Critically ill COVID-19 associated trait genetics reveals CDK6 inhibitors as potential treatment

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

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

[2]Institute of Population Health, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, United Kingdom.

*e-Mail: ms@biotx.ai

Abstract

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 the lack of clinical evidence. We here report on a methodology to find and test gene drug target candidates for drug repurposing. We matched critically ill COVID-19 cases from the UK Biobank with healthy controls and screened for significant differences in 33 blood cell types, 30 blood biochemistries, and body mass index in cases and controls. Significant differences in traits that have previously been associated with critically ill COVID-19 status, such as alanine aminotransferase, body mass index, C-reactive protein, and neutrophil cell count were further investigated. In-depth statistical analysis of COVID-19 associated traits and their genetics using regression modeling and propensity score stratification identified cyclin-dependent kinase 6 (CDK6) as a more promising drug target to selectively treat critically ill COVID-19 patients than the previously reported interleukin 6. Four existing CDK6 inhibitors abemaciclib, ribociclib, trilaciclib, and palbociclib have been approved for breast cancer. Clinical evidence for CDK6 inhibitors in treating critically ill COVID-19 has been reported. Further clinical investigations are ongoing.

Introduction

The phenotype of critically ill coronavirus disease 2019 (COVID-19) completely 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). Although the molecular mechanism leading to critically ill COVID-19 is still unclear, 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, as treating the disease remains a top goal despite the recent development of vaccines.[7–10]

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 forCOVID-19 infection in general, whereas six genome-wide significant associations were found to be associated with critical illness due to COVID-19. The six loci comprise 3p21.31 close to *CXCR6* that plays a role in chemokine signaling and *LZTFL1* that has been implicated in lung cancer; 12q24.13 in a gene cluster that encodes antiviral restriction enzyme activators; 17q21.31 with the *KANSL1* gene has been previously reported for reduced lung function; 19p13.3 within the gene that encodes dipeptidyl peptidase 9 (*DPP9*); 19p13.2 that encodes tyrosine kinase 2 (*TYK2*); 21q22.11 encoding for the interferon receptor gene *IFNAR2*. The functions of the identified genes are either related to host antiviral defense mechanisms or are mediators of inflammatory organ damage. These results are a

good starting point for a better understanding of host genetics in viral infections. Nevertheless, none of the found genes encodes for an established drug target. Consequently, these studies provide no evidence that supports drug repurposing.

We here present an approach for drug repurposing based on genetics of disease associated traits instead of disease genetics. First, critically ill COVID-19 cases are matched with healthy controls, and the two cohorts are investigated for significant differences in previously reported traits. Traits that differ in cases and controls and that have been associated with critically ill COVID-19 are further investigated to find and test established drug target genes for drug repurposing.

Results

Screening for critically ill COVID-19 associated traits

Critically ill COVID-19 cases were selected from the UK Biobank[11] based on their previously described phenotype definition. Critically ill COVID-19 cases are patients who were hospitalized due to confirmed SARS-CoV-2 infection and who required respiratory support and/or died due to infection.[12] We identified 8,153 cases in the UK Biobank and selected age and sex matched healthy controls. In order to explore how the critically ill cohort differed in general 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 by COVID-19 (Fig. 1). We observed statistically significant differences in 64 traits confirmed by multiple statistical tests, independent two-sample t-test and Mann-Whitney U-test[13], and Bonferroni correction ($p < \alpha/n = 0.05/64$)[14]. In these measures taken years prior to infection, cases showed significant differences in various traits that have been described as phenotypes for critically ill COVID-19. For instance, cases

99 had higher body mass index (BMI), higher reticulocyte cell count, higher inflammatory
100 markers such as alanine aminotransferase, C-reactive protein, cystatin C, neutrophil
101 cell count, and higher glycated hemoglobin (HbA1c), but lower HDL and LDL
102 cholesterol as well as lower vitamin D level than healthy controls.
103 As an additional comparison group we selected infection disease cases from the UK
104 Biobank to assess to what extent these traits are general indicators of predisposition
105 to severe infections. In total 23,348 participants were selected who had been
106 diagnosed with a respiratory infection, acute respiratory distress syndrome, influenza
107 or pneumonia and were hospitalized or had died as a result. Again, cases were
108 matched with healthy controls in order to screen 33 blood cell types, 30 blood
109 biochemistries, and body mass index for statistically significant differences in these
110 two cohorts. As in the COVID-19 investigation, we here found that cases had higher
111 body mass index (BMI), higher reticulocyte cell count, higher inflammatory markers
112 such as alanine aminotransferase, C-reactive protein, cystatin C, neutrophil cell count,
113 and higher glycated hemoglobin (HbA1c), but lower HDL and LDL cholesterol as well
114 as lower vitamin D levels than healthy controls (Supplementary Information Fig. 1 and
115 Fig. 2).
116 Traits that have been previously associated with critically ill COVID-19 and infection
117 diseases such as alanine aminotransferase, BMI, C-reactive protein, and neutrophil
118 cell count were further investigated.[15,16]
119

120 GWAS results

121 We next focused on trait genetics of reported critically ill COVID-19 associated
122 phenotypes such as BMI, neutrophil cell count, C-reactive protein, and alanine
123 aminotransferase. We ran GWA analyses for these four traits and compared our

124 results with previously reported statistics available from the NHGRI-EBI GWAS
125 Catalog[17]. The identified genes were further investigated for already approved drug
126 molecules. We found *IL-6*, encoding for interleukin 6, reported for C-reactive protein[18]
127 and *CDK6,* encoding for cyclin-dependent kinase 6 (CDK6), reported for BMI[19] and
128 neutrophil cell count[20]. Unfortunately, we could not confirm the *IL-6* signal (rs2097677)
129 in our GWA analysis for C-reactive protein (Supplementary Information Fig. 3). We
130 continued our research with the reported drug target for interleukin 6 and interleukin 6
131 receptor (IL-6R). We furthermore confirmed single nucleotide polymorphisms (SNPs)
132 rs42044 and rs445 in the *CDK6* gene as significant in our GWA analyses for BMI and
133 neutrophil cell count (Supplementary Information Fig. 4). The allele distributions of
134 rs2097677, rs42044, and rs445 in cases and controls can be found in the
135 Supplementary Information Tab. 1 and 2.
136

137 **Regression modeling**

138 Logistic regression models were built to examine the relationship between a series of
139 candidate predictive traits (age, alanine aminotransferase, BMI, C-reactive protein,
140 and neutrophil cell count) and critically ill COVID-19 status. All of these traits apart
141 from age were significant predictors of critically ill COVID-19 (as illustrated for
142 neutrophil cell count in Fig. 2). As expected, no relationship between age and disease
143 status could be found as the members of the control group were matched with the
144 reported cases by age and sex. A drop one analysis revealed that all traits explain
145 unique variance in critically ill COVID-19 status.
146

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 and would otherwise make it impossible to determine causation. We specifically employed propensity score stratification using the propensity function of Imai and van Dyk[21] in order to determine whether a causal relationship exists between both C-reactive protein and neutrophil cell count and the prevalence of critically ill COVID-19 independently from other covariates. A weighted average across propensity deciles indicated a significant effect of C-reactive protein (Log Odds Ratio = 0.04; SE = 0.004; p = 1.3288e-26) and neutrophil cell count (Log Odds Ratio = 0.3; SE = 0.018; p = 2.7354e-61) on disease status. Tab. 2 shows the neutrophil cell count in the critically ill COVID-19 patients and the control group across the propensity deciles.

