR1*, *CCR3* & *CCR2*. The genes *OAS1*, *OAS2*, *OAS3* are grouped in locus q24.13 on chromosome 12. Gene set analysis of gene-level burden of significance did not identify any significantly enriched pathways or gene ontology terms after correction multiple comparisons (FDR<0.05).

Mendelian randomisation

Given a set of assumptions, discussed extensively elsewhere,[PMID 32628676] Mendelian randomisation provides evidence for a causal relationship between an exposure variable and an outcome. We employ it here to assess the evidence in support of causal effects of RNA expression of various genes on the

odds of critical Covid-19.

We specified an *a priori* list of target genes that relate to the mechanism of action of many host-targeted drugs that have been proposed for the treatment of Covid-19 (Supp Table XXX). Seven of these targets had a suitable locally-acting eQTL. Of these, *IFNAR2* remained significant after Bonferroni correcting for multiple testing for 7 tests (beta -1.49, standard error 0.52, p-value 0.0043), with equivocal evidence of heterogeneity (HEIDI[PMID 27019110] p-value = 0.0150; $0.05/7 <$ p-value < 0.05; 6 SNPs). This Mendelian randomisation result succesfully replicated in the results of COVID19-hg ('ANA_B2_V2'; hospitalized covid vs. population): beta -1.14, standard error 0.40, p-value 0.0066 (1 test). Whilst not a complete replication - due to the repeated use of GTEx v7 Whole Blood data - we believe that this adds considerable weight to the causal evidence in support of *IFNAR2*.

We then performed transcriptome-wide Mendelian randomisation to quantify support for *unselected* genes as potential therapeutic targets. Instruments were available for 4,614 unique Ensembl gene IDs. No genes were statistically significant after correcting for multiple comparisons in this analysis (4,614 tests). After conservative filtering for evidence of significant heterogeneity using HEIDI (p-value $<$ 0.05), the smallest Mendelian randomisation p-value was 0.00049 for a variant at chr19:10466123 affecting expression of *TYK2*. 9 other genes with Mendelian randomisation $p < 0.05$ were then tested for independent external evidence (Supp tab XXX). We found that *TYK2* had an significant independent Mendelian randomisation $p = 0.0022$ in this second set (Bonferroni-corrected significance threshold = 0.006).

TWAS

Transcriptome-wide association studies (TWAS) link GWAS results to tissue-specific gene expression data by inferring gene expression from known genetic variants that are associated with transcript abundance (expression quantitative trait loci, eQTL).[19,20] We performed TWAS to look for associations with gene expression using GTExv8 [ref] data for two disease-relevant tissues chosen *a priori*: whole blood and lung.

Genetic correlations, tissue, and cell-type associations

Using high-definition likelihood (HDL),[21] we tested for genetic correlations with other traits, that is the degree to which the underyling genetic components are shared with severe Covid-19. This revealed no independently-significant genetic correlations after correcting for multiple comparisons (Supplementary Figure XXX, Supplementary Table XXX). Consistent with GWAS results from other infectious and inflammatory diseases,[22] there was a significant enrichment of strongly-associated variants in enhancers, particularly those identified by the EXaC study as under strong evolutionary selection[23] (Supplementary Figure XXX). The strongest tissue type enrichment was in spleen, followed by pancreas (Supplementary Figure XXX).

Discussion

We have discovered and replicated robustly significant associations with susceptibility to life-threatening Covid-19. Our focus on critical illness increases the probability that some of these associations relate to the later, immune-mediated disease associated with respiratory failure requiring invasive mechanical ventilation.[2]

Patients admitted to intensive care units in the UK during the first wave of Covid-19 were, on average, younger and less burdened by comorbid illness than the hospitalised population.[15] Compared to other

countries, UK ICU admission tends to occur at a higher level of illness severity,[24] reflected in the high rate of invasive mechanical ventilation use in our cohort (73%; Table 1). Therefore, the population studied here are defined by extreme susceptibility to severe Covid-19. GenOMICC recruited in 208 intensive care units (covering more than 95% of UK ICU capacity), ensuring that a broad spread across the genetic ancestry of UK patients was included (Figure **??**).

Our key findings are consistent across 4 ancestry groups and 3 control groups (Table XXX). The nearest comparison is the hospitalised vs population analysis in the Covid-19 Host Genetics initiative, which has been generously shared with the international community. Likewise, full summary statistics from GenOMICC have been made openly available in order to advance the rate of discovery in this area.

Despite the differences in case definitions, novel associations from our study of critical illness replicate robustly in the hospitalised case study: rs2109069 on 19p13.3, xxx and xxx (Figure **??**). The Mendelian randomisation association with *IFNAR2* is also replicated.

The association in 19p13.3 (rs2109069, p = 3.98 x10^{12}) is an intronic variant in the gene encoding dipeptidyl peptidase 9 (*DPP9*). Variants in this locus are associated with idiopathic pulmonary fibrosis[25] and interstitial lung disease.[26] *DPP9* encodes a serine protease with diverse intracellular functions, including cleavage of the key antiviral signalling mediator CXCL10,[27]

We replicate the finding of Ellinghaus *et al.* at 3p21.31.[11] The extremely small p-value at this locus (p=1.25 x 10^{-27}) may reflect the strength of the signal, the large size of our study, our focus on extreme severity, and our inclusion of ethnic groups in which the risk allele is more prevalent (28% in South Asian populations[28]). While this effect size is surprisingly large (minor allele frequency = 7.5% for lead variant rs73064425 in Europeans), this is consistent with effect sizes reported previously (OR 2.1[11] and 1.7[10]).

The 3p21.31 locus is populated by a number of genes with mechanisms of action that could plausibly explain an association, including multiple chemokine receptors and genes involved in intracellular transport. Our meta-TWAS[ref erola] results show that variants in this region confer genome-wide significant differences in predicted expression of *CXCR6*, *CCR2* and *CCR3*. Meta-analysis of experimental data on betacoronavirus infection from other sources provides strongest support for *FYCO1*.[29]

The *ABO* locus was also previously associated with severe Covid-19,[11] but was not significant in our study (smallest p=1.30 x 10^{-3}, chr9:136115876_A/G; Supp Figure XXX). This does not rule out the possibility of a true association, but other possible explanations include differences between the control populations in each study.

Mendelian randomisation results suggest that increased expression of the interferon receptor subunit *IFNAR2* reduced the odds of severe Covid-19 with discovery p = 0.0043 (7 tests); replication p = 0.0066 (1 test). Within the assumptions of Mendelian randomisation, this represents evidence for a protective role for IFNAR2 in Covid-19. We deemed this gene to be therapeutically-informative *a priori* because it is a target for exogenous interferon treatment. Loss-of-function variants in *IFNAR2* have previously been associated with fatal sequelae from live-attenuated measles virus in humans[30,31] and influenza in mice.[32]

TYK2 is one of 4 gene products listed in the druggable genome Targets Central Resource Database[33] as a target for baricitinib, one of the nine candidate drugs we used in the creation of our *a priori* target list (Supplementary Table 1). However, since we did not *a priori* include *TYK2* on the final set of genes for focused Mendelian Randomisation, we use a significance threshold corrected for the full set of comparisons: discovery p = 0.00049 (4614 tests); replication p = 0.0022 (9 tests).

