Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity

Satoshi Yoshiji[1,2,3,4], Guillaume Butler-Laporte[1,5], Tianyuan Lu[1,6], Julian Daniel Sunday Willett[1,6], Chen-Yang Su[1,7], Tomoko Nakanishi[1,2,3,4], David Morrison[1], Yiheng Chen[1,2], Kevin Liang[1,5], Michael Hultström[1,8], Zaman Afrasiabi[1], Shanshan Lan[1], Naomi Duggan[1], Chantal DeLuca[1], Mitra Vaezi[1], Chris Tselios[1], Xiaoqing Xue[1], Meriem Bouab[1], Fangyi Shi[1], Laetitia Laurent[1], Hans Markus Münter[9], Marc Afilalo[1,10], Jonathan Afilalo[1,5,11], Vincent Mooser[2,9], Nicholas J Timpson[12], Hugo Zeberg[13,14], Sirui Zhou[1,2,9], Vincenzo Forgetta[1,15], Yossi Farjoun[1,15], J. Brent Richards[1,2,5,15,16*]

[1] Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada
[2] Department of Human Genetics, McGill University, Montréal, Québec, Canada
[3] Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[4] Japan Society for the Promotion of Science, Japan
[5] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
[6] Quantitative Life Sciences Program, McGill University, Montréal, Québec, Canada
[7] Department of Computer Science, McGill University, Montréal, Québec, Canada
[8] Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Sweden
[9] McGill Genome Centre, McGill University, Montréal, Québec, Canada
[10] Department of Emergency Medicine, Jewish General Hospital, McGill University, Montréal, Québec, Canada
[11] Division of Cardiology, Jewish General Hospital, McGill University, Montréal, Québec, Canada
[12] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
[13] Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[14] Department of Neuroscience, Karolinska Institute, Stockholm, Sweden
[15] 5 Prime Sciences, Montréal, Québec, Canada
[16] Department of Twin Research, King's College London, London, United Kingdom

***Correspondence:**
J. Brent Richards
Professor of Medicine, McGill University, Senior Lecturer, King's College London (Honorary), Pavilion H-413, Jewish General Hospital, 3755 Côte-Ste-Catherine Montréal, Québec, H3T 1E2, Canada.
Tel: +1-514-340-8222 Fax: +1-514-340-7529

Abstract

Obesity is a major risk factor for COVID-19 severity; however, the underlying mechanism is not fully understood. Considering that obesity influences the human plasma proteome, we sought to identify circulating proteins mediating the effects of obesity on COVID-19 severity. We first screened 4,907 plasma proteins to identify proteins influenced by body mass index (BMI) using Mendelian randomization (MR). This yielded 1,216 proteins, whose effects on COVID-19 severity were assessed, again using MR. This two-step approach identified nephronectin (NPNT), for which a one standard deviation increase was associated with severe COVID-19 (odds ratio = 1.71, 95% CI: 1.45–2.02, $P = 1.63 \times 10^{-10}$). Colocalization analyses indicated that an NPNT splice isoform drove this effect. Overall, NPNT mediates 3.7% of the total effect of BMI on severe COVID-19. Finally, we found that decreasing body fat mass and increasing fat-free mass can lower NPNT levels and thus may improve COVID-19 outcomes. These findings provide actionable insights into how obesity influences COVID-19 severity.

52 ## Introduction

53 Coronavirus disease (COVID-19) has claimed more than 6 million lives globally since the start of
54 the pandemic[1], and obesity increases the risk of severe COVID-19[2,3]. To date, multiple pathways
55 have been explored as mechanisms linking obesity to COVID-19 severity, including metabolic
56 abnormalities, systemic inflammation, and respiratory dysfunction (e.g., impaired gas exchange)[3-
57 5]. However, the underlying mediators through which obesity influences COVID-19 outcomes are
58 not fully understood. One strategy for disentangling this relationship is to identify circulating
59 proteins mediating the effects of obesity on COVID-19 outcomes. As circulating proteins can be
60 measured and, in some cases modulated, identification of mediator proteins may provide insights
61 on the pathways whereby obesity increases the risk of severe COVID-19 and offer potential
62 targets for therapeutic interventions.
63
64 A previous large cross-sectional study showed that body mass index (BMI) is significantly
65 associated with changes in the plasma levels of 1,576 proteins[6], and another study supported the
66 considerable influence of obesity on the plasma proteome[7]. However, as proteins are intricately
67 involved in complex biological processes, observational studies may be biased by unmeasured
68 confounding factors and reverse causation. Considering that COVID-19 has been associated with
69 substantial changes in the levels of circulating proteins[8], the effects of circulating proteins on
70 COVID-19 severity are also subject to such biases.
71
72 Mendelian randomization (MR) can help to protect against these biases. MR is a method
73 that uses genetic variants as instrument variables to evaluate the causal effects of exposures (risk
74 factors) on outcomes. As genetic variants are randomly allocated at the time of conception, they
75 are largely independent of environmental and other potential confounders, thereby decreasing
76 the risk of confounding. Additionally, they are not subject to reverse causation as the allocation of
77 genetic variants always precedes the onset of diseases[9,10].
78
79 From the beginning of the pandemic, MR has played a critical role in providing evidence
80 of modifiable risk factors for COVID-19[2,11,12]. Furthermore, multiple MR studies identified potential
81 therapeutic targets for COVID-19, including OAS1, ABO, IFNAR2, IL-6, ELF5, and FAS[12-18].
82 Indeed, some MR findings have been validated by randomized controlled trials, demonstrating
83 the utility of MR[19-24].
84

Notably, MR is based on several instrumental variable assumptions:[9,10] (I) the genetic variants used as instrumental variables are associated with the exposure; (II) they are not associated with factors that confound the relationship between the exposure and the outcome and; (III) they influence the outcome only through exposure (also known as a lack of horizontal pleiotropy). The most problematic of these assumptions is the last one; consequently, careful assessment of horizontal pleiotropy is required. Nevertheless, with proper use, MR can serve as a powerful tool to help understand the causal mechanisms in humans[9,10], and in the case of obesity and COVID-19, can be used to screen thousands of proteins that may help to explain this relationship.

Here, we integrated proteome-wide MR using the large-scale aptamer-based plasma protein measurements, multiple sensitivity analyses, mediation analysis, colocalization, fine-mapping, and single-cell RNA-sequencing analysis to identify plasma proteins mediating the effect of BMI on COVID-19 severity.

Results

Study overview

The study was conducted in the following manner (**Figure 1**):

1) **Step 1 MR**: We performed two-sample MR to estimate the effect of BMI on circulating protein levels using BMI genome-wide association study (GWAS) data for 694,649 individuals from GIANT and UK Biobank by Yengo *et al*.[25] in addition to the largest proteomic GWAS data from the deCODE study[26] that measured plasma protein abundance using 4907 aptamers in 35,559 individuals. Hereafter, “protein” is used when referring to an aptamer targeting a protein[26].
2) **Step 2 MR**: Next, we performed two-sample MR to estimate the causal effects of the above-identified proteins (BMI-driven proteins) on COVID-19 outcomes. To ensure that the MR findings were not biased by linkage disequilibrium (LD), we used colocalization analyses to evaluate whether the cis-acting protein quantitative loci (cis-pQTLs) of the identified proteins and COVID-19 outcomes shared the same causal variant. For validation, we repeated MR using cis-pQTLs from two independent studies (the FENLAND study[27] and the AGES Reykjavik study[28]). For nephronectin (NPNT), which was robustly associated with COVID-19 severity, we also investigated the consistency between MR findings and observational associations between BMI and NPNT and NPNT and COVID-19 using independent cohorts.

3) **Mediation analysis**: for NPNT, we performed an MR mediation analysis to quantify the extent to which the total effect of BMI on severe COVID-19 was mediated by plasma NPNT levels.
4) **Follow-up analysis for NPNT**: Furthermore, to better understand whether the total levels of the NPNT protein, NPNT transcripts, or NPNT splice isoforms influenced COVID-19 severity, we performed colocalization analyses to identify whether the same causal variant was shared between COVID-19 severity and NPNT protein level, NPNT transcript level, and NPNT splice isoform level. Moreover, we analyzed single-cell RNA-sequencing data of the lung autopsy samples from patients who died due to COVID-19 to gain insights into the biological role of NPNT. Finally, we estimated the independent causal effect of body fat and fat-free mass on plasma NPNT levels to identify a way of modulating plasma NPNT levels.

Identifying the effects of BMI on plasma proteins (Step 1 MR)

We performed a two-sample MR using BMI as the exposure and 4,907 plasma protein levels as the outcomes (see **Methods** and **Supplementary Table 1** for details). Out of 4,907 proteins screened, 1,304 were estimated to be influenced by BMI using a Bonferroni-adjusted threshold of $P < 1.0 \times 10^{-5}$ (0.05/4907), highlighting the substantial influence of obesity on plasma protein levels (**Figure 2**). Among them, 1,229 proteins showed no apparent sign of horizontal pleiotropy using the MR Egger test[29] ($P_{\text{Egger intercept}} > 0.05$) (**Supplementary Table 2)**.

To assess potential reverse causation, whereby the proteins influenced BMI, we performed a bidirectional MR that used protein levels as exposures and BMI as the outcome (see **Methods**). Thirteen proteins exhibited bidirectional effects (**Supplementary Table 3**) and were excluded from further analyses. Hence, a total of 1,216 protein levels were identified as BMI-driven proteins, which were estimated to be influenced by BMI with no detectable horizontal pleiotropy or reverse causation. These protein levels, including NPNT and HSD17B14 (hydroxysteroid 17-beta dehydrogenase 14), were used as exposures in the second step of our study.

Identifying the effects of BMI-driven proteins on COVID-19 severity (Step 2 MR)

Next, we evaluated the causal effects of the above-obtained, BMI-driven proteins on COVID-19 outcomes. We used cis-acting protein quantitative loci (cis-pQTLs) for these proteins as instrumental variables and COVID-19 outcomes from the COVID-19 Host Genetics Initiative GWASs (release 7)[2] as outcomes (**Supplementary Table 1)**. We did so to protect against bias

153 from horizontal pleiotropy. This is because cis-pQTLs reside near the transcription start site of the
154 protein-coding gene and are more likely to directly influence the protein levels than the trans-
155 pQTLs, thus reducing the risk of horizontal pleiotropy[30,31]. If such cis-pQTLs only influence
156 transcription or translation of their associated gene, then the risk of horizontal pleiotropy would be
157 importantly reduced. We searched for the cis-pQTLs for each of 1,216 BMI-driven proteins using
158 the deCODE study[26]. Following the cis-pQTL search and data harmonization, 334, 335, and 336
159 proteins were able to be tested in MR for their estimated causal effects on severe COVID-19,
160 COVID-19 hospitalization, and COVID-19 susceptibility, respectively. Based on a Bonferroni-
161 adjusted threshold of $P < 1.49 \times 10^{-4}$, MR revealed that a one standard deviation (SD) increase
162 in genetically predicted NPNT levels was associated with severe COVID-19 (OR = 1.71, 95% CI:
163 1.45–2.02, $P = 1.63 \times 10^{-10}$) and COVID-19 hospitalization (OR = 1.36, 95% CI: 1.22–1.53, $P =$
164 4.52×10^{-8}) (**Figure 3** and **Supplementary Table 4 and 5**). Similarly, a one SD increase in
165 genetically predicted HSD17B14 levels was associated with severe COVID-19 (OR = 1.92, 95%
166 CI: 1.39–2.65, $P = 6.85 \times 10^{-5}$). Neither of the proteins was convincingly associated with COVID-
167 19 susceptibility (**Supplementary Table 6**).
168
169 To verify the assumption of lack of horizontal pleiotropy, we checked whether the cis-
170 pQTLs for NPNT and HSD17B14 were associated with any traits or diseases using the
171 PhenoScanner[32] database at the genome-wide significant threshold of $P = 5\times10^{-8}$. In
172 PhenoScanner, the lead cis-pQTL for NPNT (rs34712979) from the deCODE study was
173 associated with lung-related traits (**Supplementary Table 7**). However, because NPNT has an
174 established role as an extracellular matrix protein and fibrosis in the lung[33-35], it is possible that
175 the NPNT cis-pQTL affects such traits through alteration in NPNT levels, and thus, should not
176 violate the assumption of no horizontal pleiotropy. Indeed, MR showed that NPNT levels influence
177 FEV1/FVC (key lung function index used for the definition of chronic obstructive lung disease
178 (COPD)[36]) and asthma risk (**Supplementary Table 8**). No other cis-pQTL for NPNT or HSD17B14
179 was associated with any trait or disease, reducing the possibility of horizontal pleiotropy.
180
181 Next, to assess potential bias from reverse causation, we tested whether COVID-19
182 severity influences NPNT levels. Since MR requires that the outcome GWAS is performed in
183 populations that could be exposed to the exposure of interest, we used a GWAS of plasma protein
184 levels in 264 individuals during acute infection of SARS-CoV-2 (severe acute respiratory
185 syndrome coronavirus 2) from the Biobanque Québécoise de la COVID-19 (BQC19)

