medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Neutrophils and triglycerides predict critical illness in COVID-19 and reveal CDK6 inhibitors as potential preventive treatment

Hannes A. Baukmann[1], Justin L. Cope[1], Charles N. J. Ravarani[1], Colin Bannard[2], Margaretha R. J. Lamparter[1], Alexander R. E. C. Schwinges[1], Joern E. Klinger[1], Marco F. Schmidt[1*]

[1]biotx.ai GmbH, Am Mühlenberg 11, 14476 Potsdam, Germany.

[2]University of Manchester, Oxford Road, Manchester, M13 9PL. United Kingdom.

*e-Mail: ms@biotx.ai

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Summary

Despite the recent development of vaccines and monoclonal antibodies preventing SARS-CoV-2 infection, treating critically ill COVID-19 patients still remains a top goal. In principle, drug repurposing – the use of an already existing drug for a new indication – could provide a shortcut to a treatment. However, drug repurposing is often very speculative due to lack of clinical evidence. We report here on a methodology to find and test drug target candidates for drug repurposing. Using UK Biobank data, we screened for significant differences in 33 blood cell types, 30 blood biochemistries, and body mass index between an infectious disease phenotype and healthy controls. We then matched critically ill COVID-19 cases with controls that exhibited mild or no symptoms after SARS-CoV-2 infection. Using data from the UK Biobank, we describe a workflow to find evidence for high neutrophil cell count and high concentrations of blood triglycerides as predictors of the immune overreaction in critical illness due to COVID-19. Based on these findings, we identified the enzyme CDK6 as a potential drug target to prevent in high risk individuals with high neutrophil cell count the immune overreaction in critical illness due to COVID-19. Three existing CDK4/6 inhibitors -- abemaciclib, ribociclib, and palbociclib -- have been approved for the treatment of breast cancer. Clinical evidence for CDK4/6 inhibitors in treating critically ill COVID-19 patients has been reported. Further clinical investigations are ongoing.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Introduction

The phenotype of critically ill coronavirus disease 2019 (COVID-19) status substantially differs from mild or moderate disease, even among hospitalized cases, by an uncontrolled overreaction of the host's immune system[1–3] – a so-called virus-induced immunopathology[4] – resulting in acute respiratory distress syndrome (ARDS). The molecular mechanism leading to critical illness due to COVID-19 is still unclear. Identifying causal risk factors is central for prevention and treatment. Nonetheless, there is evidence that susceptibility and overreaction of the immune system to respiratory infections are both strongly heritable.[5,6] A series of genome-wide association (GWA) studies have been conducted to investigate disease pathogenesis in order to find mechanistic targets for therapeutic development or drug repurposing.[7–10] Treating the disease remains a top priority despite the recent development of vaccines preventing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection due to the threat of new vaccine-resistant variants.

The results of 46 GWA studies comprising 46,562 COVID-19 patients from 19 countries have been combined in three meta-analyses by the COVID-19 Host Genetics Initiative.[10] Overall, 15 independent genome-wide significant loci associations were reported for COVID-19 infection in general, of which six were found to be associated with critical illness due to COVID-19: 3p21.31 close to *CXCR6*, which plays a role in chemokine signaling, and *LZTFL1*, which has been implicated in lung cancer; 12q24.13 in a gene cluster that encodes antiviral restriction enzyme activators; 17q21.31, containing the *KANSL1* gene, which has been previously reported for reduced lung function; 19p13.3 within the gene that

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

encodes dipeptidyl peptidase 9 (*DPP9*); 19p13.2 encoding tyrosine kinase 2 (*TYK2*); and 21q22.11 encoding the interferon receptor gene *IFNAR2*. The functions of the genes associated with these six loci are either related to host antiviral defense mechanisms or are mediators of inflammatory organ damage. These results are a very good starting point for a better understanding of host genetics in viral infections. Unfortunately, none of these genes encodes for an established drug target. Consequently, these studies provide no starting point for drug repurposing.

We present here an approach for drug repurposing based not on disease genetics but on the genetics of disease-causing traits. Using UK Biobank data[11], critically ill COVID-19 cases are matched with mild COVID-19 cases as controls. Traits that significantly differ in cases and controls are further investigated for their relationship to critical illness in COVID-19 (Fig. 1). The genetics of these traits may be further investigated to identify and test established target genes for drug repurposing.

Results

Screening for traits associated with infectious disease

Using UK Biobank data[11], we identified 42,065 individuals with respiratory infections, acute respiratory distress syndrome (ARDS), influenza and pneumonia, which serve as our infectious disease cohort. In order to explore how the infectious disease cohort differs from healthy controls, we screened 64 candidate predictive traits (33 blood cell types, 30 blood biochemistries, and body mass index) that had been measured years before the individuals were affected. We observed Bonferroni-corrected statistically significant differences ($p < \alpha/n = 0.05/64$)[12] in 53

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

traits confirmed by independent two-sample t-test and Mann-Whitney U-test[13] (Fig. 1 and SI Fig. 1).

Regression modeling

Furthermore, we identified 1,505 patients who were hospitalized due to SARS-CoV-2 infection and who required respiratory support and/or died due to infection.[14] These patients were defined as cases and matched to controls that were infected with SARS-CoV-2, but showed no and only mild symptoms. Carrying over the 53 traits identified in the previous step, we used regression modeling to investigate the effect of these traits on critically ill COVID-19 status. Out of the 53 traits, 21 traits were significant predictors of critical illness due to COVID-19 with a Bonferroni-corrected significance threshold of $p < \alpha/n = 0.05/53$ (Fig. 1 and SI Tab. 1).

Propensity score analysis

Propensity score analysis is a technique for estimating the effect of a treatment on an outcome independent of any observed factors that covary with that treatment. We employed propensity score stratification using the propensity function of Imai and van Dyk[15] in order to estimate the effect of the treatment on critical illness in COVID-19 independent of the covariates. We iteratively defined each of the 21 predictive traits as treatment and then determined the effect of that treatment when setting one of the remaining traits as covariate, totalling 420 analyses (see Supplementary Information). We built models with pairs of traits rather than moving straight to an all-predictor model, because of the close relationship between some

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

predictor pairs (such as leukocyte and neutrophil count, as neutrophils are leukocytes). The results revealed eleven independent traits that showed a significant effect on severe COVID-19 independently of any other trait (Fig. 2): Body mass index (BMI), neutrophil cell count, cystatin C, glucose, glycated haemoglobin, triglycerides, and five traits related to reticulocytes.