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.[22] We employed two-sample summary data Mendelian randomization to further validate causal effects of neutrophil cell count genes on the outcome of critically ill COVID-19. We used independent GWAS summary data for neutrophil cell count (exposure) published by Vuckovic *et al.*[23] and summary data for critically ill COVID-19 (outcome) published by the COVID-19 Host Genetics Initiative[10]. As shown in the Supplementary Information Tab. 3, 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 (Supplementary Information Tab. 3).

Discussion

We have described a method for identifying drug targets for the treatment of disease based not directly on the genetics of the disease itself but rather on the genetics of disease-associated traits. Using data from the UK Biobank, we found evidence of a causal relationship between a series of traits and critical illness due to COVID-19. Using genome-wide associations we identified genetic markers associated with these traits. Based on these two steps we were able to identify CDK6 as a potential drug target for critically ill COVID-19. The four CDK6 inhibitors abemaciclib, palbociclib, ribociclib, and trilaciclib have been already approved for breast cancer and can potentially be repurposed to treat critically ill COVID-19.

Our procedure worked as follows. We matched reported critically ill COVID-19 cases from the UK Biobank with healthy controls and checked for significant differences between cases and controls in 64 candidate predictive traits. Cases showed significant differences in alanine aminotransferase, BMI, C-reactive protein, and neutrophil cell count. These measures were taken from the individuals concerned years before infection. We hypothesize that the genetic drivers of a disease associated trait displays a good drug target to treat the disease. In the literature we found the genes *IL-6*, encoding interleukin 6, as a driver of C-reactive protein, and *CDK6*, encoding cyclin-dependent kinase 6, as a driver of BMI and neutrophil cell count. Three drugs have been already approved that inhibit interleukin 6 either directly (siltuximab for Castleman's Disease) or indirectly via the interleukin 6 receptor (tocilizumab and sarilumab for rheumatoid arthritis). The four drugs abemaciclib, palbociclib, ribociclib, and trilaciclib are already approved for treatment of breast cancer target CDK4/6.

196 Regression models that tested all traits together showed that C-reactive protein and
197 neutrophil cell count explained independent variance in critically ill COVID-19 status
198 in the presence of other traits such as BMI. Propensity score stratification found
199 evidence of a causal relationship between both C-reactive protein and neutrophil cell
200 count and the prevalence of critically ill COVID-19 status.
201 It is important to note that Mendelian randomization results did not confirm a causal
202 role for either C-reactive protein or neutrophil cell count. However, MR is typically used
203 where there is a direct relationship between gene and outcome. In our case we are
204 looking for a relationship that is mediated by a viral infection, adding a great deal more
205 noise. This, compounded by the fact that MR needs a larger sample size[24] than we
206 have available, might account for our not finding evidence of a relationship in this
207 analysis.
208 The role of C-reactive protein in COVID-19 can be explained by the previously-
209 reported disease mechanism. The phenotype of critically ill COVID-19 completely
210 differs from mild or moderate disease, even among hospitalized cases, by an
211 uncontrolled overreaction of the host's immune system.[1–3] The most prominent
212 difference between critically ill and moderate COVID-19 cases is the response to
213 immunosuppressive therapy. In patients without respiratory failure, there is a trend
214 indicating that treatment with corticosteroid dexamethasone is harmful, whereas
215 among patients with critical respiratory failure, it has substantial benefit.[25] Therefore,
216 immunomodulatory therapies, such as interleukin 6 receptor (IL-6R) antagonists
217 tocilizumab and sarilumab have been successfully tested to block the immune
218 system's overreaction in form of a cytokine storm in critically ill COVID-19 patients.[26]
219 The role of neutrophil count in the disease mechanism can also be explained.
220 Neutrophils are white blood cells and an important component of our host defense

221 against invading pathogens. Critically ill COVID-19 status is characterized by
222 infiltration of the lungs with macrophages and neutrophils that cause diffuse lung
223 alveolar damage, the histological equivalent to ARDS.[27–29] Neutrophils develop a
224 sophisticated network of DNA called neutrophil extracellular traps (NETs) through
225 NETosis, a liberation of web-like structures of nucleic acids wrapped with histones that
226 detain viral particles.[30] However, ineffective clearance and regulation of NETs result
227 in pathological effects such as thromboinflammation as described above.[31] On the one
228 hand, NETs are essential and efficient for trapping the virus. And on the other, they
229 cause damage to the organism by triggering highly intense immunological and
230 inflammatory processes. CDK4/6 have been previously described as regulators of
231 NETosis. CDK4/6 inhibitors abemaciclib and palbociclib block NET formation in a
232 dose-responsive manner but do not inhibit the oxidative burst, phagocytosis, or
233 degranulation, indicating that CDK4/6 inhibition specifically affects NET production,
234 rather than generally modulating inflammatory pathways as IL-6 inhibitors do.[32]
235 There is good reason to think that IL-6 and CDK4/6 might both be good drug targets
236 for the treatment of COVID-19. However, given that neutrophils have a direct role in
237 the thromboinflammatory process in critically ill COVID-19 patients, we believe that
238 blocking neutrophils represents a more selective strategy than suppressing the
239 immune system as a whole. There are two phases of COVID-19 infection (Fig. 3). In
240 the beginning of infection, the virus enters the cell and starts viral replication. Here,
241 vaccines and monoclonal antibodies block the viral entry. Nonetheless, after entry into
242 the cell via the ACE2 receptor, currently, no therapeutic option for intervention exists.
243 After viral infection first immune reactions are observed such as a decrease of
244 lymphocyte cell count (lymphopenia). This phase is called the viral response phase.
245 In the beginning of the next phase, so-called host response phase, an overreaction of