186 (**Supplementary Table 1**) (see **Methods**). We did not find evidence that severe COVID-19 and
187 hospitalization influenced plasma NPNT or HSD17B14 levels (**Supplementary Table 9**).
188

189 **Colocalization analyses investigating whether cis-pQTLs for NPNT and HSD17B14 and**
190 **COVID outcomes share the same causal variant**

191 Considering that MR analyses from Step 1 and Step 2 implicated NPNT and HSD17B14 as
192 candidate proteins mediating the effect of BMI on COVID-19 severity, we performed colocalization
193 to assess whether the cis-pQTLs for NPNT and HSD17B14 shared the same single causal variant
194 with severe COVID-19 and hospitalization. The colocalization analyses revealed that NPNT had
195 a high posterior probability of colocalization with severe COVID-19 and hospitalization (posterior
196 probability (PP_{shared}) = 100.0% and 99.7%, respectively) (**Figure 4**). Moreover, using Combined
197 Annotation Dependent Depletion (CADD)-scores[37] as priors, fine-mapping further confirmed that
198 rs34712979, the lead cis-pQTL for NPNT, was fine-mapped for severe COVID-19 and
199 hospitalization (**Supplementary Table 10 and 11**). On the other hand, cis-pQTLs for HSD17B14
200 did not colocalize with either of the outcomes (**Extended Figure 1**), indicating that MR findings
201 for HSD17B14 may be confounded by LD.
202

203 **Validation analyses using cis-pQTLs for NPNT and HSD17B14 from different cohorts**

204 To validate the findings that plasma NPNT levels were causal for severe COVID-19, we performed
205 another MR analysis using cis-pQTLs from different cohorts. Additional cis-pQTLs for NPNT and
206 HSD17B14 were identified based on the FENLAND study by Pietzner *et al*. (n = 10,708)[27] and
207 the AGES Reykjavik study by Emilsson *et al*. (n = 3,200)[28]. Again, the PhenoScanner database
208 did not reveal evidence of horizontal pleiotropy for these cis-pQTLs. MR analyses using these
209 cis-pQTLs estimated a consistent causal effect of NPNT on severe COVID-19 (the AGES
210 Reykjavik study: OR = 1.25, 95% CI:1.06–1.49, $P = 8.26 \times 10^{-3}$; the FENLAND study: OR = 1.89,
211 95% CI: 1.04–1.31, $P = 7.30\times 10^{-3}$) and COVID-19 hospitalization (the AGES Reykjavik study,
212 OR = 1.17, 95% CI:1.06–1.49, $P = 8.26 \times 10^{-3}$; the FENLAND study, OR =1.45, 95% CI: 1.28–
213 1.66, $P = 2.17\times 10^{-8}$). In contrast, the estimated causal effects of HSD17B14 on these COVID-19
214 outcomes were not supported (**Supplementary Table 12**). Hence, considering the lack of
215 replication and colocalization of HSD17B14, we excluded this protein from further analyses.
216

217 **Observational evaluation of the associations between BMI and NPNT, and NPNT and**
218 **severe COVID-19**

Association between BMI and plasma proteins in 2,729 individuals from the INTERVAL study

We compared our MR findings for NPNT with results from the INTERVAL study, which evaluated the associations between BMI and 3,622 plasma protein levels among 2,729 individuals[6]. In the INTERVAL study, a one SD increase in BMI was observationally associated with NPNT (beta = 0.127, 95% CI: 0.088–0.167, $P = 4.52 \times 10^{-10}$), which was directionally consistent with the MR findings.

Association between plasma proteins and risk of COVID-19 in 463 individuals from the BQC19 cohort

Considering the MR analyses (Step 2 MR) used plasma protein levels in a non-infectious state from the deCODE study as the exposures, we observationally assessed whether the plasma levels of NPNT in a non-infectious state was associated with the risk of COVID-19 using logistic regression in 463 non-infectious state individuals from our BQC19 cohort (see **Methods** for a definition of non-infectious state**)**. Logistic regression analysis adjusting for age, sex, and sample collection batch showed that non-infectious plasma levels of NPNT were associated with increased risk of COVID-19 susceptibility (OR = 1.71 95% CI: 1.40–2.12 $P = 3.57 \times 10^{-7}$) (**Supplementary Table 13**). While effect estimates for severe and hospitalized COVID-19 were directionally consistent with COVID-19 susceptibility, the confidence intervals were wide and included the null, perhaps reflecting a small sample size of cases.

Mediation analysis of NPNT on the effect of obesity on severe COVID-19

We then performed a mediation analysis using network MR to understand the extent to which plasma NPNT levels mediate the association between BMI and severe COVID-19. For this purpose, we estimated the effect of BMI on severe COVID-19 mediated by plasma NPNT levels by deploying the product of coefficients method via the two-step MR (**Extended Figure 2**). We first estimated the effect of BMI on plasma NPNT levels and then multiplied this estimate by the effect of plasma NPNT levels on severe COVID-19 outcome adjusted for BMI using multivariable MR (see **Methods** for further details). The ratio of the overall effect mediated by NPNT was calculated by dividing the NPNT-mediated effect estimate by the total effect estimate of BMI on severe COVID-19. We found that plasma NPNT levels mediated 3.7% (95% CI: 0.7–6.7%, $P = 1.75 \times 10^{-2}$) of the effect of BMI on severe COVID-19.

252 **Colocalization analyses investigating whether cis-pQTL for NPNT and eQTLs and sQTLs**
253 **of the lung share the same causal variant**
254 Since proteins were measured using SomaScan aptamer–based measurements, the observed
255 differences in NPNT levels measured by the assay may be explained not only by total NPNT
256 levels, but also by levels of particular isoform. Thus, we evaluated whether the cis-pQTL
257 influenced either total RNA expression level or a specific isoform level of NPNT. Given that the
258 lung is a primary target organ in the context of COVID-19 severity, NPNT is highly expressed in
259 the lung, NPNT splice isoforms have previously been implicated as a risk factor for COPD and
260 other lung outcomes[38,39,40], and SNPs influencing a specific isoform level would be a splice QTLs
261 (sQTLs), we performed colocalization analysis of the cis-pQTL with eQTLs and sQTLs for NPNT
262 in lung tissue from GTEx[41].
263
264 Colocalization analysis of NPNT pQTL and sQTL within a one-megabase (1-Mb) region
265 (±500 kb) surrounding the lead cis-pQTL (rs34712979) revealed that there was a high probability
266 of the pQTL and sQTL sharing a single causal variant (posterior probability for a shared causal
267 signal (PP_{shared}) = 100.0%). However, this was not true for eQTL (PP_{shared} = 7.4%) (**Figure 5**).
268 Given that the thyroid, lung, and arteries are the top three *NPNT*-expressing tissues according to
269 the GTEx[41], as additional analyses, we performed the same colocalization analyses in thyroid and
270 arteries (aorta, tibial, and coronary) and found that pQTL of NPNT also colocalized with sQTL in
271 these tissues (PP_{shared} = 100.0% for all). It colocalized with eQTL in aorta and tibial artery (PP_{shared}
272 = 100.0% for both) but not in thyroid or coronary artery (PP_{shared} = 11.9% and 5.8%, respectively).
273 Notably, the lead cis-pQTL from the deCODE study, i.e., rs34712979, was also identified by the
274 FENLAND study[27] and has been reported to create a cryptic splice acceptor site, which inserts a
275 three-nucleotide sequence coding a serine residue at the 5'-splice site of exon 2, resulting in
276 perturbations of the alpha-helix motif[38]. Further, this lead cis-pQTL (rs34712979) and another cis-
277 pQTL (rs78213340) from the AGES Reykjavik study—that was not in high LD with rs34712979
278 (r^2 = 0.234)—were both associated with the same exon-skipping splicing in lung in GTEx
279 (**Supplementary Table 14**). Hence, two different cis-pQTLs of NPNT from three different
280 studies—all using SomaScan—were associated with the same splicing pattern, likely suggesting
281 that the SomaScan assay measures a specific isoform of the NPNT protein.
282
283 Collectively, these findings indicate that the NPNT-targeting aptamer of SomaScan could
284 reflect the specific isoform levels and that the specific isoform of NPNT with a serine insertion at
285 the N-terminus influences the effect of NPNT on COVID-19 severity.

286
287
288 **Analysis of single-cell RNA-sequencing data in lung autopsy samples of patients who died**
289 **of COVID-19**
290 To gain insights into the biological role of NPNT in SARS-CoV-2-infected lungs, we explored lung
291 cell types that significantly expressed *NPNT* gene by analyzing single-cell RNA-sequencing data
292 from lung autopsy samples (106,792 cells) of 16 patients who died of COVID-19[42] (Single Cell
293 Portal of the Broad Institute (Accession ID: SCP1052)).
294
295 We found that *NPNT* is widely expressed in lung cell types, including epithelial cells (type
296 1 and type 2 alveolar cells) and fibroblasts, highlighting its role in air exchange and fibrosis.
297 Furthermore, we conducted a subgroup analysis of SARS-CoV-2-positive and -negative cells and
298 found that *NPNT* was significantly expressed in SARS-CoV2-positive alveolar cells and fibroblasts
299 (permutation test $P < 0.001$, see **Methods** for details) (**Extended Figure 3**).
300
301 **Multivariable MR analyses to assess the effects of body fat and fat-free mass on plasma**
302 **NPNT levels**
303 Given consistent evidence that BMI influenced NPNT levels, which in turn influence COVID-19
304 severity, we aimed to identify a way of modulating plasma NPNT levels, by better understanding
305 if estimated causal effect of BMI on plasma NPNT levels was influenced by fat or fat-free mass in
306 humans. In Step 1 MR, we showed that increased BMI is estimated to increase plasma levels of
307 NPNT. However, BMI is a function of height and weight and does not take into account body
308 compartment: body fat and fat-free mass. Thus, we specifically assessed their independent
309 effects on plasma NPNT levels. For this, we performed multivariable MR using body fat and fat-
310 free mass as the exposures and plasma NPNT levels as the outcome (see **Methods**). We found
311 that a one SD increase in body fat mass was associated with increased levels of NPNT (beta =
312 0.23, 95% CI: 0.13–0.33, $P = 6.86 \times 10^{-6}$), whereas a one SD increase in body fat-free mass was
313 associated with decreased levels of NPNT (beta = -0.15, 95% CI: -0.26, -0.04, $P = 5.83 \times 10^{-3}$)
314 (**Figure 6**). These findings suggest that decreasing body fat mass and increasing fat-free mass
315 (e.g., thorough appropriate diet and exercise) can reduce plasma NPNT levels and thus reduce
316 the risk of severe COVID-19.
317
318 **Discussion**