Drop1 analysis

The drop1 analysis compares all possible models that can be constructed by dropping a single model term and evaluating its impact on the regression model. As shown in Fig. 2 we here investigated the remaining independent traits obtained by the propensity score analysis. As the five measures of reticulocytes are highly interdependent, immature reticulocyte fraction can be used as a proxy. This leads to seven traits in our drop1 analysis: BMI, neutrophil cell count, immature reticulocyte fraction, Cystatin C, glucose, glycated hemoglobin and triglycerides. The drop1 analysis revealed that only neutrophil count and triglycerides explain unique variance in critically ill COVID-19 status to a Bonferroni-corrected significance threshold of $p < \alpha/n = 0.05/7$ (Fig. 1 and Tab. 1).

Trait genetics analysis

We next focused on the genetics of neutrophil cell count and triglycerides. We ran GWA analyses for these traits and compared our results with previously reported statistics available from the NHGRI-EBI GWAS Catalog[16] (SI Fig. 2). The identified genes were further investigated for already approved drug molecules. We did not find a gene reported for triglycerides measurement and an approved drug molecule.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

More importantly, elevated triglyceride levels have been previously described to increase neutrophil cell count[17] giving evidence that triglycerides regulate neutrophils. Therefore, we focused on the genetics of neutrophil cell count. We found *CDK6,* encoding for cyclin-dependent kinase 6 (CDK6), reported for neutrophil cell count[18]. Therefore, we envision CDK6 as a potential drug target to decrease neutrophil cell count and, from there, to prevent in high risk individuals with high neutrophil cell count the immune overreaction in critical illness due to COVID-19. Three existing CDK6 inhibitors - abemaciclib, ribociclib, and palbociclib - have been approved for the treatment of breast cancer.

Mendelian randomization

Mendelian randomization (MR) is a robust and accessible tool to examine the causal relationship between an exposure variable and an outcome from GWAS summary statistics.[19] We employed two-sample summary data Mendelian randomization to further validate causal effects of neutrophil cell count genes on the outcome of critical illness due to COVID-19. We used independent GWAS summary data for neutrophil cell count (exposure) published by Vuckovic *et al.[20]* and summary data for critical illness in COVID-19 (outcome) published by the COVID-19 Host Genetics Initiative[10]. As shown in the Supplementary Information Tab. 2, instrumental variable weight (IVW) was significant with a *p* value of 0.01199 when we used a lenient clumping parameter of r = 0.2 and 1,581 SNPs whereas we observed no significant IVW when we used strict clumping parameters of r = 0.01 and 567 SNPs.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Discussion

Using data from the UK Biobank, we describe a workflow to find hints for a causal relationship between high neutrophil cell count and high concentrations of blood triglycerides as a predisposition of the immune overreaction in critical illness due to COVID-19. Based on these findings, we identified the enzyme CDK6 as a potential drug target to prevent in high risk individuals with high neutrophil cell count and triglycerides the immune overreaction in critical illness due to COVID-19. Three existing CDK4/6 inhibitors - abemaciclib, ribociclib, and palbociclib - have been approved for the treatment of breast cancer.

Our procedure worked as follows. First, we identified significant differences in 64 candidate predictive traits between an infectious disease cohort and healthy controls. We used regression models to investigate the effect of these traits on critically ill COVID-19 cases compared to asymptomatic controls. Because highly dependent traits (such as leukocyte and neutrophil count, as neutrophils are leukocytes) would not be significant in drop1 analysis, we first used propensity-score-based multi-model analysis to filter for independently predictive traits. The obtained seven traits then underwent a drop1 analysis. We here identified neutrophil cell count and triglycerides as traits that have a unique effect on critical illness in COVID-19 independent of other traits. These traits are connected as it has been shown that triglycerides activate neutrophils.[17] It is important to note that our Mendelian randomization (MR) results do not confirm a causal role of neutrophil count genes similar to previous reports.[10] However, MR is typically used where there is a direct relationship between gene and outcome.[21] In our case, we are looking for a trait to predict disease progression, and it is irrelevant whether the trait is triggered

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

genetically or by other factors such as prior disease. Type 2 diabetes, cardiovascular disease, and obesity have previously been described as risk factors for the severe course of COVID-19.[22] We could show here that these diseases are confounders of high triglycerides and neutrophil cell count. Especially a high neutrophil cell count already before infection seems to be the reason for critical illness in COVID-19 and should therefore be in the focus of possible preventive therapies.

The role of neutrophil cell count in COVID-19 can be explained by the previously reported disease mechanism.[23] Neutrophils are white blood cells and an important component of our host defense against invading pathogens. Critical illness in COVID-19 is characterized by infiltration of the lungs with macrophages and neutrophils that cause diffuse lung alveolar damage, the histological equivalent to ARDS (Fig. 3).[24–26] Neutrophils develop so-called neutrophil extracellular traps (NETs), web-like structures of nucleic acids wrapped with histones that detain viral particles, through NETosis, a regulated form of neutrophil cell death.[27] However, ineffective clearance and regulation of NETs result in pathological effects such as thromboinflammation.[28]

Cyclin-dependent kinases (CDK) 4 and 6 have been previously described as regulators of NETosis. CDK4/6 inhibitors block NETs formation in a dose-responsive manner but does not impair oxidative burst, phagocytosis, or degranulation.[29] This indicates that CDK4/6 inhibition specifically affects NET production rather than universally modulating inflammatory pathways (in contrast to immunosuppressants such as dexamethasone or interleukin-6 inhibitors). This is supported by Grinshpun *et al.*'s report that COVID-19 progression was halted for a breast cancer patient on CDK4/6 inhibitor therapy. Once the drug was withdrawn, the full classic spectrum of

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

illness appeared, including oxygen desaturation necessitating a prolonged hospital stay for close monitoring of the need for invasive ventilations.[30] Selective inhibition of NETosis is a particularly attractive treatment because CDK4/6 inhibitors can prevent the cytokine storm and, thus, later intensive care.