246 the host immune system occurs. The detailed mechanism is still unknown why some
247 patients' host immune systems overreact. From our results here we assume that
248 already manifested inflammations observed by high C-reactive protein levels and high
249 neutrophil cell count trigger the immune system's overreaction resulting in thrombotic
250 inflammation in the lungs and, from there, respiratory failure. In summary, the host
251 response decides whether an infection disease like COVID-19 has a mild course or
252 leads to respiratory failure.
253 A central role in host response displays the cytokine IL-6. On one hand, IL-6 binds to
254 liver cells inducing the release of C-reactive protein that binds to phosphocholine of
255 dead cells and recruits phagocytes. On the other hand, IL-6 stimulates the production
256 of neutrophils and, thus, indirectly induces NETosis. Therefore, it is reasonable to
257 inhibit IL-6 in a therapeutic intervention. However, immunomodulators can only be
258 administered to tackle the overreaction of the immune system. Immunomodulators
259 given in the early infection phase are harmful for patients. In contrast to IL-6 inhibitors,
260 CDK4/6 inhibitors selectively block the NET formation and have no impact on other
261 important host immune reactions such as phagocytosis so that they can be given
262 earlier in the infection than IL-6 inhibitors filling the therapeutic gap between vaccines
263 and monoclonal antibodies in early infection and immunomodulators in the late stage.
264 Our statistical analyses identified an independent effect of neutrophil cell count on
265 critically ill COVID-19 status providing evidence that therapeutic intervention at this
266 later stage is still effective. We therefore hypothesize that CDK4/6 inhibitors are
267 superior to IL-6 inhibitors in the treatment of critically ill COVID-19 supported by the
268 reported cases of breast cancer patients and their disease course during a CDK4/6
269 therapy.

270 In view of the evidence we have presented here it is interesting to note the report of
271 Grinshpun *et al.*[33] that a breast cancer patient on a CDK4/6 inhibitor therapy had a
272 unique disease course, halting the full presentation of the disease. Once the drug was
273 withdrawn, the full classic spectrum of illness appeared, including a bothering
274 desaturation necessitating prolongations of hospital stay for close monitoring of the
275 need for invasive ventilations.[33]
276 To conclude – we propose CDK6 as a new and plausible drug target and the
277 repurposing of already-approved breast cancer drugs abemaciclib, ribociclib,
278 trilaciclib, and palbociclib as a possibly treatment against critically ill COVID-19.
279 CDK4/6 inhibitors have the advantage of targeting the thrombo inflammation earlier
280 and more selectively than reported IL-6 inhibitors. Additionally, CDK4/6 inhibitors are
281 chemical compounds and, from there, easier to store and administer than monoclonal
282 antibodies. Clinical evidence for CDK6 inhibitors in treating critically ill COVID-19 have
283 been already reported. Further clinical investigations are ongoing. We have also
284 presented a novel methodology to find and test gene drug target candidates for drug
285 repurposing in population data. Our approach is not limited to drug repurposing, but
286 can also be used to validate new drug targets. This highlights the importance of
287 Biobanks for global health in a pandemic.
288

289 # Methods

290 ## Recruitment of cases and controls

291 The COVID-19 phenotype was defined based on the rich information made available
292 by the UK Biobank project, which has been collecting COVID-19 outcomes for their
293 large cohort of patients. The COVID-19 outcomes up until 28th October 2020 were
294 collected and cases were defined as reported previously.[8] Briefly, severe cases were

295 defined as patients who died or were hospitalized (cause of death or diagnosis
296 contains an ICD10 code for COVID-19 U071) or were on a ventilator (operation code
297 contains an E85*). Cases were not filtered based on COVID-19 test outcomes. The
298 remaining individuals from the UK Biobank were defined as potential controls.
299 Subsequently, patients of European ancestry were selected, and cases and controls
300 were matched for similar age and sex distributions, resulting in a matched number of
301 cases and controls (total number of subjects 16,307 with 8,153 cases and 8,154
302 controls). The dataset was used to perform a number of analyses described below.
303 The infection disease phenotype was created based on UK Biobank data for
304 respiratory infections, acute respiratory distress syndrome, influenza, and pneumonia
305 with hospitalization or death as a result. We aggregated hospital in-patient and death
306 register data from over 337,000 individuals in UK Biobank for ICD codes
307 corresponding to J00-J06, J09-J18, J80, and J20-J22 (UK Biobank Fields 41202,
308 41204, 40001, 40002, 41201, and 41270 for diseases and 40001 and 40002 for
309 deaths). The remaining individuals from the UK Biobank were defined as potential
310 controls. European ancestry was carried over and selected based on UK Biobank
311 phenotype as well as covariates information (age and sex) for the resulting dataset.
312 Covariate distributions were matched to result in the same number of cases and
313 controls. Variants reported by Pairo-Castineira *et al.*[8] and Ellinghaus *et al.*[7] as well as
314 variants reported by the ClinVar database[35] for the genes reported by the papers were
315 included in the dataset.
316

317 **GWAS**

318 The UK Biobank genotypes for the cases and controls were extracted to create a
319 dataset that was then submitted to a series of quality control steps with an aim to

320 remove biases in the downstream analysis as described in Marees *et al.*[34] First we
321 filtered SNPs and individuals based on their missingness in the dataset. This excludes
322 SNPs that have a high proportion of subjects where genotyping information is
323 unavailable or of poor quality. Similarly, individuals where a large proportion of SNPs
324 could not be measured were excluded. This was achieved in two steps, where first a
325 lenient threshold of 0.2 (i.e. > 20%) was applied to remove the clear outliers, followed
326 by a more stringent threshold of 0.02 (i.e. > 2%). SNP filtering was performed before
327 individual filtering. Next, all variants not on autosomal chromosomes were removed.
328 Next, variants that deviate from the Hardy-Weinberg equilibrium were removed in a
329 two-step process whereby we first applied a lenient threshold of 1e-6, followed by a
330 more stringent threshold of 1e-10. This is a common indicator of genotyping errors.
331 Thereafter, individuals were filtered out based on their heterozygosity rates which can
332 indicate sample contamination. Individuals deviating by more than 3 standard
333 deviations from the mean of the rate from all samples were filtered out. To assess the
334 heterozygosity rate per sample, those variants that were in linkage disequilibrium with
335 each other were extracted by scanning the genome at a window size of 50 variants, a
336 step size of 5, and a pairwise correlation threshold of 0.2. Next, related individuals
337 were removed. To achieve this, their identity by descent coefficients (IBD) were
338 calculated and only one individual per related cluster was kept. Then, the small
339 proportion of missing genotypes were imputed and additional variants reported by
340 Pairo-Castineira *et al.*[8] and Ellinghaus *et al.*[7] as well as variants reported by ClinVar
341 database[35] were included in the dataset from the UK Biobank imputed variants. This
342 yielded a dataset with a total number of 335,332 quality controlled variants. Finally,
343 the population structure of the samples was analyzed in two stages to identify internal
344 stratifications, which was used to filter out any individuals not predicted close enough