The *TYK2* locus includes multiple ICAM (intracellular adhesion molecules) genes which play key roles in the interaction between vascular endothelium and immune cells during adhesion and extravasation(Figure 3). Infiltration of immune cells into pulmonary vessel walls is characteristic of fatal

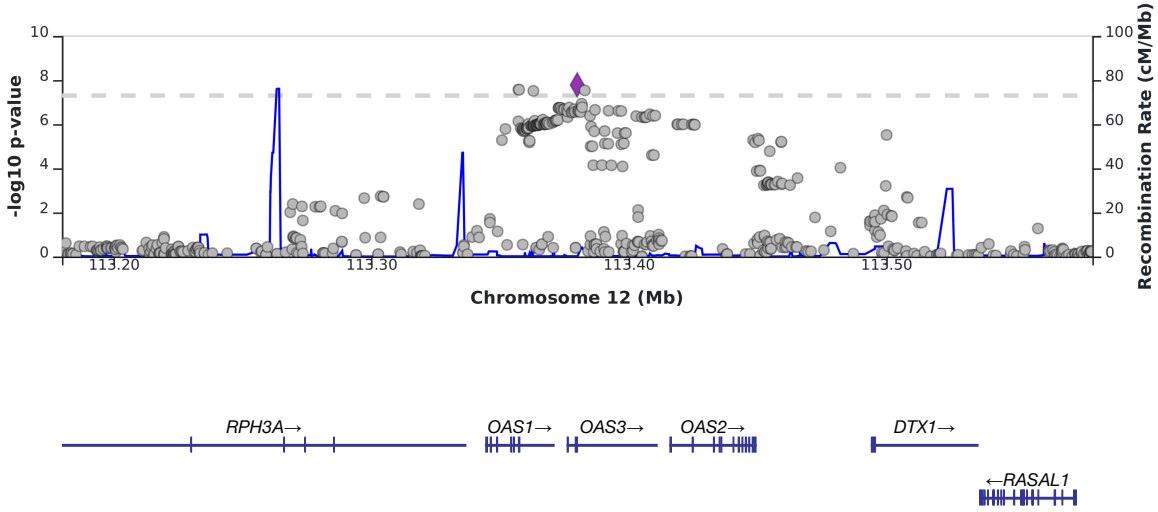

Figure 2: OAS gene cluster

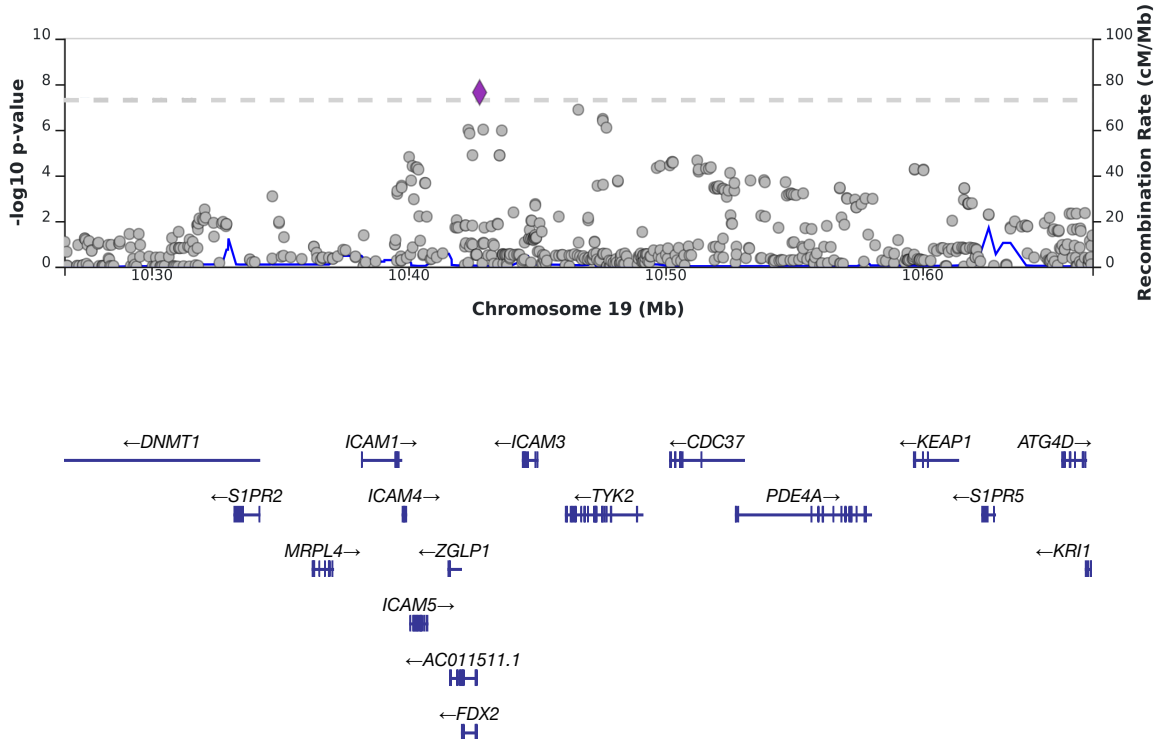

Figure 3: ICAM gene cluster

Covid-19.[4,5,34,35,36] The lead variant in the *ICAM* gene cluster is rs74956615(Figure 3), which is strongly-associated with multiple steroid-responsive autoimmune phenotypes[37] including ankylosing spondylitis, psoriasis, inflammatory bowel disease,[38] and rheumatoid arthritis.[39] A variant in *ICAM3* was associated with increased risk of SARS in Hong Kong.[40] The most abundantly-expressed gene in the region is *ICAM1*, which is involved in monocyte trans-endothelial trafficking,[41] which is a distinctive feature of fatal Covid-19.[5] It is strongly-expressed on the luminal surface of endothelial cells,[42] particularly in the pulmonary circulation.[43,44] Circulating ICAM1 levels are increased in Covid-19 and associated with worse disease.[45] Identifying the causative variant at this locus is an urgent priority for global Covid-19 research.

TWAS revealed a genome-wide significant association between predicted *CCR2* (CC-chemokine receptor 2) expression and life-threatening Covid-19(Figure **??**), particularly in lung tissue(Figure **??**). CCR2 promotes monocyte/macrophage chemotaxis towards sites of inflammation. There is increased expression of the canonical ligand for CCR2, monocyte chemoattractant protein (MCP-1), bronchoalveolar lavage fluid from the lungs of Covid-19 patients during mechanical ventilation,[46] and circulating MCP-1 concentrations are associated with more severe disease.[47] Anti-CCR2 monoclonal antibody therapy in treatment of rheumatoid arthritis is safe.[48]

Since translating these findings requires biologically-interpretable signals at the level of named genes, pathways and cell types, we performed post-GWAS analyses using MAGMA (Table **??**).[49] This reveals genome-wide significant association for the oligoadenylate synthetase (OAS) gene cluster (*OAS1*, *OAS2* and *OAS3*; {Figure 2}). OAS genes encode enzymes which activate an effector enzyme, RNAse L, which degrades double-stranded RNA,[50] a replication intermediate of coronaviruses.[51] *OAS1* variants were implicated in susceptibility to SARS-CoV in candidate gene association studies in Vietnam[52] and China.[53] The association signals at this locus did not replicate in external data(Table 2) and so we regard this result as preliminary until it is independently confirmed.

There is an urgent need to deepen these findings through further studies of this type, with harmonised integration across multiple studies. We continue to recruit to the GenOMICC study, in the expectation that additional associations exist and can be detected with larger numbers of cases. Our cohort is strongly enriched for immediately life-threatening disease in patients who are either receiving invasive mechanical ventilation, or considered by the treating physicians to be at high risk of requiring mechanical support. With 2790 cases we have statistical power to detect strong effects, such as the highly-significant locus at 3p21.31, as well as moderate genome-wide significant findings with external replication at *DPP9* and *IFNAR2*. Importantly, we cannot exclude therapeutically-informative effects at any locus on the genome: we can assert positive associations with quantifiable certainty, but we cannot determine the abscence of associations.

Because of the urgency of completing and reporting this work, we have drawn controls from population genetic studies who were genotyped using different technology from the cases. We mitigated the consequent risk of false-postive associations driven by genotyping errors by genotyping the majority of our subjects using two different methods (array+imputation, whole genome sequencing, agreement r=0.99), by drawing controls from three different population studies. The success of these mitigations is demonstrated by robust replication of our top hits in external studies.

We have discovered new and highly plausible genetic associations with critical illness in Covid-19. Some of these associations lead us to potential therapeutic approaches to augment interferon signalling, antagonise monocyte activation and infiltration into the lungs, or target anti-inflammatory pathways. This adds to the biological rationale underpinning therapeutic approaches. Where sufficient evidence exists that a given therapy has both a solid rationale and acceptable safety, each treatment must be tested in large-scale clinical trials before entering clinical practice.

Materials and methods

Recruitment

2,636 patients recruited to the GenOMICC study (genomicc.org) had confirmed Covid-19 according to local clinical testing and were deemed, in the view of the treating clinician, to require continuous cardiorespiratory monitoring. In UK practice this kind of monitoring is undertaken in high-dependency or intensive care units. An additional 134 patients were recruited through ISARIC 4C (isaric4c.net) - these individuals had confirmed Covid-19 according to local clinical testing and were deemed to require hospital admission.