Several drug classes, each with different mechanisms of action, have been postulated for the treatment of COVID-19.[31] However, clinically relevant effects were only confirmed for two drug classes: Antivirals such as passive immunity through monoclonal antibodies show an effect at the beginning of the infection,[32] while immunosuppressants are only beneficial for the treatment of the later immune overreaction.[33,34] Therefore, a therapeutic gap exists if the infection is not detected early and a possible immune overreaction is to be prevented (Fig. 3). The immune overreaction and, from there, intensive care must be circumvented in order to avoid overwhelming the health care system and triggering lockdowns in the event of further waves. This gap urgently needs to be closed in order to be prepared for any future variants able to evade vaccine protection.

In particular, CDK4/6 inhibitors represent a swift solution to this problem, as they have already been approved for the treatment of breast cancer (abemaciclib, ribociclib, and palbociclib). In the case reported by Grinshpun *et al.*,[30] the CDK4/6 inhibitor was administered prior to infection, therefore it was not harmful in the early course of the disease (like immunosuppressants[33]), but protected against thromboinflammation and thus prevented the necessity of intensive care. Another advantage rendering CDK6 an attractive drug target is that since it is a human protein, mutations of the virus do not influence drug action - in stark contrast to antivirals. Ultimately, CDK4/6 inhibitors might become so-called magic bullets, as

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

they could be used against all virus-induced immune pathologies, and thus also contain future pandemics of novel viruses. Further clinical investigation will reveal whether high neutrophil counts are causative for critical illness in COVID-19, and whether reducing neutrophil cell counts with CDK4/6 inhibitors is a therapeutic option.

Methods

Recruitment of cases and controls

We downloaded the rich information made available by the UK Biobank project on October 25, 2021. COVID-19 test results up until 18th October 2021 were collected, and cases were defined as reported previously.[8]

The infectious disease phenotype was created based on UK Biobank data for respiratory infections, acute respiratory distress syndrome (ARDS), influenza, and pneumonia with hospitalization or death as a result. We aggregated hospital in-patient and death register data for ICD codes corresponding to J00-J06 ("Acute upper respiratory infections"), J09-J18 ("Influenza and pneumonia"), J20-J22 ("Other acute lower respiratory infections"), and J80 (ARDS), yielding 42,065 cases. The remaining individuals from the UK Biobank were defined as potential controls.

Briefly, 1,505 severe cases were defined as patients who died or were hospitalized due to COVID-19 (cause of death or diagnosis containing ICD10 codes U07.1 or U07.2) or were ventilated (operation codes E85.*) in 2020 or 2021 and tested positive for SARS-CoV-2 infection. Individuals that were tested positive for SARS-CoV-2, but did not die or were critical due to COVID-19 and were not ventilated, were defined as potential controls.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

For both cohorts, cases and controls were filtered for European ancestry ("British", "Irish", and "Any other white background"), and individuals with missing age and sex information were discarded. Controls were the then randomly matched to the same number of cases based on age and sex. Variants reported by Pairo-Castineira *et al.[8]* and Ellinghaus *et al.[7]* as well as variants reported by the ClinVar database[35] for the genes reported by the papers were included in the dataset.

Screening for significant traits

The UK Biobank contains data on biological samples taken years before potential infection upon registration of individuals to the program, including blood cell counts and blood biochemistry. In order to identify traits that are significantly different between the infectious disease cohort and age- and sex-matched healthy controls, we performed independent two-sample t-tests and Mann-Whitney U-test using the function of the scipy package in Python 3. We applied a Bonferroni-corrected p-value threshold of $p < \alpha/n = 0.05/64$. In four instances, the p-values were too small to be represented properly, and were instead set to 1.0E-297.

Regression modeling

Logistic regression models were fitted using the *glm* function in R (www.R-project.org).

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Propensity score analysis

Using the method of Imai and Van Dyk[15], the individuals are split into deciles who have a similar propensity for a treatment (one of 21 predictive traits) given the covariates (another predictive trait, age, sex). We then estimated the effect of treatment on severe COVID-19 within each of the groups. The effect across these groups is examined and the average effect of treatment is calculated over the groups to give an estimate of effect of treatment independent of the covariates. The estimate of effect of treatment independent of the covariates was defined as significant if $p < \alpha/n$ = 0.05/(20*21) = 1.1905E-4. There were only slight differences in the results using quintiles or deciles.

Drop1 analysis

A drop1 model comparison procedure was performed using the *drop1()* function in R (www.R-project.org) in order to determine whether each of a set of traits accounts for unique variance in critically ill COVID-19 disease status. The formula of BMI + neutrophil cell count + immature reticulocyte fraction + cystatin C + glucose + glycated haemoglobin + triglycerides was used to predict critical illness due to COVID-19. Single terms were deleted and the F value is calculated to perform an F-test to derive the Pr(>F) value, where low values indicate that a model that does not include this term is significantly different from the full model.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

GWAS

The UK Biobank genotypes for the cases and controls were extracted to create a dataset that was then submitted to a series of quality control steps with an aim to remove biases in the downstream analysis as described in Marees *et al.[36]* First we filtered SNPs and individuals based on their missingness in the dataset. This excludes SNPs that have a high proportion of subjects where genotyping information is unavailable or of poor quality. Similarly, individuals where a large proportion of SNPs could not be measured were excluded. This was achieved in two steps, where first a lenient threshold of 0.2 (i.e. > 20%) was applied to remove the clear outliers, followed by a more stringent threshold of 0.02 (i.e. > 2%). SNP filtering was performed before individual filtering. Next, all variants not on autosomal chromosomes were removed. Next, variants that deviate from Hardy-Weinberg equilibrium were removed in a two-step process whereby we first applied a lenient threshold of 1e-6, followed by a more stringent threshold of 1e-10. This is a common indicator of genotyping errors. Thereafter, individuals were filtered out based on their heterozygosity rates which can indicate sample contamination. Individuals deviating by more than 3 standard deviations from the mean of the rate from all samples were filtered out. To assess the heterozygosity rate per sample, those variants that were in linkage disequilibrium with each other were extracted by scanning the genome at a window size of 50 variants, a step size of 5, and a pairwise correlation threshold of 0.2. Next, related individuals were removed. To achieve this, their identity by descent coefficients (IBD) were calculated and only one individual per related cluster was kept. Then, the small proportion of missing genotypes were imputed and additional variants reported by Pairo-Castineira *et al.[8]* and Ellinghaus *et al.[7]* as well as

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

variants reported by ClinVar database[35] were included in the dataset from the UK Biobank imputed variants. This yielded a dataset with a total number of 335,332 quality controlled variants. Finally, the population structure of the samples was analyzed in two stages to identify internal stratifications, which was used to filter out any individuals not predicted close enough to a European reference cohort. A genome-wide association analysis was performed with the R-package SAIGE[37].