319 In the present study, we conducted MR analyses and found that NPNT is likely to mediate the
320 effect of increased BMI on COVID-19 severity. Considering BMI is a highly polygenic trait with
321 more than 530 associated loci[25] and influences more than 1,200 circulating proteins as shown in
322 the present and previous studies[6,7], it is remarkable that a single protein accounts for ~4% of the
323 effect of BMI on severe COVID-19. Additionally, colocalization analyses provided evidence
324 that the NPNT cis-pQTL is shared with a lung sQTL for NPNT that leads to a splice isoform with
325 a serine insertion at the N-terminus of NPNT[38]. These results therefore suggest that NPNT
326 mediates a proportion of BMI's effect on severe COVID-19, and that this effect may be conferred
327 by alternative splicing of *NPNT*.
328
329 NPNT, or nephronectin, is an extracellular matrix protein that controls integrin binding
330 activity. It is known as a functional ligand of integrin α8/β-1 in kidney development and also
331 associated with the development, remodeling, and survival of various tissues through the binding
332 of integrins[43,44]. NPNT has also been implicated in inflammation and autoimmunity[45,46], which may
333 align with the suggested role of obesity-induced inflammation in COVID-19 severity[47]. In addition,
334 *NPNT* is expressed in lung alveolar cells and fibroblasts[48,49], and recent GWASs have identified
335 that the same NPNT splice variant, rs34712979, is associated with lung function-related traits
336 (e.g., forced expiratory volume (FEV1), forced vital capacity (FVC), FEV1/FVC) and COPD[39,40].
337 Our single-cell RNA-sequencing analysis on COVID-19 lung autopsy samples also found that
338 *NPNT* was expressed in SARS-CoV-2-infected alveolar cells and fibroblasts, suggesting its role
339 in air exchange and fibrosis in SARS-CoV-2-affected lungs. These findings collectively suggest
340 that NPNT alternative splicing, which results in an isoform with a serine insertion in the N-terminus
341 of the protein, increases the risk of deterioration of the lung function and lung diseases such as
342 COPD and COVID-19.
343
344 This study has both strengths and weaknesses. MR and colocalization analyses robustly
345 implicated NPNT as a causal mediator of the relationship between BMI and COVID-19 severity.
346 The robustness of the MR findings was enhanced by the large sample sizes used to derive the
347 findings. To the best of our knowledge, this is the first study to identify a mediator of BMI on
348 COVID-19 employing a two-step MR approach, although a previous study used a similar
349 framework with limited statistical power earlier in the pandemic could not identify a strong protein
350 signal[50]. Furthermore, our MR findings withstood multiple sensitivity analyses and were supported
351 by colocalization, fine-mapping, observational validation using the independent cohorts, and
352 RNA-sequencing studies in COVID-19 lung samples. Intriguingly, multivariable MR showed that

353 plasma NPNT levels were increased by body fat mass but decreased by fat-free mass, indicating
354 that the effect of BMI on plasma NPNT levels was driven by body fat mass and partially
355 counteracted by fat-free mass. These findings not only suggest that NPNT can be a potential
356 target of pharmaceutical intervention in obese individuals to prevent severe COVID-19 but also
357 highlight one of the possible mechanisms by which appropriate diet and exercise can confer risk
358 reduction of COVID-19 severity through the reduction in plasma NPNT levels.
359
360 This study also has important limitations. First, MR and colocalization analyses were
361 restricted to individuals of European ancestry to maximize statistical power, and whether NPNT
362 mediates the effect of BMI on severe COVID-19 in populations of non-European ancestries
363 requires further investigation. Second, because there were no genome-wide significant genetic
364 instrumental variables for BMI in the 1-Mb region around the cis-region of NPNT (1-Mb around
365 the transcription start site), we cannot pinpoint a single genetic variant that directly links BMI to
366 NPNT. Hence, the effect of BMI on NPNT is likely mediated by a combination of multiple trans-
367 effects, which was also indicated by the scatter plot in Step 1 MR analysis for NPNT. However,
368 this is not surprising considering BMI is a highly polygenic, complex trait, and multiple pathways
369 could confer the effect of BMI on other diseases[51]. Third, we do not know the molecular
370 mechanism by which BMI influences the splice isoform. Previous studies have suggested that
371 obesity is associated with alternative splicing[52-54] and extracellular matrix protein remodeling[55],
372 which may align with our findings. However, further investigation is required to clarify the specific
373 molecular mechanisms involved. Lastly, we do not rule out the possibility that total levels of NPNT
374 (all isoforms) measured by NPNT-targeting aptamer are mediating the effect of obesity on COVID-
375 19 severity. However, given the evidence provided by the MR, colocalization, and fine-mapping
376 as well as a biological understanding of lead cis-pQTL (i.e., the variant causes alternative splicing
377 resulting in perturbations of the alpha-helix motif in the lung), it is likely that the specific isoform
378 measured by the aptamer is driving the effect. Nevertheless, isoform-specific measurements (e.g.,
379 mass spectrometry) will be required to confirm these findings.
380
381 In conclusion, we integrated multiple lines of MR analyses and genetic colocalization
382 evidence to identify NPNT as a mediator of the effect of BMI on severe COVID-19. We also
383 disentangled the effect of BMI on NPNT and showed that NPNT is increased by body fat mass
384 and decreased by body fat-free mass, highlighting the effect of obesity on COVID-19 severity
385 through NPNT is conferred by body fat mass. These findings provide actionable insights into how
386 obesity influences COVID-19 severity.

387 ## Online Methods

388 ### Step 1: Identifying the effects of BMI on plasma proteins

389 ### Two-sample MR

390 **BMI GWAS:** We used the BMI GWAS meta-analysis with the largest sample size, comprising
391 693,529 European ancestry individuals from the GIANT consortium and UK Biobank[25]. The details
392 of the consortium are provided in **Supplementary Table 1**.
393
394 **Proteomic GWAS:** For GWAS of plasma protein levels, we used the largest proteomic GWAS
395 available[26], which measured 4907 proteins in 35,559 individuals of European ancestry using the
396 SomaLogic SomaScan assay v4 (SOMAscan, SomaLogic, Boulder, Colorado, USA).
397
398 **Two-sample MR:** The effect of BMI on plasma protein levels was assessed using the inverse
399 variance weighted method in TwoSampleMR v.0.5.6[56] (https://mrcieu.github.io/TwoSampleMR/).
400 The instrumental variables for the exposure were defined as genome-wide significant and
401 independent single nucleotide polymorphisms (SNPs) ($P < 5 \times 10^{-8}$) that were not in LD with other
402 SNPs ($r^2 < 0.001$, with a clumping window of 1 Mb). SNPs in the human major histocompatibility
403 complex (MHC) region at chromosome 6: 28,477,797–33,448,354 (GRCh38) were excluded
404 considering its complex LD structure. We used PLINK v1.9[57]
405 (http://pngu.mgh.harvard.edu/purcell/plink/) to obtain instrumental variables by clumping SNPs
406 using the 1000 Genomes Project European reference panel[58] and applying an LD threshold of r^2
407 < 0.001. The genome-wide significant independent SNPs with the lowest *P*-value were selected
408 from each LD block. Data harmonization was performed using default settings of TwoSampleMR,
409 including removal of palindromic SNPs that have minor allele frequency above 0.42. A Bonferroni
410 correction was used to set a statistical significance threshold, by dividing 0.05 by the number of
411 proteins ($P = 1.0 \times 10^{-5}$ (0.05/4907)). We note that this correction is overly-conservative since
412 many proteins are non-independent. For evaluating horizontal pleiotropy, we used the MR-Egger
413 intercept test; where horizontal pleiotropy was considered to be present when the MR-Egger
414 intercept differed from the null ($P < 0.05$). For NPNT and HSD17B14, we used MR-Egger,
415 weighted median, and weighted mode methods as additional sensitivity analyses to evaluate
416 directional consistency of beta coefficients with the inverse variance weighted method.
417
418 For reverse MR, in which the effects of plasma protein levels on BMI were assessed, we
419 used the cis-pQTLs from the deCODE study as the exposure and BMI GWAS as the outcome,
420 deploying the inverse variance weighted method or Wald ratio method when only one SNP was

available as an instrumental variable. After the cis-pQTL search using data from the deCODE study and data harmonization, 332 proteins were tested in MR for their estimated reverse effects. Results with $P = 1.51\times 10^{-4}$ (0.05/332; Bonferroni correction) were considered statistically significant. We used BMI GWAS from the UK Biobank (not the meta-analysis GWAS) because multiple variants, including cis-pQTL for NPNT from the deCODE study (rs34712979), were dropped during the stringent quality control process of the meta-analysis (e.g., rs34712979 was not present in the GIANT GWAS and thus dropped during the meta-analysis process).

Step 2: Identifying the effects of BMI-driven proteins on COVID-19 severity

Two-sample MR

Cis-pQTL GWAS: We identified cis-pQTLs from the GWAS of 4,907 proteins in 35,559 individuals of European ancestry from the deCODE study[26]. Cis-pQTLs were defined as pQTLs that were located within 1-Mb (± 500 kb) of the transcription start site of the protein-coding gene.

COVID-19 outcome GWAS: For COVID-19 outcomes, we used a GWAS meta-analysis from the COVID-19 Host Genetics Initiatives data release 7 (https://www.covid19hg.org/). The outcomes included severe COVID-19 (13,769 cases and 1,072,442 controls), COVID-19 hospitalization (32,519 cases and 2,062,805 controls), and COVID-19 susceptibility (1,22,616 cases and 2,475,240 controls). Severe COVID-19 was defined as a requirement for respiratory support among hospitalized individuals with laboratory-confirmed SARS-CoV-2 infection or death due to COVID-19. COVID-19 hospitalization was defined as a laboratory-confirmed SARS-CoV-2 infection that required hospitalization. COVID-19 susceptibility was defined as SARS-CoV-2 infection determined by laboratory testing (nucleic acid amplification or serology tests), physicians' diagnosis on electrical medical records, or self-report questionnaires.

Two-sample MR: Using the data above, we undertook two-sample MR for the effects of plasma protein levels on three COVID-19 outcomes. Results with a $P < 1.49 \times 10^{-4}$ (0.05/336; Bonferroni correction with the number of proteins tested in the Step 2 MR) were considered statistically significant. Given that certain proteins had only one genetic instrumental variable, for which the MR-Egger intercept test cannot be performed, we used the PhenoScanner database to test whether variants had potential pleiotropic associations with other diseases or traits. Associations with $P < 5 \times 10^{-8}$ were considered statistically significant.