Genotyping

DNA was extracted from whole blood using Nucleon Kit (Cytiva) with the BACC3 protocol. DNA samples were re-suspended in 1 ml TE buffer pH 7.5 (10mM Tris-Cl pH 7.5, 1mM EDTA pH 8.0). The yield of the DNA was measured using Qubit and normalised to 50ng/μl before genotyping.

Genotyping was performed using the Illumina Global Screening Array v3.0 + multi-disease beadchips (GSAMD-24v3-0-EA) and Infinium chemistry. In summary this consists of three steps: (1) whole genome amplification, (2) fragmentation followed by hybridisation, and (3) single-base extension and staining. For each of the samples, 4 μl of DNA normalised to 50ng/μl was used. Each sample was interrogated on the arrays against 730,059 SNPs. The Arrays were imaged on an Illumina iScan platform and genotypes were called automatically using GenomeStudio Analysis software v2.0.3, GSAMD-24v3-0-EA_20034606_A1.bpm manifest and cluster file provided by manufacturer.

In 1667 cases, genotypes and imputed variants were confirmed with Illumina NovaSeq 6000 whole genome sequencing. Samples were aligned to the human reference genome hg38 and variant called to GVCF stage on the DRAGEN pipeline (software v01.011.269.3.2.22, hardware v01.011.269) at Genomics England. Variants were genotyped with the GATK GenotypeGVCFs tool v4.1.8.1,[54] filtered to minimum depth 8X (95% sensitivity for heterozygous variant detection,[55]) merged and annotated with allele frequency with bcftools v1.10.2.

Quality control

Genotype calls were carefully examined within GenomeStudio using manufacturer and published[56] recommendations, after excluding samples with low initial call rate (<90%) and reclustering the data thereafter. Briefly, X and Y markers calls were all visually inspected and curated if necessary, as were those for autosomal markers with minor allele frequency > 1% displaying low Gentrain score, cluster separation, and excess or deficit of heterozygous calls. Genotype-based sex determination was performed in GenomeStudio and samples excluded if not matching records expectation. Five individuals with XXY genotypes were also detected and excluded for downstream GWAS analyses. Genotypes were exported, in genome reference consortium human build 37 (GRCHb37) and Illumina "source" strand orientation, using the GenotypeStudio `plink-input-report-plugin-v2-1-4`. A series of filtering steps was then applied using PLINK 1.9 leaving 2790 individuals and 479095 variants for further analyses (exclusion of samples with call rate < 95%, selection of variants with call rate > 99% and MAF > 1% and final samples selection using a call rate > 97%).

Kinship

Kinship and ancestry inference were calculated following UK Biobank[57] and 1M veteran program.[58] First King 2.1[59] was used to find duplicated individuals which have been recruited by two different routes. The analysis flagged 56 duplicated pairs, from which one was removed according to genotyping quality (GenomeStudio p50GC score or/and individual call rate). This leaves a set of 2734 unique individuals.

Regions of high LD defined in the UK Biobank[57] were excluded from the analysis, as well as SNPs with MAF<1% or missingness >1%. King 2.1 was used to construct a relationship matrix up to 3rd degree using the King command `--kinship --degree 3` and then the function `largest_independent_vertex_set()` from the `igraph` tool[[http://igraph.sf.net]] was used to create a first set of unrelated individuals. Principal component analysis (PCA) was conducted with gcta 1.9[60] in the set of unrelated individuals with pruned SNPs using a window of 1000 markers, a step size of 80 markers and an r^2 threshold of 0.1. SNPs with large weights in PC1, PC2 or PC3 were removed, keeping at least 2/3 of the number of pruned SNPs to keep as an input of the next round of King 2.1. The second round of King 2.1 was run using the SNPs with low weights in PC1, PC2 and PC3 to avoid overestimating kinship in non-european individuals. After this round 2718 individuals were considered unrelated up to 3rd degree.

Genetic ancestry

Unrelated individuals from the 1000 Genome Project dataset were calculated using the same procedure described above, and both datasets were merged using the common SNPs. The merged genotyped data was pruned with `plink` using a window of 1000 markers a step size of 50 and a r^2 of 0.05, leaving ~92K markers that were used to calculate the 20 first principal components with gcta 1.9. Ancestry for genomicc individuals was inferred using ADMIXTURE[61] populations defined in 1000 genomes. When one individual had a probability > 80% of pertaining to one ancestry, then the individual was assigned to this ancestry, otherwise the individual was assigned to admix ancestry as in the 1M veteran cohort.[58] According to this criterion there are 1818 individuals from European ancestry, 190 from African ancestry, 158 from East Asian ancestry, 254 from South Asian ancestry, and 301 individuals with admixed ancestry (2 or more).

Imputation

Genotype files were converted to plus strand and SNPs with Hardy-Weinberg Equilibrium (hwe) p-value<10^{-6} were removed. Imputation was calculated using the TOPMed reference panel.[18] and results were given in grch38 human reference genome and plus strand. The imputed dataset was filtered for monogenic and low imputation quality score (r^2<0.4) using BCFtools 1.9. To perform GWAS, files in VCF format were further filtered for r^2>0.9 and converted to BGEN format using QCtools 1.3.[62]

UK Biobank imputed variants with imputation score >0.9 and overlapping our set of variants (n=5,981,137) were extracted and merged with GenOMICC data into a single BGEN file containing cases and controls using QCtools 1.3.

GWAS

Individuals with a positive Covid-19 test or suspected Covid-19 when they were admitted in the hospital were included in the GWAS as cases. Related individuals to degree 3 were removed. 13 individuals with American ancestry were removed as the sample size provided insufficient power to perform a

reliable GWAS for this group. The final GWAS analysis includes 2244 individuals: 1676 individuals from European ancestry, 149 individuals from East asian ancestry, 237 individuals from South Asian ancestry and 182 individuals from African ancestry {tbl:baseline}. If age or deprivation status were missing for some individuals, the value was set to the mean of their ancestry.

Tests for association between case-control status and allele dosage at individuals SNPs were performed by fitting logistic regression models using PLINK.[63] Independent analyses were performed for each ethnic group. All models included sex, age, mean centered age squared, deprivation score decile of residential postcode, and the first 10 genomic principal components as covariates.

Genomic principal components were computed on the combined sample of all UK Biobank and GenOMICC participants. Specifically, 456,750 genetic variants were identified which were shared between the variants contained in the called genotypes in the GenOMICC dataset and imputed UK Biobank genotypes, which had an information score above 0.95 and a minor allele frequency above 1%. After merging genotypes at these variants, variants were removed which had a minor allele frequency below 2.5%, a missingness rate above 1.5%, showed departure from Hardy-Weinberg equilibrium with a p value below 10^{-50}, or which were within previously identified regions of high linkage disequilibrium within UK Biobank. After LD-pruning of the remaining variants to a maximum r^2 of 0.01 based on a 1000 variant window moving in 50 variants steps, using the PLINK indep-pairwise command and yielding 13,782 SNPs, the leading 20 genomic principal components were computed using FlashPCA2.[64]

GWAS results were filtered for maf>0.01, variant genotyping rate > 0.99 and hwe p-value > 10^{-150} for each ethnicity. An extra filter was added to avoid bias for using a different imputation panel between controls and cases. Minor allele frequencies (MAF) for each ancestry were compared between UK Biobank and gnomAD hg38 downloaded in August 2020.[28] SNPs were were removed from the GWAS results specifically for each ethnicity following these two rules: (a) In SNPs with MAF > 10% in gnomAD, an absolute difference of 5% between gnomAD and UK biobank controls MAF (b) In SNPs with MAF <10% in gnomAD, a difference > 25% gnomAD MAF, between UK Biobank controls and gnomAD. To calculate differences between UK Biobank European individuals and gnomAD allele frequencies, non Finnish-europeans gnomAD allele frequencues were used, as European UK Biobank controls are mainly non-Finnish.

Filtered GWAS for each ancestry, containing a total of ~4.7M SNPs, were combined in a trans-ethnic meta-analysis using METAL[65] standard error mode and controling for population stratification (genomiccontrol on).

Deprivation score The UK Data Service provides measures of deprivation based on Census Data and generated per postcode. The latest version of the Deprivation Scores were published in 2017 and are based on the 2011 census. Since only partial postcodes were available for most samples we were unable to use these indices directly. However, we generated an approximation to the scores by calculating an average weighted by population count across the top-level postcode areas.