Mendelian randomization

We used independent GWAS summary data for neutrophil cell count (exposure) published by Vuckovic *et al.[20]* (GCST90002398 downloaded January 15th 2021) and summary data for critically ill COVID-19 status (outcome) published by the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/results - COVID19hg GWAS meta-analyses round 5 release date January 18th 2021). Two-sample MR analyses were done as previously described.[10]

Acknowledgment

The research has been conducted using the UK Biobank Resource under Application no. 36226. We thank Radi Hilaneh for making Fig. 3. The research work was supported by the *Investitionsbank des Landes Brandenburg* (ILB), the European Regional Development Fund (ERDF), and the European Social Fund+ (ESF+). Access to the UK Biobank was funded by the EIT Health Digital Sandbox program to access European biobank data (grant number 2019-DS1001-3754). We also thank the program 'digital solutions made in Brandenburg' (digisolBB) for its continued support.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Competing interests

H.A.B., J.L.C., C.N.J.R., M.R.J.L., J.E.K., and M.F.S are employees of biotx.ai GmbH. A.R.E.S was an employee of biotx.ai GmbH.

References

1. Berlin, D.A., Gulick, R.M., and Martinez, F.J. (2020). Severe Covid-19. N. Engl. J. Med. *383*, 2451–2460.

2. Fajgenbaum, D.C., and June, C.H. (2020). Cytokine Storm. N. Engl. J. Med. *383*, 2255–2273.

3. Millar, J.E., Neyton, L., Seth, S., Dunning, J., Merson, L., Murthy, S., Russell, C.D., Keating, S., Swets, M., Sudre, C.H., *et al.* (2020). Robust, reproducible clinical patterns in hospitalised patients with COVID-19. medRxiv. Available at: https://www.medrxiv.org/content/10.1101/2020.08.14.20168088v1.

4. Rouse, B.T., and Sehrawat, S. (2010). Immunity and immunopathology to viruses: what decides the outcome? Nat. Rev. Immunol. *10*, 514–526.

5. Casanova, J.-L. (2015). Severe infectious diseases of childhood as monogenic inborn errors of immunity. Proc. Natl. Acad. Sci. U. S. A. *112*, E7128–37.

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

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

7. Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., *et al.* (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N. Engl. J. Med. *383*, 1522–1534.

8. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A.D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M.H., Russell, C.D., *et al.* (2021). Genetic mechanisms of critical illness in COVID-19. Nature *591*, 92–98.

9. Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I.K.D., Hodeib, S., Korol, C., *et al.* (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science *370*. Available at: http://dx.doi.org/10.1126/science.abd4570.

10. COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19. Nature *600*, 472–477.

11. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., *et al.* (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature *562*, 203–209.

12. Dunn, O.J. (1961). Multiple Comparisons among Means. J. Am. Stat. Assoc. *56*, 52–64.

13. Fay, M.P., and Proschan, M.A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat. Surv. *4*, 1–39.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

14. Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 7) (2020). Chin. Med. J. *133*, 1087–1095.

15. Imai, K., and van Dyk, D.A. (2004). Causal inference with general treatment regimes. J. Am. Stat. Assoc. *99*, 854–866.

16. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., *et al.* (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012.

17. Wanten, G.J.A., Janssen, F.P., and Naber, A.H.J. (2002). Saturated triglycerides and fatty acids activate neutrophils depending on carbon chain-length. Eur. J. Clin. Invest. *32*, 285–289.

18. Lin, B.D., Carnero-Montoro, E., Bell, J.T., Boomsma, D.I., de Geus, E.J., Jansen, R., Kluft, C., Mangino, M., Penninx, B., Spector, T.D., *et al.* (2017). 2SNP heritability and effects of genetic variants for neutrophil-to-lymphocyte and platelet-to-lymphocyte ratio. J. Hum. Genet. *62*, 979–988.

19. Katan, M.B. (1986). Apolipoprotein E isoforms, serum cholesterol, and cancer. Lancet *1*, 507–508.

20. Vuckovic, D., Bao, E.L., Akbari, P., Lareau, C.A., Mousas, A., Jiang, T., Chen, M.-H., Raffield, L.M., Tardaguila, M., Huffman, J.E., *et al.* (2020). The Polygenic and Monogenic Basis of Blood Traits and Diseases. Cell *182*, 1214–1231.e11.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

21. Burgess, S. (2014). Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int. J. Epidemiol. *43*, 922–929.

22. Know, W.Y.N.T. People with certain medical conditions. Available at: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

23. Bonaventura, A., Vecchié, A., Dagna, L., Martinod, K., Dixon, D.L., Van Tassell, B.W., Dentali, F., Montecucco, F., Massberg, S., Levi, M., *et al.* (2021). Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. Nat. Rev. Immunol. *21*, 319–329.

24. Schaller, T., Hirschbühl, K., Burkhardt, K., Braun, G., Trepel, M., Märkl, B., and Claus, R. (2020). Postmortem Examination of Patients With COVID-19. JAMA *323*, 2518–2520.

25. Nicholls, J.M., Poon, L.L.M., Lee, K.C., Ng, W.F., Lai, S.T., Leung, C.Y., Chu, C.M., Hui, P.K., Mak, K.L., Lim, W., *et al.* (2003). Lung pathology of fatal severe acute respiratory syndrome. Lancet *361*, 1773–1778.

26. Tian, S., Hu, W., Niu, L., Liu, H., Xu, H., and Xiao, S.-Y. (2020). Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus (COVID-19) Pneumonia in Two Patients With Lung Cancer. J. Thorac. Oncol. *15*, 700–704.

27. Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., Weinrauch, Y., and Zychlinsky, A. (2004). Neutrophil extracellular traps kill

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

bacteria. Science *303*, 1532–1535.

28. Cheng, O.Z., and Palaniyar, N. (2013). NET balancing: a problem in inflammatory lung diseases. Front. Immunol. *4*, 1.