454 We assessed reverse causation, wherein the effect of COVID-19 may influence plasma
455 protein levels, using data hosted on a COVID-19 biobank (the BQC19 cohort) because MR
456 requires the outcome GWAS to be performed in populations potentially exposed to the exposure
457 of interest. Toward this, we used GWAS of plasma protein levels in 487 SARS-CoV-2-infected
458 individuals from the BQC19 cohort (described below). Results with $P < 1.49 \times 10^{-4}$ (0.05/336;
459 Bonferroni correction with the number of BMI-driven proteins evaluated) were considered
460 statistically significant.
461
462 **The BQC-19 cohort:** BQC19 is a province-wide biobank that provides global access to important
463 biological and clinical data from patients with COVID-19 as well as control subjects in Québec,
464 Canada (https://www.bqc19.ca/). Blood samples were collected in acid citrate dextrose (ACD)
465 tubes from 264 SARS-CoV-2 infectious and 463 non-infectious patients of European ancestry
466 (see below for the detailed definition of infectious or non-infectious state). The samples underwent
467 proteomic profiling on the SomaScan v4 assay. Up to 5,284 aptamers were measured for these
468 samples. After removing any aptamers that represented non-human proteins or controls, we
469 retained 4,907 aptamers for analysis, consistent with the deCODE study[26]. SomaLogic performed
470 normalization and calibration steps to remove systematic biases, which are detailed in their
471 technical note (https://www.mcgill.ca/genepi/files/genepi/bqc19_jgh_prt_tech_note_0.pdf). Blood
472 samples were sent to SomaLogic at two-time points during the pandemic, providing two batches
473 of proteomic measurements. For quality control, protein levels were natural log-transformed and
474 then batch corrected using the ComBat function implemented in the sva R package[59], which uses
475 an empirical Bayesian framework to perform batch effect removal.
476
477 **Definition of SARS-CoV-2 infectious or non-infectious state in the BQC-19 cohort:** We
478 defined infectious and non-infectious state as follows (all date ranges are inclusive): infectious
479 samples were defined as blood samples collected from individuals who tested positive for SARS-
480 CoV-2 test from 7 days before and up to and including 14 days after the first date of SARS-CoV-
481 2-associated symptoms. Non-infectious samples were defined as those meeting either of the
482 following three criteria: (1) Samples were collected from individuals who tested negative for
483 SARS-CoV-2; (2) samples were collected within 31 days after but 90 days before the first date of
484 symptoms from patients whose positive SARS-CoV-2 test was confirmed within 7 days from the
485 first symptom onset; (3) samples were collected at least 15 days before or at least 31 days after
486 the positive SARS-CoV-2 test from patients whose positive SARS-CoV-2 test was confirmed at

487 least 7 days before or after the first symptom onset. Further details can be found at
488 https://github.com/richardslab/BQC19_phenotypeQC/blob/main/src/COVID-19%20Omics.svg.
489
490 **GWAS of plasma protein levels in SARS-CoV-2 infectious individuals from the BQC19**
491 **cohort:** GWAS of plasma protein levels in SARS-CoV-2 infectious individuals of European
492 ancestry were performed using fastGWA from the GCTA package
493 (https://yanglab.westlake.edu.cn/software/gcta/)[60] after rank-based inverse normal
494 transformation of protein levels and adjusting for the first 10 principal components, batch, age,
495 and sex as well as restricting genetic variants to those with MAF > 0.05. Genotyping was
496 performed with Axiom Precision Medicine genotyping array. Genotypic imputation was performed
497 using the TOPMed Imputation Reference panel[61] while retaining only SNPs with MAF > 0.01 and
498 INFO score > 0.6.
499
500 **Colocalization analysis of cis-pQTL for NPNT and HSD17B14 and COVID outcomes**
501 To identify potential bias caused by confounding owing to LD, we performed colocalization to
502 evaluate whether cis-pQTLs shared a single causal variant between the three COVID-19
503 outcomes for the causal proteins and NPNT or HSD17B14. We used the coloc R package v5.1.0[62]
504 (https://chr1swallace.github.io/coloc/) to evaluate all SNPs in the 1-Mb region (± 500 kb) around
505 the top cis-pQTL for the NPNT (defined as the cis-pQTL with the smallest *P*-value). The posterior
506 probability of hypothesis 4 or PP_{shared} (two traits sharing a single causal variant) > 0.8 was
507 considered to indicate strong evidence of colocalization. Colocalization analyses were conducted
508 using the coloc R package v5.1.0[62], using default priors of $p1 = p2 = 10^{-4}$ and $p12 = 10^{-5}$ where
509 $p1$ is a prior probability that only trait 1 has a genetic association in the region, $p2$ is a prior
510 probability that only trait 2 has a genetic association in the region, and $p12$ is a prior probability
511 that both trait 1 and trait 2 share the same genetic association in the region.
512
513 **Fine-mapping of the NPNT region for severe COVID-19 and hospitalization GWASs**
514 The COVID-19 Host Genetics Initiatives release 7 data for those of European ancestry with
515 severe COVID-19 and hospitalization were inputted into FINEMAP v1.4[63]
516 (http://www.christianbenner.com/) using default options. We assumed a maximum of one causal
517 SNP given a single GWAS significant SNP in the NPNT region. Summary statistics were filtered
518 for SNPs present in European ancestry, as done by Huffman *et al*[64]. We set prior probabilities
519 using scaled CADD-PHRED scores[37] (https://cadd.gs.washington.edu/). The scaled CADD-
520 PHRED scores were normalized so their sum equaled one. The variance of effect size prior was

521 set such that the maximum odds ratio a variant can have with 95% probability is two, then scaled
522 using the case-control ratio as described previously[64].
523
524 ## Validation analyses using cis-pQTL for NPNT and HSD17B14 from different cohorts
525 We also obtained cis-pQTLs and their beta estimates for plasma protein levels of NPNT and
526 HSD17B14 from the FENLAND study and AGES Reykjavik study. We repeated the two-sample
527 MR using these cis-pQTLs as instrumental variables and the three COVID-19 outcomes as
528 described above.
529
530 ## Observational evaluation of associations between BMI and NPNT, and NPNT and COVID-
531 19
532 **Observational evaluation of associations between BMI and plasma NPNT levels in 2737**
533 **individuals from the INTERVAL study**
534 The INTERVAL study was a prospective cohort study conducted in England. The study recruited
535 ~50,000 participants of mostly European ancestry without a self-reported history of major disease.
536 The recruitment took place between 2012 and 2014, before the COVID-19 pandemic. The study
537 measured 3622 plasma protein levels in 2737 individuals using the SomaScan assay. The study
538 performed a linear regression to evaluate associations between BMI and plasma protein levels
539 adjusting for age and sex. The derived beta estimates represented normalized SD-unit difference
540 in each protein level per one SD (4.8 kg/m^2) increase in BMI. Further details of the study have
541 been described in detail previously[6].
542
543 **Observational evaluation of associations between plasma NPNT levels in the non-**
544 **infectious state and COVID-19 severity in 463 individuals from the BQC19 cohort**
545 Given that the MR analyses used plasma protein levels in a non-infectious state as the exposures,
546 we observationally assessed an association between plasma NPNT and risk of COVID-19
547 outcomes in the BQC19 cohort. Non-infectious individuals were defined above. We defined the
548 COVID-19 outcomes in accordance with those for the GWAS from COVID-19 Host Genetics
549 Initiative. (I) Severe COVID-19: cases were defined as laboratory-confirmed SARS-CoV-2
550 infection by PCR or serology testing combined with the requirement for respiratory support or
551 death. Controls were defined as those who did not meet the criteria. (II) Hospitalization: cases
552 were defined as those who were hospitalized due to COVID-19-related symptoms. Those who
553 did not fulfill the case definition were defined as controls. (III) Susceptibility: we define cases as

554 those with SARS-CoV-2 infection confirmed in the laboratory testing. Those who tested negative
555 were used as controls.
556 We performed logistic regression analysis using one of the COVID-19 outcomes as the
557 dependent variable and standardized plasma NPNT levels as the independent variable while
558 adjusting for sex, age, and batch number.
559
560 **Mediation analysis of NPNT on the effect of obesity on COVID-19 severity**
561 We undertook mediation analysis to calculate the proportion of BMI's effect on severe COVID-19
562 mediated by NPNT using network MR (**Extended Figure 2**). We used the product of coefficient
563 method to estimate the NPNT-mediated effect (i.e., the effect of BMI on severe COVID-19 that
564 was accounted for by NPNT): First, we estimated the effect of BMI on NPNT, then multiplied this
565 with the effect of NPNT on severe COVID-19. Subsequently, the proportion of the total effect of
566 BMI on COVID-19 mediated by NPNT was estimated by dividing the NPNT-mediated effect ($\beta_{\text{NPNT-}}$
567 $_{\text{to-severe COVID-19}}$) by the total effect ($\beta_{\text{BMI-to-NPNT}}$), as described previously[65,66].
568
569 For this mediation analysis, we used the BMI GWAS from the UK Biobank
570 (https://gwas.mrcieu.ac.uk/datasets/ukb-b-19953/), which was less powered than the meta-
571 analysis of GIANT consortium and UK Biobank (sample size: 681,275 vs. 456,426, respectively).
572 This was because the lead cis-pQTL (rs34712979) was present in UK Biobank but absent in the
573 GIANT GWAS and their meta-analysis GWAS. Given that multivariable MR requires at least one
574 genetic variant shared among the exposures, multivariable MR using BMI and NPNT as the
575 exposures was possible only with the BMI GWAS from UK Biobank, which contains the two cis-
576 pQTLs from the deCODE study, where rs34712979 is the lead cis-pQTL.
577
578
579 **Follow-up analyses for NPNT**
580 **Colocalization analysis of cis-pQTL with eQTL and sQTL from GTEx**
581 **sQTL and eQTL GWAS:** We used sQTL and eQTL GWASs of the lung as well as those of thyroid
582 and arteries (i.e., the top three *NPNT*-expressing tissues) in European-ancestry individuals from
583 the GTEx Portal V8 dataset[41] (https://gtexportal.org/).
584 **Colocalization analyses:** To evaluate whether cis-pQTL is more affected by sQTL, rather than
585 total expression, we also undertook colocalization of the cis-pQTL with eQTLs and sQTLs of
586 NPNT using the GTEx V8 dataset. Colocalization analyses were performed in the 1 Mb region
587 around the cis-pQTL for NPNT (rs34712979) using the coloc v5.1.0[62] with default priors of $p1$ =

588 $p2 = 10^{-4}$ and $p12 = 10^{-5}$ where $p1$ is a prior probability that only trait 1 has a genetic association
589 in the region, $p2$ is a prior probability that only trait 2 has a genetic association in the region, and
590 $p12$ is a prior probability that both trait 1 and trait 2 share the same genetic association in the
591 region. $PP_{shared} > 0.8$ was considered to indicate strong evidence of colocalization. We also
592 evaluated whether the lead cis-pQTL from the deCODE study (rs34712979), the FENLAND study
593 (rs34712979), and AGES Reykjavik study (rs78213340) were associated with the same splice
594 pattern in the lung using GTEx. For eQTL and sQTL, we considered associations with $P < 1 \times 10^{-}$
595 ^{7} statistically significant, as did the deCODE study[26].
596
597
598 **Single-cell RNA-sequencing data analysis**
599 To understand the *NPNT* expression pattern in the lungs of SARS-CoV-2 infected individuals, we
600 obtained single-cell transcriptomic data of SARS-CoV-2-infected lungs by Delorey *et al*[42]. from
601 the Single Cell Portal of the Broad Institute (https://singlecell.broadinstitute.org/single_cell/)
602 (Accession ID: SCP1052). The data contained 106,792 single cells from the lungs of 16 autopsy
603 donors aged 30 to older than 89 years who died due to COVID-19. For data processing, briefly,
604 the study aligned FASTQ files to the reference genome of Homo sapiens, and SARS-CoV-2 and
605 ambient RNA were removed using CellBender. Quality control was performed under the Cumulus
606 single-cell analysis workflow. Cell types were assigned using an automatic prediction approach,
607 followed by manually curating marker genes and gene signatures[42].
608
609 For NPNT expression analysis**,** we reanalyzed the data focusing on NPNT expression
610 status. We analyzed the gene expression matrix and the associated metadata using R v4.1.2
611 and Seurat R package v4.1.1[67] (https://satijalab.org/seurat/). For visualization, NPNT
612 expression levels were represented on a log-transcript per 10 thousand + 1, i.e., log (TP10K+1)
613 scale. To cluster the cells, we adopted the clustering annotation from the original study[42]. To
614 test whether *NPNT* expression was enriched in a cell type, we calculated the proportion of
615 *NPNT*-expression cells in this cell type. Subsequently, we permuted the cell type labels 1,000
616 times, and obtained the frequency (permutation p-value) of the same cell type containing the
617 same or a larger proportion of *NPNT*-expression cells. Additionally, we compared the *NPNT*
618 expression level between a target cell type and the other cell types using Wilcoxon rank-sum
619 test. This enrichment analysis was performed separately in SARS-CoV-2-infected and
620 uninfected cells.
621

Multivariable MR analysis using body fat and fat-free mass

Body fat and fat-free mass GWAS: We obtained GWAS of body fat and fat-free mass from UK Biobank using the IEU OpenGWAS project (https://gwas.mrcieu.ac.uk/). Accession ID for each GWAS was 'ukb-b-19393' and 'ukb-b-13354', respectively.