The initial input file was part of the aggregated census data identified by DOI http://dx.doi.org/10.5257/census/aggregate-2011-2. Specifically the postcode data were downloaded from:

http://s3-eu-west-1.amazonaws.com/statistics.digitalresources.jisc.ac.uk/dkan/files/Postcode_Counts_and_Deprivation_Ranks/postcodes.zip

Population count and deprivation score for each published postcode were extracted and weighted average score calculated for each top-level postcode. We further categorised each top-level postcode score into decile and quintile bins for more coarse-grained analyses.

Whole Genome Sequencing

Whole Genome Sequencing (WGS) gVCF files were obtained for the 1667 individuals for which we had whole genome sequence data. Variants overlapping the positions of the imputed variants were called using GATk and variants with depth<8 (the minimum depth for which 95% coverage can be expected) were filtered. Individual VCF files were joined in a multi-sample VCF file for comparison with imputed variants. 1613 of these 1667 were used in the final GWAS. Samples were filtered and variants annotaed using bcftools 1.9. VCF files obtained from imputation were processed in an identical manner. Alternative allele frequency was calculated with PLINK 2.0[66] for both WGS and imputed data. From the 4469187 imputed variants that passed all filters after GWAS, 72658 did not pass QC filtering in WGS data and were removed. Further filtering of the data was applied, comparing the allele frequencies of each SNP between WGS and imputation, the correlation of allele frequencies was r^2=0.9994, and all variants with a difference of $> 5\%$ were removed from the analysis, leaving 4396207 imputed variants.

Controls

UK Biobank

UK Biobank participants were were considered as potential controls if they were not identified by the UK Biobank as outliers based on either genotyping missingness rate or heterogeneity, and their sex inferred from the genotypes matched their self-reported sex. For these individuals, information on sex (UKBID 31), age, ancestry, and residential postcode deprivation score decile was computed. Specifically, age was computed as age on April 1st, 2020 based on the participants birth month (UKBID 34) and year (UKBID 52). The first part of the residential postcode of participants was computed based on the participant's home location (UKBID 22702 and 22704) and mapped to a deprivation score decile as previously described for GenOMICC participants. Ancestry was inferred as previously described for GenOMICC participants.

After excluding participants who had received PCR tests for Covid-19, based on information downloaded from the UK Biobank in August 2020, five individuals with matching inferred ancestry were sampled for each GenOMICC participant as controls. After sampling each control, individuals related up to 3rd degree were removed from the pool of potential further controls.

Generation Scotland

Generation Scotland: Scottish Family Health Study (hereafter referred to as Generation Scotland) is a population-based cohort of 24 084 participants sampled from five regional centers across Scotland.[67][***http://www.generationscotland.org] A large subset of participants were genotyped using either Illumina HumanOmniExpressExome-8v1_A or v1-2, and 20 032 passed QC criteria previously described.[68] Genotype imputation using the TOPMed reference panel was recently performed (freeze 5b) using Minimac4 v1.0 on the University of Michigan serverhttps://imputationserver.sph.umich.edu.[69] Imputation data from XXXX unrelated (genomic sharing identical by descent estimated using PLINK1.9 $< 5\%$) participants were used as control genotypes in a GWAS using the XXXXX GenOMICC cases of European ancestry, for quality check purpose of associated variants. The GWAS was performed in a mixed linear model implemented by fastGWA[70] from the GCTA suite, [https://cnsgenomics.com/software/gcta/] fitting 10 first principal component coordinates computed using the SNP-loads from the 1000G+COVID principal component analysis, age and sex as fixed effects and a polygenic effect with relationship matrix for the merged cases and controls as a random effect.

Replication

No GWAS has been reported of critical illness or mortality in Covid-19. As a surrogate, to provide some replication for our findings, replication analyses were performed using Host Genetics Initiative build 37, version 2 (July 2020) B2 (hospitalised Covid-19 vs population) v2 GWAS. Summary statistics were used from the full analysis, including all cohorts and GWAS without UK Biobank, to avoid sample overlap. Replication p-value was set to $0.05/n$, where n is the number of loci significant in the discovery.

Meta-analysis

To extend the list of genes associated with Covid-19 fixed-variance inverse-variance meta-analysis of GenOMICC trans-ethnic GWAMA and Host Genetics Initiative build 37, version 2 (July 2020) B2 (hospitalised Covid-19 vs population) v2 was performed using METAL,[65] with correction for genomic inflation factor.

Post-GWAS analyses

Gene-level

Gene-level burden of significance in the EUR ancestry group result was calculated using MAGMA v1.08.[49] SNPs were annotated to genes by mapping based on genomic location. SNPs were assigned to a gene if the SNPs location is within 5 kb up- or down-stream of the gene region (defined as the transcription start site to transcription stop site). The MAGMA SNP-wise mean method was applied which utilises the sum of squared SNP Z-statistics as the test statistic. The 1000 Genomes Project European reference panel was used to estimate LD between SNPs.

Auxiliary files were downloaded from https://ctg.cncr.nl/software/magma on 1st September 2020. Gene location files for protein-coding genes were obtained from NCBI (ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO/Ma
on 29/04/2015 and ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/ARCHIVE/ANNOTATION_RELEASE.105/mapview/
on 25/05/2016). The reference data files used to estimate LD are derived from Phase 3 of the 1000 Genomes Project.

Competitive gene set enrichment analysis was conducted in MAGMA using a regression model that accounts for gene-gene correlations, to reduce bias resulting from clustering of functionally similar genes on the genome.[49] Gene sets were queried from the databases KEGG 2019, Reactome 2016, GO Biological Process 2018, Biocarta 2016 and WikiPathways 2019. The Benjamini-Hochberg procedure was used to control false discovery rate (<0.05).

TWAS

We performed transcriptome-wide association using the MetaXcan framework[71] and the GTExv8 eQTL MASHR-M models available for download (http://predictdb.org/) . First GWAS results were harmonised, lifted over to hg38 and linked to 1000 Genomes reference panel using GWAS tools (https://github.com/hakyimlab/summary-gwas-imputation/wiki/GWAS-Harmonization-And-Imputation). TWAS for whole blood and lung were calculated using GWAS summary statistics for the European population GWAS and S-PrediXcan. Resulting p-values were corrected using the Bonferroni correction to find significant gene associations.

Mendelian randomisation

Two-sample Summary data based Mendelian randomisation [PMID 27019110] was performed using the results of GenOMICC and the Genotype-Tissue expression project, GTEx v7 [PMID 29022597], with Generation Scotland [PMID 22786799; PMID 17014726] forming a linkage disequilibrium reference. GenOMICC results from those of European ancestry were used as the outcome; and GTEx (v7) whole blood expression results as the exposure. Data pertaining to GTEx v7 were downloaded from the GTEx portal - https://gtexportal.org/ (accessed 20 Feb 2020, 05 Apr 2020, and 04 Jul 2020), and SMR/HEIDI from https://cnsgenomics.com/software/smr/ (accessed 03 Jul 2020). Analyses were conducted using Python 3.7.3 and SMR/HEIDI v1.03. An LD reference was created using data from the population-based Generation Scotland cohort (used with permission; described previously [PMID 28270201]): from a random set of 5,000 individuals, using Plink v1.9 (www.cog-genomics.org/plink/1.9/), a set of individuals with a genomic relatedness cutoff < 0.01 was extracted; 2,778 individuals remained in the final set. All data used for the SMR/HEIDI analyses were limited to autosomal biallelic SNPs: 4,264,462 variants remained in the final merged dataset.

Significant (as per GTEx v7; nominal p-value below nominal p-value threshold) local (distance to transcriptional start site $<$ 1Mb) eQTL from GTEx v7 whole blood for protein coding genes (as per GENCODE v19) with a MAF > 0.01 (GTEx v7 and GenOMICC) were considered as potential instrumental variables. Per variant, we first selected the Ensembl gene ID to which it was most strongly associated (so as to ensure that each variant can only be considered as an instrument for the gene to which it is most strongly associated) followed by selecting the variant to which each Ensembl gene ID was most strongly associated. Instruments were available for 4,614 unique Ensembl gene IDs.

Results were assessed based upon a list of genes selected *a priori* as of interest (TABLE XXX), and together as a whole.