29. Amulic, B., Knackstedt, S.L., Abu Abed, U., Deigendesch, N., Harbort, C.J., Caffrey, B.E., Brinkmann, V., Heppner, F.L., Hinds, P.W., and Zychlinsky, A. (2017). Cell-Cycle Proteins Control Production of Neutrophil Extracellular Traps. Dev. Cell *43*, 449–462.e5.

30. Grinshpun, A., Merlet, I., Fruchtman, H., and Nachman, D. (2020). A Protracted Course of COVID19 Infection in a Metastatic Breast Cancer Patient During CDK4/6 Inhibitor Therapy. Front. Oncol. *10*, 1085.

31. Lamontagne, F., Agoritsas, T., Siemieniuk, R., Rochwerg, B., Bartoszko, J., Askie, L., Macdonald, H., Amin, W., Bausch, F.J., Burhan, E., *et al.* (2021). A living WHO guideline on drugs to prevent covid-19. BMJ *372*, n526.

32. Taylor, P.C., Adams, A.C., Hufford, M.M., de la Torre, I., Winthrop, K., and Gottlieb, R.L. (2021). Neutralizing monoclonal antibodies for treatment of COVID-19. Nat. Rev. Immunol. *21*, 382–393.

33. RECOVERY Collaborative Group, Horby, P., Lim, W.S., Emberson, J.R., Mafham, M., Bell, J.L., Linsell, L., Staplin, N., Brightling, C., Ustianowski, A., *et al.* (2021). Dexamethasone in Hospitalized Patients with Covid-19. N. Engl. J. Med. *384*, 693–704.

34. REMAP-CAP Investigators, Gordon, A.C., Mouncey, P.R., Al-Beidh, F., Rowan,

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

K.M., Nichol, A.D., Arabi, Y.M., Annane, D., Beane, A., van Bentum-Puijk, W., *et al.* (2021). Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. N. Engl. J. Med. Available at: http://dx.doi.org/10.1056/NEJMoa2100433.

35. Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Hoffman, D., Jang, W., Kaur, K., Liu, C., *et al.* (2020). ClinVar: improvements to accessing data. Nucleic Acids Res. *48*, D835–D844.

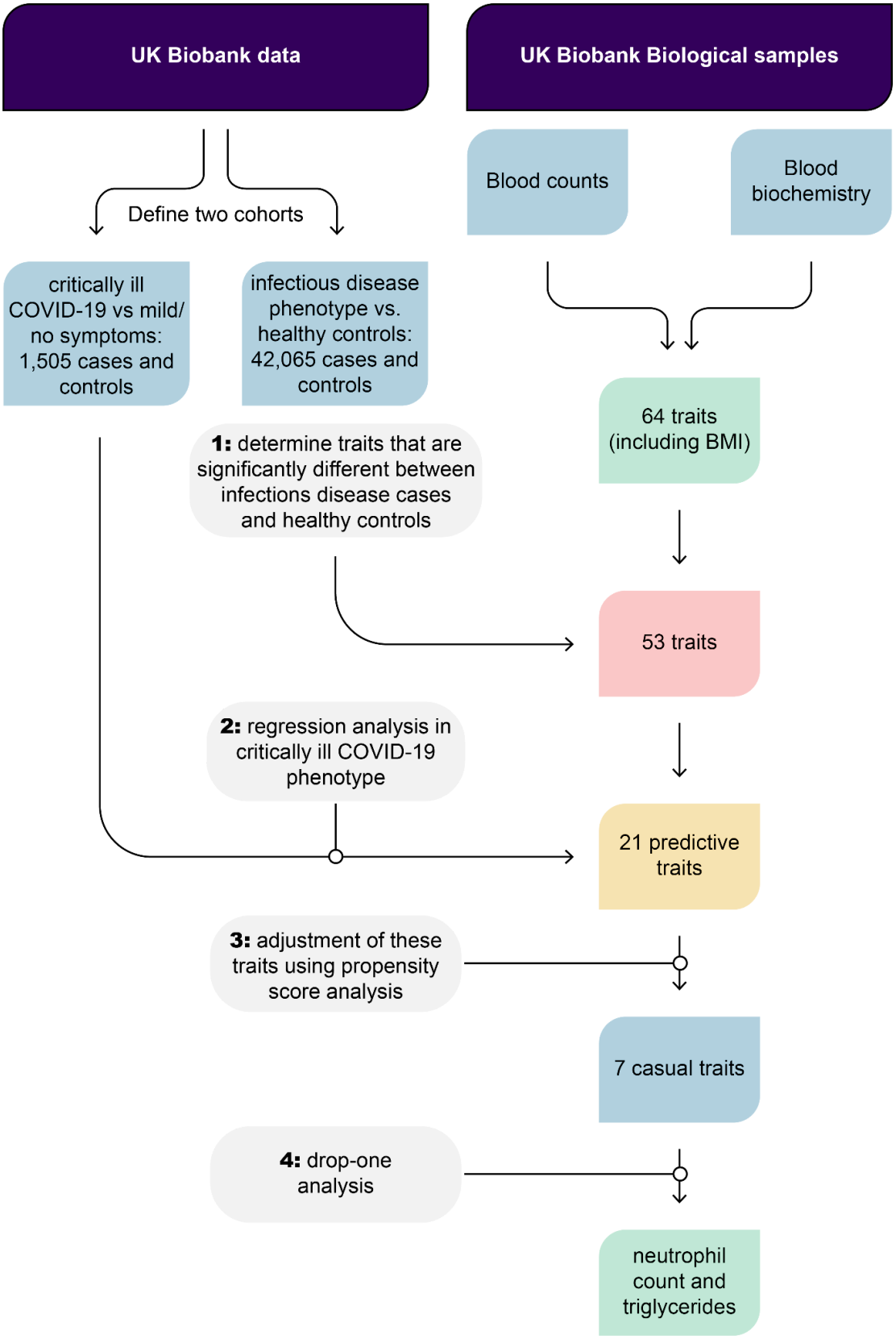
36. Marees, A.T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., and Derks, E.M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int. J. Methods Psychiatr. Res. *27*, e1608.

37. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., *et al.* (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figures & Tables

Figure 1



medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Fig. 1. Workflow to identify traits leading to critical illness due to COVID-19. We identified significant differences in 64 candidate predictive traits between an infectious disease cohort and healthy controls. We used regression models to investigate the effect of these traits on critically ill COVID-19 cases compared to asymptomatic controls. Because highly dependent traits (such as leukocyte and neutrophil count, as neutrophils are leukocytes) would not be significant in drop1 analysis, we first used propensity score analysis to filter for independently predictive traits. The obtained seven traits underwent a drop1 analysis. We here identified neutrophil cell count and triglycerides as traits that have a unique effect on critical illness in COVID-19 independent of other traits.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 2

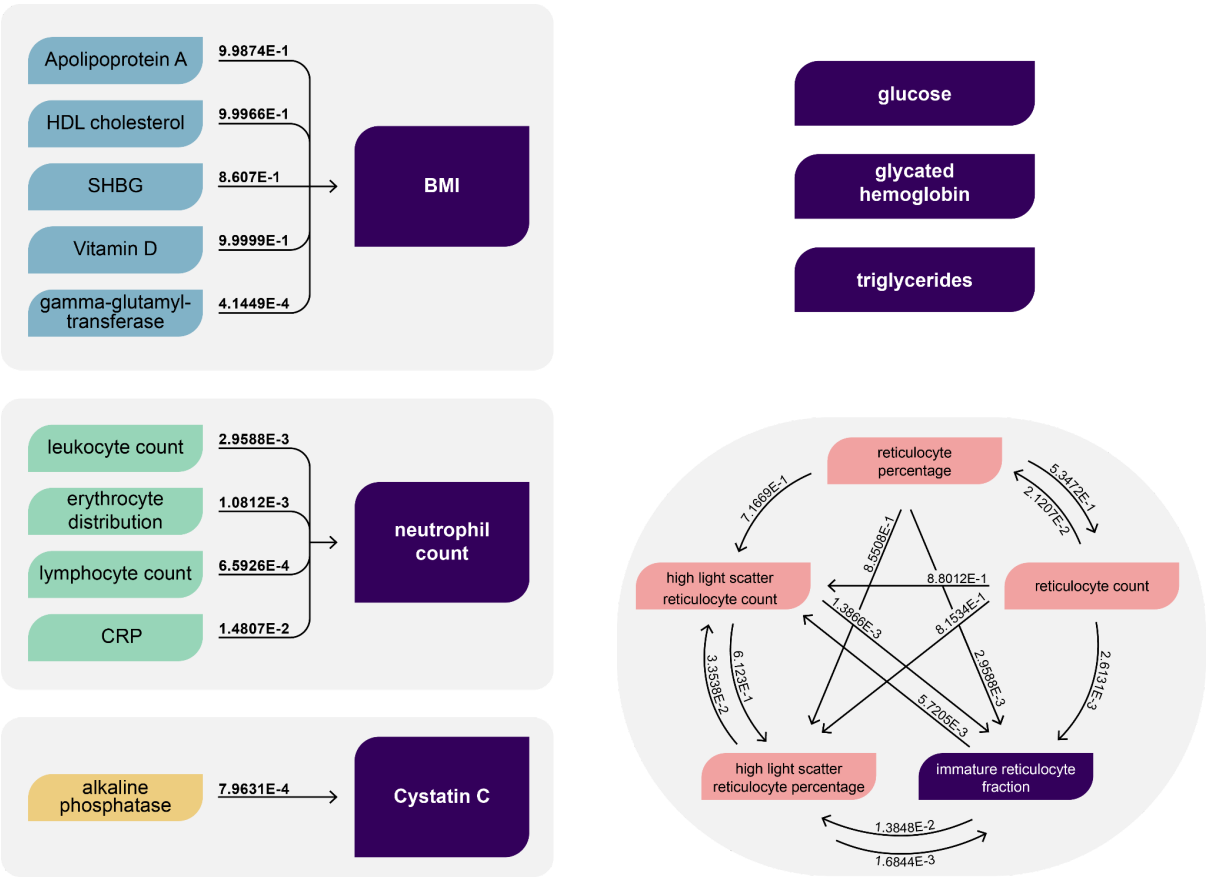


Fig. 2. Propensity score analysis to identify independent traits. Arrows indicate that the covariate in the donor node (treatment) does not have a significant effect on the outcome (severe COVID-19) independent of the covariate in the acceptor node (and age and sex) as indicated by p-values above the threshold of 1.1905E-4. In total, seven traits were identified as independent (in purple) and were further analyzed in drop1 analysis. The five reticulocyte-related covariates cannot be accounted for by any other covariate. Each of the three covariates immature reticulocyte fraction, high light-scatter count/percentage can control for all five covariates though. It is important to note that not all statistical relationships are depicted.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Figure 3