NPNT GWAS: For GWAS of plasma NPNT levels, we used the same GWAS from the deCODE study[26] as the one used in Step 1 MR, which measured plasma NPNT levels in 35,559 individuals of European ancestry by using the SomaLogic SomaScan assay v4.

Multivariable MR: For instrumental variables, we identified genome-wide significant and independent SNPs for body fat and fat-free mass using the same criteria as in the Step1 MR (i.e., $P < 5 \times 10^{-8}$ for significance and $r^2 < 0.001$ with a clumping window of 1 Mb for independence). SNPs in the MHC region were excluded. After data harmonization, we undertook multivariable MR with the inverse variance weighted method using body fat and fat-free mass as the exposures and plasma NPNT levels as the outcome. Results with $P < 0.025$ (0.05/2; Bonferroni correction) were considered statistically significant. LD clumping was performed using PLINK v1.9[57]. Data harmonization and multivariable MR were performed using TwoSampleMR v.0.5.6[56].

Ethical approval

For summary-level data, all contributing cohorts obtained ethical approval from their intuitional ethics review boards. The contributing cohorts include: UK Biobank, GIANT consortium, deCODE study, FENLAND study, AGES Reykjavik study, INTERVAL study, COVID-19 Host Genetics Initiative, and BQC19. For individual-level data in BQC19, BQC19 received ethical approval from the Jewish General Hospital research ethics board (2020-2137) and the Centre hospitalier de l'Université de Montréal institutional ethics board (MP-02-2020-8929, 19.389).

Data availability

·GWAS summary statistics for each trait are available as follows:

BMI (https://portals.broadinstitute.org/collaboration/giant/),

Plasma proteome from the deCODE study (https://www.decode.com/summarydata/),

COVID-19 outcomes (https://www.covid19hg.org/results/r7/),

GTEx Portal V8 (https://gtexportal.org/home/datasets/).

Plasma proteome from BQC19 (https://www.mcgill.ca/genepi/mcg-covid-19-biobank). Access to the data of BQC19 can be obtained upon approval of requests.

656 ·Cis-pQTLs of each study are available in the corresponding publications' supplementary
657 materials[26-28].
658 Body fat and fat-free mass GWASs are available at IEU OpenGWAS project with Accession ID of
659 ukb-b-19393 and ukb-b-13354, respectively (https://gwas.mrcieu.ac.uk/).
660 ·Single-cell RNA-sequencing data of COVID-19 lung autopsy samples are available at the Single
661 Cell Portal under the Accession ID of SCP1052 (https://singlecell.broadinstitute.org/single_cell/).
662 ·CADD-scores v.1.6 can be accessed at https://cadd.gs.washington.edu/score.
663 ·Genotype data from 1000G genomes project is available at
664 https://www.internationalgenome.org/data.
665

666 **Acknowledgments**

667 We thank the COVID-19 Host Genetics Initiative for providing the latest summary statistics for
668 COVID-19 outcomes. The Richards research group is supported by the Canadian Institutes of
669 Health Research (CIHR: 365825; 409511, 100558, 169303), the McGill Interdisciplinary Initiative
670 in Infection and Immunity (MI4), the Lady Davis Institute of the Jewish General Hospital, the
671 Jewish General Hospital Foundation, the Canadian Foundation for Innovation, the NIH
672 Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada, McGill
673 University, Cancer Research UK [grant number C18281/A29019] and the Fonds de Recherche
674 Québec Santé (FRQS). JBR is supported by an FRQS Mérite Clinical Research Scholarship.
675 Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the
676 Welcome Trust, Medical Research Council, European Union, the National Institute for Health
677 Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research
678 Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College
679 London. S.Y. is supported by the Japan Society for the Promotion of Science. T.L. has been
680 supported by a Vanier Canada Graduate Scholarship, an FRQS doctoral training fellowship, and
681 a McGill University Faculty of Medicine Studentship. These funding agencies above had no role
682 in the design, implementation, or interpretation of this study.
683

684 **Author contributions**

685 Conception and design: S.Y. and J.B.R. Methodology: S.Y., T.L., and J.B.R. Data analysis: S.Y.,
686 T.L., J.D.S.W., C.Y.S., and J.B.R. Visualization: S.Y. and T.L. Writing – original draft: S.Y. Writing
687 – Review and editing: S.Y., G.B.L., T.L., J.D.S.W., C.Y.S., T.N., D.M., Y.C., K.L., M.H., Z.A., S.L.,
688 N.D., C.D., M.V., C.T., X.X., M.B., F.S., L.L., H.M.M., M.A., J.A., V.M., N.J.T., H.Z., S.Z., V.F.,
689 Y.F., and J.B.R.

690 ## Disclosures

691 J.B.R. has served as an advisor to GlaxoSmithKline and Deerfield Capital. J.B.R.'s institution has
692 received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline, and Biogen for
693 projects unrelated to this research. He is the CEO of 5 Prime Sciences
694 (www.5primesciences.com), which provides research services for biotech, pharma, and venture
695 capital companies for projects unrelated to this research. T.L., V.F., and Y.F are employees of 5
696 Prime Sciences.

697 ## References

698 1 Johns Hopkins University. COVID-19 global map. https://coronavirus.jhu.edu/map.html.
699 Accessed on June 3, 2022.
700 2 COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19.
701 *Nature* **600**, 472-477 (2021).
702 3 Stefan, N., Birkenfeld, A. L., Schulze, M. B. & Ludwig, D. S. Obesity and impaired
703 metabolic health in patients with COVID-19. *Nat Rev Endocrinol* **16**, 341-342 (2020).
704 4 Foulkes, A. S. *et al.* Understanding the link between obesity and severe COVID-19
705 outcomes: Causal mediation by systemic inflammatory response. *J Clin Endocrinol Metab*,
706 e698–e707, https://doi.org/10.1210/clinem/dgab629 (2022).
707 5 Zickler, M. *et al.* Replication of SARS-CoV-2 in adipose tissue determines organ and
708 systemic lipid metabolism in hamsters and humans. *Cell Metab* **34**, 1-2 (2022).
709 6 Goudswaard, L. J. *et al.* Effects of adiposity on the human plasma proteome: observational
710 and Mendelian randomisation estimates. *Int J Obes (Lond)* **45**, 2221-2229 (2021).
711 7 Zaghlool, S. B. *et al.* Revealing the role of the human blood plasma proteome in obesity
712 using genetic drivers. *Nat Commun* **12**, 1279, https://doi.org/10.1038/s41467-021-21542-
713 4 (2021).
714 8 Filbin, M. R. *et al.* Longitudinal proteomic analysis of severe COVID-19 reveals survival-
715 associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep Med*
716 **2**, 100287, https://doi.org/10.1016/j.xcrm.2021.100287 (2021).
717 9 Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in
718 Epidemiology Using Mendelian Randomization. *JAMA* **326**, 1614 (2021).
719 10 Skrivankova, V. W. *et al.* Strengthening the reporting of observational studies in
720 epidemiology using mendelian randomisation (STROBE-MR): explanation and
721 elaboration. *BMJ* **375**, n2233, https://doi.org/10.1136/bmj.n2233 (2021).
722 11 Ponsford, M. J. *et al.* Cardiometabolic Traits, Sepsis, and Severe COVID-19: A Mendelian
723 Randomization Investigation. *Circulation* **142**, 1791-1793, (2020).
724 12 Luo, S., Liang, Y., Wong, T. H. T., Schooling, C. M. & Au Yeung, S. L. Identifying factors
725 contributing to increased susceptibility to COVID-19 risk: a systematic review of Mendelian
726 randomization studies. *Int J Epidemiol,* dyac076, https://doi.org/10.1093/ije/dyac076
727 (2022).
728 13 Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry
729 against COVID-19 susceptibility and severity. *Nat Med* **27**, 659-667, (2021).

14 Gaziano, L. *et al.* Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med* **27**, 668-676, (2021).

15 Bovijn, J., Lindgren, C. M. & Holmes, M. V. Genetic variants mimicking therapeutic inhibition of IL-6 receptor signaling and risk of COVID-19. *Lancet Rheumatol* **2**, e658-e659, https://doi.org/10.1016/s2665-9913(20)30345-3 (2020).

16 Pietzner, M. *et al.* ELF5 is a respiratory epithelial cell-specific risk gene for severe COVID-19. *medRxiv*, https://doi.org/10.1101/2022.01.17.22269283 (2022).

17 Klaric, L. *et al.* Mendelian randomisation identifies alternative splicing of the FAS death receptor as a mediator of severe COVID-19. *medRxiv*, https://doi.org/10.1101/2021.04.01.21254789 (2021).

18 Niemi, M. E. K., Daly, M. J. & Ganna, A. The human genetic epidemiology of COVID-19. *Nat Rev Genet*, https://doi.org/10.1038/s41576-022-00478-5 (2022).

19 Recovery Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet* **397**, 1637-1645 (2021).

20 Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M. & Davey Smith, G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat Rev Cardiol* **18**, 435-453 (2021).

21 Manousaki, D., Mokry, L. E., Ross, S., Goltzman, D. & Richards, J. B. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. *Circ. Cardiovasc Genet* **9**, 349-356 (2016).

22 Jiang, X. *et al.* Circulating vitamin D concentrations and risk of breast and prostate cancer: a Mendelian randomization study. *Int J Epidemiol* **48**, 1416-1424 (2019).

23 Meng, X. *et al.* Phenome-wide Mendelian-randomization study of genetically determined vitamin D on multiple health outcomes using the UK Biobank study. *Int J Epidemiol* **48**, 1425-1434 (2019).

24 Manson, J. E. *et al.* Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. *N Engl J Med* **380**, 33-44 (2019).

25 Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Mol Genet* **27**, 3641-3649 (2018).

26 Ferkingstad, E. *et al.* DECODE: Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* **53**, 1712-1721 (2021).

27 Pietzner, M. *et al.* Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun* **12**, 6822, https://doi.org/10.1038/s41467-021-27164-0 (2021).

28 Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769-773 (2018).

29 Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-525 (2015).

30 Swerdlow, D. I. *et al.* Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* **45**, 1600-1616 (2016).

773 31 Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African ancestry
774 identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet*,
775 https://doi.org/10.1038/s41588-022-01051-w (2022).
776 32 Staley, J. R. *et al.* PhenoScanner: a database of human genotype-phenotype associations.
777 *Bioinformatics* **32**, 3207-3209 (2016).
778 33 Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct
779 localization and relevance to fibrosis. *Nat Commun* **11**, 1920,
780 https://doi.org/10.1038/s41467-020-15647-5 (2020).
781 34 Xie, T. *et al.* Single-Cell Deconvolution of Fibroblast Heterogeneity in Mouse Pulmonary
782 Fibrosis. *Cell Rep* **22**, 3625-3640 (2018).
783 35 Obeidat, M. e. *et al.* Molecular mechanisms underlying variations in lung function: a
784 systems genetics analysis. *Lancet Respir Med* **3**, 782-795 (2015).
785 36 Bhatt, S. P. *et al.* Discriminative Accuracy of FEV1:FVC Thresholds for COPD-Related
786 Hospitalization and Mortality. *JAMA* **321**, 2438-2447,
787 https://doi.org/10.1001/jama.2019.7233 (2019).
788 37 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human
789 genetic variants. *Nat Genet* **46**, 310-315 (2014).
790 38 Saferali, A. *et al.* Characterization of a COPD-associated NPNT functional splicing genetic
791 variant in human lung tissue via long-read sequencing. *medRxiv*,
792 https://doi.org/10.1101/2020.10.20.20203927 (2020).
793 39 Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic
794 obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**,
795 481-493, https://doi.org/10.1038/s41588-018-0321-7 (2019).
796 40 Sakornsakolpat, P. *et al.* Genetic landscape of chronic obstructive pulmonary disease
797 identifies heterogeneous cell-type and phenotype associations. *Nat Genet* **51**, 494-505
798 (2019).
799 41 GTEx Consortium. The GTEx Consortium atlas of genetic regulatoryeffects across human
800 tissues. *Science* **369**, 1318-1330 (2020).
801 42 Delorey, T. M. *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular
802 targets. *Nature* **595**, 107-113 (2021).
803 43 Brandenberger, R. *et al.* Identification and characterization of a novel extracellular matrix
804 protein nephronectin that is associated with integrin alpha8beta1 in the embryonic kidney.
805 *J Cell Biol* **154**, 447-458 (2001).
806 44 Morimura, N. *et al.* Molecular cloning of POEM: a novel adhesion molecule that interacts
807 with alpha8beta1 integrin. *J Biol Chem* **276**, 42172-42181 (2001).
808 45 Inagaki, F. F. *et al.* Nephronectin is upregulated in acute and chronic hepatitis and
809 aggravates liver injury by recruiting CD4 positive cells. *Biochem Biophys Res Commun*
810 **430**, 751-756 (2013).
811 46 Kon, S., Honda, M., Ishikawa, K., Maeda, M. & Segawa, T. Antibodies against
812 nephronectin ameliorate anti-type II collagen-induced arthritis in mice. *FEBS Open Bio* **10**,
813 107-117 (2020).
814 47 O'Rourke, R. W. & Lumeng, C. N. Pathways to severe COVID-19 for people with obesity.
815 *Obesity (Silver Spring)* **29**, 645-653 (2021).