Partial replication of Bonferroni-corrected significant results was attempted in the results of COVID19-Host Genetics Initiative - https://www.covid19hg.org/ - with UK Biobank excluded (accessed 21 Sep 2020). Hospitalized covid vs. population (ANA_B2_V2) was selected as the phenotype most similar to our own, and therefore the most appropriate for use as a replication cohort. This is not a complete replication - due to the repeated use of GTEX v7 Whole Blood results in both analyses - yet remains informative as to the strength of assoiation between the genetic variant and COVID19, and repeated consistency with a non-zero Mendelian Randomisation effect-size estimate.

Meta-analysis by information content (MAIC)

Multiple *in vitro* and *in vivo* studies have identified key host genes that either directly interact with SARS-CoV-2, or define the host response to SARS-CoV-2. We have previously reported a systematic review of these studies.[29] In order to put the new associations from this GWAS into context, we performed a data-driven meta-analysis of gene-level results combined with pre-existing biological data using meta-analysis by information content (MAIC).[72] Briefly, MAIC combines experimental results from diverse sources in the form of ranked or unranked gene lists. The algorithm assigns a weighting to each input gene list, derived from the degree of overlap with other input lists. Each gene is then assigned a score calculated from the weightings for each gene list on which it appears. This process is repeated iteratively until all scores converge on a stable value. In order to prevent a single type of experiment from unduly biasing the results, input gene lists are assigned to categories, and a rule applied that only one weighting from each category can contribute to the score for any given gene.

Cell-type enrichment

Genetic correlation with other disease phenotypes

Genome build

Results are presented using Genome Reference Consortium Human Build 37. Imputed genotypes and whole-genome sequence data were lifted over from Genome Reference Consortium Human Build 38 using Picard liftoverVCF mode from GATK 4.0 which is based on the UCSC liftover tool(chain file obtained from ftp://ftp.ensembl.org/pub/assembly_mapping/homo_sapiens/GRCh38_to_GRCh37.chain.gz.[73]

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 we have ever witnessed in UK hospitals.

This research has been conducted using the UK Biobank Resource under project 788.

GenOMICC was funded by Sepsis Research (the Fiona Elizabeth Agnew Trust), the Intensive Care Society, a Wellcome-Beit Prize award to J. K. Baillie (Wellcome Trust 103258/Z/13/A) and a BBSRC Institute Program Support Grant to the Roslin Institute (BBS/E/D/20002172). Whole-genome sequencing was done in partnership with Genomics England and was funded by UK Department of Health and Social Care, UKRI and LifeArc. The ISARIC 4C study is funded by UKRI MC_PC_19059.

This study owes a great deal to the National Institute of Healthcare Research Clinical Research Network (NIHR CRN) and the Chief Scientist Office (Scotland), who facilitate recruitment into research studies in NHS hospitals, and to the global ISARIC and InFACT consortia.

References

1.Carsana, L., Sonzogni, A., Nasr, A., Rossi, R.S., Pellegrinelli, A., Zerbi, P., Rech, R., Colombo, R., Antinori, S., Corbellino, M., Galli, M., Catena, E., Tosoni, A., Gianatti, A. & Nebuloni, M. Pulmonary post-mortem findings in a series of covid-19 cases from northern italy: A two-centre descriptive study. *The Lancet. Infectious diseases* (2020).doi:10.1016/S1473-3099(20)30434-5

2.Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., Elmahi, E., Prudon, B., Green, C., Felton, T., Chadwick, D., Rege, K., Fegan, C., Chappell, L.C., Faust, S.N., Jaki, T., Jeffery, K., Montgomery, A., Rowan, K., Juszczak, E., Baillie, J.K., Haynes, R. & Landray, M.J. Dexamethasone in hospitalized patients with covid-19 - preliminary report. *The New England journal of medicine* (2020).doi:10.1056/NEJMoa2021436

3.Baillie, J.K. Translational genomics. Targeting the host immune response to fight infection. *Science (New York, N.Y.)* **344**, 807–8(2014).

4.Carvelli, J., Demaria, O., Vély, F., Batista, L., Benmansour, N.C., Fares, J., Carpentier, S., Thibult, M.-L., Morel, A., Remark, R., André, P., Represa, A., Piperoglou, C., Cordier, P.Y., Le Dault, E., Guervilly, C., Simeone, P., Gainnier, M., Morel, Y., Ebbo, M., Schleinitz, N. & Vivier, E. Association of covid-19 inflammation with activation of the c5a-c5ar1 axis. *Nature* (2020).doi:10.1038/s41586-020-2600-6

5.Dorward, D.A., Russell, C.D., Um, I.H., Elshani, M., Armstrong, S.D., Penrice-Randal, R., Millar, T., Lerpiniere, C.E., Tagliavini, G., Hartley, C.S., Randall, N.P., Gachanja, N.N., Potey, P.M., Anderson, A.M., Campbell, V.L., Duguid, A.J., Qsous, W.A., BouHaidar, R., Baillie, J.K., Dhaliwal, K., Wallace, W.A., Bellamy, C.O., Prost, S., Smith, C., Hiscox, J.A., Harrison, D.J., Lucas, C.D. & ICECAP Tissue-specific tolerance in fatal Covid-19. *medRxiv* 2020.07.02.20145003(2020).doi:10.1101/2020.07.02.20145003

6.Sørensen, T.I., Nielsen, G.G., Andersen, P.K. & Teasdale, T.W. Genetic and environmental influences on premature death in adult adoptees. *The New England journal of medicine* **318**, 727–32(1988).

7.Patarčić, I., Gelemanović, A., Kirin, M., Kolčić, I., Theodoratou, E., Baillie, K.J., Jong, M.D. de, Rudan, I., Campbell, H. & Polašek, O. The role of host genetic factors in respiratory tract infectious diseases: Systematic review, meta-analyses and field synopsis. *Scientific reports* **5**, 16119(2015).

8.Horby, P., Nguyen, N.Y., Dunstan, S.J. & Baillie, J.K. An updated systematic review of the role of host genetics in susceptibility to influenza. *Influenza and other respiratory viruses* **7 Suppl 2**, 37–41(2013).

9.Clohisey, S. & Baillie, J.K. Host susceptibility to severe influenza A virus infection. *Critical Care* **23**, 303(2019).

10.Shelton, J.F., Shastri, A.J., Ye, C., Weldon, C.H., Filshtein-Somnez, T., Coker, D., Symons, A., Esparza-Gordillo, J., Team, T.2.C.-1., Aslibekyan, S. & Auton, A. Trans-ethnic analysis reveals genetic and non-genetic associations with COVID-19 susceptibility and severity. *medRxiv* 2020.09.04.20188318(2020).doi:10.1101/2020.09.04.20188318

11.Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M.M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., Yi, X., Pablo, R. de, Chercoles, A.G., Palom, A., Garcia-Fernandez, A.-E., Rodriguez-Frias, F., Zanella, A., Bandera, A., Protti, A., Aghemo, A., Lleo, A., Biondi, A., Caballero-Garralda, A., Gori, A., Tanck, A., Carreras Nolla, A., Latiano, A., Fracanzani, A.L., Peschuck, A., Julià, A., Pesenti, A., Voza, A., Jiménez, D., Mateos, B., Nafria Jimenez, B., Quereda, C., Paccapelo, C., Gassner, C., Angelini, C., Cea, C., Solier, A., Pestaña, D., Muñiz-Diaz, E., Sandoval, E., Paraboschi, E.M., Navas, E., García Sánchez, F., Ceriotti, F., Martinelli-Boneschi, F., Peyvandi, F., Blasi, F., Téllez, L., Blanco-Grau, A., Hemmrich-Stanisak, G., Grasselli, G., Costantino, G., Cardamone, G., Foti, G., Aneli, S., Kurihara, H., ElAbd, H., My, I., Galván-Femenia, I., Martín, J., Erdmann, J., Ferrusquía-Acosta, J., Garcia-Etxebarria, K., Izquierdo-Sanchez, L., Bettini, L.R., Sumoy, L., Terranova, L., Moreira, L., Santoro, L., Scudeller, L., Mesonero, F., Roade, L., Rühlemann, M.C., Schaefer, M., Carrabba, M., Riveiro-Barciela, M., Figuera Basso, M.E., Valsecchi, M.G., Hernandez-Tejero, M., Acosta-Herrera, M., D'Angiò, M., Baldini, M., Cazzaniga, M., Schulzky, M., Cecconi, M., Wittig, M., Ciccarelli, M., Rodríguez-Gandía, M., Bocciolone, M., Miozzo, M., Montano, N., Braun, N., Sacchi, N., Martínez, N., Özer, O., Palmieri, O., Faverio, P., Preatoni, P., Bonfanti, P., Omodei, P., Tentorio, P., Castro, P., Rodrigues, P.M., Blandino Ortiz, A., Cid, R. de, Ferrer, R., Gualtierotti, R., Nieto, R., Goerg, S., Badalamenti, S., Marsal, S., Matullo, G., Pelusi, S., Juzenas, S., Aliberti, S., Monzani, V., Moreno, V., Wesse, T., Lenz, T.L., Pumarola, T., Rimoldi, V., Bosari, S., Albrecht, W., Peter, W., Romero-Gómez, M., D'Amato, M., Duga, S., Banales, J.M., Hov, J.R., Folseraas, T., Valenti, L., Franke, A. & Karlsen, T.H. Genomewide association study of severe covid-19 with respiratory failure. *The New England journal of medicine* (2020).doi:10.1056/NEJMoa2020283