345 to a European reference cohort. A genome-wide association analysis was performed
346 with the R-package SAIGE[36] to score individual variants for their association with
347 critically ill COVID-19.
348

349 Regression modeling

350 Logistic regression models were fitted using the *glm* function in R (www.R-project.org).
351 A drop one model comparison procedure was performed in order to determine whether
352 each of a set of traits account for unique variance in critically ill COVID-19 disease
353 status.
354

355 Propensity score analysis

356 Using the method of Imai and Van Dyk[21], we regressed neutrophil count on BMI, age,
357 and gender and then used the resulting model predictively in order to generate
358 propensity values for our population. We then stratified the population into propensity
359 deciles (the members of which are approximately matched in terms of their
360 propensity), estimated the effects of z-transformed neutrophil count on disease status
361 within each band using logistic regression, and calculated a weighted average.
362

363 Mendelian randomization

364 We used independent GWAS summary data for neutrophil cell count (exposure)
365 published by Vuckovic *et al.*[23] (GCST90002398 downloaded January 15th 2021) and
366 summary data for critically ill COVID-19 (outcome) published by the COVID-19 Host
367 Genetics Initiative (https://www.covid19hg.org/results - COVID19hg GWAS meta-
368 analyses round 5 release date January 18th 2021). Two-sample MR analyses were
369 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.

Competing interests

J.E.K., C.N.J.R., M.R.J.L., H.A.B., J.L.C., 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. & Martinez, F. J. Severe Covid-19. *N. Engl. J. Med.* **383**, 2451–2460 (2020).
2. Fajgenbaum, D. C. & June, C. H. Cytokine Storm. *N. Engl. J. Med.* **383**, 2255–2273 (2020).
3. Millar, J. E. *et al.* Robust, reproducible clinical patterns in hospitalised patients with COVID-19. *medRxiv* (2020).
4. Rouse, B. T. & Sehrawat, S. Immunity and immunopathology to viruses: what decides the outcome? *Nat. Rev. Immunol.* **10**, 514–526 (2010).

393 5. Casanova, J.-L. Severe infectious diseases of childhood as monogenic inborn
394 errors of immunity. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E7128–37 (2015).
395 6. Horby, P., Nguyen, N. Y., Dunstan, S. J. & Baillie, J. K. An updated systematic
396 review of the role of host genetics in susceptibility to influenza. *Influenza Other*
397 *Respi. Viruses* **7 Suppl 2**, 37–41 (2013).
398 7. Severe Covid-19 GWAS Group *et al.* Genomewide Association Study of Severe
399 Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
400 8. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19.
401 *Nature* **591**, 92–98 (2021).
402 9. Zhang, Q. *et al.* Inborn errors of type I IFN immunity in patients with life-
403 threatening COVID-19. *Science* **370**, (2020).
404 10. The COVID-19 Host Genetics Initiative & Ganna, A. Mapping the human genetic
405 architecture of COVID-19 by worldwide meta-analysis. *medRxiv*
406 2021.03.10.21252820 (2021).
407 11. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic
408 data. *Nature* **562**, 203–209 (2018).
409 12. Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial
410 Version 7). *Chin. Med. J.* **133**, 1087–1095 (2020).
411 13. Fay, M. P. & Proschan, M. A. Wilcoxon-Mann-Whitney or t-test? On assumptions
412 for hypothesis tests and multiple interpretations of decision rules. *Stat. Surv.* **4**,
413 1–39 (2010).
414 14. Dunn, O. J. Multiple Comparisons among Means. *J. Am. Stat. Assoc.* **56**, 52–64
415 (1961).

416 15. Osuchowski, M. F. *et al.* The COVID-19 puzzle: deciphering pathophysiology and
417 phenotypes of a new disease entity. *Lancet Respir Med* (2021)
418 doi:10.1016/S2213-2600(21)00218-6.
419 16. Zhang, B. *et al.* Clinical characteristics of 82 cases of death from COVID-19.
420 *PLOS ONE* vol. 15 e0235458 (2020).
421 17. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide
422 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids*
423 *Res.* **47**, D1005–D1012 (2019).
424 18. Okada, Y. *et al.* Genome-wide association study for C-reactive protein levels
425 identified pleiotropic associations in the IL6 locus. *Hum. Mol. Genet.* **20**, 1224–
426 1231 (2011).
427 19. Hübel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in
428 anorexia nervosa. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **180**, 428–438
429 (2019).
430 20. Lin, B. D. *et al.* 2SNP heritability and effects of genetic variants for neutrophil-to-
431 lymphocyte and platelet-to-lymphocyte ratio. *J. Hum. Genet.* **62**, 979–988 (2017).
432 21. Imai, K. & van Dyk, D. A. Causal inference with general treatment regimes. *J.*
433 *Am. Stat. Assoc.* **99**, 854–866 (2004).
434 22. Katan, M. B. Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet*
435 **1**, 507–508 (1986).
436 23. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and
437 Diseases. *Cell* **182**, 1214–1231.e11 (2020).
438 24. Burgess, S. Sample size and power calculations in Mendelian randomization with
439 a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–
440 929 (2014).

441 25. RECOVERY Collaborative Group *et al.* Dexamethasone in Hospitalized Patients
442 with Covid-19. *N. Engl. J. Med.* **384**, 693–704 (2021).
443 26. REMAP-CAP Investigators *et al.* Interleukin-6 Receptor Antagonists in Critically
444 Ill Patients with Covid-19. *N. Engl. J. Med.* (2021) doi:10.1056/NEJMoa2100433.
445 27. Schaller, T. *et al.* Postmortem Examination of Patients With COVID-19. *JAMA*
446 **323**, 2518–2520 (2020).
447 28. Nicholls, J. M. *et al.* Lung pathology of fatal severe acute respiratory syndrome.
448 *Lancet* **361**, 1773–1778 (2003).
449 29. Tian, S. *et al.* Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus
450 (COVID-19) Pneumonia in Two Patients With Lung Cancer. *J. Thorac. Oncol.* **15**,
451 700–704 (2020).
452 30. Brinkmann, V. *et al.* Neutrophil extracellular traps kill bacteria. *Science* **303**,
453 1532–1535 (2004).
454 31. Cheng, O. Z. & Palaniyar, N. NET balancing: a problem in inflammatory lung
455 diseases. *Front. Immunol.* **4**, 1 (2013).
456 32. Amulic, B. *et al.* Cell-Cycle Proteins Control Production of Neutrophil
457 Extracellular Traps. *Dev. Cell* **43**, 449–462.e5 (2017).
458 33. Grinshpun, A., Merlet, I., Fruchtman, H. & Nachman, D. A Protracted Course of
459 COVID19 Infection in a Metastatic Breast Cancer Patient During CDK4/6 Inhibitor
460 Therapy. *Front. Oncol.* **10**, 1085 (2020).
461 34. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies:
462 Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, e1608
463 (2018).
464 35. Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic Acids*
465 *Res.* **48**, D835–D844 (2020).