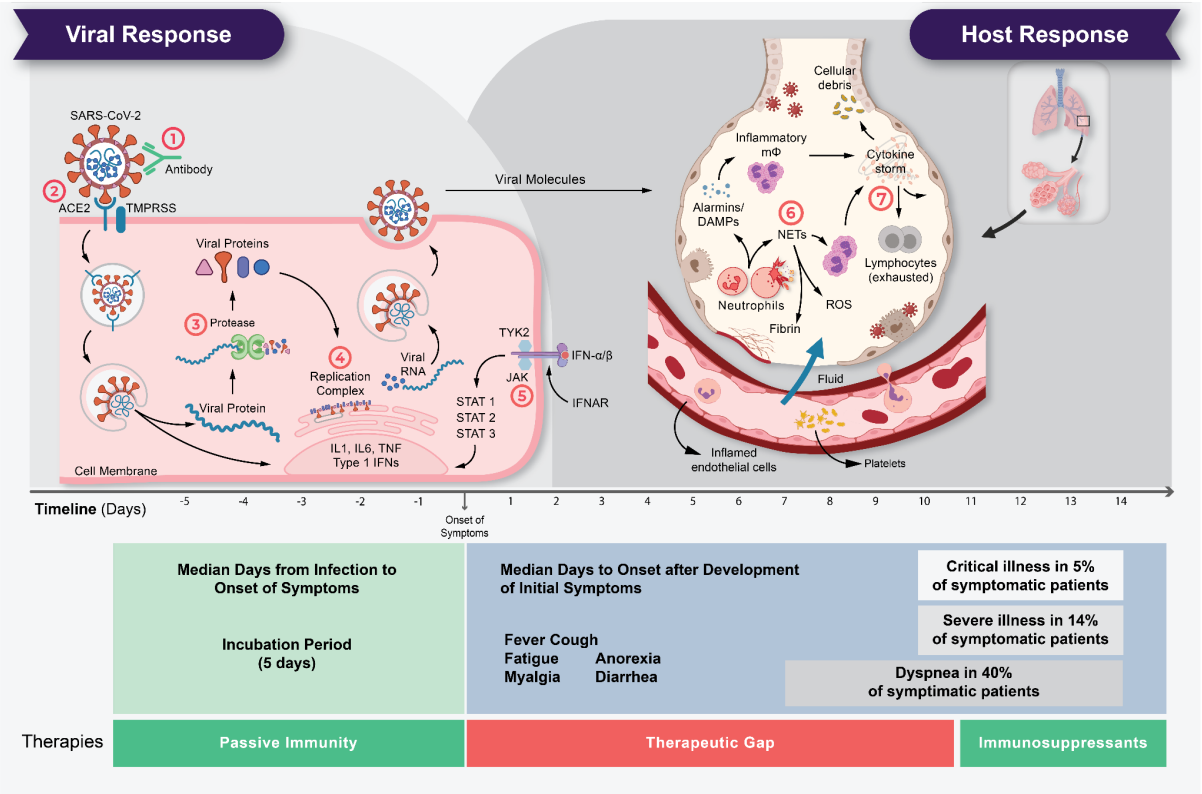


Fig. 3. The life cycle of SARS-CoV-2 and the corresponding pathogenesis of COVID-19 display two phases: a viral response and a host-response phase. In the viral response phase, the virus enters the host cell and viral replication begins. Approximately five days after infection and successful replication, initial mild and moderate symptoms such as fever, cough, fatigue, anorexia, myalgia, and diarrhea are observed in conjunction with a decrease in lymphocyte cell count (lymphopenia). The following host-response phase determines the severity of the disease: in some patients, uncontrolled overreaction of the immune system – so-called virus-induced immunopathology – requires hospitalization and respiratory support due to acute respiratory distress syndrome (ARDS). Thus, severe cases of COVID-19 originate from an immune overreaction rather than from the viral infection itself. Currently, there are seven drug mechanisms described: ① Passive immunity; ② Entry

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

inhibitors; ③ Protease inhibitors; ④ Polymerase inhibitors; ⑤ JAK inhibitors; ⑥ NETosis inhibitors; ⑦ Immunosuppressants.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 1

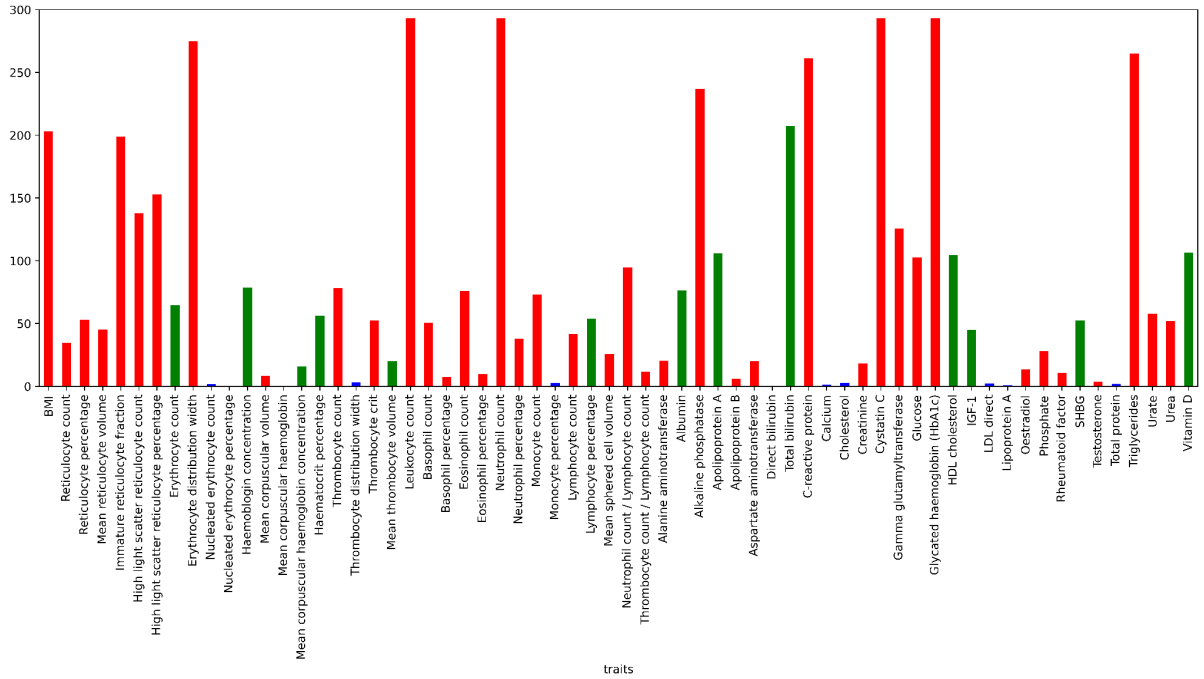
Tab. 1. Pr(>F) values of seven traits determined in drop1 analysis. Significance thresholds are indicated by asterisks, where two asterisks indicate p-values below 0.05/7.

Trait	**Pr(>F)**
BMI	0.109666
Neutrophil cell count	0.000252 **
Immature reticulocyte fraction	0.028504
Cystatin C	0.054143
Glucose	0.630619
Glycated haemoglobin	0.013758
Triglycerides	0.000179 **