12.Casanova, J.-L. Severe infectious diseases of childhood as monogenic inborn errors of immunity. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E7128–37(2015).

13.Plenge, R.M. Molecular underpinnings of severe coronavirus disease 2019. *JAMA* (2020).doi:10.1001/jama.2020.14015

14.Millar, J.E., Neyton, L., Seth, S., Dunning, J., Merson, L., Murthy, S., Russell, C.D., Keating, S., Swets, M., Sudre, C.H., Spector, T.D., Ourselin, S., Steves, C.J., Wolf, J., In-

vestigators, I., Docherty, A.B., Harrison, E.M., Openshaw, P.J., Semple, M.G. & Baillie, J.K. Robust, reproducible clinical patterns in hospitalised patients with COVID-19. *medRxiv* 2020.08.14.20168088(2020).doi:10.1101/2020.08.14.20168088

15.Docherty, A.B., Harrison, E.M., Green, C.A., Hardwick, H.E., Pius, R., Norman, L., Holden, K.A., Read, J.M., Dondelinger, F., Carson, G., Merson, L., Lee, J., Plotkin, D., Sigfrid, L., Halpin, S., Jackson, C., Gamble, C., Horby, P.W., Nguyen-Van-Tam, J.S., Ho, A., Russell, C.D., Dunning, J., Openshaw, P.J., Baillie, J.K. & Semple, M.G. Features of 200.167em133 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).

16.Angus, D.C., Derde, L., Al-Beidh, F., Annane, D., Arabi, Y., Beane, A., Bentum-Puijk, W. van, Berry, L., Bhimani, Z., Bonten, M., Bradbury, C., Brunkhorst, F., Buxton, M., Buzgau, A., Cheng, A.C., Jong, M. de, Detry, M., Estcourt, L., Fitzgerald, M., Goossens, H., Green, C., Haniffa, R., Higgins, A.M., Horvat, C., Hullegie, S.J., Kruger, P., Lamontagne, F., Lawler, P.R., Linstrum, K., Litton, E., Lorenzi, E., Marshall, J., McAuley, D., McGlothin, A., McGuinness, S., McVerry, B., Montgomery, S., Mouncey, P., Murthy, S., Nichol, A., Parke, R., Parker, J., Rowan, K., Sanil, A., Santos, M., Saunders, C., Seymour, C., Turner, A., Veerdonk, F. van de, Venkatesh, B., Zarychanski, R., Berry, S., Lewis, R.J., McArthur, C., Webb, S.A. & Gordon, A.C. Effect of hydrocortisone on mortality and organ support in patients with severe covid-19: The remap-cap covid-19 corticosteroid domain randomized clinical trial. *JAMA* (2020).doi:10.1001/jama.2020.17022

17.Baillie, J.K. & Digard, P. Influenza Time to Target the Host? *New England Journal of Medicine* **369**, 191–193(2013).

18.D, T., Dn, H., Md, K., J, C., Za, S., R, T., Sag, T., A, C., Sm, G., Hm, K., An, P., J, L., S, L., X, T., Bl, B., S, D., A, E., We, C., Dp, L., Ac, S., Tw, B., Q, W., F, A., C, A., A, A., Kg, A., S, A., Pl, A., J, B., Rg, B., Lc, B., Rl, B., Ej, B., Lf, B., J, B., M, B., Dw, B., Ja, B., Eg, B., Be, C., Jf, C., B, C., Yi, C., Mh, C., Sh, C., Mk, C., Cb, C., A, C., Je, C., B, C., D, D., M, D., Md, A., Dl, D., Sk, D., Pt, E., Ls, E., D, F., L, F., M, F., N, F., C, F., Sm, F., S, G., Mt, G., Dj, G., X, G., Me, H., J, H., Nl, H.-C., Sr, H., Mr, I., Jm, J., Ad, J., Sl, K., T, K., S, K., Ee, K., Dp, K., R, K., Ba, K., C, K., A, K., La, L., J, L.-S., D, L., X, L., K, L., C, L., Rj, L., L, G., R, G., Sa, L., Kl, L., Ac, M., A, M., Ak, M., Ra, M., Dd, M., St, M., Jb, M., Da, M., Jl, M., Ma, M., B, M., S, M., Me, M., C, M., Ac, M., Jm, M., A, N., P, N., Sc, N., Ke, N., Jr, O., Nd, P., N, P., Gm, P., Pa, P., Ws, P., Bm, P., D, R., S, R., Ap, R., D, R., Ji, R., I, R., C, S., S, S., J, S., S, S., Va, S., Mb, S., Av, S., Nl, S., Ja, S., N, S., Am, S., W, T., Kd, T., M, T., Ta, T., Rp, T., Djvd, B., Rs, V., Ka, V.-M., S, V., De, W., Bs, W., St, W., L, W., Cj, W., Y, Z., X, Z., Dk, A., Ae, A.-K., Kc, B., E, B., S, G., R, G., Km, R., Ss, R., E, S., P, Q., W, G., Gj, P., Da, N., Sr, B., Mc, Z., S, Z., Jg, W., La, C., Cc, L., Ce, J., Rd, H., Td, O., Gr, A. & undefined Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. (2019).doi:10.1101/563866

19.Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., Geus, E.J.C. de, Boomsma, D.I., Wright, F.A., Sullivan, P.F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A.J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O.T., Kuusisto, J., Laakso, M., Price, A.L., Pajukanta, P. & Pasaniuc, B. Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics* **48**, 245–52(2016).

20.Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J. & Im, H.K. A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091–8(2015).

21.Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nature genetics* **52**, 859–864(2020).

22.Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F.O., Jørgensen, M., Andersen, P.R., Bertin, N., Rackham, O.,

Burroughs, A.M., Baillie, J.K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C.J., Meehan, T.F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C.O., Heutink, P., Hume, D.A., Jensen, T.H., Suzuki, H., Hayashizaki, Y., Müller, F., Consortium, T.F., Forrest, A.R.R., Carninci, P., Rehli, M. & Sandelin, A. An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461(2014).

23.Villar, D., Berthelot, C., Aldridge, S., Rayner, T.F., Lukk, M., Pignatelli, M., Park, T.J., Deaville, R., Erichsen, J.T., Jasinska, A.J., Turner, J.M.A., Bertelsen, M.F., Murchison, E.P., Flicek, P. & Odom, D.T. Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–66(2015).

24.Wunsch, H., Linde-Zwirble, W.T., Angus, D.C., Hartman, M.E., Milbrandt, E.B. & Kahn, J.M. The epidemiology of mechanical ventilation use in the united states. *Critical care medicine* **38**, 1947–53(2010).