466 36. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample
467 relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–
468 1341 (2018).

Figures & Tables

Figure 1

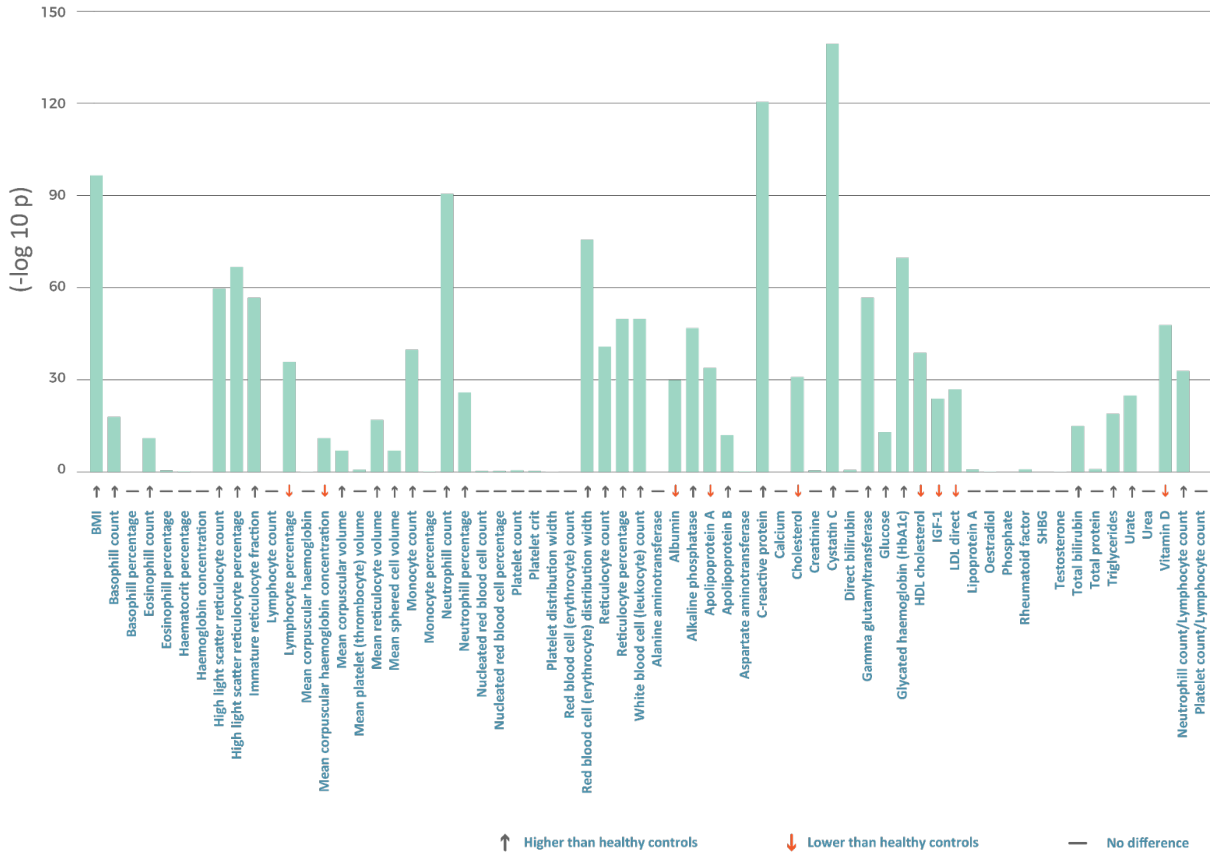
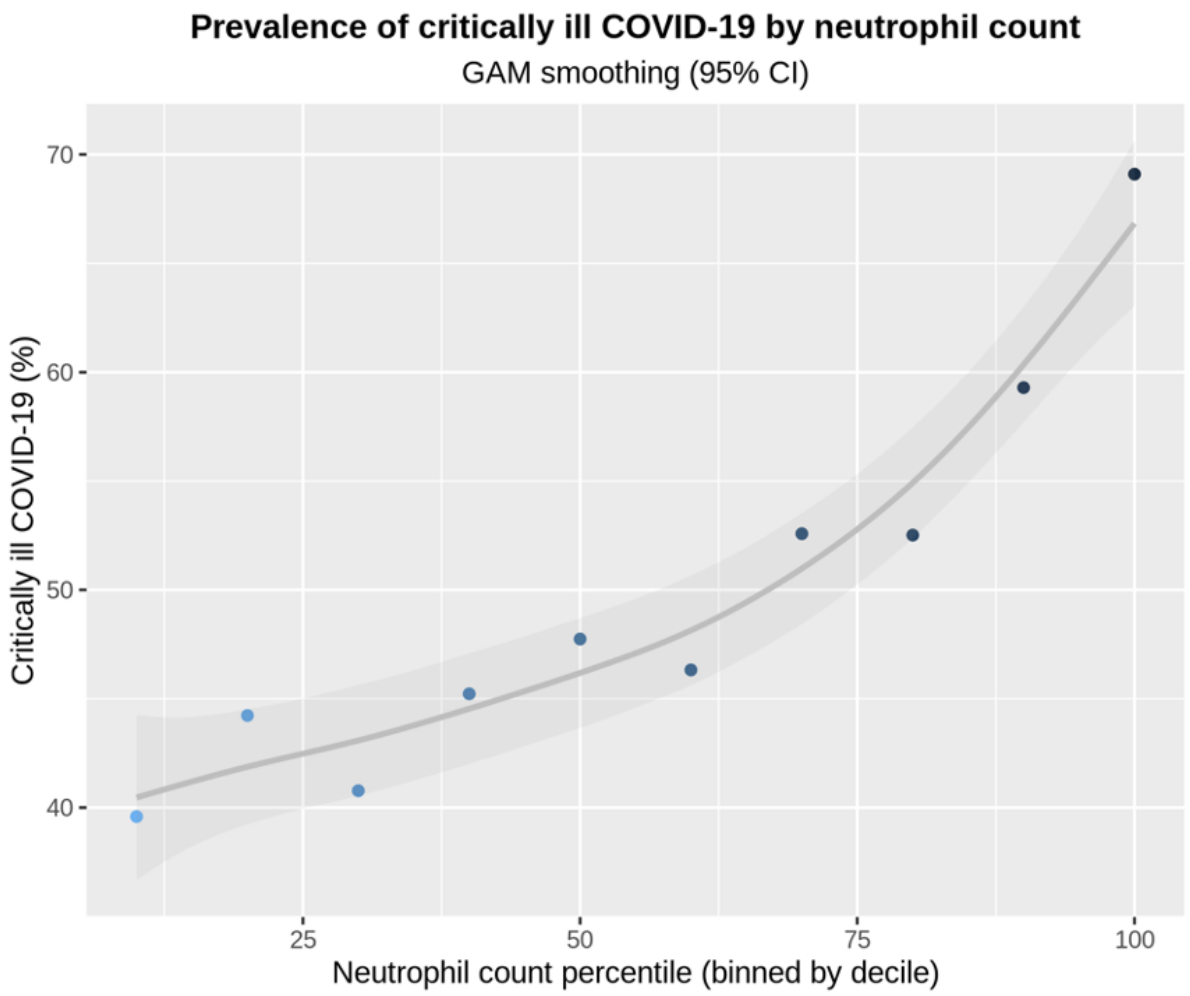