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

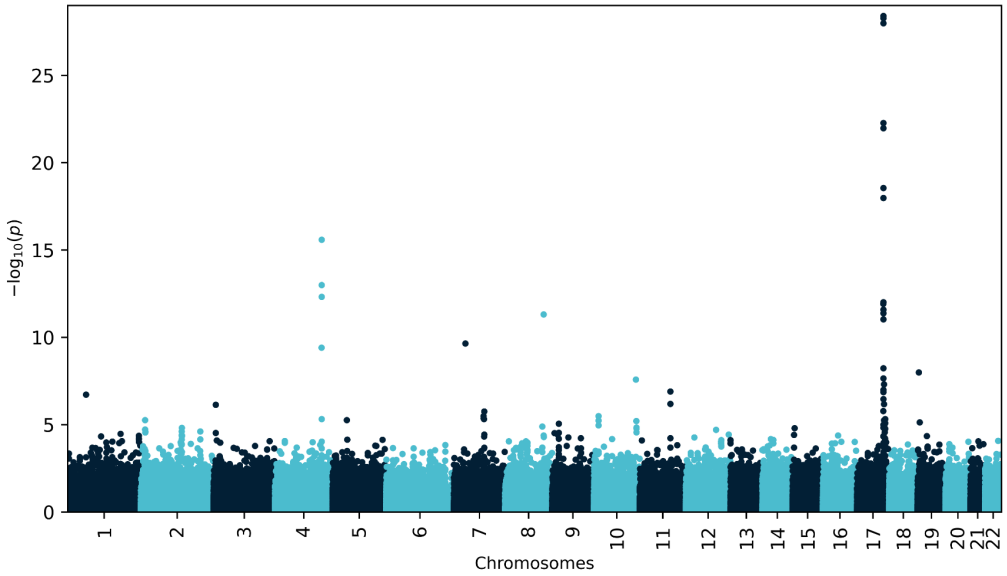
Supplementary information

SI Figure 1



SI Fig. 1. Bonferroni-corrected statistically significant differences in 64 traits identified using independent two-sample t-test and confirmed by Mann-Whitney U-test. Red and green columns indicate traits that are significantly increased in infectious disease cases or healthy controls, respectively. In these measures, taken years prior to infection, cases showed significant differences in the characteristics in various traits that have been later described as phenotypes associated with critical illness due to COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



SI Fig. 2. Manhattan plot of neutrophil cell count showing that we reproduce the reported *CDK6* signal (rs445) on chromosome 7.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

SI Table 1

SI Tab. 1. Critical illness in COVID-19 was regressed on the traits significantly different between infectious disease cases and healthy controls. Traits age, alanine aminotransferase, BMI, C-reactive protein, and neutrophil cell count. All traits other than age were found to explain unique variance in disease status. This table reports log likelihood ratios, standard errors and likelihood ratio tests from the drop one procedure for each. Significance thresholds are indicated by asterisks, where three asterisks indicate p-values below 0.001/53, two indicate p-values below 0.01/53, and one asterisk indicates p-values below 0.05/53.

Trait	Estimate	SE	p-value
BMI	0.04719	0.00703	1.890E-11 ***
Reticulocyte count	7.60920	1.49690	3.710E-07 ***
Reticulocyte percentage	0.37950	0.07100	9.020E-08 ***
Mean reticulocyte volume	0.00414	0.00467	3.750E-01
Immature reticulocyte fraction	3.64820	0.61050	2.290E-09 ***
High light scatter reticulocyte count	22.60823	3.76537	1.920E-09 ***
High light scatter reticulocyte percentage	1.06815	0.17382	7.980E-10 ***
Erythrocyte count	-0.01703	0.08626	8.430E-01
Erythrocyte distribution width	0.16615	0.03953	2.630E-05 **
Haemoblogin concentration	-0.00891	0.02967	7.640E-01
Mean corpuscular volume	-0.00217	0.00805	7.880E-01
Mean corpuscular haemoglobin concentration	0.02570	0.03528	4.660E-01
Haematocrit percentage	-0.00465	0.01027	6.510E-01
Thrombocyte count	0.00168	0.00060	5.250E-03
Thrombocyte crit	1.72370	0.73100	1.840E-02
Mean thrombocyte volume	-0.04106	0.03445	2.330E-01
Leukocyte count	0.14942	0.01993	6.500E-14 ***
Basophil count	2.35277	0.79549	3.100E-03
Basophil percentage	0.08412	0.06923	2.240E-01
Eosinophil count	0.23842	0.26624	3.710E-01
Eosinophil percentage	-0.03716	0.02074	7.320E-02

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Neutrophil count	0.17078	0.02531	1.500E-11 ***
Neutrophil percentage	0.00943	0.00422	2.520E-02
Monocyte count	0.33506	0.15561	3.130E-02
Lymphocyte count	0.23706	0.05576	2.120E-05 **
Lymphocyte percentage	-0.00802	0.00487	9.960E-02
Mean sphered cell volume	0.00048	0.00680	9.440E-01
Neutrophil count / Lymphocyte count	0.08736	0.02874	2.370E-03
Thrombocyte count / Lymphocyte count	-0.00045	0.00059	4.450E-01
Alanine aminotransferase	0.00507	0.00247	4.040E-02
Albumin	-0.00503	0.01399	7.190E-01
Alkaline phosphatase	0.00637	0.00144	9.620E-06 ***
Apolipoprotein A	-0.53990	0.14550	2.070E-04 *
Apolipoprotein B	0.18340	0.15460	2.360E-01
Aspartate aminotransferase	0.00525	0.00352	1.356E-01
Total bilirubin	-0.02864	0.00906	1.560E-03
C-reactive protein	0.03173	0.00831	1.350E-04 **
Creatinine	-0.00022	0.00204	9.140E-01
Cystatin C	1.09560	0.19520	2.000E-08 ***
Gamma glutamyltransferase	0.00315	0.00080	8.300E-05 **
Glucose	0.13716	0.02671	2.810E-07 ***
Glycated haemoglobin (HbA1c)	0.03611	0.00502	6.380E-13 ***
HDL cholesterol	-0.47520	0.10700	8.920E-06 ***
IGF-1	-0.01185	0.00650	6.820E-02
Oestradiol	-0.00038	0.00041	3.580E-01
Phosphate	0.01507	0.23191	9.480E-01
Rheumatoid factor	0.00997	0.00376	8.030E-03
SHBG	-0.00587	0.00162	2.840E-04 *
Testosterone	0.00502	0.00644	4.355E-01
Triglycerides	0.24894	0.03707	1.870E-11 ***
Urate	0.00142	0.00045	1.610E-03
Urea	0.04115	0.02234	6.550E-02
Vitamin D	-0.00998	0.00177	1.790E-08 ***

medRxiv preprint doi: https://doi.org/10.1101/2021.05.18.21256584; this version posted February 14, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

SI Table 2

SI Tab. 2. The two sample MR analyses here showed that for neutrophil cell count as exposure and critically ill COVID-19 status as outcome no significant effect was detected while using strict clumping parameters.

Clumping	SNPs	Beta	SE	IVW p-value	Pleiotropy test
lenient (r = 0.2)	1,581	-0.11139	0.04433	0.01199*	negative
strict (r = 0.01)	567	0.01135	0.06987	0.87095	negative