25.Fingerlin, T.E., Murphy, E., Zhang, W., Peljto, A.L., Brown, K.K., Steele, M.P., Loyd, J.E., Cosgrove, G.P., Lynch, D., Groshong, S., Collard, H.R., Wolters, P.J., Bradford, W.Z., Kossen, K., Seiwert, S.D., Bois, R.M. du, Garcia, C.K., Devine, M.S., Gudmundsson, G., Isaksson, H.J., Kaminski, N., Zhang, Y., Gibson, K.F., Lancaster, L.H., Cogan, J.D., Mason, W.R., Maher, T.M., Molyneaux, P.L., Wells, A.U., Moffatt, M.F., Selman, M., Pardo, A., Kim, D.S., Crapo, J.D., Make, B.J., Regan, E.A., Walek, D.S., Daniel, J.J., Kamatani, Y., Zelenika, D., Smith, K., McKean, D., Pedersen, B.S., Talbert, J., Kidd, R.N., Markin, C.R., Beckman, K.B., Lathrop, M., Schwarz, M.I. & Schwartz, D.A. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nature genetics* **45**, 613–20(2013).

26.Allen, R.J., Guillen-Guio, B., Oldham, J.M., Ma, S.-F., Dressen, A., Paynton, M.L., Kraven, L.M., Obeidat, M., Li, X., Ng, M., Braybrooke, R., Molina-Molina, M., Hobbs, B.D., Putman, R.K., Sakornsakolpat, P., Booth, H.L., Fahy, W.A., Hart, S.P., Hill, M.R., Hirani, N., Hubbard, R.B., McAnulty, R.J., Millar, A.B., Navaratnam, V., Oballa, E., Parfrey, H., Saini, G., Whyte, M.K.B., Zhang, Y., Kaminski, N., Adegunsoye, A., Strek, M.E., Neighbors, M., Sheng, X.R., Gudmundsson, G., Gudnason, V., Hatabu, H., Lederer, D.J., Manichaikul, A., Newell, J.D., O'Connor, G.T., Ortega, V.E., Xu, H., Fingerlin, T.E., Bossé, Y., Hao, K., Joubert, P., Nickle, D.C., Sin, D.D., Timens, W., Furniss, D., Morris, A.P., Zondervan, K.T., Hall, I.P., Sayers, I., Tobin, M.D., Maher, T.M., Cho, M.H., Hunninghake, G.M., Schwartz, D.A., Yaspan, B.L., Molyneaux, P.L., Flores, C., Noth, I., Jenkins, R.G. & Wain, L.V. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine* **201**, 564–574(2020).

27.Zhang, H., Maqsudi, S., Rainczuk, A., Duffield, N., Lawrence, J., Keane, F.M., Justa-Schuch, D., Geiss-Friedlander, R., Gorrell, M.D. & Stephens, A.N. Identification of novel dipeptidyl peptidase 9 substrates by two-dimensional differential in-gel electrophoresis. *The FEBS journal* **282**, 3737–57(2015).

28.Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Neale, B.M., Daly, M.J. & MacArthur, D.G. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443(2020).

29.Parkinson, N., Rodgers, N., Fourman, M.H., Wang, B., Zechner, M., Swets, M., Millar, J.E., Law, A., Russell, C., Baillie, J.K. & Clohisey, S. Systematic review and meta-analysis identifies potential host

therapeutic targets in COVID-19. *medRxiv* 2020.08.27.20182238(2020).doi:10.1101/2020.08.27.20182238

30.Duncan, C.J.A., Mohamad, S.M.B., Young, D.F., Skelton, A.J., Leahy, T.R., Munday, D.C., Butler, K.M., Morfopoulou, S., Brown, J.R., Hubank, M., Connell, J., Gavin, P.J., McMahon, C., Dempsey, E., Lynch, N.E., Jacques, T.S., Valappil, M., Cant, A.J., Breuer, J., Engelhardt, K.R., Randall, R.E. & Hambleton, S. Human ifnar2 deficiency: Lessons for antiviral immunity. *Science translational medicine* **7**, 307ra154(2015).

31.Hambleton, S., Goodbourn, S., Young, D.F., Dickinson, P., Mohamad, S.M.B., Valappil, M., McGovern, N., Cant, A.J., Hackett, S.J., Ghazal, P., Morgan, N.V. & Randall, R.E. STAT2 deficiency and susceptibility to viral illness in humans. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3053–8(2013).

32.Shepardson, K.M., Larson, K., Johns, L.L., Stanek, K., Cho, H., Wellham, J., Henderson, H. & Rynda-Apple, A. IFNAR2 is required for anti-influenza immunity and alters susceptibility to post-influenza bacterial superinfections. *Frontiers in immunology* **9**, 2589(2018).

33.Nguyen, D.-T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A., Liu, G., Ma'ayan, A., Mandava, G., Mani, S., Mehta, S., Overington, J., Patel, J., Rouillard, A.D., Schürer, S., Sheils, T., Simeonov, A., Sklar, L.A., Southall, N., Ursu, O., Vidovic, D., Waller, A., Yang, J., Jadhav, A., Oprea, T.I. & Guha, R. Pharos: Collating protein information to shed light on the druggable genome. *Nucleic acids research* **45**, D995–D1002(2017).

34.Ackermann, M., Verleden, S.E., Kuehnel, M., Haverich, A., Welte, T., Laenger, F., Vanstapel, A., Werlein, C., Stark, H., Tzankov, A., Li, W.W., Li, V.W., Mentzer, S.J. & Jonigk, D. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19. *The New England journal of medicine* **383**, 120–128(2020).

35.Menter, T., Haslbauer, J.D., Nienhold, R., Savic, S., Hopfer, H., Deigendesch, N., Frank, S., Turek, D., Willi, N., Pargger, H., Bassetti, S., Leuppi, J.D., Cathomas, G., Tolnay, M., Mertz, K.D. & Tzankov, A. Postmortem examination of covid-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology* (2020).doi:10.1111/his.14134

36.Barnes, B.J., Adrover, J.M., Baxter-Stoltzfus, A., Borczuk, A., Cools-Lartigue, J., Crawford, J.M., Daßler-Plenker, J., Guerci, P., Huynh, C., Knight, J.S., Loda, M., Looney, M.R., McAllister, F., Rayes, R., Renaud, S., Rousseau, S., Salvatore, S., Schwartz, R.E., Spicer, J.D., Yost, C.C., Weber, A., Zuo, Y. & Egeblad, M. Targeting potential drivers of covid-19: Neutrophil extracellular traps. *The Journal of experimental medicine* **217**, (2020).

37.Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P.L., Amode, R., Guillen, J.A., Riat, H.S., Trevanion, S.J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorff, L.A., Cunningham, F. & Parkinson, H. The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research* **47**, D1005–D1012(2019).

38.Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hübenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H.-J., Franke, L., Pers, T.H., Weersma, R.K., Collij, V., D'Amato, M., Halfvarson, J., Jensen, A.B., Lieb, W., Degenhardt, F., Forstner, A.J., Hofmann, A., Schreiber, S., Mrowietz, U., Juran, B.D., Lazaridis, K.N., Brunak, S., Dale, A.M., Trembath, R.C., Weidinger, S., Weichenthal, M., Ellinghaus, E., Elder, J.T., Barker, J.N.W.N., Andreassen, O.A., McGovern, D.P., Karlsen, T.H., Barrett, J.C., Parkes, M., Brown, M.A. & Franke, A. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature genetics* **48**, 510–8(2016).

39.Márquez, A., Kerick, M., Zhernakova, A., Gutierrez-Achury, J., Chen, W.-M., Onengut-Gumuscu, S., González-Álvaro, I., Rodriguez-Rodriguez, L., Rios-Fernández, R., González-Gay, M.A., Mayes,

M.D., Raychaudhuri, S., Rich, S.S., Wijmenga, C. & Martín, J. Meta-analysis of immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome medicine* **10**, 97(2018).

40.Chan, K.Y.K., Ching, J.C.Y., Xu, M.S., Cheung, A.N.Y., Yip, S.-P., Yam, L.Y.C., Lai, S.-T., Chu, C.-M., Wong, A.T.Y., Song, Y.-Q., Huang, F.-P., Liu, W., Chung, P.H., Leung, G.M., Chow, E.Y.D., Chan, E.Y.T., Chan, J.C.K., Ngan, H.Y.S., Tam, P., Chan, L.-C., Sham, P., Chan, V.S.F., Peiris, M., Lin, S.C.L. & Khoo, U.-S. Association of icam3 genetic variant with severe acute respiratory syndrome. *The Journal of infectious diseases* **196**, 271–80(2007).

41.Gerhardt, T. & Ley, K. Monocyte trafficking across the vessel wall. *Cardiovascular research* **107**, 321–30(2015).