48 Strunz, M. *et al.* Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* **11**, 3559, https://doi.org/10.1038/s41467-020-17358-3 (2020).
49 Xie, T. *et al.* Mesenchymal growth hormone receptor deficiency leads to failure of alveolar progenitor cell function and severe pulmonary fibrosis. *Sci Adv* **7**, eabg6005, https://doi.org/10.1126/sciadv.abg6005 (2021).
50 Richardson, T. G., Fang, S., Mitchell, R. E., Holmes, M. V. & Davey Smith, G. Evaluating the effects of cardiometabolic exposures on circulating proteins which may contribute to severe SARS-CoV-2. *EBioMedicine* **64**, 103228, https://doi.org/10.1016/j.ebiom.2021.103228 (2021).
51 Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206, https://doi.org/10.1038/nature14177 (2015).
52 Pihlajamaki, J. *et al.* Expression of the splicing factor gene SFRS10 is reduced in human obesity and contributes to enhanced lipogenesis. *Cell Metab* **14**, 208-218 (2011).
53 Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res* **24**, 1403-1419 (2014).
54 Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187-1200 (2017).
55 Kim, M., Lee, C. & Park, J. Extracellular matrix remodeling facilitates obesity-associated cancer progression. *Trends Cell Biol* https://doi.org/10.1016/j.tcb.2022.02.008 (2022).
56 Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408, https://doi.org/10.7554/elife.34408 (2018).
57 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
58 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
59 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
60 Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**, 1749-1755 (2019).
61 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
62 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383, https://doi.org/10.1371/journal.pgen.1004383 (2014).
63 Benner, C. *et al.* FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493-1501 (2016).
64 Huffman, J. E. *et al.* Multi-ancestry fine mapping implicates OAS1 splicing in risk of severe COVID-19. *Nat Genet* **54**, 125-127 (2022).
65 Burgess, S., Daniel, R. M., Butterworth, A. S., Thompson, S. G. & Consortium, E. P.-I. Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol* **44**, 484-495 (2015).

66 Carter, A. R. *et al.* Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. *BMJ* **365**, l1855 (2019).
67 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e29, https://doi.org/10.1016/j.cell.2021.04.048 (2021).

Figure legends

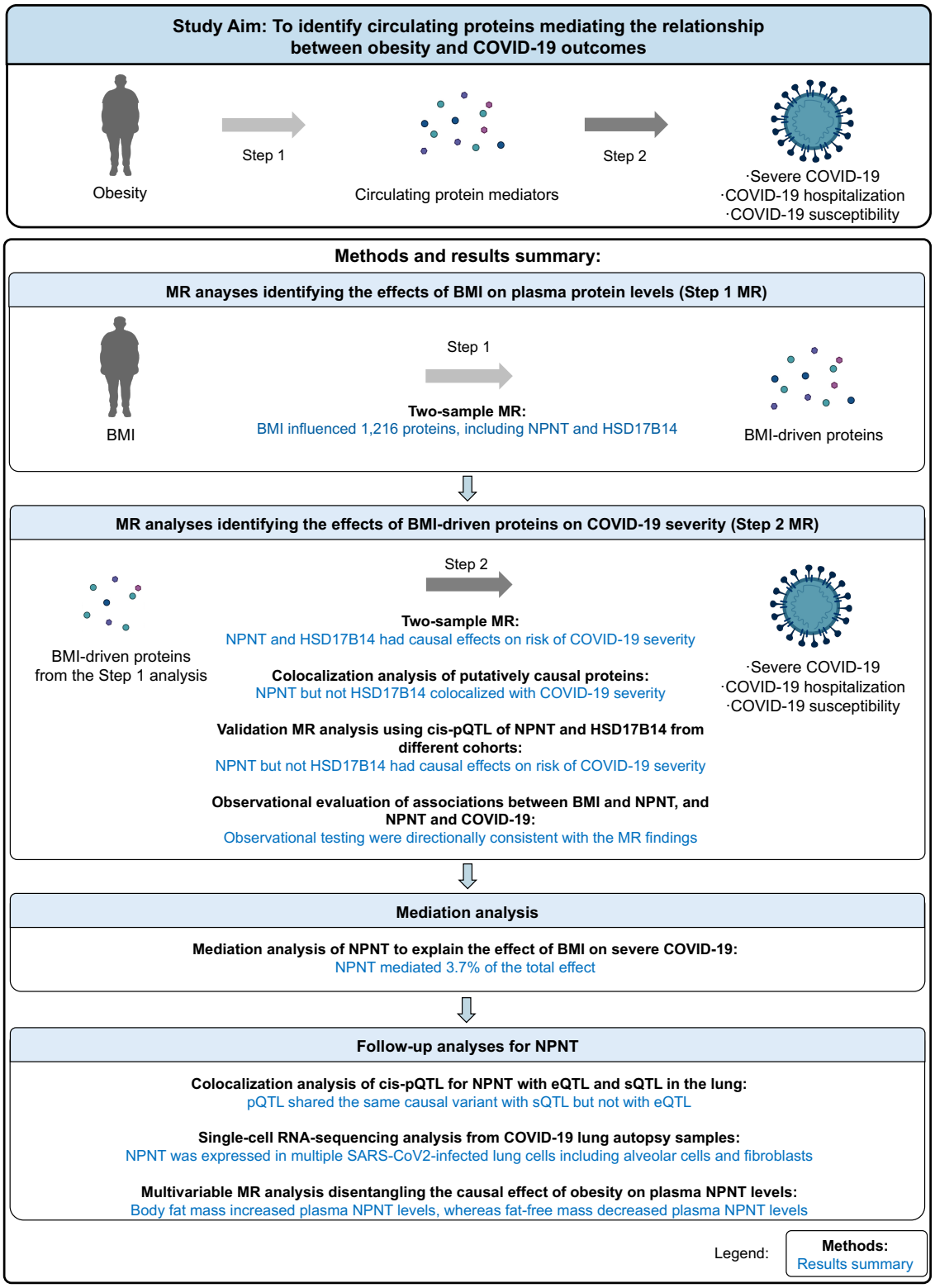


Figure 1. Study overview.

MR: Mendelian randomization, BMI: body mass index, NPNT: nephronectin, HSD17B14: hydroxysteroid 17-beta dehydrogenase 14, cis-pQTL: cis-acting quantitative trait loci, e-QTL: expression quantitative trait loci, sQTL: splicing quantitative trait loci.

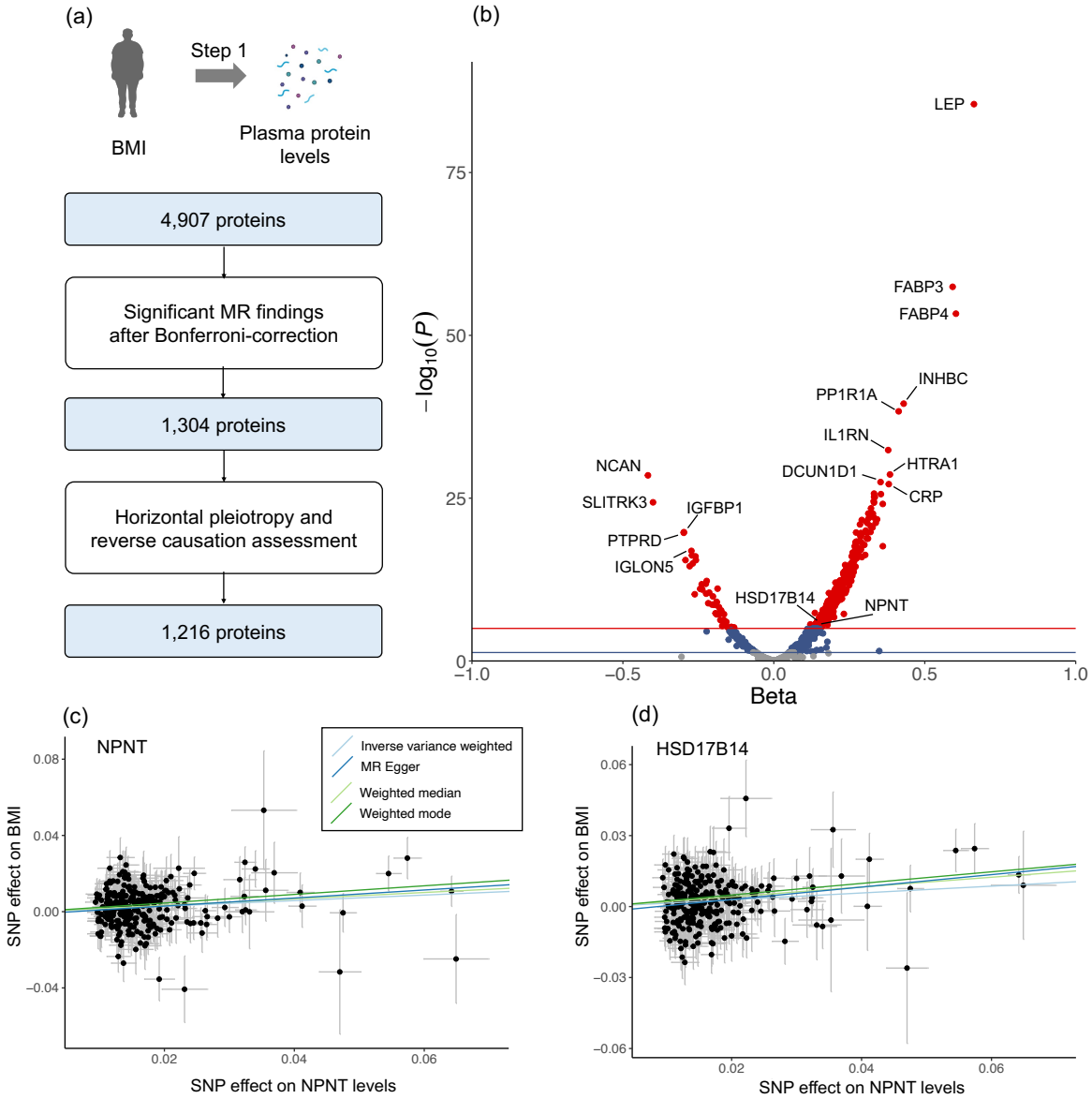


Figure 2. MR analyses for the effect of body mass index on plasma protein levels. (a) Flow diagram of the Step 1 analyses. (b) Volcano plot illustrating the effect of BMI on each plasma protein from the MR analyses. Red and blue vertical lines represent $P = 1.0 \times 10^{-5}$ and 0.05, respectively. (c) MR scatter plot for the effect of BMI on plasma NPNT levels. (d) MR scatter plot for the effect of BMI on plasma HSD17B14 levels. A genetically predicted increase in BMI by one standard deviation was associated with increased levels of NPNT (beta = 0.145, 95% CI: 0.0841–0.206, $P = 3.03 \times 10^{-6}$) and HSD17B14 (beta = 0.14, 95% CI: 0.0848–0.202, $P = 1.70 \times 10^{-6}$) using the inverse variance weighted method. Sensitivity analyses with MR-Egger, weighted median, and weighted mode methods yielded directionally consistent results with the inverse variant weighted method.