Fig. 1. Statistically significant differences in 36 traits confirmed by Mann-Whitney U-test and Bonferroni correction. In these measures taken years prior to infection, cases showed significant differences in the characteristics in various traits that have been later described as phenotypes for critically ill COVID-19. For instance, cases had higher body mass index (BMI), higher reticulocyte cell count, higher inflammatory markers such as alanine aminotransferase, C-reactive protein, cystatin C, neutrophil cell count, and higher glycated hemoglobin (HbA1c), but lower HDL and LDL cholesterol as well as lower vitamin D level than healthy controls.

483 **Figure 2**



484

485 **Fig. 2.** Regression model uncovered that neutrophil cell count measured years prior

486 to infection correlates with the prevalence of critically ill COVID-19.

487

488

489

490

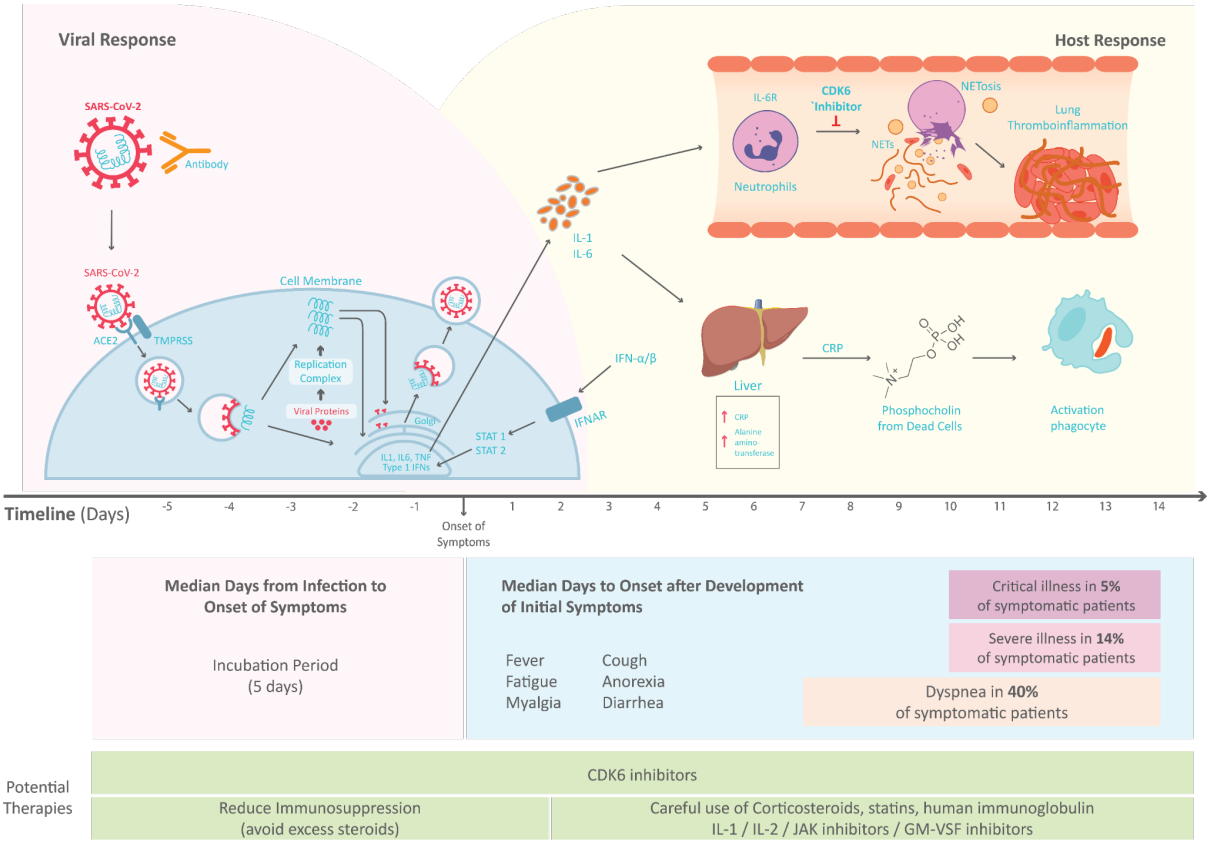
491

492

493

494

495 **Figure 3**



496

497 **Fig. 3.** COVID-19 infection stages. In the beginning of the infection the virus enters
498 the cell and starts viral replication. Here, vaccines and monoclonal antibodies block
499 the viral entry. After entry into the cell via the ACE2 receptor, currently, no therapeutic
500 option for intervention exists. First immune reactions are observed such as a decrease
501 of lymphocyte cell count (lymphopenia). This phase is called the viral response phase.
502 In the beginning of the next phase, so-called host response phase, an overreaction of
503 the host immune system occurs. The detailed mechanism is still unknown why some
504 patients' host immune systems overreact. A central role in host response displays the
505 cytokine IL-6. On one hand, IL-6 binds to liver cells inducing the release of C-reactive
506 protein that binds to phosphocholine of dead cells and recruits phagocytes. On the
507 other hand, IL-6 stimulates the production of neutrophils and, thus, indirectly induces
508 NETosis. Therefore, it is reasonable to inhibit IL-6 in a therapeutic intervention.

However, immunomodulators can only be administered to tackle the overreaction of the immune system. Immunomodulators given in the early infection phase are harmful for patients. In contrast to IL-6 inhibitors, CDK4/6 inhibitors selectively block the NET formation and have no impact on other important host immune reactions such as phagocytosis so that they can be given earlier in the infection than IL-6 inhibitors filling the therapeutic gap between vaccines and monoclonal antibodies in early infection and immunomodulators in the late stage.