42.Marlin, S.D. & Springer, T.A. Purified intercellular adhesion molecule-1 (icam-1) is a ligand for lymphocyte function-associated antigen 1 (lfa-1). *Cell* **51**, 813–9(1987).

43.Lu, Y.-T., Chen, P.-G. & Liu, S.F. Time course of lung ischemia-reperfusion-induced icam-1 expression and its role in ischemia-reperfusion lung injury. *Journal of applied physiology (Bethesda, Md. : 1985)* **93**, 620–8(2002).

44.Beck-Schimmer, B., Schimmer, R.C., Warner, R.L., Schmal, H., Nordblom, G., Flory, C.M., Lesch, M.E., Friedl, H.P., Schrier, D.J. & Ward, P.A. Expression of lung vascular and airway icam-1 after exposure to bacterial lipopolysaccharide. *American journal of respiratory cell and molecular biology* **17**, 344–52(1997).

45.Tong, M., Jiang, Y., Xia, D., Xiong, Y., Zheng, Q., Chen, F., Zou, L., Xiao, W. & Zhu, Y. Elevated expression of serum endothelial cell adhesion molecules in covid-19 patients. *The Journal of infectious diseases* **222**, 894–898(2020).

46.Zhou, Z., Ren, L., Zhang, L., Zhong, J., Xiao, Y., Jia, Z., Guo, L., Yang, J., Wang, C., Jiang, S., Yang, D., Zhang, G., Li, H., Chen, F., Xu, Y., Chen, M., Gao, Z., Yang, J., Dong, J., Liu, B., Zhang, X., Wang, W., He, K., Jin, Q., Li, M. & Wang, J. Heightened innate immune responses in the respiratory tract of covid-19 patients. *Cell host & microbe* **27**, 883–890.e2(2020).

47.Zhao, Y., Qin, L., Zhang, P., Li, K., Liang, L., Sun, J., Xu, B., Dai, Y., Li, X., Zhang, C., Peng, Y., Feng, Y., Li, A., Hu, Z., Xiang, H., Ogg, G., Ho, L.-P., McMichael, A., Jin, R., Knight, J.C., Dong, T. & Zhang, Y. Longitudinal covid-19 profiling associates il-1RA and il-10 with disease severity and rantes with mild disease. *JCI insight* **5**, (2020).

48.Vergunst, C.E., Gerlag, D.M., Lopatinskaya, L., Klareskog, L., Smith, M.D., Bosch, F. van den, Dinant, H.J., Lee, Y., Wyant, T., Jacobson, E.W., Baeten, D. & Tak, P.P. Modulation of ccr2 in rheumatoid arthritis: A double-blind, randomized, placebo-controlled clinical trial. *Arthritis and rheumatism* **58**, 1931–9(2008).

49.Leeuw, C.A. de, Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized gene-set analysis of gwas data. *PLoS computational biology* **11**, e1004219(2015).

50.Choi, U.Y., Kang, J.-S., Hwang, Y.S. & Kim, Y.-J. Oligoadenylate synthase-like (oasl) proteins: Dual functions and associations with diseases. *Experimental & molecular medicine* **47**, e144(2015).

51.Hagemeijer, M.C., Vonk, A.M., Monastyrska, I., Rottier, P.J.M. & Haan, C.A.M. de Visualizing coronavirus rna synthesis in time by using click chemistry. *Journal of virology* **86**, 5808–16(2012).

52.Hamano, E., Hijikata, M., Itoyama, S., Quy, T., Phi, N.C., Long, H.T., Ha, L.D., Ban, V.V., Matsushita, I., Yanai, H., Kirikae, F., Kirikae, T., Kuratsuji, T., Sasazuki, T. & Keicho, N. Polymorphisms of interferon-inducible genes oas-1 and mxa associated with sars in the vietnamese population. *Biochemical and biophysical research communications* **329**, 1234–9(2005).

53.He, J., Feng, D., Vlas, S.J. de, Wang, H., Fontanet, A., Zhang, P., Plancoulaine, S., Tang, F., Zhan, L., Yang, H., Wang, T., Richardus, J.H., Habbema, J.D.F. & Cao, W. Association of sars susceptibility with single nucleic acid polymorphisms of oas1 and mxa genes: A case-control study. *BMC infectious diseases* **6**, 106(2006).

54.McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M.A. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome research* **20**, 1297–303(2010).

55.Meynert, A.M., Ansari, M., FitzPatrick, D.R. & Taylor, M.S. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC bioinformatics* **15**, 247(2014).

56.Guo, Y., He, J., Zhao, S., Wu, H., Zhong, X., Sheng, Q., Samuels, D.C., Shyr, Y. & Long, J. Illumina human exome genotyping array clustering and quality control. *Nature protocols* **9**, 2643–62(2014).

57.Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T. & Collins, R. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* **12**, e1001779(2015).

58.Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., Guarino, P., Aslan, M., Anderson, D., LaFleur, R., Hammond, T., Schaa, K., Moser, J., Huang, G., Muralidhar, S., Przygodzki, R. & O'Leary, T.J. Million veteran program: A mega-biobank to study genetic influences on health and disease. *Journal of clinical epidemiology* **70**, 214–23(2016).

59.Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. & Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics (Oxford, England)* **26**, 2867–73(2010).

60.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).

61.Alexander, D.H. & Lange, K. Enhancements to the admixture algorithm for individual ancestry estimation. *BMC bioinformatics* **12**, 246(2011).

62.Wigginton, J.E., Cutler, D.J. & Abecasis, G.R. A note on exact tests of hardy-weinberg equilibrium. *American journal of human genetics* **76**, 887–93(2005).

63.Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. Second-generation plink: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7(2015).

64.Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: Principal component analysis of biobank-scale genotype datasets. *Bioinformatics (Oxford, England)* **33**, 2776–2778(2017).

65.Willer, C.J., Li, Y. & Abecasis, G.R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190–1(2010).

66.Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W. de, Daly, M.J. & Sham, P.C. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–75(2007).

67.Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., Deary, I.J., Macintyre, D.J., Campbell, H., McGilchrist, M., Hocking, L.J., Wisely, L., Ford, I., Lindsay, R.S., Morton, R., Palmer, C.N.A., Dominiczak, A.F., Porteous, D.J. & Morris, A.D. Cohort profile: Generation scotland: Scottish family health study (gs:SFHS). The study, its participants and their potential for genetic research on health and illness. *International journal of epidemiology* **42**, 689–700(2013).

68.Amador, C., Huffman, J., Trochet, H., Campbell, A., Porteous, D., Wilson, J.F., Hastie, N., Vitart, V., Hayward, C., Navarro, P. & Haley, C.S. Recent genomic heritage in scotland. *BMC genomics* **16**, 437(2015).

69.Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R. & Fuchsberger, C. Next-generation genotype imputation service and methods. *Nature genetics* **48**, 1284–1287(2016).

70.Jiang, L., Zheng, Z., Qi, T., Kemper, K.E., Wray, N.R., Visscher, P.M. & Yang, J. A resource-efficient tool for mixed model association analysis of large-scale data. *Nature genetics* **51**, 1749–1755(2019).

71.Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., Stahl, E.A., Huckins, L.M., Nicolae, D.L., Cox, N.J. & Im, H.K. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from gwas summary statistics. *Nature communications* **9**, 1825(2018).

72.Li, B., Clohisey, S.M., Chia, B.S., Wang, B., Cui, A., Eisenhaure, T., Schweitzer, L.D., Hoover, P., Parkinson, N.J., Nachshon, A., Smith, N., Regan, T., Farr, D., Gutmann, M.U., Bukhari, S.I., Law, A., Sangesland, M., Gat-Viks, I., Digard, P., Vasudevan, S., Lingwood, D., Dockrell, D.H., Doench, J.G., Baillie, J.K. & Hacohen, N. Genome-wide CRISPR screen identifies host dependency factors for influenza A virus infection. *Nature Communications* **11**, 164(2020).

73.Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., Hillman-Jackson, J., Kuhn, R.M., Pedersen, J.S., Pohl, A., Raney, B.J., Rosenbloom, K.R., Siepel, A., Smith, K.E., Sugnet, C.W., Sultan-Qurraie, A., Thomas, D.J., Trumbower, H., Weber, R.J., Weirauch, M., Zweig, A.S., Haussler, D. & Kent, W.J. The ucsc genome browser database: Update 2006. *Nucleic acids research* **34**, D590–8(2006).