MR: Mendelian randomization, BMI: body mass index, NPNT: nephronectin, HSD17B14: hydroxysteroid 17-beta dehydrogenase 14.

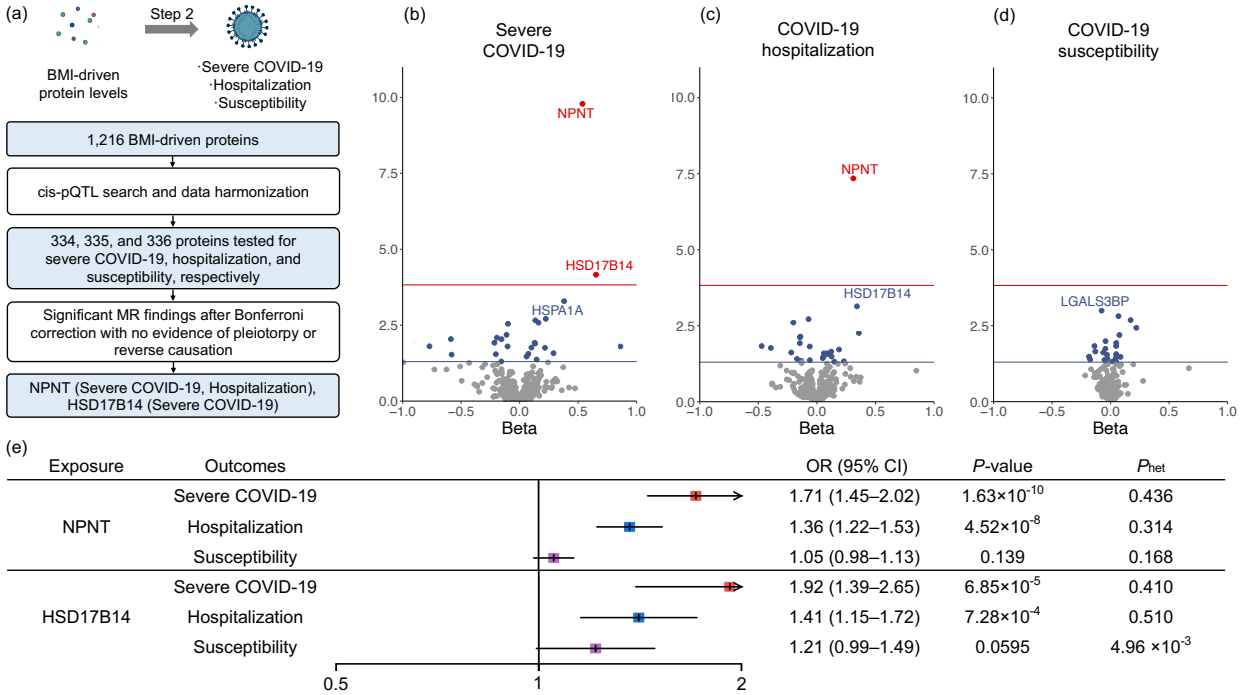


Figure 3. MR analyses of BMI-driven proteins on COVID-19 outcomes.
(a) Flow diagram of the Step 2 analyses. (b, c, d) Volcano plot illustrating the effect of BMI on severe COVID-19 (b), COVID-19 hospitalization (c), and COVID-19 susceptibility (d) from the MR analyses. Red and blue vertical lines represent $P = 1.0 \times 10^{-5}$ and 0.05, respectively. (e) Forest plot showing the MR results for NPNT and HSD17B14.
MR: Mendelian randomization, BMI: body mass index, NPNT: nephronectin, HSD17B14: hydroxysteroid 17-beta dehydrogenase 14, OR: odds ratio, 95% CI: 95% confidence intervals, P_{het} = *P*-value for heterogeneity test.

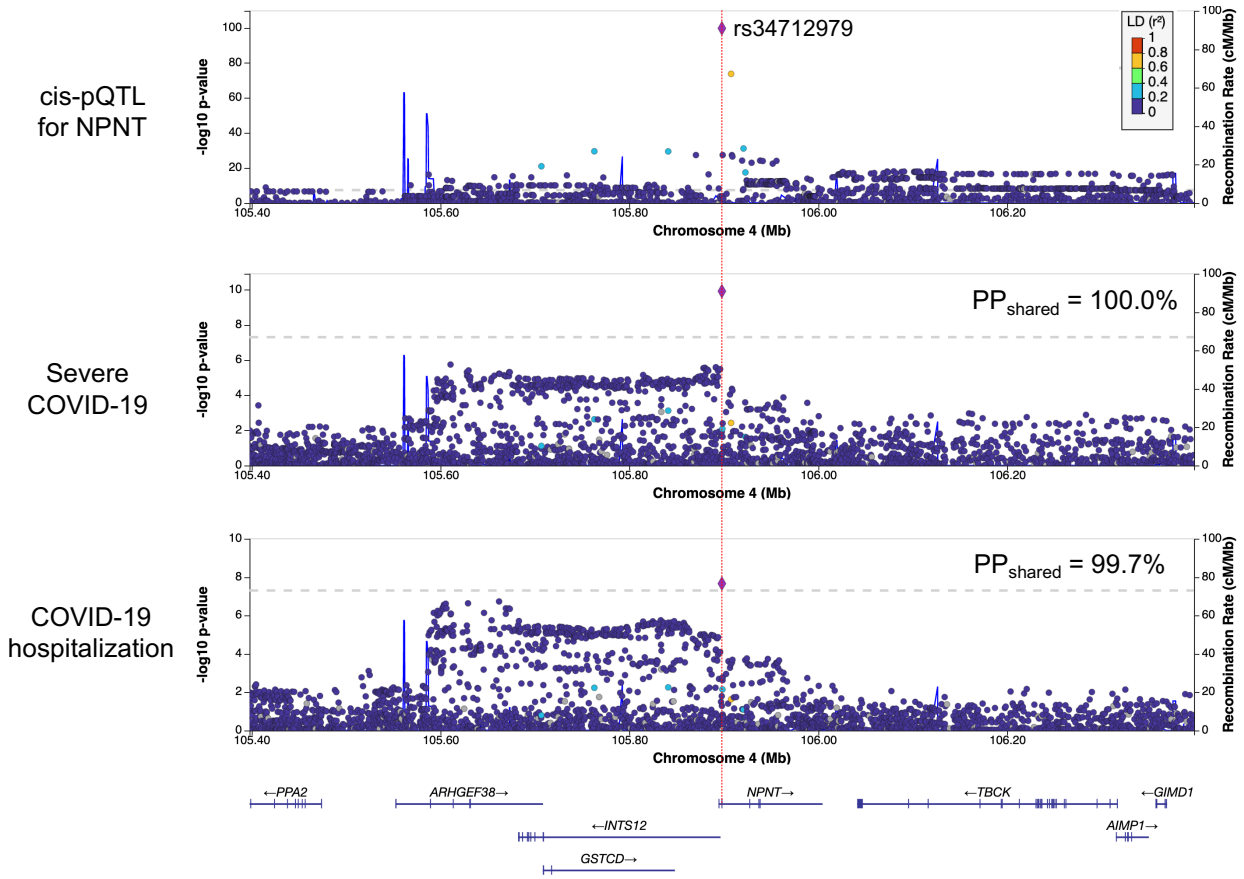


Figure 4. Colocalization analyses of cis-pQTL for NPNT and COVID outcomes in the 1-Mb region around rs34712979.

PP_{shared}: Posterior probability that cis-pQTL for NPNT shares a single causal signal with the COVID-19 outcome.

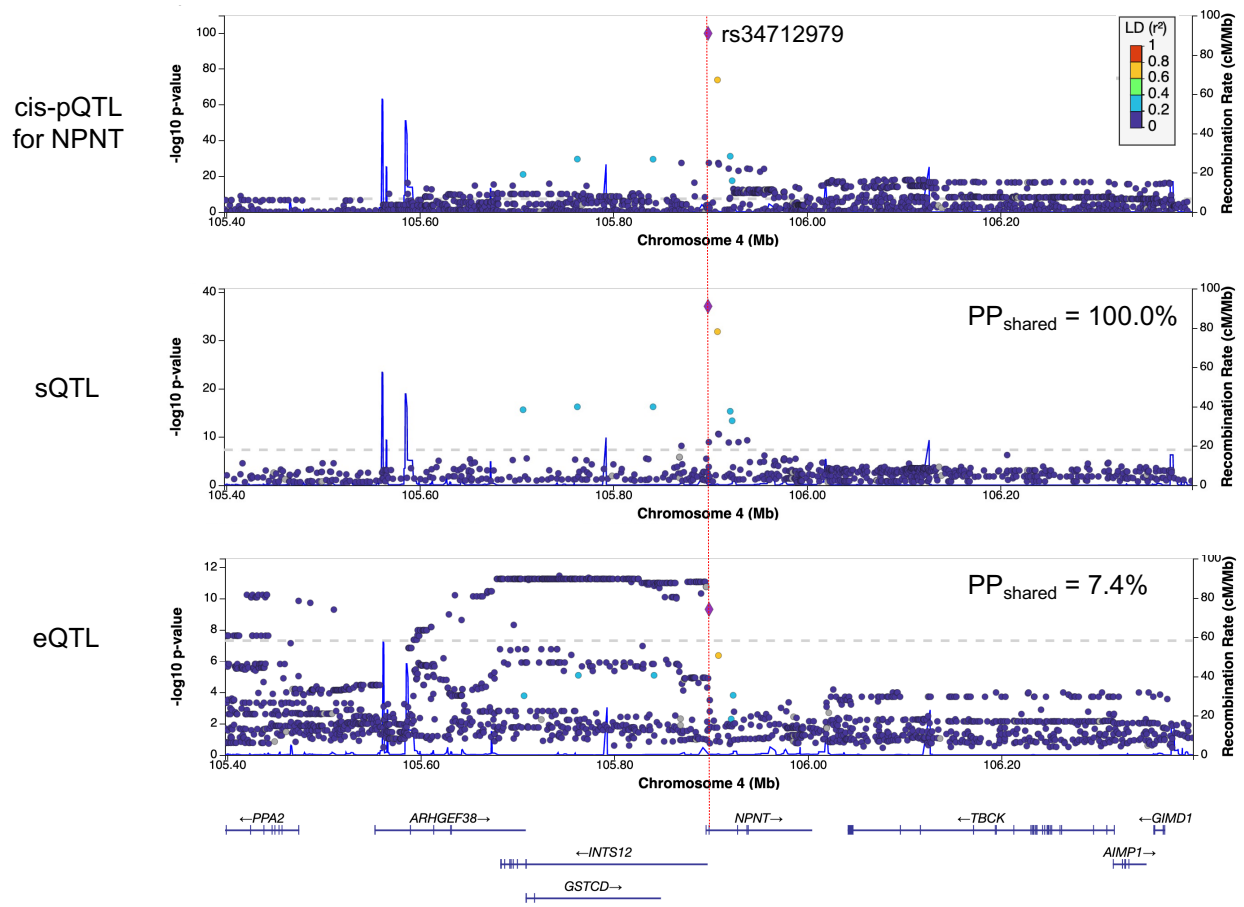


Figure 5. Colocalization analyses of cis-pQTL with sQTL and eQTL for NPNT. PP_{shared}: Posterior probability that cis-pQTL for NPNT shares a single causal signal with its sQTLs or eQTLs in the lung.