533 **Table 1**

534 **Tab. 1.** Critically ill COVID-19 status was regressed on age, alanine
535 aminotransferase, BMI, C-reactive protein, and neutrophil cell count. All traits other
536 than age were found to explain unique variance in disease status. This table reports
537 log likelihood ratios, standard errors and likelihood ratio tests from the drop one
538 procedure for each.

	Estimate	**Std. Error**	χ^2	***p* value**
(Intercept)	-2.550189	0.106980		
Alanine aminotransferase	0.004023	0.001190	11.94	0.00055
BMI	0.055657	0.003507	264.15	< 2.2e-16
C-reactive protein	0.026724	0.003629	59.61	1.155e-14
Neutrophil cell count	0.167327	0.011667	215.99	< 2.2e-16

539
540
541
542
543
544
545
546

547 ## Table 2

548 **Tab. 2.** Propensity score matching was employed in order to determine whether a
549 causal relationship exists between neutrophil cell count and the prevalence of critically
550 ill COVID-19. A weighted average across propensity deciles indicates a significant
551 effect of neutrophil cell count (Log Odds Ratio = 0.3; SE = 0.018; *p* = 2.7354e-61) on
552 disease status. The table shows the neutrophil cell count in the critically ill COVID-19
553 and control groups across the propensity deciles.

Propensity score decile	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Controls neutrophil cell count in 10^9 cells/liter	4.08	3.98	4.1	4.17	4.3	4.39	4.39	4.49	4.58	4.8
Cases neutrophil cell count in 10^9 cells/liter	4.61	4.68	4.55	4.6	4.77	4.84	4.85	4.85	4.9	5.2

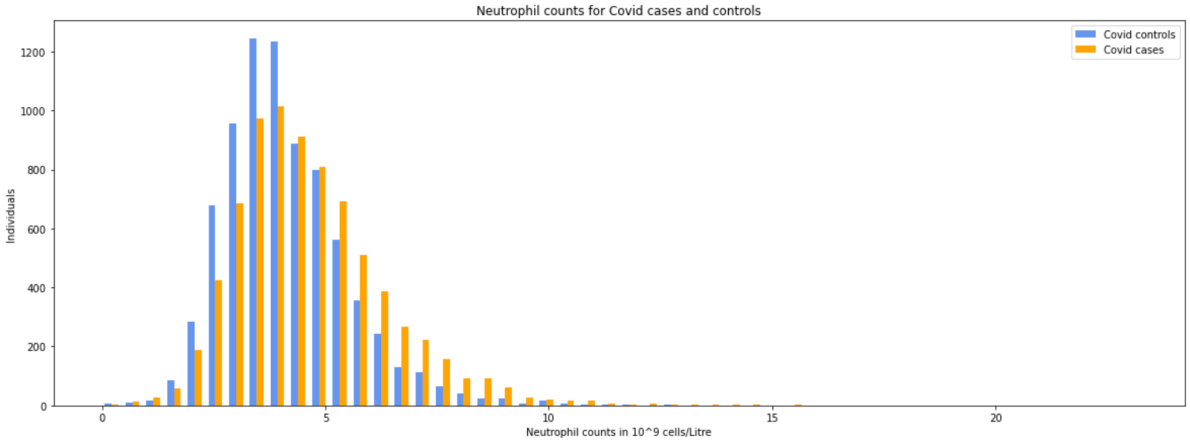
554

555

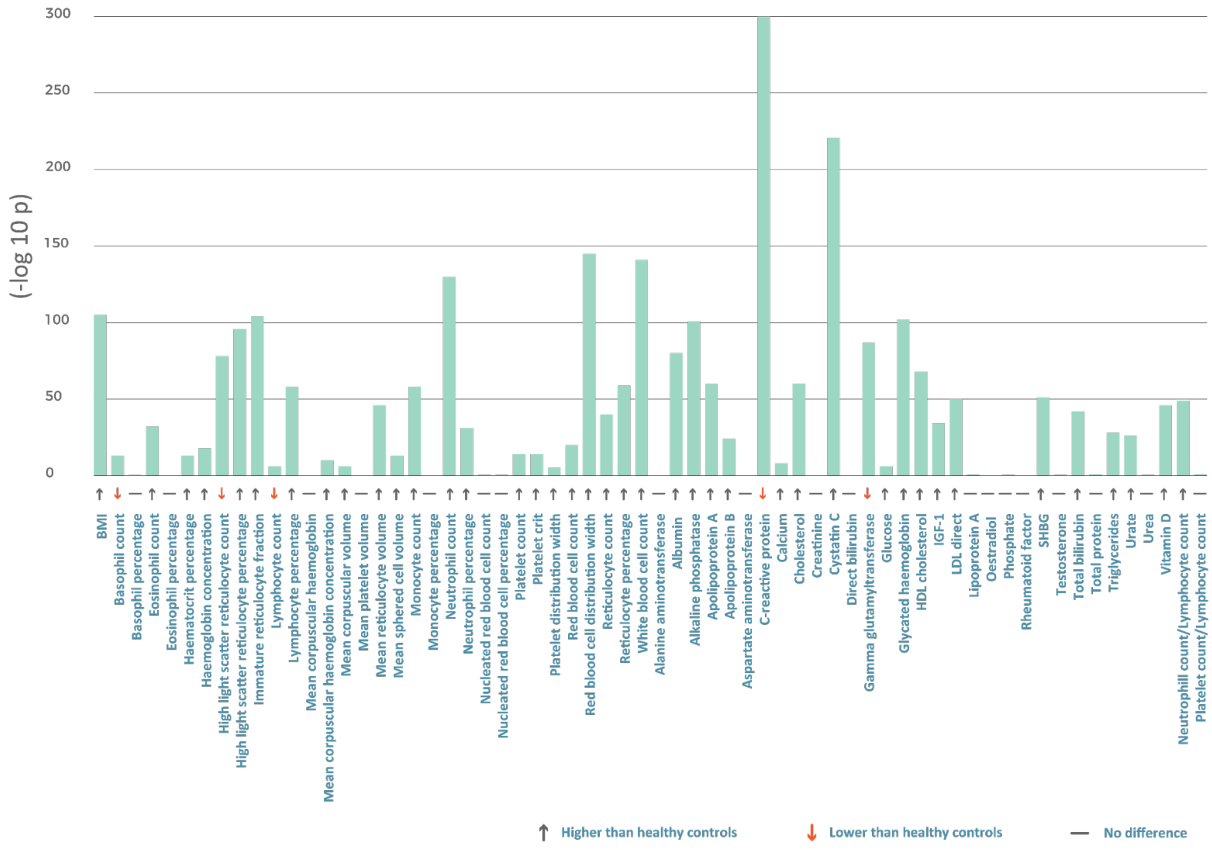
556

557

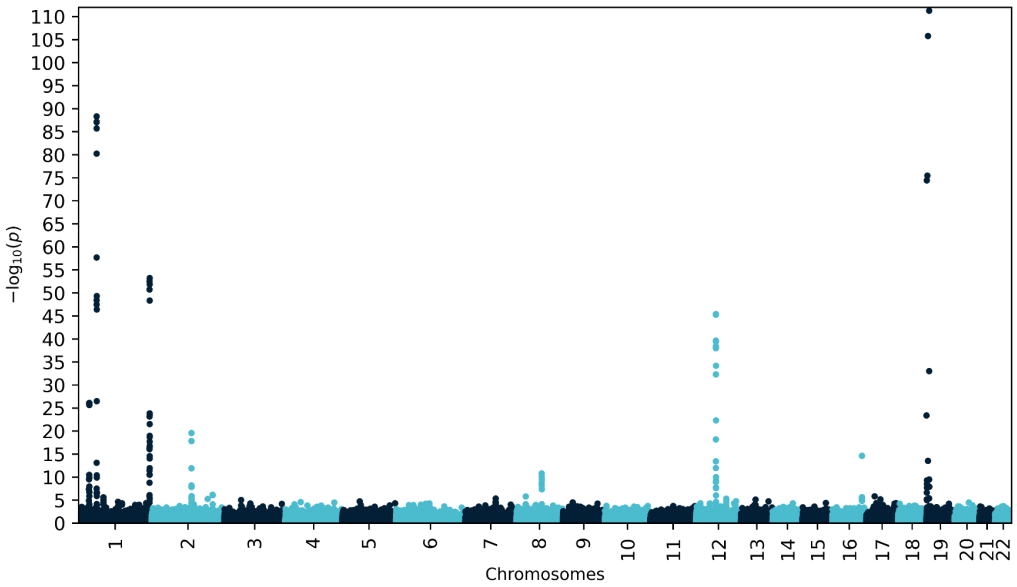
Supplementary information



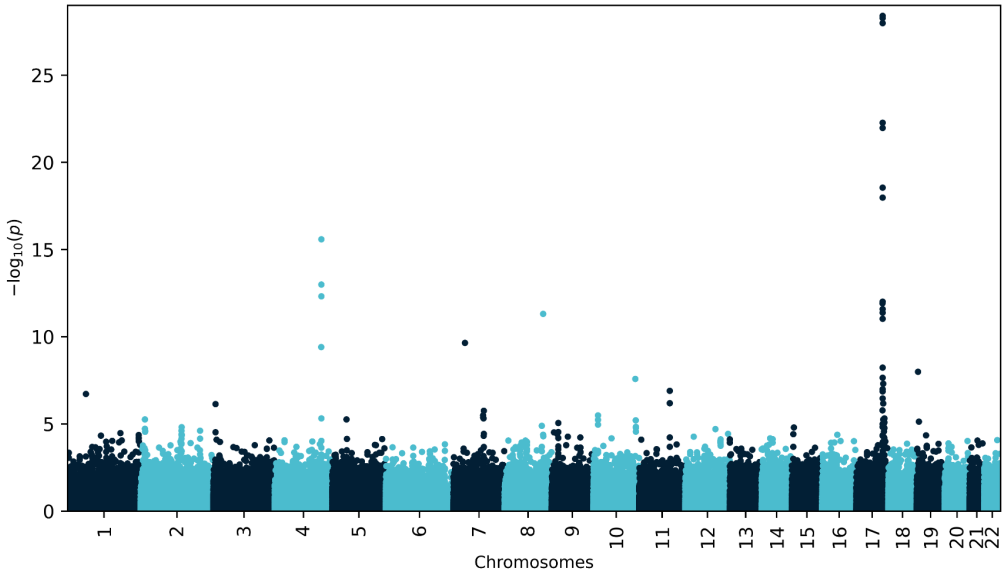
SI Fig. 1. Distribution of neutrophil cell count in critically ill COVID-19 cases (orange) and healthy controls (blue).



SI Fig. 2. As an additional comparison group we selected infection disease cases from the UK Biobank. In total 23,348 participants had been diagnosed with a respiratory infection, acute respiratory distress syndrome, influenza, or pneumonia and were hospitalized or died as a result. Again, the cases were matched with healthy controls in order to screen 64 candidate predictive traits for statistically significant differences in these two groups. Analogously to the COVID-19 investigations we here also found that cases had higher body mass index (BMI), higher reticulocyte cell count, higher inflammatory markers such as alanine aminotransferase, C-reactive protein, cystatin C, neutrophil cell count, and higher glycated hemoglobin (HbA1c), but lower HDL and LDL cholesterol as well as lower vitamin D level than healthy controls.



SI Fig. 3. Manhattan plot of C-reactive protein indicating that we did not reproduce the reported *IL-6* signal (rs2097677) on chromosome 7.



SI Fig. 4. Manhattan plot of neutrophil cell count showing that we reproduce the reported *CDK6* signal (rs445) on chromosome 7.

606 **SI Tab. 1.** Distribution of allele carriers rs2097677 (*IL-6*), rs445 (*CDK6*), and rs 42044

607 (*CDK6*) in cases and controls.

	Cases	**Controls**
rs2097677 AA	492	504
rs2097677 AG	2944	2981
rs2097677 GG	4418	4445
rs445 CC	81	74
rs445 CT	1365	1419
rs445 TT	6417	6446
rs42044 GG	581	632
rs42044 GT	3140	3102
rs42044 TT	4133	4191

608

609

610

611

612

613

614

615

616

617

618

619

620

621 **SI Tab. 2.** Distribution of allele combination carriers rs445 (*CDK6*), and rs 42044
622 (*CDK6*) in cases and controls.

	rs42044 GG	**rs42044 GT**	**rs42044 TT**
rs445 CC	352 cases / 372 ctrls	2348 cases / 2295 ctrls	3709 case / 3768 ctrls
rs445 TC	199 cases / 230 ctrls	754 cases / 778 ctrls	411 cases / 408 ctrls
rs445 TT	30 cases / 30 ctrls	38 cases / 29 ctrls	13 cases / 15 ctrls

623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641

642 **SI Tab. 3.** The two sample MR analyses here showed that for neutrophil cell count as
643 exposure and critically ill COVID-19 as outcome while using strict clumping
644 parameters, no significant effect was detected. This approach is limited by the quality
645 and sample size of critically ill COVID-19 GWASs.

	Clumping parameters	**SNPs used**	**beta**	**SE**	**IVW p-value**	**Pleiotropy test**
Exposure: neutrophil cell count Outcome: critically ill COVID-19	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

646

647

648