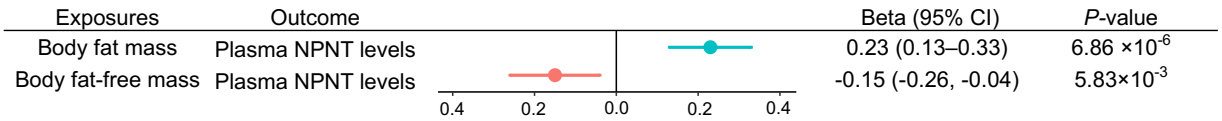
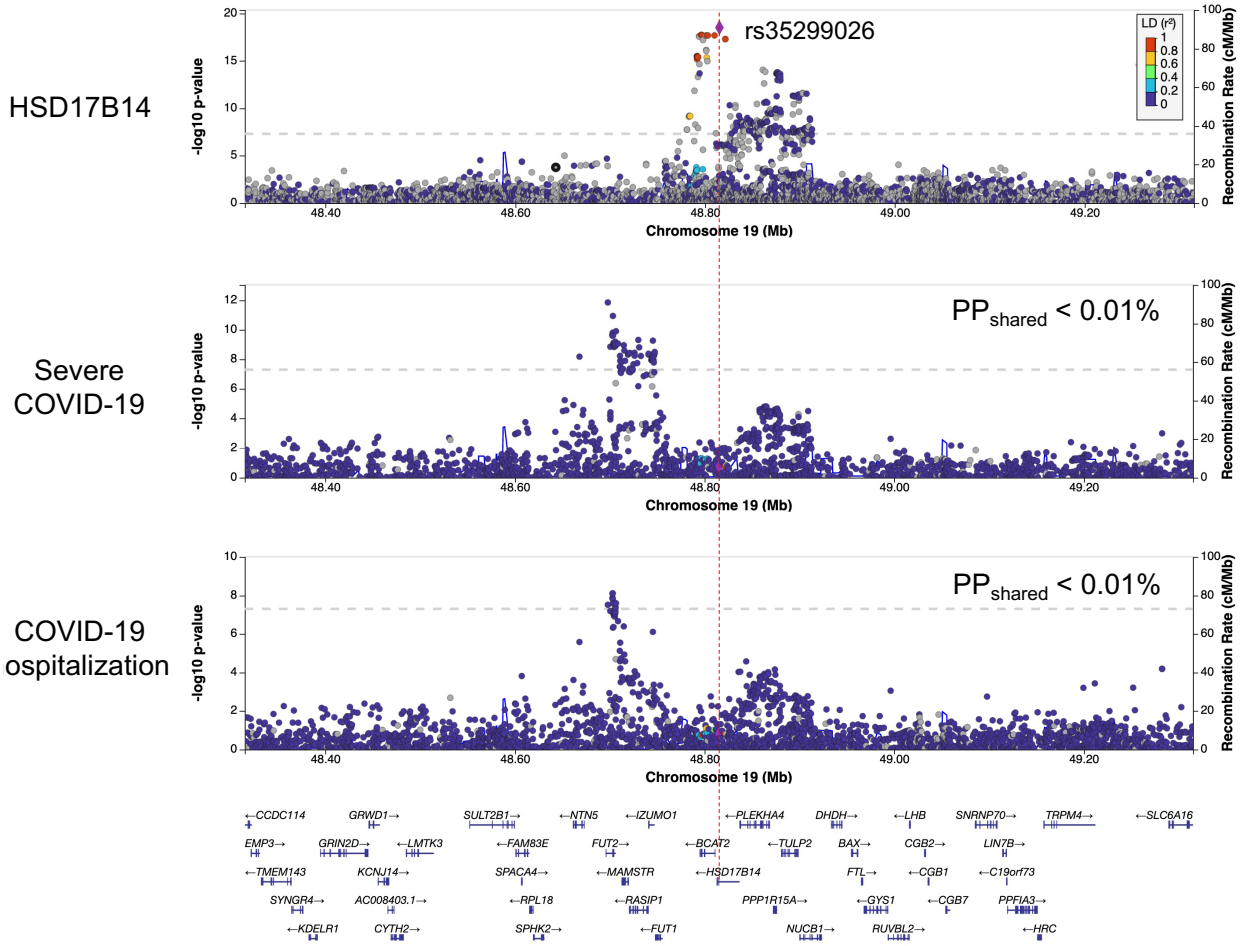


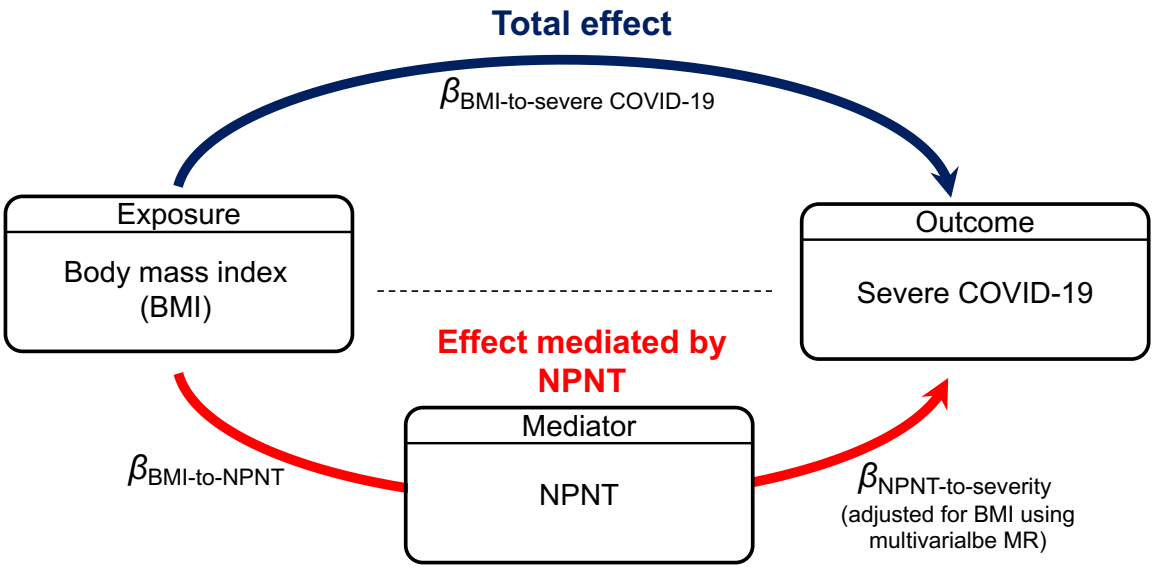
Figure 6. Multivariable MR analysis for evaluating independent effects of body fat and fat-free mass on plasma NPNT levels.

MR: Mendelian randomization; NPNT: nephronectin.



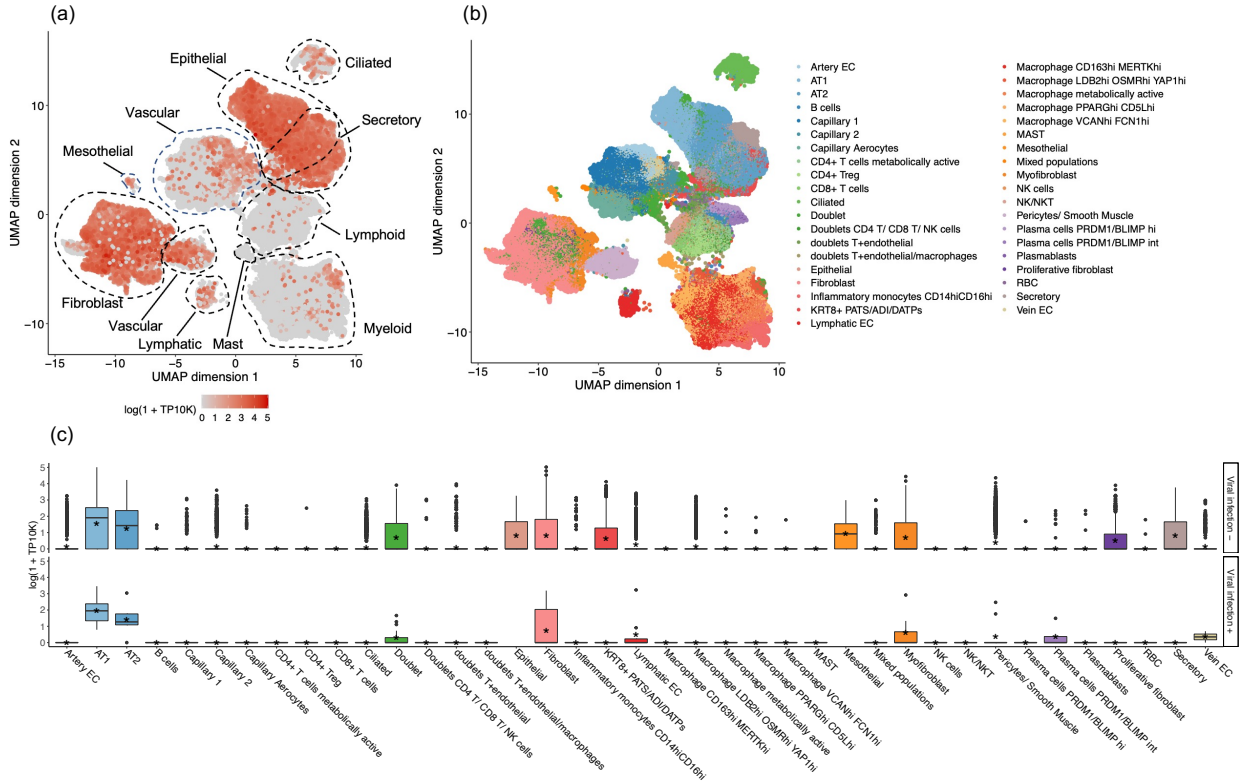
Extended Figure 1. Colocalization analysis of the cis-pQTL for HSD17B14 and the COVID-19 outcomes in the 1Mb region around rs150400299.

PP_{shared}: Posterior probability that the cis-pQTL for NPNT shares a single causal signal with COVID-19 outcomes, 1Mb: one-megabase.



$$\text{The ratio mediated} = \frac{\text{Effect mediated by NPNT}}{\text{Total effect}} = \frac{\beta_{\text{BMI-to-NPNT}} \times \beta_{\text{NPNT-to-severity adjusted for BMI}}}{\beta_{\text{BMI-to-NPNT}}} = 3.7\% \ (95\%\ \text{CI: } 0.7\text{–}6.7\%)$$

Extended Figure 2. MR mediation analysis illustrated by the directed acyclic graph. Dark blue arrow represents the total effect of BMI on severe COVID-19. Red arrow represents the effect of BMI on severe COVID-19 mediated by NPNT. The NPNT-mediated effect can be calculated by multiplying $\beta_{\text{BMI-to-NPNT}}$ and $\beta_{\text{NPNT-to-severity}}$, where $\beta_{\text{BMI-to-NPNT}}$ is the effect of BMI on NPNT and $\beta_{\text{NPNT-to-severity}}$ is the effect of NPNT on severe COVID-19 adjusted for BMI. MR: Mendelian randomization; NPNT: nephronectin.



Extended Figure 3. NPNT expression levels in lung cell types from COVID-19 lung autopsy samples.

(a) NPNT expression levels of each cell type at the single-cell resolution in the 16 lung donors with COVID-19. (b) 28 annotated cell types of the lung. (c) NPNT expression status in SARS-CoV-2 non-infected cells (viral infection −, top panel) or SARS-CoV-2 infected cells (viral infection +, bottom panel). NPNT expression levels of 28 cell types in the two groups are shown in a box plot. In each box, the horizontal line denotes a median value of the expression levels, and the asterisk inside each box denotes the mean value. Each box extends from the 25th to the 75th percentile of each group. Log(TP10K+1) was calculated by normalizing original gene counts by total unique molecular identifiers (UMI) counts, multiplying by 10,000 (TP10K), and then taking the natural logarithm.